STAT 333

Nonparametric Statistics Methods

Prepared by

Abdulrahman Alfaifi
alfaifi@ksu.edu.sa

King Saud University
Department of Statistics and Operation Research

3 Feb 2026

بِسْمِ ٱللَّهِ ٱلرَّحْمَٰنِ ٱلرَّحِيمِ

Chapter 1

INTRODUCTION

- Some types of tests presented and their parametric counterparts.

Type of analysis	Nonparametric test	Parametric equivalent
Comparing two related samples	Wilcoxon signed ranks test and sign test	*t*-Test for dependent samples
Comparing two unrelated samples	Mann–Whitney *U*-test and Kolmogorov–Smirnov two-sample test	*t*-Test for independent samples
Comparing three or more related samples	Friedman test	Repeated measures, analysis of variance (ANOVA)
Comparing three or more unrelated samples	Kruskal–Wallis *H*-test	One-way ANOVA
Comparing categorical data	Chi square χ^2 tests and Fisher exact test	None
Comparing two rank-ordered variables	Spearman rank-order correlation	Pearson product–moment correlation
Comparing two variables when one variable is discrete dichotomous	Point-biserial correlation	Pearson product–moment correlation
Comparing two variables when one variable is continuous dichotomous	Biserial correlation	Pearson product–moment correlation
Examining a sample for randomness	Runs test	None

Parametric tests rely on six main assumptions. If these assumptions are not met, the results may be misleading, and nonparametric tests should be used instead.

1. The data are randomly drowned from a normally distributed population.
2. Observations are independent of each other, expect for paired values.
3. The data are measured on an interval or ratio scale.
4. The respective populations are approximately equal variances.
5. The sample size is sufficiently large.
6. The population distribution is approximately normal.

Exercise 1:

1. Which of the following is not true of parametric statistics?

A	They are inferential tests.
B	They assume certain characteristics of population parameters.
C	They assume normality of the population.
D	They are distribution-free.

2. A collection of statistical methods that generally requires very few, if any assumptions about the population distribution is known as……………………………

A	Parametric methods	B	Nonparametric methods
C	Semiparametric	D	None of these

3. A nonparametric method for determining the differences between two populations based on two matched samples where only preference data is required is the

A	Mann-Whitney-Wilcoxon test	B	Wilcoxon signed-rank test
C	Sign test	D	Kruskal-Wallis Test

4. Parametric tests are based on some restrictive assumptions about the ………………

A	Random sample	B	Census	C	Sample	D	Population

5. Nonparametric tests for examining a sample for randomness ………………

A	Fridman test	B	Runs test	C	T - test	D	U - test

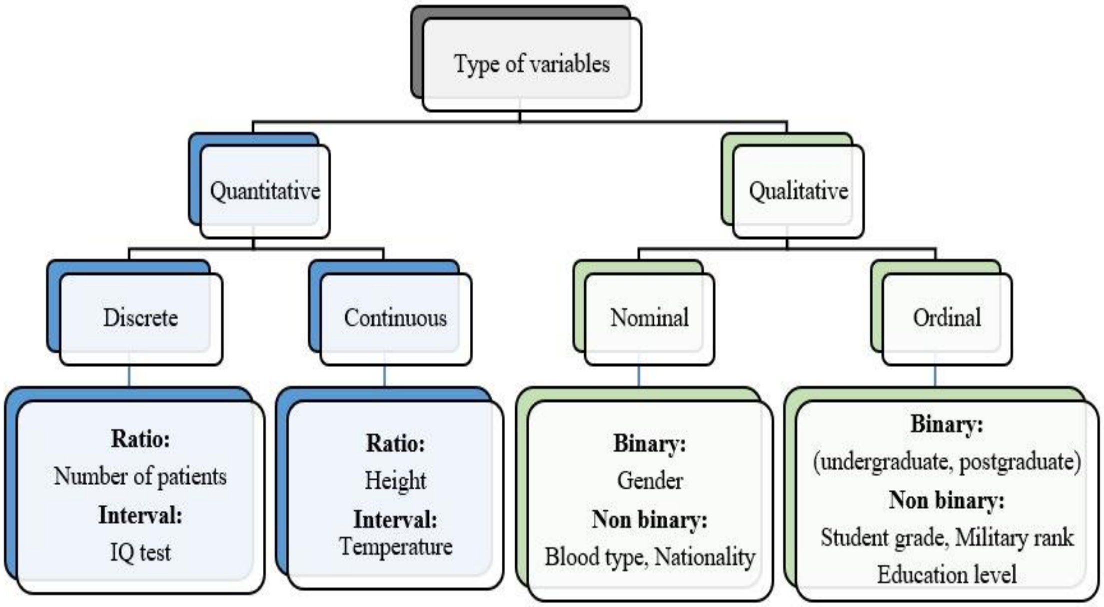
6. Point- biserial correlation is used for ………………

