Homework 5

October 20, 2019

Math A7800: Homework 5
Name: Jorge Monzon Diaz
Email: jorgemd503@gmail.com

0.1 Dependencies

[1]:
```
import numpy as np
import matplotlib.pyplot as plt
from matplotlib.patches import Ellipse, Rectangle
import pandas as pd
from scipy.stats import t
```

1 Exercise 5.6

Verify the Bonferroni inequality in (5-28) for m=3.

$$P[\text{all } C_i \text{ true}] = 1 - P[\text{at least one } C_i \text{ false}] \geq 1 - \sum_{i=1}^{m} P(C_i \text{ false}) = 1 - \sum_{i=1}^{m} (1 - P(C_i \text{ true}))$$

$$= 1 - (\alpha_1 + \alpha_2 + ... + \alpha_m)$$

$$P[C_i] = 1 - \alpha_i$$

$$P[C_1 \cap C_2 \cap C_3) \geq [1 - P[\overline{C_1}] - P[\overline{C_2}] - P[\overline{C_3}]]$$

$$P[C_1 \cap C_2 \cap C_3) \geq 1 - \alpha_1 - \alpha_2 - \alpha_3$$

2 Excercise 5.7

Use the sweat data in Table 5.1 (See Example 5.2). Find simultaneous 95% T^2 confidence intervals for μ_1, μ_2, and μ_3 using Result 5.3. Construct the 95% Bonferroni intervals using (5-29). Compare the two sets of intervals.

[2]:
```
x = pd.read_excel (r'hw5.xlsx', sheet_name='data 57')

#print means
print("\nx-mean:")
print(x.mean(axis = 0))
```

```
x-mean:
x1      4.640
x2     45.400
x3      9.965
dtype: float64
```

2.1 Simultaneous 95% T^2 confidence intervals for μ_1, μ_2, and μ_3 using Result 5.3:

$$\bar{x}_i - \sqrt{\frac{p(n-1)}{n(n-p)}F_{p,n-p}(\alpha)}\sqrt{s_{ii}} \leq \mu_i \leq \bar{x}_i + \sqrt{\frac{p(n-1)}{n(n-p)}F_{p,n-p}(\alpha)}\sqrt{s_{ii}}$$

$$\bar{x}_i - \sqrt{\frac{3(20-1)}{20(20-3)}F_{3,20-3}(0.05)}\sqrt{s_{ii}} \leq \mu_i \leq \bar{x}_i + \sqrt{\frac{3(20-1)}{20(20-3)}F_{3,20-3}(0.05)}\sqrt{s_{ii}}$$

$$\bar{x}_i - 0.732442\sqrt{s_{ii}} \leq \mu_i \leq \bar{x}_i + 0.732442\sqrt{s_{ii}}$$

$$s_{ii} = \frac{1}{n-1}\sum_{i=1}^{n}(x_i - \bar{x})$$

[3]:
```
print("s_(11):",np.var(x["x1"],ddof=1))
print("s_(22):",np.var(x["x2"],ddof=1))
print("s_(33):",np.var(x["x3"],ddof=1))
```

```
s_(11): 2.8793684210526322
s_(22): 199.78842105263155
s_(33): 3.627657894736842
```

$$4.640 - 0.732442\sqrt{2.8794} \leq \mu_1 \leq 4.640 + 0.732442\sqrt{2.8794}$$

$$3.3971 \leq \mu_1 \leq 5.8829$$

$$45.400 - 0.732442\sqrt{199.7884} \leq \mu_2 \leq 45.400 + 0.732442\sqrt{199.7884}$$

$$35.0472 \leq \mu_2 \leq 55.7528$$

$$9.965 - 0.732442\sqrt{3.6277} \leq \mu_3 \leq 9.965 + 0.732442\sqrt{3.6277}$$

$$8.5700 \leq \mu_3 \leq 11.3600$$

2.2 95% Bonferroni intervals:

$$\bar{x}_i - t_{n-1}(\frac{\alpha}{2p})\sqrt{\frac{s_{ii}}{n}} \leq \mu_i \leq \bar{x}_i + t_{n-1}(\frac{\alpha}{2p})\sqrt{\frac{s_{ii}}{n}}$$

