Practice Exams and Their Solutions

Based on

A Course in Probability and Statistics

Department of Statistics
University of California, Berkeley
Berkeley, CA 94720-3860

Please email corrections and other comments to stone@stat.berkeley.edu.

Probability (Chapters 1–6)

Practice Exams

First Practice First Midterm Exam

1. Write an essay on variance and standard deviation.

2. Let W have the exponential distribution with mean 1. Explain how W can be used to construct a random variable $Y = g(W)$ such that Y is uniformly distributed on $\{0, 1, 2\}$.

3. Let W have the density function f given by $f(w) = 2/w^3$ for $w > 1$ and $f(w) = 0$ for $w \leq 1$. Set $Y = \alpha + \beta W$, where $\beta > 0$. In terms of α and β, determine

   (a) the distribution function of Y;

   (b) the density function of Y;

   (c) the quantiles of Y;

   (d) the mean of Y;

   (e) the variance of Y.

4. Let Y be a random variable having mean μ and suppose that $E[(Y - \mu)^4] \leq 2$. Use this information to determine a good upper bound to $P(|Y - \mu| \geq 10)$.

5. Let U and V be independent random variables, each uniformly distributed on $[0, 1]$. Set $X = U + V$ and $Y = U - V$. Determine whether or not X and Y are independent.

6. Let U and V be independent random variables, each uniformly distributed on $[0, 1]$. Determine the mean and variance of the random variable $Y = 3U^2 - 2V$.

Second Practice First Midterm Exam

7. Consider the task of giving a 15–20 minute review lecture on the role of *distribution functions* in probability theory, which may include illustrative figures and examples. Write out a complete set of lecture notes that could be used for this purpose by yourself or by another student in the course.

8. Let W have the density function given by $f_W(w) = 2w$ for $0 < w < 1$ and $f_W(w) = 0$ for other values of w. Set $Y = e^W$.

   (a) Determine the distribution function and quantiles of W.

   (b) Determine the distribution function, density function, and quantiles of Y.

   (c) Determine the mean and variance of Y directly from its density function.

   (d) Determine the mean and variance of Y directly from the density function of W.

9. Let W_1 and W_2 be independent discrete random variables, each having the probability function given by $f(0) = \frac{1}{2}$, $f(1) = \frac{1}{3}$, and $f(2) = \frac{1}{6}$. Set $Y = W_1 + W_2$.

   (a) Determine the mean, variance, and standard deviation of Y.

   (b) Use Markov's inequality to determine an upper bound to $P(Y \geq 3)$.

   (c) Use Chebyshev's inequality to determine an upper bound to $P(Y \geq 3)$.

   (d) Determine the exact value of $P(Y \geq 3)$.

Third Practice First Midterm Exam

10. Consider the task of giving a 15–20 minute review lecture on the role of *independence* in that portion of probability theory that is covered in Chapters 1 and 2 of the textbook. Write out a complete set of lecture notes that could be used for this purpose by yourself or by another student in the course.

11. Let $W_1, W_2, \ldots$ be independent random variables having the common density function f given by $f(w) = w^{-2}$ for $w > 1$ and $f(w) = 0$ for $w \leq 1$.

    (a) Determine the common distribution function F of $W_1, W_2, \ldots$.

    Given the positive integer n, let $Y_n = \min(W_1, \ldots, W_n)$ denote the minimum of the random variables $W_1, \ldots, W_n$.

    (b) Determine the distribution function, density function, and pth quantile of Y_n.

    (c) For which values of n does Y_n have finite mean?

    (d) For which values of n does Y_n have finite variance?

12. Let W_1, W_2 and W_3 be independent random variables, each having the uniform distribution on $[0, 1]$.

    (a) Set $Y = W_1 - 3W_2 + 2W_3$. Use Chebyshev's inequality to determine an upper bound to $P(|Y| \geq 2)$.

    (b) Determine the probability function of the random variable

$$Y = \mathrm{ind}\Big(W_1 \geq \frac{1}{2}\Big) + \mathrm{ind}\Big(W_2 \geq \frac{1}{3}\Big) + \mathrm{ind}\Big(W_3 \geq \frac{1}{4}\Big).$$

Fourth Practice First Midterm Exam

13. Consider the following terms: distribution; distribution function; probability function; density function; random variable. Consider also the task of giving a 20 minute review lecture on the these terms, including their definitions or other explanations, their properties, and their relationships with each other, as covered in Chapter 1 of the textbook and in the corresponding lectures. Write out a complete set of lecture notes that could be used for this purpose by yourself or by another student in the course.

14. Let Y be a random variable having the density function f given by $f(y) = y/2$ for $0 < y < 2$ and $f(y) = 0$ otherwise.

    (a) Determine the distribution function of Y.

    (b) Let U be uniformly distributed on $(0, 1)$. Determine an increasing function g on $(0, 1)$ such that $g(U)$ has the same distribution as Y.

    (c) Determine constants a and $b > 0$ such that the random variable $a + bY$ has lower quartile 0 and upper quartile 1.

    (d) Determine the variance of the random variable $a + bY$, where a and b are determined by the solution to (c).

15. A box has 36 balls, numbered from 1 to 36. A ball is selected at random from the box, so that each ball has probability 1/36 of being selected. Let Y denote the number on the randomly selected ball. Let I_1 denote the indicator of the event that $Y \in \{1, \dots, 12\}$; let I_2 denote the indicator of the event that $Y \in \{13, \dots, 24\}$; and let I_3 denote the indicator of the event that $Y \in \{19, \dots, 36\}$.

    (a) Show that the random variables I_1, I_2 and I_3 are NOT independent.

    (b) Determine the mean and variance of $I_1 - 2I_2 + 3I_3$.

First Practice Second Midterm Exam

16. Write an essay on multiple linear prediction.

17. Let Y have the gamma distribution with shape parameter 2 and scale parameter β. Determine the mean and variance of Y^3.

18. The negative binomial distribution with parameters $\alpha > 0$ and $\pi \in (0, 1)$ has the probability function on the nonnegative integers given by

$$f(y) = \frac{\Gamma(\alpha + y)}{\Gamma(\alpha) y\,!}(1 - \pi)^\alpha \pi^y, \qquad y = 0, 1, 2, \dots .$$

    (a) Determine the mode(s) of the probability function.

    (b) Let Y_1 and Y_2 be independent random variables having negative binomial distributions with parameters α_1 and π and α_2 and π, respectively, where $\alpha_1, \alpha_2 > 0$. Show that $Y_1 + Y_2$ has the negative binomial distribution with parameters $\alpha_1 + \alpha_2$ and π. *Hint:* Consider the power series expansion

$$(1 - t)^{-\alpha} = \sum_{x=0}^{\infty} \frac{\Gamma(\alpha + x)}{\Gamma(\alpha) x\,!} t^x, \qquad |t| < 1,$$

    where $\alpha > 0$. By equating coefficients in the identity $(1-t)^{-\alpha_1}(1-t)^{-\alpha_2} = (1 - t)^{-(\alpha_1+\alpha_2)}$, we get the new identity

$$\sum_{x=0}^{y} \frac{\Gamma(\alpha_1 + x)}{\Gamma(\alpha_1) x\,!} \frac{\Gamma(\alpha_2 + y - x)}{\Gamma(\alpha_2)(y - x)!} = \frac{\Gamma(\alpha_1 + \alpha_2 + y)}{\Gamma(\alpha_1 + \alpha_2) y\,!}, \qquad y = 0, 1, 2, \dots,$$

    where $\alpha_1, \alpha_2 > 0$. Use the later identity to get the desired result.

19. Let $\boldsymbol{W}$ have the multivariate normal distribution with mean vector $\boldsymbol{\mu}$ and positive definite $n \times n$ variance-covariance matrix $\boldsymbol{\Sigma}$.

    (a) In terms of $\boldsymbol{\mu}$ and $\boldsymbol{\Sigma}$, determine the density function of $\boldsymbol{Y} = \exp(\boldsymbol{W})$ (equivalently, the joint density function of $Y_1, \ldots, Y_n$, where $Y_i = \exp(W_i)$ for $1 \le i \le n$).

    (b) Let $\mu_i = E(W_i)$ denote the ith entry of $\boldsymbol{\mu}$ and let $\sigma_{ij} = \operatorname{cov}(W_i, W_j)$ denote the entry in row i and column j of $\boldsymbol{\Sigma}$. In terms of these entries, determine the mean μ and variance σ^2 of the random variable $W_1 + \cdots + W_n$.

    (c) Determine the density function of $Y_1 \cdots Y_n = \exp(W_1 + \cdots + W_n)$ in terms of μ and σ.

Second Practice Second Midterm Exam

20. Consider the task of giving a twenty minute review lecture on the basic properties and role of *the Poisson distribution and the Poisson process* in probability theory. Write out a complete set of lecture notes that could be used for this purpose by yourself or by another student in the course.

21. Let W_1, W_2, and W_3 be random variables, each of which is greater than 1 with probability 1, and suppose that these random variables have a joint density function. Set $Y_1 = W_1$, $Y_2 = W_1W_2$, and $Y_3 = W_1W_2W_3$. Observe that $1 < Y_1 < Y_2 < Y_3$ with probability 1.

    (a) Determine a formula for the joint density function of Y_1, Y_2, and Y_3 in terms of the joint density function of W_1, W_2, and W_3.

    (b) Suppose that W_1, W_2, and W_3 are independent random variables, each having the density function that equals w^{-2} for $w > 1$ and equals 0 otherwise. Determine the joint density function of Y_1, Y_2, and Y_3.

    (c) (Continued) Are Y_1, Y_2, and Y_3 independent (why or why not)?

22. (a) Let Z_1, Z_2, and Z_3 be uncorrelated random variables, each having variance 1, and set $X_1 = Z_1$, $X_2 = X_1 + Z_2$, and $X_3 = X_2 + Z_3$. Determine the variance-covariance matrix of X_1, X_2, and X_3.

    (b) Let W_1, W_2, and W_3 be uncorrelated random variables having variances σ_1^2, σ_2^2, and σ_3^2, respectively, and set $Y_2 = W_1$, $Y_3 = \alpha Y_2 + W_2$, and $Y_1 = \beta Y_2 + \gamma Y_3 + W_3$. Determine the variance-covariance matrix of Y_1, Y_2, and Y_3.

    (c) Determine the values of α, β, σ_1^2, σ_2^2, and σ_3^2 in order that the variance-covariance matrices in (a) and (b) coincide.

Third Practice Second Midterm Exam

23. Consider the task of giving a 15–20 minute review lecture on the *gamma distribution* in that portion of probability theory that is covered in Chapters 3 and 4 of the textbook, including normal approximation to the gamma distribution

and the role of the gamma distribution in the treatment of the homogeneous Poisson process on $[0, \infty)$. Write out a complete set of lecture notes that could be used for this purpose by yourself or by another student in the course.

24. Let the joint distribution of Y_1, Y_2 and Y_3 be multinomial (trinomial) with parameters $n = 100$, $\pi_1 = .2$, $\pi_2 = .35$ and $\pi_3 = .45$.

    (a) Justify normal approximation to the distribution of $Y_1 + Y_2 - Y_3$.

    (b) Use normal approximation to determine $P(Y_3 \geq Y_1 + Y_2)$.

25. Let X and Y be random variables each having finite variance, and suppose that X is not zero with probability one. Consider linear predictors of Y based on X having the form $\widehat{Y} = bX$.

    (a) Determine the best predictor $\widehat{Y} = \beta X$ of the indicated form, where best means having the minimum mean squared error of prediction.

    (b) Determine the mean squared error of the best predictor of the indicated form.

26. Let $\boldsymbol{Y} = \begin{bmatrix} Y_1, & Y_2, & Y_3 \end{bmatrix}^T$ have the multivariate (trivariate) normal distribution with mean vector $\boldsymbol{\mu} = \begin{bmatrix} 1, & -2, & 3 \end{bmatrix}^T$ and variance-covariance matrix

$$\boldsymbol{\Sigma} = \begin{bmatrix} 1 & -1 & 1 \\ -1 & 2 & -2 \\ 1 & -2 & 3 \end{bmatrix}.$$

    (a) Determine $P(Y_1 \geq Y_2)$.

    (b) Determine a and b such that $[Y_1, Y_2]^T$ and $Y_3 - aY_1 - bY_2$ are independent.

Fourth Practice Second Midterm Exam

27. Consider the task of giving a 20 minute review lecture on topics involving multivariate normal distributions and random vectors having such distributions, as covered in Chapter 5 of the textbook and the corresponding lectures. Write out a complete set of lecture notes that could be used for this purpose by yourself or by another student in the course.

28. A box has three balls: a red ball, a white ball, and a blue ball. A ball is selected at random from the box. Let $I_1 = \text{ind(red ball)}$ be the indicator random variable corresponding to selecting a red ball, so that $I_1 = 1$ if a red ball is selected and $I_1 = 0$ if a white or blue ball is selected. Similarly, set $I_2 = \text{ind(white ball)}$ and $I_3 = \text{ind(blue ball)}$. Note that $I_1 + I_2 + I_3 = 1$.

    (a) Determine the variance-covariance matrix of I_1, I_2 and I_3.

    (b) Determine (with justification) whether or not the variance-covariance matrix of I_1, I_2 and I_3 is invertible.

    (c) Determine β such that I_1 and $I_1 + \beta I_2$ are uncorrelated.

    (d) For this choice of β, are I_1 and $I_1 + \beta I_2$ independent random variables (why or why not)?

29. Let W_1 and W_2 be independent random variables each having the exponential distribution with mean 1. Set $Y_1 = \exp(W_1)$ and $Y_2 = \exp(W_2)$. Determine the density function of $Y_1 Y_2$.

30. Consider a random collection of particles in the plane such that, with probability one, there are only finitely many particles in any bounded region. For $r \geq 0$, let $N(r)$ denote the number of particles within distance r of the origin. Let D_1 denote the distance to the origin of the particle closest to the origin, let D_2 denote the distance to the origin of the next closest particle to the origin, and define D_n for $n = 3, 4, \dots$ in a similar manner. Note that $D_n > r$ if and only if $N(r) < n$. Suppose that, for $r \geq 0$, $N(r)$ has the Poisson distribution with mean r^2. Determine a reasonable normal approximation to $P(D_{100} > 11)$.

First Practice Final Exam

31. Write an essay on the multivariate normal distribution, including conditional distributions and connections with independence and prediction.

32. (a) Let V have the exponential distribution with mean 1. Determine the density function, distribution function, and quantiles of the random variable $W = e^V$.

    (b) Let V have the gamma distribution with shape parameter 2 and scale parameter 1. Determine the density function and distribution function the random variable $Y = e^V$.

33. (a) Let W_1 and W_2 be positive random variables having joint density function f_{W_1,W_2}. Determine the joint density function of $Y_1 = W_1 W_2$ and $Y_2 = W_1/W_2$.

    (b) Suppose, additionally, that W_1 and W_2 are independent random variables and that each of them is greater than 1 with probability 1. Determine a formula for the density function of Y_1.

    (c) Suppose additionally, that W_1 and W_2 have the common density function $f(w) = 1/w^2$ for $w > 1$. Determine the density function of $Y_1 = W_1 W_2$.

    (d) Explain the connection between the answer to (c) and the answers for the density functions in Problem 32.

    (e) Under the same conditions as in (c), determine the density function of $Y_2 = W_1/W_2$.

34. (a) Let X and Y be random variables having finite variance and let c and d be real numbers. Show that $E[(X-c)(Y-d)] = \mathrm{cov}(X,Y) + (EX - c)(EY - d)$.

    (b) Let $(X_1, Y_1), \dots, (X_n, Y_n)$ be independent pairs of random variables, with each pair having the same distribution as (X, Y), and set $\bar{X} = (X_1 + \cdots + X_n)/n$ and $\bar{Y} = (Y_1 + \cdots + Y_n)/n$. Show that $\mathrm{cov}(\bar{X}, \bar{Y}) = \mathrm{cov}(X,Y)/n$.

35. Consider a box having N objects labeled from 1 to N. Let the lth such object have primary value $u_l = u(l)$ and secondary value $v_l = v(l)$. (For example, the primary value could be height in inches and the secondary value could be weight in pounds.) Let the objects be drawn out the box one-at-a-time, at random, by sampling without replacement, and let L_i denote the label of the object selected on the ith trial. Then $U_i = u(L_i)$ is the primary value of the objected selected on the ith trial and $V_i = v(L_i)$ is the secondary value of the object selected on that trial. (Here $1 \le i \le N$.) Also, $S_n = U_1 + \cdots + U_n$ is the sum of the primary values of the objects selected on the first n trials and $T_n = V_1 + \cdots + V_n$ is the sum of the secondary values of the objects selected on these trials. (Here $1 \le n \le N$.) Set $\bar{u} = (u_1 + \cdots + u_N)/N$ and $\bar{v} = (v_1 + \cdots + v_N)/N$.

   (a) Show that $\mathrm{cov}(U_1, V_1) = C$, where $C = N^{-1}\sum_{l=1}^{N}(u_l - \bar{u})(v_l - \bar{v})$.

   (b) Show that $\mathrm{cov}(U_i, V_i) = C$ for $1 \le i \le N$.

   (c) Let $D = \mathrm{cov}(U_1, V_2)$. Show that $\mathrm{cov}(U_i, V_j) = D$ for $i \ne j$.

   (d) Express $\mathrm{cov}(S_n, T_n)$ in terms of C, D, and n.

   (e) Show that $\mathrm{cov}(S_N, T_N) = 0$.

   (f) Use some of the above results to solve for D in terms of C.

   (g) Use some of the above results to show that

$$\mathrm{cov}(S_n, T_n) = nC\left(1 - \frac{n-1}{N-1}\right) = nC\frac{N-n}{N-1}.$$

36. Let Z_1, Z_2, and Z_3 be independent, standard normal random variables and set $Y_1 = Z_1$, $Y_2 = Z_1 + Z_2$, and $Y_3 = Z_1 + Z_2 + Z_3$.

   (a) Determine the joint distribution of Y_1, Y_2, and Y_3.

   (b) Determine the conditional distribution of Y_2 given that $Y_1 = y_1$ and $Y_3 = y_3$.

   (c) Determine the best predictor of Y_2 based on Y_1 and Y_3 and determine the mean squared error of this predictor.

37. (a) Let n and s be positive integers with $n \le s$. Show that (explain why) there are $\binom{s}{n}$ distinct sequences $s_1, \ldots, s_n$ of positive integers such that $s_1 < \cdots < s_n \le s$.

   Let $0 < \pi < 1$. Recall that the negative binomial distribution with parameters $\alpha > 0$ and π has the probability function given by

$$f(y) = \frac{\Gamma(\alpha + y)}{\Gamma(\alpha) y\,!}(1-\pi)^{\alpha}\pi^{y}, \qquad y = 0, 1, 2, \ldots,$$

   and that, when $\alpha = 1$, it coincides with the geometric distribution with parameter π. Let $n \ge 2$ and let $Y_1, \ldots, Y_n$ be independent random variables such each of the random variables $Y_i - 1$ has the geometric distribution with parameter π. It follows from Problem 18 that $(Y_1 - 1) + \cdots + (Y_n - 1) = Y_1 + \cdots + Y_n - n$ has the negative binomial distribution with parameteres $\alpha = n$ and π.

(b) Let $y_1, \ldots, y_n$ be positive integers and set $s = y_1 + \cdots + y_n$. Show that

$$\begin{aligned} & P(Y_1 = y_1, \ldots, Y_{n-1} = y_{n-1} \,|\, Y_1 + \cdots + Y_n = s) \\ &= P(Y_1 - 1 = y_1 - 1, \ldots, Y_{n-1} - 1 = y_{n-1} - 1 \,|\, Y_1 + \cdots + Y_n - n = s - n) \\ &= \frac{1}{\binom{s-1}{n-1}}. \end{aligned}$$

(c) Use the above to explain why the conditional distribution of

$$\{Y_1, Y_1 + Y_2, \ldots, Y_1 + \cdots + Y_{n-1}\}$$

given that $Y_1 + \cdots + Y_n = s$ coincides with the uniform distribution on the collection of all subsets of $\{1, \ldots, s-1\}$ of size $n-1$.

Second Practice Final Exam

38. Consider the task of giving a thirty minute review lecture that is divided into two parts. Part II is to be a fifteen minute lecture on the role/properties of the multivariate normal distribution (*i.e.*, on the material in Sections 5.7 and 6.7 of the textbook) and Part I is to be a fifteen minute review of other material in probability that serves as a necessary background to Part II. Assume that someone has already reviewed the material in Chapters 1–4 and in Section 5.1 (Covariance, Linear Prediction, and Correlation). Thus your task in Part II is restricted to the necessary background to your review in Part I that is contained in Sections 5.2–5.6 on multivariate probability and in relevant portions of Sections 6.1 and 6.4–6.6 on conditioning. As usual, write out a complete set of lecture notes that could be used for this purpose by yourself or by another student in the course.

39. A box has 20 balls, labeled from 1 to 20. Ten balls are selected from the box one-at-a-time by sampling WITH replacement. On each trial, Frankie bets 1 dollar that the label of the selected ball will be between 1 and 10 (including 1 and 10). If she wins the bet she wins 1 dollar; otherwise, she wins -1 dollars (*i.e.*, she loses 1 dollar). Similarly, Johnny repeatedly bets 1 dollar that the label of the selected ball will be between 11 and 20, and Sammy repeatedly bets 1 dollar that it will be between 6 and 15. Let X, Y, and Z denote the amounts (in dollars) won by Frankie, Johnny, and Sammy, respectively, on the ten trials.

    (a) Determine the means and variances of X, Y, and Z.

    (b) Determine $\operatorname{cov}(X,Y)$, $\operatorname{cov}(X,Z)$, and $\operatorname{cov}(Y,Z)$.

    (c) Determine the mean and variance of the combined amount $X + Y + Z$ won by Frankie, Johnny, and Sammy on the ten trials.

40. (a) Let $U_1, U_2, \ldots$ be independent random variables each having the uniform distribution on $[0,1]$, and let A, B, and C be a partition of $[0,1]$ into three disjoint intervals having respective lengths π_1, π_2, and π_3 (which necessarily add up to 1). Let n be a positive integer. Let

$Y_1 = \#\{i : 1 \le i \le n \text{ and } U_i \in A\}$ denote the number of the first n trials in which the random variable U_i lies in the set A. Similarly, let Y_2 and Y_3 denote the numbers of the first n trials in which U_i lies in B and C, respectively. Explain convincingly why the joint distribution of Y_1, Y_2, and Y_3 is trinomial with parameters n, π_1, π_2, and π_3.

(b) Let N be a random variable having the Poisson distribution with mean λ, and suppose that N is independent of $U_1, U_2, \ldots$. Let $Y_1 = \#\{i : 1 \le i \le N \text{ and } U_i \in A\}$ now denote the number of the first N trials in which U_i lies in the set A. Similarly, let Y_2 and Y_3 denote the numbers of the first N trials in which U_i lies in B and C respectively. Observe that $Y_1 + Y_2 + Y_3 = N$. (If $N = 0$, then $Y_1 = Y_2 = Y_3 = 0$; the conditional joint distribution of Y_1, Y_2, and Y_3 given that $N = n \ge 1$ is the trinomial distribution discussed in (a).) Show that Y_1, Y_2, and Y_3 are independent random variables each having a Poisson distribution, and determine the means of these random variables.

41. Let X and Y have the joint density function given by

$$f(x, y) = 2ye^{-(x^2+y^2)/x}, \qquad x, y > 0,$$

and $f(x, y) = 0$ otherwise.

(a) Determine the marginal density function, mean, and variance of X.

(b) Determine the conditional density function of Y given $X = x > 0$.

(c) Determine the mean and variance of this conditional density function.

(d) Determine the best predictor of Y based on X, and determine the mean squared error of this predictor.

(e) Use the answers to (a) and (c) to determine the mean and variance of Y.

42. Consider a box having 10 balls, of which 3 are red and 7 are white.

(a) Let the balls be drawn out of the box by sampling WITHOUT replacement. Let N_1 denote the first trial in which a red ball is selected, N_2 the second trial in which a red ball is selected, and N_3 the third trial in which a red ball is selected, so that $1 \le N_1 < N_2 < N_3 \le 10$. Determine $P(N_1 = 2, N_2 = 5, N_3 = 9)$ by using conditional probabilities.

(b) (Continued) Determine $P(N_1 = n_1, N_2 = n_2, N_3 = n_3)$ for $1 \le n_1 < n_2 < n_3 \le 10$ and justify your answer.

(c) Let the balls now be selected by sampling WITH replacement, and let N_1, N_2, N_3 be as in (a), so that $1 \le N_1 < N_2 < N_3$. Determine $P(N_1 = 2, N_2 = 5, N_3 = 9)$.

(d) (Continued) Determine $P(N_1 = n_1, N_2 = n_2, N_3 = n_3)$ for $1 \le n_1 < n_2 < n_3$.

43. Let Y_1 be normally distributed with mean 0 and variance σ_1^2, let the conditional distribution of Y_2 given that $Y_1 = y_1$ be normally distributed with mean $\beta_1 y_1$ and variance σ_2^2, and let the conditional distribution of Y_3 given that $Y_1 = y_1$ and $Y_2 = y_2$ be normally distributed with mean $\beta_2 y_1 + \beta_3 y_2$ and variance σ_3^2. Use Theorem 6.7 or its corollary to determine

(a) the joint distribution of Y_1 and Y_2;

(b) the joint distribution of Y_1, Y_2, and Y_3.

Third Practice Final Exam

44. The Summary of Probability in Appendix B of the textbook omits a summary of conditioning. Write out a set of notes that could be used to prepare a coherent draft of a summary of a substantial portion of the material on conditioning in Chapter 6. (Don't include the material in Sections 6.2 and 6.3 on sampling without replacement and the hypergeometric distribution or the material in Section 6.8 on random parameters.)

45. Let $U_1, U_2, \ldots$ be independent random variables, each having the uniform distribution on $[0, 1]$. Set

$$N_1 = \min\{i \geq 1 : U_i \geq 1/4\} \quad \text{and} \quad N_2 = \min\{i \geq N_1 + 1 : U_i \geq 1/4\}.$$

(a) Show that N_1 and $N_2 - N_1$ are independent.

(b) Determine the probability function of N_2.

46. Let f be the function given by $f(y) = cy^2(1-y^2)$ for $-1 < y < 1$ and $f(y) = 0$ elsewhere.

(a) Determine the constant c such that f is a density function.

(b) Let c be such that f is a density function, and let Y be a random variable having this density function. Determine the mean and variance of Y.

(c) Let $Y_1, \ldots, Y_{100}$ be independent random variables each having the same distribution as Y, and set $\bar{Y} = (Y_1 + \cdots + Y_{100})/100$. Determine (and justify) a reasonable approximation to the upper quartile of $\bar{Y}$.

47. Let X and Y be random variables, each having finite, positive variance. Set $\mu_1 = E(X)$, $\sigma_1 = \mathrm{SD}(X)$, $\mu_2 = E(Y)$, $\sigma_2 = \mathrm{SD}(Y)$, and $\rho = \mathrm{cor}(X, Y)$. Let $(X_1, Y_1), \ldots, (X_n, Y_n)$ be independent pairs of random variables, with each pair having the same joint distribution as (X, Y). Set $\bar{X} = (X_1 + \cdots + X_n)/n$ and $\bar{Y} = (Y_1 + \cdots + Y_n)/n$. Suppose μ_1, σ_1, σ_2 and ρ are known, but μ_2 is unknown. Consider the unbiased estimates $\widehat{\mu}_2^{(1)} = \bar{Y}$ and $\widehat{\mu}_2^{(2)} = \bar{Y} - \bar{X} + \mu_1$ of μ_2. Determine a simple necessary and sufficient condition on σ_1, σ_2, and ρ in order that $\mathrm{var}(\widehat{\mu}_2^{(2)}) < \mathrm{var}(\widehat{\mu}_2^{(1)})$.

48. Let $Y_1, Y_2, \ldots$ be uncorrelated random variables and suppose that $E(Y_i) = \mu$ and $\mathrm{var}(Y_i) = i^b$ for $i \geq 1$, where b is a positive constant. Recall from calculus that

$$1 = \lim_{n\to\infty} \frac{1^b + \cdots + n^b}{\int_1^n t^b\, dt} = \lim_{n\to\infty} \frac{1^b + \cdots + n^b}{n^{b+1}/(b+1)} = (b+1) \lim_{n\to\infty} \frac{1^b + \cdots + n^b}{n^{b+1}}.$$

Set $\bar{Y}_n = (Y_1 + \cdots + Y_n)/n$ for $n \geq 1$.

(a) Use the Chebyshev inequality to show that if $b < 1$, then

$$\lim_{n\to\infty} P(|\bar{Y}_n - \mu| \geq c) = 0, \qquad c > 0. \tag{1}$$

(b) Show that if (1) holds, if $Y_1, Y_2, \ldots$ are independent, and if these random variables are normally distributed, then $b < 1$.

(c) Explain how the assumptions of independence and normality are needed in the verification of the conclusion in (b).

49. Let X_1, X_2 and X_3 be positive random variables having joint density f_{X_1,X_2,X_3}. Set

$$Y_1 = X_1 + X_2 + X_3, \quad Y_2 = \frac{X_1 + X_2}{X_1 + X_2 + X_3} \quad \text{and} \quad Y_3 = \frac{X_1}{X_1 + X_2}.$$

(a) Determine the joint density f_{Y_1,Y_2,Y_3} of Y_1, Y_2 and Y_3 in terms of the joint density of X_1, X_2 and X_3.

(b) Suppose that X_1, X_2 and X_3 are independent random variables, each having the exponential distribution with inverse-scale parameter λ. Determine the individual densities/distributions of Y_1, Y_2 and Y_3 and show that these random variables are independent.

50. Let $\boldsymbol{Y} = \begin{bmatrix} X_1, & X_2, & Y \end{bmatrix}^T$ have the multivariate (trivariate) normal distribution with mean vector $\begin{bmatrix} 1, & -2, & 3 \end{bmatrix}^T$ and variance-covariance matrix

$$\begin{bmatrix} 1 & -1 & 1 \\ -1 & 2 & -2 \\ 1 & -2 & 3 \end{bmatrix}.$$

(a) Determine the distribution of X_1.

(b) Determine the conditional distribution of Y given that $X_1 = x_1$.

(c) Determine the joint distribution of X_1 and X_2.

(d) Determine the conditional distribution of Y given that $X_1 = x_1$ and $X_2 = x_2$.

(e) Let $\widehat{X}_2$ be the best predictor of X_2 based on X_1; let $\widehat{Y}^{(1)}$ be the best predictor of Y based on X_1; and let $\widehat{Y}^{(2)}$ be the best predictor of Y based on X_1 and X_2. Determine $\widehat{X}_2$, $\widehat{Y}^{(1)}$ and $\widehat{Y}^{(2)}$ explicitly, and use these results to illustrate Theorem 5.17.

Fourth Practice Final Exam

51. Consider the task of giving a half-hour review lecture on prediction, including both linear prediction and nonlinear (i.e., not necessarily linear) prediction, as covered in Chapters 5 and 6 of the textbook and the corresponding lectures. Write out a set of lecture notes that could be used for this purpose by yourself or by another student in the course.

52. Let $Y_1, ..., Y_{100}$ be independent random variables each having the density function f given by $f(y) = 2y\exp(-y^2)$ for $y > 0$ and $f(y) = 0$ for $y \leq 0$. Determine a reasonable approximation to the upper decile of $Y_1 + \cdots + Y_{100}$.

53. A box starts out with 1 red ball and 1 white ball. At each trial a ball is selected from the box. Then it is returned to the box along with another ball of the same color (from another box). Let Y_n denote the number of red balls selected in the first n trials.

    (a) Show that Y_n is uniformly distributed on $\{0, \ldots, n\}$ for $n = 1, 2, 3, 4$ by computing its probability function for each such n.

    (b) Show that Y_n is uniformly distributed on $\{0, \ldots, n\}$ for $n \geq 1$. *Hint:* Given $y \in \{0, \ldots, n\}$, first determine the probability that the first y balls selected are red and the remaining $n - y$ balls selected are white.

54. Let $N, X_1, X_2, \ldots$ be independent random variables such that $N - 1$ has the geometric distribution with parameter $\pi \in (0, 1)$ (so that $P(N = n) = (1 - \pi)\pi^{n-1}$ for $n \geq 1$) and each of the random variables $X_1, X_2, \ldots$ has the exponential distribution with scale parameter β. Set $Y_n = X_1 + \cdots + X_n$ for $n \geq 1$.

    (a) Determine the density function f_{Y_n} of Y_n for $n \geq 1$.

    (b) Determine the density function and distribution of Y_N. *Hint:* Use an appropriate analog to (10) in Chapter 6.

    (c) Determine the mean and variance of Y_N.

    (d) Consider the above setup, but suppose now that each of the random variables $X_1, X_2, \ldots$ has the gamma distribution with shape parameter 2 and scale parameter β. Determine the mean and variance of Y_n for $n \geq 1$.

    (e) (Continued) Determine the mean and variance of Y_N.

55. Let W_1, W_2, W_3, and W_4 be independent random variables, each having the exponential distribution with inverse-scale parameter λ. Set $T_1 = W_1$, $T_2 = W_1 + W_2$, $T_3 = W_1 + W_2 + W_3$, and $T_4 = W_1 + W_2 + W_3 + W_4$.

    (a) Determine the joint density function of T_1, T_2, T_3, and T_4.

    (b) Determine the conditional density function of T_4 given that $T_1 = t_1$, $T_2 = t_2$, and $T_3 = t_3$, where $0 < t_1 < t_2 < t_3$.

    (c) Determine the variance-covariance matrix of T_1, T_2, T_3, and T_4.

    (d) Determine whether or not this variance-covariance matrix is invertible (justify your answer, preferably by something more elegant than a brute-force calculation of its determinant).

Brief Solutions to Practice Exams

Solutions to First Practice First Midterm Exam

2. Let $0 < w_1 < w_2 < \infty$. Set $g(w) = 0$ for $0 \le w < w_1$, $g(w) = 1$ for $w_1 \le w < w_2$, and $g(w) = 2$ for $w \ge w_2$. Then the probability function of $Y = g(W)$ is given by $P(Y = 0) = P(0 \le W < w_1) = 1 - e^{-w_1}$, $P(Y = 1) = P(w_1 \le W < w_2) = e^{-w_1} - e^{-w_2}$, and $P(Y = 2) = P(W \ge w_2) = e^{-w_2}$. Thus Y is uniformly distributed on $\{0, 1, 2\}$ provided that $1 - e^{-w_1} = 1/3$ and $e^{-w_2} = 1/3$; that is, $w_1 = \log(3/2)$ and $w_2 = \log 3$.

3. (a) The distribution function of W is given by $F_W(w) = -w^{-2}\big|_1^w = 1 - w^{-2}$ for $w > 1$ and $F_W(w) = 0$ for $w \le 1$. Thus the distribution function of Y is given by $F_Y(y) = P(Y \le y) = P(\alpha + \beta W \le y) = P(W \le (y - \alpha)/\beta)$ $= F_W((y - \alpha)/\beta) = 1 - \beta^2/(y - \alpha)^2$ for $y > \alpha + \beta$ and $F_Y(y) = 0$ for $y \le \alpha + \beta$.

