

Homework 6

November 11, 2019

Math A7800: Homework 6
Name: Jorge Monzon Diaz
Email: jorgemd503@gmail.com

0.1 Dependencies

```
[1]: import numpy as np
from IPython.display import Math
import pandas as pd
import scipy.stats as stats
```

1 Exercise 5.15

Let X_{ji} and X_{jk} be the i th and k th components, respectively of \mathbf{X}_j

1.1 (a) Show that $\mu_i = E(X_{ji}) = p_i$ and $\sigma_{ii} = Var(X_{ji}) = p_i(1 - p_i), i = 1, 2, \dots, p$

Since we are given discrete random variables

$$\mu = E(X_{ji}) = 1(p_i) + 0(1 - p_i) = p_i$$

$$\sigma_{ii} = Var(X_{ji}) = (p_i)(1 - \mu)^2 + (1 - p_i)(0 - \mu)^2 = (p_i)(1 - p_i)^2 + (1 - p_i)(p_i)^2$$

$$= p_i(1 - p_i)(1 - p_i + p_i) = p_i(1 - p_i)$$

1.2 (b) Show that $\sigma_{ik} = Cov(X_{ji}, X_{jk}) = -p_i p_k, i \neq k$. Why must this covariance be necessarily negative?

Due to independence

$$\sigma_{ik} = Cov(X_{ji}, X_{jk}) = E(X_{ji}, X_{jk}) - E(X_i)E(X_j) = 0 - p_i p_k = -p_i p_k$$

This covariance is then always negative as there is no correlation between the variables.

2 Exercise 5.16

Given:

\hat{p}_1	\hat{p}_2	\hat{p}_3	\hat{p}_4	\hat{p}_5
.30	.33	.16	.07	.14

$$p_5 = 1 - (p_1 + p_2 + p_3 + p_4)$$

2.1 (a) Construct simultaneous 95% confidence intervals for p_1, p_2, \dots, p_5 .

$$\hat{p} \pm \sqrt{\chi_q^2(\alpha)} \sqrt{\frac{\hat{\Sigma}}{n}}$$

Since we have 5 categories, $q=4$. We also know $n=355$.

$$\hat{p}_i \pm \sqrt{\chi_4^2(0.05)} \sqrt{\frac{\hat{p}_i(1 - \hat{p}_i)}{355}}$$

```
[2]: p=[.30,.33,.16,.07,.14]

for i in range(0,5):
    #display CI for each p
    j=i+1
    display(Math('\text{CI for }p_{\%i}' %j))
    display(Math('%.2f\pm\sqrt{\chi_4^2(0.05)}\sqrt{\frac{\%{.2f}(1-\%{.2f})}{355}}' % (p[i],p[i],p[i])))

    #compute CI bounds and display them
    lower=p[i]-np.sqrt(stats.chi2.ppf(1-.05,4))*np.sqrt((p[i]*(1-p[i]))/355)
    upper=p[i]+np.sqrt(stats.chi2.ppf(1-.05,4))*np.sqrt((p[i]*(1-p[i]))/355)
    display(Math('(%{.4f},\%{.4f})' %(lower,upper)))
    print("\n")
```

CI for p_1

$$0.30 \pm \sqrt{\chi_4^2(0.05)} \sqrt{\frac{0.30(1 - 0.30)}{355}}$$

(0.2251, 0.3749)

CI for p_2

$$0.33 \pm \sqrt{\chi_4^2(0.05)} \sqrt{\frac{0.33(1 - 0.33)}{355}}$$

(0.2531, 0.4069)

CI for p_3

$$0.16 \pm \sqrt{\chi_4^2(0.05)} \sqrt{\frac{0.16(1-0.16)}{355}}$$

(0.1001, 0.2199)

CI for p_4

$$0.07 \pm \sqrt{\chi_4^2(0.05)} \sqrt{\frac{0.07(1-0.07)}{355}}$$

(0.0283, 0.1117)

CI for p_5

$$0.14 \pm \sqrt{\chi_4^2(0.05)} \sqrt{\frac{0.14(1-0.14)}{355}}$$

(0.0833, 0.1967)

2.2 (b) Construct a simultaneous 95% confidence interval that allows a comparison of the Bank of Shorewood with its major competitor, Bank B. Interpret this interval.

