Categorical Data Analysis

Report

1.9
After tying the new analgesic, 40 women reported greater relief with the standard analgesic, and 60 reported greater relief with the new one.

a. Test the hypothesis that the probability of greater relief with the standard analgesic is the same as the probability of greater relief with the new analgesic.

   Since we are trying to determine if one probability is greater, we will do a proportion test with our p_o as 0.5 and our complement probability q_o (which is equal to $1-p_o$) as 0.5.

   Prior to performing a proportion test, we must first check to see if our n is large enough. To continue, the product of p_o and n must be greater than 5, and the product of the complement probability q_o must be greater than 5.

   In R, I set pie= p_o=0.5 and n to 100. After, I created checks to find the products mentioned above. Checkone is equal to the product of n and pie, and checktwo is equal to the product of n and q_o (which is 1-pie).

```
> pie=0.5
> n=100
> checkone=n*pie
> checktwo=n*(1-pie)
> print("checkone equals");checkone
[1] "checkone equals"
[1] 50
> print("checktwo equals");checktwo
[1] "checktwo equals"
[1] 50
```

   From the output we know that checkone is 50, which is greater than 5, and checktwo is 50, which is also greater than 5. Because both checks are greater than 5 we can continue to the proportion test.

   In the problem, our p hat is equal to the proportion of women that reported greater relief with the new analgesic. With n=100, our proportion is equal to 60/100=0.6.

   In R, I will first input all of the variables. Notice that Variable names slightly vary from what I have called them above. Although they vary, they are still recognizable as the same name such as p_hat is p hat, and po is p_o which can also be called pie.

```
> p_hat=0.6
> po=0.5
> qo=1-0.5
```

The next step is to calculate the z for which we can test our hypothesis:
H_o: pie= 0.5 and H_a: pie $\neq$ 0.5

To do this we use the equation: (p hat- p_o)/(standard error)

The equation for standard error: $\sqrt{((p_o q_o)/n)}$

The calculated standard error is shown below.

```
> std_error=sqrt( (po*qo)/n)
> print("The null standard error is");std_error
[1] "The null standard error is"
[1] 0.05
```

Next, we calculate z from the standard error calculated above. This calculation is shown below.

```
> z=(p_hat-po)/std_error
> print("The zscore is");z;print("The z of alpha 0.025 is");z_alpha
[1] "The zscore is"
[1] 2
```

Using the equation mentioned above, we were able to calculate the z score.

Next we will compare the z score calculated above to the standardized z for an alpha level of 0.05. Since this is a two tailed test, we split the alpha level in half. We now have an alpha level of 0.025. To find the critical value in which to prove or disprove the null hypothesis, we will use the qnorm function in R. I set z_alpha to qnorm of 0.025 to find the upper tail z.

```
> z_alpha=qnorm(0.025,lower.tail=FALSE)
> print("The z of alpha 0.025 is");z_alpha
[1] "The z of alpha 0.025 is"
[1] 1.959964
```

Since our z score of 2 is greater than our critical value of 1.9599 (it falls into the rejection region), we reject the null hypothesis. There is sufficient evidence to prove that the probability of greater relief with the standard analgesic and not equal to the probability of greater relief with the new analgesic.

b. Construct and interpret a 95% confidence interval for the probability of greater relief with the new analgesic.

For this problem, we want the proportion of women with greater relief with the new analgesic. This will be the same proportion (p hat) from part a: 60/100=0.6. Our alpha level will also be the same from part a because 100%-95%confidence=5%error. With our current alpha level of 0.05, we must still split the alpha because confidence intervals are two sided. So, our resulting alpha level will be 0.025.

For good measures I will recalculate the z of 0.025 in addition to restating the variables that will be reused from part a.

```
> z_alpha=qnorm(0.025,lower.tail=FALSE)
> p_hat=0.6
> n=100
```

The next step is plugging the information into the confidence interval formula for a single proportion. The formula for this is: p hat ± (z of alpha x standard error). The formula for standard error is the similar to part a: √((p hat x q hat)/n). For good measures, I will also recalculate the standard error.

The following is the output of the confidence interval using variables created in R:

```
> z_alpha=qnorm(0.025,lower.tail=FALSE)
> p_hat=0.6
> n=100

> q_hat=1-p_hat
> std_error_two=sqrt((p_hat*q_hat)/n)

> left=p_hat
> right=z_alpha*std_error_two
> lower=p_hat-right
> upper=p_hat+right
> range<-c(lower,upper)
> print("The true population proportion falls between");range
[1] "The true population proportion falls between"
[1] 0.5039818 0.6960182
```

In order to simplify the equation in R, I split the equation into two pieces (left and right sides). This makes it easy to add and subtract the right side from the left. The left side is p hat while the right side is the z of alpha times the standard error. I stored the lower and upper bracket into a vector (range<-c(lower,upper)) so I could print the numbers next to each other.

According to the results, we are 95% confident that the true proportion of women that reported greater relief with the new analgesic falls between 0.5039 and 0.6960.

<u>1.10</u>

Refer to the previous exercise. The researchers worked a sufficiently large sample to be able to estimate the probability of preferring the new analgesic to within 0.08, with a confidence of 0.95. If the true probability is 0.75, how large a sample is needed to achieve this accuracy?

Because the confidence level is .95, our alpha level is 0.05 (1-.95 = 0.05). To compute n for a true probability of 0.75 and a margin of error of 0.08, we use the Margin of error formula:

E = z alpha x standard error

Again, standard error is equal to: √((p hat x q hat)/n).

Since our true probability is 0.75, p hat is equal to 0.75 and p hat's complement, q hat, is equal to 0.25. These probabilities have been labeled as variables in R.

Again, we will need to compute z of 0.025 (0.05/2=0.025 because margin of error considers both tails for this problem). For this, we will use the qnorm function in R. This qnorm function gives us the standard z for the upper tail, and also alleviates the use of a z table.

For the sake of space, I pre-rearranged the Margin of error formula to directly compute n. The rearranged formula that I input into R is:

$$n = (z^2 \, (\text{p hat} \times \text{q hat})) / E^2$$

Notice: Variable names differ slightly in R, but are still easily recognizable. For example q hat is named comp for complement and E is named margin_error.

The output for this calculation is as follows:

```
> z_alpha=qnorm(0.025,lower.tail=FALSE)
> margin_error=0.08
> p_hat=0.75
> comp=1-p_hat

> n=((z_alpha^2*(p_hat*comp))/margin_error^2)
> print("The sample size that we need to achieve is");n
[1] "The sample size that we need to achieve is"
[1] 112.5427
```

According to this calculation, we need a sample size of 112.54, roughly 113, in order to estimate the probability of women preferring the new analgesic to within a 0.08 margin of error with a confidence of 0.95 and a true probability of 0.75.