$$4.640 - t_{19}(\frac{0.05}{6})\sqrt{\frac{2.8794}{20}} \leq \mu_1 \leq 4.640 + t_{19}(\frac{0.05}{6})\sqrt{\frac{2.8794}{20}}$$

$$3.6440 \leq \mu_1 \leq 5.6360$$

$$45.400 - t_{19}(\frac{0.05}{6})\sqrt{\frac{199.7884}{20}} \leq \mu_2 \leq 45.400 + t_{19}(\frac{0.05}{6})\sqrt{\frac{199.7884}{20}}$$

$$37.1031 \leq \mu_2 \leq 53.6969$$

$$9.965 - t_{19}(\frac{0.05}{6})\sqrt{\frac{3.6277}{20}} \leq \mu_3 \leq 9.965 + t_{19}(\frac{0.05}{6})\sqrt{\frac{3.6277}{20}}$$

$$8.8470 \leq \mu_3 \leq 11.0830$$

The intervals are similar to one another using both methods, but the Bonferonni have tighter bounds.

3 Exercise 5.10

Refer to the bear growth data in Example 1.10 (see Table 1.4). Restrict your attention to the measurements of length.

[4]:
```
x = pd.read_excel (r'hw5.xlsx', sheet_name='data 510')
x['Length 2-3'] = x['Length 3']-x['Length 2']
x['Length 3-4'] = x['Length 4']-x['Length 3']
x['Length 4-5'] = x['Length 5']-x['Length 4']
print(x)

#print means
print("\nx-mean:")
print(x.mean(axis = 0))

#print var
print("\nVar:\ns_(22):",np.var(x["Length 2"],ddof=1))
print("s_(33):",np.var(x["Length 3"],ddof=1))
print("s_(44):",np.var(x["Length 4"],ddof=1))
print("s_(55):",np.var(x["Length 5"],ddof=1))
print("s_((2-3)(2-3)):",np.var(x["Length 2-3"],ddof=1))
print("s_((3-4)(3-4)):",np.var(x["Length 3-4"],ddof=1))
print("s_((4-5)(4-5)):",np.var(x["Length 4-5"],ddof=1))
```

```
   Length 2  Length 3  Length 4  Length 5  Length 2-3  Length 3-4  Length 4-5
0       141       157       168       183          16          11          15
1       140       168       174       170          28           6          -4
2       145       162       172       177          17          10           5
3       146       159       176       171          13          17          -5
4       150       158       168       175           8          10           7
5       142       140       178       189          -2          38          11
6       139       171       176       175          32           5          -1

x-mean:
Length 2      143.285714
Length 3      159.285714
Length 4      173.142857
Length 5      177.142857
Length 2-3     16.000000
Length 3-4     13.857143
Length 4-5      4.000000
dtype: float64

Var:
s_(22): 15.238095238095239
s_(33): 99.90476190476188
s_(44): 15.809523809523808
s_(55): 45.476190476190474
```

```
s_((2-3)(2-3)): 133.0
s_((3-4)(3-4)): 128.47619047619048
s_((4-5)(4-5)): 58.333333333333336
```

3.1 (a) Obtain the 95% T^2 simultaneous confidence intervals for the four population means for length

$$\bar{x}_i - \sqrt{\frac{p(n-1)}{n(n-p)}F_{p,n-p}(\alpha)}\sqrt{s_{ii}} \leq \mu_i \leq \bar{x}_i + \sqrt{\frac{p(n-1)}{n(n-p)}F_{p,n-p}(\alpha)}\sqrt{s_{ii}}$$

$$\bar{x}_i - \sqrt{\frac{4(7-1)}{7(7-4)}F_{4,7-4}(0.05)}\sqrt{s_{ii}} \leq \mu_i \leq \bar{x}_i + \sqrt{\frac{4(7-1)}{7(7-4)}F_{4,7-4}(0.05)}\sqrt{s_{ii}}$$

$$143.2857 - 3.2280\sqrt{15.2381} \leq \mu_2 \leq 143.2857 + 3.2280\sqrt{15.2381}$$

$$130.6849 \leq \mu_2 \leq 155.8865$$

$$159.2857 - 3.2280\sqrt{99.9048} \leq \mu_3 \leq 159.2857 + 3.2280\sqrt{99.9048}$$

