Graduate Course – Term 1, Sept-Dec 2018
30.502: Research Methods

SINGAPORE UNIVERSITY OF TECHNOLOGY AND DESIGN
Established in collaboration with MIT

Assessed Problem Set 5

Due: 23^{rd} Nov, 23:59PM
Delivery method: E-dimension
Format: Portable Document Format (PDF)

For all the python based questions in this problem set, please copy and past the code that you used for each function/operation and the result of the calculation.

Problem 1: MLE

I analysed the flow of students into SUTD's student shop at appromimately 3 pm last Friday afternoon. I recorded the following times (in minutes) between the arrivals of 10 students: 3.2, 2.1, 5.3, 4.2, 1.2, 2.8, 6.4, 1.5, 1.9, and 3.0. Let's assume that the interarrival time can be modelled by the following PDF:

$$f(x;\beta) = \begin{cases} \frac{1}{\beta}e^{-\frac{x}{\beta}} & ;x \geq 0, \\ 0 & ;x < 0 \end{cases}$$

Write a python code that plots the likelihood as a function of β and state the maximum likelihood estimate of β. Please include a copy of the plot in your solution sheet.

```
import numpy as np

def myexp(t, bet):
    return (1/bet)*np.exp(-(1/bet)*t)

times=[3.2, 2.1, 5.3, 4.2, 1.2, 2.8, 6.4, 1.5, 1.9, 3.0]
guess=np.linspace(0, 10, 101)

likely=[]
for g in guess:
    temp=1
    for t in times:
        temp= temp*myexp(t, g)

    likely.append(temp)
plt.figure()
plt.plot(guess, likely)
plt.ylabel('Likelihood')
plt.xlabel('beta')
plt.grid()
```

The most likely vale is: $\beta = 3.2$

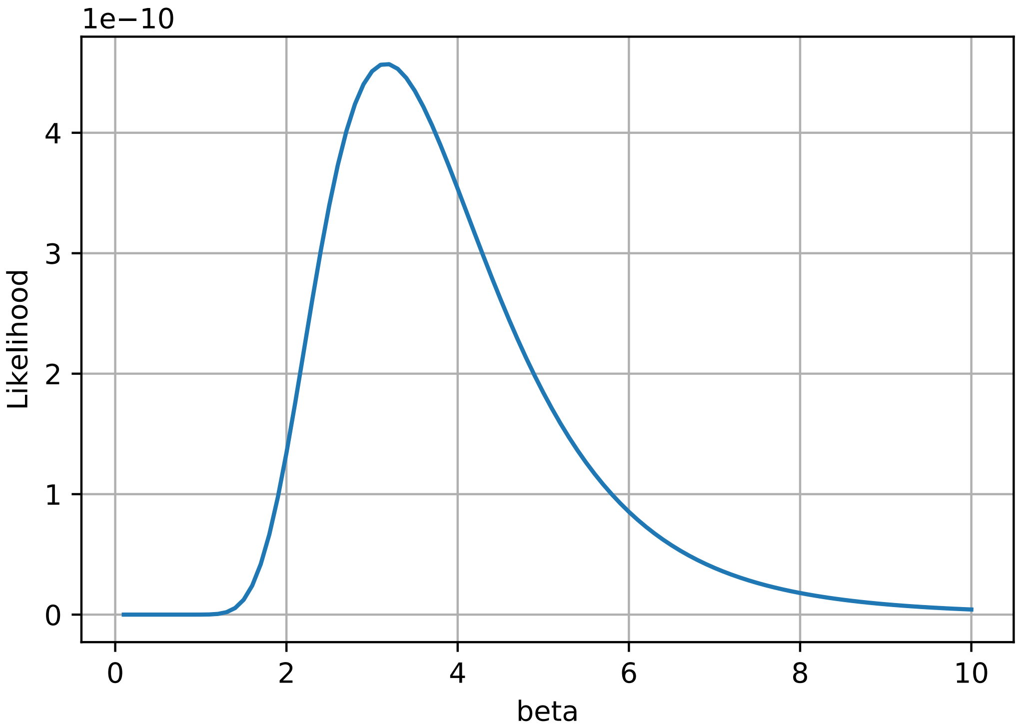

Figure 1: Likelihood vs beta

Problem 2: Hotelling

The population means for a multivariate dataset are: weight= 60 kg, shoe size= 24 cm, and height= 161 cm. A random sample of people were analysed and the data in Table 1 were obtained:

Weight (kg)	Shoe Size (cm)	Height (cm)
65	28	175
80	26	159
70	27	180
62	29	167
74	24	170

Table 1: Body Measurement Data

(a) Evaluate T^2.

