Review for Final Exam: Ch.1 – 12

Reminder: You will be provided the “Formulas and Tables” handout as before. In addition, you may bring one page (8.5” x 11”) with any notes you like (both sides). As always, your calculator can help you with many of the calculations. The exam will feel weighted towards confidence intervals and hypothesis testing because so many more chapters dealt with these concepts.

On the test, I’ll put the new material first (Ch. 9-12), then a clear break, then the older material.

Start by reviewing each of your exams. Then try reading each chapter review. I also list many review problems below. I do not expect you to do them all, but you might look at them and ask yourself if you know how to get started on each part of each question. If you do work some of the problems, be sure to do a variety. Sections 15-2 and 15-3 are also useful sections to skim!

Chapter 1 Introduction to Statistics
Study the key points listed below. Mostly vocabulary here…
- Types of data (nominal, ordinal, interval, ratio) (1-2)
- Continuous vs. Discrete (1-2)
- Recognizing non-random samples (self-selecting/volunteer, loaded questions, etc) (1-3, 1-4)
- Observational study vs. experiment (1-4)
Chapter Review Exercises # 1, 3, 4(skip c), 7, 8

Chapter 2 Summarizing and Graphing Data
Here are our descriptive statistics methods. How do we organize and describe a collection of data? Being familiar with your calculator will help with many of the calculations. Study the calculator instructions (usually at the end of each section).
- Frequency distributions, relative freq. distributions and cumulative freq. distributions (2-2)
- You may be asked to construct a histogram (doing it by hand may be the easiest!), and you may be asked about class size, limits, etc. (2-3)
- Out of all the graph types, focus on scatterplots and Stem-and-leaf plots (2-4)
Chapter Review Exercises # 4, 5.
Cumulative Review Exercises # 1 – 3.

Chapter 3 Describing, Exploring and Comparing Data
- Mean of a distribution. Is the distribution skewed or symmetric? (3-2)
- Variation and standard deviation of a distribution (STATS → CALC → 1-VarStats) (3-3)
- Range rule of thumb, Empirical Rule, Chebyshev’s Theorem [all ways to give ballpark estimates about a distribution] (3-3)
- z-scores, quartiles (skip percentiles) (3-4)
- Boxplots – construct a boxplot, or compare two boxplots (3-5)
Chapter Review Exercises: # 1 – 8
Cumulative Review Exercises: # 1
Chapter 4 Probability

Do study and understand the formulas here, but often it boils down to counting: the probability of an event is:

\[ P(A) = \frac{\# \text{ ways the event can occur}}{\text{total # of possible outcomes}}. \]

- Complementary events (4-2, 4-3, 4-5)
- Addition Rule – be able to use this with a table of data. (4-3)
- Multiplication Rule – be able to interpret the rule in the context of a table. We often don’t actually have to multiply! (4-4, 4-5)
- Conditional probabilities – be able to compute these from a table (see example on p.178-9). (4-5)
- Combinations and Permutations will not be on the final. (4-7)

Chapter Review Exercises # 1 – 10
Cumulative Review Exercises # 1, 2

Chapter 5 Probability Distributions

As you review 5-2, think about all the distributions you have now seen: Binomial (our special discrete distribution) and then uniform, normal, t-distribution, \( \chi^2 \)-distribution, F-distribution (some of the many continuous distributions).

- Review the basic requirements to be a probability distribution. Find an expected value (generally the same as the mean!). Notice also, that here is where much of the discussion of recognizing an unusual event (low probability or the range rule of thumb) begins. (5-2)
- Be familiar with your calculator, either DISTR \rightarrow \text{binompdf} / \text{binomcdf} OR the PRGM BINOML83. (Alternatively, there is the formula, but this is limited.) Know the requirements for identifying binomial problems (that is, recognize a problem as solvable using the binomial distribution – main key fact: there can only be two outcomes). Find the mean and SD for a binomial distribution and use in conjunction with the Range Rule of Thumb. (5-3, 5-4)

Chapter Review Exercises # 1 – 3
Cumulative Review Exercises # 1 – 3

Chapter 6 Normal Probability Distributions

6-1 is a good reminder of the transition between discrete distributions and continuous distributions. We see that in a continuous distribution, probabilities correspond with areas under a curve. We first look at uniform distributions, then the rest of the chapter focuses on the normal distribution. We look at the standard normal distribution, which has mean = 0 and SD = 1, which gives us z-scores, and then proceed to more general variants, which could have any mean and any standard deviation. Our calculators make switching from standard normal to any normal distribution relatively easy!

- Be able to find the area under a uniform distribution. Be able to find a z-score for a data value. Find the corresponding probability for a z-score (NORMAL83 or normalcdf). Also, given a probability (area) find the corresponding z-score or data value (INVNOR83 or invNorm). (6-2, 6-3)
- Next we tackle sampling distributions. This is what allows us to start making inferences about a population given only a sample of size \( n \). We consider the collection of all possible samples of size \( n \). Each of these samples has its own mean (or proportion). This collection of means (or proportions) is itself a distribution that can be analyzed. As it turns out the mean of our sampling distribution (collection of means or proportions) is equal to the population mean or proportion! (6-4)
- As our sample size increases, we find that our sampling distribution becomes normal in shape [that is, clustered around the mean], (know the specifics here p.293). That’s the Central Limit
Theorem. Once we know we can use the normal distribution, we just need a mean and a standard deviation. We saw earlier that \( \mu_x = \mu \). We must modify the SD: \( \sigma_x = \frac{\sigma}{\sqrt{n}} \). (6-5)

- We saw that under certain conditions, we can use the continuous normal distribution to approximate the discrete binomial distribution. Know the conditions for this (check \( np \) and \( nq \)) and recognize the key word “Estimate” that signals to do this approximation. Be sure to make the correct “continuity correction.” (6-6)

