Group Work - Week 10

1 Conduct tests for the scenarios below at a $\alpha = 0.05$ level of significance. Be sure to state your conclusion in the context of the question.

(a) Researchers discover a new gene which, under the right circumstances, could lead to a mildly inconvenient, but chronic, disease. 10% of the general population have the gene. One of the researchers thinks that people with naturally red hair are more likely to have the gene. Genetic tests are conducted on a sample of 65 redheads and it is found that 11 of them have the gene.

```
> prop.test(11, 65, .1, alternative = "greater", correct = FALSE)

1-sample proportions test without continuity correction

data:  11 out of 65, null probability 0.1
X-squared = 3.4615, df = 1, p-value = 0.03141
alternative hypothesis: true p is greater than 0.1
95 percent confidence interval:
 0.1063379 1.0000000
sample estimates:
        p
0.1692308
```

The test statistic is $z = 1.861$. The p-value is $p = 0.0314 < \alpha = 0.05$. Reject the null hypothesis. There is evidence that redheads have a higher occurrence of the gene.

(b) A coffee shop is interested in the proportion of decaf coffee drinkers on Sunday and Monday mornings. The manager thinks they have a lower proportion of decaf drinkers on Monday. They examine a random sample of coffee orders and find that on Sunday 52 out of 156 orders are for decaffeinated coffee and on Monday 43 out 174 are decaf orders.

```
> prop.test(c(52, 43), c(156, 174), alternative="greater", correct=FALSE)

2-sample test for equality of proportions without continuity correction

data:  c(52, 43) out of c(156, 174)
X-squared = 2.9818, df = 1, p-value = 0.0421
alternative hypothesis: greater
95 percent confidence interval:
 0.004066543 1.000000000
sample estimates:
   prop 1    prop 2
0.3333333 0.2471264
```

$z = 1.73$, $p = 0.0421 < \alpha = 0.05$. Reject H_0.
There is evidence that the proportion of decaf coffee orders is lower on Mondays.

2 M&Ms are expected to have the following distribution:

Color	Blue	Brown	Green	Orange	Red	Yellow
Percent	24%	14%	15%	20%	13%	14%

(a) What is the minimum number of M&Ms needed to do a valid goodness-of-fit test against the expected distribution. In other words, what is the sample size n so that the smallest expected value is at least 5?

Since red is the least frequent color in the expected distribution, we need to find a sample size so that there are at least 5 expected red M&M's.

$\boldsymbol{n \times 0.13 \geq 5, \qquad n \geq \frac{5}{0.13} = 38.46}$

So, the sample size should be at least 39 M&M's.

(b) Conduct a goodness-of-fit test of whether the distribution of M&Ms is what is claimed by the company at a significance level of $\alpha = 0.05$. Make sure to state the null and alternative hypotheses, and your conclusion in context of question.

(There are about 16 M&M's in a regular fun size pack and about 7 M&M's in a peanut fun size pack.)

The results will vary depending on the exact frequency of colors each finds.

Example:

```
> expected.dist <- c(0.24, 0.14, 0.15, 0.2, 0.13, 0.14)
> color.counts <- c(12, 9, 3, 11, 8, 5)

> chisq.test(color.counts, p=expected.dist)

Chi-squared test for given probabilities

data:  color.counts
X-squared = 4.3844, df = 5, p-value = 0.4955
```

$\boldsymbol{\chi^2 = 4.38}$, $\boldsymbol{p = 0.50 > \alpha = 0.05}$

Fail to reject H_0. There is no evidence that M&M colors don't follow the given distribution.

2 The Tortilla and Cheese Organization (TACO) thinks that preferences for types of tacos are the same for men and women. They conduct a survey and collect the following data ("taco_preference.csv" on D2L):

	Type of taco			
Gender	Beef	Pork	Chicken	Fish
Men	105	34	56	27
Women	83	29	75	35

Test the claim the taco preference is the same for men and women at $\alpha = 0.05$ level of significance. Make sure to state the null and alternative hypotheses, and your conclusion in context of question.

H_0 : There is no association between gender and taco preference, or gender and taco preference are independent
H_a : There is an association between gender and taco preference, or gender and taco preference are not independent

```
> taco <- matrix(c(105, 83, 34, 29, 56, 75, 27, 35), nrow=2)
> taco
     [,1] [,2] [,3] [,4]
[1,]  105   34   56   27
[2,]   83   29   75   35

> chisq.test(taco)

Pearson's Chi-squared test

data:  taco
X-squared = 6.7593, df = 3, p-value = 0.07998
```

$\chi^2 = 6.76$, $p = 0.08 > \alpha = 0.05$
Fail to reject H_0. There is no evidence that taco preference and gender are not independent.