Midterm 3 Suggested Review Problems

Here are problems that are similar to the ones you might see on the exam. Be sure to also review old quiz and workshop questions too.

Confidence Intervals

1. (Exercise 6.45 from OpenIntro Statistics) We are interested in estimating the proportion of graduates at a mid-sized university who found a job within one year of completing their undergraduate degree. Suppose we conduct a survey and find out that 348 of the 400 randomly sampled graduates found jobs. The graduating class under consideration included over 4500 students.
   (a) Describe the population parameter of interest. What is the value of the point estimate of this parameter?
   (b) Check if the conditions for constructing a confidence interval based on these data are met.
   (c) Calculate a 95% confidence interval for the proportion of graduates who found a job within one year of completing their undergraduate degree at this university, and interpret it in the context of the data.
   (d) What does "95% confidence" mean?
   (e) Now calculate a 99% confidence interval for the same parameter and interpret it in the context of the data.
   (f) Compare the widths of the 95% and 99% confidence intervals. Which one is wider? Explain.

Hypothesis Testing

Concepts to review: (i) The definition of a p-value (memorize this): *A **p-value** is the probability of getting a result at least as extreme as what happened, if the null hypothesisis true.* (ii) Make sure you understand when to reject a null hypothesis and what statistically signficant means.

2. (Exercise 5.25 from OpenIntro Statistics) A patient named Diana was diagnosed with Fibromyalgia, a long-term syndrome of body pain, and was prescribed anti-depressants. Being the skeptic that she is, Diana didn't initially believe that anti-depressants would help her symptoms. However after a couple months of being on the medication she decides that the anti-depressants are working, because she feels like her symptoms are in fact getting better.
   (a) Write the hypotheses in words for Diana's skeptical position when she started taking the anti-depressants.
   (b) What is a Type 1 Error in this context?
   (c) What is a Type 2 Error in this context?

3. (Exercise 6.49 from OpenIntro Statistics) A survey of 2,254 American adults indicates that 17% of cell phone owners browse the internet exclusively on their phone rather than a computer or other device.
   (a) According to an online article, a report from a mobile research company indicates that 38 percent of Chinese mobile web users only access the internet through their cell phones. Conduct a hypothesis test to determine if these data provide strong evidence that the proportion of Americans who only use their cell phones to access the internet is different than the Chinese proportion of 38%.
   (b) Interpret the p-value in this context.
   (c) Calculate a 95% confidence interval for the proportion of Americans who access the internet on their cell phones, and interpret the interval in this context.

Differences in Two Proportions

Concepts to review: (i) Meaning of two-sample confidence intervals. (ii) Pooled proportions.

4. (Exercise 6.18 from OpenIntro Statistics) The Stanford University Heart Transplant Study was conducted to determine whether an experimental heart transplant program increased lifespan. Each patient entering the program was officially designated a heart transplant candidate, meaning that he was gravely ill and might benefit from a new heart. Patients were randomly assigned into treatment and control groups. Patients in the treatment group received a transplant, and those in the control group did not. The table below displays how many patients survived and died in each group.

	control	treatment
alive	4	24
dead	30	45

   Suppose we are interested in estimating the difference in survival rate between the control and treatment groups using a confidence interval. Explain why we cannot construct such an interval using the normal approximation. What might go wrong if we constructed the confidence interval despite this problem?

5. (Exercise 6.22 from OpenIntro Statistics) According to a report on sleep deprivation by the Centers for Disease Control and Prevention, the proportion of California residents who reported insufficient rest or sleep during each of the preceding 30 days is 8.0%, while this proportion is 8.8% for Oregon residents. These data are based on simple random samples of 11,545 California and 4,691 Oregon residents. Calculate a 95% confidence interval for the difference between the proportions of Californians and Oregonians who are sleep deprived and interpret it in context of the data.

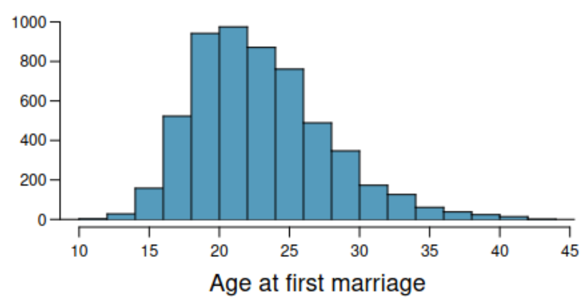
6. (Exercise 6.24 from OpenIntro Statistics) Using the same data as the previous problem:
   (a) Conduct a hypothesis test to determine if these data provide strong evidence the rate of sleep deprivation is different for the two states. (Reminder: Check conditions)
   (b) It is possible the conclusion of the test in part (a) is incorrect. If this is the case, what type of error was made (Type I error or type II)?

Inference About Means

Concepts to review: (i) Degrees of freedom and how to use the t-distribution table.