$$127.0229 \leq \mu_3 \leq 191.5521$$

$$173.1429 - 3.2280\sqrt{15.8095} \leq \mu_4 \leq 173.1429 + 3.2280\sqrt{15.8095}$$

$$160.3080 \leq \mu_4 \leq 185.9778$$

$$177.1429 - 3.2280\sqrt{45.4762} \leq \mu_5 \leq 177.1429 + 3.2280\sqrt{45.4762}$$

$$155.3745 \leq \mu_5 \leq 198.9113$$

3.2 (b) Refer to Part a. Obtain the 95% T^2 simultaneous confidence intervals for the three successive yearly increases in mean length

$$\bar{x}_i - \sqrt{\frac{4(7-1)}{7(7-4)}F_{4,7-4}(0.05)}\sqrt{s_{ii}} \leq \mu_i \leq \bar{x}_i + \sqrt{\frac{4(7-1)}{7(7-4)}F_{4,7-4}(0.05)}\sqrt{s_{ii}}$$

$$16.0000 - 3.2280\sqrt{133.0000} \leq \mu_{2-3} \leq 16.0000 + 3.2280\sqrt{133.0000}$$

$$-21.2271 \leq \mu_{2-3} \leq 53.2271$$

$$13.8571 - 3.2280\sqrt{128.4762} \leq \mu_{3-4} \leq 13.8571 + 3.2280\sqrt{128.4762}$$

$$-22.7314 \leq \mu_{3-4} \leq 50.4456$$

$$4.0000 - 3.2280\sqrt{58.3333} \leq \mu_{4-5} \leq 4.0000 + 3.2280\sqrt{58.3333}$$

$$-20.6543 \leq \mu_{4-5} \leq 28.6543$$

3.3 (d) Refer to parts a and b. Construct the 95% Bonferroni confidence intervals for the set consisting of four mean lengths and three successive yearly increases in mean length

Since we are finding CI's for a set of size 7, $m = p = 7$

$$\bar{x}_i - t_{n-1}(\frac{\alpha}{2p})\sqrt{\frac{s_{ii}}{n}} \leq \mu_i \leq \bar{x}_i + t_{n-1}(\frac{\alpha}{2p})\sqrt{\frac{s_{ii}}{n}}$$

$$143.2857 - t_6(\frac{0.05}{14})\sqrt{\frac{15.2381}{7}} \leq \mu_2 \leq 143.2857 + t_6(\frac{0.05}{14})\sqrt{\frac{15.2381}{7}}$$

$$137.3884 \leq \mu_2 \leq 149.1831$$

$$159.2857 - t_6(\frac{0.05}{14})\sqrt{\frac{99.9048}{7}} \leq \mu_3 \leq 159.2857 + t_6(\frac{0.05}{14})\sqrt{\frac{99.9048}{7}}$$

$$144.1854 \leq \mu_3 \leq 174.3860$$

$$173.1429 - t_6(\frac{0.05}{14})\sqrt{\frac{15.8095}{7}} \leq \mu_4 \leq 173.1429 + t_6(\frac{0.05}{14})\sqrt{\frac{15.8095}{7}}$$

$$167.1359 \leq \mu_4 \leq 179.1498$$

$$177.1429 - t_6(\frac{0.05}{14})\sqrt{\frac{45.4762}{7}} \leq \mu_5 \leq 177.1429 + t_6(\frac{0.05}{14})\sqrt{\frac{45.4762}{7}}$$

$$166.9550 \leq \mu_5 \leq 187.3307$$

$$16.0000 - t_6(\frac{0.05}{14})\sqrt{\frac{133.0000}{7}} \leq \mu_{2-3} \leq 16.0000 + t_6(\frac{0.05}{14})\sqrt{\frac{133.0000}{7}}$$

$$-1.4228 \leq \mu_{2-3} \leq 33.4228$$

$$13.8571 - t_6(\frac{0.05}{14})\sqrt{\frac{128.4762}{7}} \leq \mu_{3-4} \leq 13.8571 + t_6(\frac{0.05}{14})\sqrt{\frac{128.4762}{7}}$$

$$-3.2668 \leq \mu_{3-4} \leq 30.9811$$

$$4.0000 - t_6(\frac{0.05}{14})\sqrt{\frac{58.3333}{7}} \leq \mu_{4-5} \leq 4.0000 + t_6(\frac{0.05}{14})\sqrt{\frac{58.3333}{7}}$$