CI for difference in proportions:

$$\hat{p}_i - \hat{p}_j \pm \sqrt{\chi_q^2(\alpha)} \sqrt{\frac{\hat{p}_i(1-\hat{p}_i) + \hat{p}_j(1-\hat{p}_j) - 2\hat{p}_i\hat{p}_j}{n}}$$

Bank of Shorewood: p_1

Bank B: p_2

$$\hat{p}_1 - \hat{p}_2 \pm \sqrt{\chi_4^2(0.05)} \sqrt{\frac{\hat{p}_1(1-\hat{p}_1) + \hat{p}_2(1-\hat{p}_2) - 2\hat{p}_1\hat{p}_2}{355}}$$

```
[3]: #display CI formula
display(Math('%.2f - %.2f\pm\sqrt{\chi_4^2(0.05)}\sqrt{\frac{%.2f(1-%.2f)+%.2f(1-%.2f) - 2 ( %.2f)( %.2f)}{355}}'
    \rightarrow(p[0],p[1],p[0],p[0],p[1],p[1],p[0],p[1])))

#compute CI bounds and display them
lower=p[0]-p[1]-np.sqrt(stats.chi2.ppf(1-.05,4))*np.
    \rightarrow sqrt(((p[0]*(1-p[0]))+(p[1]*(1-p[1]))-2*p[0]*p[1])/355)
upper=p[0]-p[1]+np.sqrt(stats.chi2.ppf(1-.05,4))*np.
    \rightarrow sqrt(((p[0]*(1-p[0]))+(p[1]*(1-p[1]))-2*p[0]*p[1])/355)
display(Math(' ( %.4f, %.4f ) ' %(lower,upper)))
```

$$0.30 - 0.33 \pm \sqrt{\chi_4^2(0.05)} \sqrt{\frac{0.30(1 - 0.30) + 0.33(1 - 0.33) - 2(0.30)(0.33)}{355}}$$

$$(-0.1089, 0.0489)$$

Since the simultaneous confidence is so small, we can claim there is no significant difference between the two proportions.

3 Exercise 5.17

Given:

\hat{p}_1	\hat{p}_2	\hat{p}_3
.585	.310	.105

Construct the 95% simultaneous confidence intervals for the three proportions p_1, p_2 , and $p_3 = 1 - (p_1 + p_2)$

$$\hat{p} \pm \sqrt{\chi_q^2(\alpha)} \sqrt{\frac{\hat{\Sigma}}{n}}$$

Since we have 3 categories, $q=2$. We also know $n=200$.

$$\hat{p}_i \pm \sqrt{\chi_2^2(0.05)} \sqrt{\frac{\hat{p}_i(1 - \hat{p}_i)}{200}}$$

```
[4]: p=[.585,.310,.105]

for i in range(0,3):
    #display CI for each p
    j=i+1
    display(Math('\text{CI for }p_{\%i}' %j))
    display(Math('%.2f\pm\sqrt{\chi_2^2(0.05)}\sqrt{\frac{%.2f(1-%.
→2f)}{355}}' %(p[i],p[i],p[i])))

    #compute CI bounds and display them
    lower=p[i]-np.sqrt(stats.chi2.ppf(1-.05,2))*np.sqrt((p[i]*(1-p[i]))/200)
    upper=p[i]+np.sqrt(stats.chi2.ppf(1-.05,2))*np.sqrt((p[i]*(1-p[i]))/200)
    display(Math('%.4f, %.4f' %(lower,upper)))
    print("\n")
```

CI for p_1

$$0.58 \pm \sqrt{\chi_2^2(0.05)} \sqrt{\frac{0.58(1 - 0.58)}{355}}$$

(0.4997, 0.6703)

CI for p_2

$$0.31 \pm \sqrt{\chi_2^2(0.05)} \sqrt{\frac{0.31(1 - 0.31)}{355}}$$

(0.2300, 0.3900)

CI for p_3

$$0.10 \pm \sqrt{\chi^2_2(0.05)} \sqrt{\frac{0.10(1 - 0.10)}{355}}$$

(0.0519, 0.1581)