   (b) The density function of Y is given by $f_Y(y) = F'_Y(y) = 2\beta^2/(y - \alpha)^3$ for $y > \alpha + \beta$ and $f_Y(y) = 0$ for $y \le \alpha + \beta$.

   (c) The pth quantile of Y is given by $1 - \beta^2/(y_p - \alpha)^2 = p$ or $y_p = \alpha + \beta/\sqrt{1 - p}$ for $0 < p < 1$.

   (d) The mean of W is given by $E(W) = \int_1^\infty w(2/w^3)\, dw = 2\int_1^\infty 1/w^3\, dw = 2(-1/w)\big|_1^\infty = 2$. Thus the mean of Y is given by $E(Y) = E(\alpha + \beta W) = \alpha + \beta EW = \alpha + 2\beta$.

   (e) The second moment of W is given by $E(W^2) = \int_1^\infty w^2(2/w^3)\, dw = 2\int_1^\infty 1/w\, dw = 2(\log w)\big|_1^\infty = \infty$. Thus the second moment of Y is given by $E(Y^2) = E[(\alpha + \beta W)^2] = E(\alpha^2 + 2\alpha\beta W + \beta^2 W^2) = \infty$. Consequently, Y has infinite variance.

4. As in the proof of the Markov and Chebyshev inequalities, $E[(Y - \mu)^4] \ge 10^4 P(|Y - \mu| \ge 10)$. Thus $P(|Y - \mu| \ge 10) \le 10^{-4} E[(Y - \mu)^4] \le 2 \times 10^{-4}$. Alternatively, the same result follows by applying the Markov inequality to the random variable $(Y - \mu)^4$.

5. Observe that if $u + v > 1.5$ and $u - v > 0.5$, then $u > 1$. Consequently, $P(U + V > 1.5 \text{ and } U - V > 0.5) = 0$. On the other hand, $P(U + V > 1.5) > P(U > 0.75, V > 0.75) = P(U > 0.75)P(V > 0.75) = 0.0625$ and $P(U - V > 0.5) > P(U > 0.75, V < 0.25) = P(U > 0.75)P(V < 0.25) = 0.0625$, so $P(U + V > 1.5 \text{ and } U - V > 0.5) \ne P(U + V > 1.5)P(U - V > 0.5)$. Therefore, $U + V$ and $U - V$ are not independent random variables.

6. Now $E(U^2) = \int_0^1 u^2\, du = 1/3$ and $E(U^4) = \int_0^1 u^4\, du = 1/5$, so $\operatorname{var}(U^2) = E(U^4) - [E(U^2)]^2 = 1/5 - 1/9 = 4/45$. Also, $E(V) = 1/2$ and $E(V^2) = 1/3$, so $\operatorname{var}(V) = E(V^2) - [E(V)]^2 = 1/3 - 1/4 = 1/12$. Consequently, $E(Y) = E(3U^2 - 2V) = 3E(U^2) - 2E(V) = 1 - 1 = 0$ and $\operatorname{var}(Y) = \operatorname{var}(3U^2 - 2V) = 9\operatorname{var}(U^2) + 4\operatorname{var}(V) = 9(4/45) + 4(1/12) = 4/5 + 1/3 = 17/15$.

Solutions to Second Practice First Midterm Exam

8. (a) $F_W(w) = w^2$ for $0 < w < 1$, $F_W(w) = 0$ for $w < 0$, and $F_W(w) = 1$ for $w \geq 1$. Also, $w_p = \sqrt{p}$ for $0 < p < 1$.

   (b) $F_Y(y) = P(Y \leq y) = P(e^W \leq y) = P(W \leq \log Y) = F_W(\log y) = \log^2 y$ for $1 < y < e$, $F_Y(y) = 0$ for $y \leq 1$, and $F_Y(y) = 1$ for $y \geq e$. Thus $f_Y(y) = \frac{2\log y}{y}$ for $1 < y < e$ and $f_Y(y) = 0$ elsewhere. Also, $\log^2 y_p = p$, so $y_p = \exp(\sqrt{p})$.

   (c) $EY = \int_1^e 2\log y\, dy = 2y\log y\big|_1^e - \int_1^e 2\, dy = 2e - 2(e-1) = 2$ and $E(Y^2) = \int_1^e 2y\log y\, dy = \int_1^e \log y\, dy^2 = y^2\log y\big|_1^e - \int_1^e y\, dy = e^2 - \frac{e^2-1}{2} = \frac{e^2+1}{2}$, so $\operatorname{var}(Y) = \frac{e^2+1}{2} - 4 = \frac{e^2-7}{2}$.

   (d) $EY = \int_0^1 2we^w\, dw = 2we^w\big|_0^1 - \int_0^1 2e^w\, dw = 2e - 2(e-1) = 2$ and $E(Y^2) = \int_0^1 2we^{2w}\, dw = we^{2w}\big|_0^1 - \int_0^1 e^{2w}\, dw = e^2 - \frac{e^2-1}{2} = \frac{e^2+1}{2}$, so $\operatorname{var}(Y) = \frac{e^2+1}{2} - 4 = \frac{e^2-7}{2}$.

9. (a) Now $EW_1 = EW_2 = 0\cdot\frac{1}{2} + 1\cdot\frac{1}{3} + 2\cdot\frac{1}{6} = \frac{2}{3}$, so $EY = \frac{4}{3}$. Also, $E(W_1^2) = E(W_2^2) = 0^2\cdot\frac{1}{2} + 1^2\cdot\frac{1}{3} + 2^2\cdot\frac{1}{6} = 1$, so $\operatorname{var}(W_1) = \operatorname{var}(W_2) = 1 - \frac{4}{9} = \frac{5}{9}$ and hence $\operatorname{var}(Y) = \frac{10}{9}$ and $\operatorname{SD}(Y) = \frac{\sqrt{10}}{3}$.

   (b) $P(Y \geq 3) \leq \frac{EY}{3} = \frac{4}{9}$.

   (c) $P(Y \geq 3) = P\Big(Y - \frac{4}{3} \geq \frac{5}{3}\Big) \leq \frac{10/9}{25/9} = \frac{2}{5}$.

   (d) $P(Y \geq 3) = P(W_1 = 1, W_2 = 2) + P(W_1 = 2, W_2 = 1) + P(W_1 = 2, W_2 = 2) = 2\cdot\frac{1}{3}\cdot\frac{1}{6} + \frac{1}{6}\cdot\frac{1}{6} = \frac{5}{36}$.

Solutions to Third Practice First Midterm Exam

11. (a) $F(w) = \int_1^w t^{-2}\, dt = -t^{-1}\big|_1^w = 1 - w^{-1}$ for $w > 1$ and $F(w) = 0$ for $w \leq 1$.

   (b) $P(Y_n > y) = P(\min(W_1, \dots, W_n) > y) = P(W_1 > y, \dots, W_n > y) = [P(W_1 > y)]^n = y^{-n}$ for $y > 1$ and $P(Y_n > y) = 1$ for $y \leq 1$. Thus the distribution function of Y_n is given by $F_{Y_n}(y) = 0$ for $y \leq 1$ and $F_{Y_n}(y) = 1 - y^{-n}$ for $y > 1$. Consequently, the density function of Y_n is given by $f_{Y_n}(y) = 0$ for $y \leq 1$ and $f_{Y_n}(y) = F'_{Y_n}(y) = ny^{-n-1}$ for $y > 1$. The pth quantile of Y_n is given by $1 - y_p^{-n} = p$ or $y_p = 1/(1-p)^{1/n}$.

   (c) $E(Y_n) = \int_1^\infty ny^{-n}\, dy = \frac{n}{1-n}y^{1-n}\big|_1^\infty < \infty$ for $n \geq 2$ and $E(Y_1) = \int_1^\infty \frac{1}{y}\, dy = \log y\big|_1^\infty = \infty$. Consequently, Y_n has finite mean if and only if $n \geq 2$.

   (d) $E(Y_n^2) = \int_1^\infty ny^{1-n}\, dy = \frac{n}{2-n}y^{2-n}\big|_1^\infty < \infty$ if $n \geq 3$, $E(Y_1^2) = \int_1^\infty dy = \infty$, and $E(Y_2^2) = \int_1^\infty 2y^{-1}\, dy = \infty$. Thus Y_n has finite second moment and hence finite variance if and only if $n \geq 3$.

12. (a) $E(Y) = E(W_1 - 3W_2 + 2W_3) = E(W_1) - 3E(W_2) + 2E(W_3) = \frac{1}{2}(1-3+2) = 0$ and $\operatorname{var}(Y) = \operatorname{var}(W_1 - 3W_2 + 2W_3) = \operatorname{var}(W_1) + 9\operatorname{var}(W_2) + 4\operatorname{var}(W_3) = \frac{14}{12} = \frac{7}{6}$. Thus, by Chebyshev's inequality, $P(|Y| \geq 2) \leq \frac{7/6}{4} = \frac{7}{24}$.

   (b) $P(Y = 0) = P(W_1 < 1/2, W_2 < 1/3, W_3 < 1/4) = P(W_1 < 1/2)P(W_2 < 1/3)P(W_3 < 1/4) = (1/2)(1/3)(1/4) = \frac{1}{24}$; $P(Y = 1) = P(W_1 \geq$

$1/2)P(W_2 < 1/3)P(W_3 < 1/4) + P(W_1 < 1/2)P(W_2 \geq 1/3)P(W_3 < 1/4) + P(W_1 < 1/2)P(W_2 < 1/3)P(W_3 \geq 1/4) = (1/2)(1/3)(1/4) + (1/2)(2/3)(1/4) + (1/2)(1/3)(3/4) = \frac{6}{24}$; $P(Y = 2) = P(W_1 \geq 1/2)P(W_2 \geq 1/3)P(W_3 < 1/4) + P(W_1 \geq 1/2)P(W_2 < 1/3)P(W_3 \geq 1/4)P(W_1 < 1/2)P(W_2 \geq 1/3)P(W_3 \geq 1/4) = (1/2)(2/3)(1/4) + (1/2)(1/3)(3/4) + (1/2)(2/3)(3/4) = \frac{17}{24}$; $P(Y = 3) = P(W_1 \geq 1/2)P(W_2 \geq 1/3)P(W_3 \geq 1/4) = (1/2)(2/3)(3/4) = \frac{6}{24}$.

Solutions to the Fourth Practice First Midterm Exam

14. (a) The distribution function F of Y is given by $F(y) = 0$ for $y \leq 0$, $F(y) = \int_0^y (z/2)\,dz = y^2/4$ for $0 < y < 2$ and $F(y) = 1$ for $y \geq 2$.

(b) The function g is given by $g(u) = F^{-1}(u)$. The solution to the equation $u = F(y) = y^2/4$ is given by $y = F^{-1}(u) = g(u) = \sqrt{4u}$ for $0 < u < 1$.

(c) The lower quartile of Y is given by $y_{.25} = F^{-1}(.25) = \sqrt{4(.25)} = 1$. The upper quartile of Y is given by $y_{.75} = F^{-1}(.75) = \sqrt{4(.75)} = \sqrt{3}$. Thus, by Theorem 1.4, the lower and upper quartiles of $a + bY$ are given by $a + by_{.25} = a + b$ and $a + by_{.75} = a + b\sqrt{3}$, respectively. The solution to the equations $a + b = 0$ and $a + b\sqrt{3} = 1$ is given by $b = 1/(\sqrt{3} - 1) = (1 + \sqrt{3})/2 \doteq 1.366$ and $a = -(1 + \sqrt{3})/2 \doteq -1.366$.

(d) Now $E(Y) = \int_0^2 y(y/2)\,dy = (y^3/6)\big|_0^2 = 4/3$ and $E(Y^2) = \int_0^2 y^2(y/2)\,dy = (y^4/8)\big|_0^2 = 2$, so $\mathrm{var}(Y) = E(Y^4) - [E(Y)]^2 = 2 - 16/9 = 2/9$. Thus the variance of $a + bY$ is given by

$$b^2 \cdot \frac{2}{9} = \Big(\frac{1 + \sqrt{3}}{2}\Big)^2 \cdot \frac{2}{9} = \frac{4 + 2\sqrt{3}}{4} \cdot \frac{2}{9} = \frac{2 + \sqrt{3}}{9} \doteq 0.415$$

or, alternatively, by $b^2(2/9) \doteq (1.866)(0.2222) \doteq 0.415$.

15. (a) Now $P(I_1 = 1) = P(I_2 = 1) = 1/3$, but $P(I_1 = 1, I_2 = 1) = P(Y \in \{1, \dots, 12\} \cap \{13, \dots, 24\}) = P(Y \in \emptyset) = 0$, so I_1 and I_2 are not independent and hence I_1, I_2 and I_3 are not independent.

(b) Now $E(I_1 - 2I_2 + 3I_3) = P(I_1 = 1) - 2P(I_2 = 1) + 3P(I_3 = 1) = (1/3) - 2(1/3) + 3(1/2) = 7/6$. Note that $E(I_1 I_2) = P(I_1 = 1, I_2 = 1) = 0$ and similarly that $E(I_1 I_3) = 0$. Also, $E(I_2 I_3) = P(I_2 = 1, I_3 = 1) = P(Y \in \{13, \dots, 24\} \cap \{19, \dots, 36\}) = P(Y \in \{19, \dots, 24\}) = 1/6$. Moreover, $E(I_1^2) = E(I_1)$, $E(I_2^2) = E(I_2)$, and $E(I_3^2) = E(I_3)$. Thus $E[(I_1 - 2I_2 + 3I_3)^2] = E(I_1) + 4E(I_2) + 9E(I_3) - 12E(I_2 I_3) = 1/3 + 4(1/3) + 9(1/2) - 12(1/6) = 25/6$, so $\mathrm{var}\big((I_1 - 2I_2 + 3I_3)^2\big) = E[(I_1 - 2I_2 + 3I_3)^2] - [E(I_1 - 2I_2 + 3I_3)]^2 = 25/6 - (7/6)^2 = 101/36$. Alternatively, set $V = I_1 - 2I_2 + 3I_3$. Observe that if $Y \in \{1, \dots, 12\}$, then $I_1 = 1$, $I_2 = 0$, and $I_3 = 0$, so $V = 1$; if $Y \in \{13, \dots, 18\}$, then $I_1 = 0$, $I_2 = 1$, and $I_3 = 0$, so $V = -2$; if $Y \in \{19, \dots, 24\}$, then $I_1 = 0$, $I_2 = 1$, and $I_3 = 1$, so $V = 1$; and if $Y \in \{25, \dots, 36\}$, then $I_1 = 0$, $I_2 = 0$, and $I_3 = 1$, so $V = 3$. Consequently, $P(V = -2) = 1/6$, $P(V = 1) = 1/2$, and $P(V = 3) = 1/3$. Therefore, $E(V) = (-2)(1/6) + 1/2 + 3(1/3) = 7/6$ and $E(V^2) = (-2)^2(1/6) + 1^2(1/2) + 3^2(1/3) = 25/6$, so $\mathrm{var}(V) = E(V^2) - [E(V)]^2 = 25/6 - (7/6)^2 = 101/36$.

Solutions to First Practice Second Midterm Exam

17. $E(Y^3) = \int_0^\infty y^3 \frac{y}{\beta^2} e^{-y/\beta}\,dy = \frac{1}{\beta^2}\int_0^\infty y^4 e^{-y/\beta}\,dy = \frac{\beta^5\Gamma(5)}{\beta^2} = 4!\beta^3 = 24\beta^3$; $E(Y^6) = \int_0^\infty y^6 \frac{y}{\beta^2} e^{-y/\beta}\,dy = \frac{1}{\beta^2}\int_0^\infty y^7 e^{-y/\beta}\,dy = \frac{\beta^8\Gamma(8)}{\beta^2} = 7!\beta^6 = 5040\beta^6$; $\operatorname{var}(Y^3) = E(Y^6) - [E(Y^3)]^2 = 5040\beta^6 - 576\beta^6 = 4464\beta^6$.

18. (a) Now $\frac{f(y)}{f(y-1)} = \frac{\frac{\Gamma(\alpha+y)}{\Gamma(\alpha)y!}}{\frac{\Gamma(\alpha+y-1)}{\Gamma(\alpha)(y-1)!}}\frac{\pi^y}{\pi^{y-1}} = \frac{(\alpha+y-1)\pi}{y} = \Bigl(1+\frac{\alpha-1}{y}\Bigr)\pi$ for $y = 1, 2, \ldots$. If $\alpha \le 1$, this ratio is always less than 1, so 0 is the unique mode of the probability function. If $\alpha > 1$, this ratio is a decreasing function of y (viewed as a positive real number). It equals 1 if and only if $(\alpha+y-1)\pi = y$ or, equivalently, if and only if $y = (\alpha-1)\pi/(1-\pi)$. Suppose that $(\alpha-1)\pi/(1-\pi)$ is a positive integer r. Then the probability function has two modes: r and $r-1$. Otherwise, the probability function has a unique mode at $[(\alpha-1)\pi/(1-\pi)]$.

    (b) The probability function of $Y_1 + Y_2$ is given by

$$
\begin{aligned}
P(Y_1+Y_2=y) &= \sum_{x=0}^{y} P(Y_1=x)P(Y_2=y-x) \\
&= \sum_{x=0}^{y} \frac{\Gamma(\alpha_1+x)}{\Gamma(\alpha_1)x!}(1-\pi)^{\alpha_1}\pi^x \frac{\Gamma(\alpha_2+y-x)}{\Gamma(\alpha_2)(y-x)!}(1-\pi)^{\alpha_2}\pi^{y-x} \\
&= (1-\pi)^{\alpha_1+\alpha_2}\pi^y \sum_{x=0}^{y} \frac{\Gamma(\alpha_1+x)}{\Gamma(\alpha_1)x!}\,\frac{\Gamma(\alpha_2+y-x)}{\Gamma(\alpha_2)(y-x)!} \\
&= \frac{\Gamma(\alpha_1+\alpha_2+y)}{\Gamma(\alpha_1+\alpha_2)y!}(1-\pi)^{\alpha_1+\alpha_2}\pi^y, \qquad y = 0, 1, 2, \ldots .
\end{aligned}
$$

19. (a) Now $f_{\boldsymbol{W}}(\boldsymbol{w}) = \frac{1}{(2\pi)^{n/2}(\det\boldsymbol{\Sigma})^{1/2}}\exp\bigl(-(\boldsymbol{w}-\boldsymbol{\mu})^T\boldsymbol{\Sigma}^{-1}(\boldsymbol{w}-\boldsymbol{\mu})/2\bigr)$. The solution to $y_i = \exp(w_i)$ for $1 \le i \le n$ is given by $w_i = \log y_i$ for $1 \le i \le n$. Observe that $\frac{\partial(w_1,\ldots,w_n)}{\partial(y_1,\ldots,y_n)} = \frac{1}{y_1\cdots y_n}$. Thus $f_{\boldsymbol{Y}}(\boldsymbol{y}) = \frac{f_{\boldsymbol{W}}(\log\boldsymbol{y})}{y_1\cdots y_n} = \frac{1}{(2\pi)^{n/2}(\det\boldsymbol{\Sigma})^{1/2}y_1\cdots y_n}\exp\bigl(-(\log\boldsymbol{y}-\boldsymbol{\mu})^T\boldsymbol{\Sigma}^{-1}((\log\boldsymbol{y}-\boldsymbol{\mu})/2\bigr)$.

    (b) $\mu = E(W_1+\cdots+W_n) = \mu_1+\cdots+\mu_n$; $\sigma^2 = \operatorname{var}(W_1+\cdots+W_n) = \sum_i\sum_j \operatorname{cov}(W_i,W_j) = \sum_i\sum_j \sigma_{ij}$.

    (c) Now $W = W_1+\cdots+W_n$ is normally distributed with mean μ and variance σ^2. Thus the distribution function of $Y = e^W$ is given by $F_Y(y) = P(Y \le y) = P(e^W \le y) = P(W \le \log y) = \Phi\Bigl(\frac{\log y-\mu}{\sigma}\Bigr)$. Consequently, the density function of Y is given by $f_Y(y) = F_Y'(y) = \frac{1}{y\sigma\sqrt{2\pi}}\exp\bigl(-\frac{(\log y-\mu)^2}{2\sigma^2}\bigr)$. Alternatively, this result can be obtained from Theorem 5.20.

Solutions to Second Practice Second Midterm Exam

21. (a) Let $w_1, w_2, w_3 > 1$. The equations $y_1 = w_1$, $y_2 = w_1w_2$, and $y_3 = w_1w_2w_3$ have the unique solution given by $w_1 = y_1$, $w_2 = y_2/y_1$, and $w_3 =$

y_3/y_2. Observe that $\frac{\partial(w_1,w_2,w_3)}{\partial(y_1,y_2,y_3)} = \begin{vmatrix} 1 & 0 & 0 \\ -y_2/y_1^2 & 1/y_1 & 0 \\ 0 & -y_3/y_2^2 & 1/y_2 \end{vmatrix} = \frac{1}{y_1 y_2}$. Thus $f_{Y_1,Y_2,Y_3}(y_1,y_2,y_3) = \frac{1}{y_1 y_2} f_{W_1,W_2,W_3}(y_1, y_2/y_1, y_3/y_2)$ for $1 < y_1 < y_2 < y_3$, and this joint density function equals zero elsewhere.

(b) The joint density function of Y_1, Y_2, and Y_3 is given by $f_{Y_1,Y_2,Y_3}(y_1,y_2,y_3 = \frac{1}{y_1 y_2 y_1^2 (y_2/y_1)^2 (y_3/y_2)^2} = \frac{1}{y_1 y_2 y_3^2}$ for $1 < y_1 < y_2 < y_3$, and this joint density function equals zero elsewhere.

(c) The random variables Y_1, Y_2, and Y_3 are dependent. Their joint density function appears to factor as $\left(\frac{1}{y_1}\right)\left(\frac{1}{y_2}\right)\left(\frac{1}{y_3^2}\right)$, but this reasoning ignores the fact that the range $1 < y_1 < y_2 < y_3$ is not a rectangle with sides parallel to the coordinate axes and hence that the corresponding indicator function does not factor. Alternatively, $P(2 < Y_1 < 3) > 0$ and $P(1 < Y_2 < 2) > 0$, but $P(2 < Y_1 < 3$ and $1 < Y_2 < 2) = 0$, so Y_1 and Y_2 are dependent.

22. (a) Now $X_1 = Z_1$, $X_2 = Z_1 + Z_2$, and $x_3 = Z_1 + Z_2 + Z_3$, so $\mathrm{var}(X_1) = 1$, $\mathrm{var}(X_2) = 2$, $\mathrm{var}(X_3) = 3$, $\mathrm{cov}(X_1, X_2) = 1$, $\mathrm{cov}(X_1, X_3) = 1$, and $\mathrm{cov}(X_2, X_3) = 2$. Thus the desired variance-covariance matrix equals

$$\begin{bmatrix} 1 & 1 & 1 \\ 1 & 2 & 2 \\ 1 & 2 & 3 \end{bmatrix}.$$

(b) Now $Y_2 = W_1$, $Y_3 = \alpha W_1 + W_2$, and $Y_1 = (\beta + \alpha\gamma)W_1 + \gamma W_2 + W_3$. Thus $\mathrm{var}(Y_1) = (\beta+\alpha\gamma)^2\sigma_1^2 + \gamma^2\sigma_2^2 + \sigma_3^2$, $\mathrm{var}(Y_2) = \sigma_1^2$, $\mathrm{var}(Y_3) = \alpha^2\sigma_1^2 + \sigma_2^2$, $\mathrm{cov}(Y_1, Y_2) = (\beta + \alpha\gamma)\sigma_1^2$, $\mathrm{cov}(Y_1, Y_3) = \alpha(\beta + \alpha\gamma)\sigma_1^2 + \gamma\sigma_2^2$, and $\mathrm{cov}(Y_2, Y_3) = \alpha\sigma_1^2$. Thus the desired variance-covariance matrix equals

$$\begin{bmatrix} (\beta+\alpha\gamma)^2\sigma_1^2 + \gamma^2\sigma_2^2 + \sigma_3^2 & (\beta+\alpha\gamma)\sigma_1^2 & \alpha(\beta+\alpha\gamma)\sigma_1^2 + \gamma\sigma_2^2 \\ (\beta+\alpha\gamma)\sigma_1^2 & \sigma_1^2 & \alpha\sigma_1^2 \\ \alpha(\beta+\alpha\gamma)\sigma_1^2 + \gamma\sigma_2^2 & \alpha\sigma_1^2 & \alpha^2\sigma_1^2 + \sigma_2^2 \end{bmatrix}.$$

(c) Clearly, $\sigma_1^2 = 2$, $\alpha = 1$, and $\sigma_2^2 = 1$. From the formula for $\mathrm{cov}(Y_1, Y_2)$, we now get that $\beta + \gamma = \frac{1}{2}$. Next, from the formula for $\mathrm{cov}(Y_1, Y_3)$, we get that $\gamma = 0$ and hence that $\beta = \frac{1}{2}$. Finally, from the formula for $\mathrm{var}(Y_3)$, we get that $\sigma_3^2 = \frac{1}{2}$.

Solutions to Third Practice Second Midterm Exam

24. (a) Now $Y_1 + Y_2 + Y_3 = 100$, so $Y_1 + Y_2 - Y_3 = Y_1 + Y_2 + Y_1 + Y_2 - 100 = 2(Y_1 + Y_2) - 100$. (Alternatively, $Y_1 + Y_2 - Y_3 = 100 - 2Y_3$.) Observe that $Y_1 + Y_2$ has the binomial distribution with parameters $n = 100$ and $\pi = .55$, which is approximately normally distributed by the Central Limit Theorem, so $Y_1 + Y_2 - Y_3 = 2(Y_1 + Y_2) - 100$ is approximately normally distributed.

(b) Now $P(Y_3 \geq Y_1+Y_2) = P(Y_1+Y_2-Y_3 \leq 0) = P(2(Y_1+Y_2)-100 \leq 0) = P(Y_1+Y_2 \leq 50)$, while Y_1+Y_2 has mean $100(.55) = 55$ and standard deviation $\sqrt{100(.55)(.45)} \doteq 4.975$. Thus, by the Central Limit Theorem with the half-integer correction, $P(Y_3 \geq Y_1+Y_2) = P(Y_1+Y_2 \leq 50) \approx \Phi\left(\frac{50+\frac{1}{2}-55}{4.975}\right) \doteq \Phi(-0.9045) = 1-\Phi(0.9045) \doteq 1-.8171 = .1829$.

25. (a) Now $E[(Y-\widehat{Y})^2] = E[(Y-bX)^2] = E(Y^2-2bXY+b^2X^2) = E(Y^2)-2bE(XY)+b^2E(X^2)$. Since $\frac{d^2}{db^2}E[(Y-\widehat{Y})^2] = 2E(X^2) > 0$, $E[(Y-\widehat{Y})^2]$ is maximized when $0 = \frac{d}{db}E[(Y-\widehat{Y})^2] = 2bE(X^2)-2E(XY)$. Thus, the unique best predictor of the indicated form is given by $\widehat{Y} = \beta X$, where $\beta = \frac{E(XY)}{E(X^2)}$.

(b) The mean squared error of the best predictor of the indicated form is given by $E[(Y-\widehat{Y})^2] = E(Y^2)-2\beta E(XY)+\beta^2E(X^2) = E(Y^2)-\frac{2[E(XY)]^2}{E(X^2)}+\frac{[E(XY)]^2}{E(X^2)} = E(Y^2)-\frac{[E(XY)]^2}{E(X^2)}$.

26. (a) Now $E(Y_2-Y_1) = -2-1 = -3$ and $\operatorname{var}(Y_2-Y_1) = \operatorname{var}(Y_1)+\operatorname{var}(Y_2)-2\operatorname{cov}(Y_1,Y_2) = 1+2+2 = 5$. Since Y_2-Y_1 is normally distributed, $P(Y_1 \geq Y_2) = P(Y_2-Y_1 \leq 0) = \Phi\left(\frac{0-(-3)}{\sqrt{5}}\right) = \Phi\left(\frac{3}{\sqrt{5}}\right) \doteq \Phi(1.342) \doteq .910$.

(b) Now Y_1, Y_2 and $Y_3-aY_1-bY_2$ have a trivariate normal distribution, so $[Y_1,Y_2]^T$ and $Y_3-aY_1-bY_2$ are independent if and only if $\operatorname{cov}(Y_1,Y_3-aY_1-bY_2) = 0$ and $\operatorname{cov}(Y_2,Y_3-aY_1-bY_2) = 0$. Observe that $\operatorname{cov}(Y_1,Y_3-aY_1-bY_2) = \operatorname{cov}(Y_1,Y_3)-a\operatorname{var}(Y_1)-b\operatorname{cov}(Y_1,Y_2) = 1-a+b$ and $\operatorname{cov}(Y_2,Y_3-aY_1-bY_2) = \operatorname{cov}(Y_2,Y_3)-a\operatorname{cov}(Y_1,Y_2)-b\operatorname{var}(Y_2) = -2+a-2b$. The unique solution to the equations $a-b = 1$ and $a-2b = 2$ is given by $b = -1$ and $a = 0$. Thus $[Y_1,Y_2]^T$ and $Y_3-aY_1-bY_2$ are independent if and only if $a = 0$ and $b = -1$.

Solutions to Fourth Practice Second Midterm Exam

28. (a) Now $E(I_1^2) = E(I_1) = 1/3$, $E(I_2^2) = E(I_2) = 1/3$, and $E(I_3^2) = E(I_3) = 1/3$. Thus $\operatorname{var}(I_1) = E(I_1^2)-[E(I_1)]^2 = 1/3-[1/3]^2 = 2/9$ and, similarly, $\operatorname{var}(I_2) = \operatorname{var}(I_3) = 2/9$. Also, $\operatorname{cov}(I_1,I_2) = E(I_1I_2)-[E(I_1)][E(I_2)] = 0-(1/3)(1/3) = -1/9$ and, similarly, $\operatorname{cov}(I_1,I_3) = \operatorname{cov}(I_2,I_3) = -1/9$. Thus the variance-covariance matrix of I_1, I_2 and I_3 equals

$$\begin{bmatrix} 2/9 & -1/9 & -1/9 \\ -1/9 & 2/9 & -1/9 \\ -1/9 & -1/9 & 2/9 \end{bmatrix}.$$

(b) Since $I_1+I_2+I_3 = 1$, it follows from Corollary 5.6 of the textbook that the variance-covariance matrix of I_1, I_2 and I_3 is noninvertible. Alternatively,

$$\begin{bmatrix} 2/9 & -1/9 & -1/9 \\ -1/9 & 2/9 & -1/9 \\ -1/9 & -1/9 & 2/9 \end{bmatrix} \begin{bmatrix} 1 \\ 1 \\ 1 \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \\ 0 \end{bmatrix},$$

which implies that the matrix is noninvertible. As a third approach, the determinant of the matrix is given by

$$\begin{vmatrix} 2/9 & -1/9 & -1/9 \\ -1/9 & 2/9 & -1/9 \\ -1/9 & -1/9 & 2/9 \end{vmatrix} = \frac{1}{9^3}\begin{vmatrix} 2 & -1 & -1 \\ -1 & 2 & -1 \\ -1 & -1 & 2 \end{vmatrix} = \frac{1}{9^3}(8-1-1-2-2-2) = 0,$$

which implies that the matrix is noninvertible.

(c) The random variables I_1 and $I_1 + \beta I_2$ are uncorrelated if and only if

$$0 = \operatorname{cov}(I_1, I_1 + \beta I_2) = \operatorname{var}(I_1) + \beta \operatorname{cov}(I_1, I_2) = \frac{2}{9} - \frac{\beta}{9}$$

and hence if and only if $\beta = 2$.

(d) If a red ball is chosen, then $I_1 = 1$ and $I_1 + 2I_2 = 1$; if a white ball is chosen, then $I_1 = 0$ and $I_1 + 2I_2 = 2$; if a blue ball is chosen, then $I_1 = 0$ and $I_1 + 2I_2 = 0$. Thus $P(I_1 = 1, I_1 + 2I_2 = 0) = 0$, but $P(I_1 = 1) = 1/3$ and $P(I_1 + 2I_2 = 0) = 1/3$, so $P(I_1 = 1, I_1 + 2I_2 = 0) \neq P(I_1 = 1)P(I_1 + 2I_2 = 0)$ and hence I_1 and $I_1 + 2I_2$ are not independent random variables.

29. Now $Y = Y_1Y_2 = \exp(W_1)\exp(W_2) = \exp(W_1 + W_2) = \exp(W)$, where the random variable $W = W_1 + W_2$ has the gamma distribution with shape parameter 2 and scape parameter 1, whose density function is given by $f_W(w) = we^{-w}$ for $w > 0$ and $f_W(w) = 0$ for $w \le 0$. If $y = e^w$, then $w = \log y$ and $dw/dy = 1/y$. Thus $f_Y(y) = f_W(\log y)\frac{dw}{dy} = \frac{1}{y}(\log y)e^{-\log y} = \frac{\log y}{y^2}$ for > 1 and $f_Y(y) = 0$ for $y \le 1$. Alternatively, we can first find the joint density function of $Z_1 = Y_1$ and $Z_2 = Y_1Y_2$. Note that if $z_1 = y_1 = \exp w_1$ and $z_2 = y_1y_2 = \exp(w_1 + w_2)$, then $w_1 = \log z_1$ and $w_2 = \log z_2 - \log z_1$, so $\frac{\partial(w_1,w_2)}{\partial(z_1,z_2)} = \begin{vmatrix} z_1^{-1} & 0 \\ -z_1^{-1} & z_2^{-1} \end{vmatrix} = \frac{1}{z_1z_2}$. Consequently, $f_{Z_1,Z_2}(z_1, z_2) = f_{W_1,W_2}(w_1, w_2)\frac{\partial(w_1,w_2)}{\partial(z_1,z_2)} = \frac{\exp(-(w_1+w_2))}{z_1z_2} = \frac{1}{z_1z_2^2}$ for $1 < z_1 < z_2$ and $f_{Z_1,Z_2}(z_1, z_2) = 0$ elsewhere. Thus the marginal density of $Z_2 = Y_1Y_2$ is given by $f_{Z_2}(z_2) = \int_1^{z_2} \frac{1}{z_1z_2^2}\,dz_1 = \frac{\log z_2}{z_2^2}$ for $z_2 > 1$ and $f_{Z_2}(z_2) = 0$ for $z_2 \le 1$. As a variant of this approach we can first find that the joint density function of Y_1 and Y_2 is given by $f_{Y_1,Y_2}(y_1, y_2) = \frac{1}{y_1^2y_2^2}$ for $y_1, y_2 > 1$ and $f_{Y_1,Y_2}(y_1, y_2) = 0$ elsewhere. Set $Z_1 = Y_1$ and $Z_2 = Y_1Y_2$. Note that if $z_1 = y_1$ and $z_2 = y_1y_2$, then $\frac{\partial(z_1,z_2)}{\partial(y_1,y_2)} = \begin{vmatrix} 1 & 0 \\ y_2 & y_1 \end{vmatrix} = y_1$. Consequently,

$$f_{Z_1,Z_2}(z_1, z_2) = \frac{f_{Y_1,Y_2}(y_1, y_2)}{\frac{\partial(z_1,z_2)}{\partial(y_1,y_2)}} = \frac{1}{y_1^3y_2^2} = \frac{1}{z_1^3(z_2/z_1)^2} = \frac{1}{z_1z_2^2}, \qquad 1 < z_1 < z_2,$$

and $f_{Z_1,Z_2} = 0$ elsewhere, and so forth. As another alternative, we could first find the joint density of $Z_1 = W_1$ and $Z_2 = Y_1Y_1 = \exp(W_1 + W_2)$. Note that if $z_1 = w_1$ and $z_2 = \exp(w_1 + w_2)$, then $w_1 = z_1$ and $w_2 = \log z_2 - z_1$, so $\frac{\partial(w_1,w_2)}{\partial(z_1,z_2)} = \begin{vmatrix} 1 & 0 \\ -1 & z_2^{-1} \end{vmatrix} = \frac{1}{z_2}$. Consequently,

$$f_{Z_1,Z_2}(z_1, z_2) = f_{W_1,W_2}(w_1, w_2)\frac{\partial(w_1, w_2)}{\partial(z_1, z_2)} = \frac{\exp(-(w_1 + w_2))}{z_2} = \frac{1}{z_2^2}$$

for $0 < z_1 < \log z_2$ and $f_{Z_1,Z_2}(z_1, z_2) = 0$ elsewhere. Thus the marginal density of $Z_2 = Y_1Y_2$ is given by $f_{Z_2}(z_2) = \int_0^{\log z_2} \frac{1}{z_2^2}\, dz_1 = \frac{\log z_2}{z_2^2}$ for $z_2 > 1$ and $f_{Z_2}(z_2) = 0$ for $z_2 \le 1$.

