Problems Section 4C

(#1-10) Use each of the following two-population proportion Z-test statistics and the corresponding critical values to fill out the table.

	Z-test stat	Sentence to explain the two-population Z-test statistic.	Critical Value (+ right tail, − left tail, ± two tail)	Does the Z-test statistic fall in a tail determined by a critical value? (In Tail or Not in tail)	Does sample data significantly disagree with H_0?
1.	−1.835		±1.645		
2.	+0.974		+2.576		
3.	−1.226		−1.96		
4.	−3.177		±1.96		
5.	+2.244		+1.645		
6.	+1.448		±2.576		
7.	−0.883		−2.576		
8.	+1.117		+1.96		
9.	+2.139		±2.576		
10.	−0.199		−1.645		

(#11-20) Use each of the following P-values and corresponding significance levels to fill out the table.

	P-value Proportion	P-value %	Sentence to explain the P-value	Significance Level %	Significance level Proportion	If H_0 is true, could the sample data occur by random chance or is it unlikely?	Reject H_0 or Fail to reject H_0?
11.	0.728			10%			
12.	0.0421			1%			
13.	2.11×10^{-4}			5%			
14.	0.0033			1%			
15.	0.176			5%			
16.	0			10%			
17.	0.0628			5%			
18.	0.277			10%			
19.	3.04×10^{-6}			1%			
20.	0			5%			

21. Explain the difference between random samples and random assignment.

22. List the assumptions that we need to check for a two-population proportion hypothesis test.

23. List the assumptions that we need to check for a two-population proportion hypothesis test that is using experimental design.

24. Explain how to use a two-population proportion hypothesis test to show that two categorical variables are related.

25. Explain how to use a two-population proportion hypothesis test to show there is a cause and effect between two categorical variables.

(#26-30) Directions: Use the following Statcato printouts to answer the following questions.

a) Write the null and alternative hypothesis. Include relationship implications. Is this a left-tailed, right-tailed, or two-tailed test?

b) Check all of the assumptions for a two-population proportion Z-test. Explain your answers. Does the problem meets all the assumptions?

d) Write a sentence to explain the Z-test statistic in context.

e) Use the test statistics and the critical value to determine if the sample data significantly disagrees with the null hypothesis. Explain your answer.

f) Write a sentence to explain the P-value.

g) Use the P-value and significance level to determine if the sample data could have occurred by random chance (sampling variability) or is it unlikely to random chance? Explain your answer.

h) Should we reject the null hypothesis or fail to reject the null hypothesis? Explain your answer.

i) Write a conclusion for the hypothesis test. Explain your conclusion in plain language.

j) Is the population proportion related to the categorical variable or not? Explain your answer.

26. The United States has the highest teen pregnancy rate in the industrialized world. In 2008, a random sample of 1014 teenage girls found that 326 of them were pregnant before the age of 20. In 2012, a random sample of 1025 teenage girls was taken and 334 were found to be pregnant before the age of 20. Let population proportion 1 represent 2008 and population proportion 2 represent 2012. Use a 10% significance level and the following Statcato printout to test the claim that the population percentage of teen pregnancies in the U.S. is lower in 2008 than it is in 2012. This claim would indicate that the population percentage of U.S. teen pregnancies is related to the year.

	Number of Events	Number of trials	Proportion
Sample 1	334	1025	0.326
Sample 2	326	1014	0.321

Significance Level	Critical Value	Test Statistic Z	p-Value
0.05	-1.645	-0.210	0.4168

27. While many Americans favor the legalization of marijuana, opponents of legalization argue that marijuana may be a gateway drug. They believe that if a person uses marijuana, then they are more likely to use other more dangerous illegal drugs. Use the table of random sample data given below and a 5% significance level to test the claim that marijuana users have a higher percentage of other drug use than non-marijuana users. This claim also would indicate that using Marijuana is related to using other drugs.

	Uses Other Drugs	Total
Uses Marijuana	87	213
Does not use Marijuana	26	219

	Number of Events	Number of trials	Proportion
Sample 1	87	213	0.408
Sample 2	26	219	0.119

Significance Level	Critical Value	Test Statistic Z	p-Value
0.05	1.645	6.850	$3.6839 \cdot 10^{-12}$

28. Use a 1% significance level and the following Statcato printout to test this claim that gender is not related to abstaining from drinking alcohol. If this is the case, then the percentage of men and women that do not drink alcohol should be the same. We took a random sample of 190 men and found that 66 of them never drink alcohol. We took a random sample of 250 women and found that that 137 of them never drink alcohol. We designated the proportion of men that never drink alcohol as population 1 and the women as population 2.

