Final exam: Tuesday May 16, 8 a.m. – 11 a.m., 1 Pimentel.

Omitted from final. Chapters 6, 7, 22, 24, 25; Sections 3 and 4 of Chapter 28.

Summary of major topics:

1. Design of experiments, and sampling. Chapters 1, 2, 19. Randomized controlled experiments, observational studies, confounding factors, random samples, biases. Simple random samples.

2. Fundamentals. Chapters 3, 4, 5. Histograms, averages, SDs, the normal curve, percentiles.

3. Correlation and regression. Chapters 8, 9, 10, 11, 12. Correlation coefficient, what it measures, how to compute it, common misinterpretations: Ch 8, 9. Regression method for estimating the “average of a strip,” and the regression effect: Ch 10. The r.m.s. error of regression, its use as the “SD of a strip” and using the normal curve within a strip: Ch 11. Equation of regression line, interpretation of slope: Ch 12.

4. Probability: small samples. Chapters 13, 14, 15. A few draws from a box; charts; multiplying probabilities; “at least one;” “both cards show heart” versus “both cards show the same suit”: Ch 13, 14. Binomial formula for the chance of a certain number of successes in a fixed number of independent repeated trials: Ch 15.

5. Probability: large samples. Chapters 16, 17, 18, 20; additional SE formulae in Chs 23, 27. Law of averages – with more trials, the proportion of successes tends to get closer to the theoretical probability of success, but the chance of hitting exactly that probability gets smaller: Ch 16. In Chapters 17, 18, and 20, you know the contents of the box and are asked to calculate probabilities for how the sample may come out. Box models, expected values and standard errors for:
   • The sum of the draws: Ch 17.
   • The average of the draws: Ch 23.
   • For 0-1 boxes:
     • The count of 1’s among draws: Ch 17; same as SE for sum.
     • The percent of 1’s among the draws: Ch 20.
   • For the difference between two independent random quantities:
     • SE for the difference: Ch 27.
     Probability histograms and the normal curve; the Central Limit Theorem: Ch 18.
     Approximating the chance of exactly so many successes; the continuity correction: Ch 18.
     SE correction factor for SRS, relative accuracy of samples, square root law: Ch 20.

6. Confidence intervals: Using a random sample to estimate a population percent or population average. Chapters 21 (confidence intervals for percents), 23 (confidence intervals for averages). With large samples, the normal curve can be used for confidence intervals even if the data aren’t normal; the difference between the SD of the sample and the SE for the sample average; our SE formulae don’t work for other kinds of random samples such as cluster samples: Ch 23.

7. Using random samples to test hypotheses about populations. Chapters 26, 27, 28, 29. All involve null and alternative hypotheses, a test statistic (z, t, or $\chi^2$), and a $P$-value. One large random sample used to test hypotheses about the population average (or the population percent): One–sample z–test. Testing for a population average (not a percent!) using a small sample: could be z-test, t-test, or neither. The t-test involves SD+, degrees of freedom = sample size –1, and the t table. Ch 26. Testing for a population percent using a small sample with replacement: exact binomial test. Lecture 4/25. Large independent random samples used to test a hypothesis about the difference between two population averages (or population percents): Two–sample z–test, involves SE for the difference. This test can also be used when individuals are randomized into treatment and control groups. It does not work when you are comparing two observations from each sampled unit. Ch 27. Testing whether or not your data look like a random sample from a population of several categories of individuals. $\chi^2$–test, degrees of freedom = number of categories –1, and the $\chi^2$ table. Ch 28.

That’s it!