Abe Mirza **Topics Review** **Statistics**

Part I

Topics	Page
Learning Objectives	2
How to study	2
General Introduction	3
Qualitative Data	5
Descriptive Statistics	6
Empirical Rules and Z -Score	10
Grouped Data (Freq. Table)/ Histogram	11
Grouped Data (Freq. Table) Mean/St. Deviation	12
Correlation and Regression	14
Steps to do a Correlation and Regression problem	15
Basic Probability	22
Multiplication Rule	24
Probability Practice Problems	27
Formula Sheet	30

Quizzes on Part 1

You are allowed to use formula sheet on the quiz and the test.

Quiz # 1: Study pages **3 through 7** of topics review (not measure of variation) and do related practice problems

Quiz # 2: Study pages **4 through 13** of topics review (not z-score on page 10) and do related practice problems
.

Quiz # 3: Study pages **14 through 22** of topics review and do related practice problems
.

Quiz # 4: Study pages **23 through 28** of topics review, also z-score on page 10 and do related practice problems
.

Learning Objectives

Definition of Statistics (**CODA**)
Difference between **Population** and **Sample**. Difference between **Parameter** and **Statistic** (not **Statistics**)
When to use the right notation (Greek or Lower case English) for **Parameter** and **Statistic**
Difference between **Qualitative** and **Quantitative** data.
Difference between **Discrete** and **Continuous** data.
5 Types of Data Sampling: Knowing their definitions and their examples

---

Difference between **Ungrouped** and **Grouped Data (Frequency Table)**

Ungrouped Data

How to find **Mean**, **Mode**, **Median** , **Quartiles** (Q1, Q2, Q3), **Range**, **Standard Deviation** and **Variance** by formula and by **TI**.
How to do apply **3 Empirical Rules**. How to do **Box-plot** by **hand** and by **TI.**

Grouped Data (Frequency Table)

How to find **Mean**, **Standard Deviation** and **Variance** by formula and by **TI**.
How to do **Histogram** by hand**.**
How to do apply **3 Empirical Rules**.

Regression and Correlation:
How to do **Scattered -plot** for (X,Y) variables by hand and by **TI** and being able to **explain** that.
How to find **Correlation Coefficient** between (X,Y) by **TI** and being able to **explain** that.
How to find **Slope** and **Y-intercept** between (X,Y) by **TI ,** so you can write the **Regression Equation.**
Use the **Regression Equation** to **Estimate** or **Predict.**

Basic Probability Rules and its applications.
Multiplication Rule and its applications.

How to Study Part 1

Print all the materials for part 1,
Watch the PowerPoint (PowerPoint Link) slides for chapter 1.
Watch the PowerPoint (PowerPoint Link) slides for chapter 2, for descriptive statistics.
Read the topics review and starting from page 4 at the end of each topic there is a reference to extra practice problems, you **need to do all related-problems from Practice Problem** link and check the answers within the next few pages.
Learn how to use your TI calculator for problems that are computational such as how to compute mean, median, quartiles, standard deviation look at formula sheet, **TI 83/TI 84 links** and also the YouTube links.
You **must** be ready to **start** and **do HW problems**, as you finish reading pages from Topics Review and doing related-practice problems

---

Watch the PowerPoint (PowerPoint Link) slides for chapter 9 **regression topics**
For **regression topics** starting on page 14 of topics review, all the computational work **must** be done by **TI**.

Project one

Work on **Project one** after studying regression topics and doing related-practice problems.

Watch the PowerPoint slides for chapter 3 for probability discussed at the end of part 1

General Introduction

The Purpose of statistics: Statistics has many uses, but perhaps its most important purpose is to help us make decisions about issues that involve uncertainty. DDD (Data Driven Decisions), nowadays many decisions are solely made by the analysis of collection data and the entire process from collection through analysis of data is the subject of this study.

Definition of Statistics:

1. Numerical Facts

1. Average price for one bedroom apartment at the city of Rocklin is $ 895.
2. 80% of Sierra students graduate in 2 years.

2. C O D A **C**ollection, **O**rganization, **D**escription, **A**nalysis and interpretation of data.

Collection: **D**ata **S**ampling

Organization: **F**requency Table (Bar-chart, Pie-chart),
Histogram, **F**requency **P**olygon, **O**give **C**urve

Description: **M**ean, **M**ode, **M**edian,
Range, **V**ariance, **S**tandard **D**eviation **SD**,
Quartiles, **P**ercentiles, **B**ox Plot

Analysis: **C**orrelation and **R**egression, **E**stimation,
Test of **H**ypothesis, Analysis of **V**ariance

Types of Statistics:

Descriptive: **C**ollection, **O**rganization, **D**escription

Inferential: **A**nalysis and interpretation of data

What is the statistics all about?