4 Exercise 5.30

Refer to the data on energy consumption in Exercise 3.18

x_1 = petroleum

x_2 = natural gas

x_3 = hydroelectric power

x_4 = nuclear electric power

$$\bar{\mathbf{x}} = \begin{bmatrix} 0.766 \\ 0.508 \\ 0.438 \\ 0.161 \end{bmatrix}$$

$$\mathbf{S} = \begin{bmatrix} 0.856 & 0.635 & 0.173 & 0.096 \\ 0.635 & 0.568 & 0.128 & 0.067 \\ 0.173 & 0.127 & 0.171 & 0.039 \\ 0.096 & 0.067 & 0.039 & 0.043 \end{bmatrix}$$

4.1 (a) Obtain the large sample 95% Bonferroni confidence intervals for the mean consumption of each of the four types, the total of the four, and the difference, petroleum minus natural gas

Bonferroni CI:

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

$$\left[\bar{x}_i - t_{n-1}\left(\frac{\alpha}{2m}\right)\sqrt{\frac{s_{ii}}{n}}, \bar{x}_i + t_{n-1}\left(\frac{\alpha}{2m}\right)\sqrt{\frac{s_{ii}}{n}} \right]$$

For large samples

$$\left[\bar{x}_i - z\left(\frac{\alpha}{2m}\right)\sqrt{\frac{s_{ii}}{n}}, \bar{x}_i + z\left(\frac{\alpha}{2m}\right)\sqrt{\frac{s_{ii}}{n}} \right]$$

m=6

For $\mu_1, \mu_2, \mu_3, \mu_4$, their respective corresponding variance is $s_{11}, s_{22}, s_{33}, s_{44}$

Let μ_{tot} be the mean for the total, and s_{tot} be its corresponding variance.

$$\overline{x_{tot}} = \sum_{i=1}^4 (\bar{x}_i) = 1.873$$

$$s_{tot} = Var(x_1 + x_2 + x_3 + x_4) = \sum_{i=1}^4 \sum_{j=1}^4 (s_{i,j}) = 3.913$$

Let μ_{1-2} be the mean for petroleum minus natural gas, and s_{1-2} be its corresponding variance.

$$\overline{x_{1-2}} = \bar{x}_1 - \bar{x}_2 = 0.258$$

$$s_{1-2} = Var(x_1 - x_2) = Var(x_1) + Var(x_2) - 2Cov(x_1, x_2) = s_{11} + s_{22} - 2s_{12} = 0.154$$


```
[5]: labels=["petroleum","natural gas", "hydroelectric power", "nuclear electric_
      ↪power", "total", "petroleum minus natural gas"]
xbar=[0.766,0.508,0.438,0.161,1.873,0.258]
s=[0.856,0.568,0.171,0.043,3.913,0.154]

for i in range(len(labels)):
    #print equations
    display(Math('\text{Large sample Bonferroni CI for %s}' %labels[i]))
    display(Math('\left[%.3f-z(\frac{0.05}{12})\sqrt{\frac{%.3f}{50}},%.
      ↪3f+z(\frac{0.05}{12})\sqrt{\frac{%.3f}{50}}\right]'_
      ↪%(xbar[i],s[i],xbar[i],s[i]))

    #compute CI bounds and display them
    lower=xbar[i]-stats.norm.ppf(1-0.05/12)*np.sqrt(s[i]/50)
    upper=xbar[i]+stats.norm.ppf(1-0.05/12)*np.sqrt(s[i]/50)
    display(Math('("%.3f, %.3f)' %(lower,upper)))
    print("\n")
```

Large sample Bonferroni CI for petroleum

$$\left[0.766 - z\left(\frac{0.05}{12}\right)\sqrt{\frac{0.856}{50}}, 0.766 + z\left(\frac{0.05}{12}\right)\sqrt{\frac{0.856}{50}} \right]$$

(0.421, 1.111)

Large sample Bonferroni CI for natural gas

$$\left[0.508 - z\left(\frac{0.05}{12}\right)\sqrt{\frac{0.568}{50}}, 0.508 + z\left(\frac{0.05}{12}\right)\sqrt{\frac{0.568}{50}} \right]$$

