Your Signature _____

Instructions:
1. For writing your answers use both sides of the paper in the answer booklet.

2. Additional sheets taken, if any, should be properly attached to the main answer booklet.

3. **Please write your name on every page of this booklet and every additional sheet taken.**

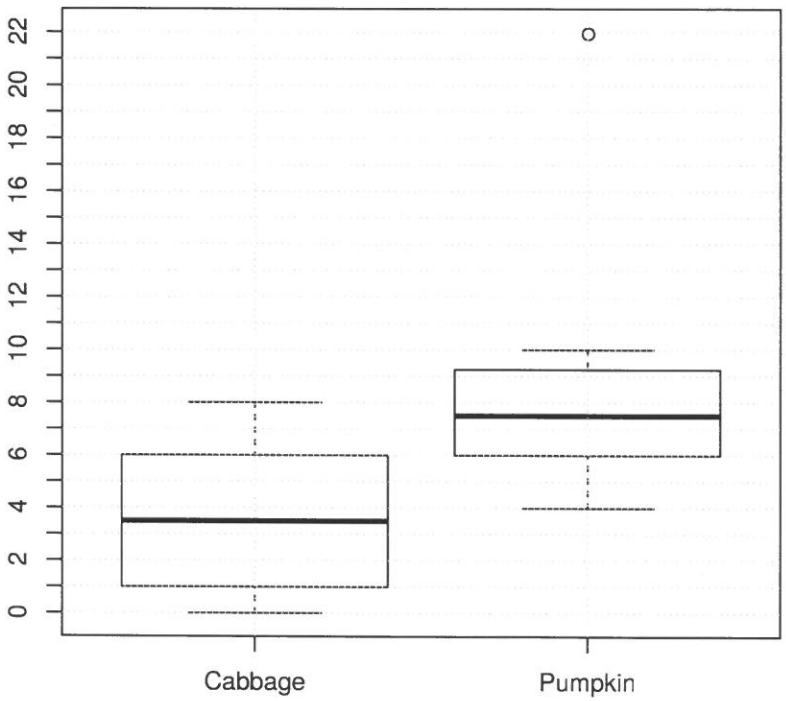
Score

Q.No.	Alloted Score	Score
1.	10	
2.	20	
3.	15	
4.	20	
5.	20	
6.	15	
Total	100	

Number of Extra sheets attached to the answer script: _____

1.(a)Below there are two box plots that are shown.

Comparison of weights in Pounds



(i) Which vegetable has a higher median weight ?

(ii) What is the approximate median weight of pumpkins ?

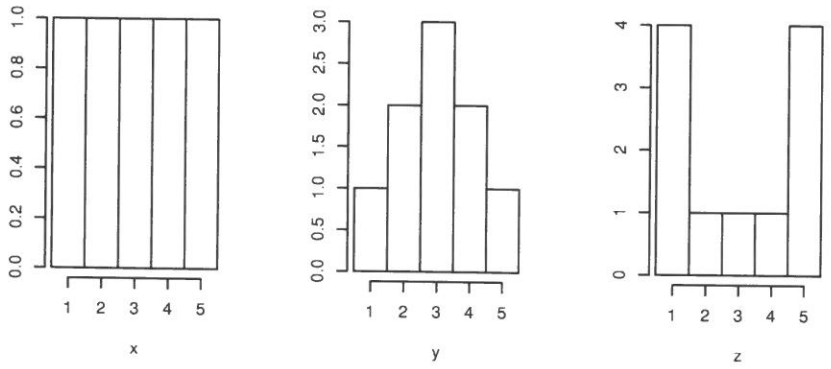
(iii) What is the IQR for Cabbages ?

(iv) How many outliers are there in this data set ?

(v) Which vegetable has a larger range of weights ?

1(b). In each of the following below, please circle the correct choice. No justification is required.

