Inference on Proportions Homework- week 9

1. An investigator has developed a new diagnostic procedure. He would like to test to see if this new procedure is superior to the old procedure which has a sensitivity of 75%. A sample of 100 patients that are known to have a specific disease are evaluated with this new procedure. Of the 100, 86 of them are classified by it as a positive.
a. List the null and alternative hypotheses
b. Is the z-test valid here? Explain your answer
c. Test the hypothesis that you created in part using α =0.05
d. Give the real-world answer

Solution:

a. List the null and alternative hypotheses
N=100
P=X/N

P=86/100=0.86

The null hypothesis is,
$H_0: \pi_0 = 0.75$

The alternative hypothesis is,
$H_A: \pi > 0.75$

First we need to check the assumption that the new procedure is superior to the old procedure..

b. Is the z-test valid here? Explain your answer

Since we have sample size given in the question and enough data is available to calculate null and alternative hypotheses, Z-test is valid in this case.

c. Test the hypothesis that you created in part using α =0.05

$Z_{stat} = (P - \pi 0)/[\sqrt{\pi} 0((1 - \pi 0)/N)]$

$Z_{stat} = (0.86 - 0.75)/[\sqrt{0.75}((1 - 0.75)/100)]$
$Z_{stat} = 0.11/[\sqrt{0.75} * 0.25/100)]$
$Z_{stat} = 0.11/[\sqrt{0.1875/100})]$
$Z_{stat} = 0.11/[\sqrt{0.001875}$
$Z_{stat} = 0.11/0.043$
$Z_{stat} = 2.55$

For p-value approach,
P-value= P(Z≤z)
P-value= P(Z≤2.55)
P-value= 0.0053

For the rejection region approach, since the confidence level is 0.95 and the test is 1-sided ,the critical point becomes,

$Z_{0.957}$ = 1.96

Decision rule based on p-value,
Reject H_0 = p-value ≤ α
Fail to reject H_0 = p-value >α

p-value=0.0053
α=0.05

Since P-value < α i.e (0.0053<0.05), we reject the null hypothesis.

d. Give the real-world answer
we reject the null hypothesis and there is strong evidence to support alternative hypotheses. We can conclude that there is strong evidence to support the assumption that new procedure is superior to old procedure.

2. Assume that you now want to create a confidence interval/bound for number one.
a. Would you create a confidence interval, upper bound or lower bound and why
b. Using =0.05, create what you decided on in part a
c. Interpret the interval
d. Use the interval to test the hypotheses from #1, did you get the same result?

Solution:

a. Would you create a confidence interval, upper bound or lower bound and why

We will be creating a confidence interval (upper bound) since we are rejecting the null hypothesis in favour of the lower tailored alternative. As 86 percent of patients classified as positive, we should calculate the upper bound.

b. Using =0.05, create what you decided on in part a

Since we decided that we will be creating a confidence interval:

We have n=100

X=86

α=0.05

p=X/n= 86/100= 0.86

np=100*0.86 = 86

n(1-p) = 100 (1-0.86) = 100 * 0.14 = 14

Since, both the above values are greater than 10,

π ∈ p±Z_((1-α/2)) √((p(1-p))/n)

$\pi \in 0.86 \pm Z_{0.975} \sqrt{(0.86 (1 - 0.86))/100}$

$\pi \in 0.86 \pm 1.96 \sqrt{0.86 (0.14) /100}$

$\pi \in 0.86 \pm 1.96 * 0.03469$

$\pi \in 0.86 \pm 0.06799$

$\pi \in (0.9279, 0.79201)$

c. Interpret the interval
$\pi \in (0.9279, 0.79201)$
We have the upper bound as 0.9279 and lower bound as 0.79201 and we can say that we are 95% confident that the value will lie between both bounds i.e 0.9279 and 0.79201.

d. Use the interval to test the hypotheses from #1, did you get the same result?

The answer would be the same as in part 1, we reject the null hypothesis and there is strong evidence to support the alternate hypothesis and we can say that the new procedure is superior to the old one.

From part 1 we can say that p-value is less than α and hence we reject the null hypothesis.

3. The following are the results of a study on exposure to a particulate and the larynx cancer in roughly 2000 people that lived in a certain city.

Disease Outcome

Exposure	**No Cancer**	**Larnyx Cancer**
Low	**350**	**680**
High	**320**	**780**

a. This study was run as a case control study. Explain what that means

b. Explain how you think that this study was run

c. Explain why we consider this an observational study

d. Explain why running this as an experiment would be morally wrong

Solution:

a. This study was run as a case control study. Explain what that means

This study can be considered as a case-control study because we are comparing the conditional distribution of Larynx cancer based on exposure to a certain particulate.

Here there are 2 cases where a person among roughly 2000 that lived in that city will have Larynx cancer or not and for case-control studies, it makes sense to treat each column as a binomial sample.

A case-control study always and only provides us with the opposite conditional distributions.

b. Explain how you think that this study was run

I think this study was run by taking some volunteers and exposing them to a certain particulate and experimenting whether those particular people will be affected with Larynx cancer or not.Some of them were exposed low to the particulate whereas others were exposed very high and observation was conducted to see how many people got affected with cancer and how many did not.

c. Explain why we consider this an observational study

An observational study is a type of study where researchers observe which group each subject falls in and who has the outcome of interest.

In our case, the study is considered as an observational study because the researchers made sure that subjects were randomly exposed low and high to a certain particulate and the outcome of interest was how many people got affected with Larnyx cancer and how many did not**.**

d. Explain why running this as an experiment would be morally wrong

Running this experiment is morally wrong because exposing someone to a certain particulate just to have a study whether that person gets affected with a certain type of cancer is cruel. In the name of experiment and study, nobody should put someone else's life in the hands of death.

4. Given the situation in number 3, assume that the analyst would like to run a test to see if the proportion of people that had cancer was higher for the high exposure group than the low exposure group. Use =0.05

a. Give the null and alternative in this situation

b. Should the z-test be used here? Justify your answer

c. Test the hypothesis that you created in part a

d. Draw conclusions to the real-world

Solution:

a. Give the null and alternative in this situation

p1= 780/(780+320)

p1= 0.709

p2= 680/(680+350)

p2 = 0.660

P = (780+680)/(1100+1030) = 0.685

The null hypothesis can be stated as follows:

H_0: $\pi_{low} = \pi_{high}$

The Alternative hypothesis can be stated as follows:

H_A= $\pi_{low} < \pi_{high}$

b. Should the z-test be used here? Justify your answer

Z-test is used to test when the sample size is large and since our given sample size is large along with 2 populations we can use Z test to test the hypothesis which we have stated in the above sub-question.

c. Test the hypothesis that you created in part a

So, Z_0= $p1 - p2 - \triangle 0/\sqrt{}p(1 - p)[(1/n1) + (1/n2)]$

Z_0= $0.709 - 0.660/\sqrt{}0.685(1 - 0.685)[(1/1100) + (1/1030)]$

Z_0=2.4185

P – value = 2 * P (Z > 2.4185)

So, P (Z > 2.4185) = 0.0078 Now, as P - value is less than α = 0.05, our decision is to reject the Null Hypothesis.

So, $Z_{0.025}$ = 1.64.

Now, as the Test Statistic is more extreme than Critical Value, our decision is to reject the Null Hypothesis.

d. Draw conclusions to the real-world

Hence, from the conclusion we can say that we reject the null hypothesis from sufficient evidence, we can conclude, at α = 0.05 level of significance, the proportion of people that had cancer was higher for the high exposure group than the low exposure group.