3. Let $x_1', \ldots, x_d'$ be the design points and let $g_1, \ldots, g_p$ be a basis of the linear space $G$. Then the corresponding design matrix $X$ is the $d \times p$ matrix given by

$$X = \begin{bmatrix}
g_1(x_1') & \cdots & g_p(x_1') \\
\vdots & & \vdots \\
g_1(x_d') & \cdots & g_p(x_d')
\end{bmatrix}.$$ 

Given $g = b_1g_1 + \cdots + b.pg_p \in G$, set $b = [b_1, \ldots, b_p]^T$. Then

$$\begin{bmatrix}
g(x_1') \\
\vdots \\
g(x_d')
\end{bmatrix} = Xb.$$ 

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

$$A = \begin{bmatrix} a_{11} & \cdots & a_{1p} \\
\vdots & & \vdots \\
a_{p1} & \cdots & a_{pp} \end{bmatrix}$$ 

be the corresponding coefficient matrix. Also, let $\tilde{X}$ be the design matrix for the alternative basis. Then $\tilde{X} = XA^T$. The Gram matrix $M$ corresponding to the original basis is given by $M = X^TWX$, where $W = \text{diag}(w_1, \ldots, w_d)$ is the weight matrix. Suppose that $G$ is identifiable. Given a function $h$ that is defined at least on the design set, set $h = [h(x_1'), \ldots, h(x_d')]^T$, 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 $b^* = [b_1^*, \ldots, b_p^*]^T$. Then $b^*$ satisfies the normal equation $X^TWXb^* = X^TWh$, whose unique solution is given by $b^* = (X^TWX)^{-1}X^TWh$. In the context of the linear experimental model, let $\overline{Y}_k$ denote the average response at the $k$th design point, which has mean $\mu(x_k')$, where $\mu(\cdot) = \beta_1g_1 + \cdots + \beta_pg_p$ is the regression function. Set $\hat{Y} = [\overline{Y}_1, \ldots, \overline{Y}_k]^T$ and $\beta = [\beta_1, \ldots, \beta_p]^T$. Then $E(\hat{Y}) = X\beta$.

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 $\tau = \mu(3/2) = \mu_1g_1(3/2) + \mu_2g_2(3/2) + \mu_3g_3(3/2) + \mu_4g_4(3/2) + \mu_5g_5(3/2)$. Here

$$g_1(3/2) = \frac{(3/2-2)(3/2-3)(3/2-4)(3/2-5)}{(1-2)(1-3)(1-4)(1-5)} = \frac{(-1/2)(-3/2)(-5/2)(-7/2)}{(-1)(-2)(-3)(-4)} = \frac{35}{128}.$$
Statistics

\[ g_2(3/2) = \frac{(3/2-1)(3/2-3)(3/2-4)(3/2-5)}{(2-1)(2-3)(2-4)(2-5)} = \frac{(1/2)(-3/2)(-5/2)(-7/2)}{128} = \frac{140}{128}, \]

\[ g_3(3/2) = \frac{(3/2-1)(3/2-2)(3/2-4)(3/2-5)}{(3-1)(3-2)(3-4)(3-5)} = \frac{(1/2)(-1/2)(-3/2)(-5/2)}{128} = \frac{-70}{128}, \]

\[ g_4(3/2) = \frac{(3/2-1)(3/2-2)(3/2-3)(3/2-5)}{(4-1)(4-2)(4-3)(4-5)} = \frac{(1/2)(-1/2)(-3/2)(-7/2)}{128} = \frac{28}{128}, \]

\[ g_5(3/2) = \frac{(3/2-1)(3/2-2)(3/2-3)(3/2-4)}{(5-1)(5-2)(5-3)(5-4)} = \frac{(1/2)(-1/2)(-3/2)(-5/2)}{128} = \frac{-5}{128}. \]

(b) \[ \hat{\tau} \equiv \left( \frac{45}{128} \right)(18.136) + \left( \frac{140}{128} \right)(17.58) - \left( \frac{70}{128} \right)(18.32) + \left( \frac{28}{128} \right)(18.06) - \left( \frac{5}{128} \right)(18.13) \equiv 17.411. \]

(c) Now \[ \left( \frac{45}{128} \right)^2 \cdot \frac{1}{11} + \left( \frac{140}{128} \right)^2 \cdot \frac{1}{10} + \left( \frac{70}{128} \right)^2 \cdot \frac{1}{10} + \left( \frac{28}{128} \right)^2 \cdot \frac{1}{10} + \left( \frac{5}{128} \right)^2 \cdot \frac{1}{10} \equiv 0.161, \]

so \[ SE(\hat{\tau}) \equiv 0.423 \sqrt{0.161} \equiv 0.170. \]

(d) Since \[ t_{975,46} \equiv 2.013, \] the 95% confidence interval for \( \tau \) is given by \( 17.411 \pm (2.013)(0.170) = 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 \( \hat{\tau}_1 = \hat{\tau} - \hat{Y}_1 \equiv 17.411 - 18.136 = -0.725. \) Now \( \left( \frac{93}{128} \right)^2 \cdot \frac{1}{11} + \left( \frac{140}{128} \right)^2 \cdot \frac{1}{10} + \left( \frac{70}{128} \right)^2 \cdot \frac{1}{10} + \left( \frac{28}{128} \right)^2 \cdot \frac{1}{10} \equiv 0.202, \)

so \( SE(\hat{\tau}_1) \equiv 0.423 \sqrt{0.202} \equiv 0.190. \)

