STAT120 HW 8

Victor Huang

5/24/2021

Section 7.2

7.32

Expected count: $\frac{330 \cdot 160}{600} = 88$

chi-square: $\frac{(89-88)^2}{88} = 0.01136$

7.41

```
##      [,1] [,2] [,3]
## [1,]   10   40   50
## [2,]   18   32   50
## [3,]   28   72  100
```

```
##
##  Pearson's Chi-squared test
##
## data:  out1
## X-squared = 3.1746, df = 4, p-value = 0.529
```

$H_0 : p_m = p_e; H_a : p_m \neq p_e$ (p_e is the proportion of electric tagged penguins that survived, p_m is the proportion of magnet tagged penguins that survived)

```
##      [,1] [,2]
## [1,]   14   36
## [2,]   14   36
```

```
##
##  Pearson's Chi-squared test
##
## data:  out1
## X-squared = 3.1746, df = 4, p-value = 0.529
```

The chi-square value is: 3.1746

Since our p-value is larger than the significance value of 0.05, we fail to reject the null hypothesis.

7.46

H_0 : The method of choosing skittles does not affect their favorability; H_a: The method of choosing skittles does affect their favorability

```
##          [,1]     [,2]     [,3]     [,4]      [,5]
## [1,] 13.03185 10.50955 14.29299 19.75796  8.407643
## [2,] 17.96815 14.49045 19.70701 27.24204 11.592357
```

```
##
##  Pearson's Chi-squared test
```

```
##
## data:  out3
## X-squared = 9.0691, df = 4, p-value = 0.0594
```