1. It is about how we test if a new drug is effective in treating cancer.
2. It is about opinion polls, pre-election polls, and exit polls.
3. It is about sports, where we rank players and teams primarily through their statistics.
4. It is about the market research and the effectiveness of advertising
5. It is about how agricultural inspectors ensure the safety of the food supply.

Population versus **Sample**:

Population: Entire elements or subjects under study that share one or more **common characteristic** such as age, gender, major or race. (Keyword all/every), All college students, All Sierra College students, All male Sierra College Students who are taking statistics and majoring in business. Two Elements: **T**ime and **P**lace

Sample: A portion of population.

Census: The collection of data from every element in a population.

Parameter vs. statistic: A numerical measurement describing some characteristic of a **P**opulation (called Parameter) vs. a **S**ample (called statistic)
Hint: Use Greek Alphabet for parameter and lower case English for statistic.
μ = avg. σ (sigma) = st. dev χ^2 = Chi-squared…. $\overline{X}$, s, r

Extra Practice:** Answer questions **A** from **page 1** of practice problem **part 1

Types of Data:

Qualitative (Names, Labels ...) pass / fail, democrat/republican/independent, yes/no, grades (A,B,C,D,F)

Quantitative: **1**. **D**iscrete (Countable): number of accidents in Rocklin each day, number of emergency call to 911 center each day, number of students that will pass Abe stat class

2. **C**ontinuous (Measurable): Speed, weight, time, capacity, length, volume, area

Extra Practice:** Answer questions **B** from **page 1** of practice problem **part 1

5 Types of Sampling: **R_S_S_C_C**

1. **R**andom: Every member of population has equal chance to be selected.
How? Every member will be assigned a different number, and we select random numbers by a computer or a table and match those with the members' numbers.

2. **S**ystematic: We select some starting point and then select every kth (such as every 20^{th}) member in the population.
How? Every 10^{th} customer or client will be selected to be asked questions.

3. **S**tratified: We use stratified (subgrouping) sampling, when stata (subgroup) is of importance. We begin by **Subdividing** the population into at least two different subgroups (strata) sharing the same
characteristics (such as gender or age bracket), then we draw a sample from each stratum.
How? **a)**divide the police officers in Sacramento into male and female group
b)select a random sample of each and collect data regarding the years in service.

4. **C**luster: We use cluster sampling when data are located in some type of clusters like, restuants, classes, stations. We begin by dividing the population into sections (or clusters), and then randomly select some of those clusters, then choose all the members from those selected clusters.
How? To see the customer feedback to a new menu
a) divide Sacramento in different zones,
b) randomly select some of those zones
c) collect data from **all** fast-food branches in those selected zones.

5.**C**onvinence: Use the results are readily available.
How? A math instructor ask some of his students if they use student solution manual to do their homework.

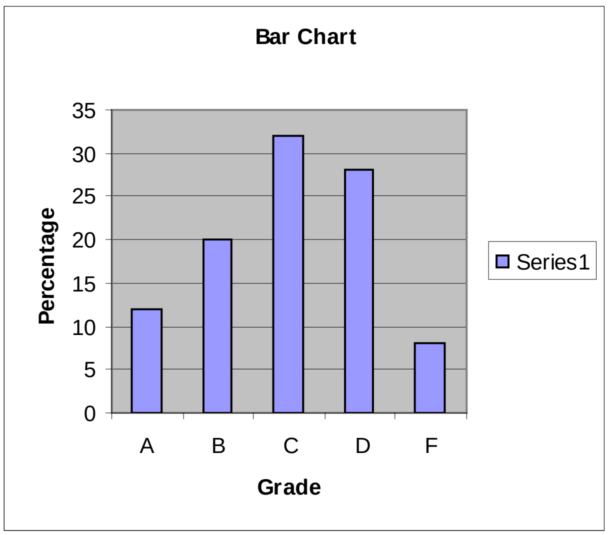
Extra Practice:** Answer questions **C** on **page 1** of practice problem **part 1

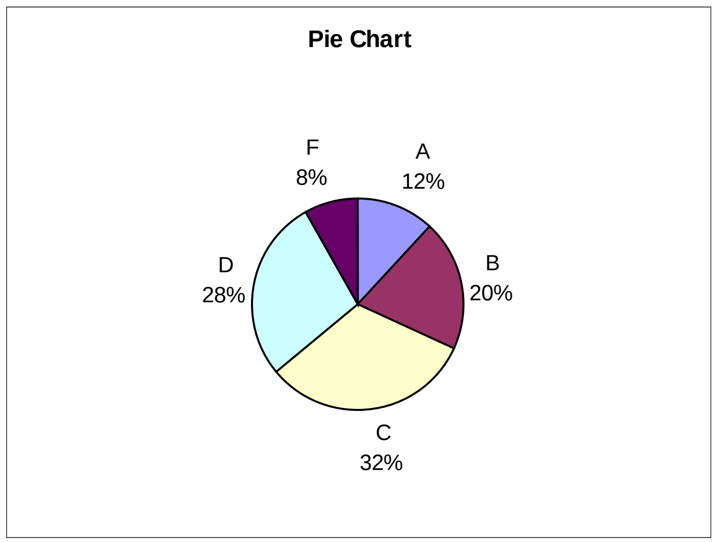
Qualitative Data

Example 1.

