STAT537: Statistics for Research I: HW#6

Due on Sep. 20, 2016

Dr. Schmidhammer TR 11:10am – 12:25pm

Wenqiang Feng

Contents

Problem 1

Ott Exercises 10.12

Solution. Since the number of successes is 0, hence it is not proper to use normal distribution. We use binomial distribution and the formula for the confidence interval is

$$[0, 1 - (\alpha/2)^{1/n}].$$

Since we need to compute 95% confidence interval, hence $\alpha = .05$. According to the above formula, we get the 95% confidence interval is [0,0.1684335]. □

Problem 2

Ott Exercises 10.14

Solution. (a) **Would it be appropriate to use a normal approximation in conducting a statistical test of the research hypothesis that over half of persons suffering from chronic pain are over 50 years of age?**

From the problem, we get the sample size is 800, and the chronic pain found is 424, hence the probability is $\hat{\pi} = \frac{424}{800} = 0.53$, $n\hat{\pi} = 424 > 5$ and $n(1-\hat{\pi}) = 376 > 5$. Hence, we can claim that the normal distribution is valid.

(b) **Using the data in the survey, is there substantial evidence ($\alpha = 0.05$) that more than half of persons suffering from chronic pain are over 50 years of age?**

Let $H_0 : \pi \leq 0.5$, $H_\alpha : \pi > 0.5$, $\alpha = 0.05$. Since the large sample is valid, we can use z-score test. $\hat{\pi} = 0.53, \pi_0 = 0.5$, hence $\hat{\sigma} = \sqrt{\frac{\pi_0(1-\pi_0)}{n}} = 0.01767767$. Therefore,

$$z = \frac{\hat{\pi} - \pi_0}{\hat{\sigma}} = 1.697056$$

so the p-value is

$$\begin{aligned} p - value &= P(z > \frac{\hat{\pi} - \pi_0}{\hat{\sigma}}) \\ &= P(z > 1.697056) \\ &= 0.04484301 < 0.05. \end{aligned}$$

or

$$z > z_\alpha = 1.645.$$

Reject H_0. Hence we can conclude the data provide significant evidence that more than half of persons suffering from chronic pain are over 50 years of age at a level of significance $\alpha = 0.05$.

(c) **Place a 95% confidence interval on the proportion of persons suffering from chronic pain that are over 50 years of age.**

Since the sample size is large to construct a large sample confidence interval. And the formula is

$$\hat{\pi} \pm z_{\alpha/2}\hat{\sigma}.$$

Hence the 95% confidence interval on the proportion of persons suffering from chronic pain that are over 50 years of age is [0.4953524, 0.5646476].

□

Problem 3

Ott Exercises 10.18

Solution. (a) **Place a 95% confidence interval on the difference $\pi_1 - \pi_2$ in the proportion of customers purchasing lawn mowers with and without the warranty.**

$$\hat{\pi}_1 = \frac{91}{250} = 0.364, \quad \hat{\pi}_2 = \frac{53}{250} = 0.212.$$

And the formula for 95% confidence interval on the difference $\pi_1 - \pi_2$ in the proportion of customers purchasing lawn mowers with and without the warranty is as follows

$$\hat{\pi}_1 - \hat{\pi}_2 \pm z_{\alpha/2}\hat{\sigma},$$

where

$$\hat{\sigma} = \sqrt{\frac{\hat{\pi}_1(1-\hat{\pi}_1)}{n_1} + \frac{\hat{\pi}_2(1-\hat{\pi}_2)}{n_2}} = \sqrt{\frac{0.364(1-0.364)}{250} + \frac{0.212(1-0.212)}{250}} = 0.03992794,$$

and

$$z_{\alpha/2} = z_{0.05/2} = 1.959964.$$

Hence, the 95% confidence interval on the difference $\pi_1 - \pi_2$ is $[0.07374269, 0.23025731]$.

(b) **Test the research hypothesis that offering the warranty will increase the proportion of customers who will purchase a mower. Use $\alpha = 0.01$.**

Let $H_0 : \pi_1 \leq \pi_2$, $H_\alpha : \pi_1 > \pi_2$, $\alpha = 0.01$. Since $91 > 5$ and $53 > 5$, then the large sample is valid for using z-score test. And

$$z = \frac{\hat{\pi}_1 - \hat{\pi}_2}{\hat{\sigma}} = 3.806859$$

so the p-value is

$$\begin{aligned} p - value &= P(z > \frac{\hat{\pi}_1 - \pi_2}{\hat{\sigma}}) \\ &= P(z > 3.806859) \\ &= 7.03716e - 05 < 0.01. \end{aligned}$$

Reject H_0. Hence we can conclude the data provides significant evidence that offering the warranty will increase the proportion of customers who will purchase a mower at a level of significance $\alpha = 0.01$.

(c) **Based on your results from parts (a) and (b) should the dealer offer the warranty?**

- Based on (a) the 95% confidence interval on the difference $\pi_1 - \pi_2$ is $[0.07374269, 0.23025731]$ which do not contains the 0. That is to way π_1 will always lager than π_2. Hence the dealer should offer the warranty.
- Based on (b), Since the data provides significant evidence that offering the warranty will increase the proportion of customers who will purchase a mower at a level of significance $\alpha = 0.01$. Therefore the dealer should offer the warranty.