30. Now $P(D_{100} > 11) = P(N(11) < 100)$. The random variable $N(100)$ has the Poisson distribution with mean $11^2 = 121$ and variance 121, so it has standard deviation 11. By normal approximation to the Poisson distribution with large mean, $P(D_{100} > 11) = P(N(11) < 100) \approx \Phi\Big(\frac{100-121}{11}\Big) \doteq \Phi(-1.909) \doteq .028$. Alternatively, using the half-integer correction, we get that $P(D_{100} > 11) \approx \Phi\Big(\frac{99.5-121}{11}\Big) \approx \Phi(-1.955) \doteq .0253$. (Actually, $P(N(11) < 100) \doteq .0227$.)

    Observe that D_n does not have a gamma distribution. In particular, D_1 does not have an exponential or other gamma distribution. To see this, note that $P(D_1 > r) = P(N(r) < 1) = P(N(r) = 0) = \exp(-r^2)$, so the density f_1 of D_1 is given by $f_1(r) = 2r\exp(-r^2)$ for $r > 0$ and $f_1(r) = 0$ for $r \le 0$. Thus D_1 has a Weibull distribution, but not an exponential or other gamma distribution.

Solutions to First Practice Final Exam

32. (a) The distribution function of $W = e^V$ is given by $F_W(w) = P(W \le w) = P(e^V \le w) = P(V \le \log w) = 1 - e^{-\log w} = 1 - 1/w$ for $w > 1$ and $F_W(w) = 0$ for $w \le 1$. The density function of W is given by $f_W(w) = 0$ for $w \le 1$ and $f_W(w) = F'_W(w) = 1/w^2$ for $w > 1$. The pth quantile of W is given by $1 - 1/w_p = p$ or $w_p = 1/(1-p)$.

    (b) The distribution function of V is given by $F_V(v) = 1 - e^v(1+v)$ for $v > 0$ and $F_V(v) = 0$ for $v \le 0$. Thus the distribution function of Y is given by $F_Y(y) = P(Y \le y) = P(e^V \le y) = P(V \le \log y) = 1 - e^{-\log y}(1+\log y) = 1 - \frac{1+\log y}{y}$ for $y > 1$ and $F_Y(y) = 0$ for $y \le 1$. The density function of Y is given by $f_Y(y) = F'_Y(y) = -\frac{1}{y^2} + \frac{1+\log y}{y^2} = \frac{\log y}{y^2}$ for $y > 1$ and $f_Y(y) = 0$ for $y \le 1$.

33. (a) Suppose that $y_1 = w_1w_2$ and $y_2 = w_1/w_2$, where $w_1, w_2 > 0$. Then $w_1 = \sqrt{y_1y_2}$ and $w_2 = \sqrt{y_1/y_2}$. Thus

$$\frac{\partial(w_1, w_2)}{\partial(y_1, y_2)} = \begin{vmatrix} \frac{1}{2}\sqrt{y_2/y_1} & \frac{1}{2}\sqrt{y_1/y_2} \\ \frac{1}{2}\sqrt{1/(y_1y_2)} & -\frac{1}{2}\sqrt{y_1/y_2^3} \end{vmatrix} = -\frac{1}{2y_2},$$

    so $f_{Y_1,Y_2}(y_1, y_2) = \frac{1}{2y_2} f_{W_1,W_2}(\sqrt{y_1y_2}, \sqrt{y_1/y_2})$ for $y_1, y_2 > 0$.

    (b) Suppose $y_1 = w_1w_2$ and $y_2 = w_1/w_2$, where $w_1, w_2 > 1$. Then $y_1 > 1$ and $\frac{1}{y_1} < y_2 < y_1$, so it follows from the solution to (a) that $f_{Y_1}(y_1) = \int_{1/y_1}^{y_1} \frac{1}{2y_2} f_{W_1}(\sqrt{y_1y_2} f_{W_2}(\sqrt{y_1/y_2})\, dy_2$ for $y_1 > 1$.

    (c) $f_{Y_1}(y_1) = \int_{1/y_1}^{y_1} \frac{1}{2y_2} \frac{1}{y_1y_2} \frac{1}{y_1/y_2}\, dy_2 = \frac{1}{2y_1^2} \int_{1/y_1}^{y_1} \frac{1}{y_2}\, dy_2 = \frac{\log y_1}{y_1^2}$ for $y > 1$.

(d) Let V_1 and V_2 be independent random variables each having the exponential distribution with mean 1. Then $V_1 + V_2$ has the gamma distribution with shape parameter 2 and scale parameter 1. According to Problem 32(a), $W_1 = e^{V_1}$ and $W_2 = e^{V_2}$ are independent random variables having the common density function given by $f(w) = 1/w^2$ for $w > 1$. Thus, by Problem 32(b), the density function of $Y_1 = W_1 W_2 = \exp(V_1 + V_2)$ is given by $f_{Y_1}(y_1) = (\log y_1)/y_1^2$ for $y_1 > 1$, which agrees with the answer to (c).

(e) Observe that $y_1 > 1$ and $1/y_1 < y_2 < y_1$ if and only if $y_2 > 0$ and $y_1 > \max(y_2, 1/y_2)$. Thus if $0 < y_2 \le 1$, then $y_1 \ge 1/y_2$ and if $y_2 > 1$, then $y_1 > y_2$. Consequently, $f_{Y_2}(y_2) = \int_{1/y_2}^{\infty} \frac{1}{2y_1^2 y_2}\, dy_1 = \frac{1}{2}$ for $0 < y_2 \le 1$ and $f_{Y_2}(y_2) = \int_{y_2}^{\infty} \frac{1}{2y_1^2 y_2}\, dy_1 = \frac{1}{2y_2^2}$ for $y_2 > 1$.

34. (a) $E[(X-c)(Y-d)] = E[(X - EX + EX - c)(Y - EY + EY - d)] = E[(X-EX)(Y-EY)] + (EX-c)(EY-d) + (EX-c)E(Y-EY) + (EY-d)E(X-EX) = \operatorname{cov}(X,Y) + (EX-c)(EY-d)$.

(b) $\operatorname{cov}(\bar{X}, \bar{Y}) = \frac{1}{n^2}\operatorname{cov}(X_1 + \cdots + X_n, Y_1 + \cdots + Y_n) = \frac{1}{n^2}[\operatorname{cov}(X_1, Y_1) + \cdots + \operatorname{cov}(X_n, Y_n)] = \frac{1}{n^2}\, n \operatorname{cov}(X,Y) = \frac{1}{n}\operatorname{cov}(X,Y)$.

35. (a) Since L_1 is uniformly distributed on $\{1, \ldots, N\}$, $EU_1 = E[u(L_1)] = \bar{u}$ and $EV_1 = E[v(L_1)] = \bar{v}$, so $C = \operatorname{cov}(U_1, V_1) = E[(U_1 - \bar{u})(V_1 - \bar{v})] = \frac{1}{N}\sum_{l=1}^{N}(u_l - \bar{u})(v_l - \bar{v})$.

(b) L_i has the same distribution as L_1 for $2 \le i \le N$.

(c) (L_i, L_j) has the same distribution as (L_1, L_2) for $i \ne j$.

(d) $\operatorname{cov}(S_n, T_n) = \operatorname{cov}(U_1 + \cdots + U_n, V_1 + \cdots + V_n) = n\operatorname{cov}(U_1, V_1) + n(n-1)\operatorname{cov}(U_1, V_2) = nC + n(n-1)D$.

(e) Since $S_N = U_1 + \cdots + U_N = u_1 + \cdots + u_N$ is a constant random variable, $\operatorname{cov}(S_N, T_N) = 0$.

(f) $0 = NC + N(N-1)D$, so $D = -C/(N-1)$.

(g) $\operatorname{cov}(S_n, T_n) = nC + n(n-1)D = nC\left(1 - \frac{n-1}{N-1}\right) = nC\frac{N-n}{N-1}$.

36. (a) The random vector $[Y_1, Y_2, Y_3]^T$ has the trivariate normal distribution with mean vector $[0,0,0]^T$ and variance-covariance matrix $\begin{bmatrix} 1 & 1 & 1 \\ 1 & 2 & 2 \\ 1 & 2 & 3 \end{bmatrix}$.

(b) The inverse of the variance-covariance matrix of $[Y_1, Y_3]^T$ is given by $\begin{bmatrix} 1 & 1 \\ 1 & 3 \end{bmatrix}^{-1} = \frac{1}{2}\begin{bmatrix} 3 & -1 \\ -1 & 1 \end{bmatrix} = \begin{bmatrix} 3/2 & -1/2 \\ -1/2 & 1/2 \end{bmatrix}$. Thus the conditional distribution of Y_2 given that $Y_1 = y_1$ and $Y_2 = y_2$ is normal with mean $\beta_1 y_1 + \beta_3 y_3$ and variance σ^2, where $\begin{bmatrix} \beta_1 \\ \beta_3 \end{bmatrix} = \begin{bmatrix} 3/2 & -1/2 \\ -1/2 & 1/2 \end{bmatrix}\begin{bmatrix} 1 \\ 2 \end{bmatrix} = \begin{bmatrix} 1/2 \\ 1/2 \end{bmatrix}$ and $\sigma^2 = 2 - \begin{bmatrix} 1/2 & 1/2 \end{bmatrix}\begin{bmatrix} 1 \\ 1 \end{bmatrix} = 2 - \frac{3}{2} = \frac{1}{2}$.

(c) The best predictor of Y_2 based on Y_1 and Y_3 coincides with the corresponding best linear predictor, which is given by $\widehat{Y}_2 = (Y_1 + Y_3)/2$. The mean squared error of this predictor equals $1/2$.

37. (a) There is a one-to-one correspondence between the distinct sequences $s_1, \ldots, s_n$ of positive integers such that $s_1 < \cdots < s_n \le s$ and the subsets of $\{1, \ldots, s\}$ of size n. Since there are $\binom{s}{n}$ the indicated subsets, there are the same number of the indicated sequences.

(b) $P(Y_1 - 1 = y_1 - 1, \ldots, Y_{n-1} - 1 = y_{n-1} - 1 \mid Y_1 + \cdots + Y_n - n = s - n) = \frac{P(Y_1-1=y_1-1,\ldots,Y_n-1=y_n-1)}{P(Y_1+\cdots+Y_n-n=s-n)} = \frac{\prod_{i=1}^n [(1-\pi)\pi^{y_i-1}]}{\frac{\Gamma(s)}{\Gamma(n)(s-n)!}(1-\pi)^n \pi^{s-n}} = \frac{(n-1)!\,(s-n)!}{(s-1)!} = 1/\binom{s-1}{n-1}$.

(c) Let $s_1, \ldots, s_{n-1}$ be positive integers such that $s_1 < \cdots < s_{n-1} \le s - 1$ and set $y_1 = s_1, y_2 = s_2 - s_1, \ldots, y_{n-1} = s_{n-1} - s_{n-2}, y_n = s - s_{n-1}$. Then $y_1, \ldots, y_n$ are positive integers adding up to s. Thus, by (b), $P(\{Y_1, Y_1 + Y_2, \ldots, Y_1 + \cdots + Y_{n-1}\} = \{s_1, s_2, \ldots, s_{n-1}\} \mid Y_1 + \cdots + Y_n = s\} = P(Y_1 = y_1, \ldots, Y_{n-1} = y_{n-1} \mid Y_1 + \cdots + Y_n = s) = 1/\binom{s-1}{n-1}$.

Solutions to Second Practice Final Exam

39. (a) Let X_i denote the amount won by Frankie on the ith trial, and let Y_i and Z_i be defined similarly for Johnny and Sammy. Then each of these random variables takes on each of the values -1 and 1 with probability 1/2 and hence has mean zero and variance 1. Since $X_1, \ldots, X_{10}$ are independent random variables, we conclude that $X = X_1 + \cdots + X_{10}$ has mean 0 and variance 10. Similarly, Y and Z have mean zero and variance 10.

(b) Now $E(X_i Y_i) = -1$, so $\operatorname{cov}(X_i, Y_i) = -1$. Since the random pairs (X_i, Y_i), $1 \le i \le 10$, are independent, we conclude that $\operatorname{cov}(X, Y) = -10$. On the other hand, $E(X_i Z_i) = 0$, so $\operatorname{cov}(X_i, Z_i) = 0$ and hence $\operatorname{cov}(X, Z) = 0$. Similarly, $\operatorname{cov}(Y, Z) = 0$.

(c) Now $E(X+Y+Z) = EX + EY + EZ = 0$. Moreover, $\operatorname{cov}(X+Y+Z) = \operatorname{var}(X) + \operatorname{var}(Y) + \operatorname{var}(Z) + 2\operatorname{cov}(X,Y) + 2\operatorname{cov}(X,Z) + 2\operatorname{cov}(Y,Z) = 10 + 10 + 10 - 2 \cdot 10 = 10$. Alternatively, $Y = -X$, so $X + Y + Z = Z$ and hence $E(X+Y+Z) = E(Z) = 0$ and $\operatorname{var}(X+Y+Z) = \operatorname{var}(Z) = 10$.

40. (a) Think of the outcome of each trial as being of one of three mutually exclusive types: A, B, and C. Then we have n independent trials and the outcome of each trial is of type A, B, or C with probability π_1, π_2, or π_3, respectively. Since Y_1 is the number of type A outcomes on the n trials, Y_2 is the number of type B outcomes, and Y_3 is the number of type C outcomes, we conclude that the joint distribution of Y_1, Y_2, and Y_3 is trinomial with parameters n, π_1, π_2, and π_3.

(b) Let y_1, y_2, and y_3 be nonnegative integers, and set $n = y_1 + y_2 + y_3$. Then $P(Y_1 = y_1, Y_2 = y_2, Y_3 = y_3) = P(Y_1 = y_1, Y_2 = y_2, Y_3 = y_3, N = n) = P(N = n)P(Y_1 = y_1, Y_2 = y_2, Y_3 = y_3 | N = n) = \frac{\lambda^n}{n!}e^{-\lambda}\frac{n!}{y_1!\,y_2!\,y_3!}\pi_1^{y_1}\pi_2^{y_2}\pi_3^{y_3} = \frac{(\lambda\pi_1)^{y_1}}{y_1!}e^{-\lambda\pi_1}\frac{(\lambda\pi_2)^{y_2}}{y_2!}e^{-\lambda\pi_2}\frac{(\lambda\pi_3)^{y_3}}{y_3!}e^{-\lambda\pi_3}$, so Y_1, Y_2, and Y_3 are independent random random variables having Poisson distributions with means $\lambda\pi_1$, $\lambda\pi_2$, and $\lambda\pi_3$, respectively.

41. (a) The marginal density function of X is given by

$$f_X(x) = 2e^{-x}\int_0^\infty ye^{-y^2/x}\,dy = xe^{-x}\int_0^\infty d(-e^{-y^2/x}) = xe^{-x}, \qquad x > 0,$$

and $f_X(x) = 0$ for $x \le 0$. Thus X has the gamma distribution with shape parameter 2 and scale parameter 1, which has mean 2 and variance 2.

(b) The conditional density function of Y given that $X = x > 0$ is given by $f_{Y|X}(y|x) = \frac{2ye^{-(x^2+y^2)/x}}{xe^{-x}} = \frac{2y}{x}e^{-y^2/x}$ for $y > 0$ and $f_{Y|X}(y|x) = 0$ for $y \le 0$.

(c) Let $x > 0$. It follows from the answer to (b) that $E(Y|X = x) = \int_0^\infty \frac{2y^2}{x}e^{-y^2/x}\,dy$. Setting $t = y^2/x$, we get that $dt = 2y/x\,dy$ and $y = (xt)^{1/2}$ and hence that $E(Y|X = x) = \int_0^\infty (xt)^{1/2}e^{-t}\,dt = x^{1/2}\Gamma(3/2) = \frac{\sqrt{\pi x}}{2}$. Similarly, $E(Y^2|X = x) = \int_0^\infty \frac{2y^3}{x}e^{-y^2/x}\,dy = \int_0^\infty (xt)e^{-t}\,dt = x$ and hence $\mathrm{var}(Y|X = x) = x(1 - \pi/4)$.

(d) The best predictor of Y based on X is given by $\widehat{Y} = E(Y|X) = \sqrt{\pi X}/2$. The mean squared error of this predictor is given by

$$\mathrm{MSE}(\widehat{Y}) = E[\mathrm{var}(Y|X)] = E[X(1 - \pi/4)] = 2 - \pi/2.$$

(e) The mean of Y is given by

$$EY = E[E(Y|X)] = E[\sqrt{\pi X}/2] = (\sqrt{\pi}/2)E(\sqrt{X}).$$

Since $E(\sqrt{X}) = \int_0^\infty x^{1/2}xe^{-x}\,dx = \int_0^\infty x^{3/2}e^{-x}\,dx = \Gamma(5/2) = 3\sqrt{\pi}/4$, we conclude that $EY = 3\pi/8$. By Corollary 6.2, $\mathrm{var}(Y) = E[\mathrm{var}(Y|X)] + \mathrm{var}(E(Y|X)) = 2 - \pi/2 + (\pi/4)\mathrm{var}(\sqrt{X})$. Now $\mathrm{var}(\sqrt{X}) = EX - (E\sqrt{X})^2 = 2 - (3\sqrt{\pi}/4)^2 = 2 - 9\pi/16$, so $\mathrm{var}(Y) = 2 - \pi/2 + \pi(2 - 9\pi/16)/4 = 2 - 9\pi^2/64$.

42. (a) The desired probability is given by $P(N_1 = 2, N_2 = 5, N_3 = 9) = \frac{7}{10}\cdot\frac{3}{9}\cdot\frac{6}{8}\cdot\frac{5}{7}\cdot\frac{2}{6}\cdot\frac{4}{5}\cdot\frac{3}{4}\cdot\frac{2}{3}\cdot\frac{1}{2} = \frac{6}{10\cdot 9\cdot 8} = \frac{1}{120}$.

(b) Observe that $P(N_1 = n_1, N_2 = n_2, N_3 = n_3)$ is the probability that the set of ten trials on which the red balls are drawn equals $\{n_1, n_2, n_3\}$. Since there are $\binom{10}{3}$ such subsets, which are equally likely,

$$P(N_1 = n_1, N_2 = n_2, N_3 = n_3) = \frac{1}{\binom{10}{3}} = \frac{6}{10\cdot 9\cdot 8} = \frac{1}{120}.$$

(c) The desired probability is given by $P(N_1 = 2, N_2 = 5, N_3 = 9) = \left(\frac{7}{10}\right)\left(\frac{3}{10}\right)\left(\frac{7}{10}\right)^2\left(\frac{3}{10}\right)\left(\frac{7}{10}\right)^3\left(\frac{3}{10}\right)\left(\frac{7}{10}\right)^3\left(\frac{3}{10}\right) = \left(\frac{3}{10}\right)^3\left(\frac{7}{10}\right)^6$.

(d) The desired probability is given by $P(N_1 = n_1, N_2 = n_2, N_3 = n_3) = \left(\frac{7}{10}\right)^{n_1-1}\left(\frac{3}{10}\right)\left(\frac{7}{10}\right)^{n_2-n_1-1}\left(\frac{3}{10}\right)\left(\frac{7}{10}\right)^{n_3-n_2-1}\left(\frac{3}{10}\right) = \left(\frac{3}{10}\right)^3\left(\frac{7}{10}\right)^{n_3-3}$.

43. (a) By Corollary 6.8, the joint distribution of Y_1 and Y_2 is bivariate normal with mean vector $\mathbf{0}$ and variance-covariance matrix

$$\begin{bmatrix} \sigma_1^2 & \beta_1\sigma_1^2 \\ \beta_1\sigma_1^2 & \beta_1^2\sigma_1^2 + \sigma_2^2 \end{bmatrix}.$$

(b) Observe that $\begin{bmatrix} \sigma_1^2 & \beta_1\sigma_1^2 \\ \beta_1\sigma_1^2 & \beta_1^2\sigma_1^2+\sigma_2^2 \end{bmatrix} \begin{bmatrix} \beta_2 \\ \beta_3 \end{bmatrix} = \begin{bmatrix} (\beta_2+\beta_1\beta_3)\sigma_1^2 \\ (\beta_1\beta_2+\beta_1^2\beta_3)\sigma_1^2+\beta_3\sigma_2^2 \end{bmatrix}$ and hence that $\begin{bmatrix} \beta_2 & \beta_3 \end{bmatrix} \begin{bmatrix} \sigma_1^2 & \beta_1\sigma_1^2 \\ \beta_1\sigma_1^2 & \beta_1^2\sigma_1^2+\sigma_2^2 \end{bmatrix} \begin{bmatrix} \beta_2 \\ \beta_3 \end{bmatrix} = (\beta_2^2+2\beta_1\beta_2\beta_3+\beta_1^2\beta_3^2)\sigma_1^2+\beta_3^2\sigma_2^2$. Consequently, by Theorem 6.7, the joint distribution of Y_1, Y_2, and Y_3 is trivariate normal with mean vector $\mathbf{0}$ and variance-covariance matrix

$$\begin{bmatrix} \sigma_1^2 & \beta_1\sigma_1^2 & (\beta_2+\beta_1\beta_3)\sigma_1^2 \\ \beta_1\sigma_1^2 & \beta_1^2\sigma_1^2+\sigma_2^2 & (\beta_1\beta_2+\beta_1^2\beta_3)\sigma_1^2+\beta_3\sigma_2^2 \\ (\beta_2+\beta_1\beta_3)\sigma_1^2 & (\beta_1\beta_2+\beta_1^2\beta_3)\sigma_1^2+\beta_3\sigma_2^2 & (\beta_2^2+2\beta_1\beta_2\beta_3+\beta_1^2\beta_3^2)\sigma_1^2+\beta_3^2\sigma_2^2+\sigma_3^2 \end{bmatrix}.$$

Solutions to Third Practice Final Exam

45. (a) Observe that N_1 and N_2 are positive integer-valued random variables. Let n_1 and m_2 be positive integers. Then $P(N_1 = n_1,\ N_2 - N_1 = m_2) = P(U_i < 1/4 \text{ for } 1 \le i < n_1,\ U_{n_1} \ge 1/4,\ U_i < 1/4 \text{ for } n_1+1 \le i < n_1+m_2,\ U_{n_1+m_2} \ge 1/4) = \big(1/4\big)^{n_1+m_2-2}\big(3/4\big)^2 = \big(1/4\big)^{n_1-1} 1/4 \,\big(1/4\big)^{m_2-1} 3/4$. Thus N_1 and $N_2 - N_1$ are independent. (Moreover, each of $N_1 - 1$ and $N_2 - 1$ has the geometric distribution with parameter $\pi = 1/4$.)

    (b) Observe that N_2 is an integer-valued random variable and that $N_2 \ge 2$. Let n_2 be an integer with $n_2 \ge 2$. Then $P(N_2 = n_2) = \sum_{n_1=1}^{n_2-1} P(N_1 = n_1,\ N_2 - N_1 = n_2 - n_1)$ $= \sum_{n_1=1}^{n_2-1} \big(1/4\big)^{n_2-2}\big(3/4\big)^2 = (n_2-1)\big(1/4\big)^{n_2-2}\big(3/4\big)^2$. Thus $N_2 - 2$ has the negative binomial distribution with parameters $\alpha = 2$ and $\pi = 1/4$.

46. (a) $1 = c\int_{-1}^1 y^2(1-y^2)\,dy = c\int_{-1}^1 (y^2-y^4)\,dy = c\big(y^3/3 - y^5/5\big)\big|_{-1}^1 = (4c)/15$, so $c = 15/4$.

    (b) $EY = (15/4)\int_{-1}^1 y^3(1-y^2)\,dy = 0$ and $\mathrm{var}(Y) = E(Y^2) = (15/4)\int_{-1}^1 y^4(1-y^2)\,dy = (15/4)\big(y^5/5 - y^7/7\big)\big|_{-1}^1 = 3/7$.

    (c) Now $\bar{Y}$ has mean 0 and variance $3/700$. The upper quartile of the standard normal distribution is given by $z_{.75} \doteq 0.674$. Thus, by the central limit theorem, the upper quartile of $\bar{Y}$ is given approximately by $0.674\sqrt{3/700} \doteq 0.0441$.

47. Now $\mathrm{var}(\widehat{\mu}_2^{(1)}) = \mathrm{var}(\bar{Y}) = n^{-1}\sigma_2^2$ and $\mathrm{var}(\widehat{\mu}_2^{(2)}) = \mathrm{var}(\bar{Y} - \bar{X}) = n^{-1}\mathrm{var}(Y - X) = n^{-1}(\sigma_1^2 + \sigma_2^2 - 2\rho\sigma_1\sigma_2)$. Thus $\mathrm{var}(\widehat{\mu}_2^{(2)}) < \mathrm{var}(\widehat{\mu}_2^{(1)})$ if and only if $\sigma_1^2 + \sigma_2^2 - 2\rho\sigma_1\sigma_2 < \sigma_2^2$ or, equivalently, $\sigma_1 < 2\rho\sigma_2$.

48. (a) Now $E(\bar{Y}_n) = \mu$ and $\mathrm{var}(\bar{Y}_n) = \frac{\mathrm{var}(Y_1)+\cdots+\mathrm{var}(Y_n)}{n^2} = \frac{1^b+\cdots+n^b}{n^2}$. Let $c > 0$. Then, by Chebyshev's inequality, $P(\bar{Y}_n - \mu| \ge c) \le \frac{\mathrm{var}(\bar{Y}_n)}{c^2} = \frac{1^b+\cdots+n^b}{c^2n^2}$. Thus if $b < 1$, then $\lim_{n\to\infty} P(|\bar{Y}_n - \mu| \ge c) \le \frac{1}{c^2(b+1)}\lim_{n\to\infty}\Big(\frac{1^b+\cdots+n^b}{n^{b+1}/(b+1)}\cdot n^{b-1}\Big) = \frac{1}{c^2(b+1)}\lim_{n\to\infty} n^{b-1} = 0$.

    (b) Suppose that $Y_1, Y_2, \ldots$ are independent and normally distributed. Then $\bar{Y}_n - \mu$ is normally distributed with mean 0 and variance $(1^b+\cdots+n^b)/n^2$. Thus $P(|\bar{Y}_n - \mu| \ge c) = 2\Phi\Big(-\frac{c}{\sqrt{(1^b+\cdots+n^b)/n^2}}\Big)$. Consequently, if (1) holds, then $0 = \lim_{n\to\infty} \frac{1^b+\cdots+n^b}{n^2} = \frac{1}{b+1} \lim_{n\to\infty} \left(\frac{1^b+\cdots+n^b}{n^{b+1}/(b+1)} \cdot n^{b-1}\right) = \frac{1}{b+1} \lim_{n\to\infty} n^{b-1}$ and hence $b < 1$.

(c) The assumptions of independence and normality are required to conclude that $Y_1 + \cdots + Y_n$ is normally distributed and hence that $\bar{Y}_n$ is normally distributed.

49. (a) The solution to the equations $y_1 = x_1 + x_2 + x_3$, $y_2 = \frac{x_1+x_2}{x_1+x_2+x_3}$, and $y_3 = \frac{x_1}{x_1+x_2}$ is given by $x_1 = y_1 y_2 y_3$, $x_2 = y_1 y_2(1-y_3)$, and $x_3 = y_1(1-y_2)$. Thus

$$\frac{\partial(x_1, x_2, x_3)}{\partial(y_1, y_2, y_3)} = \begin{vmatrix} y_2 y_3 & y_1 y_3 & y_1 y_2 \\ y_2(1-y_3) & y_1(1-y_3) & -y_1 y_2 \\ 1-y_2 & -y_1 & 0 \end{vmatrix} = -y_1^2 y_2.$$

Consequently,

$$f_{Y_1,Y_2,Y_3}(y_1, y_2, y_3) = y_1^2 y_2 f_{X_1,X_2,X_3}(y_1 y_2 y_3, y_1 y_2(1-y_3), y_1(1-y_2))$$

for $y_1 > 0$ and $0 < y_1, y_2 < 1$, and $f_{Y_1,Y_2,Y_3}(y_1, y_2, y_3) = 0$ otherwise.

(b) Suppose that X_1, X_2 and X_3 are independent random variables, each having the exponential distribution with inverse-scale parameter λ. Then Y_1 is a positive random variables and y_2 and Y_3 are $(0,1)$-valued random variables. Let $y_1 > 0$ and $0 < y_2, y_3 < 1$. Then $f_{Y_1,Y_2,Y_3}(y_1, y_2, y_3) = y_1^2 y_2 \lambda^3 e^{-\lambda[y_1 y_2 y_3 + y_1 y_2(1-y_3) + y_1(1-y_2)]} = y_1^2 y_2 \lambda^3 e^{-\lambda y_1} = \frac{\lambda^3 y_1^2 e^{-\lambda y_1}}{2} \cdot 2y_2 \cdot 1$. Thus Y_1, Y_2 and Y_3 are independent, Y_1 has the gamma distribution with shape parameter 3 and inverse-scale parameter λ, Y_2 has the density $2y_2$ on $(0,1)$, and Y_3 is uniformly distributed on $(0,1)$.

50. (a) The distribution of X_1 is normal with mean 1 and variance 1.

(b) The conditional distribution of Y given that $X_1 = x_1$ is normal with mean $3 + (x_1 - 1) = 2 + x_1$ and variance $3 - 1 = 2$.

(c) The joint distribution of X_1 and X_2 is bivariate normal with mean vector $[1, -2]^T$ and variance-covariance matrix $\begin{bmatrix} 1 & -1 \\ -1 & 2 \end{bmatrix}$.

(d) The inverse of the variance-covariance matrix of X_1 and X_2 is given by $\begin{bmatrix} 1 & -1 \\ -1 & 2 \end{bmatrix}^{-1} = \begin{bmatrix} 2 & 1 \\ 1 & 1 \end{bmatrix}$. Thus the conditional distribution of Y given that $X_1 = x_1$ and $X_2 = x_2$ is normal with mean

$$3 + \begin{bmatrix} 1 & -2 \end{bmatrix} \begin{bmatrix} 2 & 1 \\ 1 & 1 \end{bmatrix} \begin{bmatrix} x_1 - 1 \\ x_2 + 2 \end{bmatrix} = 3 + 0(x_1 - 1) + (-1)(x_2 + 2) = 1 - x_2$$

and variance $3 - (-1)\text{cov}(X_2, Y) = 3 - 2 = 1$.

(e) It follows from the solution to (b) that $\widehat{Y}^{(1)} = 2 + X_1$. It follows from the solution to (d) that $\widehat{Y}^{(2)} = 1 - X_2$. Moreover, $\widehat{X}_2 = -2 + (-1)(X_1 - 1) = -1 - X_1$. Thus $1 - \widehat{X}_2 = 1 - (-1 - X_1) = 2 + X_1 = \widehat{Y}^{(1)}$, which illustrates Theorem 5.17.

Solutions to Fourth Practice Final Exam

52. The common mean of $Y_1, \ldots, Y_n$ is given by $\mu = \int_0^\infty 2y^2 \exp(-y^2)\,dy$. Making the change of variables $t = y^2$ with $dt = 2y\,dy$ and $y = \sqrt{t}$, we get that $\mu = \int_0^\infty \sqrt{t}\exp(-t)\,dt = \Gamma(3/2) = \frac{1}{2}\sqrt{\pi} \doteq 0.8862$. The common second moment of $Y_1, \ldots, Y_n$ is given by $\int_0^\infty 2y^3 \exp(-y^2)\,dy = \int_0^\infty t\exp(-t)\,dt = \Gamma(2) = 1$. Thus the common of these random variables is given by $\sigma = \sqrt{1 - \pi/4} \doteq 0.4633$. According to the central limit theorem, the upper decile of $Y_1 + \cdots + Y_{100}$ can be reasonably well approximated by that of the normal distribution with mean 100μ and standard deviation 10σ; that is, by $100\mu + 10\sigma z_{.9} \doteq 100(0.8862) + 10(0.4633)(1.282) \doteq 94.56$.

53. (a) Now $P(Y_1 = 1) = P(\mathrm{R}) = 1/2$ and $P(Y_1 = 0) = P(\mathrm{W}) = 1/2$, so Y_1 is uniformly distributed on $\{0, 1\}$. Next, $P(Y_2 = 2) = P(\mathrm{RR}) = \frac{1}{2} \cdot \frac{2}{3} = \frac{1}{3}$. Similarly, $P(Y_2 = 0) = P(\mathrm{WW}) = \frac{1}{3} \cdot \frac{2}{3} = \frac{2}{3}$, so $P(Y_2 = 1) = 1 - \frac{1}{3} - \frac{1}{3} = \frac{1}{3}$ and hence Y_2 is uniformly distributed on $\{0, 1, 2\}$. Further, $P(Y_3 = 3) = P(\mathrm{RRR}) = \frac{1}{2} \cdot \frac{2}{3} \cdot \frac{3}{4} = \frac{1}{4}$ and $P(Y_3 = P(\mathrm{WWW}) = \frac{1}{3} \cdot \frac{2}{3} \cdot \frac{3}{4} = \frac{1}{4}$. By symmetry, $P(2$ red balls and 1 white ball $= P(1$ red ball and 2 white balls$) = \frac{1 - 1/4 - 1/4}{2} = \frac{1}{4}$, so Y_3 is uniformly distributed on $\{0, 1, 2, 3\}$. Finally, $P(Y_4 = 4) = P(\mathrm{RRRR}) = \frac{1}{2} \cdot \frac{2}{3} \cdot \frac{3}{4} \cdot \frac{4}{5} = \frac{1}{5}$. By symmetry, $P(Y_4 = 0) = P(\mathrm{WWWW}) = \frac{1}{5}$. Also, $P(\mathrm{RRRW}) = \frac{1}{2} \cdot \frac{2}{3} \cdot \frac{3}{4} \cdot \frac{1}{5} = \frac{3!\cdot 1!}{5!} = \frac{1}{20}$. Similarly, $P(\mathrm{RRWR}) = P(\mathrm{RWRR}) = P(\mathrm{WRRR}) = \frac{1}{20}$, so $P(Y_4 = 3) = 4 \cdot \frac{1}{20} = \frac{1}{5}$. Consequently, $P(Y_4 = 2) = 1 - \frac{1}{5} - \frac{1}{5} - \frac{1}{5} - \frac{1}{5} = \frac{1}{5}$ and hence Y_4 is uniformly distributed on $\{0, 1, 2, 3, 4\}$. Alternatively, $P(\mathrm{RRWW}) = \frac{1}{2} \cdot \frac{2}{3} \cdot \frac{1}{4} \cdot \frac{2}{5} = \frac{2!\cdot 2}{5!} = \frac{1}{30}$. Similarly, each of the orderings involving 2 reds and 2 whites has probability $\frac{1}{30}$. Since there are $\binom{4}{2} = \frac{4!}{2!\,2!} = 6$ such orderings, we conclude that $P(Y_4 = 2) = 6 \cdot \frac{1}{30} = \frac{1}{5}$.

