Homework 4

November 12, 2019

[1]:
```
import numpy as np
import pandas as pd
import scipy.stats as st
```

1 Problem 7.2-1

The length of life of brand X light bulbs is assumed to be $N(\mu_x, 784)$. The length of life of brand Y light bulbs is assumed to be $N(\mu_y, 627)$ and independent of X. If a random sample of $n = 56$ brand X light bulbs yielded a mean of $\bar{x} = 937.4$ hours and a random sample of size $m = 57$ brand Y light bulbs yielded a mean of $\bar{y} = 988.9$ hours, find a 90% confidence interval for $\mu_x - \mu_y$.

[2]:
```
#x_1 is x bar one, x_2 is x bar two, z is the z-score
#sigma_1 is variance of sample one, sigma_2 is variance of sample two
#n is the amount of values in sample one, and m for sample two
def conf_end_points(x_1,x_2,z,sigma_1,sigma_2,n,m):
    solution = [0,0]
    solution[0]=(x_1-x_2)-z*(np.sqrt(sigma_1/n+sigma_2/m))
    solution[1]=(x_1-x_2)+z*(np.sqrt(sigma_1/n+sigma_2/m))
    return solution

x_var = 784
x_bar = 937.4
x_n = 56

y_var = 627
y_bar = 988.9
y_n = 57
conf_interval = 0.9
sigma = np.round((1-conf_interval),1)/2 #need to round
z_value_right = (st.norm.ppf((1-sigma))).round(3)

answer = conf_end_points(x_bar,y_bar,z_value_right,x_var,y_var,x_n,y_n)

print("The endpoints are",answer)
```

```
The endpoints are [-59.725, -43.275]
```

2 Problem 7.2-2

Let $X_1, X_2, ..X_5$ be a random sample of SAT mathematics scores, assumed to be $N(\mu_x, \sigma^2)$ and let $Y_1, Y_2, .., Y_8$ be an independent random sample of SAT verbal scores, assumed to be $N(\mu_y, \sigma^2)$. If the following data are observed, find a 90% confidence interval for $\mu_x - \mu_y$.

[3]:
```
Problem7_2_4data = pd.read_csv('E7_2-02.txt',sep="\t",header=None)
Problem7_2_4data.rename(columns={0:"X",1:"Y"},inplace=True)
print(Problem7_2_4data)
```

```
       X    Y
0  644.0  623
1  493.0  472
2  532.0  492
3  462.0  661
4  565.0  540
5    NaN  502
6    NaN  549
7    NaN  518
```

[4]:
```
x_bar = Problem7_2_4data['X'].mean()

#Need to exclude NaN values
x_n = 0
for a in Problem7_2_4data['X']:
    if ~np.isnan(a):
        x_n += 1

y_bar = Problem7_2_4data['Y'].mean()
y_n = len(Problem7_2_4data['Y'])

#Getting standard deviations
temp = 0
for a in Problem7_2_4data['Y']:
    temp += np.square(a-y_bar)
std_y = np.sqrt(temp/(y_n-1)).round(4)

temp = 0
for a in Problem7_2_4data['X']:
    if ~np.isnan(a):
        temp += np.square(a-x_bar)
std_x = np.sqrt(temp/(x_n-1)).round(4)

#Getting t-value of 11 degrees of freedom
deg_freedom = x_n + y_n - 2
alpha = np.round(((1-0.9)/2),2)
t_value_right = st.t.ppf(1-alpha,deg_freedom).round(3)
```

```
#function to find the interval
def conf_end_points2(x_1,x_2,t,sigma_1,sigma_2,n,m):
    solution = [0,0]
    common_sd = np.sqrt(((n-1)*np.square(sigma_1)+(m-1)*np.square(sigma_2))/
 ↪(n+m-2))
    solution[0]=((x_1-x_2)-t*common_sd*np.sqrt(1/n+1/m)).round(3)
    solution[1]=((x_1-x_2)+t*common_sd*np.sqrt(1/n+1/m)).round(3)
    return solution

answer = conf_end_points2(x_bar,y_bar,t_value_right,std_x,std_y,x_n,y_n)
print("The interval of solution is",answer)
```

```
The interval of solution is [-74.517, 63.667]
```