A	Comparing two rank-ordered variables.
B	Comparing two variables when one variable is discrete dichotomous.
C	Comparing two variables when one variable is continuous dichotomous.
D	Comparing two related samples.

7. Parametric test equivalent to Kruskal–Wallis *H*-test is: ………………

A	One-way ANOVA	B	Repeated measures
C	T-test for dependent samples	D	Fridman test

- **Measurement scales:**



<u>Interval scale:</u> Consider as pertinent information not only the relative order of the measurements as in the ordinal scale but also the size of the interval between measurements.

For example:

- Temperature.
- TQ-test.
- Time of the day (00:00 midnight, 14:00 afternoon)

<u>Ratio scale:</u> Not only the order and interval size are important, but also the ratio between two measurements is meaningful.

For example:

- Crop yields (إنتاج المحاصيل).
- Distances.
- Time to finish an exam.
- Weights.
- Heights.
- Income.

Exercise 2: Choose the correct measurement scale.

1. Gender (Male, Female) A) Nominal B) Ordinal C) Interval D) Ratio	8. Height in centimeters A) Ordinal B) Interval C) Ratio D) Nominal
2. Blood type (A, B, AB, O) A) Ordinal B) Nominal C) Ratio D) Interval	9. Employee ID number A) Interval B) Ratio C) Nominal D) Ordinal
3. Satisfaction level (Low, Medium, High) A) Nominal B) Ordinal C) Interval D) Ratio	10. Calendar year (2015, 2020, 2025) A) Ratio B) Ordinal C) Interval D) Nominal
4. Age in years A) Nominal B) Ordinal C) Interval D) Ratio	11. Marital status (Single, Married, Divorced) A) Ordinal B) Interval C) Ratio D) Nominal
5. Temperature in Celsius (°C) A) Nominal B) Ordinal C) Interval D) Ratio	12. Pain level (Mild, Moderate, Severe) A) Nominal B) Ordinal C) Interval D) Ratio
6. Exam grades (A, B, C, D) A) Ratio B) Interval C) Ordinal D) Nominal	13. Distance traveled (km) A) Ordinal B) Interval C) Nominal D) Ratio
7. Number of students in a class A) Nominal B) Ordinal C) Interval D) Ratio	14. Eye color A) Ratio B) Interval C) Ordinal D) Nominal

- **Ranking data:**

Example: Rank the following data:

Students who ate breakfast	Students who skipped breakfast
90	75
85	80
95	55
70	90

After ordering	Rank ignoring ties values		Rank accounting for ties values
55	1		1
70	2		2
75	3		3
80	4		4
85	5		5
90	6	$\frac{6+7}{2} = 6.5$	6.5
90	7		6.5
95	8		8

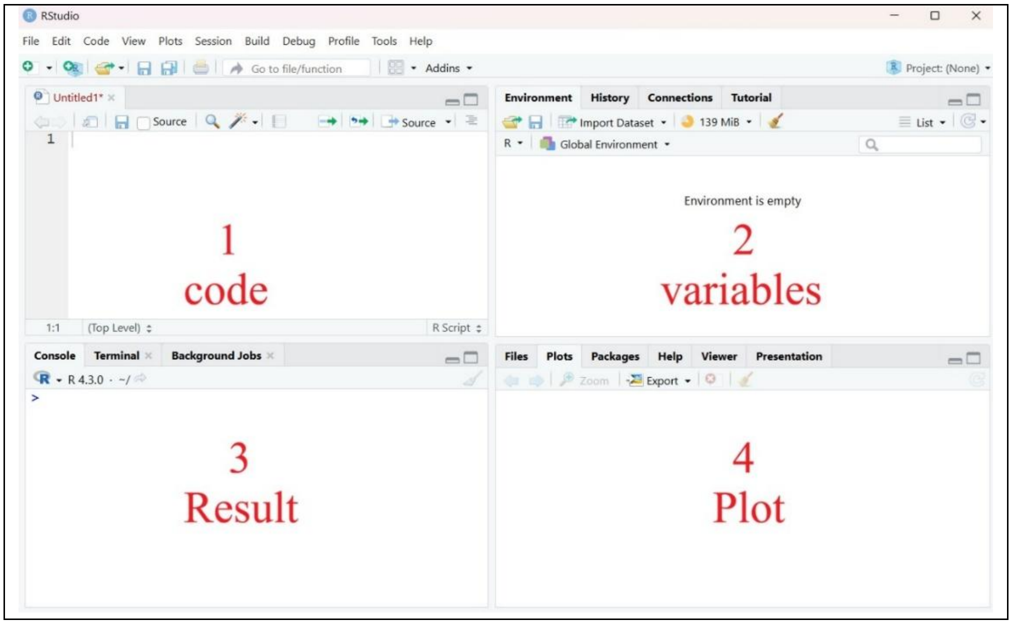
Example: the following data represent quiz score om math.