The \( t \) statistic for testing the hypothesis that \( \tau = \mu_1 \) or, equivalently, that \( \tau_1 = 0 \) is given by \( t = \frac{\hat{\tau} - \tau_0}{SE(\hat{\tau})} = \frac{-0.556}{0.185} = 2.4. \) Thus \( P \)-value is given by \( P \)-value \( \equiv 2t_{46}(2.4) < 0.01, \) so the test of size \( \alpha = 0.01 \) rejects the hypothesis.

(f) Now \( \tau = \mu_1 = \ldots = \mu_5 \) if and only if \( \mu_1 = \ldots = \mu_5. \) 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 \( \hat{\tau} = \hat{Y}_2 - \hat{Y}_1 \equiv 17.580 - 18.136 = -0.556 \) and \( SE(\hat{\tau}) = S \sqrt{\frac{1}{11} + \frac{1}{10}} = 0.423 \sqrt{\frac{1}{11} + \frac{1}{10}} \equiv 0.185. \) Consequently, the \( t \) statistic is given by \( t = \frac{\hat{\tau} - \tau_0}{SE(\hat{\tau})} = \frac{-0.556}{0.185} = 2.4. \) Thus \( P \)-value \( \equiv 1 - t_{46}(2.4) < 0.01, \) so the test of size \( \alpha = 0.01 \) rejects the hypothesis.

(b) The \( F \) statistic is given by \( F = \frac{\frac{[11(18.136 - 18)^2 + 10(17.580 - 17)^2 + 10(18.320 - 19)^2]}{3}}{0.179} \equiv 15.254. \)

Thus \( P \)-value \( \equiv 1 - F_{3,46}(15.254) \equiv 0.01, \) so the hypothesis is rejected.

7. (a) \( P(x) = x^2 - (\frac{1}{3}x^2) - (\frac{1}{3}x^2) \equiv x^2 - \frac{70}{6} = x^2 - \frac{35}{3}. \)

(b) \( \hat{\beta}_0 = \frac{(\bar{y})}{\frac{(\bar{x})}{2}} = \frac{y_1 + y_2 + y_3 + y_4 + y_5 + y_6}{6} = \hat{\beta} \) and \( \hat{\beta}_1 = \frac{(x \bar{y})}{\frac{(\bar{x})}{2}} = \frac{x_1 \bar{y} + x_2 \bar{y} + x_3 \bar{y} + x_4 \bar{y} + x_5 \bar{y} + x_6 \bar{y}}{6} = \frac{-3y_1 - 3y_2 - y_3 + 3y_4 + 3y_5 + 3y_6}{70}. \)

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 \( \| P(x) \|^2 = \frac{2(40^2 + 8^2 + 32^2)}{9} = 1292. \) Consequently, \( \hat{\beta}_2 = \frac{(P(x) \bar{y})}{\| P(x) \|^2} = \frac{40y_1 - 8y_2 - 32y_3 - 32y_4 + 32y_5 + 32y_6}{1792} = \frac{5y_1 - 2y_2 - 4y_3 - 4y_4 - y_5 + 5y_6}{224}. \)

Solutions to Fourth Practice First Midterm Exam
9. (a) According to Table 7.3, \(\hat{\tau} = \frac{\bar{Y}_1 + \bar{Y}_2}{3} - \frac{\bar{Y}_1 + \bar{Y}_2}{2} = \frac{18.136 + 18.330 + 18.130}{3} - \frac{17.580 + 18.060}{2} = 18.199 - 17.820 = 0.379\). Also, \(\text{var}(\hat{\tau}) = \sigma^2 \left( \frac{1}{9} \cdot \frac{1}{11} + \frac{1}{9} \cdot \frac{1}{10} + \frac{1}{10} \cdot \frac{1}{11} \right) = 0.08232\sigma^2\). According to page 361 of the textbook, the pooled sample standard deviation is given by \(S = 0.413\). Thus the standard error of \(\hat{\tau}\) is given by \(\text{SE}(\hat{\tau}) = 0.413\sqrt{0.08232} = 0.1185\). According to the table on page 830 of the textbook, \(t_{.95,46} \approx t_{.95,45} \approx 1.679\). Thus the 95% lower confidence bound for \(\tau\) is given by \(0.379 - (1.679)(0.1185) \approx 0.379 - 0.199 = 0.180\).

(b) The \(t\) statistic for the test is given by \(t = 3.20\). According to the table on page 830, \(.001 < P\)-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 \(\tau\) was selected after examining the data. Thus, it cannot be concluded that the probability (suitably interpreted) that the 95% lower confidence bound for \(\tau\) would be less than \(\tau\) 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 \(\tau \leq 0\) would be less than \(.05\) would be \(.05\).

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

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