3 Problem 7.2-3

Independent random samples of the heights of adult males living in two countries yielded the following results; $n = 12, \bar{x} = 65.7$ inches, $s_x = 4$ inches and $m = 15, \bar{y} = 68.2$ inches, $s_y = 3$ inches. Find an approximate 98% confidence interval for the difference $\mu_x - \mu_y$ of the means of the populations of heights. Assume that $\sigma_x^2 = \sigma_y^2$

[5]:
```
ci = .98 #conf interveral
x_bar = 65.7 #mean of X
n = 12 #number of elements for X
x_std = 4 #standard deviation of x

y_bar = 68.2 #mean of Y
m = 15 #number of elements for Y
y_std = 3 #standard deviation of y

df = m + n - 2 #degrees of freedom
alpha = 1-(1-.98)/2

t_value_right = st.t.ppf(alpha,df).round(3) #t-value

answer = conf_end_points2(x_bar,y_bar,t_value_right,x_std,y_std,n,m)

print("The end points for the confidence interval are",answer)
```

```
The end points for the confidence interval are [-5.845, 0.845]
```

4 Problem 7.3-1

A machine shop manufactures toggle levers. A lever is flawed if a standard nut cannot be screwed onto the threads. Let p equal the proportion of flawed toggle levers that the shop manufactures. If there were 24 flawed levers out of a sample of 642 that were selected randomly from the production line,

4.1 Part a

Give a point estimate of p.

$\hat{p} = \frac{y}{n} = \frac{24}{642} \approx 3.74\%$

4.2 Part b

Using equation 7.3-2 to find an approximate 95% confidence interval for p.

Equation 7.3-2: $\frac{y}{n} \pm z_{\alpha/2}\sqrt{\frac{(y/n)(1-y/n)}{n}}$

$\alpha = 0.05$ thus $z_{\alpha/2} = z_{0.025}$

$0.0374 \pm 1.96 \times \sqrt{\frac{(0.0374)(1-0.0374)}{642}}$

The two solutions are $(0.0227, 0.0521)$

4.3 Part c

Use equation 7.3-4 to find an approximate 90% confidence interval for p.

Equation 7.3-4: $\dfrac{\hat{p} + z_{\alpha/2}^2/(2n) \pm z_{\alpha/2}\sqrt{\frac{\hat{p}(1-\hat{p})}{n} + \frac{z_{\alpha/2}^2}{4n^2}}}{1 + \frac{z_{\alpha/2}^2}{n}}$

$$\frac{0.0374 + 1.96^2/(2 \times 642) \pm 1.96\sqrt{\frac{0.0374(1-0.0374)}{642} + \frac{1.96^2}{4 \times 642^2}}}{1 + \frac{1.96^2}{642}}$$

The two solutions are $(0.0253, 0.055)$

4.4 Part d

Use equation 7.3-5 to find an approximate 90% confidence interval for p.

Equation 7.3-5: $\tilde{p} \pm z_{\alpha/2}\sqrt{\frac{\tilde{p}(1-\tilde{p})}{n+4}}$

$\tilde{p} = \frac{y+2}{n+4} = 0.0402$

$z_{\alpha/2} = 1.96$

$0.0402 \pm 1.96\sqrt{\frac{0.0402(1-0.0402)}{642+4}}$

The two solutions are $(0.0251, 0.055)$.

4.5 Part e

Find a one-sided 90% confidence interval for p that provides a lower bound for p.