□

Problem 4

Ott Exercises 10.20

Solution. (a) **For both treatments, place 95% confidence intervals on the proportion of patients who experienced a significant reduction in pain.**

From the problem, we get $n_1 = 1000, n_2 = 1000$ and

$$\hat{\pi}_1 = \frac{560}{1000} = 0.56, \quad \hat{\pi}_2 = \frac{680}{1000} = 0.68.$$

Hence, we get $n_1\hat{\pi}_1 = 560 > 5, n_1(1-\hat{\pi}_1) = 440 > 5, n_2\hat{\pi}_2 = 680 > 5, n_2(1-\hat{\pi}_2) = 320 > 5$. Therefore, the sample size is large enough for using z-score test. And the formula for 95% confidence interval is given by

$$\hat{\pi} \pm z_{\alpha/2}\hat{\sigma},$$

with $\alpha = 0.05$ and

$$\hat{\sigma} = \sqrt{\frac{\hat{\pi}(1-\hat{\pi})}{n}}.$$

- **For Biofeedback:** we get

$$\begin{aligned}\hat{\pi}_1 \pm z_{\alpha/2}\hat{\sigma}_1 &= 0.56 \pm 1.96 * \sqrt{\frac{0.56(1-0.56)}{1000}} \\ &= 0.56 \pm 0.03076582 \\ &= [0.5292342, 0.5907658].\end{aligned}$$

- **NSAID:** Similarly, we get

$$\begin{aligned}\hat{\pi}_2 \pm z_{\alpha/2}\hat{\sigma}_2 &= 0.68 \pm 1.96 * \sqrt{\frac{0.68(1-0.68)}{1000}} \\ &= 0.68 \pm 0.02891196 \\ &= [0.651088, 0.708912].\end{aligned}$$

Hence, the 95% confidence interval are [0.5292342, 0.5907658] and [0.651088, 0.708912] for Biofeedback and NSAID, respectively.

(b) **Is there significant evidence ($\alpha = 0.05$) of a difference in the two treatments relative to the proportions of patients who experienced a significant reduction in pain?**

As we mentioned in part, we get $n_1\pi_1 = 560 > 5, n_1(1-\pi_1) = 440 > 5, n_2\pi_2 = 680 > 5, n_2(1-\pi_2) = 320 > 5$. Therefore, the sample size is large enough for using z-score test.

Let $H_0 : \pi_1 = \pi_2$, $H_\alpha : \pi_1 \neq \pi_2$, $\alpha = 0.05$.

$$z = \frac{\hat{\pi}_1 - \hat{\pi}_2}{\hat{\sigma}} = -5.57086$$

and

$$\hat{\sigma} = \sqrt{\frac{\hat{\pi}_1(1-\hat{\pi}_1)}{n_1} + \frac{\hat{\pi}_2(1-\hat{\pi}_2)}{n_2}} = 0.02154066,$$

so the p-value is

$$\begin{aligned}p - value &= P(z < \frac{\hat{\pi}_1 - \pi_2}{\hat{\sigma}}) \\ &= P(z < -5.57086)\end{aligned}$$

$$= \quad 1.267424e - 08 < 0.05.$$

or

$$|z| = 5.57086 > z_{\alpha/2} = 1.96$$

Reject H_0. Hence we can conclude there is significant evidence ($\alpha = 0.05$) of a difference in the two treatments relative to the proportions of patients who experienced a significant reduction in pain.

(c) **Place a 95% confidence interval on the difference in the two proportions.** The formula for 95% confidence interval on the difference $\pi_1 - \pi_2$ is as follows

$$\hat{\pi}_1 - \hat{\pi}_2 \pm z_{\alpha/2}\hat{\sigma},$$

where

$$\hat{\sigma} = \sqrt{\frac{\hat{\pi}_1(1-\hat{\pi}_1)}{n_1} + \frac{\hat{\pi}_2(1-\hat{\pi}_2)}{n_2}} = .02154066,$$

and

$$z_{\alpha/2} = z_{0.05/2} = 1.959964.$$

Hence, the 95% confidence interval on the difference $\pi_1 - \pi_2$ is $[-0.16221892, -0.07778108]$.

□

Problem 5

Ott Exercises 10.78

Solution. (a) **Give the percentage of rats with one or more tumors for each of the three treatment groups.** The percentage of rats with one or more tumors for each of the three treatment groups is as follows

- **Control:** 10%
- **Low dose:** 14%
- **High dose:** 19%

(b) **Conduct a test of whether there is a significant difference in the proportion of rats having one or more tumors for the three treatment groups with $\alpha = 0.05$.**

Let $H_0 : \pi_1 = \pi_2 = \pi_3$, $H_\alpha : \pi_i \neq \pi_j, i, j = 1, 2, 3, i \neq j$, $\alpha = 0.05$. The Pearson's Chi-squared test result can be found in the following:

```
> chisq.test(ctbl)

Pearson's Chi-squared test

data:  ctbl
X-squared = 3.3119, df = 2, p-value = 0.1909
```

The the p-value is $0.1909 > 0.05$, failed to reject to the H_0 at a level of significance $\alpha = 0.05$. Hence , we may conclude that there is no significant difference in the proportion of rats having one or more tumors for the three treatment groups with $\alpha = 0.05$.