Grade	f = Students	Rel. freq % $\frac{f}{n}\times 100$	Angles 360° (Rel. freq)
A	**6**	$(6/50)\times 100 =$ **12**	$.12\times 360 =$ **43.2**$^\circ$
B	**10**	$(10/50)\times 100 =$ **20**	$.20\times 360 =$ **72**$^\circ$
C	**16**	$(16/50)\times 100 =$ **32**	$.32\times 360 =$ **115.2**$^\circ$
D	**14**	$(14/50)\times 100 =$ **28**	$.28\times 360 =$ **100.8**$^\circ$
F	**4**	$(4/50)\times 100 =$ **8** $+$	$.8\times 360 =$ **28.8**$^\circ$ $+$
	$n=\sum f = 50$	100 ?	360° ?

Using Excel to graph the followings





Practice 1:

Complete the table and draw the bar chart and the pie chart.(You can use Microsoft Excel to do the graphs)

Grade	f = Students	Rel. freq % $\frac{f}{n}\times 100$	Angles 360° (Rel. freq)
A	**22**		
B	**26**		
C	**20**		
D	**8**		
F	**4**	$+$	$+$
	$n=\sum f =$	100 ?	360° **?**

Descriptive Statistics

A) Measure of Central Tendency (**M**ean, **M**edian, **M**ode)

Mean (μ , $\bar{x}$) $x = data$ $\sum = Sum$ N or n = Number of data points

Data: 5, 6, 3 , 9, 7 $\bar{x} = \frac{\sum x}{n} = \frac{5+6+3+9+7}{5} = \frac{30}{5} = 6$

Median: The middle data point in a ranked (largest to smallest or smallest to largest) data, **or**
The median cuts the ranked data in half **one half below** it and **one half above** it.

Example1: Suppose the median score for the first test was 73, it simply means half the class got below 73 and the other half above it. Examples for odd and even numbered of data.

2, 5, 11, 16, 8, 9, 3, 7, 5 Ranked 2, 3, 5, 5, **7**, 8, 9, 11, 16, **Median** = **7**

2, 3, 5, 5, **7**, 8, 9, 11, 16, 4 Ranked 2, 3, 4, 5, **5**, **7**, 8, 9, 11, 16, **Median** = $(5+7)/2 = 6$

Hint: If there are extreme values in data set (too large or too low with respect of the rest of data) then median is a better than mean to identify the measure of central tendency.

Mode: The data value(s) with the highest occurrence, bimodal, multimodal

2, **8**, 11, 7, **8**, 13 **Mode** = 8

3, **12**, 5, 14, 9, **12**, **7**, 16, **7** **Bimodal** = 7, 12

11, **15**, 7, **2**, 6, 16, **15**, 3, **2**, **11**, 19, 5, 4 **Multimodal =** 2, 11, 15

B) Measure of Positions (**Q**uartiles, **B**ox-Plot, **P**ercentile, **Z**-score)

Quartiles: Breaking the **ranked data** in 3 quartiles (**Q1, Q2, Q3**)

Data: ______25%_______**Q1**_______ 25%_________**Q2**________25%_______**Q3**_______25%________

How to find quartiles? 3 steps
Rank the data, Find **Q2** = Median, Find the new medians **Q1, Q3** on either side of Q2.

Example 1: **Odd** number of data Data: 2, 5, 11, 16, 8, 9, 3, 7, 5, 4, 13

Ranked Data: 2, 3, ,**4**, 5, 5, **7**, 8, 9, **11**, 13, 16,
Q1 Q2 Q3

Example 2: **Even** number of data Data: 2, 3, 5, 5, 7, 8, 9, 11, 16, 4
Ranked Data 2, 3, **4**, 5, 5, 7, 8, **9**, 11, 16, **Q2 = Median** = $(5+7)/2 = 6$
Q1 **Q2** = 6 **Q3**

TI-83/84 *Inputting data in **L1** (stat → Option 1 → enter)*
then *stat → calc → Option 1 → enter → 2n d → 1 → enter*