$\alpha = 0.05$, therefore $z_\alpha = 1.645$

Upper bound of confidence interval is

$0.0374 + 1.645\sqrt{\frac{0.0374(1-0.0374}{642}}$

Therefore, the solution is (0,0.0497)

5 Problem 7.3-2

Let p equal the proportion of letters mailed in the Netherlands that are delivered the next day. Suppose that $y = 142$ out of a random sample of $n = 200$ letters were delivered the day after they were mailed.

5.1 Part a

Give a point estimate of p.

$\hat{p} = \frac{y}{n} = \frac{142}{200} = 71\%$

5.2 Part b

Use equation 7.3-2 to find an approximate 90% confidence interval for p.

Equation 7.3-2: $\frac{y}{n} \pm z_{\alpha/2}\sqrt{\frac{(y/n)(1-y/n)}{n}}$

$\alpha = 0.10$ thus $z_{\alpha/2} = z_{0.05} = 1.645$

$0.71 \pm 1.645 \times \sqrt{\frac{(0.71)(1-0.71)}{200}}$

The two solutions are $(0.6572, 0.7628)$

5.3 Part c

Use equation 7.3-4 to find an approximate 90% confidence interval for p.

Equation 7.3-4: $\dfrac{\hat{p} + z^2_{\alpha/2}/(2n) \pm z_{\alpha/2}\sqrt{\frac{\hat{p}(1-\hat{p})}{n} + \frac{z^2_{\alpha/2}}{4n^2}}}{1 + \frac{z^2_{\alpha/2}}{n}}$

$$\frac{0.71 + 1.645^2/(2 \times 200) \pm 1.645\sqrt{\frac{0.71(1-0.71)}{200} + \frac{1.645^2}{4 \times 200^2}}}{1 + \frac{1.645^2}{200}}$$

The two solutions are $(0.6545, 0.7599)$

5.4 Part d

Use equation 7.3-5 to find an approximate 90% confidence interval for p.

Equation 7.3-5: $\tilde{p} \pm z_{\alpha/2}\sqrt{\frac{\tilde{p}(1-\tilde{p})}{n+4}}$

$\tilde{p} = \frac{y+2}{n+4} = 0.7059$

$0.7059 \pm 1.645\sqrt{\frac{0.7059(1-0.7059)}{200+4}}$

The two solutions are $(0.6534, 0.7584)$.

5.5 Part e

Find a one-sided 90% confidence interval for p that provides a lower bound for p.

$\alpha = 0.10$, therefore $z_\alpha = 1.28$

Upper bound of confidence interval is

$0.71 + 1.28\sqrt{\frac{0.71(1-0.71}{200}}$

Therefore, the solution is (0.6689,1)

6 Problem 7.3-8

A proportion, p, that many public opinion polls estimate is the number of Americans who would say yes to a certain question. In one such random sample of 1022 adults, 388 said yes.

6.1 Part a

On the basis of the given data, find a point estimate of p.

$\hat{p} = \frac{y}{n} = \frac{388}{1022} \approx 37.96\%$

6.2 Part b

Find an approximate 90% confidence interval for p.

Equation 7.3-2: $\frac{y}{n} \pm z_{\alpha/2}\sqrt{\frac{(y/n)(1-y/n)}{n}}$

$\alpha = 0.10$ thus $z_{\alpha/2} = z_{0.05} = 1.645$

$0.3796 \pm 1.645 \times \sqrt{\frac{(0.3796)(1-0.3796)}{1022}}$

The two solutions are $(0.3546, 0.4046)$

7 Problem 7.4-3

A company packages powdered soap in "6-pound" boxes. The sample mean and standard deviation of the soap in tese boxes are currently 6.09 pounds and 0.02 pound, respectively. If the mean fill can be lowered by 0.01 pound, $14,000 would be saved per year. Adjustments were made in the filling equipment, but it can be assumed that the standard deviation remains unchanged.