(c) **Does there appear to be a drug-related problem regarding tumors for this drug product? That is, as the dose is increased, does there appear to be an increase in the proportion of rats with tumors?**

No, since there is no significant difference in the proportion of rats having one or more tumors for the three treatment groups with $\alpha = 0.05$.

□

Problem 6

Ott Exercises 10.78

Solution. (a) Since one cell of four cells is less than 5, So the percentage is $1/4 = 25\%$ which is more than 20% of the $E_{i,j}$ less than 5. Hence the χ^2 is not valid. The Fisher Exact Test will be applied to this data set. The Fisher Exact Test is given in the following:

```
> fisher.test(ctbl,alternative="two.sided", conf.level = 0.99)

Fisher's Exact Test for Count Data

data:  ctbl
p-value = 2.029e-15
alternative hypothesis: true odds ratio is not equal to 1
99 percent confidence interval:
 0.0005700799 0.0834071130
sample estimates:
odds ratio
0.01247519
```

From the above information, we can see that the p-value is $2.029e - 15$ which is less than 0.01. Hence reject the H_0. Therefore, there is significant evidence to claim that voting for Judge Thomas is dependent of political party affiliation.

(b) The estimated the difference between the proportion of Republicans voting for Judge Thomas and the proportion of Democrats voting for Judge Thomas at 99% confidence interval is [0.0005700799, 0.0834071130]. And the odds ratio is 0.01247519.

□

Appendix

R code for HW#6

Listing 1: Source code for problem 1

```
rm(list = ls())
# set the path or enverionment
setwd("/home/feng/Dropbox/UTK_Course/Stat537/Excel/CH10")


#install.packages("readxl") # CRAN version
library(readxl)
```

```
#install.packages("moments")
library(moments)
rawdata = read_excel("ex6-44.xls",sheet = 1)
#attach(rawdata)

######################################################
# Ott Exercises 10.12
alpha=0.05
n=20

val = 1-(alpha/2)^(1/n)
val

######################################################
# Ott Exercises 10.14
p = 424/800
p
right = 800*(1-p)
right

n=800
pi0=0.5
sigma = sqrt(pi0*(1-pi0)/n)
sigma

z= (p-pi0)/sigma
z
pvalue=1-pnorm(z)
pvalue


pval <- 0.05
z = qnorm(1 - (pval/2))
z
ci= c(p-z*sigma,p+z*sigma)
ci

######################################################
# Ott Exercises 10.18

pi1=91/250
pi1
pi2=53/250
pi2

n1=250
n2=250
sigma = sqrt(pi1*(1-pi1)/n1+pi2*(1-pi2)/n2)
sigma
pval <- 0.05
z = qnorm(1 - (pval/2))
z
ci= c(pi1-pi2-z*sigma,pi1-pi2+z*sigma)
```

```
ci

z= (pi1-pi2)/sigma
z
pvalue=1-pnorm(z)
pvalue

##################################################
# Ott Exercises 10.20
# a
pval <- 0.05
z = qnorm(1 - (pval/2))
z
pi =0.56
n = 1000
sigma =sqrt(pi*(1-pi)/n)
sigma
z*sigma
ci= c(pi-z*sigma,pi+z*sigma)
ci

pval <- 0.05
z = qnorm(1 - (pval/2))
z
pi =0.68
n = 1000
sigma =sqrt(pi*(1-pi)/n)
sigma
z*sigma
ci= c(pi-z*sigma,pi+z*sigma)
ci

# b

n1=1000
n2=1000
pi1=560/n1
pi1
pi2=680/n2
pi2

sigma = sqrt(pi1*(1-pi1)/n1+pi2*(1-pi2)/n2)
sigma
pval <- 0.05
z = qnorm(1 - (pval/2))
z
ci= c(pi1-pi2-z*sigma,pi1-pi2+z*sigma)
ci

z= (pi1-pi2)/sigma
z
pvalue=pnorm(z)
pvalue
```

```
#####################################################
# Ott Exercises 10.20
group = c("control", "low_dose", "high_dose")
one_or_more = c(10,14,19)
None = c(90,86,81)
data = data.frame(group,one_or_more,None)
data
percentage = one_or_more/(one_or_more+None)*100
percentage

ctbl=cbind(data$one_or_more,data$None)
ctbl
chisq.test(ctbl)


#####################################################
# extra
group = c("Democrat", "Republic")
For = c(11,14)
Against = c(46,2)
data = data.frame(For,Against)
data

ctbl=cbind(data$For,data$Against)
ctbl
#chisq.test(ctbl)
fisher.test(ctbl, conf.level = 0.99)
```