(0.227, 0.789)

Large sample Bonferroni CI for hydroelectric power

$$\left[0.438 - z\left(\frac{0.05}{12}\right)\sqrt{\frac{0.171}{50}}, 0.438 + z\left(\frac{0.05}{12}\right)\sqrt{\frac{0.171}{50}} \right]$$

(0.284, 0.592)

Large sample Bonferroni CI for nuclear electric power

$$\left[0.161 - z\left(\frac{0.05}{12}\right)\sqrt{\frac{0.043}{50}}, 0.161 + z\left(\frac{0.05}{12}\right)\sqrt{\frac{0.043}{50}} \right]$$

(0.084, 0.238)

Large sample Bonferroni CI for total

$$\left[1.873 - z\left(\frac{0.05}{12}\right)\sqrt{\frac{3.913}{50}}, 1.873 + z\left(\frac{0.05}{12}\right)\sqrt{\frac{3.913}{50}} \right]$$

(1.135, 2.611)

Large sample Bonferroni CI for petroleum minus natural gas

$$\left[0.258 - z\left(\frac{0.05}{12}\right)\sqrt{\frac{0.154}{50}}, 0.258 + z\left(\frac{0.05}{12}\right)\sqrt{\frac{0.154}{50}} \right]$$

(0.112, 0.404)

4.2 (b) Obtain the large sample 95% simultaneous T^2 intervals for the means consumption of each of the four types, the total of the four, and the difference, petroleum minus natural gas. Compare with your results for Part a.

For large samples

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

$$\left[\bar{x}_i - \sqrt{\chi_p^2(\alpha)} \sqrt{\frac{s_{ii}}{n}}, \bar{x}_i + \sqrt{\chi_p^2(\alpha)} \sqrt{\frac{s_{ii}}{n}} \right]$$

We already computed the remaining means and sample variances in the previous problem so now we can just plug into our equation and compute

```
[6]: for i in range(len(labels)):
    #print equations
    display(Math('\text{Large simultaneous } T^2 \text{-interval for } %s' \_
    \rightarrow \%labels[i]))
    display(Math('\left[ %.3f - \sqrt{\chi_4^2(0.05)} \sqrt{\frac{%.3f}{50}}, \_
    \rightarrow 3f + \sqrt{\chi_4^2(0.05)} \sqrt{\frac{%.3f}{50}} \right]' \_
    \rightarrow \%(\bar{x}[i], s[i], \bar{x}[i], s[i])))

    #compute CI bounds and display them
    lower = \bar{x}[i] - np.sqrt(stats.chi2.ppf(1-0.05, 4)) * np.sqrt(s[i]/50)
    upper = \bar{x}[i] + np.sqrt(stats.chi2.ppf(1-0.05, 4)) * np.sqrt(s[i]/50)
    display(Math(' (%.3f, %.3f)' \% (lower, upper)))
    print("\n")
```

Large simultaneous T^2 -interval for petroleum

$$\left[0.766 - \sqrt{\chi_4^2(0.05)} \sqrt{\frac{0.856}{50}}, 0.766 + \sqrt{\chi_4^2(0.05)} \sqrt{\frac{0.856}{50}} \right]$$

(0.363, 1.169)

Large simultaneous T^2 -interval for natural gas

$$\left[0.508 - \sqrt{\chi_4^2(0.05)} \sqrt{\frac{0.568}{50}}, 0.508 + \sqrt{\chi_4^2(0.05)} \sqrt{\frac{0.568}{50}} \right]$$

(0.180, 0.836)

Large simultaneous T^2 -interval for hydroelectric power

$$\left[0.438 - \sqrt{\chi_4^2(0.05)} \sqrt{\frac{0.171}{50}}, 0.438 + \sqrt{\chi_4^2(0.05)} \sqrt{\frac{0.171}{50}} \right]$$

(0.258, 0.618)

Large simultaneous T^2 -interval for nuclear electric power

$$\left[0.161 - \sqrt{\chi_4^2(0.05)} \sqrt{\frac{0.043}{50}}, 0.161 + \sqrt{\chi_4^2(0.05)} \sqrt{\frac{0.043}{50}} \right]$$

