Homework 2

STA 138 | Camden Possinger | Winter 2022

1. A winemaker claims that one fifth of her wine barrels are infected with Brettanomyces. You will independently sample 6 of her barrels, and use a two-sided Binomial test with = 0.01 to evaluate this claim.

(a) Using the p-value to conclude your test, what is the probability of a type I error?

```
X_i <- 0:6
p_values <- sapply(X_i, \(x){binom.test(x,6,1/5,"two.sided")$p.val})
nullProbs <- dbinom(0:6,6,1/5)


wine_frame <- cbind("X Values" = X_i, "P Values" = p_values, "Null Probabilities" = nullProbs)

kbl(wine_frame)
```

Answer:

X Values	P Values	Null Probabilities
0	0.606784	0.262144
1	1.000000	0.393216
2	0.344640	0.245760
3	0.098880	0.081920
4	0.016960	0.015360
5	0.001600	0.001536
6	0.000064	0.000064

```
nullProbs[p_values <= 0.01] %>% sum %>% kbl(col.names = c("Type I Error"))
```

Type I Error
0.0016

In this case the probability of a Type I error is : 0.0016

(b) Using the mid-p-value instead, what is the probability of a type I error?

```
p_values_mid <- p_values - 0.5 * nullProbs

wine_frame_mid <- cbind("X Values" = X_i, "Mid P Values" = p_values_mid, "Null Probabilities" = nullPro
kbl(wine_frame_mid)
```

Answer:

X Values	Mid P Values	Null Probabilities
0	0.475712	0.262144
1	0.803392	0.393216
2	0.221760	0.245760
3	0.057920	0.081920
4	0.009280	0.015360
5	0.000832	0.001536
6	0.000032	0.000064

```
nullProbs[p_values_mid <= 0.01] %>% sum %>% kbl(col.names = c("Type I Error"))
```

Type I Error
0.01696

In this case the probability of a Type I error is : 0.01696

---

2. Continue from the previous problem; suppose that, in truth, 40% of the wine barrels are infected. Suppose also that the p-value will be used to conclude the test.

(a) Is the probability of rejecting the null hypothesis the same as it was in problem I.(a) above? Explain.

```
trueProbs <- dbinom(0:6,6,2/5)


wine_frame_true <- cbind("X Values" = X_i, "P Values" = p_values, "True Probabilities" = trueProbs)

kbl(wine_frame_true)
```

Answer:

X Values	P Values	True Probabilities
0	0.606784	0.046656
1	1.000000	0.186624
2	0.344640	0.311040
3	0.098880	0.276480
4	0.016960	0.138240
5	0.001600	0.036864
6	0.000064	0.004096

```
trueProbs[p_values <= 0.01] %>% sum %>% kbl(col.names = c("Power"))
```

Power
0.04096

The probability of rejecting the null hypothesis is different when we consider the true proportion because we change the underlying binomial distribution when constructing the p-values. In this case the probability of rejecting the null hypothesis is a little bit larger which makes sense given that the tails of the true distribution are more spread out. In this case we are also looking for the power not the chance of a Type I error.

(b) What is the power of this test?

```
pow_sml_samp <- 1-(trueProbs[p_values > 0.01] %>% sum)
pow_sml_samp %>% kbl(col.names = c("Power"))
```

Answer:

Power
0.04096

(c) What would the power of this test be if, instead of 6, the sample size were 60? Explain how this differs from your result in part (b). Is this what you expect from a larger sample?

```
X_i <- 0:60

lrg_true_probs <- dbinom(X_i,60,2/5)

lrg_p_vals <- X_i %>% sapply(\(x){binom.test(x,60,0.2,c("two.sided"))$p.val})


pow_lrg_samp <- 1-(lrg_true_probs[lrg_p_vals > 0.01] %>% sum)


pow_lrg_samp %>% kbl(col.names = c("Power"))
```

Answer:

Power
0.8214298

The power is larger with a larger sample size compared to a smaller sample size which makes sense since the mass function will be more spread out.

---

3. Data on the longevity of 409 restaurants started in Poughkeepsie, New York around the turn of the twentieth century was collected. Of these businesses, some went out of business within the first three years (I), some went out of business after between three to six years (II), and some went out of after between six to nine years (III). The remainder (IV) survived for more than 9 years. Assume that the restaurants were independently sampled.