Chapter Review Exercises # 1 – 5
Cumulative Review Exercises # 1, 2

Chapter 7 Estimates and Sample Sizes
I’ll refer you to the previous Exam review sheet for this chapter.
Cumulative Review Exercises: # 1 – 3

Chapter 8 Hypothesis Testing
Again, see the previous Exam review sheet. 8-2 will make more sense in hindsight. See again my handout for how to show your work for any hypothesis test.
Cumulative Review Exercises # 1, 3

Chapter 9 Inferences from Two Samples
As the chapter title infers, we now work with two samples, either comparing proportions, means or variances/standard deviations. We’ll again use the normal z-distribution for proportions and z- or t-distributions for means (depending on if we know the pop. SD or not – though also watch out for independent vs. matched pairs – see below). We have a new distribution, the F-distribution, for comparing variation in two samples – there is a built in distribution and test. The program FINV83 may be useful for finding critical values. The calculator instructions at the end of each section are detailed – read them carefully.

- Be able to find a confidence interval about the difference between two proportions. [2-PropZInt]
- Be able to test a hypothesis about comparing two proportions. [2-PropZTest] (9-2)
- Be able to recognize when you have two independent samples (no connection between the samples). (9-3)
- Be able to do confidence intervals and hypothesis tests about two means. Know when to use the normal z-distribution and when to use the t-distribution. The calculator generally does all the work! [2-SampZInt, 2-SampTInt, 2-SampZTest, 2-SampTTest] (9-3)
- Be able to work with matched pairs. Be sure to see the cautionary notes in the calculator instructions! We must do a test on the set of differences between each pair, so in fact we are testing about a single collection of numbers (the differences) rather than two independent samples. Use the correct notation. [T-Interval, T-Test] (9-4)
- When comparing variation in two samples, we use a new distribution – the F-distribution! The principles are the same. [2-SampFTest] (9-5)

Chapter Review Exercises # 1 – 8
Cumulative Review Exercises # 1 – 3

Chapter 10 Correlation and Regression (10.1 – 10.3 only)

- Understand what it means to have a correlation between two variables. Be able to interpret a scatterplot to recognize a strong positive or negative correlation. Be able to calculate the linear correlation coefficient, \( r \), using your calculator. You’ll have Table A-6 in the packet of notes I
give you, so you’ll be able to interpret \( r \). Skim the subsection “Formal Hypothesis Test”, but it will not be tested. Instead, skip to the end of the section where it describes using the calculator to find \( r \). See also my Instructor Video/Notes for another method for this

\[ \text{[STAT} \rightarrow \text{CALC} \rightarrow \text{LinReg]} \] (10-2)

- Use your calculator to find the **equation of the regression line** from given data, and use your equation to make **appropriate predictions** (or use \( \bar{y} \)). I will not ask about residuals and the least-squares property.

\[ \text{[STAT} \rightarrow \text{CALC} \rightarrow \text{LinReg]} \] (10-3)

Chapter Review Exercises # 1, 2.
Cumulative Review Exercises # 1, 2, 4.

**Chapter 11 Multinomial Experiments and Contingency Tables**

All three tests here (Goodness-of-Fit, Independence, and Homogeneity) use the \( \chi^2 \)-distribution and are right-tailed. The latter two tests are very similar – it’s mostly about where the data is coming from. Again, review the 5 essential parts of a hypothesis test summarized in the Chapter 7 handout.

- Multinomial experiments have more than 2 outcomes. A **goodness-of-fit test** measures how well our observed frequencies of each outcome match up with a set of given expected values. For example, rolling a die 24 times, we’d expect the numbers 1 – 6 to each come up 1/6 of the time. That is we’d expect to see 4 ones, 4 twos, etc. (In some problems, we might have to do some minor calculations to get the expected frequencies.) We might have a situation where all the expected frequencies are equal (as with the die above), or we might be given specific percentages (as in the M&Ms problem #15). \[ \text{[PRGM} \rightarrow \text{GOODFT83 or } \chi^2 \text{GOF-Test]} \] (11-2)

- Know what a **contingency table** is (just a table of data, really, with at least two rows and two columns). (11-3)

- Be able to enter a matrix (table) into your calculator \[ \text{[MATRIX} \rightarrow \text{Edit]} \] and then do a \( \chi^2 \) TEST on it (p.635).

- Understand that a **test for independence** is looking at relationships within a single population. \( H_0 \) is always that the rows and columns of the table are independent, and \( H_1 \) is always that the rows and columns are dependent. (11-3)

- A **test for homogeneity** is identical to that for independence, but it is looking at two populations. \( H_0 \) is always that the two populations have the same proportions, and \( H_1 \) is always that the two populations have different proportions. (11-3)

Chapter Review Exercises: # 1 – 4

The Chapter 11 Cumulative Review (#1-6) is a particularly good set of problems to try.

**Chapter 12 Analysis of Variance (ANOVA)**

We cover only **One-Way ANOVA**, which uses the F-distribution again. Read the Overview (12-1) carefully – it reminds us that the calculations here are so complex that we focus on using the calculator. ANOVA is only for collections of 3 or more data sets. We are comparing means by looking at variances. Once again, be able to present the 5 essential steps of a hypothesis test.

- Be able to find the important numbers in a calculation display. Mostly, we look for the p-value! (12-2)

- Be able to enter data into various lists, and then perform the built-in ANOVA test and interpret the results. The F-distribution is used here. \[ \text{[STAT} \rightarrow \text{TESTS} \rightarrow \text{ANOVA]} \] (12-2)

Chapter Review Exercises # 1, 2