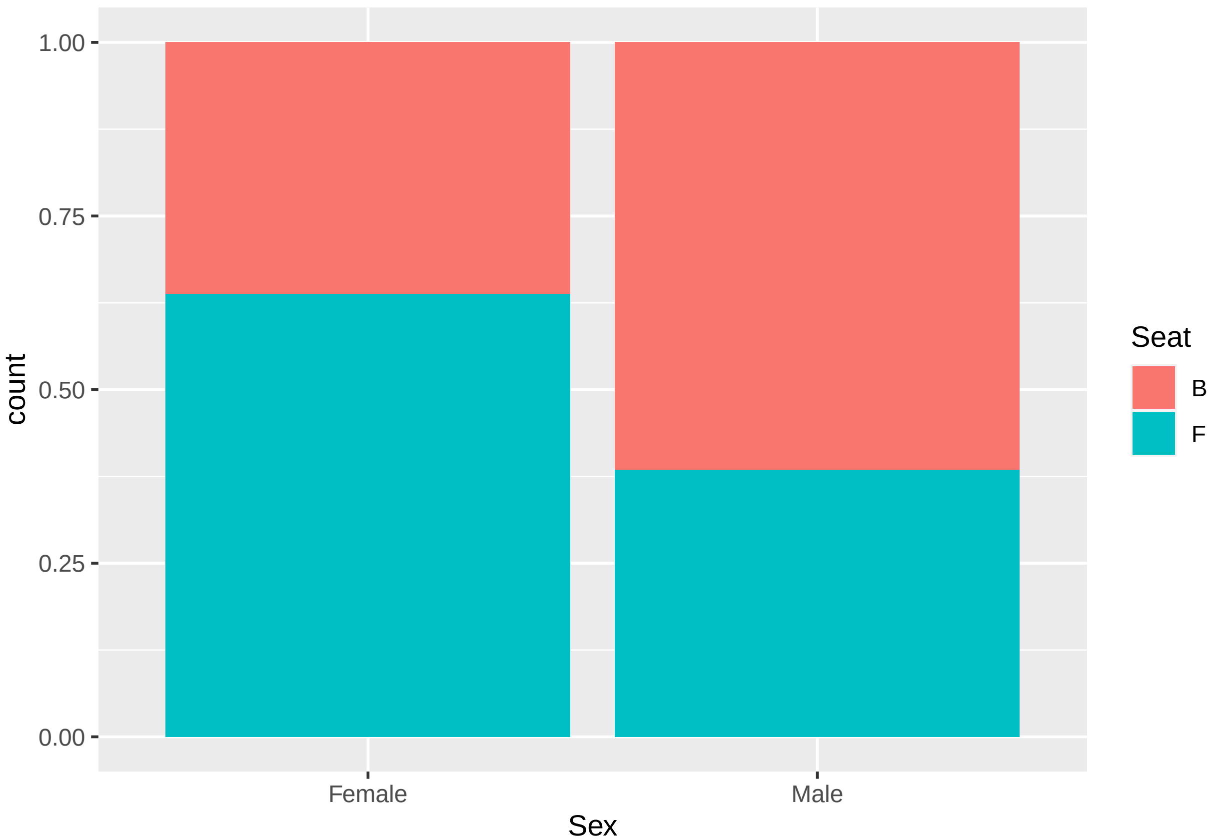
Yes, the chi-squared value is 9.0691

The degrees of freedom is 1 times 4, which is 4.

We get a p-value of 0.0594, which is larger than a 0.05 significance value. As such, we fail to reject the null hypothesis. As such, we find no evidence that method of choosing effects skittle favorability.

Section 8.6

1




```
##
##  Pearson's Chi-squared test with Yates' continuity correction
##
## data:  out
## X-squared = 5.0414, df = 1, p-value = 0.02475
```

It appears the females favor sitting in the front while males favor sitting in the back.

Creating a table with the Students data set and preforming a chi-squared test on it. We get a p-value of 0.02475, which is much less than a 0.05 significance level. As such, we reject the null hypothesis and conclude with their being an association of sitting in the front or back based on your gender.

4

```
##
##     away home
##   L   16   15
```

```
##   T    3    4
##   W   20   24
```

```
## Warning in chisq.test(out): Chi-squared approximation may be incorrect
```

```
##
##  Pearson's Chi-squared test
##
## data:  out
## X-squared = 0.34445, df = 2, p-value = 0.8418
```

```
##
##  Pearson's Chi-squared test with Yates' continuity correction
##
## data:  (out1)
## X-squared = 0.11889, df = 1, p-value = 0.7302
```

With a p-value of 0.8418, we fail to reject the null hypothesis. As suchm we can not conclude that there is an association between location and outcome. As such, we can draw the conclusion that location does not affect the outcome of the game. The warning sign is due to the small count (the 4 and 3). Chi-square tests are usually preformed on 2x2 or larger matrices with each element being larger or equal to 5.

With the modified table, we get a different chi-squared, df, and p value. However, our p-value still remains larger than 0.05. As such, we still draw the same conclusion that location does not affect outcome.

Section 8.1

8.4

Dataset A, while both datasets have the same difference in means, dataset A has a higher variability within each group (i.e. a more elongated box).