(a) If half of startup restaurants were to go out of business every three years, what proportions of them would be categorized into the time periods I, II, III, and IV?

Answer: I 204 would go out of business or $\frac{204}{409} \approx 50\%$

II 102 would go out of business or $\frac{102}{409} \approx 25\%$

III 51 would go out of business or $\frac{51}{409} \approx 12.5\%$

IV 52 would be remaining after 9 years or $\frac{52}{409} \approx 12.5\%$

(b) Suppose that Pearson's test of goodness of fit will be used to evaluate the claim that half of restaurants go out of business every three years, with a significance level of 0.01. (Approximately) what will the chance of a type I error for this test be?

Answer: The chance of a Type I error is ≈ 0.01 since we are using the continous chi-squared distribution in this test.

(c) If I, II, III, and IV were actually equally likely for the restaurants, what would the power of the test in (b) be?

```
crit_val <- qchisq(1-0.01, 4-1,)

n <- 409
pi0 <- c(0.5,0.25,0.125,0.125)
pi1 <- rep(0.25,4)

nonCentParam <- sum((n*pi1-n*pi0)^2 / (n*pi0))

power <- 1-pchisq(crit_val,4-1,nonCentParam)


nonCentParam %>% kbl(col.names = c("NCP"))
```

Answer:

NCP
153.375

```
power %>% kbl(col.names = c("Power"))
```

Power
1

(d) If I, II, and III were equally likely, but IV was actually the most likely possibility (more likely than each of the others), would the power of the test be higher or lower than in (c)? Explain.

```
pi0 <- c(0.5,0.25,0.125,0.125)
pi1 <- c(0.2,0.2,0.2,0.4)

nonCentParam <- sum((n*pi1-n*pi0)^2 / (n*pi0))

power <- 1-pchisq(crit_val,4-1,nonCentParam)


nonCentParam %>% kbl(col.names = c("NCP"))
```

Answer:

NCP
343.56

```
power %>% kbl(col.names = c("Power"))
```

Power
1

The power should be larger than in (c) because the non-centrality parameter is larger in (d). This shifts the Chi-Squared pdf more to the right creating a greater area for the area that we calculate power from. This isn't evident in our result because of the large sample size, but we can tell from the comparative sizes of the non-centrality parameters.

(e) The study found 241 businesses in category I, 69 in category II, and 41 in category III. Evaluate the claim that half of restaurants go out of business every three years. Use a significance level of 0.01, and Pearson's test of goodness of fit.

```
pi0 <- c(0.5,0.25,0.125,0.125)
Xi <- c(241,69,41,58)

chisq.test(Xi,p = pi0,correct = FALSE)$p.value %>% kbl(col.names = c("p value"))
```

Answer:

p value
0.0001502

The resulting Pearson's Goodness of Fit Test provides evidence against the null hypothesis that half of restaurants go out of business every three years. Here we reject the null hypothesis

---

4. A study was carried out to determine whether shirt color has an effect on online dating results. To this end, a group of volunteers tried online dating with six different shirt colors visible in their profile pictures (1=black, 2=white, 3=red, 4=yellow, 5=blue, 6=green). Each volunteer used a profile picture with a randomly chosen color for two weeks; then, they were randomly assigned a different color to use for the next two weeks, and this was repeated for a total of 36 weeks. In each case, the only change to the dating profile was the shirt color in the profile picture. Messages received by shirt color were recorded, and you can find them on Canvas in colorshirt.dat

(a) Is there evidence at significance level 0.01 to conclude that the colors in this study differ in the attention

```
that they attract? Use a Pearson test of goodness of fit.
```

```
raw_counts <- read.table(paste0(getwd(),"/colorshirt.dat"),col.names = c("messages"))

message_freq <- raw_counts %>% table %>% as.data.frame()

colnames(message_freq) <- c("shirt_color","number_of_messages")

pi0 <- rep(1/6,6)

chisq.test(message_freq$number_of_messages, p = pi0, correct = FALSE)$p.value %>% kbl(col.names = c("p v
```

Answer:

p value
9.7e-06

The resulting Pearson's Goodness of Fit Test provides evidence against the null hypothesis that there is no shirt color effect on the number of messages sent. We can reject the null hypothesis and conclude that there is a shirt color effect on the number of messages sent.

