Hi Class, here are some homework problems on the two topics we discussed on Monday and Wednesday. These were

A confidence interval for the difference of population proportions. The short summary that if we have two independent random samples producing *sample* proportions $\hat{p}_1$ and $\hat{p}_2$ then a 95% CI for the difference in the population proportions is given by

$$(\hat{p}_1 - \hat{p}_2) \pm 1.96\mathsf{SE}(\hat{p}_1 - \hat{p}_2),$$

where the standard error is

$$\mathsf{SE}(\hat{p}_1 - \hat{p}_2) = \sqrt{\hat{p}_1(1-\hat{p}_1)/n_1 + \hat{p}_2(1-\hat{p}_2)/n_2}.$$

A significance test for a population proportion A significance test is a different form of statistical inference. Rather than finding an interval where we are confident a population parameter is in, we *assume* some fixed value of the parameter and then calculate a p-value, which is the probability a statistic is as extreme or more than the one observed value of the statistic *assuming* the fixed value of the parameter.

We use the language H_0, the null hypothesis, for the value we assume for the parameter. Whereas, H_A is the alternative hypothesis. This gives us an idea of what "extreme" means. Basically any value more in line with H_A than H_0.

For a test about a population proportion

$$H_0 : p = p_0, \quad H_A : p < p_0$$

we can use the test statistic (the z-score of the observed value of $\hat{p}$)

$$Z = \frac{\hat{p} - p}{\sqrt{p(1-p)/n}}$$

to define the p-value as

$$\text{p-value} = P(Z \leq \text{observed value of } Z | H_0 \text{ is true})$$

This is some probability computed with the normal distribution. The observed value is the z-score of the number produced from the data assuming the value of p in H_0.

If the alternative were $H_A : p > p_0$ then the p-value is

$$\text{p-value} = P(Z \geq \text{observed value of } Z | H_0 \text{ is true})$$

FInally, if the alternative were $H_A : p \neq p_0$ then the p-value is

$$text p - value = P(|Z| \geq |\text{observed value of } Z||H_0 \text{ is true})$$

which is a two sided picture when the density is drawn.

Excercise 1. To what extent do syntax textbooks, which analyze the structure of sentences, illustrate gender bias. A study sample sentences from 10 texts. A gender reference can be to an adult or a juvenile (Man or boy, or Woman or girl). Suppose the data found the following

gender	n	no. of juvenile references out of n
Female	60	48
Male	132	52

Find a 95% CI for the difference of population proportions. Does it include 0? If it does not include 0, explain why some might call this a "bias."

Excercise 2. Is there a gender difference in the rate of drunk riding among bicycle fatalities? Here is some data for bicycle fatalities accounting for gender and whether alcohol was involved

gender	n	no. tested positive
Female	192	27
Male	1520	515

Find a 90% CI for the difference in population proportions. (You will need to replace 1.96 in the formula above with an appropriate number.)

Is 0 included in your confidence interval?

Excercise 3. Is there a gender difference between students who seek Summer Employment? Suppose a financial aid office surveyed students and found the following

	Men	Women
Employed	728	603
Not Employed	89	149
Total	817	752

Find a 95% CI for the difference in the population proportions (for the proportion who are employed.). Does it include 0?

Excercise 4. In order for a software upgrade to be profitable for a company, more than 20% of its customers would need to purchase it. In a sample 11 of 40 indicated that they would. Perform a significance test (That is find a p-value) of

$$H_0 : p = 0.2 \qquad H_A : p > 0.2$$

Is the p-value less than 0.05? What does the p-value indicate to the company?

Excercise 5. Large trees growing near power lines can be an issue for power companies. A chemical treatment to inhibit a trees growth is tried out on a sample of trees. It will be used as long as the mortality rate is less than 15%. In the sample of 230 trees, 30 died. Perform a significance test (That is find a p-value) of

$$H_0 : p = 0.15 \qquad H_A : p < 0.15$$

Is the p-value less than 0.05? What does the p-value found indicate to the power company?

Excercise 6. A mathematician, John Kerrich, tossed a coin 10,000 times and found 5067 heads. Is this statistically significant evidence that the coin was biased?

(To answer this, find a p-value for the significance test

$$H_0 : p = 0.5, \qquad H_A : p \neq 0.5$$

If the p value is less than 0.05 the difference is statistically significant. Otherwise it is not.)

Excercise 7. Todd claims that he gets a strike 50% of the time. In 100 tries, he got 40 strikes. Find a p-value for the significance test

$$H_0 : p = 0.5, \qquad H_A : p < 0.5$$

Is there statistically significant evidence that Todd's claim is not true?