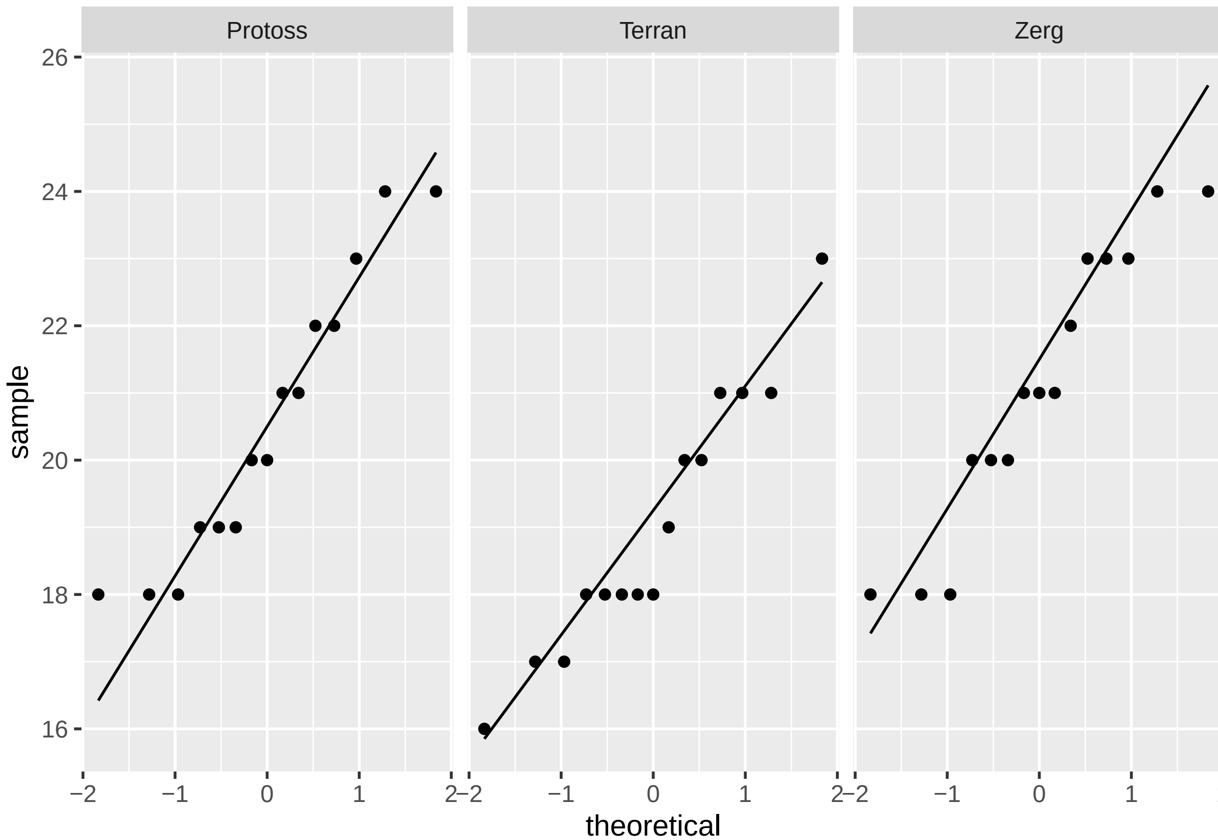
8.16

a. Since both groups ahve a positive mean value, we can see that synchronization does indeed go up.
b. Since the group with HE has a positive mean, their synchronization goes up while the group with LE has a negative mean, as such their synchronization does down.
c. Given a DF of 259, we can find the total by adding 1 to that value. As such, we can conclude that the total number of students that participated is 260.
d. Since our p-value is 0.042, it is less than the 0.05 significance level. As such we reject the null hypothesis. As such, we can conclude that their is indeed a difference in means.
e. Since out p-value is 0.042, it is greater than the 0.01 significance level. As such, we fail to reject the null hypothesis. As such, we can not conclude that their is a difference in means.

Section 9.3

1

There are 15 Protoss, Terran, and Zerg players with a sample mean of 20.5333, 19, and 21.0667 for their average age

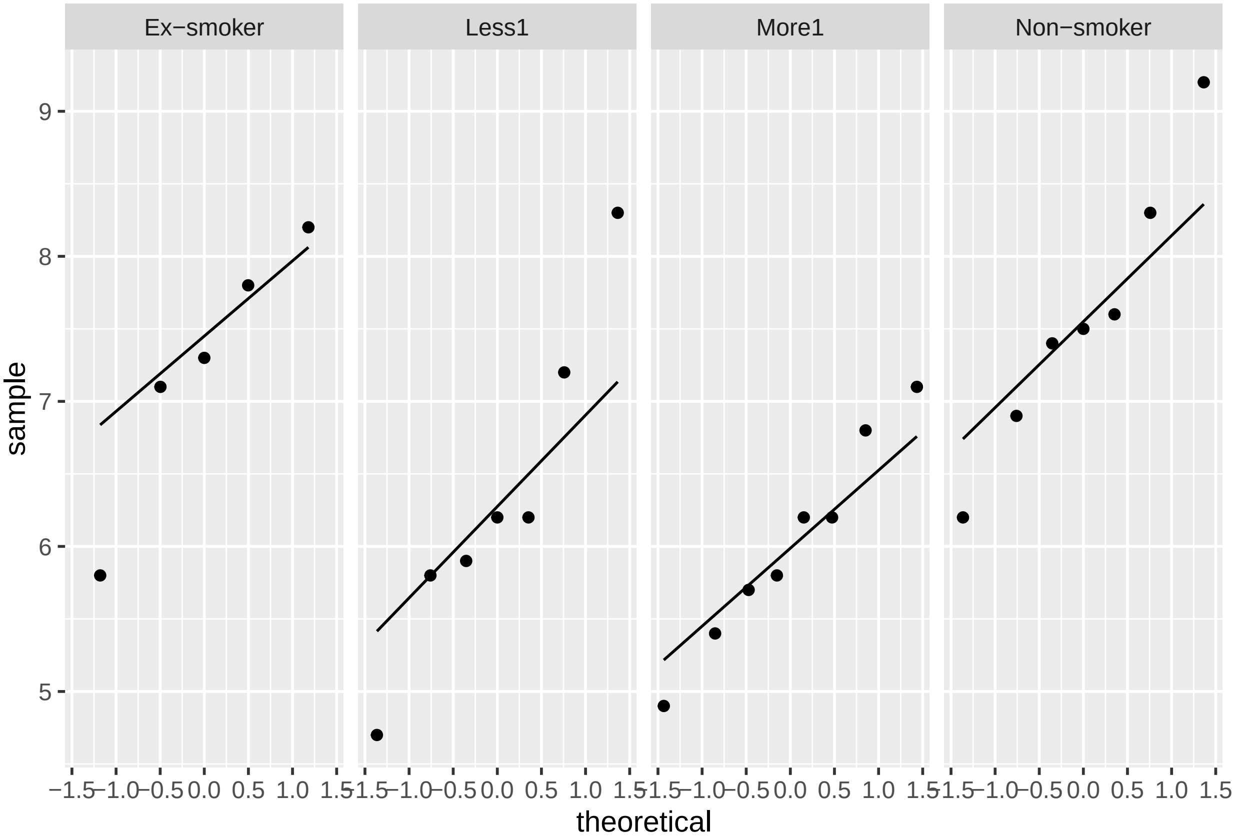



```
##             Df Sum Sq Mean Sq F value Pr(>F)
## Race         2  34.53  17.267   4.152 0.0226 *
## Residuals   42 174.67   4.159
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

The conditions are satisfied for preforming an ANOVA test

Since we have such a small p-value we can conclude that there is a difference in mean age across the three races

2