(b) Based on the calculated T^2, is this sample significantly different from the population means at the $\alpha = 0.01$ significance level?

```
print "==================="
print "Hotelling␣T-squard"
print "==================="
import scipy.stats as ss

x=np.array([[65, 80, 70, 62, 74],[28, 26, 27, 29, 24],
   [175,159,180,167,170]]) # the vector of observarions
xbar=np.array([np.mean(x[0,:]),np.mean(x[1,:]), np.mean(x[1,:])])
    #sample mean vector

n=x.shape[1] #number of obersvations for each variable
k=x.shape[0]
Cov=np.cov(x) # compute the covariance matrix of the observations

mu=np.array([60,24,161])

invCov=np.linalg.inv(Cov)
Tsqd=n* np.matrix.transpose(np.array(xbar)-np.array(mu))  .dot(
   invCov) .dot(np.array(xbar)-np.array(mu))


F=(n-k)*Tsqd/(k*(n-1))
Fc=ss.f.ppf(0.99, k,(n-k))  # One right hand tail test
   signifficance is 0.05, and 95% confidence
print "T-squared=", Tsqd
print "F-value", F
print "Critical␣F=", Fc  # One right hand tail test signifficance
    is 0.05, and 95% confidence
if F>Fc:
    print "F>Fc,␣Reject␣H0"
else:
    print "F<Fc,␣Accept␣H0"
```

The function outputs the following:

(a)

```
===================
Hotelling T-squard
===================
T-squared= 1767.9229724823929
F-value 294.65382874706546
Critical F= 99.16620137447148
F>Fc, Reject H0
```

(b)

Our null hypothesis is H_0: The random sample is representative of the population. The Hotelling-T^2 analysis shows that the F-value is much larger than the $\alpha = 0.01$ critical F-value (99.166). Therefore, there is statistical evidence to reject H_0 and we conclude that the random sample is significantly different from the population means.

Problem 3: ANOVA

(a) Fill in the missing data in the one-way ANOVA Table 3:

Source	DoF	SS	MS	F	P
Factor (between)	3	36.15	?	?	?
Error (within)	?	?	?		
Total	19	196.04			

Table 2: One way ANOVA table for problem 3(a)

(a) Fill in the missing data in the one-way ANOVA Table 3:

Source	DoF	SS	MS	F	P
Factor (between)	3	36.15	36.15/3=12.05	$\frac{12.05}{9.99} = 1.21$	0.343 (dof=3,16)
Error (within)	19-3=16	196.04-36.15=159.89	9.99		
Total	19	196.04			

Table 3: One way ANOVA table for problem 3(a)

A textile company weaves a fabric on a large number of looms. It would like the looms to be homogeneous so that it obtains a fabric of uniform strength. The process engineer suspects that, in addition to the usual variation in strength within samples of fabric from the same loom, there may also be significant variations in strength between looms. To investigate this, she selects four looms at random and makes four strength determinations on the fabric manufactured on each loom. This experiment is run in random order, and the data obtained are shown in Table 4.

	Observations				
Loom	1	2	3	4	y_i
1	98	97	99	96	390
2	91	90	93	92	366
3	96	95	97	95	383
4	95	96	99	98	388

Table 4: Strength Data for problem 3(b-f)

(b) Calculate the variation due to the different groups (SSB)

(c) Calculate the mean square variation between the groups (MSB)

(d) Calculate the variation (error) within the groups (SSE)

(e) Calculate the mean square variation within the groups (MSE)

(f) Is there a significant difference between the looms at the $\alpha = 0.05$ significance level?

A textile company weaves a fabric on a large number of looms. It would like the looms to be homogeneous so that it obtains a fabric of uniform strength. The process engineer suspects that, in addition to the usual variation in strength within samples of fabric from the same loom, there may also be significant variations in strength between looms. To investigate this, she selects four looms at random and makes four strength determinations on the fabric manufactured on each loom. This experiment is run in random order, and the data obtained are shown in Table 3.17. The ANOVA is conducted and is shown in Table 3.18. From the ANOVA, we conclude that the looms in the plant differ significantly.

■ TABLE 3.17
Strength Data for Example 3.11

	Observations				
Looms	**1**	**2**	**3**	**4**	$y_{i.}$
1	98	97	99	96	390
2	91	90	93	92	366
3	96	95	97	95	383
4	95	96	99	98	388
					$1527 = y_{..}$

The variance components are estimated by $\hat{\sigma}^2 = 1.90$ and

$$\hat{\sigma}_\tau^2 = \frac{29.73 - 1.90}{4} = 6.96$$

Therefore, the variance of any observation on strength is estimated by

$$\hat{\sigma}_y = \hat{\sigma}^2 + \hat{\sigma}_\tau^2 = 1.90 + 6.96 = 8.86.$$

Most of this variability is attributable to differences *between* looms.

■ TABLE 3.18
Analysis of Variance for the Strength Data

Source of Variation	Sum of Squares	Degrees of Freedom	Mean Square	F_0	*P*-Value
Looms	89.19	3	29.73	15.68	<0.001
Error	22.75	12	1.90		
Total	111.94	15			