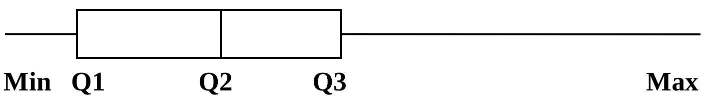
Extra Practice:** Answer questions on columns **A-G** on **page 3** of practice problem **part 1

Box-Plot: is mainly used for ungrouped data to show how the data are distributed by showing **center**, **spread**, and **skewness**. **Center** is the **Q2**, **Spread** is how wide the box is, **Skewness** explains the distribution of the data by using the longer tail to describe the **Skewness** (for example if the longer tail is on the right, it is called skewed to the right)

To construct a box-plot
1. Find the **5-number summery** of the data that are **Min, Q1, Q2, Q3, Max**
2. Plot these points on a **scaled** number line.
3. Construct a box by using Q1, Q2, Q3

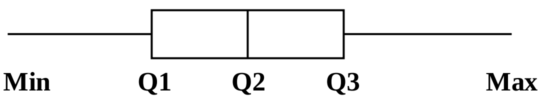
There are many possibilities of where the box in box-plot may be located.

If the box in box-plot is located to the **far Left**, it suggests that distribution of data are **skewed** to the **Right**



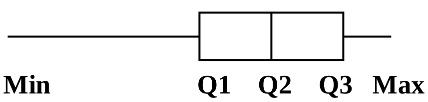
Skewed to the **Right**

If the box in box-plot is located to the **Center**, it suggests that distribution of data are **Centered**.



Centered

If the box in box-plot is located to the **far Right**, it suggests that distribution of data are **skewed** to the **Left**



Skewed to the **Left**

To see how to do Box-Plot by TI, look at the Youtube links on class website.

Extra Practice: *Answer questions on columns* ***A-G*** *on* ***page 3*** *of practice problem* ***part 1***

C) Measure of Variation (**R**ange, **S**tandard Deviation, **V**ariance)

Range: It shows how far apart the data points are? **Range** = the highest value **-** the smallest value

Standard Deviation (σ, s) : It measures the **average dispersion** of data **around the mean.**

Example: Consider the 3 random delivery time (in days) taken by 2 different companies A, and B

	A	**B**
Mean	5	5
Median	5	5
Mode	5	none

At first it seems there are not that much of difference between the delivery times of these two companies but let's look at their actual data and their plot on Dot-Plot.

	A	B
Delivery time	5	5
Delivery time	5	0
Delivery time	5	10

A **Dot Plot** **B**

A:
```
            x
            x
____ ____x________
0           5          10
```

B:
```
x________x________x
0        5        10
```

Now, it seems that there is **no dispersion** for company A, but an **average dispersion of 5** for company B.

The formula for the Standard Deviation or average dispersion of data around mean $s = \sqrt{\frac{\sum (x-\bar{x})^2}{n-1}}$

Company A

x	$\bar{x}$	$(x-\bar{x})$	$(x-\bar{x})^2$
5	5	0	0
5	5	0	0
5	5	0	0
			$\sum (x-\bar{x})^2 = 0$

Company B

x	$\bar{x}$	$(x-\bar{x})$	$(x-\bar{x})^2$
5	5	0	0
0	5	-5	25
10	5	5	25
			$\sum (x-\bar{x})^2 = 50$

$$s = \sqrt{\frac{\sum (x-\bar{x})^2}{n-1}} = \sqrt{\frac{0}{3-1}} = \sqrt{0} = 0$$

$$s = \sqrt{\frac{\sum (x-\bar{x})^2}{n-1}} = \sqrt{\frac{50}{3-1}} = \sqrt{25} = 5$$

Find the mean and standard deviation for 5, 6, 3 , 9, 10, 3, and also draw the **dot-plot**.

x	$\bar{x} = \frac{\sum x}{n} =$	$(x-\bar{x})$	$(x-\bar{x})^2$
5			
6			
3			
9			
10			
3			
$\sum x =$			$\sum (x-\bar{x})^2 =$

$s = \sqrt{\frac{\sum (x-\bar{x})^2}{n-1}} = \sqrt{\frac{\quad}{\quad}} = \sqrt{\quad} = 2.97$ **Variance** $= s^2 = 8.8$

Variance (σ^2, s^2): Variance is the **square of standard deviation.**

TI-83/84 *Inputting data in **L1** (stat $\rightarrow$ Option 1 $\rightarrow$ enter)*

then *stat $\rightarrow$ calc $\rightarrow$ Option 1 $\rightarrow$ enter $\rightarrow$ 2n d $\rightarrow$ 1 $\rightarrow$ enter*

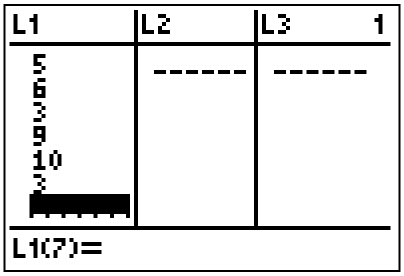
Rule of thumb to **estimate s**: $s = \frac{Range}{4}$ Generally the larger the data set the closer the estimate will be to the exact value.