```
##             Df Sum Sq Mean Sq F value Pr(>F)
## status       3  11.67   3.891   4.408 0.0137 *
## Residuals   23  20.30   0.883
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

OUr hypothesis are as follows:

H_0: Smoking status of mothers do not affect the birthweight of their babies vs. H_a: Smoking status of mothers do affect the birthweight of their babies.

After doing an ANOVA test on the data given, we get a p-value of 0.0137, which is smaller than the 0.05 significance level. As such, we can conclude that there is a relationship between smoking status and mean birth weights of their babies. THis is concluded on the basis of assuming that each group is normally distributed, populations have a common variance, samples are independent from each other, and that observations are random and independent.

Section 8.2

8.41

Yes, the p-value is 0.002, much lower than the 0.05 sig level.

8.42

Pooled SD: $\sqrt{6.20} = 2.490$

DF: 12

8.43

95% confidence interval: $10.200 \pm 2.178813\frac{\sqrt{6.20}}{\sqrt{5}} = 10.200 \pm 2.426 = 7.774$ to 12.626

8.44

90% confidence interval: $(16.8 - 10.8) \pm 1.782288\sqrt{\frac{6.20^2}{5^2} + \frac{6.20^2}{5^2}} = 6 \pm 2.68 = 3.32$ to 8.68

8.45

With our hypothesis as follows:

$H_0 : \mu_A = \mu_C; H_a : \mu_A \neq \mu_C$

we use a chi-squared test to find if there is any evidence for difference. After the chi-squared test we get a p-value of 0.71. Since the p-value is greater than the 0.05 sig level, we fail to reject the null hypothesis and conclude that there is no difference in means between A and C.

8.54

a. $H_0 : \mu_{aa} = \mu_{np}; H_a : \mu_{aa} \neq \mu_{np}$

Using a chi-squared test, we find that the p-value is 0.51. Since it is greater than the 0.05 sig level, we fail to reject the null hypothesis and find no difference in means for students wanting the two awards in question.

b. $H_0 : \mu_{aa} = \mu_{ogm}; H_a : \mu_{aa} \neq \mu_{ogm}$

Using a chi-squared test, we find that the p-value is 0.134. Since it is greater than the 0.05 sig level, we fail to reject the null hypothesis and find no difference in means for students wanting the two awards in question.