	Number of Events	Number of trials	Proportion
Sample 1	66	190	0.347
Sample 2	137	250	0.548

Significance Level	Critical Value	Test Statistic Z	p-Value
0.01	-2.576, 2.576	-4.182	$2.8937 \cdot 10^{-5}$

29. A health magazine claims that marriage status is one of the most telling factors for a person's happiness. Use a 10% significance level and the Statcato printout below to test the claim that the percent of married people that are unhappy is lower than the percent of single or divorced people that are unhappy. If this is the case, then perhaps being married, single or divorced is related to being unhappy. The following sample data was collected randomly. Population 1 represented married adults and population 2 represented single or divorced adults.

	Unhappy	Total
Married	74	200
Single or Divorced	97	200

	Number of Events	Number of trials	Proportion
Sample 1	74	200	0.37
Sample 2	97	200	0.485

Significance Level	Critical Value	Test Statistic Z	p-Value
0.10	-1.282	-2.325	0.0100

30. A tattoo magazine claimed that the percent of men that have at least one tattoo is greater than the percent of women with at least one tattoo. If this were true, then gender would be related to having a tattoo. Use a 5% significance level and the following Statcato printout to test this claim. A random sample of 857 men found that 146 of them had at least one tattoo. A random sample of 794 women found that 137 of them had at least one tattoo. Population 1 was the proportion of men with at least one tattoo and population 2 was the proportion of women with at least one tattoo.

	Number of Events	Number of trials	Proportion
Sample 1	146	857	0.170
Sample 2	137	794	0.173

Significance Level	Critical Value	Test Statistic Z	p-Value
0.05	1.645	-0.118	0.5468

(#31-34) Directions: go to www.lock5stat.com and click on StatKey. Then under the "Randomization Hypothesis Test" menu click on "Test for Difference in Proportions". Create a randomized simulation of the null hypothesis to answer the following questions.

a) Write the null and alternative hypothesis. Include relationship implications. Is this a left-tailed, right-tailed, or two-tailed test?

b) What is the difference between the sample proportions? Adjust the tails of your simulation to reflect the significance level. Did your sample proportion difference fall in the tail?

c) Does the sample data significantly disagree with the null hypothesis? Explain your answer.

d) Put the sample proportion difference into the bottom box in the appropriate tail of your simulation in order to calculate the P-value. What was the P-value? (Answers will vary.) Write a sentence to explain the P-value.

e) Use the P-value and significance level to determine if the sample data could have occurred by random chance (sampling variability) or is it unlikely to random chance? Explain your answer.

f) Should we reject the null hypothesis or fail to reject the null hypothesis? Explain your answer.

g) Write a conclusion for the hypothesis test. Explain your conclusion in plain language.

h) Is the population proportion related to the categorical variable or not? Explain your answer.

i) Use the following formula to calculate the Z-test statistic. Write a sentence to explain the Z-test statistic in context. (Answers will vary.)

$$Z\ test\ stat = \frac{Sample\ Proportion\ Difference}{Standard\ Error}$$

31. A body mass index of 20-25 indicates that a person is of normal weight for their height and body type. A random sample of 760 women found that 198 of the women had a normal BMI. A random sample of 745 men found that 273 of them had a normal BMI. A fitness magazine claims that the percent of women with a normal BMI is lower than the percent of men with a normal BMI. This would imply that gender is related to having a normal BMI. Let population 1 be the proportion of women with a normal BMI and population 2 be the proportion of men with a normal BMI. Use a 10% significance level and a randomized simulation in StatKey.

32. A new medicine has been developed that treats high cholesterol. An experiment was conducted and adults were randomly assigned into two groups. The groups had similar gender, ages, exercise patterns and diet. Of the 420 adults in the placebo group, 38 of them showed a decrease in cholesterol. Of the 410 adults in the treatment group, 49 of them showed a decrease in cholesterol. The FDA claims that the medicine is not effective in lowering cholesterol since the proportion for the placebo group and the treatment groups are about the same. Use a randomized simulation in StatKey, and a 1% significance level to test this claim.

33. A study was done to see if there is a relationship between smoking and being able to get pregnant. Two random samples of women trying to get pregnant were compared. A random sample of 135 women that smoke (population 1) found that 38 were able to get pregnant in the allotted amount of time. A random sample of 543 women that do not smoke (population 2) found that 206 were able to get pregnant in the allotted amount of time. Test the claim that the population percent of smoking women that were able to get pregnant is lower than the population percent of non-smoking women. This claim also implies that smoking is related to getting pregnant. Use a randomized simulation in StatKey, and a 5% significance level to test this claim.