100	60	70	90	80	100	80	20	100	50

Rank the quiz score.

Quiz score	After ordering	Rank ignoring ties values		Rank accounting for ties values
100	20	1		1
60	50	2		2
70	60	3		3
90	70	4		4
80	80	5	$\frac{5+6}{2} = 5.5$	5.5
100	80	6		5.5
80	90	7		7
20	100	8		9
100	100	9	$\frac{8+9+10}{3} = 9$	9
50	100	10		9

Using R studio:

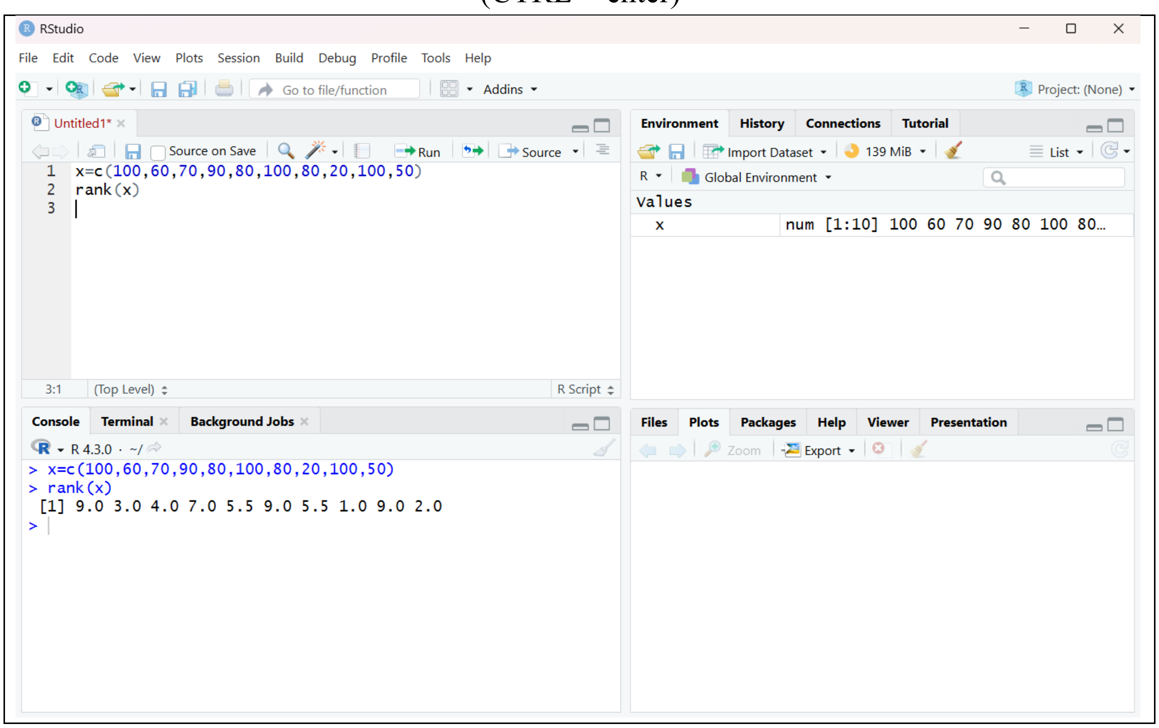


في حال ظهرت 3 شاشات إضغط (CTRL + shift + N)