(0.071, 0.251)

Large simultaneous T^2 -interval for total

$$\left[1.873 - \sqrt{\chi_4^2(0.05)} \sqrt{\frac{3.913}{50}}, 1.873 + \sqrt{\chi_4^2(0.05)} \sqrt{\frac{3.913}{50}} \right]$$

(1.011, 2.735)

Large simultaneous T^2 -interval for petroleum minus natural gas

$$\left[0.258 - \sqrt{\chi_4^2(0.05)} \sqrt{\frac{0.154}{50}}, 0.258 + \sqrt{\chi_4^2(0.05)} \sqrt{\frac{0.154}{50}} \right]$$

(0.087, 0.429)

5 Exercise 5.18a

Use the college test data in Table 5.2

Test the null hypothesis $H_0 : \boldsymbol{\mu}^T = [500, 50, 30]$ versus $H_1 : \boldsymbol{\mu}^T \neq [500, 50, 30]$ at the $\alpha = .05$ level of significance. Suppose $[500, 50, 30]^T$ represent average scores for thousands of college students over the last 10 years. Is there reason to believe that the group of students represented by the scores in Table 5.2 is scoring differently? Explain.

$$T^2 = n(\bar{\mathbf{X}} - \boldsymbol{\mu}_0)^T \mathbf{S}^{-1}(\bar{\mathbf{X}} - \boldsymbol{\mu}_0)$$

$$\mathbf{S} = \frac{1}{n-1}(\mathbf{X} - \mathbf{1}\bar{\mathbf{x}}^T)^T(\mathbf{X} - \mathbf{1}\bar{\mathbf{x}}^T)$$

```
[7]: #read data
x = pd.read_excel (r'hw6.xlsx', sheet_name='data 518')

#compute mean
xmean=x.mean(axis = 0)
print("x-mean:")
print(xmean, "\n")

#strip values from table
X=x.to_numpy()

#find S
R=X-np.ones((87,1))*[xmean[0],xmean[1],xmean[2]]
S=1/86*(np.matmul(R.transpose(),R))
print('S:')
print(S)

#invert S
invS=np.linalg.inv(S)
print('\nS^(-1):')
print(invS)

#compute T^2
T2=87*np.matmul(np.matmul(np.
    →array([xmean[0]-500,xmean[1]-50,xmean[2]-30]),invS),np.
    →array([xmean[0]-500,xmean[1]-50,xmean[2]-30]).transpose())
print('\nT^2:')
print(T2)
```

```
x-mean:
x1    526.586207
x2     54.689655
x3     25.126437
dtype: float64
```

```
S:
[[5808.05934242  597.83520449  222.02967121]
 [ 597.83520449  126.05372895   23.38853248]
 [ 222.02967121   23.38853248   23.11173483]]
```

```
S^(-1):
[[ 0.00043194 -0.00157423 -0.0025565 ]
 [-0.00157423  0.0155044  -0.00056675]
 [-0.0025565  -0.00056675  0.06840139]]
```

```
T^2:
223.3101756848916
```

Our critical point at $\alpha = 0.05$ is

$$\frac{(n-1)p}{(n-p)} F_{p, n-p}(\alpha) = \frac{(87-1)3}{(87-3)} F_{3, 87-3}(0.05) = \frac{258}{84} F_{3, 84}(0.05)$$

```
[8]: print("Critical point:", 258/84*stats.f.ppf(1-.05, 3, 84))
```

```
Critical point: 8.3334833261061
```

We reject our null hypothesis at $\alpha = .05$ because $223.3102 > 8.3335$.

