## PART I

## Due Date: May 12th,2022

Problems due: 2,3,5, 7(b)ii, 8

- 1. Distinguishing between Students-t distribution and Normal distribution:
  - (a) Using rnorm and rt generate 100 samples of Normal(0, 1) and  $t_{25}$  distribution. Compare them using the inbuilt boxplot, qqnorm and qqline functions.
  - (b) Using range of [-4.4] (in same frame) plot the densities of Normal(0,1) and  $t_k$  distributions for k = 3, 33, 66 and 99 using the dnorm, dt and plot function.
- 2. Suppose we wish to test if the coin given to us is fair.
  - (a) We toss it a 100 times and find that there are 45 heads. Using the inbuilt prop.test in R, describe each output of the command prop.test(45, 100).
  - (b) Suppose we toss the coin a 10000 times and find that there are 4500 heads. Then will you conclude that the coin is fair ?
- 3. Suppose Doddapple manufactures claims that their batteries last 25 years. Students from B.Nothing Hons. sample 10 users and find the sample mean time for battery life was 21 with a sample standard deviation of 1.7. Is Doddapple claim believable ?
- 4. Super-shakti-malt is supposed to help patients recover faster from common cold. Recovery time is measured in Days. A placebo group is also used. The data is as follows:

Super-shakti-malt	:	9	1	4	1	3	2	15	3	8	2
placebo	:	9	8	6	2	8	1	10	4	9	6

- (a) Perform a Box-plot to see if the equal means and variances are valid.
- (b) Using the inbuilt t.test under the assumptions that variances are equal, decide if the means are equal. Describe each output of the command. Please explain all the inferences you can make from the output.
- (c) Using the inbuilt t.test under the assumptions that variances are unequal, decide if the means are equal. Describe each output of the command. Please explain all the inferences you can make from the output.
- (d) Explain if any the differences in the results in part (b) and part (c).
- 5. Super-hero-motor-cycle driver always believes helmets are of no use and they have no effect on injury during an accident. The super-boring-traffic-police-commissioner decides to prove such super-heros wrong. Past 5 years data is tabulated below:

	None	Slight Injury	Minor Injury	Major Injury
Helmet worn	13995	6192	6007	5804
Helmet not worn	17491	7767	7554	7187

Using the inbuilt command chisq.test decide if the super-boring-traffic-police-commissioner canclude that the two are not independent.

6. Consider Multinomial distribution with  $p_1 = \frac{1}{8}$ ,  $p_2 = \frac{1}{2}$  and  $p_3 = \frac{3}{8}$ 

(a) Simulate 100 samples from Multinomial with size 30 and compute the Pearson- $\chi^2$  statistic

$$\mathbf{X}^2 := \sum_j \frac{(\text{Observed} - \text{Expected})^2}{\text{Expected}}$$

via an R-code for each repetition, saving it in a variable Xsquared.

- (b) On the same graph: plot the histogram of Xsquared and chi-squared density with 3 degrees of freedom.
- 7. Two sample test of means for paired data Let  $n \ge 1, X_1, X_2, \ldots, X_n$  and  $Y_1, Y_2, \ldots, Y_n$  betwo samples, that are paired. Then the test Statistic is given by:

$$T := \frac{\bar{X} - \bar{Y} - (\mu_x - \mu_Y)}{\frac{S}{\sqrt{n}}} \quad \text{where} \quad S^2 = \frac{1}{n-1} (\sum_{i=1}^n (z_i - \bar{z})^2) \quad \text{with } z_i = x_i - y_i$$

with  $T \sim t_{n-1}$ 

- (a) Write a function called pairedttest that performs the above test for paired data x and y.
- (b) Consider the shoes dataset from the package MASS. It contains shoe wear data. There is a list of two vectors, giving the wear of shoes of materials A and B for one foot each of ten boys.
  - i. Using the function pairedttest perform the paired t-test on the above data to see if the two types of shoes have different mean wear amounts.
  - ii. Verify the same using the inbuilt function t.test(A,B, paired=TRUE).
  - iii. If you were to use the command t.test without the paired=TRUE command then can you still perform the above test ?
- 8. At the ISI student mess the cook has invented a new hot drink, and would like to find out if it will be as popular as the existing Tea. For this purpose, the students arrange 18 participants for taste testing. Each participant tries both drinks in random order before giving his or her opinion. It turns out that 5 of the participants like the new drink better, and the rest prefer the existing Tea. Using the inbuilt sign test function binom.test, decide if we can reject the notion that the two drinks are equally popular at 0.05 significance level ?