R - code

```
> x=c(100,60,70,90,80,100,80,20,100,50)
> rank(x)
```

(CTRL + enter)



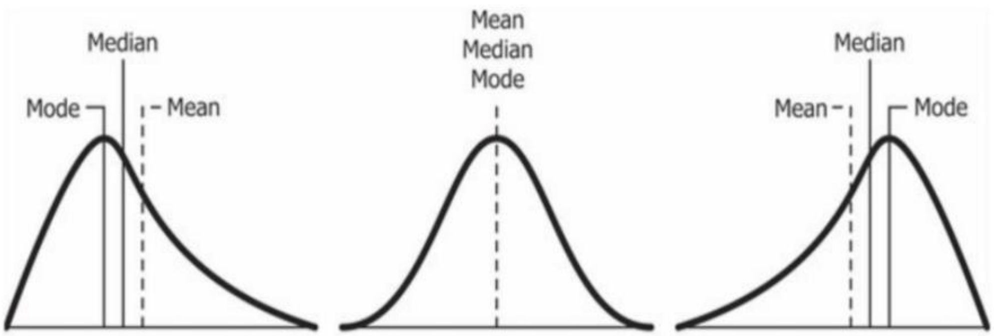
Chapter 2

TESTING DATA FOR NORMALITY

- **Skewness and Kurtosis:**

	Skewness:	Kurtosis:
Formula	$S_k = \frac{n}{(n-1)(n-2)}\sum\left(\frac{x_i-\bar{x}}{s}\right)^3$	$K = \left[\frac{n(n+1)}{(n-1)(n-2)(n-3)}\sum\left(\frac{x_i-\bar{x}}{s}\right)^4\right] - \frac{3(n-1)^2}{(n-2)(n-3)}$ $s = \sqrt{\frac{\sum(x_i-\bar{x})^2}{n-1}}$
Standard error (SE)	$SE_{S_k} = \sqrt{\frac{6n(n-1)}{(n-2)(n+1)(n+3)}}$	$SE_K = \sqrt{\frac{24n(n-1)^2}{(n-2)(n-3)(n+5)(n+3)}}$
Z - score	$Z_{S_k} = \frac{S_k-0}{SE_{S_k}}$ If $Z_{S_k} \in (-1.96, 1.96)$ pass the normality assumption	$Z_k = \frac{K-0}{SE_K}$ If $Z_k \in (-1.96, 1.96)$ pass the normality assumption

Skewness:

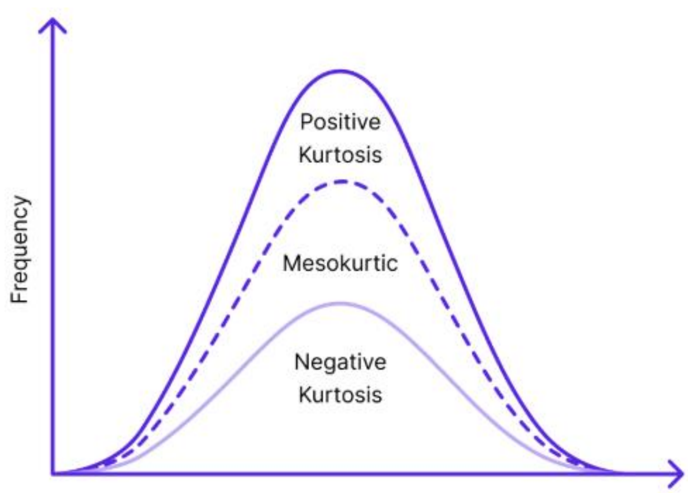


Positive skew
Right skew

No skewness

Negative skew
Left skew

Kurtosis:



Positive kurtosis

Kurtosis = 0

Negative kurtosis

Exercise 1:

The following data represent a samples of week 1 quiz score.
Calculate the skewness and kurtosis.