c. $H_0 : \mu_{np} = \mu_{ogm}; H_a : \mu_{np} \neq \mu_{ogm}$

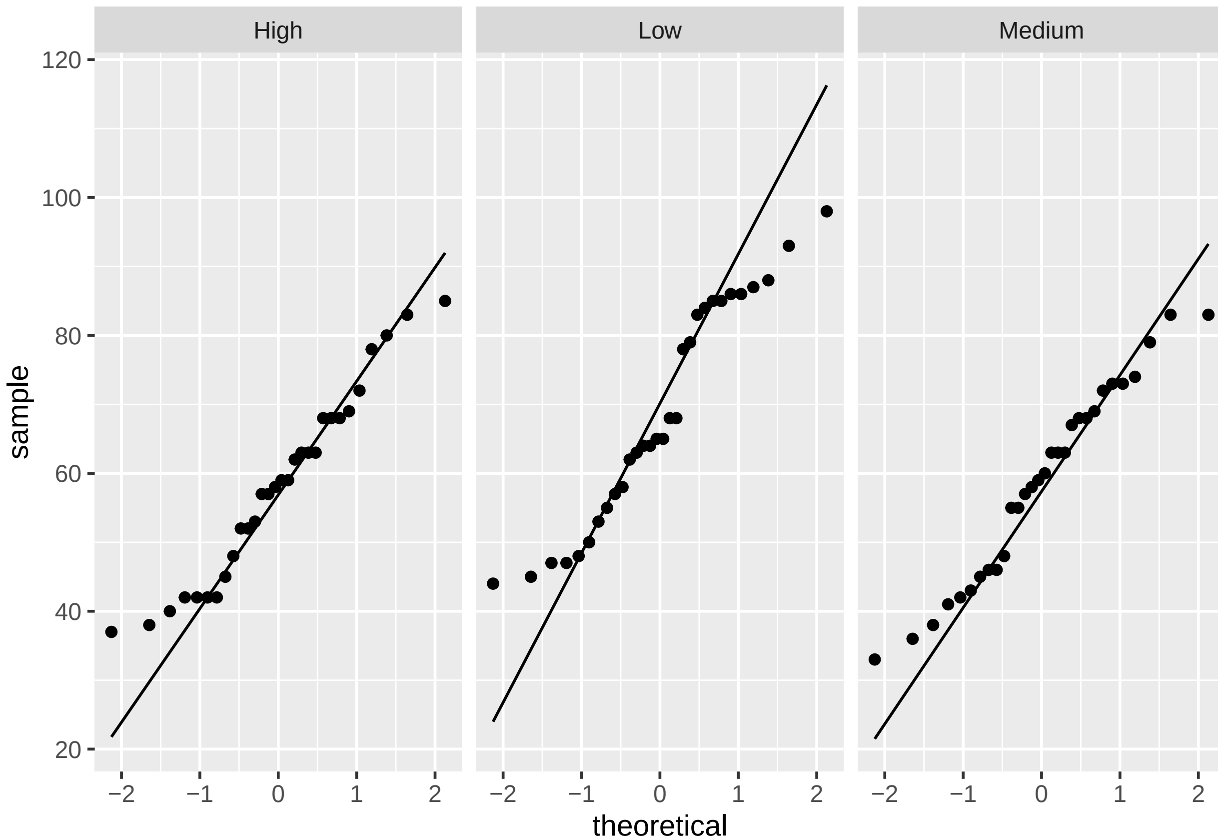
Using a chi-squared test, we find that the p-value is 0.0002. Since it is less than the 0.05 sig level, we are able to reject the null hypothesis and find evidence that there is a difference in means for students wanting the two awards in question.

8.62

```
##
## Attaching package: 'dplyr'

## The following objects are masked from 'package:stats':
##
##     filter, lag

## The following objects are masked from 'package:base':
##
##     intersect, setdiff, setequal, union
```




```
##             Df Sum Sq Mean Sq F value Pr(>F)
## Calcium      2   2037  1018.6   4.648 0.0121 *
## Residuals   87  19064   219.1
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```
## `summarise()` ungrouping output (override with `.groups` argument)
```

```
## # A tibble: 3 x 3
##   Calcium `mean(GillRate)` `sd(GillRate)`
##   <chr>              <dbl>          <dbl>
## 1 High                58.2           13.8
## 2 Low                 68.5           16.2
## 3 Medium              58.7           14.3
```

After calculating the respective p-values for each pair (low vs. medium, low vs. high, medium vs. high) we get that the p-values for low vs. medium and low vs. high to be smaller than the 0.05 sigificance level. As such, those are the levels that have a difference in means.