STA 100 Fall 2018
Homework 7 - Due Thursday, December 6^{th} by 11PM.

Book Homework

These prompts correspond to "Book Homework" portion of the homework on Canvas. You turn in the answers to these questions online.

1. An experiment was conducted in which two types of acorn squash were crossed. According to the genetic model, 1/2 of the resulting plants should have dark stems and dark fruit (A), 1/4 should have light stems and dark fruit (B), and 1/4 should have light stems and light fruit (C). The actual recorded counts are:

	A	B	C
count	220	129	105

   (a) State the null hypothesis for testing the genetic model.
   (b) State the alternative hypothesis.
   (c) Calculate the test-statistic.
   (d) Estimate the p-value.
   (e) Do we reject, or fail to reject the null if $\alpha = 0.05$?
   (f) State your conclusion in terms of the problem.

2. At a Midwestern hospital there were a total of 932 births in 20 consecutive weeks. Of those births, 216 occurred on weekends (Saturday and Sunday) and 716 occurred on weekdays (Monday, Tuesday, Wednesday, Thursday, Friday). The goal is to determine if approximately the same about of births occur on the weekends that would happen by chance (by chance is 2/7).

   (a) State the null hypothesis.
   (b) State the alternative hypothesis.
   (c) Calculate the test-statistic.
   (d) Estimate the p-value.
   (e) Interpret your p-value in terms of the problem.
   (f) Do we reject, or fail to reject the null?
   (g) State your conclusion in terms of the problem.

3. Mongolian gerbils are thought to be equally likely to be brown, white, or black in color. A random sample showed the following frequencies:

	Black	Brown	White
count	40	59	42

   (a) State the null hypothesis.
   (b) State the alternative hypothesis.
   (c) Calculate the test-statistic.
   (d) Estimate the p-value.
   (e) Using $\alpha = 0.1$, do we reject or fail to reject the null?
   (f) State your conclusion in terms of the problem.
   (g) What category differed the most from what we expected if the null hypothesis was true?

4. Men with prostate cancer were randomly assigned to undergo surgery, or "watchful waiting"(WW). The results over the next several years are found below:

	Surgery	WW
Died	83	106
Alive	264	242

   Assume we want to test if death was independent of what group they were in (Surgery, WW) with $\alpha = 0.05$.

   (a) State the appropriate null hypothesis.
   (b) State the appropriate alternative hypothesis.
   (c) Calculate the test-statistic.
   (d) Find the range for the p-value.
   (e) State your decision. Did you reject, or fail to reject the null?
   (f) State your conclusion in terms of the problem.
   (g) What was the sample estimate of the probability of death, if you were in group Surgery?
   (h) What was the sample estimate of the probability of death, if you were in group WW?

5. A randomized experiment was conducted in which patients with coronary artery disease either had angioplasty (A) or bypass surgery (B). The accompanying table shows the treatment type and if chest pain occurred over the next 5 years:

	A	B
Pain	111	74
No Pain	402	441

   Assume we want to test if chest pain was independent of treatment, with $\alpha = 0.01$.

   (a) State the appropriate null hypothesis.
   (b) State the appropriate alternative hypothesis.
   (c) Calculate the test-statistic.
   (d) Find the range for the p-value.
   (e) Interpret your p-value in terms of the problem.
   (f) State your decision. Did you reject, or fail to reject the null?
   (g) State your conclusion in terms of the problem.
   (h) What was the sample estimate of the probability of pain, if you were in group A?
   (i) What was the sample estimate of the probability of pain, if you were in group B?

6. Answer the following questions with TRUE or FALSE. It is good practice to explain your answers.

   (a) When you are testing if the probability of being in k categories of a categorical variable is equal, $p_i = 1/\text{k}$ for all i.

R Homework

These prompts correspond to "R Portion" of the homeworks on Canvas. You use R to find the answers to the following questions, and submit your answers online.

I. Biologists wanted to know if a particular parasite (weevils) had a preference for a particular bean plant. The results are stored in `beans.csv`, where the result was which plant the weevil preferred (`Pinto`, `Cowpea`, `Navy`, `Northern`). Use this data and R, and assume the claim is that weevils are equally likely to select all bean plants.

(a) Find the test statistic.

(b) Find the p-value.

(c) Did you reject or fail to reject the null if $\alpha = 0.05$?

(d) Which bean plant(s) did the weevils prefer more than expected if the null was true?

(e) What category contributed most to the value of the test-statistic?

II. On Canvas you will find the file `blood.csv`, which has two columns: `Type` (for blood type), and `Disease` (either yes or no), which is data taken from a random sample from a particular area. Use R and this data to answer the following questions:

(a) State the appropriate null for testing if having this particular disease is independent of blood type.

(b) State the appropriate alternative hypothesis for testing if having this particular disease is independent of blood type.

(c) Find the test-statistic.

(d) Find the p-value.

(e) State your conclusion in terms of the problem, assuming $\alpha = 0.01$.

(f) Were blood type A individuals more or less likely to have the disease than what we expected if the null was true?

(g) Were blood type O individuals more or less likely to have the disease than what we expected if the null was true?

(h) Which group contributed most to the value of the test-statistic?