90	72	90
64	95	89
74	88	100
77	57	35
100	64	95
65	80	84
90	100	76

	data	$x_i - \bar{x}$	$(x_i - \bar{x})^2$	$\left(\frac{x_i-\bar{x}}{S}\right)^3$	$\left(\frac{x_i-\bar{x}}{S}\right)^4$
1	90	9.761905	95.29478	0.202561	0.118962
2	64	-16.2381	263.6757	-0.9323	0.91077
3	74	-6.2381	38.91383	-0.05286	0.019837
⋮	⋮	⋮	⋮	⋮	⋮
19	95	14.7619	217.9138	0.700452	0.622068
20	84	3.761905	14.15193	0.011592	0.002624
21	76	-4.2381	17.96145	-0.01658	0.004226
Total	1685		5525.81	-18.4149	69.01972

$$\bar{x} = \frac{\sum x}{n} = \frac{1685}{21} = 80.2381$$

$$S = \sqrt{\frac{\sum(x_i-\bar{x})^2}{n-1}} = \sqrt{\frac{5525.80}{20}} = 6.62199$$

$$\sum \left(\frac{x_i-\bar{x}}{S}\right)^3 = -18.4149$$

$$\sum \left(\frac{x_i-\bar{x}}{S}\right)^4 = 69.01972$$

$$S_k = \frac{n}{(n-1)(n-2)} \sum \left(\frac{x_i-\bar{x}}{S}\right)^3$$

$$= \frac{21}{(21-1)(21-2)} \times -18.4149 = -1.01766$$

$$SE_{S_k} = \sqrt{\frac{6n(n-1)}{(n-2)(n+1)(n+3)}} = \sqrt{\frac{6\times 21(21-1)}{(21-2)(21+1)(21+3)}} = 0.50119$$

$$Z_{S_k} = \frac{S_k - 0}{SE_{S_k}}$$

$$= \frac{-1.01766-0}{0.50119} = -2.032$$

$$Z_{S_k} \notin (-1.96, 1.96)$$

$$K = \left[\frac{n(n+1)}{(n-1)(n-2)(n-3)} \sum \left(\frac{x_i-\bar{x}}{S}\right)^4\right] - \frac{3(n-1)^2}{(n-2)(n-3)}$$

$$= \left[\frac{21(21+1)}{(21-1)(21-2)(21-3)} \times 69.02\right] - \frac{3(21-1)^2}{(21-2)(21-3)} = 1.153$$

$$SE_K = \sqrt{\frac{24n(n-1)^2}{(n-2)(n-3)(n+5)(n+3)}}$$

$$= \sqrt{\frac{24\times 21(21-1)^2}{(21-2)(21-3)(21+5)(21+3)}} = 0.971941$$

$$Z_k = \frac{K - 0}{SE_K}$$

$$= \frac{1.153-0}{0.971941} = 1.186$$

$$Z_k \in (-1.96, 1.96)$$

R – code

```
x=c(90,72,90,64,95,89,74,88,100,77,57,35,100,64,95,65,80,84,90,100,76)
m=mean(x)
s=sd(x)
n=length(x)

i3=sum(((x-m)/s)^3)
i4=sum(((x-m)/s)^4)

sk=n/((n-1)*(n-2))*i3
SEs=sqrt(6*n*(n-1)/(n-2)/(n+1)/(n+3))

kur=(n*(n+1)/(n-1)/(n-2)/(n-3)*i4)-(3*(n-1)^2/(n-2)/(n-3))
SEk=sqrt((24*n*(n-1)^2)/(n-2)/(n-3)/(n+5)/(n+3))
```

```
x=c(90,72,90,64,95,89,74,88,100,77,57,35,100,64,95,65,80,84,90,100,76)
m=mean(x)
s=sd(x)
```

$$S = \sqrt{\frac{\sum (x_i - \bar{x})^2}{n-1}}$$

```
n=length(x)
i3=sum(((x-m)/s)^3)
```

$$\sum \left(\frac{x_i - \bar{x}}{s}\right)^3$$

```
i4=sum(((x-m)/s)^4)
```

$$\sum \left(\frac{x_i - \bar{x}}{s}\right)^4$$

```
sk=n/((n-1)*(n-2))*i3
```

$$S_k = \frac{n}{(n-1)(n-2)} \sum \left(\frac{x_i - \bar{x}}{s}\right)^3$$

```
SEs=sqrt(6*n*(n-1)/(n-2)/(n+1)/(n+3))
```