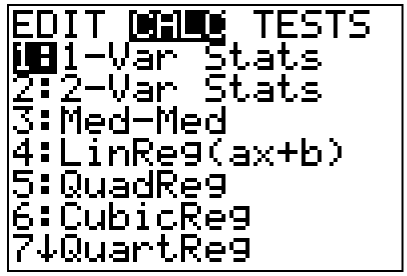
Extra Practice:** Answer questions on columns **A-G** on **page 3** of practice problem **part 1

TI-83/84

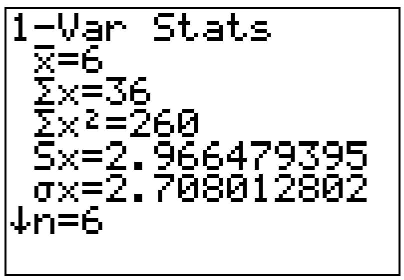
Find the mean, median, Q1, Q3 and standard deviation for 5, 6, 3 , 9, 10, 3, and also draw the Box_Plot.

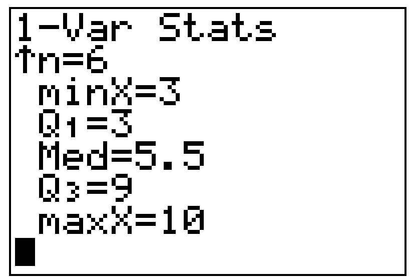
*Inputting data in **L1** (stat → Option 1 →enter*   *stat → calc→ Option 1 →enter*





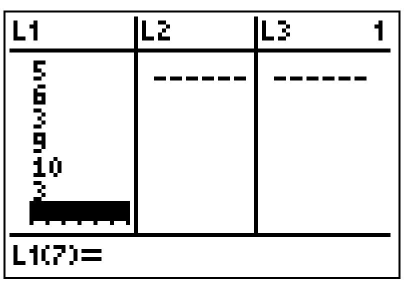
2n d →1 → enter   *Results*   *Use down arrow for more Results*

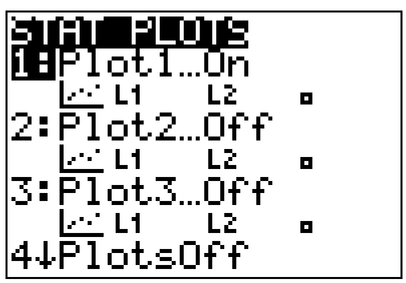


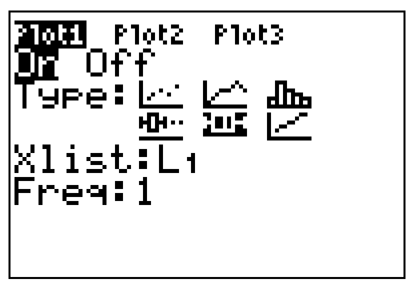


Doing the Box Plot by TI

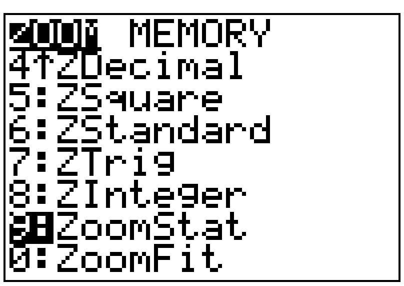
Inputting data in L1   ***2nd STAT Plots***   ***Choose the fifth option***

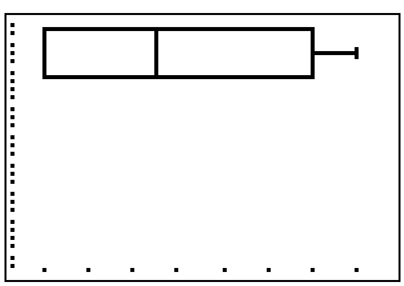






Press ZOOM 9   ***Result***





Empirical Rules: If and only if the **box-plot or histogram is centered** then we can apply the **three** following empirical rules.

$68\% = \bar{x} \pm s$ — **68 %** of data are within 1 s of the mean ($\bar{x}$)

$95\% = \bar{x} \pm 2s$ — **95 %** of data are within 2 s of the mean ($\bar{x}$)

$99.7\% = \bar{x} \pm 3s$ — **99.7 %** of data are within 3 s of the mean ($\bar{x}$)

Example: Find all three empirical rules for Abe Stat class if the average was 72 and the standard deviation was 8, assuming that Box-plot was centered.