34. A study was done to see if there is a relationship between the age of a person (teen or adult) and using text messages to communicate. A random sample of 800 teens (population 1) found that 696 of them use text messages regularly to communicate. A random sample of 2252 adults (population 2) found that 1621 of them use text messages regularly to communicate. Test the claim that population percentages are equal for the two groups implying that age is not related to using text messages. Use a randomized simulation in StatKey, and a 10% significance level to test this claim.

Social Justice Questions

35. An experiment was done on labor market discrimination. The statisticians created fictitious resumes to help-wanted adds in Boston and Chicago newspapers with both male and female sounding names. Each resume was randomly assigned either a very black sounding name or a very white sounding male or female name. Other confounding variables like experience, age, type of job, where the person lives, etc. were controlled. The data met the assumptions for a two-proportion hypothesis test. The following table summarizes the findings of multiple two-population proportion hypothesis tests including P-values.

Table 1
Mean Call-Back Rates By Racial Soundingness of Names [a]

	Call-Back Rate for White Names	*Call-Back Rate for African American Names*	*Ratio*	*Difference (p-value)*
Sample:				
All sent resumes	**10.06%** [2445]	**6.70%** [2445]	**1.50**	**3.35%** (.0000)
Chicago	**8.61%** [1359]	**5.81%** [1359]	**1.48**	**2.80%** (.0024)
Boston	**11.88%** [1086]	**7.83%** [1086]	**1.52**	**4.05%** (.0008)
Females	**10.33%** [1868]	**6.87%** [1893]	**1.50**	**3.46%** (.0001)
Females in administrative jobs	**10.93%** [1363]	**6.81%** [1364]	**1.60**	**4.12%** (.0001)
Females in sales jobs	**8.71%** [505]	**6.99%** [529]	**1.25**	**1.72%** (.1520)
Males	**9.19%** [577]	**6.16%** [552]	**1.49**	**3.03%** (.0283)

[a] Notes:

1. The table reports, for the entire sample and different subsamples of sent resumes, the call-back rates for applicants with a White sounding name (column 1) and an African American sounding name (column 2), as well as the ratio (column 3) and difference (column 4) of these call-back rates. In brackets in each cell is the number of resumes sent in that cell.
2. Column 4 also reports the p-value for a test of proportion testing the null hypothesis that the call-back rates are equal across racial groups.

a) Let's look at all the resumes (male and female) from Chicago. When comparing callbacks for resumes with white sounding names to resumes with black sounding names, the P-value was 0.0024 (0.24%). Assuming the significance level was 5%, would this P-value indicate that there was a significant difference? The experiment controlled confounding variables. So does the experiment show that there is racial discrimination in the labor market in Chicago?

b) Let's look at all the resumes (male and female) from Boston. When comparing callbacks for resumes with white sounding names to resumes with black sounding names, the P-value was 0.0008 (0.08%). Assuming the significance level was 5%, would this P-value indicate that there was a significant difference? The experiment controlled confounding variables. So does the experiment show that there is racial discrimination in the labor market in Boston?

c) Let's look at all the resumes from males applying for jobs in Boston and Chicago. When comparing callbacks for resumes with white sounding names to resumes with black sounding names, the P-value was 0.0283 (2.83%). Assuming the significance level was 5%, would this P-value indicate that there was a significant difference? The experiment controlled confounding variables. So does the experiment show that there is racial discrimination in the labor market in Boston and Chicago for males attempting to get jobs?

d) Let's look at all the resumes from females applying for administrative jobs in Boston and Chicago. When comparing callbacks for resumes with white sounding names to resumes with black sounding names, the P-value was 0.0001 (0.01%). Assuming the significance level was 5%, would this P-value indicate that there was a significant difference? The experiment controlled confounding variables. So does the experiment show that there is racial discrimination in the labor market in Boston and Chicago for females attempting to get administrative jobs?

e) Let's look at all the resumes from females applying for sales jobs in Boston and Chicago. When comparing callbacks for resumes with white sounding names to resumes with black sounding names, the P-value was 0.1520 (15.2%). Assuming the significance level was 5%, would this P-value indicate that there was a significant difference? The experiment controlled confounding variables. So does the experiment show that there is racial discrimination in the labor market in Boston and Chicago for females attempting to get sales jobs?