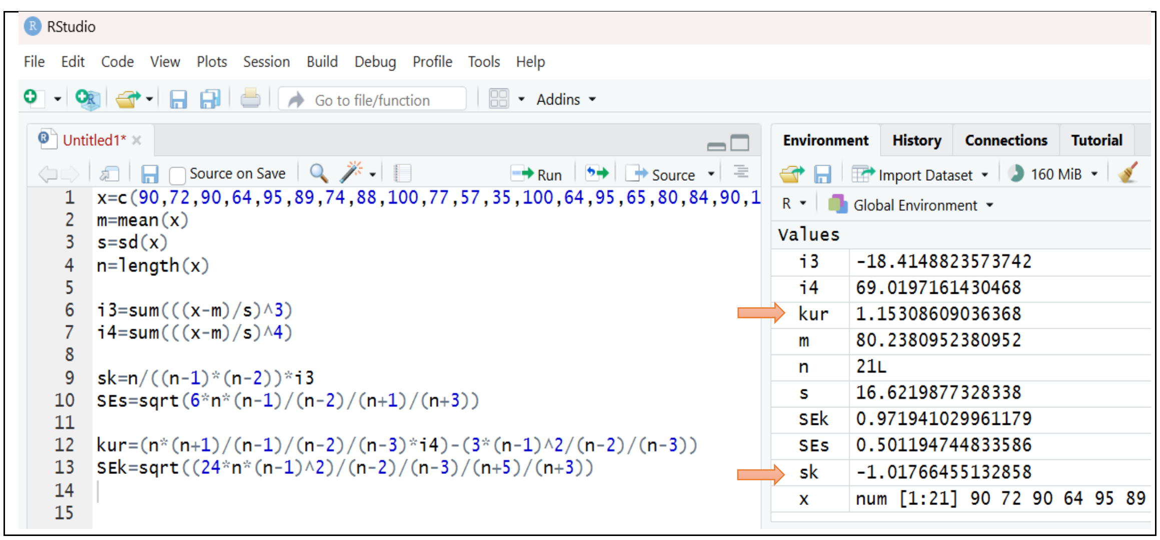
$$SE_{S_k} = \sqrt{\frac{6n(n-1)}{(n-2)(n+1)(n+3)}}$$

```
kur=(n*(n+1)/(n-1)/(n-2)/(n-3)*i4)-(3*(n-1)^2/(n-2)/(n-3))
```

$$K = \left[\frac{n(n+1)}{(n-1)(n-2)(n-3)} \sum \left(\frac{x_i - \bar{x}}{s}\right)^4\right] - \frac{3(n-1)^2}{(n-2)(n-3)}$$

```
SEk=sqrt((24*n*(n-1)^2)/(n-2)/(n-3)/(n+5)/(n+3))
```

$$SE_K = \sqrt{\frac{24n(n-1)^2}{(n-2)(n-3)(n+5)(n+3)}}$$



Exercise 2:

A department store has decided to evaluate customer satisfaction. The store provides customers with a survey to rate employee friendliness. The survey uses a scale of 1–10. The survey results are:

7	3	3	6
4	4	4	5
5	5	8	9
5	5	5	7
6	8	6	2

Calculate the skewness and kurtosis.

	data	$x_i - \bar{x}$	$(x_i - \bar{x})^2$	$\left(\frac{x_i-\bar{x}}{s}\right)^3$	$\left(\frac{x_i-\bar{x}}{s}\right)^4$
1	7				
2	4				
3	5				
⋮	⋮	⋮	⋮	⋮	⋮
18	9				
19	7				
20	2				
Total	107		62.55	4.09576	45.0707

$\bar{x} =$ ……………………………

$s =$ ……………………………

$\sum \left(\frac{x_i-\bar{x}}{s}\right)^3 =$ ………………

$\sum \left(\frac{x_i-\bar{x}}{s}\right)^4 =$ ………………

$$S_k = \frac{n}{(n-1)(n-2)} \sum \left(\frac{x_i-\bar{x}}{s}\right)^3 = \text{……………………………}$$

$$SE_{S_k} = \sqrt{\frac{6n(n-1)}{(n-2)(n+1)(n+3)}} = \text{……………………………}$$

$$Z_{S_k} = \frac{S_k - 0}{SE_{S_k}} =$$

$$K = \left[\frac{n(n+1)}{(n-1)(n-2)(n-3)} \sum \left(\frac{x_i-\bar{x}}{s}\right)^4\right] - \frac{3(n-1)^2}{(n-2)(n-3)} =$$

……………………………………………………………………

$$SE_K = \sqrt{\frac{24n(n-1)^2}{(n-2)(n-3)(n+5)(n+3)}} = \text{……………………………}$$

$$Z_k = \frac{K - 0}{SE_K} =$$

R – code

```
x=c(7,4,5,5,6,3,4,5,5,8,3,4,8,5,6,6,5,9,7,2)
```

Exercise 3:

Calculate the skewness and kurtosis for the following data.

	data	$x_i - \bar{x}$	$(x_i - \bar{x})^2$	$\left(\frac{x_i-\bar{x}}{S}\right)^3$	$\left(\frac{x_i-\bar{x}}{S}\right)^4$
1	25				
2	30				
3	12				
4	18				
5	20				
Total					

$\bar{x} =$ ....