$68\% = 72 \pm 1(8) = 72 \pm 8$ — $64 <$ **68 %** of class got scores < 80

$95\% = 72 \pm 2(8) = 72 \pm 16$ — $56 <$ **95 %** of class got scores < 88

$99.7\% = 72 \pm 3(8) = 72 \pm 24$ — $48 <$ **99.7 %** of class got scores < 96

Extra Practice: *Answer questions* ***C*** *on* ***page 3*** *of practice problem* ***part 1*** ***Answers are on p.18***

Z-score: is used to show the relative position of a data points with respect of the rest of data by measuring how many standard deviation the point is away from the mean. **To apply the z-score the box-plot or histogram must be centered**.

$$Z = \frac{x - \bar{x}}{s} \quad \textbf{or} \quad Z = \frac{x - \mu}{\sigma}$$

The possible range of Z-values;

------------ -2 --------------------- 0 -------------------- 2 -----------------------------

Unusual Values: $Z < -2$ **Ordinary Values:** $-2 \le Z \le 2$ **Unusual Values:** $Z > 2$

Example 1: Find the z-score of final exam for Tommy Yank in stat class at CSUS, if his score was 87, when the class average was 72 and the standard deviation was 8**.**

$$Z = \frac{x - \mu}{\sigma} = \frac{87 - 72}{8} = \frac{15}{8} = 1.875$$

Ordinary Or Unusual Value?

So, he does relatively an ordinary performance relative to the rest of his class.

Example 2: Find the z-score of final exam for Marcy Tank in stat class at UC Davis, if his score was 82, when the class average was 71 and the standard deviation was 4**.**

$$Z = \frac{x - \mu}{\sigma} = \frac{82 - 71}{4} = \frac{11}{4} = 2.75$$

Ordinary Or Unusual Value?

So, she does relatively better than the rest of her class.

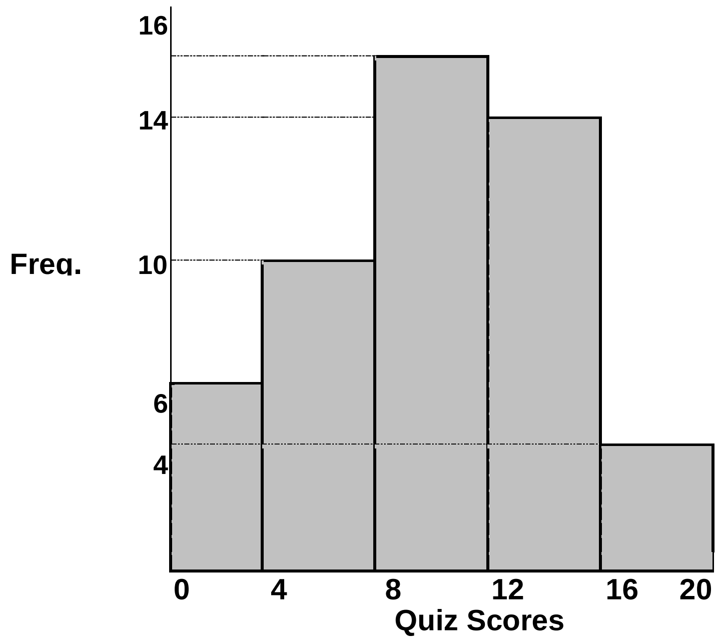
Extra Practice: *Answer questions* ***D*** *on* ***page 2*** *of practice problem* ***part 1***

Grouped Data (Freq. Table)

The table below shows the quiz scores of 50 students that are given in group.

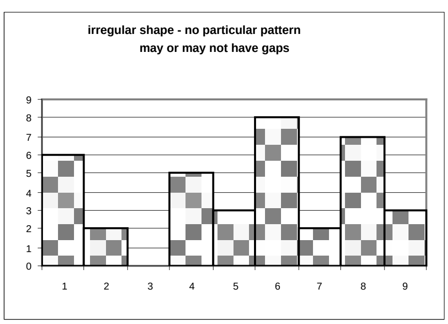
Quiz Score	Freq (f) = Students			
0 – 4	6			
4 - 8	10			
8 – 12	16			
12 – 16	14			
16 – 20	4			

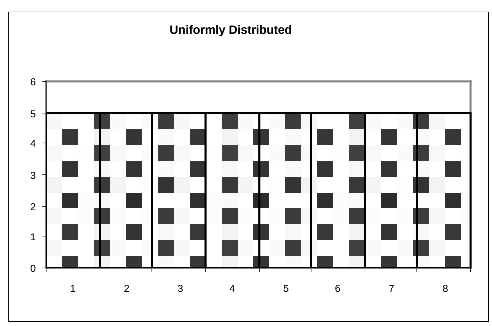
Use the quiz scores on x-axis, frequency on the Y-axis to draw blocks for a shape that is called **Histogram**