(b) Observe that the probability of getting R y times followed by W $n - y$ times is given by $\frac{1}{2} \cdot \frac{2}{3} \cdots \frac{y}{y+1} \cdot \frac{1}{y+2} \cdot \frac{2}{y+3} \cdots \frac{n-y}{n+1} = \frac{y!\,(n-y)!}{(n+1)!}$. Similarly, any other ordering of y reds and $n - y$ whites has this probability. Since there are $\binom{n}{y}$ such orderings, $P(Y_n = y) = \binom{n}{y}\frac{y!\,(n-y)!}{(n+1)!} = \frac{n!}{y!\,(n-y)!}\frac{y!\,(n-y)!}{(n+1)!} = \frac{1}{n+1}$. Thus Y_n is uniformly distributed on $\{0, \ldots, n\}$.

Alternatively, the desired result can be proved by induction. To this end, we note first that, by (a), the result is true for $n = 1$. Suppose the result is true for the positive integer n; that is, that $P(Y_n = i) = 1/(n+1)$ for $1 \le i \le n$. Let $0 \le i \le n+1$. Then $P(Y_{n+1} = i) = P(Y_n = i-1)\frac{i}{n+2} + P(Y_n = i)\frac{n+1-i}{n+2} = \frac{1}{n+1}\frac{n+1}{n+2} = \frac{1}{n+2}$. Therefore Y_{n+1} is uniformly distributed on $\{0, \ldots, n+1\}$. Thus the result is true for $n \ge 1$ by induction.

54. (a) By Corollary 3.1, Y_n has the gamma distribution with shape parameter n and scale parameter β, whose density function is given by $f_{Y_n}(y) = \frac{y^{n-1}}{\beta^n (n-1)!} e^{-y/\beta}$ for $y > 0$ and $f_{Y_n}(y) = 0$ for $y \le 0$.

(b) The density function of Y_N is given by $f_{Y_N}(y) = \sum_{n=1}^\infty P(N = n) f_{Y_n}(y) = \sum_{n=1}^\infty (1-\pi)\pi^{n-1} \frac{y^{n-1}}{\beta^n (n-1)!} e^{-y/\beta} = \frac{1-\pi}{\beta} e^{-y/\beta} \sum_{n=1}^\infty \frac{(\pi y/\beta)^{n-1}}{(n-1)!}$ $= \frac{1-\pi}{\beta} e^{-y/\beta} \sum_{n=0}^\infty \frac{(\pi y/\beta)^n}{n!} = \frac{1-\pi}{\beta} e^{-y/\beta} e^{\pi y/\beta} = \frac{1-\pi}{\beta} e^{-y(1-\pi)/\beta}$ for $y > 0$ and $f_{y_N}(y) = 0$ for $y \le 0$. Thus Y_N has the exponential distribution with scale parameter $\beta/(1-\pi)$.

(c) $E(Y_N) = \beta/(1-\pi)$ and $\mathrm{var}(Y_N) = [\beta/(1-\pi)]^2$. Alternatively, Y_n has mean $n\beta$ and variance $n\beta^2$. Also $E(N) = 1 + \pi/(1-\pi) = 1/(1-\pi)$ and $\mathrm{var}(N) = \pi/(1-\pi)^2$. Thus $E(Y_N) = E[E(Y_N|N)] = E(N\beta) = \beta/(1-\pi)$ and $\mathrm{var}(Y_N) = E[\mathrm{var}(Y_N|N)] + \mathrm{var}[E(Y_N|N)] = E[\beta^2 N] + \mathrm{var}(\beta N) = \frac{\beta^2}{1-\pi} + \beta^2 \frac{\pi}{(1-\pi)^2} = \left(\frac{\beta}{1-\pi}\right)^2$.

(d) Since each of the random variables $X_1, X_2, \dots$ has mean 2β and variance $2\beta^2$, Y_n has mean $2\beta n$ and variance $2\beta^2 n$.

(e) Using the alternative solution to (c), we see that $E(Y_N) = E[E(Y_N|N)] = E(2\beta N) = 2\beta/(1-\pi)$ and $\mathrm{var}(Y_N) = E[\mathrm{var}(Y_N|N)] + \mathrm{var}[E(Y_N|N)] = E(2\beta^2 N) + \mathrm{var}(2\beta N) = 2\beta^2/(1-\pi) + 4\beta^2\pi/(1-\pi)^2 = \frac{2\beta^2(1+\pi)}{(1-\pi)^2}$.

55. (a) Let $w_1, w_2, w_3, w_4 > 0$ and set $t_1 = w_1$, $t_2 = w_1 + w_2$, $t_3 = w_1 + w_2 + w_3$ and $t_4 = w_1 + w_2 + w_3 + w_4$. Then $w_1 = t_1$, $w_2 = t_2 - t_1$, $w_3 = t_3 - t_2$ and $w_4 = t_4 - t_3$. Consequently,

$$\frac{\partial(w_1, w_2, w_3, w_4)}{\partial(t_1, t_2, t_3, t_4)} = \begin{vmatrix} 1 & 0 & 0 & 0 \\ -1 & 1 & 0 & 0 \\ 0 & -1 & 1 & 0 \\ 0 & 0 & -1 & 1 \end{vmatrix} = 1.$$

Thus $f_{T_1,T_2,T_3,T_4}(t_1, t_2, t_3, t_4) = f_{W_1,W_2,W_3,W_4}(t_1, t_2 - t_1, t_3 - t_2, t_4 - t_3) = \lambda e^{-\lambda t_1} \lambda e^{-\lambda(t_2 - t_1)} \lambda e^{-\lambda(t_3 - t_2)} \lambda e^{-\lambda(t_4 - t_3)} = \lambda^4 e^{-\lambda t_4}$ for $0 < t_1 < t_2 < t_3 < t_4$ and $t_{T_1,T_2,T_3,T_4}(t_1, t_2, t_3, t_4) = 0$ otherwise.

(b) Observe that $T_4 = T_3 + W_4$, where $[T_1, T_2, T_3]$ and W_4 are independent. Thus, by Theorem 6.2, the conditional distribution of T_4 given that $T_1 = t_1$, $T_2 = t_2$ and $T_3 = t_3$ coincides with the distribution of $t_3 + W_4$, so the corresponding conditional density function is given by

$$f_{T_4|T_1,T_2,T_3}(t_4|t_1, t_2, t_3) = \lambda \exp(-\lambda(t_4 - t_3)), \qquad 0 < t_3 < t_4,$$

and $f_{T_4|T_1,T_2,T_3}(t_4|t_1, t_2, t_3) = 0$ elsewhere. Alternatively, arguing as in (a) we conclude that the joint density of T_1, T_2 and T_3 is given by

$$f_{T_1,T_2,T_3}(t_1, t_2, t_3) = \lambda \exp(-\lambda t_3), \qquad 0 < t_1 < t_2 < t_3,$$

and $f_{T_1,T_2,T_3}(t_1, t_2, t_3) = 0$ elsewhere. Consequently, $f_{T_4|T_1,T_2,T_3}(t_4|t_1, t_2, t_3) = \frac{f_{T_1,T_2,T_3,T_4}(t_1,t_2,t_3,t_4)}{f_{T_1,T_2,T_3}(t_1,t_2,t_3)} = \frac{\lambda^4 \exp(-\lambda t_4)}{\lambda^3 \exp(-\lambda t^3)}$ $= \lambda \exp(-\lambda(t_4 - t_3))$ for $0 < t_3 < t_4$, and $f_{T_4|T_1,T_2,T_3}(t_4|t_1, t_2, t_3) = 0$ elsewhere.

(c) Now $\mathrm{var}(T_1) = \mathrm{var}(W_1) = \lambda^{-2}$ and $\mathrm{var}(T_2) = \mathrm{var}(W_1 + W_2) = \mathrm{var}(W_1) + \mathrm{var}(W_2) = 2\lambda^{-2}$. Similarly, $\mathrm{var}(T_3) = 3\lambda^{-3}$ and $\mathrm{var}(T_4) = 4\lambda^{-4}$. Also, $\mathrm{cov}(T_1, T_2) = \mathrm{cov}(W_1, W_1 + W_2) = \mathrm{var}(W_1) = \lambda^{-2}$. Similarly, $\mathrm{cov}(T_1, T_3) = \mathrm{cov}(T_1, T_4) = \lambda^{-2}$. Moreover, $\mathrm{cov}(T_2, T_3) = \mathrm{cov}(W_1 + W_2, W_1 + W_2 + W_3) = \mathrm{var}(W_1 + W_2) = 2\lambda^{-2}$. Similarly, $\mathrm{cov}(T_2, T_4) = 2\lambda^{-2}$ and $\mathrm{cov}(T_3, T_4) = 3\lambda^{-2}$. Thus the variance-covariance matrix of T_1, T_2, T_3 and T_4 equals

$$\begin{bmatrix} \lambda^{-2} & \lambda^{-2} & \lambda^{-2} & \lambda^{-2} \\ \lambda^{-2} & 2\lambda^{-2} & 2\lambda^{-2} & 2\lambda^{-2} \\ \lambda^{-2} & 2\lambda^{-2} & 3\lambda^{-2} & 3\lambda^{-2} \\ \lambda^{-2} & 2\lambda^{-2} & 3\lambda^{-2} & 4\lambda^{-2} \end{bmatrix}.$$

(d) Since T_1, T_2, T_3 and T_4 have a joint density function, it follows from Theorem 5.19 that their variance-covariance matrix is positive definite and hence invertible. Alternatively,

$$\begin{bmatrix} T_1 \\ T_2 \\ T_3 \\ T_4 \end{bmatrix} = \begin{bmatrix} 1 & 0 & 0 & 0 \\ 1 & 1 & 0 & 0 \\ 1 & 1 & 1 & 0 \\ 1 & 1 & 1 & 1 \end{bmatrix} \begin{bmatrix} W_1 \\ W_2 \\ W_3 \\ W_4 \end{bmatrix}.$$

Thus the variance-covariance matrix of T_1, T_2, T_3 and T_4 equals

$$\lambda^{-2} \begin{bmatrix} 1 & 0 & 0 & 0 \\ 1 & 1 & 0 & 0 \\ 1 & 1 & 1 & 0 \\ 1 & 1 & 1 & 1 \end{bmatrix} \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} 1 & 1 & 1 & 1 \\ 0 & 1 & 1 & 1 \\ 0 & 0 & 1 & 1 \\ 0 & 0 & 0 & 1 \end{bmatrix}.$$

Since the three indicated matrices are invertible (e.g., because the determinant of each of them equals 1), so is their product.

Statistics (Chapters 7–13)

First Practice First Midterm Exam

1. Write an essay on Chapter 7.
2. Write an essay on one of the following three topics and its connection to statistics:
   (a) identifiability;
   (b) saturated spaces;
   (c) orthogonal projections.

Second Practice First Midterm Exam

3. Consider the task of giving a twenty minute review lecture on the basic properties and role of *the design matrix in linear analysis.* The review lecture should include connections of the design matrix with identifiability, saturatedness, change of bases, Gram matrices, the normal equations for the least squares approximation and the solution to these equations, and the linear experimental model. Write out a complete set of lecture notes that could be used for this purpose by yourself or by another student in the course. [To be effective, your lecture notes should be broad but to the point, insightful, clear, interesting, legible, and not too short or too long; perhaps $2\frac{1}{2}$ pages is an appropriate length. Include as much of the relevant background material as is appropriate given the time and space constraints. Points will be subtracted for material that is extraneous or, worse yet, incorrect.]
4. Consider the sugar beet data shown on page 360 of the textbook. Suppose that the five varieties of sugar beet correspond to five equally spaced levels of a secret factor [with the standard variety (variety 1) having the lowest level of the secret factor, variety 2 having the next lowest level, and so forth]. Suppose that the assumptions of the homoskedastic normal linear experimental model are satisfied with G being the space of polynomials of degree 4 (or less) in the level of the secret factor. Let τ be the mean percentage of sugar content for a sixth variety of sugar beet, the level of whose secret factor is midway between that of the standard variety and that of variety 2.
   (a) Express τ as a linear combination of the mean percentages $\mu_1, \dots, \mu_5$ corresponding to the five varieties used in the experiment (see Corollary 9.7).
   (b) Determine a reasonable estimate $\hat{\tau}$ of τ.
   (c) Determine the standard error of $\hat{\tau}$.
   (d) Determine the 95% confidence interval for τ.
   (e) Determine the P-value for testing the hypothesis that $\tau = \mu_1$.
   (f) Carry out a reasonable test of the hypothesis that $\tau = \mu_1 = \cdots = \mu_5$.

Third Practice First Midterm Exam

5. Consider the task of giving a twenty-minute review lecture *on the connection between the least squares problem in statistics and the least squares problem in linear analysis and on the solution to the latter problem based on orthogonal projections*. The coverage should be restricted to the material in Chapters 8 and 9 of the textbook and the corresponding lectures. Write out a complete set of lecture notes that could be used for this purpose by yourself or by another student in the course.

6. Consider the sugar beet data shown on page 360 of the textbook.

   (a) Use all of the data to carry out a t test of the hypothesis that $\mu_2 - 17 \le \mu_1 - 18$.

   (b) Use all of the data to carry out an F test of the hypothesis that $\mu_1 = 18$, $\mu_2 = 17$, and $\mu_3 = 19$.

7. Consider the data

x	-5	-3	-1	1	3	5
y	y_1	y_2	y_3	y_4	y_5	y_6

   Let G be the space of quadratic polynomials on $\mathbb{R}$ and consider the basis 1, x, x^2 of G.

   (a) Determine a quadratic polynomial $P(x)$ such that 1, x, $P(x)$ is an orthogonal basis of G (relative to the inner product with unit weights).

   (b) Let the least squares quadratic polynomial fit to the data be written as $\hat{b}_0 + \hat{b}_1 x + \hat{b}_2 P(x)$. Determine $\hat{b}_0$, $\hat{b}_1$ and $\hat{b}_2$ explicitly in terms of $y_1, \ldots, y_6$.

Fourth Practice First Midterm Exam

8. Consider the task of giving a fifteen-minute review lecture on *the role and properties of orthogonal and orthonormal bases in linear analysis and their connection with the least squares problem in statistics*. The coverage should be restricted to the material in Chapters 8 and 9 of the textbook and the corresponding lectures. Write out a complete set of lecture notes that could be used for this purpose by yourself or by another student in the course.

9. Consider the sugar beet data shown in Table 7.3 on page 360 of the textbook, the corresponding homoskedastic normal five-sample model, and the linear parameter τ defined by

$$\tau = \frac{\mu_1 + \mu_3 + \mu_5}{3} - \frac{\mu_2 + \mu_4}{2}.$$

   (a) Determine the 95% lower confidence bound for τ.

   (b) Determine the P-value for the t test of the null hypothesis that $\tau \le 0$, the alternative hypothesis being that $\tau > 0$. (You may express your answer in terms of a short numerical interval.)

(c) Suppose that the parameter τ was defined by noticing that $\bar{Y}_2$ and $\bar{Y}_4$ are smaller than $\bar{Y}_1$, $\bar{Y}_3$ and $\bar{Y}_5$. Discuss the corresponding practical statistical implications in connection with (a) and (b).

10. Consider the data

x_1	-5	-3	-1	1	3	5
x_2	2	-1	-1	-1	-1	2
y	y_1	y_2	y_3	y_4	y_5	y_6

Let G be the space of functions on $\mathbb{R}^2$ of the form $b_0 + b_1x_1 + b_2x_2$.

(a) Determine the design matrix $\boldsymbol{X}$.

(b) Determine whether or not G is saturated.

(c) Determine whether or not G is identifiable.

(d) Determine whether or not the basis functions 1, x_1 and x_2 are orthogonal.

(e) Express the least squares estimates $\widehat{b}_0$, $\widehat{b}_1$ and $\widehat{b}_2$ of b_0, b_1 and b_2, respectively, explicitly in terms of y_1, y_2, y_3, y_4, y_5 and y_6.

Fifth Practice First Midterm Exam

11. Consider the task of giving a fifteen-minute review lecture on *the properties of the t distribution and the role of this distribution in statistical inference.* The coverage should be restricted to the material in Chapter 7 of the textbook and the corresponding lectures. Write out a complete set of lecture notes that could be used for this purpose by yourself or by another student in the course.

12. Let G be the space of functions on $\mathcal{X} = \mathbb{R}$ that are linear on $(-\infty, 0]$ and on $[0, \infty)$, but whose slope can have a jump at $x = 0$. (Observe that the functions in G are continuous everywhere, in particular at $x = 0$. The point $x = 0$ is referred to as a *knot*, and the functions in G are referred to as *linear splines.* More generally, a space of linear splines can have any specified finite set of knots.)

(a) Show that (explain why) a function g on $\mathbb{R}$ is in G if and only if it is of the form

$$g(x) = \begin{cases} A + Bx & \text{for } x \le 0, \\ A + Cx & \text{for } x \ge 0, \end{cases}$$

for some choice of the constants A, B and C.

(b) Show that G is a linear space.

Consider the function $x_+ \in G$ defined by

$$x_+ = \begin{cases} x & \text{for } x \ge 0, \\ 0 & \text{for } x \le 0. \end{cases}$$

(c) Show that 1, x and x_+ are linearly independent functions in G.

(d) Show that these three functions form a basis of G and hence that $\dim(G) = 3$.

(e) Show that 1, x and $|x|$ form an alternative basis of G, and determine the coefficient matrix $\boldsymbol{A}$ for this new basis relative to the old basis specified in (c).

Consider the design points $x'_1 = -2$, $x'_2 = -1$, $x'_3 = 0$, $x'_4 = 1$ and $x'_5 = 2$ and unit weights.

(f) Determine the Gram matrix $\boldsymbol{M}$ for the basis in (e).

(g) Determine a function $g \in G$ such that 1, x and g form an orthogonal basis of G.

(h) Determine the squared norms of the basis functions in (g).

(i) Use the orthogonal basis in (g) to determine the orthogonal projection $P_G(x^2)$ of the function x^2 onto G, and check that $x^2 - P_G(x^2)$ is orthogonal to 1.

First Practice Second Midterm Exam

13. Write an essay using the lube oil data to tie together many of the main ideas from Chapters 10 and 11 of the textbook.

14. Consider the following experimental data.

x_1	x_2	x_3	x_4	Y
-1	-1	-1	0	Y_1
-1	-1	0	1	Y_2
-1	-1	1	-1	Y_3
-1	1	-1	-1	Y_4
-1	1	0	0	Y_5
-1	1	1	1	Y_6
1	-1	-1	0	Y_7
1	-1	0	-1	Y_8
1	-1	1	1	Y_9
1	1	-1	1	Y_{10}
1	1	0	0	Y_{11}
1	1	1	-1	Y_{12}

Think of the values of Y as specified but not shown. Instead of using their actual numerical values on the exam, use the summary statistics

$$
\begin{aligned}
n &= 12 \\
\bar{Y} &= \frac{1}{n}\sum_i Y_i \\
\text{TSS} &= \sum_i (Y_i - \bar{Y})^2 \\
T_1 &= \langle x_1, Y(\cdot)\rangle \\
T_2 &= \langle x_2, Y(\cdot)\rangle \\
T_3 &= \langle x_3, Y(\cdot)\rangle \\
T_4 &= \langle x_4, Y(\cdot)\rangle \\
T_5 &= \langle x_4^2 - 2/3, Y(\cdot)\rangle
\end{aligned}
$$

Similarly, any numerical value defined in terms of these summary statistics can be treated as a numerical value later on. You can also define numerical values (e.g., P-values) by using functions such as $t_\nu(y)$, $t_{p,\nu}$, $F_{\nu_1,\nu_2}(y)$, F_{p,ν_1,ν_2} or the equivalent S/S-PLUS functions.

(a) Are the design random variables X_1, X_2, X_3 and X_4 uniformly distributed (why or why not)?

(b) Are these design random variables pairwise independent (why or why not)?

(c) Does the design have the form of an orthogonal array (why or why not)?

(d) Show that the design random variables X_1, X_2 and X_3 are independent, as are X_1, X_2 and X_4.

(e) Consider the following linear spaces:

G_1: linear functions on $\{-1, 1\}$
G_2: linear functions on $\{-1, 1\}$
G_3: linear functions on $\{-1, 0, 1\}$
G_4: quadratic polynomials on $\{-1, 0, 1\}$

Think of $G_1, \ldots, G_4$ as spaces of functions on $x_1, \ldots, x_4$, respectively. Let $G_j^0 = G_j \ominus G_{\text{con}}$ as usual. Then these spaces have the following orthogonal bases:

G_1^0: x_1
G_2^0: x_2
G_3^0: x_3
G_4^0: x_4, $x_4^2 - 2/3$

Determine the squared norms of the indicated basis functions. Also determine which two of these basis functions fail to be orthogonal and determine their inner product.

(f) Consider the linear space

$$G = G_{\text{con}} \oplus G_1^0 \oplus G_2^0 \oplus G_3^0 \oplus G_4^0.$$

Write the least squares estimate of the regression function in this space as

$$\hat{\beta}_0 + \hat{\beta}_1 x_1 + \hat{\beta}_2 x_2 + \hat{\beta}_3 x_3 + \beta_4 x_4 + \hat{\beta}_5 (x_4^2 - 2/3).$$

Determine the least squares estimates of the regression coefficients in terms of the summary statistics $T_1, \ldots, T_5$.

(g) Determine the corresponding fitted sum of squares FSS (in terms of $\hat{\beta}_1, \ldots, \hat{\beta}_5$) and residual sum of squares RSS (recall that you "know" TSS). Show your results in the form of an ANOVA table, including the proper degrees of freedom and F statistic. Describe the null hypothesis that is being tested by this statistic.

(h) Determine the usual estimates S^2 and S of σ^2 and σ, respectively. Determine the standard error of $\hat{\beta}_5$. Determine the 95% confidence interval for β_5. Determine the P-value for testing the hypothesis that $\beta_5 = 0$.

(i) Let RSS_0 be the residual sum of squares for the least squares estimate of the regression function in the subspace of G spanned by 1, x_1, x_2, x_3 and x_4. Determine a simple formula for $\text{RSS}_0 - \text{RSS}$ in terms of implicit numerical values.

Second Practice Second Midterm Exam

15. Consider the task of giving a twenty minute review lecture on *hypothesis testing in the context of the normal linear model*. The review lecture should include the description of the model, t tests, F tests, their equivalence, tests of constancy, goodness-of-fit tests, and connection with coefficient tables and ANOVA tables. Write out a complete set of lecture notes that could be used for this purpose by yourself or by another student in the course.

16. Consider the following orthogonal array and response data.

x_1	x_2	x_3	x_4	Y
-3	-1	1	-1	232
-3	-1	-1	1	190
-3	0	-1	-1	204
-3	0	1	1	138
-3	1	-1	1	174
-3	1	1	-1	148
-1	-1	-1	-1	224
-1	-1	1	1	222
-1	0	-1	1	176
-1	0	1	-1	202
-1	1	-1	-1	232
-1	1	1	1	138
1	-1	-1	1	228
1	-1	1	-1	226
1	0	-1	-1	214
1	0	1	1	184
1	1	1	-1	194
1	1	-1	1	208
3	-1	-1	-1	298
3	-1	1	1	148
3	0	1	-1	180
3	0	-1	1	222
3	1	-1	-1	198
3	1	1	1	220

(a) Let X_1, X_2, X_3 and X_4 be the corresponding design random variables. Which of the following triples of design random variables are independent and which are dependent? X_1, X_2 and X_3; X_1, X_2 and X_4; X_1, X_3 and X_4.

(b) Let G be the 7-dimensional linear space of functions on $\mathbb{R}^5$ having the basis functions 1, x_1, $x_1^2 - 5$, x_2, x_3, x_4, x_1x_3. Determine the squared norms of these basis functions.

(c) With one exception, these basis functions are orthogonal to each other. Determine the exceptional pair and determine the inner product of the basis functions of this pair.

(d) Determine the Gram matrix of the basis functions, and determine the inverse of this matrix.

(e) Here are some useful summary statistics: $\sum_i Y_i = 4800$; $\sum_i x_{i1}Y_i = 600$; $\sum_i (x_{i1}^2 - 5)Y_i = -384$; $\sum_i x_{i2}Y_i = -256$; $\sum_i x_{i3}Y_i = -336$; $\sum_i x_{i4}Y_i = -304$; $\sum_i x_{i1}x_{i3}Y_i = -336$. Use these summary statistics to determine the least squares estimates of the regression coefficients.

(f) Determine the fitted sum of squares.

(g) The total sum of squares is given by TSS $= 30480$. Determine the residual sum of squares.

(h) Test the hypothesis that the regression function is a linear function of x_1, x_2, x_3 and x_4.

(i) Consider the possibility of removing one of the basis functions of G. Which basis function should be removed and why?

Third Practice Second Midterm Exam

17. Consider the task of giving a twenty-minute review lecture *on randomization, blocking, and covariates*. The coverage should be restricted to the relevant material in the textbook and the corresponding lectures. Write out a complete set of lecture notes that could be used for this purpose by yourself or by another student in the course.

18. Consider the following orthogonal array and response data.

x_1	x_2	x_3	x_4	x_5	x_6	Y	x_1	x_2	x_3	x_4	x_5	x_6	Y
−1	−1	−1	−1	−1	−1	262	1	−1	1	−1	−1	−1	642
−1	−1	−1	0	0	0	649	1	−1	1	0	0	0	315
−1	−1	−1	1	1	1	854	1	−1	1	1	1	1	520
−1	−1	−1	−1	0	−1	879	1	−1	1	−1	0	0	497
−1	−1	−1	0	1	0	461	1	−1	1	0	1	1	793
−1	−1	−1	1	−1	1	698	1	−1	1	1	−1	−1	460
−1	−1	1	−1	1	0	856	1	−1	−1	−1	1	0	754
−1	−1	1	0	−1	1	197	1	−1	−1	0	−1	1	809
−1	−1	1	1	0	−1	637	1	−1	−1	1	0	−1	535
−1	1	−1	−1	0	1	273	1	1	1	−1	−1	0	582
−1	1	−1	0	1	−1	713	1	1	1	0	0	1	255
−1	1	−1	1	−1	0	236	1	1	1	1	1	−1	604
−1	1	1	−1	−1	1	16	1	1	−1	−1	0	1	315
−1	1	1	0	0	−1	547	1	1	−1	0	1	−1	755
−1	1	1	1	1	0	38	1	1	−1	1	−1	0	992
−1	1	1	−1	1	1	60	1	1	−1	−1	1	−1	552
−1	1	1	0	−1	−1	259	1	1	−1	0	−1	0	607
−1	1	1	1	0	0	555	1	1	−1	1	0	1	903

Consider the linear space $\mathbb{G}$ spanned by the linearly independent functions 1, x_1, x_2, x_3, x_1x_2, x_4, x_5, x_6 and x_4x_5. The corresponding Gram matrix is given by

$$\boldsymbol{M} = \begin{bmatrix} 36 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 36 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 36 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 36 & -12 & 0 & 0 & 0 & 6 \\ 0 & 0 & 0 & -12 & 36 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 24 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 24 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 24 & 0 \\ 0 & 0 & 0 & 6 & 0 & 0 & 0 & 0 & 16 \end{bmatrix}.$$

(a) Verify the correctness of the entry 16 in the last row and column of $\boldsymbol{M}$.

(b) Explain why it follows from the fact that the design has the form of an orthogonal array that the entry in row 3 and column 5 of the Gram matrix equals zero.

(c) Are the design random variables X_3, X_4 and X_5 independent?

(d) Are the design random variables X_4, X_5 and X_6 independent?

The inverse Gram matrix can be written as

$$\boldsymbol{M}^{-1} = \frac{1}{36 \cdot 238} \begin{bmatrix} 238 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 238 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 238 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 288 & 96 & 0 & 0 & 0 & -108 \\ 0 & 0 & 0 & 96 & 270 & 0 & 0 & 0 & -36 \\ 0 & 0 & 0 & 0 & 0 & 357 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 357 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 357 & 0 \\ 0 & 0 & 0 & -108 & -36 & 0 & 0 & 0 & 576 \end{bmatrix}.$$

Here are some potentially useful summary statistics: $\sum_i Y_i = 19080$, $\sum_i x_{i1}Y_i = 2700$, $\sum_i x_{i2}Y_i = -2556$, $\sum_i x_{i3}Y_i = -3414$, $\sum_i x_{i1}x_{i2}Y_i = 3036$, $\sum_i x_{i4}Y_i = 1344$, $\sum_i x_{i5}Y_i = 1200$, $\sum_i x_{i6}Y_i = -1152$, $\sum_i x_{i4}x_{i5}Y_i = -1090$, and $\sum_i (Y_i - \bar{Y})^2 = 2363560$. Assume that the normal linear regression model corresponding to $\mathbb{G}$ is valid. Write the regression function and its least squares estimate as $\mu(\cdot) = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \beta_3 x_3 + \beta_4 x_1 x_2 + \beta_5 x_4 + \beta_6 x_5 + \beta_7 x_6 + \beta_8 x_4 x_5$ and $\hat{\mu}(\cdot) = \hat{\beta}_0 + \hat{\beta}_1 x_1 + \hat{\beta}_2 x_2 + \hat{\beta}_3 x_3 + \hat{\beta}_4 x_1 x_2 + \hat{\beta}_5 x_4 + \hat{\beta}_6 x_5 + \hat{\beta}_7 x_6 + \hat{\beta}_8 x_4 x_5$.

(e) Determine $\hat{\beta}_4$.

The residual sum of squares is given by RSS = 1325184.

(f) Determine the usual estimate S of σ.

(g) Test the hypothesis that $\beta_4 = 0$.

Let $\mathbb{G}_0$ be the subspace of $\mathbb{G}$ spanned by 1, x_1, x_2, x_3, and $x_1 x_2$.

(h) Determine the numerical values of the regression coefficients for the least squares fit in $\mathbb{G}_0$.

(i) Determine the corresponding fitted and residual sums of squares.

(j) Test the hypothesis that the regression function is in $\mathbb{G}_0$. (Show the appropriate ANOVA table.)

Fourth Practice Second Midterm Exam

19. Section 10.8 of the textbook is entitled "The F Test." It includes subsections entitled "Tests of Constancy," "Goodness-of-Fit Test," and "Equivalence of t Tests and F Tests." Consider the task of giving a fifteen-minute review lecture on this material. Write out a complete set of lecture notes that could be used for this purpose by yourself or by another student in the course.

20. The orthogonal array in the following dataset is taken from Table 9.18 on Page 210 of *Orthogonal Arrays: Theory and Applications* by A. S. Hedayat, N. J. A. Sloane and John Stufken, Springer: New York, 1999. Note that each of the first six factors take on two levels. In the indicated table, the two levels are 0 and 1, but for purpose of analysis it is easier to think of them as being -1 and 1 (so that $\bar{x}_1 = \cdots = \bar{x}_6 = 0$). Similarly, the last three factors take on four levels. In the indicated table, they are indicated as 0, 1, 2 and 3, but for present purposes, it is more convenient to think of them as being -3, -1, 1 and 3:

x_1	x_2	x_3	x_4	x_5	x_6	x_7	x_8	x_9	y
-1	-1	-1	-1	-1	-1	-3	-3	-3	16
1	1	-1	1	-1	1	-3	1	1	26
-1	1	1	1	1	-1	1	-3	1	34
1	-1	1	-1	1	1	1	1	-3	28
-1	-1	1	1	1	1	-3	3	3	29
1	1	1	-1	1	-1	-3	-1	-1	27
-1	1	-1	-1	-1	1	1	3	-1	11
1	-1	-1	1	-1	-1	1	-1	3	41
1	1	-1	-1	1	1	3	-3	3	38
-1	-1	-1	1	1	-1	3	1	-1	24
1	-1	1	1	-1	1	-1	-3	-1	40
-1	1	1	-1	-1	-1	-1	1	3	26
1	1	1	1	-1	-1	3	3	-3	27
-1	-1	1	-1	-1	1	3	-1	1	33
1	-1	-1	-1	1	-1	-1	3	1	25
-1	1	-1	1	1	1	-1	-1	-3	15

Here are some (mostly) relevant summary statistics:

$\sum_i x_{i1} = \cdots = \sum_i x_{i9} = 0$;

$\sum_i x_{i1}x_{i7}x_{i8} = \sum_i x_{i2}x_{i7}x_{i8} = \sum_i x_{i3}x_{i7}x_{i8} = \sum_i x_{i4}x_{i7}x_{i8} = 0$;

$\bar{y} = 27.5$; $\sum_i (y_i - \bar{y})^2 = 1108$;

$\sum_i x_{i1}y_i = 64$; $\sum_i x_{i2}y_i = -32$; $\sum_i x_{i7}y_i = 80$; $\sum_i x_{i8}y_i = -120$;

$\sum_i x_{i7}x_{i8}y_i = -288$.

(a) Are the design random variables X_1, X_7 and X_8 independent? (Justify your answer.)

(b) Are x_1 and x_7x_8 orthogonal?

(c) Determine $\langle x_6, x_7x_8 \rangle$.

(d) Determine the least squares estimate of the regression function of the form $\widehat{\beta}_0 + \widehat{\beta}_1 x_1 + \widehat{\beta}_2 x_2 + \widehat{\beta}_7 x_7 + \widehat{\beta}_8 x_8 + \widehat{\beta}_{78} x_7 x_8$.

(e) Determine the corresponding fitted and residual sums of. (Show your work. As a check, I got that FSS = 787.36.)

(f) (Under the assumptions of the corresponding homoskedastic normal linear regression model) Determine the t statistic and determine (an appropriate interval for) the corresponding P-value for testing the hypothesis that $\beta_2 = 0$.

(g) Determine the residual sum of squares corresponding to the hypothesis that $\beta_2 = 0$, and use it to carry out the F test of this hypothesis.

(h) Determine the 95% confidence interval for the change in the mean response caused by changing x_8 from its lowest level (-3) to its highest level (3) with x_1, x_2 and x_7 all held fixed at the level -1.

Fifth Practice Second Midterm Exam

21. Consider the task of giving a fifteen-minute review lecture on *tensor-product spaces and their application to interactions in linear regression models and, in particular, to the corresponding ANOVA tables.* Write out a complete set of lecture notes that could be used for this purpose by yourself or by another student in the course.