$S =$ ....

$\sum \left(\frac{x_i-\bar{x}}{S}\right)^3 =$ ....

$\sum \left(\frac{x_i-\bar{x}}{S}\right)^4 =$ ....

$S_k =$ ....
....
....

$Z_{S_k} = \frac{S_k - 0}{SE_{S_k}} =$

$SE_{S_k} =$ ....
....
....

$K =$ ....
....
....

$Z_k = \frac{K - 0}{SE_K} =$

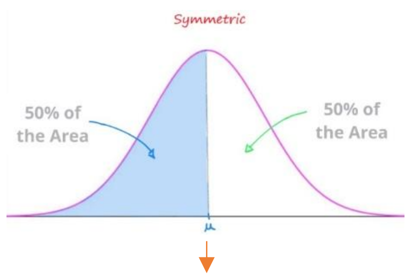
$SE_K =$ ....
....
....

R – code

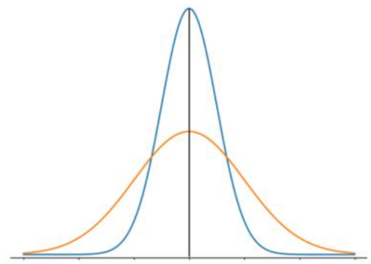
```
x=c(25,30,12,18,20)
```

The Normal Distribution

$$f(x) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2}; \quad -\infty < X < \infty$$



Mean = Median = Mode



μ

$\sigma > \sigma$

Normal distribution	$X \sim N(\mu, \sigma^2)$	
Standard normal	$Z \sim N(0, 1)$	$Z = \frac{X-\mu}{\sigma} \sim N(0, 1)$

Testing Data for Normality: Kolmogorov–Smirnov

H_0: The data approximately follow normal distribution.
H_A: The data do not follow normal distribution.

Question: Suppose that we have a random sample of size n. For α = 0.05, if the Z-score of the skewness of the sample is (−2.032), then the sample has Therefore, either the sample must be modified and rechecked or you must use a nonparametric statistical test.

A	Pass the normality assumption for kurtosis
B	Pass the normality assumption for skewness.
C	Failed the normality assumption for kurtosis
D	Failed the normality assumption for skewness.

Exercise 4:

For the following data:

8.1	8.2	8.2	8.7	8.7	8.8	8.8	8.9	8.9	8.9
9.2	9.2	9.2	9.3	9.3	9.3	9.4	9.4	9.4	9.4
9.5	9.5	9.5	9.5	9.6	9.6	9.6	9.7	9.7	9.9

(a) Find: Skewness, Standard error of the skewness, Kurtosis, Standard error of the kurtosis

Using R

```
y=c(8.1,9.2,9.5,8.2,9.2,9.5,8.2,9.2,9.5,8.7,9.3,9.5 ,8.7,9.3,9.6,8.8,9.3,9.6,8.8,9.4,9.6,8.9,9.4,9.7
        ,8.9,9.4,9.7,8.9,9.4,9.9)
m=mean(y)
s=sd(y)
n=length(y)
i3=sum(((y-m)/s)^3)
i4=sum(((y-m)/s)^4)
sk=n/((n-1)*(n-2))*i3
SEs=sqrt(6*n*(n-1)/(n-2)/(n+1)/(n+3))
kur=(n*(n+1)/(n-1)/(n-2)/(n-3)*i4)-(3*(n-1)^2/(n-2)/(n-3))
SEk=sqrt((24*n*(n-1)^2)/(n-2)/(n-3)/(n+5)/(n+3))

> sk
[1] -0.9043788
> SEs
[1] 0.4268924
> kur
[1] 0.1877582
> SEk
[1] 0.8327456
```

(b) Using a Kolmogorov–Smirnov one-sample test, is the data follow normal distribution

Using R

```
ks.test(y, "pnorm", mean = mean(y), sd = sd(y))
            Asymptotic one-sample Kolmogorov-Smirnov test
data:  y
D = 0.18377, p-value = 0.263 ⇐
alternative hypothesis: two-sided
```

H_0: The data approximately follow normal distribution.
H_A: The data do not follow normal distribution.