7. (Exercise 7.12 from OpenIntro Statistics) Researchers interested in lead exposure due to car exhaust sampled the blood of 52 police officers subjected to constant inhalation of automobile exhaust fumes while working traffic enforcement in a primarily urban environment. The blood samples of these officers had an average lead concentration of 124.32 μg/L and a SD of 37.74 μg/L; a previous study of individuals from a nearby suburb, with no history of exposure, found an average blood level concentration of 35 μg/L.
   (a) Write down the hypotheses that would be appropriate for testing if the police officers appear to have been exposed to a different concentration of lead.
   (b) Explicitly state and check all conditions necessary for inference on these data.
   (c) Regardless of your answers in part (b), test the hypothesis that the downtown police officers have a higher lead exposure than the group in the previous study. Interpret your results in context.

8. (Exercise 7.56 from OpenIntro Statistics) The National Survey of Family Growth conducted by the Centers for Disease Control gathers information on family life, marriage and divorce, pregnancy, infertility, use of contraception, and men's and women's health. One of the variables collected on this survey is the age at first marriage. The histogram below shows the distribution of ages at first marriage of 5,534 randomly sampled women between 2006 and 2010. The average age at first marriage among these women is 23.44 with a standard deviation of 4.72.



   Estimate the average age at first marriage of women using a 95% confidence interval, and interpret this interval in context. Discuss any relevant assumptions.

9. (Exercise 7.58 from OpenIntro Statistics) The data in the previous problem showed that the average age of women at first marriage is 23.44. Suppose a social scientist thinks this value has changed since the survey was taken. Below is how she set up her hypotheses. Indicate any errors you see.

$$H_0 : \bar{x} \neq 23.44 \text{ years old}$$

$$H_A : \bar{x} = 23.44 \text{ years old}$$

Matched Pairs Data

Know when to use one sample method for the differences in match pairs data instead of two-sample methods.

10. (Exercise 7.18 from OpenIntro Statistics) In each of the following scenarios, determine if the data are paired.
    (a) We would like to know if Intel's stock and Southwest Airlines' stock have similar rates of return. To find out, we take a random sample of 50 days, and record Intel's and Southwest's stock on those same days.
    (b) We randomly sample 50 items from Target stores and note the price for each. Then we visit Walmart and collect the price for each of those same 50 items.
    (c) A school board would like to determine whether there is a difference in average SAT scores for students at one high school versus another high school in the district. To check, they take a simple random sample of 100 students from each high school.

11. (Exercise 7.23 from OpenIntro Statistics) In the early 1990's, researchers in the UK collected data on traffic flow, number of shoppers, and traffic accident related emergency room admissions on Friday the 13th and the previous Friday, Friday the 6th. The table below summarizes the number of cars passing by a specific intersection on Friday the 6th and Friday the 13th for many such date pairs.

	6th	13th	Diff.
$\bar{x}$	128,385	126,550	1,835
s	7,259	7,664	1,176
n	10	10	10

    (a) What is the sample size here? Who or what are the individuals in the sample?
    (b) What are the hypotheses for evaluating whether the number of people out on Friday the 6th is different than the number out on Friday the 13th?
    (c) Check conditions to carry out the hypothesis test from part (b).
    (d) Calculate the test statistic and the p-value.
    (e) What is the conclusion of the hypothesis test?
    (f) Interpret the p-value in this context.
    (g) What type of error might have been made in the conclusion of your test? Explain.

Two Sample t-Distribution Methods

12. (Exercise 7.24 from OpenIntro Statistics) Prices of diamonds are determined by what is known as the 4 Cs: cut, clarity, color, and carat weight. The prices of diamonds go up as the carat weight increases, but the increase is not smooth. For example, the difference between the size of a 0.99 carat diamond and a 1 carat diamond is undetectable to the naked human eye, but the price of a 1 carat diamond tends to be much higher than the price of a 0.99 diamond. In this question we use two random samples of diamonds,

0.99 carats and 1 carat, each sample of size 23, and compare the average prices of the diamonds. In order to be able to compare equivalent units, we first divide the price for each diamond by 100 times its weight in carats. That is, for a 0.99 carat diamond, we divide the price by 99. For a 1 carat diamond, we divide the price by 100. The distributions and some sample statistics are shown below.

	0.99 carats	1 carat
Mean	\$44.51	\$56.81
SD	\$13.32	\$16.13
n	23	23

Conduct a hypothesis test to evaluate if there is a difference between the average standardized prices of 0.99 and 1 carat diamonds. Make sure to state your hypotheses clearly, check relevant conditions, and interpret your results in context of the data.

13. (Exercise 7.26 from OpenIntro Statistics) Using the same data as the previous problem, construct a 95% confidence interval for the average difference between the standardized prices of 0.99 and 1 carat diamonds. You may assume the conditions for inference are met.