22. Consider the following "experimental" data in the form of an orthogonal array:

x_1	x_2	x_3	x_4	y
-1	-1	-1	-1	12
-1	-1	0	1	11
-1	-1	1	0	17
-1	0	-1	1	13
-1	0	0	0	19
-1	0	1	-1	17
-1	1	-1	0	13
-1	1	0	-1	11
-1	1	1	1	10
1	-1	-1	0	25
1	-1	0	-1	31
1	-1	1	1	34
1	0	-1	-1	25
1	0	0	1	20
1	0	1	0	30
1	1	-1	1	10
1	1	0	0	12
1	1	1	-1	26

Let X_1, X_2, X_3 and X_4 denote the corresponding design random variables. Observe that

- X_1, X_2 and X_3 are independent
- X_1, X_2 and X_4 are independent
- X_1, X_3 and X_4 are independent

(a) Are X_2, X_3 and X_4 independent (why or why not)?

(b) Let G be the space of functions on $\mathbb{R}^4$ spanned by the functions

$$1,\ x_1,\ x_2,\ x_2^2,\ x_3,\ x_3^2,\ x_4,\ x_4^2,\ x_1x_2,\ x_1x_3,\ x_1x_4.$$

Justify that the indicated 11 functions are linearly independent and form a basis of G.

(c) Demonstrate that the basis in (b) is not an orthogonal basis.

(d) It can be shown that

$$1,\ x_1,\ x_2,\ x_2^2-2/3,\ x_3,\ x_3^2-2/3,\ x_4,\ x_4^2-2/3,\ x_1x_2,\ x_1x_3,\ x_1x_4$$

is an orthogonal basis of G. Demonstrate, in particular, that $x_2^2-2/3$ is orthogonal to 1, x_1, x_2, x_3 and $x_3^2-2/3$. *Hint:* use the design random variables.

Here are some relevant summary statistics:

$$\begin{aligned}
\sum_i y_i &= 336\\
\sum_i (y_i-\bar{y})^2 &= 1078\\
\sum_i x_{i1}y_i &= 90\\
\sum_i x_{i2}y_i &= -48\\
\sum_i (x_{i2}^2-2/3)y_i &= -12\\
\sum_i x_{i3}y_i &= 36\\
\sum_i (x_{i3}^2-2/3)y_i &= 8\\
\sum_i x_{i4}y_i &= -24\\
\sum_i (x_{i4}^2-2/3)y_i &= -4\\
\sum_i x_{i1}x_{i2}y_i &= -36\\
\sum_i x_{i1}x_{i3}y_i &= 24\\
\sum_i x_{i1}x_{i4}y_i &= -12
\end{aligned}$$

Let the least squares estimate in G of the regression function be written as

$$\begin{aligned}
\widehat{\mu}(x_1,x_2,x_3,x_4) = \widehat{\beta}_0 &+ \widehat{\beta}_1x_1 + \widehat{\beta}_2x_2 + \widehat{\beta}_3(x_2^2-2/3)\\
&+ \widehat{\beta}_4x_3 + \widehat{\beta}_5(x_3^2-2/3) + \widehat{\beta}_6x_4 + \widehat{\beta}_7(x_4^2-2/3)\\
&+ \widehat{\beta}_8x_1x_2 + \widehat{\beta}_9x_1x_3 + \widehat{\beta}_{10}x_1x_4.
\end{aligned}$$

It follows from the above summary statistics that

$$\begin{aligned}\widehat{\mu}(x_1, x_2, x_3, x_4) &= 18\tfrac{2}{3} + 5x_1 - 4x_2 - 3(x_2^2 - 2/3) \\ &\quad + 3x_3 + 2(x_3^2 - 2/3) - 2x_4 - (x_4^2 - 2/3) \\ &\quad - 3x_1x_2 + 2x_1x_3 - x_1x_4 \\ &= 20 + 5x_1 - 4x_2 - 3x_2^2 + 3x_3 + 2x_3^2 - 2x_4 - x_4^2 - 3x_1x_2 + 2x_1x_3 - x_1x_4.\end{aligned}$$

(e) Check that the numerical values given for $\widehat{\beta}_0$, $\widehat{\beta}_1$, $\widehat{\beta}_2$, $\widehat{\beta}_3$ and $\widehat{\beta}_8$ are correct.

The numerical value of the fitted sum of squares is given by FSS = 1022.

(f) Determine the residual sum of squares and the squared multiple correlation coefficient.

(g) Determine the least squares fit $\widehat{\mu}_0(\cdot)$ in the subspace G_0 of G spanned by 1, x_1, x_2, x_3 and x_4, and determine the corresponding fitted sum of squares FSS_0.

From now on, let the assumptions of the homoskedastic normal linear regression model corresponding to G be satisfied.

(h) Determine the usual estimates s^2 and s of σ^2 and σ, respectively.

(i) Carry out the standard test of the null hypothesis that $\mu(\cdot) \in G_0$. In particular, determine the numerical value of the usual test statistic and determine (the appropriate interval containing) the corresponding P-value.

Let the regression function be written in terms of the basis in (d) as

$$\begin{aligned}\mu(x_1, x_2, x_3, x_4) &= \beta_0 + \beta_1x_1 + \beta_2x_2 + \beta_3(x_2^2 - 2/3) \\ &\quad + \beta_4x_3 + \beta_5(x_3^2 - 2/3) + \beta_6x_4 + \beta_7(x_4^2 - 2/3) \\ &\quad + \beta_8x_1x_2 + \beta_9x_1x_3 + \beta_{10}x_1x_4.\end{aligned}$$

Also, let τ denote the change in the mean response caused by changing factor 1 from low to high (i.e., from $x_1 = -1$ to $x_1 = 1$) with the other factors held at their low level (i.e., $x_2 = x_3 = x_4 = -1$).

(j) Determine the least squares estimate $\widehat{\tau}$ of τ (both in terms of $\widehat{\beta}_0, \dots, \widehat{\beta}_{10}$ and numerically).

(k) Determine the 95% confidence interval for τ.

(l) Carry out the usual test of the null hypothesis that $\tau = 0$. In particular, determine the numerical value of the usual test statistic and determine (the appropriate interval containing) the corresponding P-value.

First Practice Final Exam

23. Write out a detailed set of lecture notes for a review of HYPOTHESIS TESTING in the context of the homoskedastic normal linear regression model. Be sure to include t tests, F tests, tests of homogeneity, goodness-of-fit tests, and their relationships.

24. Write out a detailed set of lecture notes for a review of HYPOTHESIS TESTNG in the context of POISSON REGRESSION.

25. (a) Use least-squares approximation, as treated in the textbook, to determine a decent quadratic approximation to the function $\sqrt{x}$ in the interval $1 \le x \le 2$.

    (b) Determine the maximum magnitude of the error of approximation.

26. Consider an experiment involving two factors, each having the two levels 1 and 2. The experimental results are as follows:

x_1	x_2	# of runs	# of successes
1	1	50	20
1	2	25	11
2	1	25	18
2	2	50	28

Consider the corresponding linear logistic regression model in which G is the space of additive functions of x_1 and x_2. Consider also the submodel in which G_0 is the space of additive functions of the form $g(x_1) + g(x_2)$ (equivalently, of the form $g_1(x_1) + g_2(x_2)$ with $g_1 = g_2$).

(a) Determine the dimensions of G_0 and G.

(b) Determine a basis of G_0 and determine an extension of this basis to a basis of G.

(c) Determine the design matrices (as defined in Chapter 13) corresponding to G_0 and G.

(d) Using $\mathbf{0}$ as the vector of starting values for the maximum likelihood estimates of the regression coefficients for the submodel, determine the first iteration of the iteratively reweighted least-squares method for obtaining these estimates.

The maximum likelihood estimates $\widehat{\pi}$ and $\widehat{\pi}_0$ of the success probabilities at the design points corresponding to the model and submodel, respectively, are as follows:

x_1	x_2	$\widehat{\pi}$	$\widehat{\pi}_0$
1	1	.4336	.4332
1	2	.3728	.5136
2	1	.6528	.5136
2	2	.5936	.5932

(e) Determine the maximum likelihood estimate $\widehat{\boldsymbol{\beta}}_0$ of the coefficient vector corresponding to the submodel.

(f) Determine the standard errors of the entries of the MLE in (e).

(g) Determine the likelihood ratio statistic for testing the submodel under the assumption that the model is valid, and determine the corresponding P-value.

Second Practice Final Exam

27. Consider the task of giving a 45 minute review lecture that is directed at *synthesizing the material on ordinary regression and that on logistic regression.* Thus you should NOT GIVE SEPARATE TREATMENTS of the two contexts. Instead, you should take the various important topics in turn and point out the similarities and differences between ordinary regression and logistic regression in regard to each such topic. In particular, you should compare and contrast

    - the assumptions underlying ordinary regression and those underlying logistic regression;
    - the least squares method and the maximum likelihood method;
    - the joint distribution of the least squares estimates and that of the maximum likelihood estimates;
    - confidence intervals in the two contexts;
    - tests involving a single parameter in the two contexts;
    - F tests in the context of ordinary regression and likelihood ratio tests in the context of logistic regression;
    - goodness-of-fit tests in the two contexts.

    Write out a complete set of lecture notes that could be used for this purpose by yourself or by another student in the course.

28. Consider a complete factorial experiment involving two factors, the first of which is at three levels and the second of which is at two levels. The experimental data are as follows:

x_1	x_2	Time parameter	Number of events
-1	-1	25	35
-1	1	25	20
0	-1	50	75
0	1	50	50
1	-1	25	30
1	1	25	40

    Consider the corresponding Poisson linear experimental model, in which $G = G_1 + G_2$, where G_1 is the space of quadratic polynomials in x_1 and G_2 is the space of linear functions of x_2.

    (a) Determine the dimensions of G_1, G_2 and G.

    (b) Determine which of these spaces are identifiable.

    (c) Determine which of these spaces are saturated.

    (d) Consider the basis 1, x_1, $x_1^2 - 2/3$ of G_1. Determine the corresponding maximum likelihood estimate in G_1 of the regression coefficients. *Hint:* Observe that if the column vectors of a square matrix $\boldsymbol{X}$ are suitably orthogonal and nonzero then $\boldsymbol{X}^T\boldsymbol{X}$ is an invertible square matrix $\boldsymbol{D}$ and hence $\boldsymbol{X}^{-1} = \boldsymbol{D}^{-1}\boldsymbol{X}^T$.

(e) Under the assumption that the Poisson regression function is in G_1, determine a reasonable 95% confidence interval for the change in *rate* when the first factor is changed from its low level to its high level.

(f) Consider the basis 1, x_1, $x_1^2 - 2/3$, x_2 of G. Determine the corresponding maximum likelihood equations for the regression coefficients.

(g) The actual numerical values of the maximum likelihood estimates of these coefficients are given by $\hat{\beta}_0 \doteq 0.211$, $\hat{\beta}_1 \doteq 0.121$, $\hat{\beta}_2 \doteq -0.007$, and $\hat{\beta}_3 \doteq -0.121$. Determine the corresponding deviance and use it to test the goodness-of-fit of the model.

(h) Under the assumption that the Poisson regression function is in G, carry out the likelihood ratio test of the hypothesis that this function is in G_1.

29. Consider the two-sample model, in which $Y_{11}, \ldots, Y_{1n_1}, Y_{21}, \ldots, Y_{2n_2}$ are independent random variables, $Y_{11}, \ldots, Y_{1n_1}$ are identically distributed with mean μ_1 and finite, positive standard deviation σ_1, and $Y_{21}, \ldots, Y_{2n_2}$ are identically distributed with mean μ_2 and finite, positive standard deviation σ_2. Here n_1 and n_2 are known positive integers and μ_1, σ_1, μ_2, and σ_2 are unknown parameters. Do NOT assume that the indicated random variables are normally distributed or that $\sigma_1 = \sigma_2$. Let $\bar{Y}_1$, S_1, $\bar{Y}_2$, and S_2 denote the corresponding sample means and sample standard deviations as usually defined.

(a) Derive a reasonable formula for confidence intervals for the parameter $\mu_1 - \mu_2$ that is approximately valid when n_1 and n_2 are large.

(b) Derive a reasonable formula for the P-value for a test of the hypothesis that the parameter in (a) equals zero.

(c) Calculate the numerical value of the 95% confidence interval in (a) when $n_1 = 50$, $\bar{Y}_1 \doteq 18$, $S_1 \doteq 20$, $n_2 = 100$, $\bar{Y}_2 \doteq 12$, and $S_2 \doteq 10$.

(d) Determine the P-value for the test in (b) [when n_1 and so forth are as in (c)].

(e) Derive a reasonable formula for confidence intervals for the parameter μ_1/μ_2 that is approximately valid when n_1 and n_2 are large and $\mu_2 \neq 0$.

(f) Explain why $\frac{\sigma_2}{\sqrt{n_2}\,|\mu_2|} \approx 0$ is required for the validity of the confidence intervals in (e).

(g) Derive a reasonable formula for the P-value for a test of the hypothesis that the parameter in (e) equals 1.

(h) Calculate the numerical value of the 95% confidence interval in (e).

(i) Determine the P-value for the test in (g).

(j) Discuss the relationships of your answers to (c), (d), (h) and (i).

Third Practice Final Exam

30. Consider the task of giving a twenty-minute review lecture on each of the two topics indicated below. The coverage should be restricted to the relevant material in the textbook and the corresponding lectures. Write out a complete set of lecture notes that could be used for this purpose by yourself or by another student in the course.

(a) The topic is least squares estimation and the corresponding confidence intervals and t tests (not F tests).

(b) The topic is maximum likelihood estimation in the context of Poisson regression and the corresponding confidence intervals and Wald tests (not likelihood ratio tests).

31. Let Y_1, Y_2, and Y_3 be independent random variables such that Y_1 has the binomial distribution with parameters $n_1 = 200$ and π_1, Y_2 has the binomial distribution with parameters $n_2 = 300$ and π_2, and Y_3 has the binomial distribution with parameters $n_3 = 400$ and π_3. Consider the logistic regression model $\text{logit}(\pi_k) = \beta_0 + \beta_1 x_k + \beta_2 x_k^2$ for $k = 1, 2, 3$, where $x_1 = -1$, $x_2 = 0$, and $x_3 = 1$. Suppose that the observed values of Y_1, Y_2, and Y_3 are 60, 150, and 160, respectively.

(a) Determine the design matrix and its inverse.

(b) Determine the maximum likelihood estimates of β_0, β_1, and β_2.

(c) Determine a 95% confidence interval for β_1.

(d) Carry out the Wald test of the hypothesis that $\beta_1 = 0$.

Consider the submodel $\beta_1 = 0$ or, equivalently, that $\text{logit}(\pi_k) = \beta_0 + \beta_2 x_k^2$ for $k = 1, 2, 3$.

(e) Determine the maximum likelihood estimates of β_0 and β_2 under the submodel. *Hint:* explain why the data for the first and third design points can be merged.

(f) Carry out the likelihood ratio test of the submodel.

32. Consider an experiment involving a single factor at three levels: $x_1 = -1$, $x_2 = 0$, and $x_3 = 1$. There are 200 runs at -1, 300 runs at 0, and 400 runs at 1. The 900 responses are independent and normally distributed with a common unknown variance σ^2. Let Y_i, $1 \le i \le 200$, denote the responses at level -1; let Y_i, $201 \le i \le 500$, denote the responses at level 0; and let Y_i, $501 \le i \le 900$, denote the responses at level 1. Let μ_1, μ_2, and μ_3 denote the mean responses at the levels -1, 0, and 1, respectively. The observed sample means at these levels are given by $\bar{Y}_1 = 0.38$, $\bar{Y}_2 = 0.62$, and $\bar{Y}_3 = 0.47$, respectively; the sample standard deviations at these levels are given by $S_1 = 1.4$, $S_2 = 1.6$, and $S_3 = 1.5$, respectively.

(a) Determine the within sum of squares WSS, between sum of squares BSS, and total sum of squares TSS.

Consider the corresponding (homoskedastic) normal linear experimental model with $\mu_k = \beta_0 + \beta_1 x_k + \beta_2 x_k^2$ for $k = 1, 2, 3$. Let $\widehat{\beta}_1$, $\widehat{\beta}_2$, and $\widehat{\beta}_3$ denote the least squares estimates of the regression coefficients β_1, β_2, and β_3.

(b) Determine the least squares estimates of the regression coefficients. *Hint:* see your solution to Problem 31(a).

(c) Determine the fitted sum of squares FSS, the residual sum of squares RSS, and the corresponding estimate S of σ.

(d) Determine a reasonable (aproximately) 95% confidence interval for the parameter $\tau = \beta_0^2 + \beta_2^2$.

Consider the submodel $\beta_1 = 0$ or, equivalently, that $\mu_k = \beta_0 + \beta_2 x_k^2$ for $k = 1, 2, 3$.

(e) Determine the least squares estimates of β_0 and β_2 under the submodel.

(f) Determine the fitted sum of squares for the submodel.

(g) Determine the F statistic for testing the hypothesis that the submodel is valid, and determine the corresponding P-value.

Fourth Practice Final Exam

33. Consider the task of giving a thirty-minute review lecture on Poisson regression (including the underlying model, maximum likelihood estimation of the regression coefficients, the corresponding maximum likelihood equations, asymptotic joint and individual distribution of the maximum likelihood estimates of the regression coefficients, confidence intervals for these coefficients, likelihood ratio and Wald tests and their relationship, and goodness of fit tests). The coverage should be restricted to the relevant material in the textbook and the corresponding lectures. Write out a complete set of lecture notes that could be used for this purpose by yourself or by another student in the course.

34. Consider a factorial experiment involving d distinct design points $\boldsymbol{x}'_1, \dots, \boldsymbol{x}'_d \in \mathcal{X}$ and n_k runs at the design point $\boldsymbol{x}'_k$ for $k = 1, \dots, d$. Let G be an identifiable and saturated d-dimensional linear space of functions on $\mathcal{X}$ containing the constant functions, and let $g_1, \dots, g_d$ be a basis of G.

(a) Determine the corresponding $d \times d$ form $\boldsymbol{X}$ of the design matrix.

(b) Determine the weight matrix $\boldsymbol{W}$ (which depends only on $n_1, \dots, n_d$).

For $k = 1, \dots, d$, let $\bar{Y}_k$ denote the sample mean of the n_k responses at the kth design point, let $\bar{y}_k$ denote the observed value of $\bar{Y}_k$, and set $\bar{\boldsymbol{Y}} = [\bar{Y}_1, \dots, \bar{Y}_d]^T$ and $\bar{\boldsymbol{y}} = [\bar{y}_1, \dots, \bar{y}_d]^T$. Also, let $\bar{Y}$ denote the sample mean of all $n = n_1 + \cdots + n_d$ responses, and let $\bar{y}$ denote the observed value of $\bar{Y}$.

(c) Express $\bar{y}$ in terms of $n_1, \dots, n_d, n, \bar{y}_1, \dots, \bar{y}_d$.

Let s_k denote the sample standard deviation (or its observed value) of the n_k responses at the kth design point.

(d) Express the within sum of squares WSS in terms of $n_1, \dots, n_d, s_1, \dots, s_d$.

Consider the corresponding homoskedastic normal linear regression model, in which the response variables have common standard deviation σ and the regression function has the form $\mu(\cdot) = \beta_1 g_1 + \cdots + \beta_d g_d$. Let $\boldsymbol{\beta} = [\beta_1, \dots, \beta_d]^T$ denote the vector of regression coefficients, let $\mu_k = \mu(\boldsymbol{x}'_k)$ denote the mean of the responses at the kth design point, and set $\boldsymbol{\mu} = [\mu_1, \dots, \mu_d]^T$.

(e) Express the distribution of $\bar{\boldsymbol{Y}}$ in terms of $\boldsymbol{\mu}$, $\boldsymbol{W}$ and σ.

(f) Express $\boldsymbol{\mu}$ in terms of $\boldsymbol{X}$ and $\boldsymbol{\beta}$.

Let $\widehat{\mu}(\cdot) = \widehat{\beta}_1 g_1 + \cdots + \widehat{\beta}_d g_d$ denote the least squares estimate of the regression function, and let $\widehat{\boldsymbol{\beta}} = [\widehat{\beta}_1, \ldots, \widehat{\beta}_d]^T$ denote the vector of least squares estimates of the regression coefficients (or the observed value of this vector).

(g) Express $\widehat{\boldsymbol{\beta}}$ in terms of $\boldsymbol{X}$ and $\bar{\boldsymbol{Y}}$.

(h) Determine the distribution of $\widehat{\boldsymbol{\beta}}$ in terms of $\boldsymbol{\beta}$, σ, $\boldsymbol{X}$, and $\boldsymbol{W}$.

(i) Determine simple and natural formulas for the lack of fit sum of squares LSS, the fitted sum of squares FSS, and the residual sum of squares RSS.

(j) Determine a simple and natural formula for the usual unbiased estimate s^2 of σ^2.

(k) Suppose that $\boldsymbol{X}^T\boldsymbol{X} = \boldsymbol{V}$, where $\boldsymbol{V}$ is a diagonal matrix having positive diagonal entries $v_1, \ldots, v_d$, and hence that $\boldsymbol{X}^{-1} = \boldsymbol{V}^{-1}\boldsymbol{X}^T$ and $(\boldsymbol{X}^T)^{-1} = \boldsymbol{X}\boldsymbol{V}^{-1}$. Then

$$\operatorname{var}(\widehat{\beta}_j) = \frac{\sigma^2}{v_j^2}\left(\sum_{k=1}^{d} \frac{1}{n_k} g_j^2(\boldsymbol{x}'_k)\right), \qquad j = 1, \ldots, d.$$

Verify this formula either in general or in the special case that $d = 2$ and $j = 2$.

(l) Use the formula in (k) to obtain a simple and natural formula for $\mathrm{SE}(\widehat{\beta}_j)$.

(m) Suppose that $\boldsymbol{X}^T\boldsymbol{X} = \boldsymbol{V}$ as in (k) and also that $n_1 = \cdots = n_d = r$ and hence that $n = dr$. Let $1 \le d_0 < d$. Consider the hypothesis that $\beta_{d_0+1} = \cdots = \beta_d = 0$ or, equivalently, that the regression function belongs to the subspace G_0 of G spanned by $g_1, \ldots, g_{d_0}$, which subspace is assumed to contain the constant functions. Let FSS_0 and RSS_0 denote the fitted and residual sums of squares corresponding to G_0. Determine a simple and natural formula for $\mathrm{RSS}_0 - \mathrm{RSS} = \mathrm{FSS} - \mathrm{FSS}_0$ in terms of $\widehat{\beta}_{d_0+1}, \ldots, \widehat{\beta}_d$ and deterministic quantities (that is, quantities that involve the experimental design but not the responses variables or their observed values).

(n) Use the answers to (i) and (m) to obtain a simple and natural formula for the F statistic for testing the hypothesis in (m).

35. Consider a success-failure experiment involving three factors, each of which takes on a low value indicated by -1 and a high value indicated by 1, and four design points. The experimental design, numbers of runs at the various design points, and corresponding numbers of successes are shown in the following table:

k	x_1	x_2	x_3	# of runs	# of successes
1	-1	-1	1	20	15
2	-1	1	-1	40	25
3	1	-1	-1	40	28
4	1	1	1	80	32

Let π_k denote the probability of success corresponding to the kth design point, and set $\theta_k = \operatorname{logit}(\pi_k)$.

(a) Consider a saturated model for the logistic regression function. Determine the corresponding maximum likelihood estimates $\widehat{\pi}_k$ of π_k and $\widehat{\theta}_k$ of θ_k for $k = 1, 2, 3, 4$.

(b) (Continued) Determine a reasonable (approximate) 95% confidence interval for the odds ratio o_4/o_1, where $o_j = \pi_j/(1 - \pi_j)$ is the odds corresponding to π_j.

(c) (Continued) Consider the model $\theta(x_1, x_2, x_3) = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \beta_3 x_3$ for the logistic regression function. Determine the corresponding design matrix $\boldsymbol{X}$.

(d) (Continued) Determine the maximum likelihood estimates $\widehat{\beta}_0, \ldots, \widehat{\beta}_3$ of the regression coefficients. (You may express your answers in terms of $\widehat{\theta}_1, \ldots, \widehat{\theta}_4$ and avoid the implied arithmetic.)

(e) Consider the submodel of the model given in (c) obtained by ignoring x_2 and x_3: $\theta(x_1) = \beta_{00} + \beta_{01} x_1$. The maximum likelihood estimates of β_{00} and β_{01} are given by $\widehat{\beta}_{00} \doteq 0.347$ and $\widehat{\beta}_{01} \doteq -0.347$. Obtain these numerical values algebraically (that is, without using an iterative numerical method).

(f) Describe a reasonable test of the null hypothesis that the submodel is valid. In particular, indicate the name of the test; give a formula for the test statistic (for example, by refering to the appropriate formula from Chapter 13); explain how to obtain the various quantities in this formula; and describe the approximate distribution of this test statistic under the null hypothesis. But do not perform the numerical calculations that would be required for carrying out the test.

Fifth Practice Final Exam

36. Consider the task of giving a thirty-minute review lecture on (i) F tests in the context of homoskedastic normal linear regression models; (ii) likelihood ratio tests in the context of logistic regression models; and (iii) the similarities and differences between the F tests in (i) and the likelihood ratio tests in (ii). The coverage should be restricted to the relevant material in the textbook and the corresponding lectures. Write out a complete set of lecture notes that could be used for this purpose by yourself or by another student in the course.

37. Let $Y_1, Y_2, Y_3, Y_4, Y_5, Y_6, Y_7, Y_8, Y_9$ be independent, normally distributed random variables having common, unknown variance σ^2.

(a) Consider the homoskedastic normal two-sample model in which Y_1, Y_2, Y_3 have common, unknown mean μ_1 and $Y_4, Y_5, Y_6, Y_7, Y_8, Y_9$ have common, unknown mean μ_2. More or less as usual, set

$$\bar{Y}_1 = \frac{Y_1 + Y_2 + Y_3}{3} \quad \text{and} \quad \bar{Y}_2 = \frac{Y_4 + Y_5 + Y_6 + Y_7 + Y_8 + Y_9}{6};$$

also, set

$$S_1^2 = \frac{(Y_1 - \bar{Y}_1)^2 + (Y_2 - \bar{Y}_1)^2 + (Y_3 - \bar{Y}_1)^2}{2}$$

and

$$S_2^2 = \frac{(Y_4 - \bar{Y}_2)^2 + (Y_5 - \bar{Y}_2)^2 + (Y_6 - \bar{Y}_2)^2 + (Y_7 - \bar{Y}_2)^2 + (Y_8 - \bar{Y}_2)^2 + (Y_9 - \bar{Y}_2)^2}{5}.$$

Consider the null hypothesis H_0: $\mu_2 = \mu_1$ and the alternative hypothesis H_a: $\mu_2 \neq \mu_1$. Obtain a formula for the corresponding t statistic in terms of $\bar{Y}_1$, $\bar{Y}_2$, S_1^2 and S_2^2. Under H_0, what is the distribution of this statistic?

(b) Consider the experimental data:

i	1	2	3	4	5	6	7	8	9
x	0	0	0	1	1	1	1	1	1
Y	Y_1	Y_2	Y_3	Y_4	Y_5	Y_6	Y_7	Y_8	Y_9

In this part, and in the rest of this problem, consider the homoskedastic normal linear regression model in which $E(Y_i) = \beta_0 + \beta_1 x_i$. Determine the design matrix $\boldsymbol{X}$ (long $n \times p$ form, not short $d \times p$ form) corresponding to the indicated basis $1, x$.

(c) Determine the Gram matrix $\boldsymbol{X}^T\boldsymbol{X}$.

(d) Determine $\left(\boldsymbol{X}^T\boldsymbol{X}\right)^{-1}$.

(e) Determine the least squares estimates $\widehat{\beta}_0$ and $\widehat{\beta}_1$ of β_0 and β_1; show your work and express your answers in terms of $\bar{Y}_1$ and $\bar{Y}_2$.

(f) Determine the corresponding fitted sum of squares FSS; show your work and express your answer in terms of $\bar{Y}_1$ and $\bar{Y}_2$.

(g) Determine the usual estimate $\widehat{\boldsymbol{V}}$ of the variance-covariance matrix of $\widehat{\beta}_0$ and $\widehat{\beta}_1$.

(h) Use your answer to (g) to determine $\mathrm{SE}(\widehat{\beta}_1)$.

(i) Consider the null hypothesis H_0: $\beta_1 = 0$ and the alternative hypothesis H_a: $\beta_1 \neq 0$. Obtain a formula for the corresponding t statistic. Under H_0, what is the distribution of this statistic?

38. Consider an experiment involving three factors and four runs. The first factor takes on two levels: -1 (low) and 1 (high). Each of the other two factors takes on four levels: -3 (low), -1 (low medium), 1 (high medium), and 3 (high). The factor levels x_{km}, $m = 1, 2, 3$, time parameter n_k, and observed average response ($\bar{y}_k = y_k/n_k$) on the kth run are as follows:

k	x_1	x_2	x_3	n	$\bar{y}$
1	-1	-3	1	40	0.5
2	-1	3	-1	100	0.2
3	1	-1	-3	50	0.4
4	1	1	3	40	0.5

(a) Let G be the linear space having basis $1, x_1, x_2, x_3$. Determine the design matrix $\boldsymbol{X}$.

(b) Are the indicated basis functions orthogonal (relative to unit weights) (why or why not)?

(c) Determine the inverse matrix $\boldsymbol{X}^{-1}$ (show your work).

(d) Is G identifiable, and is G saturated (justified your answers)?

Consider the Poisson regression model corresponding to the indicated basis of G:

$$\theta(x_1, x_2, x_3) = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \beta_3 x_3,$$

where $\theta(\cdot) = \log \lambda(\cdot)$. Let $\widehat{\theta}(\cdot)$ and $\widehat{\lambda}(\cdot)$ denote the maximum likelihood estimates of $\theta(\cdot)$ and $\lambda(\cdot)$.

(e) Determine the maximum likelihood estimates $\widehat{\beta}_0, \widehat{\beta}_1, \widehat{\beta}_2, \widehat{\beta}_3$ of $\beta_0, \beta_1, \beta_2, \beta_3$.

(f) Is it true that $\widehat{\lambda}(\boldsymbol{x}_k) = \bar{y}_k$ for $k = 1, 2, 3, 4$ (why or why not))?

(g) Determine $\mathrm{SE}(\widehat{\beta}_3)$ and the corresponding 95% confidence interval for β_3. What does this say about the test of H_0: $\beta_3 = 0$ versus H_a: $\beta_3 \neq 0$?

Let G_0 be the subspace of G spanned by 1 and x_1.

(h) Explain the sense in which G_0 is identifiable and saturated.

(i) Carry out the likelihood ratio test of H_0: $\theta(\cdot) \in G_0$ versus H_a: $\theta(\cdot) \in G \backslash G_0$. In particular, determine the P-value for the test.

Brief Solutions to Practice Exams

Solutions to Second Practice First Midterm Exam

3. Let $\boldsymbol{x}'_1, \dots, \boldsymbol{x}'_d$ be the design points and let $g_1, \dots, g_p$ be a basis of the linear space G. Then the corresponding design matrix $\boldsymbol{X}$ is the $d \times p$ matrix given by

$$\boldsymbol{X} = \begin{bmatrix} g_1(\boldsymbol{x}'_1) & \cdots & g_p(\boldsymbol{x}'_1) \\ \vdots & & \vdots \\ g_1(\boldsymbol{x}'_d) & \cdots & g_p(\boldsymbol{x}'_d) \end{bmatrix}.$$

Given $g = b_1 g_1 + \cdots + b_p g_p \in G$, set $\boldsymbol{b} = [b_1, \dots, b_p]^T$. Then

$$\begin{bmatrix} g(\boldsymbol{x}'_1) \\ \vdots \\ g(\boldsymbol{x}'_d) \end{bmatrix} = \boldsymbol{X}\boldsymbol{b}.$$

The linear space G is identifiable if and only if the only vector $\boldsymbol{b} \in \mathbb{R}^p$ such that $\boldsymbol{X}\boldsymbol{b} = \boldsymbol{0}$ is the zero vector or, equivalently, if and only if the columns of the design matrix are linearly independent, in which case $p \le d$. The space G is saturated if and only if, for every $\boldsymbol{c} \in \mathbb{R}^d$, there is at least one $\boldsymbol{b} \in \mathbb{R}^p$ such that $\boldsymbol{X}\boldsymbol{b} = \boldsymbol{c}$ or, equivalently, if and only if the column vectors of the design matrix span $\mathbb{R}^d$, in which case $p \ge d$. The space G is identifiable and saturated if and only if $p = d$ and the design matrix is invertible. Let $\widetilde{g}_1, \dots, \widetilde{g}_p$ be an alternative basis of G, where $\widetilde{g}_j = a_{j1} g_1 + \cdots + a_{jp} g_p$ for $1 \le j \le p$, and let

$$\boldsymbol{A} = \begin{bmatrix} a_{11} & \cdots & a_{pp} \\ \vdots & & \vdots \\ a_{p1} & \cdots & a_{pp} \end{bmatrix}$$

be the corresponding coefficient matrix. Also, let $\widetilde{\boldsymbol{X}}$ be the design matrix for the alternative basis. Then $\widetilde{\boldsymbol{X}} = \boldsymbol{X}\boldsymbol{A}^T$. The Gram matrix $\boldsymbol{M}$ corresponding to the original basis is given by $\boldsymbol{M} = \boldsymbol{X}^T\boldsymbol{W}\boldsymbol{X}$, where $\boldsymbol{W} = \operatorname{diag}(w_1, \dots, w_d)$ is the weight matrix. Suppose that G is identifiable. Given a function h that is defined at least on the design set, set $\boldsymbol{h} = [h(\boldsymbol{x}'_1), \dots, h(\boldsymbol{x}'_d)]^T$, let $h^* = P_G h = b_1^* g_1 + \cdots + b_p^* g_p$ be the least squares approximation in G to h, and set $\boldsymbol{b}^* = [b_1^*, \dots, b_p^*]^T$. Then $\boldsymbol{b}^*$ satisfies the normal equation $\boldsymbol{X}^T\boldsymbol{W}\boldsymbol{X}\boldsymbol{b}^* = \boldsymbol{X}^T\boldsymbol{W}\boldsymbol{h}$, whose unique solution is given by $\boldsymbol{b}^* = (\boldsymbol{X}^T\boldsymbol{W}\boldsymbol{X})^{-1}\boldsymbol{X}^T\boldsymbol{W}\boldsymbol{h}$. In the context of the linear experimental model, let $\bar{Y}_k$ denote the average response at the kth design point, which has mean $\mu(\boldsymbol{x}'_k)$, where $\mu(\cdot) = \beta_1 g_1 + \cdots + \beta_p g_p$ is the regression function. Set $\bar{\boldsymbol{Y}} = [\bar{Y}_1, \dots, \bar{Y}_k]^T$ and $\boldsymbol{\beta} = [\beta_1, \dots, \beta_p]^T$. Then $E(\bar{\boldsymbol{Y}}) = \boldsymbol{X}\boldsymbol{\beta}$.