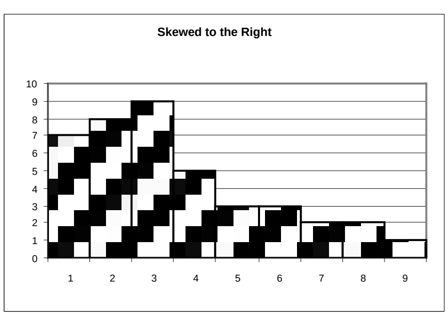


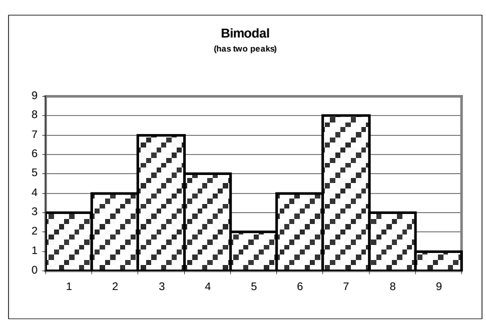
Histogram looks close to a Centered or bell-shaped distribution.

Different possible shapes of Histogram









Mean and **Standard Deviation**.

First step is to create a new column called **midpoint** (average of scores in each group). For example for 0 – 4, the midpoint will be 2, for 4 – 8, the midpoint will be 6. Next step is to open two new columns $f \times m$ and $f \times m^2$ do the necessary calculations, find the summation for each and then use them in the given formulas.

X-axis		**Midpoint is the average for each group**	Use the **summation** of this **column** for **mean** and **St, Dev** formula	Use the **summation** of this **column** for **St, Dev** formula
Quiz Scores	**Freq(f)= Students**	m	$f \times m$	$f \times m^2$
0 – 4 4 - 8 8 – 12 12 – 16 16 – 20	6 10 16 14 4	$(0+4)/2=2$ $(4+8)/2=6$ 10 14 18	12 60 160 196 72	24 360 1600 2744 1296
	$\sum f = n = 50$		$\sum(f \times m) = 500$	$\sum(f \times m^2) = 6024$

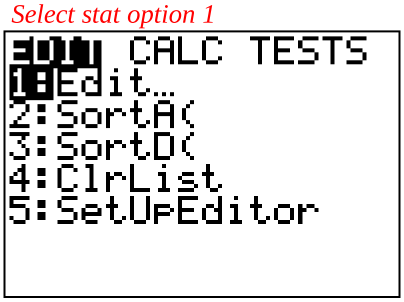
Mean: $\bar{X} = \dfrac{\sum(f \times m)}{n} = \dfrac{500}{50} = 10$

Standard deviation: $s = \sqrt{\dfrac{n\sum f \times m^2 - (\sum f \times m)^2}{n(n-1)}} = \sqrt{= \dfrac{50(6024)-(500)^2}{50(50-1)}} = \sqrt{\dfrac{51200}{2450}} = 4.57$

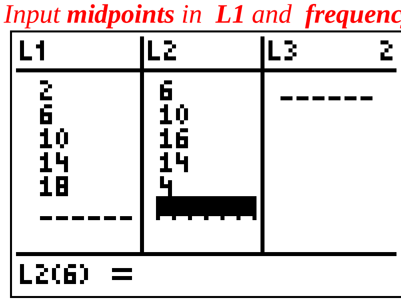
Variance: $s^2 = 4.57^2 = 20.9 =$

TI-83/84

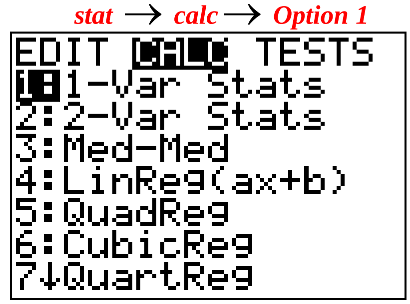
Select stat option 1



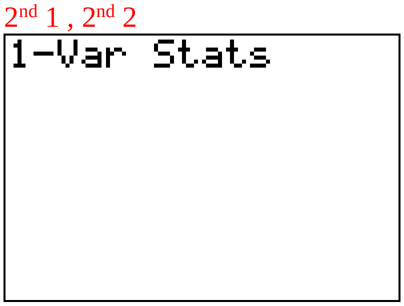
*Input **midpoints** in **L1** and **frequency** in **L2***



stat → ***calc*** → ***Option 1***

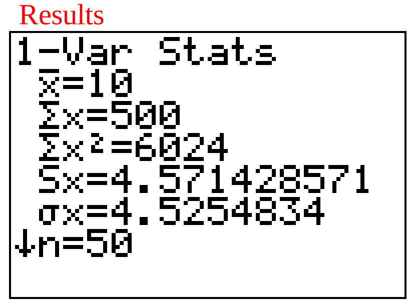


2nd 1 , 2nd 2



Press enter

Results



Practice 1: Use both formula and the Ti to find the mean, standard deviation and the variance.