$P - value = 0.263 > 0.05$,
We accept H_0, The data approximately follow normal distribution

Exercise 5:

A department store has decided to evaluate customer satisfaction. The store provides customers with a survey to rate employee friendliness. The survey uses a scale of 1–10. The survey results are:

7	3	3	6
4	4	4	5
5	5	8	9
5	5	5	7
6	8	6	2

Use the Kolmogorov–Smirnov one-sample test to decide if survey results approximately matching a normal distribution.

Using R

```
x=c(7,4,5,5,6,3,4,5,5,8,3,4,8,5,6,6,5,9,7,2)
ks.test(y, "pnorm", mean = mean(x), sd = sd(x))
        Asymptotic one-sample Kolmogorov-Smirnov test

data:  x
D = 0.17648, p-value = 0.5617
alternative hypothesis: two-sided
```

H_0: The data approximately follow normal distribution.
H_A: The data do not follow normal distribution.

$P - value = 0.5617 > 0.05$,
We accept H_0, The data approximately follow normal distribution

Chapter 3

THE WILCOXON SIGNED RANK AND THE SIGN TEST

In testing for the difference between two populations, it is possible to use …………

A	The Wilcoxon Rank-Sum test	B	The Sign test
C	Either of (A) or (B)	D	None of these

Consider the following two independent samples:

Sample A	15	17	18				
Sample B	14	16	19	19	20	22	23

The value of the test statistics for a right-tail Wilcoxon rank test is:

A	3	B	7	C	11	D	44

Consider a clinical investigation to assess the effectiveness of a new drug designed to reduce repetitive behaviors in children affected with autism. The data are shown below.

Child	1	2	3	4	5	6	7	8	9	10
Before Treatment	30	56	48	47	43	45	36	44	44	40
After 2 Weeks of Treatment	39	46	37	44	32	39	41	40	38	46

Use a one-tailed Wilcoxon signed rank test and a one-tailed sign test to assess the effectiveness of the drug (is there differences in behavior before and after taking the drug?). Use $\alpha = 0.05$.

Chapter 4

THE MANN–WHITNEY U-TEST AND THE KOLMOGOROV–SMIRNOV TWO-SAMPLE TEST

For each of the following questions (1-4), determine which would be the simplest type of statistical analysis that would be appropriate to use. Use each type of analysis only once.

(A) Paired t test	(B) Two sample t-test	(C) ANOVA
(D) Kruskal-Wallis	(E) Wilcoxon Rank-Sum Test	

Compare the average number of hours per week spent on Facebook for Freshmen, Sophomore, Juniors and Seniors at UF, based on a random sample of 100 students.	(C) ANOVA
Compare the average number of hours per week spent on Facebook during the first week in April and the first week in May (finals week) for random students at UF, measured on the same 100 students.	(A) Paired t test
Compare the distribution of the number of hours per week spent on Facebook for male and female students at UF, based on a random sample of 10 students. There was an outlier in one of the groups.	(E) Wilcoxon Rank-Sum Test
Compare the average number of hours per week spent on Facebook for male and female students at UF, based on a random sample of 100 students.	(B) Two sample t-test

The following data were obtained from a reading-level test for 1st-grade children. Compare the performance gains of the two different methods for teaching reading. Two different classes being taught a basic mathematics skills using two different methods.

Gain score (Method 1)	16	13	16	16	13	9	12	12	20	17
Gain score (Method 1)	11	2	10	4	9	8	5	6	4	16

Use two-tailed Mann–Whitney U and Kolmogorov–Smirnov two-sample tests to determine which method was better for teaching reading. Set $\alpha = 0.05$.

[1] The hypothesis associated with this test ……………………………
[2] The calculated value of the test statistic is ………………………...
[3] The critical value ……………..

The Mann-Whitney U test is preferred to a t-test when ……………………………

A	Data are paired	B	Sample sizes are small
C	Sample are dependent	D	The assumption of normality is not met

Chapter 5

THE FRIEDMAN TEST