4. (a) We can think of the five varieties used in the experiment as corresponding to levels 1, 2, 3, 4, and 5 of the second factor and of the sixth variety as corresponding to level 3/2 of this factor. Then $\tau = \mu(3/2) = \mu_1 g_1(3/2) + \mu_2 g_2(3/2) + \mu_3 g_3(3/2) + \mu_4 g_4(3/2) + \mu_5 g_5(3/2)$. Here
$g_1(3/2) = \frac{(3/2-2)(3/2-3)(3/2-4)(3/2-5)}{(1-2)(1-3)(1-4)(1-5)} = \frac{(-1/2)(-3/2)(-5/2)(-7/2)}{(-1)(-2)(-3)(-4)} = \frac{35}{128}$;

$g_2(3/2) = \frac{(3/2-1)(3/2-3)(3/2-4)(3/2-5)}{(2-1)(2-3)(2-4)(2-5)} = \frac{(1/2)(-3/2)(-5/2)(-7/2)}{(1)(-1)(-2)(-3)} = \frac{140}{128}$;
$g_3(3/2) = \frac{(3/2-1)(3/2-2)(3/2-4)(3/2-5)}{(3-1)(3-2)(3-4)(3-5)} = \frac{(1/2)(-1/2)(-5/2)(-7/2)}{(2)(1)(-1)(-2)} = -\frac{70}{128}$;
$g_4(3/2) = \frac{(3/2-1)(3/2-2)(3/2-3)(3/2-5)}{(4-1)(4-2)(4-3)(4-5)} = \frac{(1/2)(-1/2)(-3/2)(-7/2)}{(3)(2)(1)(-1)} = \frac{28}{128}$;
$g_5(3/2) = \frac{(3/2-1)(3/2-2)(3/2-3)(3/2-4)}{(5-1)(5-2)(5-3)(5-4)} = \frac{(1/2)(-1/2)(-3/2)(-5/2)}{(4)(3)(2)(1)} = -\frac{5}{128}$.

(b) $\hat{\tau} \doteq \left(\frac{35}{128}\right)(18.136)+\left(\frac{140}{128}\right)(17.58)-\left(\frac{70}{128}\right)(18.32)+\left(\frac{28}{128}\right)(18.06)-\left(\frac{5}{128}\right)(18.13) \doteq 17.411$.

(c) Now $\left(\frac{35}{128}\right)^2 \cdot \frac{1}{11} + \left(\frac{140}{128}\right)^2 \cdot \frac{1}{10} + \left(\frac{70}{128}\right)^2 \cdot \frac{1}{10} + \left(\frac{28}{128}\right)^2 \cdot \frac{1}{10} + \left(\frac{5}{128}\right)^2 \cdot \frac{1}{10} \doteq 0.161$, so $\mathrm{SE}(\hat{\tau}) \doteq 0.423\sqrt{0.161} \doteq 0.170$.

(d) Since $t_{.975,46} \doteq 2.013$, the 95% confidence interval for τ is given by $17.411 \pm (2.013)(0.170) \doteq 17.411 \pm 0.342 = (17.069, 17.753)$.

(e) Set $\tau_1 = \tau - \mu_1 = -\frac{93}{128}\mu_1 + \frac{140}{128}\mu_2 - \frac{70}{128}\mu_3 + \frac{28}{128}\mu_4 - \frac{5}{128}\mu_5$. Then $\hat{\tau}_1 = \hat{\tau} - \bar{Y}_1 \doteq 17.411 - 18.136 = -0.725$. Now $\left(\frac{93}{128}\right)^2 \cdot \frac{1}{11} + \left(\frac{140}{128}\right)^2 \cdot \frac{1}{10} + \left(\frac{70}{128}\right)^2 \cdot \frac{1}{10} + \left(\frac{28}{128}\right)^2 \cdot \frac{1}{10} + \left(\frac{5}{128}\right)^2 \cdot \frac{1}{10} \doteq 0.202$, so $\mathrm{SE}(\hat{\tau}_1) \doteq 0.423\sqrt{0.202} \doteq 0.190$. The t statistic for testing the hypothesis that $\tau = \mu_1$ or, equivalently, that $\tau_1 = 0$ is given by $t \doteq -\frac{0.725}{0.190} \doteq -3.816$. The corresponding P-value is given by P-value $\doteq 2t_{46}(-3.813) < 2(.001) = .002$.

(f) Now $\tau = \mu_1 = \cdots = \mu_5$ if and only if $\mu_1 = \cdots = \mu_5$. Thus the desired test is given by the solution to Example 7.24.

Solutions to Third Practice First Midterm Exam

6. (a) Set $\tau = \mu_2 - \mu_1$. The hypothesis is that $\tau \leq -1$. Now $\hat{\tau} = \bar{Y}_2 - \bar{Y}_1 \doteq 17.580 - 18.136 = -0.556$ and $\mathrm{SE}(\hat{\tau}) = S\sqrt{\frac{1}{11} + \frac{1}{10}} = 0.423\sqrt{\frac{1}{11} + \frac{1}{10}} \doteq 0.185$. Consequently, the t statistic is given by $t = \frac{\hat{\tau}-\tau_0}{\mathrm{SE}(\hat{\tau})} \doteq \frac{-0.556-(-1)}{0.185} = 2.4$. Thus P-value $= 1 - t_{46}(2.4) < .01$, so the test of size $\alpha = .01$ rejects the hypothesis.

(b) The F statistic is given by

$$F = \frac{[11(18.136-18)^2 + 10(17.580-17)^2 + 10(18.320-19)^2]/3}{0.179} \doteq 15.254.$$

Thus P-value $= 1 - F_{3,46}(15.254) \ll .01$, so the hypothesis is rejected.

7. (a) $P(x) = x^2 - \frac{\langle 1, x^2 \rangle}{\|1\|^2} 1 - \frac{\langle x, x^2 \rangle}{\|x\|^2} x = x^2 - \frac{70}{6} = x^2 - \frac{35}{3}$.

(b) $\hat{b}_0 = \frac{\langle 1, \bar{y}(\cdot) \rangle}{\|1\|^2} = \frac{y_1+y_2+y_3+y_4+y_5+y_6}{6} = \bar{y}$ and $\hat{b}_1 = \frac{\langle x, \bar{y}(\cdot) \rangle}{\|x\|^2} = \frac{-5y_1-3y_2-y_3+y_4+3y_5+5y_6}{70}$. Moreover, $P(-5) = P(5) = 25 - \frac{35}{3} = \frac{40}{3}$, $P(-3) = P(3) = 9 - \frac{35}{3} = -\frac{8}{3}$, and $P(-1) = P(1) = 1 - \frac{35}{3} = -\frac{32}{3}$, so $\|P(x)\|^2 = \frac{2(40^2+8^2+32^2)}{9} = \frac{1792}{3}$. Consequently, $\hat{b}_2 = \frac{\langle P(x), \bar{y}(\cdot) \rangle}{\|P(x)\|^2} = \frac{40y_1-8y_2-32y_3-32y_4-8y_5+40y_6}{1792} = \frac{5y_1-y_2-4y_3-4y_4-y_5+5y_6}{224}$.

Solutions to Fourth Practice First Midterm Exam

9. (a) According to Table 7.3, $\widehat{\tau} = \frac{\bar{Y}_1+\bar{Y}_3+\bar{Y}_5}{3} - \frac{\bar{Y}_1+\bar{Y}_2}{2} = \frac{18.136+18.330+18.130}{3} - \frac{17.580+18.060}{2} \doteq 18.199 - 17.820 = 0.379$. Also, $\mathrm{var}(\widehat{\tau}) = \sigma^2\left(\frac{1}{9}\cdot\frac{1}{11} + \frac{1}{9}\cdot\frac{1}{10} + \frac{1}{9}\cdot\frac{1}{10} + \frac{1}{4}\cdot\frac{1}{10} + \frac{1}{4}\cdot\frac{1}{10}\right) \doteq .08232\sigma^2$. According to page 361 of the textbook, the pooled sample standard deviation is given by $S \doteq 0.413$. Thus the standard error of $\widehat{\tau}$ is given by $\mathrm{SE}(\widehat{\tau}) \doteq 0.413\sqrt{.08232} \doteq 0.1185$. According to the table on page 830 of the textbook, $t_{.95,46} \approx t_{.95,45} \doteq 1.679$. Thus the 95% lower confidence bound for τ is given by $0.379 - (1.679)(0.1185) \doteq 0.379 - 0.199 = 0.180$.

(b) The t statistic for the test is given by $t \doteq 0.379/0.1185 \doteq 3.20$. According to the table on page 830, $.001 < P\text{-value} < .005$.

(c) The key point is that the usual statistical interpretations for confidence intervals, confidence bounds, and P-values are valid only when the underlying model, the parameter of interest, the form of the confidence interval or confidence bound, and the form of the null and alternative hypotheses are specified in advance of examining the data. In the context of the present problem, the parameter τ was selected after examining the data. Thus, it cannot be concluded that the probability (suitably interpreted) that the 95% lower confidence bound for τ would be less than τ is .95, and it cannot be concluded that if, say, $\tau = 0$, then the probability that the P-value for the t test of the null hypothesis that $\tau \le 0$ would be less than .05 would be .05.

10. (a) The design matrix is given by $\boldsymbol{X} = \begin{bmatrix} 1 & -5 & 2 \\ 1 & -3 & -1 \\ 1 & -1 & -1 \\ 1 & 1 & -1 \\ 1 & 3 & -1 \\ 1 & 5 & 2 \end{bmatrix}$.

(b) Since G has dimension $p = 3$ and there are $d = 6$ design points, $p < d$, so G is not saturated.

(c) Consider a function $g = b_0 + b_1x_1 + b_2x_2 \in G$ that equals zero on the design set. Then, in particular, $b_0 + b_1 - b_2 = 0$; $b_0 + 3b_1 - b_2 = 0$; $b_0 + 5b_1 - 2b_2 = 0$. It follows from the first two of these equations that $b_1 = 0$. Consequently, by the last two equations, $b_2 = 0$; hence $b_0 = 0$. Thus $g = 0$. Therefore, G is identifiable.

(d) Since $\sum x_{i1} = -5 + \cdots + 5 = 0$, $\sum x_{i2} = 2 + \cdots + 2 = 0$, and $\sum x_{i1}x_{i2} = -10 + \cdots + 10 = 0$, the indicated basis functions are orthogonal.

(e) Observe that $\sum_i 1 = 6$, $\sum_i x_{i1}^2 = 70$, and $\sum_i x_{i2}^2 = 12$. Observe also that $\sum_i y_i = y_1 + y_2 + y_3 + y_4 + y_5 + y_6 = 6\bar{y}$; $\sum_i x_{i1}y_i = 5(y_6 - y_1) + 3(y_5 - y_2) + y_4 - y_3$; $\sum_i x_{i2}y_i = 2(y_1 + y_6) - (y_2 + y_3 + y_4 + y_5)$. Thus, $\widehat{b}_0 = \bar{y}$, $\widehat{b}_1 = [5(y_6 - y_1) + 3(y_5 - y_2) + y_4 - y_3]/70$, and $\widehat{b}_2 = [2(y_1 - y_6) - (y_2 + y_3 + y_4 + y_5)]/12$.

Solutions to Fifth Practice First Midterm Exam

12. (a) Every function of the indicated form is linear on $(-\infty, 0]$ and on $[0, \infty)$. Suppose, conversely, that g is linear on $(-\infty, 0]$ and on $[0, \infty)$. Then there are constants A_0, B, A_1 and C such that $g(x) = A_0 + Bx$ for $x \le 0$ and $g(x) = A_1 + Cx$ for $x \ge 0$. Since $g(0) = A_0$ and $g(0) = A_1$, we conclude that $A_0 = A_1 = A$ and hence that g has the indicated form.

(b) Let $g \in G$ have the form in (a) and let $b \in \mathbb{R}$. Then

$$bg(x) = \begin{cases} bA + bBx & \text{for } x \le 0, \\ bA + bCx & \text{for } x \ge 0, \end{cases}$$

so bg has the indicated form and hence it is in G. Let $g_1 \in G$ and $g_2 \in G$ each have the indicated form:

$$g_1(x) = \begin{cases} A_1 + B_1 x & \text{for } x \ge 0, \\ A_1 + C_1 x & \text{for } x \le 0, \end{cases} \quad \text{and} \quad g_2(x) = \begin{cases} A_2 + B_2 x & \text{for } x \ge 0, \\ A_2 + C_2 x & \text{for } x \le 0. \end{cases}$$

Then

$$g_1(x) + g_2(x) = \begin{cases} A_1 + A_2 + (B_1 + B_2)x & \text{for } x \ge 0, \\ A_1 + A_2 + (C_1 + C_2)x & \text{for } x \le 0, \end{cases}$$

so $g_1 + g_2 \in G$. Therefore G is a linear space.

(c) Clearly, the functions 1, x and x_+ are members of G. Suppose that $b_0 + b_1 x + b_2 x_+ = 0$ for $x \in \mathbb{R}$. Then $b_0 + b_1 x = 0$ for $x \le 0$, so $b_1 = 0$ and hence $b_2 = 0$. Similarly, $b_1 + b_2 x = 0$ for $x \ge 0$, so $b_2 = 0$. Thus 1, x and x_+ are linearly independent.

(d) Let $g \in G$ have the form given in (a). Then $g = A + Bx + (C - B)x_+$. Thus 1, x and x_+ span G, Since these functions are linearly independent, they form a basis of G.

(e) Since $x_+ = (x + |x|)/2$, the functions 1, x and x_+ are linear combinations of 1, x and $|x|$, so 1, x and $|x|$ form a basis of G. Now $|x| = 2x_+ - x$. Thus the coefficient matrix of this new basis relative to the old basis 1, x and x_+ is given by

$$\boldsymbol{A} = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & -1 & 2 \end{bmatrix}.$$

(f) Now $\|1\|^2 = 5$, $\|x\|^2 = 10$ and $\|\,|x|\,\|^2 = 10$. Also, $\langle 1, x \rangle = 0$, $\langle 1, |x| \rangle = 6$, and $\langle x, |x| \rangle = 0$. Thus the Gram matrix is given by

$$\boldsymbol{M} = \begin{bmatrix} 5 & 0 & 6 \\ 0 & 10 & 0 \\ 6 & 0 & 10 \end{bmatrix}.$$

(g) One such function g is given by

$$|x| - \frac{\langle 1, |x| \rangle}{\|1\|^2} 1 - \frac{\langle x, |x| \rangle}{\|x\|^2} x = |x| - \frac{6}{5}.$$

(h) The squared norms are given by $\|1\|^2 = 5$, $\|x\|^2 = 10$ and

$$\|\,|x| - 6/5\,\|^2 = (4/5)^2 + (-1/5)^2 + (-6/5)^2 + (-1/5)^2 + (4/5)^2$$
$$= 70/25 = 14/5.$$

(i) Now $\langle 1, x^2\rangle = \sum_x x^2 = 2^2 + 1^2 + 0^2 + 1^2 + 2^2 = 10$, $\langle x, x^2\rangle = 0$, and

$$\langle |x| - 6/5, x^2\rangle = \sum_x |x|^3 - \frac{6}{5}\sum_x x^2 = (2^3 + 1^3 + 0^3 + 1^3 + 2^3) - \frac{6}{5}\cdot 10$$
$$= 18 - 12 = 6.$$

Thus

$$P_G(x^2) = \frac{\langle 1, x^2\rangle}{\|1\|^2}1 + \frac{\langle x, x^2\rangle}{\|x\|^2}x + \frac{\langle |x| - 6/5, x^2\rangle}{\|\,|x| - 6/5\,\|^2}\Big(|x| - \frac{6}{5}\Big)$$
$$= \frac{10}{5} + \frac{6}{14/5}\Big(|x| - \frac{6}{5}\Big) = \frac{15}{7}|x| - \frac{4}{7}.$$

As a check,

$$\langle\, 1,\, x^2 - P_G(x^2)\,\rangle = \Big\langle 1,\, x^2 - \frac{15}{7}|x| + \frac{4}{7}\Big\rangle = 10 - \frac{15}{7}\cdot 6 + \frac{4}{7}\cdot 5 = 0.$$

Solutions to First Practice Second Midterm Exam

14. (a) Yes, X_1 and X_2 are each uniformly distributed on $\{-1, 1\}$, while X_3 and X_4 are each uniformly distributed on $\{-1, 0, 1\}$.

(b) The design random variables X_3 and X_4 are not independent. In particular, $P(X_3 = 0, X_4 = 0) = 1/6 \neq P(X_3 = 0)P(X_4 = 0$.

(c) No, because X_3 and X_4 are not independent.

(d) Factors A, B and C take on each of the 12 possible factor level combinations once, so X_1, X_2 and X_3 are independent. Similarly, X_1, X_2 and X_4 are independent.

(e) $\|x_1\|^2 = \|x_2\|^2 = 12$, $\|x_3\|^2 = \|x_4\|^2 = 8$, $\|x_4^2 - 2/3\|^2 = 8(1/3)^2 + 4(2/3)^2 = 8/3$, and $\langle x_3, x_4^2 - 2/3\rangle = 2$.

(f) $\hat\beta_0 = \bar Y$, $\hat\beta_1 = T_1/12$, $\hat\beta_2 = T_2/12$, and $\hat\beta_4 = T_4/8$. Moreover, $8\hat\beta_3 + 2\hat\beta_5 = T_3$ and $2\hat\beta_3 + (8/3)\hat\beta_5 = T_5$, whose unique solution is given by $\hat\beta_3 = 2T_3/13 - 3T_5/26$ and $\hat\beta_5 = -3T_3/26 + 6T_5/13$.

(g) $\mathrm{FSS} = 12\hat\beta_1^2 + 12\hat\beta_2^2 + 8\hat\beta_3^2 + 8\hat\beta_4^2 + (8/3)\hat\beta_5^2 + 4\hat\beta_3\hat\beta_5$ and $\mathrm{RSS} = \mathrm{TSS} - \mathrm{FSS}$.

Source	SS	DF	MS	F	P-value
Fit	FSS	5	FSS/5	$\mathrm{MS_{fit}}/\mathrm{MS_{res}}$	$1 - F_{5,6}(F)$
Residuals	RSS	6	RSS/6		
Total	TSS	11			

The null hypothesis being tested by the F statistic is that $\beta_1 = \cdots = \beta_5 = 0$.

(h) $S^2 = \text{RSS}/6$, $S = \sqrt{S^2}$, and $\text{SE}(\hat\beta_5) = S\sqrt{6/13} \doteq 0.679S$. The 95% confidence interval for β_5 is $\hat\beta_5 \pm t_{.975,6}\text{SE}(\hat\beta_5)$. The P-value for testing the null hypothesis that $\beta_5 = 0$ is given by $P-\text{value} = 2[1-t_6(|t|)]$, where $t = \hat\beta_5/\text{SE}(\hat\beta_5)$.

(i) It follows from Theorem 10.16 that $\frac{\text{RSS}_0-\text{RSS}}{S^2} = t^2 = \left(\frac{\hat\beta_5}{\text{SE}(\hat\beta_5)}\right)^2 = \frac{13\hat\beta_5^2}{6S^2}$ and hence that $\text{RSS}_0 - \text{RSS} = 13\hat\beta_5^2/6$. Alternatively, the coefficient of x_3 corresponding to the least squares fit in the subspace is given by $\hat\beta_{30} = T_3/8$. Also, $\text{RSS}_0 - \text{RSS} = \text{FSS} - \text{FSS}_0 = 8\hat\beta_3^2 + 8\hat\beta_5^2/3 + 4\hat\beta_3\hat\beta_5 - 8\hat\beta_{30}^2$. Thus $8\hat\beta_3^2 + 16\hat\beta_5^2/6 + 4\hat\beta_3\hat\beta_5 - 8\hat\beta_{30}^2 = 13\hat\beta_5^2/6$ or, equivalently, $16\hat\beta_3^2 + 8\hat\beta_3\hat\beta_5 + \hat\beta_5^2 = 16\hat\beta_{30}^2$. As a check, the left side of this equation is given by $[16(4T_3-3T_5)^2+8(4T_3-3T_5)(12T_5-3T_3)+(12T_5-3T_3)^2]/26^2 = T_3^2/4 = 16\hat\beta_{30}^2$.

Solutions to Second Practice Second Midterm Exam

15. In the homoskedastic normal linear experimental model, G is a p-dimensional linear space of functions on $\mathcal{X}$; $Y_1,\dots,Y_n$ are independent rendom variables that are normally distributed with common, unknown variance σ^2; and $EY_i = \mu(\boldsymbol{x}_i)$ for $1 \le i \le n$, where the regression function $\mu(\cdot)$ is an unknown member of G. We also assume that G is identifiable: if $g \in G$ and $g(\boldsymbol{x}_i) = 0$ for $1 \le i \le n$, then $g = 0$.

Let $\widehat\mu(\cdot)$ denote the least squares estimate in G of the regression function, and let $\text{RSS} = \sum_i [Y_i - \widehat\mu(\boldsymbol{X}_i)]^2$ denote the corresponding residual sum of squares. Given a basis $g_1,\dots,g_p$ of G, write $\mu(\cdot) = \beta_1 g_1 + \cdot + \beta_p g_p$ and $\widehat\mu(\cdot) = \widehat\beta_1 g_1 + \cdot + \widehat\beta_p g_p$. Set $\boldsymbol{\beta} = [\beta_1,\dots,\beta_p]^T$ and $\widehat{\boldsymbol{\beta}} = [\widehat\beta_1,\dots,\widehat\beta_p]^T$. Let $\boldsymbol{X}$ denote the design matrix corresponding to the indicated basis and set $\boldsymbol{Y} = [Y_1,\dots,Y_n]^T$. Then $\widehat{\boldsymbol{\beta}} = (\boldsymbol{X}^T\boldsymbol{X})^{-1}\boldsymbol{X}^T\boldsymbol{Y}$ and $\widehat{\boldsymbol{\beta}}$ has the multivariate normal distribution with mean $\boldsymbol{\beta}$ and variance-covariance matrix $\sigma^2(\boldsymbol{X}^T\boldsymbol{X})^{-1}$. Consider a linear parameter $\tau = \boldsymbol{c}^T\boldsymbol{\beta}$ and its least squares estimate $\widehat\tau = \boldsymbol{c}^T\widehat{\boldsymbol{\beta}}$, where $\boldsymbol{c} = [c_1,\dots,c_p]^T$. The estimate $\widehat\tau$ is normally distributed with mean τ and variance $\sigma^2\boldsymbol{c}^T(\boldsymbol{X}^T\boldsymbol{X})^{-1}\boldsymbol{c}$. Its standard error is given by $\text{SE}(\widehat\tau) = S\sqrt{\boldsymbol{c}^T(\boldsymbol{X}^T\boldsymbol{X})^{-1}\boldsymbol{c}}$, where $S = \sqrt{\text{RSS}/(n-p)}$ is the usual estimate of σ.

Let $\tau_0 \in \mathbb{R}$ and set $t = (\widehat\tau - \tau_0)/\text{SE}(\widehat\tau)$. The P-value for testing the hypotheses that $\tau = \tau_0$, $\tau \le \tau_0$, and $\tau \ge \tau_0$ are given by $2t_{n-p}(-|t|)$, $1 - t_{n-p}(t)$, and $t_{n-p}(t)$, respectively. Let G_0 be a p_0-dimensional subspace of G, where $p_0 < p$ and let RSS_0 denote the residual sum of squares for the least squares fit in G_0. The F statistic for testing the hypothesis that the regression function is in G_0 is given by

$$F = \frac{(\text{RSS}_0 - \text{RSS})/(p-p_0)}{\text{RSS}/(n-p)},$$

and the corresponding P-value equals $1 - F_{p-p_0,n-p}(F)$.

Suppose, in particular, that G_0 is the $p-1$-dimensional linear subspace of G spanned by $g_1,\dots,g_p$. Then the F statistic for testing the hypothesis that the regression function is in G_0 is given by

$$F = \frac{\text{RSS}_0 - \text{RSS}}{\text{RSS}/(n-p)}$$

and the t statistic for testing the equivalent hypothesis that $\beta_p = 0$ is given by $t = \widehat{\beta}_p/\mathrm{SE}(\widehat{\beta}_p)$, where $\mathrm{SE}(\widehat{\beta}_p) = S\sqrt{[(\boldsymbol{X}^T\boldsymbol{X})^{-1}]_{pp}}$. Here $F = t^2$ and the P-values for the two tests coincide.

Suppose that G contains the constant functions and let $\mathrm{FSS} = \|\widehat{\mu}(\cdot) - \bar{Y}\|^2 = \sum_i [\widehat{\mu}(\boldsymbol{X}_i) - \bar{Y}]^2$ denote the corresponding fitted sum of squares. Then the F statistic for testing the hypothesis that the regression function is constant is given by

$$F = \frac{\mathrm{FSS}/(p-1)}{\mathrm{RSS}/(n-p)}$$

and the corresponding P-value equals $1 - F_{p-1,n-p}(F)$.

Suppose that $p < d < n$ and let $\mathrm{LSS} = \|\bar{Y}(\cdot) - \widehat{\mu}(\cdot)\|^2 = \sum_k [\bar{Y}_k - \widehat{\mu}(\boldsymbol{x}_k)]^2$ denote the lack-of-fit sum of squares for the least squares fit in G. Then the F statistic for the goodness-of-fit test of the assumption that the regression function is in G is given by

$$F = \frac{\mathrm{LSS}/(d-p)}{\mathrm{WSS}/(n-d)}$$

and the corresponding P-value equals $1 - F_{d-p,n-d}(F)$.

16. (a) The design random variables X_1, X_2 and X_3 are independent since each of the 24 factor-level combinations involving these three factors occurs on exactly one run; the design random variables X_1, X_2 and X_4 are independent for the same reason; the design random variables X_1, X_3 and X_4 are dependent since there are 16 factor-level combinations involving these three factors and 24 runs.

(b) $\|1\|^2 = 24 \cdot 1^2 = 24$, $\|x_1\|^2 = 12 \cdot 3^2 + 12 \cdot 1^2 = 120$, $\|x_1^2 - 5\|^2 = 24 \cdot 4^2 = 384$, $\|x_2\|^2 = 16 \cdot 1^2 + 8 \cdot 0^2 = 16$, $\|x_3\|^2 = 24 \cdot 1^2 = 24$, $\|x_4\|^2 = 24 \cdot 1^2 = 24$, $\|x_1x_3\|^2 = \|x_1\|^2 = 120$.

(c) $\langle x_1x_3, x_4\rangle = 4(-3) + 8(-1) + 4(1) + 8(3) = 8$.

(d) $$\boldsymbol{X}^T\boldsymbol{X} = \begin{bmatrix} 24 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 120 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 384 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 16 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 24 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 24 & 8 \\ 0 & 0 & 0 & 0 & 0 & 8 & 120 \end{bmatrix}$$

$$(\boldsymbol{X}^T\boldsymbol{X})^{-1} = \begin{bmatrix} 1/24 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 1/120 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 1/384 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1/16 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1/24 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 15/352 & -1/352 \\ 0 & 0 & 0 & 0 & 0 & -1/352 & 3/352 \end{bmatrix}$$

(e) $\widehat{\beta}_0 = \frac{4800}{24} = 200$, $\widehat{\beta}_1 = \frac{600}{120} = 5$, $\widehat{\beta}_2 = -\frac{384}{384} = -1$,
$\widehat{\beta}_3 = -\frac{256}{16} = -16$, $\widehat{\beta}_4 = -\frac{336}{24} = -14$,
$\widehat{\beta}_5 = \frac{15}{352}(-304) - \frac{1}{352}(-336) = -12$, $\widehat{\beta}_6 = -\frac{1}{352}(-304) + \frac{3}{352}(-336) = -2$.

(f) The fitted sum of squares is given by $\mathrm{FSS} = 5^2 \cdot 120 + 1^2 \cdot 384 + 16^2 \cdot 16 + 14^2 \cdot 24 + 12^2 \cdot 24 + 2^2 \cdot 120 + 2(-12)(-2)8 = 16504$.

(g) The residual sum of squares is given by $\mathrm{RSS} = 30480 - 16504 = 13976$.

(h) When we eliminate the basis functions $x_1^2 - 5$ and x_1x_3, the coefficients of 1, x_1, x_2, and x_3 in the least squares fit don't change, but the coefficient of x_4 changes to $-304/24 = -38/3$. Thus the fitted sum of squares for the least squares linear fit is given by $\mathrm{FSS}_0 = 5^2 \cdot 120 + 16^2 \cdot 16 + 14^2 \cdot 24 + (38/3)^2 \cdot 24 \doteq 15650.7$. The F statistic for testing the indicated hypothesis is given by $F \doteq \frac{(16504-15650.7)/2}{13976/17} \doteq 0.519$. Thus the F test of size .05 accepts the hypothesis.

(i) The only two basis functions that are candidates for removal, assuming that we restrict attention to hierarchical submodels, are $x_1^2 - 5$ and x_1x_3. The standard deviation σ is estimated by $S = \sqrt{13976/17} \doteq 28.673$. The standard error of $\widehat{\beta}_2$ is given by $\mathrm{SE}(\widehat{\beta}_2) \doteq 28.673/\sqrt{384} \doteq 1.463$. The corresponding t statistic is given by $t \doteq -1/1.463 \doteq -0.684$. The standard error of $\widehat{\beta}_6$ is given by $\mathrm{SE}(\widehat{\beta}_6) \doteq 28.673\sqrt{3/352} \doteq 2.647$. The corresponding t statistic is given by $t \doteq -2/2.647 \doteq -0.756$. Thus we should remove the basis function $x_1^2 - 5$ since its t statistic is smaller in magnitude and hence its P-value is larger.

Solutions to Third Practice Second Midterm Exam

18. (a) The design random variables X_4 and X_5 are independent and uniformly distributed on $\{-1, 0, 1\}$. Thus $\|x_4x_5\|^2 = 36E(X_4^2X_5^2) = 36E(X_4^2)E(X_5^2) = 36 \cdot \frac{2}{3} \cdot \frac{2}{3} = 16$.

(b) Since X_1 and X_2 are independent and uniformly distributed on $\{-1, 1\}$, $\langle x_2, x_1x_2 \rangle = 36E(X_1X_2^2) = 36E(X_1)E(X_2^2) = 0$.

(c) Since $E(X_3) = E(X_4) = E(X_5) = 0$ and $6 = \langle x_3, x_4x_5 \rangle = 36E(X_3X_4X_5) \neq 36E(X_3)E(X_4)E(X_5)$, the design random variables X_3, X_4 and X_5 are not independent. Alternatively, $P(X_3 = -1, X_4 = -1, X_5 = -1) = 1/36$, but $P(X_3 = -1)P(X_4 = -1)P(X_5 = -1) = (1/2)(1/3)(1/3) = 1/18$, so X_3, X_4 and X_5 are not independent.

(d) Since $P(X_4 = 1, X_5 = 1, X_6 = 1)$ is a multiple of $1/36$ and $P(X_4 = 1)P(X_5 = 1)P(X_6 = 1) = 1/27$, which is not a multiple of $1/36$, the design random variables X_4, X_5 and X_6 are not independent.

(e) $\hat{\beta}_4 = \frac{(96)(-3414)+(270)(3036)-36(-1090)}{36 \cdot 238} = 62$.

(f) $S = \sqrt{\frac{\mathrm{RSS}}{n-p}} = \sqrt{\frac{1325184}{27}} \doteq 221.54$.

(g) $\mathrm{SE}(\hat{\beta}_4) \doteq 221.54\sqrt{\frac{270}{36 \cdot 238}} \doteq 39.327$, $t \doteq \frac{62}{39.33} \doteq 1.577$, and P-value $\doteq 2[1 - t_{27}(1.577)] \doteq .126$. Thus the hypothesis that $\beta_4 = 0$ is accepted.

(h) $\hat\beta_0 = \frac{19080}{36} = 530$, $\hat\beta_1 = \frac{2700}{36} = 75$, and $\hat\beta_2 = \frac{-2556}{36} = -71$. Moreover, $36\hat\beta_3 - 12\hat\beta_4 = -3414$ and $-12\hat\beta_3 + 36\hat\beta_4 = 3036$, so $\hat\beta_3 = \dfrac{\begin{vmatrix} -3414 & -12 \\ 3036 & 36 \end{vmatrix}}{\begin{vmatrix} 36 & -12 \\ -12 & 36 \end{vmatrix}} = \frac{-86472}{1152} = -75.0625$ and $\hat\beta_4 = \dfrac{\begin{vmatrix} 36 & -3414 \\ -12 & 3036 \end{vmatrix}}{\begin{vmatrix} 36 & -12 \\ -12 & 36 \end{vmatrix}} = \frac{68328}{1152} = 59.3125$.

(i) $\text{FSS} = 36\hat\beta_1^2 + 36\hat\beta_2^2 + 36\hat\beta_3^2 + 36\hat\beta_4^2 - 24\hat\beta_3\hat\beta_4 \doteq 820312$, so $\text{RSS} = \text{TSS} - \text{FSS} \doteq 2363560 - 820312 = 1543248$.

(j)

Source	SS	DF	MS	F	P-value
Fit in $\mathbb{G}_0$	820312	4	205078		
$\mathbb{G}$ after $\mathbb{G}_0$	218064	4	54516	1.111	.372
Residuals	1325184	27	49081		
Total	2363560	35			

Thus the hypothesis that the regression function is in $\mathbb{G}_0$ is acceptable.

Solutions to Fourth Practice Second Midterm Exam

20. (a) If X_1, X_7 and X_8 were independent, it would follow that $P(X_1 = 1, X_7 = 1, X_8 = 1) = P(X_1 = 1)P(X_7 = 1)P(X_8 = 1) = (1/2)(1/4)(1/4) = 1/32$, which is impossible since there are 16 design points. (Indeed, $P(X_1 = 1, X_7 = 1, X_8 = 1) = 1/16$.) Thus X_1, X_7 and X_8 are not independent.

(b) According to the summary statistics shown above, $\langle x_1, x_7x_8\rangle = \sum_i x_{i1}x_{i7}x_{i8} = 0$, so x_1 and x_7x_8 are orthogonal.

(c) Now $\langle x_6, x_7x_8\rangle = \sum_i x_{i6}x_{i7}x_{i8} = (4)(-9) + (4)(-3) + (0)(-1) + (4)(1) + (4)(3) + (0)(9) = -32$.

(d) The least squares estimates of the regression coefficients are given by $\widehat\beta_0 = \bar y = 27.5$; $\widehat\beta_1 = \frac{\sum_i x_{i1}y_i}{\sum_i x_{i1}^2} = \frac{64}{16} = 4$; $\widehat\beta_2 = \frac{\sum_i x_{i2}y_i}{\sum_i x_{i2}^2} = \frac{-32}{16} = -2$; $\widehat\beta_7 = \frac{\sum_i x_{i7}y_i}{\sum_i x_{i7}^2} = \frac{80}{8\cdot 9 + 8\cdot 1} = 1$; $\widehat\beta_8 = \frac{\sum_i x_{i8}y_i}{\sum_i x_{i8}^2} = \frac{-120}{80} = -1.5$; $\widehat\beta_{78} = \frac{\sum_i x_{i7}x_{i8}y_i}{\sum_i x_{i7}^2x_{i8}^2} = \frac{-288}{4\cdot 1 + 8\cdot 9 + 4\cdot 81} = \frac{-288}{400} = -\frac{18}{25} = -0.72$. Thus the least squares estimate of the regression function equals $27.5 + 4x_1 - 2x_2 + x_7 - 1.5x_8 - 0.72x_7x_8$.