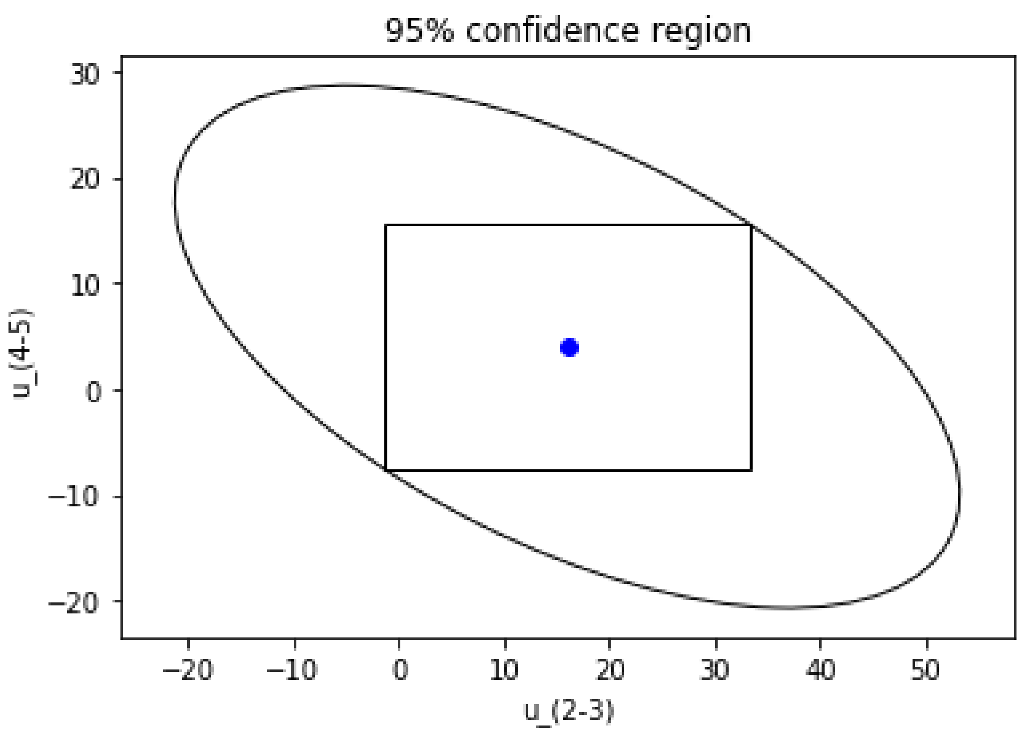
$$-7.5385 \leq \mu_{4-5} \leq 15.5385$$

3.4 (e) Refer to parts c and d. Compare the 95% Bonferroni confidence rectangle for the mean increase in length from 2 to 3 years and the mean increase in length from 4 to 5 years with the confidence ellipse produced by the T^2 procedure

[5]:
```
#create ellipse
ell=Ellipse((16,4), 2*40.5553, 2*18.6952, np.degrees(np.arctan2(149, -298.
↪4)),fill=False)
#plot center
plt.plot(16, 4, marker='o', color="blue")

#create rectangle
#recenter origin, 2*t_{n-1}(a/(2p)) width/height
rec=Rectangle((16-17.4228,4-11.5385), 2*17.4228, 2*11.5385,fill=False)

#create plot
plt.title("95% confidence region")
plt.xlabel("u_(2-3)")
plt.ylabel("u_(4-5)")
ax = plt.gca()
ax.add_patch(ell)
ax.add_patch(rec)
ax.autoscale()
plt.show()
print(ell)
print(rec)
```



```
Ellipse(xy=(16, 4), width=81.1106, height=37.3904, angle=153.4656786561316)
Rectangle(xy=(-1.4228, -7.5385), width=34.8456, height=23.077, angle=0)
```

The Bonferroni confidence rectangle fits exactly within the confidence region produced by the T^2 procedure.