6 Exercise 5.19

Measurements of x_1 = stiffness and x_2 = bending strength for a sample of $n=30$ pieces of a particular grade of lumber are given in Table 5.11

6.1 (a) Construct and sketch a 95% confidence ellipse for the pair $[\mu_1, \mu_2]^T$ where $\mu_1 = E(X_1)$ and $\mu_2 = E(X_2)$

We know that

$$T^2 = n(\bar{\mathbf{X}} - \boldsymbol{\mu}_0)^T \mathbf{S}^{-1} (\bar{\mathbf{X}} - \boldsymbol{\mu}_0)$$

and

$$T^2 \sim \frac{(n-1)p}{(n-p)} F_{p, n-p}$$

So our confidence region is given by

$$n(\bar{\mathbf{X}} - \boldsymbol{\mu}_0)^T \mathbf{S}^{-1} (\bar{\mathbf{X}} - \boldsymbol{\mu}_0) \leq \frac{(n-1)p}{(n-p)} F_{p, n-p}(0.05)$$

We are given

$$\mathbf{X} = [\mathbf{x}_1 \quad \mathbf{x}_2]$$

First we find

$$\bar{\mathbf{X}} = \begin{bmatrix} \bar{x}_1 \\ \bar{x}_2 \end{bmatrix}$$

Using the mean, we can then compute the sample variance

$$\mathbf{S} = \frac{1}{n-1} (\mathbf{X} - \mathbf{1}\bar{\mathbf{x}}^T)^T (\mathbf{X} - \mathbf{1}\bar{\mathbf{x}}^T)$$

Now we have sufficient information to plug into the equation for our confidence region

$$n(\bar{\mathbf{X}} - \boldsymbol{\mu}_0)^T \mathbf{S}^{-1} (\bar{\mathbf{X}} - \boldsymbol{\mu}_0) \leq \frac{(n-1)p}{(n-p)} F_{p, n-p}(0.05)$$

$$30 \begin{bmatrix} \bar{x}_1 - \mu_1 & \bar{x}_2 - \mu_2 \end{bmatrix} \mathbf{S}^{-1} \begin{bmatrix} \bar{x}_1 - \mu_1 \\ \bar{x}_2 - \mu_2 \end{bmatrix} \leq \frac{(29)2}{(28)} F_{2, 28}(0.05)$$

Additionally, to find the semi-major and semi-minor axis lengths of our confidence ellipse, we use \mathbf{S} to compute the eigenvalues and their respective eigenvectors.

$$|\mathbf{S} - \lambda \mathbf{I}| = 0, \text{ and } \lambda_1, \lambda_2 \text{ are our eigenvalues with their respective eigenvectors } \mathbf{u}_1, \mathbf{u}_2$$

The half-axis lengths from (μ_1, μ_2) are given by:

$$\sqrt{\frac{(29)2}{30(28)} F_{2, 28}(0.05) \lambda_i}$$

in their respective directions $\mathbf{u}_1, \mathbf{u}_2$

6.2 (b) Suppose $\mu_{10} = 2000, \mu_{20} = 10000$ represent “typical” values for the stiffness and bending strength, respectively. Given the result in (a), are the data in Table 5.11 consistent with these values?

We can perform a T^2 hypothesis test, we know that

$$T^2 = n(\bar{\mathbf{X}} - \boldsymbol{\mu}_0)^T \mathbf{S}^{-1}(\bar{\mathbf{X}} - \boldsymbol{\mu}_0) = 30 \begin{bmatrix} \bar{x}_1 - 2000 & \bar{x}_2 - 10000 \end{bmatrix} \mathbf{S}^{-1} \begin{bmatrix} \bar{x}_1 - 2000 \\ \bar{x}_2 - 10000 \end{bmatrix}$$

And our critical point is given by

$$\frac{(n-1)p}{(n-p)} F_{p, n-p}(\alpha) = \frac{(29)2}{(28)} F_{2,28}(0.05)$$

$$H_0 : [\mu_{10} = 2000, \mu_{20} = 10000]^T$$

$$H_1 : [\mu_{10} \neq 2000, \mu_{20} \neq 10000]^T$$

We reject our null hypothesis at $\alpha = .05$ if $T^2 > \frac{(29)2}{(28)} F_{2,28}(0.05)$, it is for this same reason that we also fail to reject H_0 if the point falls within the confidence ellipse.

6.3 (c) Is the bivariate normal distribution a viable population model? Explain the reference to Q-Q plots and a scatter diagram.

A scatter plot is useful in visually determining outliers in our data. We can use Q-Q plots to test for marginal normality of both the stiffness and bending strength. These three diagrams then help us to determine whether bivariate normality is a plausible population model in this scenario. If the Q-Q plots follow linear trends and the scatter plot data coincides, we can then assume it must be plausible.