## PART II

Due Date: May 19th,2022

Problems due: 1,2,5,8

1. Suppose Somadev finds that his weight in kgs during each month of year to be

757673 7574737376737977 75

- (a) Write a function called zcinf that takes in the weights above as a vector x, assumes a known standard deviation of 1.5 and produces default 95% confidence interval.
- (b) Write a function called tcinf that takes in the weights above as a vector x, assumes that variance is unknown and produces default -95% confidence interval.
- (c) Use the inbuilt t.test command on the vector x (as above), describe each output of the command t.test(x) Please explain all the inferences you can make from the output.
- 2. We are given two dice. We roll each of them 500 times and the outcomes are summarised below. Dice 1 Dice 2

Top Face	:	1	2	3	4	5	6	Top Face	:	1	2	3	4	5	6
Number	:	77	89	76	84	80	94	Number	:	50	199	102	52	44	53

- (a) Using the inbuilt chisq.test, decide if Dice 1 is fair or not. Describe each output of the command. Please explain all the inferences you can make from the output.
- (b) Using the inbuilt chisq.test, decide if Dice 2 is fair or not. Describe each output of the command. Please explain all the inferences you can make from the output.
- (c) Using the inbuilt chisq.test, decide if Dice(s) appear to be have the same distribution. Describe each output of the command. Please explain all the inferences you can make from the output.
- 3. The students hostel october mela has a game involves rolling 3 dice. The winnings are directly proportional to the total number of ones rolled. Suppose Jooa brings her own set of dice and plays the game 100 times. Her results are tabulated below:
  - Number of ones Number of Rolls
  - 0 40

1 37 Student's supreme leader, Moola, gets suspicious. He wishes  $\mathbf{2}$ 13 3 10

to determine if the dice are fair or not. Let us help him with the  $\chi^2$ -square goodness of fit test.

- (a) Compute the respective Multinomial probabilities in R.
- (b) Plot the barplot of the observed counts and expected counts.
- (c) Plot the barplot of the <u>observed count- expected count</u>
- $\sqrt{\text{expected count}}$
- (d) Compute the chi-square statistic and decide if the null hypothesis that the dice are fair can be rejected or not?
- 4. Sive decides suddenly that fairness is important when it comes to grading. He mandates that each quiz is graded twice by Siva and Sarvesh. The data is as follows

Siva	:	30	0	50	22	55	50	55	40	44	60
Sarvesh	:	20	10	40	11	44	30	33	20	33	60

- (a) Using the t.test, perform the paired t-test to see if the scores are statistically different.
- (b) Suppose we assume that two sets of students are independent, decide if the scores are statistically different.
- 5. The student body at an undergraduate university is 20% Masters, 24% third years, 26% Second year, and 30% first year students. Suppose a researcher takes a sample of 50 such students. Within the sample there are 13 Masters, 16 Third years, 10 Second years, and 11 First years. The researcher claims that his sampling procedure should have produced independent selections from the student body, with each student equally likely to be selected. Is this a plausible claim given the observed results?
- 6. Let X be a random variable with finite range  $\{c_1, c_2\}$  for which  $P(X = c_j) = p_j > 0$  for j = 1, 2. Let  $X_1, X_2, \ldots, X_n$  be an i.i.d. sample with distribution X and let  $Y_j = |\{k : X_k = c_j\}|$  for j = 1, 2. Then  $\chi^2 = \sum_{j=1}^2 \frac{(Y_j - np_j)^2}{np_j}$  has the same distribution as  $(\frac{Z - E[Z]}{SD[Z]})^2$  where  $Z \sim \text{Binomial}(n, p_1)$
- 7. In an experiment in breeding plants, a geneticist has obtained 120 brown wrinkled seeds, 48 brown round seeds, 36 white wrinkled seeds and 13 white round seeds. Theory predicts that these types of mice should be obtained in the ratios 9:3:3:1. Is the theory a valid hypothesis ?
- 8. Suppose a study of computer usage of Mahacomp for the time he plays video games:

11.5, .5, .9, .4, 7.8, 7, .2, 2.5, 0.9, 2, 3, 15

- (a) Plot the histogram of the data and obtain a guess for the median of the data set.
- (b) use the following R-code wilcox.test(x,mu=5, alt="less") to decide if the distribution is centered around 5 against the alternative that it is centered less than 5.