(b) Are there any colors that stand out to you as seemingly different than the others? Explain.

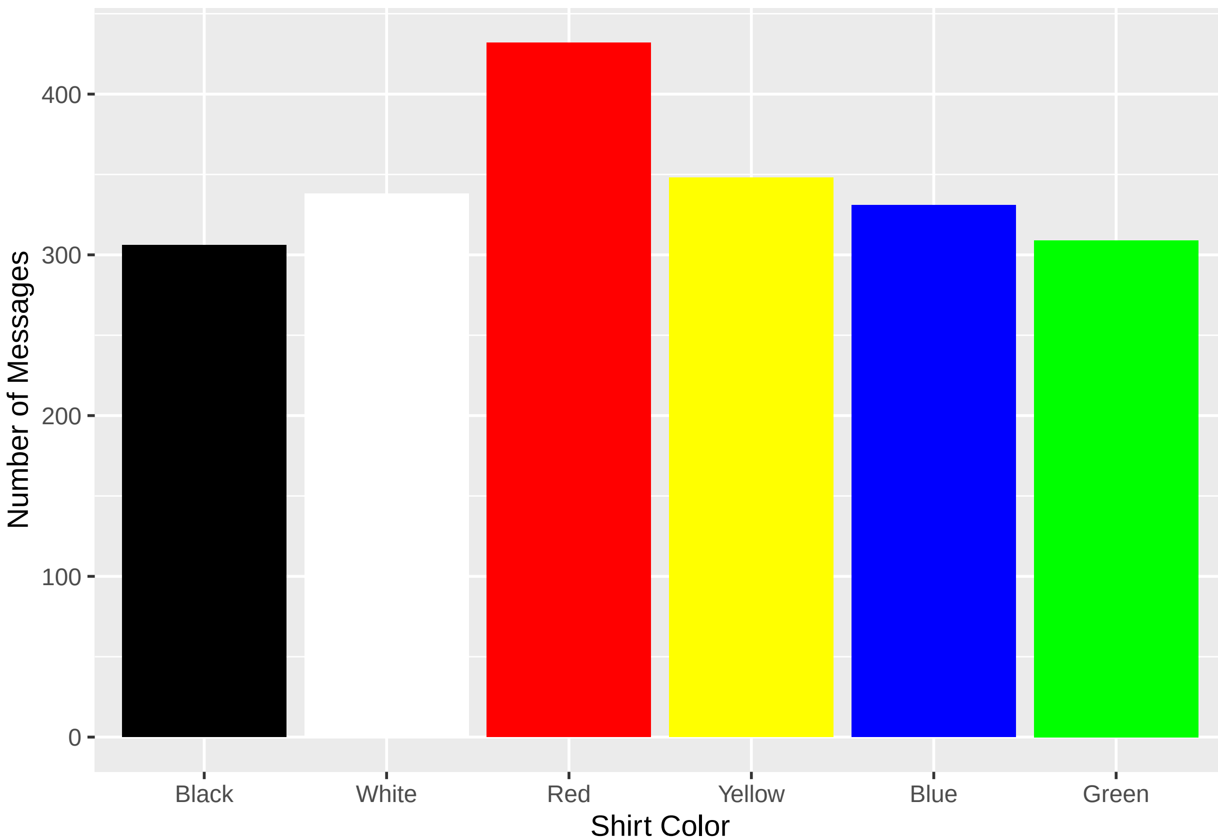
```
library(ggplot2)
```

Answer:

```
## Warning in register(): Can't find generic `scale_type` in package ggplot2 to
## register S3 method.
```

```
ggplot() + geom_histogram(data = raw_counts, aes(x = (messages %>% factor),fill = messages %>% as.facto
  ylab("Number of Messages")+xlab("Shirt Color")+
  labs(fill='Shirt Color') +
  scale_fill_manual(values = c("black","white","red","yellow","blue","green"),
                    labels = c("Black","White","Red","Yellow","Blue","Green"))+
  scale_x_discrete(labels= c("Black","White","Red","Yellow","Blue","Green"))+
  theme(legend.position="none")
```

```
## Warning: Ignoring unknown parameters: binwidth, bins, pad
```



The results of this histogram indicate that red colored shirts receive the most messages by about 80. All of the other colors pretty much receive the same number of messages, but people wearing green shirts receive the least.