PPSAS205
Introduction and Modelling in R

Time: 2 Hours
Marks: 60

Note:

1) **The candidate has the option to either attempt question 3A or question 3B. Rest all questions are mandatory.**
2) **Numbers on the right indicate full marks.**
3) **The candidate will be provided with the formula sheet and graph papers [if required] for the examination.**
4) **Use of approved scientific calculators is allowed.**

Q1. Attempt the following: **15 Marks**

A. A machine that fills bottles is known to have a mean filling amount of 125 gms and a standard deviation of 20 gms. A quality control manager took a random sample of filled bottles and found the sample mean to be 130. The quality manager assumed the sample must not be representative. Validate whether the conclusion is correct or not.

```
sample_data <- c(129, 132, 127, 134, 131, 128, 130, 133, 126, 128,
127, 129, 133, 132, 130, 129, 135, 128, 131, 129,
130, 131, 132, 128, 126, 129, 130, 133, 130, 131)
```

5 Marks

B. Given a sample variance of 127 from a set of 9 observations, construct a 95% confidence interval for the population variance. **5 Marks**

Formula for confidence interval:

$((n - 1) * s^2) / chi2(a/2, n - 1) <= sigma^2 <= ((n - 1) * s^2) / chi2(1 - a/2, n - 1)$

where n is the sample size, s^2 is the sample variance, chi2 is the chi-squared distribution function, and a is the level of significance (0.05 for a 95% confidence interval).

C. A manufacturer of pet foods was wondering whether cat owners and dog owners reacted differently to premium pet foods. They commissioned a consumer survey that yielded the following data.

Pet	Owners Surveyed	Number Using Premium Food
Cat	298	192
Dog	196	111

Is it reasonable to conclude at significance level 0.02 that cat owners are more likely than dog owners to feed their pets premium food? **5 Mark**

Q2. Attempt he following: **15 Marks**

A. The following sample of eight observations is from an infinite population with a normal distribution.

76 89 78 85 90 94 51 89

i. Find the sample mean **1 Mark.**
ii. Estimate population standard deviation **1 Mark**
iii. Construct a 95% confidence interval for the population mean **3 Marks.**

B. A study compared the effects of four 1-month intake of medications for diabetics. The sugar levels as mentioned below:

Medication_1	178	187	181	189	185
Medication_2	194	191	187	190	188
Medication12	173	178	169	183	176
Medication_3	179	183	178	169	181

Check whether there is an effect on sugar levels due to medication.

i. Create a matrix “sugarlevels” similar to the above table **1 Mark**.
ii. Print sugarlevels matrix **1 Mark**
iii. Test the hypothesis to check the effect of medication on sugarlevels **3 Marks.**

C. Perform a correlation test for the below data **5 Marks.**

```
turbo_1 <- c(13,15,16, 27, 28, 10)
turbo_12 <- c(42, 26, 17, 18, 29, 12)
```

Q3 A.

In economics, the demand function for a product is often estimated by regressing the quantity sold (Q) and the price (P). The Bamsy company is trying to estimate the demand function for new doll “Annabelle” and has collected the following data:

30 Marks.

p	30	27.5	26	24	32.5	20	18	16.5
q	225	256	283	90	212	738	350	176

i. Create a dataframe and plot the data **7 Marks.**

ii. Calculate the value of the least square regression **4 Marks.**

iii. Plot the best fit line **4 Marks**

iv. Provide the equation that best fits the model **7 Marks.**

v. Predict the amount of dolls sold when price is 15 **8 Marks**

OR

Q3 B

Sales of major appliances vary with the new housing market: When new home sales are good so are the sales of dishwashers, washing machines, driers, refrigerators. A trade association compiled the following data (in thousands of units) on major appliance sales and housing:

30 Marks

Housing	Appliance Sales
2	7
2.5	7.5
3.2	8
3.6	10
3.3	17.2
4	17.7
4.2	18.4
4.6	19
4.8	19.7
5	20

i. Develop an equation for the relationship between appliances sales and housing **7 Marks.**

ii. Interpret the slope of the regression line **3 Marks.**

iii. Compute and interpret the standard error of estimate **4 Marks.**

Iv. Plot the best fit line **4 Marks.**

iv. Housing next year may be beyond recorded range, estimates as high as 8 million units have been predicted. Compute an approximate 90% prediction interval of appliance sales, based on previous data and new prediction of housing. **12 Marks**