7.1 Part a

How large a sample is needed so that the maximum error of the estimate of the new μ is $\$\varepsilon\$ = 0.001 with 90\% confidence$?

$\bar{x} = 6.09$ and $s_x = 0.02$

$\alpha = .10$ therefore, $z_{\alpha/2} = z_{0.10/2} = 1.645$

Formula for sample size: $n = (\frac{z_{\alpha/2}\sigma}{\varepsilon})^2$

$n = (\frac{1.645 \times 0.02}{0.001})^2 = 1083$

7.2 Part b

A random sample of size $n = 1219$ yielded $\bar{x} = 6.048$ and $s = 0.022$. Calculate a 90% confidence interval for μ.

The confidence interval is
$\bar{x} \pm z_{\alpha/2} \frac{\sigma}{\sqrt{n}}$
$6.048 \pm 1.645 \frac{0.022}{\sqrt{1219}}$
Therefore the solution is $(6.047, 6.049)$

7.3 Part c

Estimate the savings per year with these new adjustments.

Difference between the means is 6.09 - 6.048 = 0.042
$\frac{0.042}{.01} = 4.2$
Because \$14,000 would be saved per year by lowering 0.01 pounds, then
$\$4.2 \times \$14,000 = \$58,800\$ would be saved$

7.4 Part d

Estimate the proportion of boxes that will now weight less than 6 pounds

$z = \frac{x - \mu}{\sigma} = \frac{6 - 6.048}{0.022} = -2.18$
$P(X < 6) = P(Z < -2.18) = 1 - P(Z < 2.18)$
$= 1 - 0.9854 = 0.0146$

8 Problem 7.4-4

Measurements of the length in centimeters of $n = 29$ fish yielded an average length of $\bar{x} = 176.82$ and $s^2 = 34.9$. Determine the size of a new sample so that $\bar{x} \pm 0.5$ is an approximate 95% confidence interval for μ.

Formula for sample size: $n = (\frac{z_{\alpha/2}\sigma}{\varepsilon})^2$
$\alpha = .05$ therefore, $z_{\alpha/2} = 1.96$
$n = \lceil (\frac{1.96 \times \sqrt{34.9}}{0.5})^2 \rceil = 537$

9 Problem 7.4-10

A seed distributor claims that 80% of its beet seeds will germinate. How many seeds must be tested for germination in order to estimate p, the true proportion that will germinate, so that the maximum error of the estimate is $\varepsilon = 0.03$ with 90% confidence?

Formula for sample size: $n = \frac{(z_{\alpha/2})^2 \hat{p}(1 - \hat{p})}{\varepsilon^2}$
$\alpha = .1$ therefore, $z_{\alpha/2} = 1.645$
$n = \frac{1.645^2 \times 0.8(0.2)}{0.03^2} = 482$

10 Problem 7.4-11

Some dentists were interested in studying the fusion of embryonic rat palates by a standard transplantation technique. When no treatment is used, the probability of fusion equals approximately 0.89. The dentists would like to estimate p, the probability of fusion, when a vitamin A is lacking.

10.1 Part a

How large a sample n of rat embryos is needed for $y/n \pm 0.10$ to be a 95% confidence interval for p?

$\hat{p} = 0.89$ and $\alpha = 0.05$

Therefore, $z_{\alpha/2} = 1.96$

Formula for sample size: $n = \dfrac{(z_{\alpha/2})^2 \hat{p}(1-\hat{p})}{\varepsilon^2}$

$n = \lceil \dfrac{1.96^2 \times 0.89(0.11)}{0.10^2} \rceil = 38$

10.2 Part b

If $y = 44$ out of $n = 60$ palates showed fusion, give a 95% confidence interval for p.

$\hat{p} = \dfrac{y}{n} \approx 0.733$

Confidence boundaries are then,

$0.733 \pm 1.96\sqrt{\dfrac{0.733(1-0.733)}{60}}$

Therefore the solution is (0.6214,0.8452)