(I): Consider the three histograms given below of datasets x, y and z



The ordering of the dataset from the smallest to biggest standard deviations is given by:

$(i)(\mathrm{x},\mathrm{y},\mathrm{z})$ $(ii)(\mathrm{x},\mathrm{z},\mathrm{y})$ $(iii)(\mathrm{y},\mathrm{x},\mathrm{z})$ $(iv)(\mathrm{y},\mathrm{z},\mathrm{x})$ $(v)(\mathrm{z},\mathrm{y},\mathrm{x})$ $(vi)(\mathrm{z},\mathrm{x},\mathrm{y})$

(II): Bivariate data $\{(x_i, y_i)\}_{i=1}^n$ is given to us and is assumed to arise from the model, $y_i = bx_i + \varepsilon_i$, where ε_i are random variables. For simple linear regression to be appropriate, it is sufficient to assume that

(i)ε_i are all Normal (μ_i, σ_i^2). (ii) ε_i are all Normal (μ_i, σ).

(iii)ε_i are all independent. (iv) ε_i are all independent with mean 0 and Variance σ^2.

(III): Sambhavi has just finished a two sample t-test for equality in means between populations x and y. She concludes that the null hypothesis can be rejected at a level of significance 0.05. A best possible estimate for the probability that the two datasets came from distributions having the same mean is :

(i) 16.5 (ii) $\frac{1}{19}$ (iii)$\frac{1}{21}$ (iv)$\frac{1}{20}$

2. At the ISI co-rec basketball league in the 10 games played team *Unit-disc* scored:

$$59, 62, 59, 74, 70, 61, 62, 66, 62, 75$$

Assume that the number of points scored by *Unit-disc* is Normally distributed.

(a) Compute a 95% confidence interval for the mean, μ.

(b) We want to test the null hypothesis that the $\mu = 63$ versus the alternative hypothesis that $\mu \neq 63$. Decide and execute a test that can check if there is enough evidence whether one can reject the null hypothesis at 5% level of significance.

3. Gobarkanth collects $X_1, X_2, X_3, \ldots, X_n$ of i.i.d measurements of radiation from Canteen's Gobar Gas plant. He assumes that the observations follow a Rayleigh distribution with parameter α, with p.d.f. given by

$$f(x) = \begin{cases} \alpha x \exp(-\frac{1}{2}\alpha x^2) & \text{if } x \geq 0, \\ \\ 0 & \text{otherwise.} \end{cases}$$

Find the maximum likelihood estimate for α.

4. The following R code simulates a random variable X

```
> L = 10
> i = 0
> U = runif(1, min=0, max =1)
> Y = -log(U)/L
> Sum = Y
> while (Sum<1) {
+     U = runif(1, min=0, max =1)
+     Y = -log(U)/L
+     Sum = Sum +Y
+     i = i + 1
+ }
> X = i
```

Find the distribution of X (*Other than p.d.f. or p.m.f. of standard distribution functions please provide adequate justification of any result that you are using*).

5. In an experiment in breeding plants, a geneticist has obtained 219 brown wrinkled seeds, 81 brown roundseeds, 69 white wrinkled seeds and 31 white round seeds. Theory predicts that these types of seeds should be obtained in the ratios $9:3:3:1$. Assuming that the null hypothesis is given by the theory, execute a test that can check if there is enough evidence to reject the null hypothesis at 5% level of significance.

6. The responses for three treatments A, B, C to a population of mice are given below. We wish to verify if the treatments are different or not. The data is entered and following test is performed by R.

```
> A = c(37, 39, 90, 92, 51)
> B = c(13, 17, 46, 30, 23)
> C= c(52, 25, 23, 43, 52)
> y= c(A,B,C)
> x = c(rep("A",5),rep("B",5),rep("C",5))
> oneway.test(y~x, var.equal=TRUE)


        One-way analysis of means

data:  y and x
F = 4.4826, num df = 2, denom df = 12, p-value = 0.03516
```

(a) Describe what test is being performed and what is the conclusion you can infer.

(b) Denote the data set by $y := (y_{ij})$ with $1 \leq i \leq I, 1 \leq j \leq J$. In terms of (y_{ij}), explain what are: F, num df, denom df and p-value in the above output.