(e) The fitted sum of squares is given by $\text{FSS} = \widehat\beta_1^2\sum_i x_{i1}^2 + \widehat\beta_2^2\sum_i x_{i2}^2 + \widehat\beta_7^2\sum_i x_{i7}^2 + \widehat\beta_8^2\sum_i x_{i8}^2 + \widehat\beta_{78}^2\sum_i x_{i7}^2x_{i8}^2 = 4^2(16) + (-2)^2(16) + 1^2(80) + (-1.5)^2(80) + (-0.72)^2(400) = 256 + 64 + 80 + 180 + 207.36 = 787.36$. Thus the residual sum of squares is given by $\text{RSS} = \text{TSS} - \text{FSS} = 1108 - 787.36 = 320.64$.

(f) The usual estimate of σ is given by $s = \sqrt{s^2} = \sqrt{\text{RSS}/(n-p)} = \sqrt{320.64/(16-6)} \doteq 5.663$. Thus $\text{SE}(\widehat\beta_2) = \frac{s}{\sqrt{\sum_i x_{i2}^2}} \doteq \frac{5.663}{4} \doteq 1.416$. Consequently, the t statistic for testing the hypothesis that $\beta_2 = 0$ is given by $t = \widehat\beta_2/\text{SE}(\widehat\beta_2) \doteq -2/1.416 \doteq -1.412$. From a table of the t distribution with 10 degrees of freedom, we conclude that $.1 = 2(1 - .95) < P\text{-value} < 2(1 - .9) = .2$.

(g) The residual sum of squares corresponding to the hypothesis that $\beta_2 = 0$ is given by $\mathrm{RSS}_0 = \mathrm{RSS} + \widehat{\beta}_2^2 \sum_i x_{i2}^2 = 320.64 + (-2)^2(16) = 384.64$. The F statistic corresponding to this hypothesis is given by $F = \frac{\mathrm{RSS}_0 - \mathrm{RSS}}{s^2} = \frac{64}{32.064} \doteq 1.996$. From a table of the F distribution with 1 degree of freedom in the numerator and 10 degrees of freedom in the denominator, P-value $> .05$, so the hypothesis is accepted by the test of size .05.

(h) The least squares estimate $\widehat{\tau}$ of the indicated change τ is given by $\widehat{\tau} = \widehat{\beta}_8(3) + \widehat{\beta}_{78}(-1)(3) - [\widehat{\beta}_8(-3) + \widehat{\beta}_{78}(-1)(-3)] = 6\widehat{\beta}_8 - 6\widehat{\beta}_{78} = 6[(-1.5) - (-0.72)] = -4.68$. Since $\widehat{\beta}_8$ and $\widehat{\beta}_{78}$ are independent (because all the basis functions are orthogonal to each other), $\mathrm{SE}(\widehat{\tau}) = 6s\sqrt{(\sum_i x_{i8}^2)^{-1} + (\sum_i x_{i7}^2 x_{i8}^2)^{-1}} \doteq 6(5.663)\sqrt{1/80 + 1/400} \doteq 4.161$. Thus the 95% confidence interval for τ is given by $\widehat{\tau} \pm t_{.975,10}\mathrm{SE}(\widehat{\tau}) \doteq (-4.68) \pm (2.228)(4.161) \doteq (-13.95, 4.59)$.

Solutions to Fifth Practice Second Midterm Exam

22. (a) Since $P(X_2 = 1, X_3 = 1, X_4 = 1) = 1/18$, whereas $P(X_2 = 1)P(X_3 = 1)P(X_4 = 1) = (1/3)^3 = 1/27$, the design random variables X_2, X_3 and X_4 are not independent.

(b) Since the indicated functions on $\mathbb{R}^4$ are distinct monomials, it follows from Theorem 9.1 on page 429 that they are linearly independent. Thus they form a basis of their span.

(c) Since $\langle 1, x_1^2 \rangle = 18$, the basis in (b) is not an orthogonal basis.

(d) Observe first that

$$\langle 1, x_2^2 - 2/3 \rangle = 18E(X_2^2 - 2/3) = 18[12/18 - 2/3] = 0$$

and

$$\langle x_2, x_2^2 - 2/3 \rangle = 18\, E[X_2(X_2^2 - 2/3)] = 18\, E\big(X_2^3\big) - 12E(X_2) = 0.$$

Since the design random variables X_1 and X_2 are independent,

$$\langle x_1, x_2^2 - 2/3 \rangle = 18\, E[X_1(X_2^2 - 2/3)] = 18\, E(X_1)E(X_2^2 - 2/3) = 0.$$

Since the design random variables X_2 and X_3 are independent,

$$\langle x_2^2 - 2/3, x_3 \rangle = 18\, E[(X_2^2 - 2/3)X_3] = 18\, E(X_1^2 - 2/3)E(X_3) = 0$$

and

$$\begin{aligned}\langle x_2^2 - 2/3, x_3^2 - 2/3 \rangle &= 18\, E[(X_2^2 - 2/3)(X_3^2 - 2/3)] \\ &= 18\, E(X_2^2 - 2/3)E(X_3^2 - 2/3) = 0.\end{aligned}$$

(e)

$$\widehat{\beta}_0 = \bar{y} = \frac{1}{18}\sum_i y_i = \frac{336}{18} = 18\tfrac{2}{3}$$
$$\widehat{\beta}_1 = \frac{\sum_i x_{i1}y_i}{\sum_i x_{i1}^2} = \frac{90}{18} = 5$$
$$\widehat{\beta}_2 = \frac{\sum_i x_{i2}y_i}{\sum_i x_{i2}^2} = \frac{-48}{12} = -4$$
$$\widehat{\beta}_3 = \frac{\sum_i (x_{i2}^2 - 2/3)y_i}{\sum_i (x_{i2}^2 - 2/3)^2} = \frac{-12}{4} = -3$$

and

$$\widehat{\beta}_8 = \frac{\sum_i x_{i1}x_{i2}y_i}{\sum_i x_{i1}^2 x_{i2}^2} = \frac{-36}{12} = -3$$

(f) The residual sum of squares is given by RSS = TSS − FSS = 1078 − 1022 = 56, so the squared multiple correlation coefficient is given by $R^2 = \dfrac{\text{FSS}}{\text{TSS}} = \dfrac{1022}{1078} \doteq .948$.

(g) The least squares fit in G_0 is given by

$$\widehat{\mu}_0(x_1,x_2,x_3,x_4) = \widehat{\beta}_0 + \widehat{\beta}_1 x_1 + \widehat{\beta}_2 x_2 + \widehat{\beta}_4 x_3 + \widehat{\beta}_6 x_4 = 18\tfrac{2}{3} + 5x_1 - 4x_2 + 3x_3 - 2x_4.$$

The corresponding fitted sum of squares is given by

$$\begin{aligned}\text{FSS} &= \widehat{\beta}_1^2\|x_1\|^2 + \widehat{\beta}_2^2\|x_2\|^2 + \widehat{\beta}_4^2\|x_3\|^2 + \widehat{\beta}_6^2\|x_4\|^2 \\ &= 5^2\cdot 18 + (-4)^2\cdot 12 + 3^2\cdot 12 + (-2)^2\cdot 12 = 798.\end{aligned}$$

(h) $s^2 = \dfrac{\text{RSS}}{n-p} = \dfrac{56}{7} = 8$, so $s = \sqrt{8} \doteq 2.828$.

(i) $\text{FSS} - \text{FSS}_0 = 1022 - 798 = 224$, so the F statistic is given by

$$F = \frac{(\text{FSS} - \text{FSS}_0)/(p - p_0)}{s^2} = \frac{224}{6\cdot 8} = 4\tfrac{2}{3} \doteq 4.667.$$

From the table of the F distribution with 6 degrees of freedom in the numerator and 7 degrees of freedom in the denominator, $.025 < P\text{-value} < .05$.

(j) $\widehat{\tau} = 2\widehat{\beta}_1 - 2\widehat{\beta}_8 - 2\widehat{\beta}_9 - 2\widehat{\beta}_{10} = 2\cdot 5 + 3\cdot 3 - 2\cdot 2 + 2\cdot 1 = 14$.

(k) Since the basis functions in (d) are orthogonal, $\widehat{\beta}_0, \dots, \widehat{\beta}_{10}$ are independent random variables. Consequently,

$$\begin{aligned}\text{var}(\widehat{\tau}) &= 4[\text{var}(\widehat{\beta}_1) + \text{var}(\widehat{\beta}_8) + \text{var}(\widehat{\beta}_9) + \text{var}(\widehat{\beta}_{10})] \\ &= 4\sigma^2\Big(\frac{1}{\|x_1\|^2} + \frac{1}{\|x_1x_2\|^2} + \frac{1}{\|x_1x_3\|^2} + \frac{1}{\|x_1x_4\|^2}\Big) = 4\Big(\frac{1}{18} + \frac{3}{12}\Big)\sigma^2 = \frac{11}{9}\,\sigma^2,\end{aligned}$$

so $\text{SE}(\widehat{\tau}) = \sqrt{\dfrac{11}{9}s^2} = \sqrt{\dfrac{88}{9}} \doteq 3.127$. From the table of the t distribution with 7 degrees of freedom, the 95% confidence interval for τ is given by

$$14 \pm (2.365)(3.127) \doteq 14 \pm 7.395 = (6.605,\ 21.395).$$

(l) The t statistic for testing $\mathrm{H_0}$: $\tau = 0$ versus $\mathrm{H_a}$: $\tau \neq 0$ is given by $t \doteq 14/3.127 \doteq 4.477$. From the table of the t distribution with 7 degrees of freedom, $.002 < 2(.001) < P\text{-value} < 2(.005) = .01$.

Solutions to First Practice Final Exam

25. (a) Consider the design points 1, 1.5 and 2, unit weights, and the orthogonal basis 1, $x-1.5$,$(x-1.5)^2-1/6$ of the identifiable and saturated space of quadratic polynomials on the interval $[1,2]$. The least squares approximation in this space to the function $h=\sqrt{x}$ is given by $g = b_0 + b_1(x-1.5) + b_2[(x-1.5)^2 - 1/6]$, where $b_0 = \frac{\langle\sqrt{x},1\rangle}{\|1\|^2} = \frac{1+\sqrt{1.5}+\sqrt{2}}{3} \doteq 1.212986$, $b_1 = \frac{\langle\sqrt{x},x-1.5\rangle}{\|x-1.5\|^2} = \frac{-0.5+(0.5)\sqrt{2}}{0.5} = \sqrt{2}-1 \doteq 0.414214$, and $b_2 = \frac{\langle\sqrt{x},(x-1.5)^2-1/6\rangle}{\|(x-1.5)^2-1/6\|^2} = 2-4\sqrt{1.5}+2\sqrt{2} \doteq -0.070552$. Alternatively, by the Lagrange interpolation formula, $g(x) = 2(x-2)(x-1.5) - 4\sqrt{1.5}(x-1)(x-2) + 2\sqrt{2}(x-1)(x-1.5)$.

(b) By trial and error, the maximum magnitude of the error of approximation, which occurs at $x \doteq 1.2$, is given by $\sqrt{1.2} - [b_0 + b_1(-0.3) + b_2(0.09-1/6)] \doteq 0.001314$ or, alternatively, by $\sqrt{1.2} - [2(-0.8)(-0.3) - 4\sqrt{1.5}(0.2)(-0.8) + 2\sqrt{2}(0.2)(-0.3)] \doteq 0.001314$.

26. (a) $\dim(G_0) = 2$ and $\dim(G) = 3$.

(b) The functions 1, $x_1 + x_2 - 3$ form a basis of G_0, which can be extended to the basis 1, x_1+x_2-3, x_1-x_2 of G.

(c) The design matrices corresponding to G_0 and G are given, respectively, by $\mathbf{X_0} = \begin{bmatrix} 1 & -1 \\ 1 & 0 \\ 1 & 0 \\ 1 & 1 \end{bmatrix}$ and $\mathbf{X} = \begin{bmatrix} 1 & -1 & 0 \\ 1 & 0 & -1 \\ 1 & 0 & 1 \\ 1 & 1 & 0 \end{bmatrix}$.

(d) Using the notation from the solution to Example 13.9, we get that $\boldsymbol{N} = \mathrm{diag}(50,25,25,50)$, $\bar{\boldsymbol{Y}} = [.4,.44,.72,.56]^T$, $\widehat{\boldsymbol{\beta}}_0 = \mathbf{0}$, $\widehat{\boldsymbol{\theta}}_0 = \mathbf{0}$, $\widehat{\boldsymbol{\pi}}_0 = [.5,.5,.5,.5]^T$, $\widehat{\boldsymbol{W}}_0 = \frac{1}{4}\mathrm{diag}(50,25,25,50)$,

$$\boldsymbol{X}_0^T\widehat{\boldsymbol{W}}_0\boldsymbol{X}_0 = \frac{1}{4}\begin{bmatrix} 1 & 1 & 1 & 1 \\ -1 & 0 & 0 & 1 \end{bmatrix}\begin{bmatrix} 50 & 0 & 0 & 0 \\ 0 & 25 & 0 & 0 \\ 0 & 0 & 25 & 0 \\ 0 & 0 & 0 & 50 \end{bmatrix}\begin{bmatrix} 1 & -1 \\ 1 & 0 \\ 1 & 0 \\ 1 & 1 \end{bmatrix}$$
$$= \begin{bmatrix} \frac{75}{2} & 0 \\ 0 & 25 \end{bmatrix},$$

$$\widehat{\boldsymbol{V}}_0 = (\boldsymbol{X}_0^T \widehat{\boldsymbol{W}}_0 \boldsymbol{X})^{-1} = \begin{bmatrix} \frac{2}{75} & 0 \\ 0 & \frac{1}{25} \end{bmatrix},$$

$$\begin{aligned}
\boldsymbol{Z}_0 &= \widehat{\boldsymbol{V}}_0 \boldsymbol{X}^T \boldsymbol{N}(\bar{\boldsymbol{Y}} - \widehat{\boldsymbol{\pi}}_0) \\
&= \begin{bmatrix} \frac{2}{75} & 0 \\ 0 & \frac{1}{25} \end{bmatrix} \begin{bmatrix} 1 & 1 & 1 & 1 \\ -1 & 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} 50 & 0 & 0 & 0 \\ 0 & 25 & 0 & 0 \\ 0 & 0 & 25 & 0 \\ 0 & 0 & 0 & 50 \end{bmatrix} \begin{bmatrix} -.10 \\ -.06 \\ .22 \\ .06 \end{bmatrix} \\
&= \begin{bmatrix} \frac{4}{75} \\ \frac{8}{25} \end{bmatrix},
\end{aligned}$$

and

$$\widehat{\boldsymbol{\beta}}_1 = \widehat{\boldsymbol{\beta}}_0 + \boldsymbol{Z}_0 = \begin{bmatrix} 0 \\ 0 \end{bmatrix} + \begin{bmatrix} \frac{4}{75} \\ \frac{8}{25} \end{bmatrix} = \begin{bmatrix} \frac{4}{75} \\ \frac{8}{25} \end{bmatrix}.$$

(e) The maximum likelihood estimate of $\boldsymbol{\theta}_0$ corresponding to the submodel is given by

$$\widehat{\boldsymbol{\theta}}_0 = \operatorname{logit}(\widehat{\boldsymbol{\pi}}_0) \doteq \begin{bmatrix} -0.2688 \\ 0.0544 \\ 0.0544 \\ 0.3772 \end{bmatrix}.$$

Since $\widehat{\boldsymbol{\theta}}_0 = \boldsymbol{X}_0 \widehat{\boldsymbol{\beta}}_0$, we conclude that

$$\begin{bmatrix} -0.2688 \\ 0.0544 \end{bmatrix} \doteq \begin{bmatrix} 1 & -1 \\ 1 & 0 \end{bmatrix} \widehat{\boldsymbol{\beta}}_0$$

and hence that

$$\widehat{\boldsymbol{\beta}}_0 \doteq \begin{bmatrix} 0 & 1 \\ -1 & 1 \end{bmatrix} \begin{bmatrix} -0.2688 \\ 0.0544 \end{bmatrix} = \begin{bmatrix} 0.0544 \\ 0.3232 \end{bmatrix}.$$

(f) The matrix $\widehat{\boldsymbol{W}}_0$ is given by

$$\begin{bmatrix} 50(.4332)(.5668) & 0 & 0 & 0 \\ 0 & 25(.5136)(.4864) & 0 & 0 \\ 0 & 0 & 25(.5136)(.4864) & 0 \\ 0 & 0 & 0 & 50(.5932)(.4068) \end{bmatrix}$$

$$\doteq \begin{bmatrix} 12.277 & 0 & 0 & 0 \\ 0 & 6.245 & 0 & 0 \\ 0 & 0 & 6.245 & 0 \\ 0 & 0 & 0 & 12.066 \end{bmatrix}.$$

Thus the estimated information matrix $\boldsymbol{X}_0^T \widehat{\boldsymbol{W}}_0 \boldsymbol{X}_0$ is given by

$$\begin{bmatrix} 1 & 1 & 1 & 1 \\ -1 & 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} 12.277 & 0 & 0 & 0 \\ 0 & 6.245 & 0 & 0 \\ 0 & 0 & 6.245 & 0 \\ 0 & 0 & 0 & 12.066 \end{bmatrix} \begin{bmatrix} 1 & -1 \\ 1 & 0 \\ 1 & 0 \\ 1 & 1 \end{bmatrix}$$

$= \begin{bmatrix} 36.833 & -0.211 \\ -0.211 & 24.343 \end{bmatrix}$, whose inverse $\widehat{\boldsymbol{V}}_0 = (\boldsymbol{X}_0^T \widehat{\boldsymbol{W}}_0 \boldsymbol{X}_0)^{-1}$ is given by

$\frac{1}{896.581}\begin{bmatrix} 24.343 & 0.211 \\ 0.211 & 36.833 \end{bmatrix} \doteq \begin{bmatrix} 0.027151 & 0.000235 \\ 0.000235 & 0.041862 \end{bmatrix}$. Consequently, the standard errors of the entries 0.0544 and 0.3232 of $\widehat{\boldsymbol{\beta}}_0$ are given, respectively, by $\sqrt{0.027151} \doteq 0.165$ and $\sqrt{0.041862} \doteq 0.205$.

(g) The likelihood ratio statistic is given by $D \doteq 2\Big(20\log\frac{.4336}{.4332}+30\log\frac{.5664}{.5668}+11\log\frac{.3728}{.5136}+14\log\frac{.6272}{.4864}+18\log\frac{.6528}{.5136}+7\log\frac{.3472}{.4864}+28\log\frac{.5936}{.5932}+22\log\frac{.4064}{.4068}\Big) \doteq 3.973$. The corresponding P-value is given by $2[1-\Phi(\sqrt{3.973})] \doteq 2[1-\Phi(1.993)] \doteq 2[1-.97685] = .0463$.

Solutions to Second Practice Final Exam

28. (a) Here $\dim(G_1) = 3$, $\dim(G_2) = 2$, and $\dim(G) = \dim(G_1)+\dim(G_2)-1 = 4$.

(b) All of these spaces are identifiable.

(c) None of these spaces are saturated.

(d) Since G_1 is a space of functions of x_1, we can ignore the factor x_2. Thus the experimental data can be shown as follows:

x_1	Time parameter	Number of events
-1	50	55
0	100	125
1	50	70

When the experimental data is viewed in this manner, the space $\mathbb{G}_1$ is saturated. The design matrix is now given by $\boldsymbol{X} = \begin{bmatrix} 1 & -1 & 1/3 \\ 1 & 0 & -2/3 \\ 1 & 1 & 1/3 \end{bmatrix}$.

Moreover, $\bar{\boldsymbol{Y}} = \begin{bmatrix} 1.10 \\ 1.25 \\ 1.40 \end{bmatrix}$, so $\hat{\boldsymbol{\theta}} = \log(\bar{\boldsymbol{Y}}) \doteq \begin{bmatrix} 0.0953 \\ 0.2231 \\ 0.3365 \end{bmatrix}$. Since $\boldsymbol{X}^T\boldsymbol{X} = \begin{bmatrix} 3 & 0 & 0 \\ 0 & 2 & 0 \\ 0 & 0 & 2/3 \end{bmatrix}$, we see that $\boldsymbol{X}^{-1} = \begin{bmatrix} 1/3 & 0 & 0 \\ 0 & 1/2 & 0 \\ 0 & 0 & 3/2 \end{bmatrix}\begin{bmatrix} 1 & 1 & 1 \\ -1 & 0 & 1 \\ 1/3 & -2/3 & 1/3 \end{bmatrix} = \begin{bmatrix} 1/3 & 1/3 & 1/3 \\ -1/2 & 0 & 1/2 \\ 1/2 & -1 & 1/2 \end{bmatrix}$. Alternatively, $\det(\boldsymbol{X}) = \frac{2}{3}+\frac{1}{3}+\frac{2}{3}+\frac{1}{3} = 2$, so $\boldsymbol{X}^{-1} = \frac{1}{2}\begin{bmatrix} 2/3 & 2/3 & 2/3 \\ -1 & 0 & 1 \\ 1 & -2 & 1 \end{bmatrix} = \begin{bmatrix} 1/3 & 1/3 & 1/3 \\ -1/2 & 0 & 1/2 \\ 1/2 & -1 & 1/2 \end{bmatrix}$. Thus the maximum likelihood estimate $\widehat{\boldsymbol{\beta}}$ of the vector $\boldsymbol{\beta}$ of regression coefficients is given by $\widehat{\boldsymbol{\beta}} = \boldsymbol{X}^{-1}\widehat{\boldsymbol{\theta}} \doteq \begin{bmatrix} 1/3 & 1/3 & 1/3 \\ -1/2 & 0 & 1/2 \\ 1/2 & -1 & 1/2 \end{bmatrix}\begin{bmatrix} 0.0953 \\ 0.2231 \\ 0.3365 \end{bmatrix} \doteq \begin{bmatrix} 0.2183 \\ 0.1206 \\ -0.0073 \end{bmatrix}$.

(e) The rate is given by $\lambda(x_1) = \exp\big(\beta_0 + \beta_1 x_1 + \beta_2(x_1^2 - 2/3)\big)$. Thus the parameter of interest is given by $\tau = \lambda(1) - \lambda(-1) = \exp(\beta_0 + \beta_1 + \beta_2/3) - \exp(\beta_0 - \beta_+ 1 + \beta_2/3)$ and its maximum likelihood estimate is given by $\hat{\tau} = \exp(\widehat{\beta}_0 + \widehat{\beta}_1 + \widehat{\beta}_2/3) - \exp(\widehat{\beta}_0 - \widehat{\beta}_1 + \widehat{\beta}_2/3) = 1.4 - 1.1 = 0.3$. The variance-covariance matrix of $\widehat{\boldsymbol{\beta}}$ is given by $\boldsymbol{V} = (\boldsymbol{X}^T\boldsymbol{W}\boldsymbol{X})^{-1} = \boldsymbol{X}^{-1}\boldsymbol{W}^{-1}(\boldsymbol{X}^{-1})^T$ and its maximum likelihood estimate is given by $\widehat{\boldsymbol{V}} = \boldsymbol{X}^{-1}\widehat{\boldsymbol{W}}^{-1}(\boldsymbol{X}^{-1})^T$. Here $\boldsymbol{W} = \begin{bmatrix} 50\sigma_1^2 & 0 & 0 \\ 0 & 100\sigma_2^2 & 0 \\ 0 & 0 & 50\sigma_3^2 \end{bmatrix} = \begin{bmatrix} 50\lambda_1 & 0 & 0 \\ 0 & 100\lambda_2 & 0 \\ 0 & 0 & 50\lambda_3 \end{bmatrix}$ and $\widehat{\boldsymbol{W}} = \begin{bmatrix} 50\widehat{\lambda}(-1) & 0 & 0 \\ 0 & 100\widehat{\lambda}(0) & 0 \\ 0 & 0 & 50\widehat{\lambda}(1) \end{bmatrix} = \begin{bmatrix} 55 & 0 & 0 \\ 0 & 125 & 0 \\ 0 & 0 & 70 \end{bmatrix}$. The gradient of $\widehat{\tau}$ with respect to $\widehat{\boldsymbol{\beta}}$ is given by $\frac{\partial\widehat{\tau}}{\partial\widehat{\beta}_0} = \exp(\widehat{\beta}_0 + \widehat{\beta}_1 + \widehat{\beta}_2/3) - \exp(\widehat{\beta}_0 - \widehat{\beta}_1 + \widehat{\beta}_2/3) = \widehat{\tau} = 0.3$, $\frac{\partial\widehat{\tau}}{\partial\widehat{\beta}_1} = \exp(\widehat{\beta}_0 + \widehat{\beta}_1 + \widehat{\beta}_2/3) + \exp(\widehat{\beta}_0 - \widehat{\beta}_1 + \widehat{\beta}_2/3) = \widehat{\tau} = 1.4 + 1.1 = 2.5$, and $\frac{\partial\widehat{\tau}}{\partial\widehat{\beta}_2} = \frac{1}{3}\exp(\widehat{\beta}_0 + \widehat{\beta}_1 + \widehat{\beta}_2/3) - \frac{1}{3}\exp(\widehat{\beta}_0 - \widehat{\beta}_1 + \widehat{\beta}_2/3) = \frac{\widehat{\tau}}{3} = 0.1$. Thus $(\boldsymbol{X}^{-1})^T\nabla\widehat{\tau} = \begin{bmatrix} 1/3 & -1/2 & 1/2 \\ 1/3 & 0 & -1 \\ 1/3 & 1/2 & 1/2 \end{bmatrix}\begin{bmatrix} 0.3 \\ 2.5 \\ 0.1 \end{bmatrix} = \begin{bmatrix} -1.1 \\ 0 \\ 1.4 \end{bmatrix}$. Consequently, the standard error of $\widehat{\tau}$ is the nonnegative square root of

$$
\begin{aligned}
&((\boldsymbol{X}^{-1})^T\nabla\widehat{\tau})^T\widehat{\boldsymbol{W}}^{-1}(\boldsymbol{X}^{-1})^T\nabla\widehat{\tau} \\
&= \begin{bmatrix} -1.1 & 0 & 1.4 \end{bmatrix}\begin{bmatrix} 1/55 & 0 & 0 \\ 0 & 1/125 & 0 \\ 0 & 0 & 1/70 \end{bmatrix}\begin{bmatrix} -1.1 \\ 0 \\ 1.4 \end{bmatrix} = \frac{1.1^2}{55} + \frac{1.4^2}{70} = 0.05,
\end{aligned}
$$

so $\mathrm{SE}(\widehat{\tau}) = \sqrt{0.05} \doteq 0.2237$. Therefore, the 95% confidence interval for τ is given by $\widehat{\tau} \pm 1.96\mathrm{SE}(\widehat{\tau}) \doteq 0.3 \pm (1.96)(0.2237) \doteq 0.3 \pm 0.438 = (-0.138, 0.738)$. Alternatively, since $\mathbb{G}_1$ is saturated, $\widehat{\tau} = \bar{Y}_3 - \bar{Y}_1 = 1.4 - 1.1 = 0.3$. Moreover $\mathrm{var}(\widehat{\tau}) = \mathrm{var}(\bar{Y}_3) - \mathrm{var}(\bar{Y}_1) = \frac{\lambda(-1)}{n_1} + \frac{\lambda(1)}{n_3}$, so $\mathrm{SE}(\widehat{\tau}) = \sqrt{\frac{\widehat{\lambda}(-1)}{n_1} + \frac{\widehat{\lambda}(1)}{n_3}} = \sqrt{\frac{1.1}{50} + \frac{1.4}{50}} = \sqrt{0.05} \doteq 0.2237$, and hence the 95% confidence interval for τ is given as above.

(f) The maximum likelihood equations for the regression coefficients are given (see (11) of Chapter 13) by $\sum_k n_k[\bar{Y}_k - \exp(\widehat{\theta}(\boldsymbol{x}_k)] = 0$, $\sum_k n_k x_{k1}[\bar{Y}_k - \exp(\widehat{\theta}(\boldsymbol{x}_k)] = 0$, $\sum_k n_k(x_{k1}^2 - 2/3)[\bar{Y}_k - \exp(\widehat{\theta}(\boldsymbol{x}_k)] = 0$, and $\sum_k n_k x_{k2}[\bar{Y}_k - \exp(\widehat{\theta}(\boldsymbol{x}_k)] = 0$. Here $\widehat{\theta}(\boldsymbol{x}_k) = \widehat{\beta}_0 + \widehat{\beta}_1 x_{k1} + \widehat{\beta}_2(x_{k1}^2 - 2/3) + \widehat{\beta}_3 x_{k2}$. Alternatively, according to Problem 13.35, the maximum likelihood equations can be written in matrix form as $\boldsymbol{X}^T\boldsymbol{N}\exp(\boldsymbol{X}\widehat{\boldsymbol{\beta}}) = \boldsymbol{X}^T\boldsymbol{N}\bar{\boldsymbol{Y}}$, where

$$\boldsymbol{X} = \begin{bmatrix} 1 & -1 & 1/3 & -1 \\ 1 & -1 & 1/3 & 1 \\ 1 & 0 & -2/3 & -1 \\ 1 & 0 & -2/3 & 1 \\ 1 & 1 & 1/3 & -1 \\ 1 & 1 & 1/3 & 1 \end{bmatrix},$$

$\boldsymbol{N} = \operatorname{diag}(25, 25, 50, 50, 25, 25)$, and $\bar{\boldsymbol{Y}} = [1.4, 0.8, 1.5, 1.0, 1.2, 1.6]^T$.

(g) Here $\widehat{\theta}(\boldsymbol{x}_1) = \widehat{\beta}_0 + \widehat{\beta}_1 x_{11} + \widehat{\beta}_2(x_{11}^2 - 2/3) + \widehat{\beta}_3 x_{12} = \widehat{\beta}_0 + \widehat{\beta}_1(-1) + \widehat{\beta}_2(1/3) + \widehat{\beta}_3(-1) \doteq 0.211 - 0.121 - 0.007/3 + 0.121 \doteq 0.209$ and so forth. Alternatively, $\widehat{\boldsymbol{\theta}} = \boldsymbol{X}\widehat{\boldsymbol{\beta}} = \begin{bmatrix} 1 & -1 & 1/3 & -1 \\ 1 & -1 & 1/3 & 1 \\ 1 & 0 & -2/3 & -1 \\ 1 & 0 & -2/3 & 1 \\ 1 & 1 & 1/3 & -1 \\ 1 & 1 & 1/3 & 1 \end{bmatrix} \doteq \begin{bmatrix} 0.211 \\ 0.121 \\ -0.007 \\ -0.121 \end{bmatrix} \doteq \begin{bmatrix} 0.209 \\ -0.033 \\ 0.337 \\ 0.095 \\ 0.451 \\ 0.209 \end{bmatrix}$, so $\widehat{\boldsymbol{\lambda}} = \exp(\widehat{\boldsymbol{\theta}}) \doteq \begin{bmatrix} 1.232 \\ 0.968 \\ 1.401 \\ 1.100 \\ 1.570 \\ 1.232 \end{bmatrix}$. Thus (see Problem 13.74) the deviance of $\mathbb{G}$ is given by $D(\mathbb{G}) \doteq 2\big(35\log\frac{1.4}{1.232} + 20\log\frac{0.8}{0.968} + 75\log\frac{1.5}{1.401} + 50\log\frac{1}{1.1} + 30\log\frac{1.2}{1.57} + 40\log\frac{1.6}{1.232}\big) \doteq 6.818$. Consequently, the P-value for the goodness-of-fit test of the model is given by P-value $\doteq 1 - \chi_2^2(6.818) \doteq .0331$. Thus there is evidence, but not strong evidence, that the model does not exactly fit the data.

(h) According to the solution to (d), the deviance of G_1 is given by $D(G_1) = 2\big(35\log\frac{1.4}{1.1} + 20\log\frac{0.8}{1.1} + 75\log\frac{1.5}{1.25} + 50\log\frac{1.0}{1.25} + 30\log\frac{1.2}{1.4} + 40\log\frac{1.6}{1.4}\big) \doteq 10.611$. According to the solution to (g), the likelihood ratio statistic for testing the hypothesis that the regression function is in G_1 under the assumption that it is in G is given by $D(G_1) - D(G) \doteq 10.611 - 6.818 = 3.793$. Under the hypothesis, this statistic has approximately the chi-square distribution with 1 degree of freedom, which is the distribution of the square of a standard normal random variable. Thus the P-value for the test is given by $2[1 - \Phi(\sqrt{3.793})] \doteq 2[1 - \Phi(1.95)] \doteq 2(.0256) = .0512$. Hence the test of size $\alpha = .05$ barely accepts the hypothesis.

29. (a) Let $0 < \alpha < 1$. Then, by the CLT, the nominal $100(1-\alpha)\%$ confidence interval $\bar{Y}_1 - \bar{Y}_2 \pm z_{(1-\alpha)/2}\mathrm{SE}(\bar{Y}_1 - \bar{Y}_2)$ for $\mu_1 - \mu_2$ has approximately the indicated coverage probability when n_1 and n_2 are large; here $\mathrm{SE}(\bar{Y}_1 - \bar{Y}_2) = \sqrt{S_1^2/n_1 + S_2^2/n_2}$.

(b) The P-value is given by P-value $= 2[1 - \Phi(|Z|)]$, where $Z = \frac{\bar{Y}_1 - \bar{Y}_2}{\mathrm{SE}(\bar{Y}_1 - \bar{Y}_2)}$.

(c) Here $\mathrm{SE}(\bar{Y}_1 - \bar{Y}_2) = \sqrt{\frac{20^2}{50} + \frac{10^2}{100}} = \sqrt{9} = 3$, so the 95% confidence interval is given by $18 - 12 \pm (1.96)3 = 6 \pm 5.88 = (0.12, 11.88)$.

(d) Here $Z = 6/3 = 2$, so P-value $= 2[1 - \Phi(2)] \doteq 2(1 - .9772) = 0.0456$.

(e) Write $\tau = h(\mu_1, \mu_2) = \mu_1/\mu_2$ and $\widehat{\tau} = h(\bar{Y}_1, \bar{Y}_2) = \bar{Y}_1/\bar{Y}_2$. Then $\mathrm{AV}(\widehat{\tau}) = \left(\frac{\partial h}{\partial \mu_1}\right)^2 \frac{\sigma_1^2}{n_1} + \left(\frac{\partial h}{\partial \mu_2}\right)^2 \frac{\sigma_2^2}{n_2} = \left(\frac{1}{\mu_2}\right)^2 \frac{\sigma_1^2}{n_1} + \left(-\frac{\mu_1}{\mu_2^2}\right)^2 \frac{\sigma_2^2}{n_2} = \frac{1}{\mu_2^2}\left(\frac{\sigma_1^2}{n_1} + \tau^2 \frac{\sigma_2^2}{n_2}\right)$, so $\mathrm{ASD}(\widehat{\tau}) = \frac{1}{|\mu_2|}\sqrt{\frac{\sigma_1^2}{n_1} + \tau^2 \frac{\sigma_2^2}{n_2}}$. Thus $\mathrm{SE}(\widehat{\tau}) = \frac{1}{\bar{Y}_2}\sqrt{\frac{S_1^2}{n_1} + \widehat{\tau}^2 \frac{S_2^2}{n_2}}$. The $100(1-\alpha)\%$ confidence interval for τ is given as usual by $\widehat{\tau} \pm z_{(1-\alpha)/2}\mathrm{SE}(\widehat{\tau})$.