7 Exercise 5.21

Using the data on bone mineral content in Table 1.8, construct the 95% Bonferroni intervals for the individual means. Also, find the 95% simultaneous T^2 -intervals. Compare the two sets of intervals.

We have 6 samples $\mathbf{X} = [\mathbf{x}_1, \dots, \mathbf{x}_6]$ and conversely $\tilde{\mathbf{X}} = [\bar{x}_1, \dots, \bar{x}_6]^T$

We can now compute the sample variance

$$\mathbf{S} = \frac{1}{n-1}(\mathbf{X} - \mathbf{1}\bar{\mathbf{x}}^T)^T(\mathbf{X} - \mathbf{1}\bar{\mathbf{x}}^T)$$

This is a 6x6 matrix from which we will need $s_{11}, s_{22}, s_{33}, s_{44}, s_{55}, s_{66}$

The 95% confidence Bonferroni intervals, given $m = p = 6$, for each μ_i where $i = 1, \dots, 6$ is given by

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

$$\bar{x}_i - t_{24}\left(\frac{0.05}{12}\right)\sqrt{\frac{s_{ii}}{25}} \leq \mu_i \leq \bar{x}_i + t_{24}\left(\frac{0.05}{12}\right)\sqrt{\frac{s_{ii}}{25}}$$

The simultaneous 95% confidence T^2 -intervals are similarly given for each μ_i where $i = 1, \dots, 6$ by

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

$$\bar{x}_i - \sqrt{\frac{6(24)}{25(19)}F_{6,19}(0.05)}\sqrt{s_{ii}} \leq \mu_i \leq \bar{x}_i + \sqrt{\frac{6(24)}{25(19)}F_{6,19}(0.05)}\sqrt{s_{ii}}$$

8 Exercise 5.29

Refer to the car body data in Exercise 5.28. These are all measured as deviations from target value so it is appropriate to test the null hypothesis that the mean vector is zero. Using the first 30 cases, test $H_0 : \boldsymbol{\mu} = \mathbf{0}$ at $\alpha = 0.05$

To test the first 30 cases using this null hypothesis, we can use a T^2 hypothesis test.

First we need to compute the mean for each of our 6 samples (each using only the first 30 observed values) and the sample variance

$$\bar{\mathbf{x}} = \begin{bmatrix} \bar{x}_1 \\ \bar{x}_2 \\ \bar{x}_3 \\ \bar{x}_4 \\ \bar{x}_5 \\ \bar{x}_6 \end{bmatrix}$$

$$\mathbf{S} = \frac{1}{n-1}(\mathbf{X} - \mathbf{1}\bar{\mathbf{x}}^T)^T(\mathbf{X} - \mathbf{1}\bar{\mathbf{x}}^T)$$

We know the formula for the Hotelling T^2 , and now have enough information to plug in all the values

$$T^2 = n(\bar{\mathbf{X}} - \boldsymbol{\mu}_0)^T \mathbf{S}^{-1}(\bar{\mathbf{X}} - \boldsymbol{\mu}_0) = 50(\bar{\mathbf{X}} - \mathbf{0})^T \mathbf{S}^{-1}(\bar{\mathbf{X}} - \mathbf{0}) = 50 \begin{bmatrix} \bar{x}_1 \\ \bar{x}_2 \\ \bar{x}_3 \\ \bar{x}_4 \\ \bar{x}_5 \\ \bar{x}_6 \end{bmatrix}^T \mathbf{S}^{-1} \begin{bmatrix} \bar{x}_1 \\ \bar{x}_2 \\ \bar{x}_3 \\ \bar{x}_4 \\ \bar{x}_5 \\ \bar{x}_6 \end{bmatrix}$$

And our critical point is given by

$$\frac{(n-1)p}{(n-p)} F_{p,n-p}(\alpha) = \frac{(29)6}{(24)} F_{6,24}(0.05)$$

We reject our null hypothesis at $\alpha = .05$ if $T^2 > \frac{(29)6}{(24)} F_{6,24}(0.05)$.