Prashant Sharma

Grading: 20 marks- Complete submission of worksheet14
40 marks- Problem 1 and 40 marks- Problem 2

Problem:1

The data is given as follows:
Prashant : 30 0 50 22 55 50 55 40 44 60
Ishan : 20 10 40 11 44 30 33 20 33 60

Problem:1(a)

We have to perform the paired t-test to see if the scores are statistically different, by using the inbuilt t.test() command.
For which the R-code is as follows:

```
> Pra_score = c(30,0,50,22,55,50,55,40,44, 60)
> Ishan_score = c(20,10,40,11,44,30,33,20,33,60)
> t.test(Pra_score,Ishan_score,paired = TRUE)


        Paired t-test

data:  Pra_score and Ishan_score
t = 3.4278, df = 9, p-value = 0.007535
alternative hypothesis: true difference in means is not equal to 0
95 percent confidence interval:
  3.570509 17.429491
sample estimates:
mean of the differences
                  10.5
```

In the paired t-test for the data; The Null Hypothesis,H_0: true difference between means of two sets of grades is 0 against
The alternative Hypothesis, H_1: true difference between means of two sets of grades is not equal to 0.
Here the $p-value$ of the test is 0.007535 which is less than 0.05. Therefore, we reject the null hypothesis at 5% level of significance. So, we can infer that the true mean difference of scores is not 0. Hence, the scores are statistically different.

Problem:1(b)

Now, we assume that two sets of students are independent. In this case we have to decide if the scores are statistically different.
As in the case of independent, data will not be paired.
Now, the R-code will be as follows:

```
> t.test(Pra_score,Ishan_score, paired = FALSE)


        Welch Two Sample t-test

data:  Pra_score and Ishan_score
t = 1.3733, df = 17.475, p-value = 0.187
alternative hypothesis: true difference in means is not equal to 0
95 percent confidence interval:
 -5.597945 26.597945
sample estimates:
mean of x mean of y
     40.6      30.1
```

As the $p-value$ for this test is 0.187 which is greater than 0.05. Thus, we accept the null hypothesis and conclude that there is no statistical difference between the two scores.

Problem:2

We are given two dice. We roll each of them 500 times and the outcomes are summarized below.
$Dice-1$
Top Face :1 2 3 4 5 6
Number : 77 89 76 84 80 94
$Dice-2$
Top Face :1 2 3 4 5 6
Number : 50 199 102 52 44 53

Problem:2(a)

We have to decide if $Dice-1$ is fair or not, by using the inbuilt chisq.test
A dice is fair if the probability of each outcome is $\frac{1}{6}$. The null hypothesis is that the die is fair.
For this purpose, R-code is as follows:

```
> dice1=c(77, 89, 76, 84, 80, 94)
> chisq.test(dice1)


        Chi-squared test for given probabilities

data:  dice1
X-squared = 3.016, df = 5, p-value = 0.6975
```

The Null Hypothesis,H_0: the dice is fair i.e., all the 6 sides have equal probability of $\frac{1}{6}$.
against
alternative Hypothesis, H_1: the dice is not fair
Here, the value of chi-squared test statistic under null hypothesis is 3.016 and df = degrees of freedom = $6-1=5$.
As we know that, p-value is the probability of obtaining the test result as extreme the result is actually

observed under the assumption null hypothesis is true.
Here, we can see that for this test $p - value = 0.6975$ which is greater than 0.05. Therefore, we accept the null hypothesis that is there is no significant difference between the observed and expected value at 5% level of significance. Hence, conclude that dice-1 is fair.

Problem:2(b)

We have to decide if $Dice - 1$ is fair or not, by using the inbuilt chisq.test
For this purpose, R-code is as follows:

```
> dice2 = c(50, 199, 102, 52, 44, 53)
> chisq.test(dice2)
```

```
        Chi-squared test for given probabilities

data:  dice2
X-squared = 219.45, df = 5, p-value < 2.2e-16
```

The Null Hypothesis,H_0: the dice is fair i.e., all the 6 sides have equal probability of $\frac{1}{6}$.
against
alternative Hypothesis, H_1: the dice is not fair
Here, the value of chi-squared test statistic under null hypothesis is 219.45 and df = degrees of freedom $= 6 - 1 = 5$.
As we know that, p-value is the probability of obtaining the test result as extreme the result is actually observed under the assumption null hypothesis is true.
Here, we can see that for this test $p - value < 2.2e - 16$ which is less than 0.05. Therefore, we reject the null hypothesis that is there is significant difference between the observed and expected value at 5% level of significance. Hence, conclude that dice-2 is not fair.

Problem:2(c)

we have to check if the Dice-1, Dice-2 appear to be have same distribution, by using the inbuilt chisq.test function in R . It is the chi-square test of homogeneity.
For which, R-code is as follows:

```
> df=data.frame(dice1,dice2)
> chisq.test(df)
```

```
        Pearson's Chi-squared test

data:  df
X-squared = 80.968, df = 5, p-value = 5.263e-16
```

To test,
The null Hypothesis,H_0: The two dices appear to have the same distribution against
alternative Hypothesis,H_1: two dices appear to have different distribution.
Here, X-squared = the value of chi-squared test statistic under null hypothesis is 80.968. and df = degrees of freedom = $(6-1)(2-1) = 5$
Here, $p - value = 5.263e - 16$ which is less than 0.05. Thus, we reject null hypothesis and we may infer that the Dice-1, Dice-2 don't appear to be have same distribution.