(f) If $\mathrm{SD}(\bar{Y}_2) = \sigma_2/\sqrt{n_2}$ is not small relative to $|\mu_2|$, then $\bar{Y}_2$ and μ_2 could well have different signs. Thus the linear Taylor approximation that underlies the confidence interval in (e) is too inaccurate for the corresponding normal approximation to be approximately valid.

(g) The Z statistic for testing the hypothesis that $\tau = 1$ is given by $Z = (\widehat{\tau} - 1)/\mathrm{SE}(\widehat{\tau})$. The corresponding P-value is given by P-value $= 2[1 - \Phi(|Z|)]$.

(h) Here $\widehat{\tau} = \bar{Y}_1/\bar{Y}_2 = 18/12 = 1.5$, so $\mathrm{SE}(\widehat{\tau}) = \frac{1}{12}\sqrt{\frac{20^2}{50} + 1.5^2 \frac{10^2}{100}} = \frac{1}{12}\sqrt{10.25} \doteq 0.2668$. Thus the desired confidence interval is given by $1.5 \pm (1.96)(0.2668) \doteq 1.5 \pm 0.523 = (0.977, 2.023)$.

(i) Here $Z \doteq (1.5 - 1)/0.2668 \doteq 1.874$, so P-value $\doteq 2[1 - \Phi(1.874)] \doteq 2[1 - .9695] = .061$.

(j) The hypothesis that $\mu_1 - \mu_2 = 0$ is equivalent to the hypothesis that $\mu_1/\mu_2 = 1$. Thus it is not surprising that the P-value in (d) is reasonably close to the P-value in (i). That the P-values are not closer together is presumably due to the inaccuracy in the linear Taylor approximation that underlies the asymptotic normality that is a basis of the solution to (e). Since μ_1/μ_2 is not a function of $\mu_1 - \mu_2$, the confidence intervals in (c) and (h) are mainly incomparable. On the other hand, the confidence interval in (c) barely excludes 0, while the confidence interval in (h) barely includes 1. Thus, the two confidence intervals lead to the conclusion that the hypothesis that $\mu_1 - \mu_2 = 1$ or, equivalently, that $\mu_1/\mu_2 = 1$ is borderline for beinga accepted or rejected at the 5% level of significance. This conclusion is in agreement with the P-values found in (d) and (i), which are both close to .05.

Solutions to Third Practice Final Exam

31. (a) The design matrix corresponding to the basis $1, x, x^2$ (any other basis would be incompatible with the definition of the regression coefficients as the coefficients of these basis functions) is given by $\boldsymbol{X} = \begin{bmatrix} 1 & -1 & 1 \\ 1 & 0 & 0 \\ 1 & 1 & 1 \end{bmatrix}$. Since $\det(\boldsymbol{X}) = 1 + 1 = 2$, $\boldsymbol{X}^{-1} = \frac{1}{2}\begin{bmatrix} 0 & 2 & 0 \\ -1 & 0 & 1 \\ 1 & -2 & 1 \end{bmatrix} = \begin{bmatrix} 0 & 1 & 0 \\ -1/2 & 0 & 1/2 \\ 1/2 & -1 & 1/2 \end{bmatrix}$.

(b) Here $\bar{\boldsymbol{Y}} = \begin{bmatrix} 60/200 \\ 150/300 \\ 160/400 \end{bmatrix} = \begin{bmatrix} .3 \\ .5 \\ .4 \end{bmatrix}$, so $\widehat{\boldsymbol{\theta}} = \mathrm{logit}(\bar{\boldsymbol{Y}}) = \begin{bmatrix} \log(3/7) \\ 0 \\ \log(2/3) \end{bmatrix} \doteq$

$\begin{bmatrix} -0.8473 \\ 0 \\ -0.4055 \end{bmatrix}$. Thus $\begin{bmatrix} \widehat{\beta}_0 \\ \widehat{\beta}_1 \\ \widehat{\beta}_2 \end{bmatrix} = \boldsymbol{X}^{-1}\widehat{\boldsymbol{\theta}} \doteq \begin{bmatrix} 0 \\ (0.8473 - 0.4055)/2 \\ -(0.8473 + 0.4055)/2 \end{bmatrix} \doteq$ $\begin{bmatrix} 0 \\ 0.2209 \\ -0.6264 \end{bmatrix}$. Consequently, $\widehat{\beta}_0 = 0$, $\widehat{\beta}_1 \doteq 0.2209$, and $\widehat{\beta}_2 \doteq -0.6264$.

(c) The asymptotic variance-covariance matrix of $\widehat{\boldsymbol{\beta}}$ equals $(\boldsymbol{X}^T\boldsymbol{W}\boldsymbol{X})^{-1} = \boldsymbol{X}^{-1}\boldsymbol{W}^{-1}(\boldsymbol{X}^{-1})^T$. Thus the asymptotic variance of $\widehat{\beta}_1$ equals

$$\frac{1}{4}\Big(\frac{1}{200\sigma_1^2} + \frac{1}{400\sigma_3^2}\Big) = \frac{1}{4}\Big(\frac{1}{200\pi_1(1-\pi_1)} + \frac{1}{400\pi_3(1-\pi_3)}\Big).$$

Since the model is saturated, the maximum likelihood estimates of π_1 and π_3 are given by $\hat{\pi}_1 = \bar{Y}_1 = .3$ and $\hat{\pi}_3 = \bar{Y}_3 = .4$. Thus the standard error of $\widehat{\beta}_1$ is given by $\mathrm{SE}(\widehat{\beta}_1) = \sqrt{\frac{1}{4}\Big(\frac{1}{200(.3)(.7)} + \frac{1}{400(.4)(.6)}\Big)} \doteq \sqrt{0.00856} \doteq 0.0925$. Hence the 95% confidence interval for $\widehat{\beta}_1$ is given by

$$\widehat{\beta}_1 \pm 1.96\,\mathrm{SE}(\widehat{\beta}_1) \doteq 0.2209 \pm (1.96)(0.0925) \doteq 0.221 \pm 0.181 = (0.040, 0.402).$$

(d) The Wald statistic is given by $W = \frac{\widehat{\beta}_1}{\mathrm{SE}(\widehat{\beta}_1)} = \frac{0.2209}{0.0925} \doteq 2.388$. The corresponding P-value is given by P-value $= 2[1-\Phi(|W|)] \doteq 2[1-\Phi(2.388)] \doteq 2[1 - .9915] \doteq .017$. Thus the hypothesis that $\beta_1 = 0$ is rejected by the test of size .02 but accepted by the test of size .01.

(e) Under the submodel, $\mathrm{logit}(\pi_3) = \mathrm{logit}(\pi_1)$, so $Y_1 + Y_3$ has the binomial distribution with parameters $200 + 400 = 600$ and probability π with $\mathrm{logit}(\pi) = \beta_0 + \beta_2$. The observed value of $Y_1 + Y_3$ is 220, and the corresponding sample proportion is $220/600 = 11/30$. Thus we can write the observed data as

x	n	$\bar{Y}$
0	300	1/2
1	600	11/30

. When the data is viewed in this manner, the design matrix is given by $\boldsymbol{X}_0 = \begin{bmatrix} 1 & 0 \\ 1 & 1 \end{bmatrix}$, which is invertible. Thus the given model is saturated. Moreover, $\boldsymbol{X}_0^{-1} = \begin{bmatrix} 1 & 0 \\ -1 & 1 \end{bmatrix}$. Also, $\widehat{\boldsymbol{\theta}}_0 = \begin{bmatrix} \mathrm{logit}(1/2) \\ \mathrm{logit}(11/30) \end{bmatrix} = \begin{bmatrix} 0 \\ \log(11/19) \end{bmatrix} \doteq \begin{bmatrix} 0 \\ -0.5465 \end{bmatrix}$. Consequently, $\begin{bmatrix} \widehat{\beta}_0 \\ \widehat{\beta}_2 \end{bmatrix} = \boldsymbol{X}_0^{-1}\widehat{\boldsymbol{\theta}}_0 = \begin{bmatrix} 1 & 0 \\ -1 & 1 \end{bmatrix}\begin{bmatrix} 0 \\ -0.5465 \end{bmatrix} = \begin{bmatrix} 0 \\ -0.5465 \end{bmatrix}$.

(f) The likelihood ratio statistic is given by $D = 2\big(60\log\frac{.3}{11/30} + 140\log\frac{.7}{19/30} + 150\log\frac{.5}{.5} + 150\log\frac{.5}{.5} + 160\log\frac{.4}{11/30} + 240\log\frac{.6}{19/30}\big) \doteq 5.834$. The corresponding P-value is given by P-value $\doteq 1 - \chi_1^2(5.834) \doteq .0157$. Alternatively, it is given by P-value $\doteq 2[1 - \Phi(\sqrt{5.834})] \doteq .0157$. Thus the hypothesis that the submodel is valid is rejected by the test of size .02 but accepted by the test of size .01.

32. (a) The within sum of squares is given by WSS $= \sum_k \sum_i (Y_{ki} - \bar{Y}_k)^2 = \sum_k (n_k - 1)S_k^2 = 199(1.4)^2 + 299(1.6)^2 + 399(1.5)^2 = 2053.23$. The overall sample mean is given by $\bar{Y} = \frac{1}{n}\sum_k n_k \bar{Y}_k = \frac{200(0.38)+300(0.62)+400(0.47)}{900} = 0.5$. Thus the between sum of squares is given by BSS $= \sum_k n_k(\bar{Y}_k - \bar{Y})^2 = 200(0.38 - 0.5)^2 + 300(0.62 - 0.5)^2 + 400(0.47 - 0.5)^2 = 7.56$. Consequently, the total sum of squares is given by TSS = BSS + WSS = $7.56 + 2053.23 = 2060.79$.

(b) The design matrix is given by $\boldsymbol{X} = \begin{bmatrix} 1 & -1 & 1 \\ 1 & 0 & 0 \\ 1 & 1 & 1 \end{bmatrix}$, which coincides with the design matrix in Problem 31(a). Thus, according to the solution to that problem, $\boldsymbol{X}^{-1} = \begin{bmatrix} 0 & 1 & 0 \\ -1/2 & 0 & 1/2 \\ 1/2 & -1 & 1/2 \end{bmatrix}$. Since $\boldsymbol{X}$ is invertible, the model is unsaturated, so $\begin{bmatrix} \widehat{\beta}_0 \\ \widehat{\beta}_1 \\ \widehat{\beta}_2 \end{bmatrix} = \boldsymbol{X}^{-1} \begin{bmatrix} \bar{Y}_1 \\ \bar{Y}_2 \\ \bar{Y}_3 \end{bmatrix} = \begin{bmatrix} 0 & 1 & 0 \\ -1/2 & 0 & 1/2 \\ 1/2 & -1 & 1/2 \end{bmatrix} \begin{bmatrix} 0.38 \\ 0.62 \\ 0.47 \end{bmatrix} = \begin{bmatrix} 0.620 \\ 0.045 \\ -0.195 \end{bmatrix}$; that is, $\widehat{\beta}_0 = 0.620$, $\widehat{\beta}_1 = 0.045$, and $\widehat{\beta}_2 = -0.195$.

(c) Since the model is saturated, the fitted and residual sums of squares are given by FSS = BSS = 7.56 and RSS = WSS = 2053.23. The standard deviation σ is estimated by $S = \sqrt{\frac{\text{RSS}}{n-p}} = \sqrt{\frac{2053.23}{897}} \doteq 1.513$.

(d) The least squares estimate of τ is given by $\widehat{\tau} = \widehat{\beta}_0^2 + \widehat{\beta}_2^2 = 0.62^2 + (0.195)^2 = 0.4224$. Write $h(\beta_0, \beta_2) = \beta_0^2 + \beta_2^2$. Then $\frac{\partial h}{\partial \beta_0} = 2\beta_0$ and $\frac{\partial h}{\partial \beta_2} = 2\beta_2$. The variance-covariance matrix of $\widehat{\beta}_0$ and $\widehat{\beta}_2$ is given by $\sigma^2 \begin{bmatrix} 0 & 1 & 0 \\ 1/2 & -1 & 1/2 \end{bmatrix} \begin{bmatrix} 1/200 & 0 & 0 \\ 0 & 1/300 & 0 \\ 0 & 0 & 1/400 \end{bmatrix} \begin{bmatrix} 0 & 1/2 \\ 1 & -1 \\ 0 & 1/2 \end{bmatrix}$ $= \sigma^2 \begin{bmatrix} 1/300 & -1/300 \\ -1/300 & 1/192 \end{bmatrix}$. Consequently, SE($\widehat{\tau}$) is given by
$2(1.513)\sqrt{\begin{bmatrix} 0.620 & -0.195 \end{bmatrix} \begin{bmatrix} 1/300 & -1/300 \\ -1/300 & 1/192 \end{bmatrix} \begin{bmatrix} 0.620 \\ -0.195 \end{bmatrix}}$
$= 3.026\sqrt{\frac{0.620^2}{300} + \frac{0.195^2}{192} + \frac{2(0.620)(0.195)}{300}} \doteq 3.026\sqrt{0.002285} \doteq 0.1446$. Thus the desired confidence interval is given by $\widehat{\tau} \pm 1.96\,\text{SE}(\widehat{\tau}) \doteq 0.4224 \pm (1.96)(0.0.1446) \doteq 0.422 \pm 0.0.283 = (0.139, 0.705)$.

(e) Under the submodel, $\mu_1 = \mu_3$ for all values of β_0 and β_2, so we can merge the data for the first and third design points, getting the observed data

x	n	$\bar{Y}$
0	300	0.62
1	600	0.44

. [Note that $\big(200(0.38) + 400(0.47)\big)/600 = 0.44$.] The corresponding design matrix is given by $\boldsymbol{X}_0 = \begin{bmatrix} 1 & 0 \\ 1 & 1 \end{bmatrix}$, whose inverse is given by $\boldsymbol{X}_0^{-1} = \begin{bmatrix} 1 & 0 \\ -1 & 1 \end{bmatrix}$. Thus the submodel is saturated and the least squares estimates of the corresponding regression coefficients are given by $\begin{bmatrix} \widehat{\beta}_{00} \\ \widehat{\beta}_{10} \end{bmatrix} = \begin{bmatrix} 1 & 0 \\ -1 & 1 \end{bmatrix} \begin{bmatrix} 0.62 \\ 0.44 \end{bmatrix} = \begin{bmatrix} 0.62 \\ -0.18 \end{bmatrix}$; that is, $\widehat{\beta}_{00} = 0.62$ and $\widehat{\beta}_{10} = -0.18$.

(f) The fitted sum of squares for the submodel is given by the between sum of squares for the merged data, that is, by $\text{FSS}_0 = \text{BSS}_0 = 300(0.62 - .5)^2 + 600(0.44 - .5)^2 = 6.48$.

(g) The F statistic is given by $F = \frac{(\text{FSS}-\text{FSS}_0)/1}{\text{RSS}/897} \doteq \frac{7.56-6.48}{2053.23/897} \doteq .0.472$. The P-value is given by $P\text{-value} \doteq 1 - F_{1,897}(0.472) \doteq 1 - \chi_1^2(0.472) = 2[1 - \Phi(\sqrt{0.472})] \doteq 2[1 - \Phi(0.687)] \doteq .492$.

Solutions to Fourth Practice Final Exam

34. (a) $\boldsymbol{X} = \begin{bmatrix} g_1(\boldsymbol{x}_1') & \cdots & g_d(\boldsymbol{x}_1') \\ \vdots & & \vdots \\ g_1(\boldsymbol{x}_d') & \cdots & g_d(\boldsymbol{x}_d') \end{bmatrix}$.

(b) $\boldsymbol{W} = \begin{bmatrix} n_1 & \cdots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \cdots & n_d \end{bmatrix}$.

(c) $\bar{y} = (n_1\bar{y}_1 + \cdots + n_d\bar{y}_d)/n$.

(d) $\text{WSS} = (n_1 - 1)s_1^2 + \cdots + (n_d - 1)s_d^2$.

(e) $\bar{\boldsymbol{Y}}$ has the multivariate normal distribution with mean vector $\boldsymbol{\mu}$ and variance-covariance matrix $\sigma^2\boldsymbol{W}^{-1}$.

(f) $\boldsymbol{\mu} = \boldsymbol{X}\boldsymbol{\beta}$.

(g) $\widehat{\boldsymbol{\beta}} = \boldsymbol{X}^{-1}\bar{\boldsymbol{Y}}$.

(h) $\widehat{\boldsymbol{\beta}}$ has the multivariate normal distribution with mean vector $\boldsymbol{\beta}$ and variance-covariance matrix $\sigma^2\boldsymbol{X}^{-1}\boldsymbol{W}^{-1}(\boldsymbol{X}^{-1})^T = \sigma^2(\boldsymbol{X}^T\boldsymbol{W}\boldsymbol{X})^{-1}$.

(i) $\text{LSS} = 0$, $\text{FSS} = \sum_{k=1}^d n_k(\bar{y}_k - \bar{y})^2$, and $\text{RSS} = \text{WSS}$.

(j) $s^2 = \text{RSS}/(n - d) = \text{WSS}/(n - d)$.

(k) Now $\text{VC}(\widehat{\boldsymbol{\beta}}) = \sigma^2(\boldsymbol{X}^T\boldsymbol{W}\boldsymbol{X})^{-1} = \sigma^2\boldsymbol{X}^{-1}\boldsymbol{W}^{-1}(\boldsymbol{X}^T)^{-1} = \sigma^2\boldsymbol{V}^{-1}\boldsymbol{X}^T\boldsymbol{W}^{-1}\boldsymbol{X}\boldsymbol{V}^{-1}$. Observe that $\text{var}(\widehat{\boldsymbol{\beta}}_j)$ is the jth diagonal entry of $\text{VC}(\widehat{\boldsymbol{\beta}})$. This equals σ^2/v_j^2 times the jth diagonal entry of $\boldsymbol{X}^T\boldsymbol{W}^{-1}\boldsymbol{X}$, which diagonal entry is $\sum_k n_k^{-1}g_j^2(\boldsymbol{x}_k')$. Thus the desired result is valid.

(l) $\text{SE}(\widehat{\beta}_j) = \frac{s}{v_j}\left(\sum_{k=1}^d \frac{1}{n_k}g_j^2(\boldsymbol{x}_k')\right)^{1/2}$.

(m) Under the indicated conditions, $\boldsymbol{X}^T\boldsymbol{W}\boldsymbol{X} = r\boldsymbol{X}^T\boldsymbol{X} = r\boldsymbol{V}$, which is a diagonal matrix. Thus the indicated basis functions are orthogonal. Moreover, $\|g_j\|^2$ is the jth diagonal entry of $r\boldsymbol{V}$, which equals rv_j. Consequently, $\text{FSS}-\text{FSS}_0 = \left\|\sum_{j=d_0+1}^d \widehat{\beta}_j g_j\right\|^2 = \sum_{j=d_0+1}^d \widehat{\beta}_j^2\|g_j\|^2 = r\sum_{j=d_0+1}^d v_j\widehat{\beta}_j^2$.

(n) $F = \frac{(\mathrm{RSS}_0 - \mathrm{RSS})/(d-d_0)}{\mathrm{WSS}/[d(r-1)]} = \frac{dr(r-1)}{(d-d_0)\mathrm{WSS}} \sum_{j=d_0+1}^{d} v_j \widehat{\beta}_j^2$.

35. (a) $\widehat{\pi}_1 = 15/20 = .75$ and $\widehat{\theta}_1 = \log 3 \doteq 1.099$; $\widehat{\pi}_2 = 25/40 = .625$ and $\widehat{\theta}_2 = \log(5/3) \doteq 0.511$; $\widehat{\pi}_3 = 28/40 = .7$ and $\widehat{\theta}_3 = \log(7/3) \doteq 0.847$; $\widehat{\pi}_4 = 32/80 = .4$ and $\widehat{\theta}_4 = \log(2/3) \doteq -0.405$.

(b) According to the solution to Example 12.16(b), $\mathrm{SE}(\widehat{\theta}_4 - \widehat{\theta}_1) = \sqrt{\frac{1}{20(3/4)(1/4)} + \frac{1}{80(2/5)(3/5)}} \doteq 0.565$. Since $\widehat{\theta}_4 - \widehat{\theta}_1 = \log(2/3) - \log 3 = \log(2/9) \doteq -1.504$, the 95% confidence interval for $\theta_4 - \theta_1$ (which is the log of the odds ratio) is given by $-1.504 \pm (1.96)(0.565) \doteq -1.504 \pm 1.107 = (-2.611, -0.397)$. Thus a reasonable 95% confidence interval for the odds ratio is given by $(e^{-2.611}, e^{-0.397}) \doteq (0.0735, 0.672)$.

(c) The design matrix is given by $\boldsymbol{X} = \begin{bmatrix} 1 & -1 & -1 & 1 \\ 1 & -1 & 1 & -1 \\ 1 & 1 & -1 & -1 \\ 1 & 1 & 1 & 1 \end{bmatrix}$.

(d) Now $\boldsymbol{X}^T\boldsymbol{X} = 4\boldsymbol{I}$, where $\boldsymbol{I}$ is the 4×4 identity matrix. Thus $\boldsymbol{X}^{-1} = (1/4)\boldsymbol{X}^T$. Consequently, $\widehat{\boldsymbol{\beta}} = \boldsymbol{X}^{-1}\mathrm{logit}(\widehat{\pi}) = (1/4)\boldsymbol{X}^T\widehat{\boldsymbol{\theta}}$. Therefore, $\widehat{\beta}_0 = [\widehat{\theta}_1 + \widehat{\theta}_2 + \widehat{\theta}_3 + \widehat{\theta}_4]/4$; $\widehat{\beta}_1 = [\widehat{\theta}_3 + \widehat{\theta}_4 - (\widehat{\theta}_1 + \widehat{\theta}_2)]/4$; $\widehat{\beta}_2 = [(\widehat{\theta}_2 + \widehat{\theta}_4) - (\widehat{\theta}_1 + \widehat{\theta}_3)]/4$; and $\widehat{\beta}_3 = [(\widehat{\theta}_1 + \widehat{\theta}_4) - (\widehat{\theta}_2 + \widehat{\theta}_3)]/4$.

(e) Ignoring x_2 and x_3, combining the first two design points, and combining the last two design points, we get the data shown in the following table:

k	x_1	# of runs	# of successes
1, 2	−1	60	40
3, 4	1	120	60

. The design matrix is given by $\boldsymbol{X}_0 = \begin{bmatrix} 1 & -1 \\ 1 & 1 \end{bmatrix}$, whose inverse is given by $\boldsymbol{X}_0^{-1} = (1/2)\begin{bmatrix} 1 & 1 \\ -1 & 1 \end{bmatrix}$. Thus $\widehat{\beta}_{00} = (\widehat{\theta}_{01} + \widehat{\theta}_{02})/2 = (\log 2 + \log 1)/2 = (\log 2)/2 \doteq 0.347$ and $\widehat{\beta}_{01} = (\widehat{\theta}_{02} - \widehat{\theta}_{01})/2 = (\log 1 - \log 2)/2 = -(\log 2)/2 \doteq -0.347$.

(f) The likelihood ratio test would provide a reasonable test of the indicated null hypothesis. The test would be based on the deviance of the submodel, which is given by (46) in Chapter 13; here $\widehat{\pi}(\boldsymbol{x}_k)$ is replaced by $\exp(\widehat{\beta}_{00} + \widehat{\beta}_{01}x_{k1})/[1 + \exp(\widehat{\beta}_{00} + \widehat{\beta}_{01}x_{k1})]$. Under the null hypothesis, this deviance has approximately the chi-square distribution with 2 degrees of freedom.

Solutions to Fifth Practice Final Exam

37. (a) We estimate $\mu_2 - \mu_1$ by $\widehat{\mu}_2 - \widehat{\mu}_1 = \bar{Y}_2 - \bar{Y}_1$. The variance of this estimate is given by

$$\mathrm{var}(\widehat{\mu}_2 - \widehat{\mu}_1) = \sigma^2\Big(\frac{1}{3} + \frac{1}{6}\Big) = \frac{\sigma^2}{2}.$$

The pooled sample variance is given by

$$S^2 = \frac{2S_1^2 + 5S_2^2}{7}.$$

Thus

$$\mathrm{SE}(\widehat{\mu}_2 - \widehat{\mu}_1) = \frac{S}{\sqrt{2}} = \sqrt{\frac{2S_1^2 + 5S_2^2}{14}}.$$

Consequently, the t statistic is given by

$$t = \frac{\widehat{\mu}_2 - \widehat{\mu}_1}{\mathrm{SE}(\widehat{\mu}_2 - \widehat{\mu}_1)} = \frac{\bar{Y}_1 - \bar{Y}_1}{\sqrt{(2S_1^2 + 5S_2^2)/14}}.$$

Under H_0, t has the t distribution with $(3-1)+(6-1) = 7$ degrees of freedom.

(b) The design matrix is given by

$$\boldsymbol{X} = \begin{bmatrix} 1 & 0 \\ 1 & 0 \\ 1 & 0 \\ 1 & 1 \\ 1 & 1 \\ 1 & 1 \\ 1 & 1 \\ 1 & 1 \\ 1 & 1 \end{bmatrix}$$

(c) The Gram matrix is given by

$$\boldsymbol{X}^T\boldsymbol{X} = \begin{bmatrix} 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 \\ 0 & 0 & 0 & 1 & 1 & 1 & 1 & 1 & 1 \end{bmatrix} \begin{bmatrix} 1 & 0 \\ 1 & 0 \\ 1 & 0 \\ 1 & 1 \\ 1 & 1 \\ 1 & 1 \\ 1 & 1 \\ 1 & 1 \\ 1 & 1 \end{bmatrix} = \begin{bmatrix} 9 & 6 \\ 6 & 6 \end{bmatrix}.$$

(d)

$$(\boldsymbol{X}^T\boldsymbol{X})^{-1} = \frac{1}{18}\begin{bmatrix} 6 & -6 \\ -6 & 9 \end{bmatrix} = \begin{bmatrix} 1/3 & -1/3 \\ -1/3 & 1/2 \end{bmatrix}.$$

(e) Observe that

$$\boldsymbol{X}^T\boldsymbol{Y} = \begin{bmatrix} 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 \\ 0 & 0 & 0 & 1 & 1 & 1 & 1 & 1 & 1 \end{bmatrix} \begin{bmatrix} Y_1 \\ Y_2 \\ Y_3 \\ Y_4 \\ Y_5 \\ Y_6 \\ Y_7 \\ Y_8 \\ Y_9 \end{bmatrix} = \begin{bmatrix} 3\bar{Y}_1 + 6\bar{Y}_2 \\ 6\bar{Y}_2 \end{bmatrix}.$$

Thus

$$\begin{bmatrix} \widehat{\beta}_0 \\ \widehat{\beta}_1 \end{bmatrix} = (\boldsymbol{X}^T\boldsymbol{X})^{-1}\boldsymbol{X}^T\boldsymbol{Y} = \begin{bmatrix} 1/3 & -1/3 \\ -1/3 & 1/2 \end{bmatrix} \begin{bmatrix} 3\bar{Y}_1 + 6\bar{Y}_2 \\ 6\bar{Y}_2 \end{bmatrix} = \begin{bmatrix} \bar{Y}_1 \\ \bar{Y}_2 - \bar{Y}_1 \end{bmatrix},$$

so $\widehat{\beta}_0 = \bar{Y}_1$ and $\widehat{\beta}_1 = \bar{Y}_2 - \bar{Y}_1$.

(f)

$$\begin{aligned} \text{FSS} &= \sum_i [\widehat{\mu}(x_i) - \bar{Y}]^2 \\ &= \sum_i \Big[\bar{Y}_1 + x_i(\bar{Y}_2 - \bar{Y}_1) - \frac{\bar{Y}_1 + 2\bar{Y}_2}{3}\Big]^2 \\ &= \sum_i (\bar{Y}_2 - \bar{Y}_1)^2 (x_i - 2/3)^2 \\ &= (\bar{Y}_2 - \bar{Y}_1)^2 \Big[3\Big(-\frac{2}{3}\Big)^2 + 6\Big(\frac{1}{3}\Big)^2\Big] \\ &= 2(\bar{Y}_2 - \bar{Y}_1)^2. \end{aligned}$$

(g)

$$\widehat{\boldsymbol{V}} = S^2(\boldsymbol{X}^T\boldsymbol{X})^{-1} = \frac{\text{WSS}}{7}\begin{bmatrix} 1/3 & -1/3 \\ -1/3 & 1/2 \end{bmatrix} = \frac{2S_1^2 + 5S_2^2}{7}\begin{bmatrix} 1/3 & -1/3 \\ -1/3 & 1/2 \end{bmatrix}.$$

(h)

$$\text{SE}(\widehat{\beta}_1) = \sqrt{\frac{2S_1^2 + 5S_2^2}{14}}.$$

(i)

$$t = \frac{\widehat{\beta}_1}{\text{SE}(\widehat{\beta}_1)} = \frac{\bar{Y}_2 - \bar{Y}_1}{\sqrt{(2S_1^2 + 5S_2^2)/14}}.$$

Under H_0: t has the t distribution with $9 - 2 = 7$ degrees of freedom.

38. (a)

$$\boldsymbol{X} = \begin{bmatrix} 1 & -1 & -3 & 1 \\ 1 & -1 & 3 & -1 \\ 1 & 1 & -1 & -3 \\ 1 & 1 & 1 & 3 \end{bmatrix}.$$

(b) Since $\bar{x}_1 = \bar{x}_2 = \bar{x}_3 = 0$, $\sum_i x_{i1}x_{i2} = 3 - 3 - 1 + 1 = 0$, $\sum_i x_{i1}x_{i3} = -1 + 1 - 3 + 3 = 0$, and $\sum_i x_{i2}x_{i3} = -3 - 3 + 3 + 3 = 0$, the indicated basis functions are orthogonal.

(c)

$$\boldsymbol{X}^T\boldsymbol{X} = \boldsymbol{D} = \begin{bmatrix} 4 & 0 & 0 & 0 \\ 0 & 4 & 0 & 0 \\ 0 & 0 & 20 & 0 \\ 0 & 0 & 0 & 20 \end{bmatrix},$$

so

$$\boldsymbol{X}^{-1} = \boldsymbol{D}^{-1}\boldsymbol{X}^T = \frac{1}{20}\begin{bmatrix} 5 & 0 & 0 & 0 \\ 0 & 5 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix}\begin{bmatrix} 1 & 1 & 1 & 1 \\ -1 & -1 & 1 & 1 \\ -3 & 3 & -1 & 1 \\ 1 & -1 & -3 & 3 \end{bmatrix}$$

$$= \frac{1}{20}\begin{bmatrix} 5 & 5 & 5 & 5 \\ -5 & -5 & 5 & 5 \\ -3 & 3 & -1 & 1 \\ 1 & -1 & -3 & 3 \end{bmatrix}.$$

(d) Since $\boldsymbol{X}$ is an invertible (square) matrix, it follows from Theorem 9.15 that G is identifiable and saturated.

(e) Since

$$\begin{bmatrix} \widehat{\beta}_0 \\ \widehat{\beta}_1 \\ \widehat{\beta}_2 \\ \widehat{\beta}_3 \end{bmatrix} = \boldsymbol{X}^{-1}\begin{bmatrix} \log \bar{y}_1 \\ \log \bar{y}_2 \\ \log \bar{y}_3 \\ \log \bar{y}_4 \end{bmatrix} \doteq \frac{1}{20}\begin{bmatrix} 5 & 5 & 5 & 5 \\ -5 & -5 & 5 & 5 \\ -3 & 3 & -1 & 1 \\ 1 & -1 & -3 & 3 \end{bmatrix}\begin{bmatrix} -0.693 \\ -1.609 \\ -0.916 \\ -0.693 \end{bmatrix} = -\frac{1}{20}\begin{bmatrix} 19.555 \\ -3.465 \\ 2.525 \\ -1.585 \end{bmatrix},$$

we see that $\widehat{\beta}_0 \doteq -0.978$, $\widehat{\beta}_1 \doteq 0.173$, $\widehat{\beta}_2 \doteq -0.126$, and $\widehat{\beta}_3 \doteq 0.0793$.

(f) Since G is saturated, $\widehat{\theta}(\boldsymbol{x}_k) = \log \bar{y}_k$ for $k = 1, 2, 3, 4$, so $\widehat{\lambda}(\boldsymbol{x}_k) = \exp \widehat{\theta}(\boldsymbol{x}_k) = \bar{y}_k$ for $k = 1, 2, 3, 4$.

(g) The asymptotic variance-covariance matrix of $\widehat{\boldsymbol{\beta}}$ is given by $(\boldsymbol{X}^T\boldsymbol{W}\boldsymbol{X})^{-1} = \boldsymbol{X}^{-1}\boldsymbol{W}^{-1}(\boldsymbol{X}^{-1})^T$, which is estimated by $\boldsymbol{X}^{-1}\widehat{\boldsymbol{W}}^{-1}(\boldsymbol{X}^{-1})^T$. Since

$$\widehat{\boldsymbol{W}} = \begin{bmatrix} n_1\bar{y}_1 & 0 & 0 & 0 \\ 0 & n_2\bar{y}_2 & 0 & 0 \\ 0 & 0 & n_3\bar{y}_3 & 0 \\ 0 & 0 & 0 & n_4\bar{y}_4 \end{bmatrix} = \begin{bmatrix} 20 & 0 & 0 & 0 \\ 0 & 20 & 0 & 0 \\ 0 & 0 & 20 & 0 \\ 0 & 0 & 0 & 20 \end{bmatrix} = 20\boldsymbol{I}_4,$$

$$\boldsymbol{X}^{-1}\widehat{\boldsymbol{W}}^{-1}(\boldsymbol{X}^{-1})^T = \frac{1}{20}(\boldsymbol{X}^T\boldsymbol{X})^{-1} = \frac{1}{400}\begin{bmatrix} 5 & 0 & 0 & 0 \\ 0 & 5 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix}.$$

Consequently, $\mathrm{SE}(\widehat{\beta}_3) = 1/20 = 0.05$. Thus the 95% confidence interval for β_3 is given by $0.0793 \pm (1.96)(0.05) = 0.0793 \pm .098 = (-0.0187, 0.1773)$. Since this confidence interval includes 0, the P-value for the Wald test of H_0 is greater than .05; hence H_0 is accepted.

(h) With regard to G_0, we can ignore x_2 and x_3 and conclude that there are just two design points, $x_1 = -1$ and $x_1 = 1$, and that the design matrix is given by

$$\boldsymbol{X}_0 = \begin{bmatrix} 1 & -1 \\ 1 & 1 \end{bmatrix}.$$

Since this (square) matrix is invertible, G_0 is identifiable and saturated.

(i) Combining the first two design point and combining the last two design points, we get the observed data

k	x_1	n	$\bar{y}$
1	-1	140	2/7
2	1	90	4/9

Consequently, the likelihood ratio statistic for the test of H_0 versus H_a is given by

$$40\Big(\log\frac{1/2}{2/7}+\log\frac{1/5}{2/7}+\log\frac{2/5}{4/9}+\log\frac{1/2}{4/9}\Big)$$
$$=40\Big(\log\frac{7}{4}+\log\frac{7}{10}+\log\frac{9}{10}+\log\frac{9}{8}\Big)\doteq 8.615.$$

From the table of the chisquare distribution with $4-2=2$ degrees of freedom, we conclude that P-value $<.005$. Thus we reject H_0 and conclude that $\theta(\cdot)$ is not a linear function of x_1.