4 Exercise 5.11(b)

A physical anthropologist performed a mineral analysis of nince ancient Peruvian hairs. The results for the chromium (x_1) and strontium (x_2) levels, in parts per million (ppm), were as follows:

[6]:
```
x = pd.read_excel (r'hw5.xlsx', sheet_name='data 511')
print(x)

#print means
print("\nx-mean:")
print(x.mean(axis = 0))

#print var
print("\nVar:\ns_(11):",np.var(x["x1(Cr)"],ddof=1))
print("s_(22):",np.var(x["x2(St)"],ddof=1))
```

```
   x1(Cr)  x2(St)
0    0.48   12.57
1   40.53   73.68
2    2.19   11.13
3    0.55   20.03
4    0.74   20.29
5    0.66    0.78
6    0.93    4.64
7    0.37    0.43
8    0.22    1.08

x-mean:
x1(Cr)     5.185556
x2(St)    16.070000
dtype: float64

Var:
s_(11): 176.0041777777778
s_(22): 527.8493000000001
```

It is known that low levels (less than or equal to .100 ppm) of chromium suggest the presence of diabetes, while strontium is an indication of animal protein intake.

Obtain the individual simultaneous 90% confidence intervals for μ_1 and μ_2 by "projecting" the ellipse constructed in Part a on each coordinate axis. (Alternatively we could use Result 5.3). Does it appear as if this Peruvian culture has a mean strontium level of 10? That is, are any of the points (μ_1 arbitrary, 10) in the confidence regions? Is $[.30, 10]^T$ a plausible value for $\boldsymbol{\mu}$?

$$\bar{x}_i - \sqrt{\frac{p(n-1)}{n(n-p)}F_{p,n-p}(\alpha)}\sqrt{s_{ii}} \leq \mu_i \leq \bar{x}_i + \sqrt{\frac{p(n-1)}{n(n-p)}F_{p,n-p}(\alpha)}\sqrt{s_{ii}}$$

$$\bar{x}_i - \sqrt{\frac{2(9-1)}{9(9-2)}F_{2,9-2}(0.10)}\sqrt{s_{ii}} \leq \mu_i \leq \bar{x}_i + \sqrt{\frac{2(9-1)}{9(9-2)}F_{2,9-2}(0.10)}\sqrt{s_{ii}}$$

$$\bar{x}_i - \sqrt{0.8278}\sqrt{s_{ii}} \leq \mu_i \leq \bar{x}_i + \sqrt{0.8278}\sqrt{s_{ii}}$$

$$5.185556 - 0.9098\sqrt{176.0042} \leq \mu_1 \leq 5.185556 + 0.9098\sqrt{176.0042}$$

$$-6.8844 \leq \mu_1 \leq 17.2556$$

$$16.07 - 0.9098\sqrt{527.8493} \leq \mu_2 \leq 16.07 + 0.9098\sqrt{527.8493}$$

$$-4.8326 \leq \mu_2 \leq 36.9726$$

From these intervals, 10 is indeed a plausible mean value for strontium and $[.30, 10]^T$ is a plausible value for $\boldsymbol{\mu}$.

5 Exercise 5.14

Create a table similar to Table 5.4 using the entries (length of one-at-a-time t-interval)/(length of Bonferroni t-interval)

$$\frac{\text{Length of one-at-a-time t-interval}}{\text{Length of Bonferroni t-interval}} = \frac{t_{n-1}(\frac{\alpha}{2})}{t_{n-1}(\frac{\alpha}{2m})}$$

[7]:
```python
m=[2,4,10]
n=[15,25,50,100,10000000000000]

#compute tabular values
num=[t.ppf(0.05/2, i-1)/t.ppf(0.05/(2*k), i-1) for i in n for k in m]
num=np.array(num).reshape((5,3))

#put into table format
table = pd.DataFrame(num,columns=['2','4','10'],␣
 ↪index=['15','25','50','100','inf'])
table.index.name = 'n'
print("Length of one-at-a-time t-interval / Length of Bonferroni t-interval:")
print("\t\tm")
print(table)
```

```
Length of one-at-a-time t-interval / Length of Bonferroni t-interval:
                m
            2         4        10
n
15   0.854643  0.748880  0.644914
25   0.863213  0.764341  0.667817
50   0.869052  0.774922  0.683592
100  0.871799  0.779910  0.691050
inf  0.874436  0.784706  0.698233
```