Quiz Scores	**Freq(f)**	m	$f \times m$	$f \times m^2$
0 – 10 10 - 20 20 – 30 30 – 40	8 12 14 6	 25 	 180 	 7350
	$\sum f = n =$		$\sum f \times m =$ **780**	$\sum f \times m^2 =$ **19000**

Mean: $\bar{X} = \frac{\sum f \times m}{n} = \frac{\quad}{\quad} = 19.5$

Standard deviation: $s = \sqrt{\frac{n\sum f \times m^2 - \left(\sum f \times m\right)^2}{n(n-1)}} = \sqrt{\frac{\quad}{\quad}} = \sqrt{\frac{\quad}{\quad}} = 9.86$

Variance: $s^2 = 9.8^2 = 97.18$

Practice 2: Use both formula and the Ti to find the mean, standard deviation and the variance

Test Scores	**Freq (f)=**	m	$f \times m$	$f \times m^2$
0 – 20	2			
20 – 40	8			
40 – 60	14	50	700	35000
60 – 80	32			
80 – 100	24			
	$\sum f = n =$		$\sum f \times m =$	$\sum f \times m^2 =$

Mean: $\bar{X} = \frac{\sum f \times m}{n} = \frac{\quad}{\quad} = 67$

Standard deviation: $s = \sqrt{\frac{n\sum f \times m^2 - \left(\sum f \times m\right)^2}{n(n-1)}} = \sqrt{\frac{\quad}{\quad}} = \sqrt{\frac{\quad}{\quad}} = 20.89$

Variance: $s^2 =$

Extra Practice: *Answer questions* ***A, B, C, D*** *on* ***pages 5, 6*** *of practice problem* ***part 1***

Regression and Correlation

It is to explore and study of the **relationship** between **two variables** (x , y) with the objective of formulating an equation between the two variables and using that equation to predict one from the other.
(x is also called independent, explanatory, or predictor variable)
(y is also called dependent, response variable). So a response variable is the variable whose value can be explained by the predictor variable.

Steps

1. To find the **nature of the relationship** (Linear or non-linear, positive or negative relationship) by doing a scattered diagram, y versus x

2. To measure the **strength of this relationship** by computing the correlation coefficient $= r$

3. Finding **slope** and **y-intercept** for equation of the best fitted- line (**regression equation =** $y = a\,x + b$) between x , y variables.

4. Using the regression equation to **estimate or predict** one variable from the other.

Nature of relationship:
Positive: Both variables either increasing or decreasing $x \uparrow\uparrow y$ **or** $x \downarrow\downarrow y$
Negative: When one variable increases the other one decreases or vice versa. $x \uparrow\downarrow y$ **or** $x \downarrow\uparrow y$

What do you think is the nature of relationship between x and y variables?

	x Independent, Explanatory, or Predictor variable	y Dependent, or response variable	Nature of relationship Positive, Negative
1	Average number of hours per week to study for stat class	Stat test score	+ , − , *None*
2	Mortgage rate	Number of loans refinanced	+ , − , *None*
3	Average height of the parents	Height of the sons or daughters	+ , − , *None*
4	No. of absences in a semester for stat class	Stat test scores	+ , − , *None*
5	Daily temperature in summer	Water or electric consumption	+ , − , *None*
6	$ amount spent on advertisement	Monthly sales	+ , − , *None*
7	Fat consumption	Cholesterol level	+ , − , *None*
8	Number of years of education	Monthly salary	+ , − , *None*
9	Number of hours watching TV/week	GPA	+ , − , *None*
10	Ice cream sales	Number of drownings	+ , − , *None*

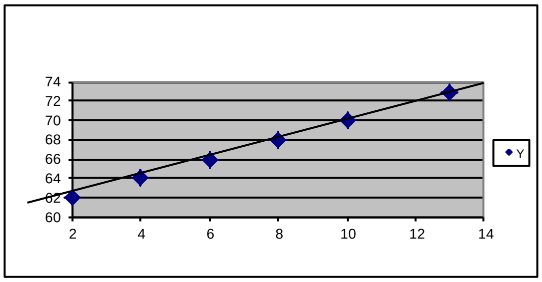
Steps to do a Correlation and Regression problem

1. Constructing a Scattered Diagram and comment on its nature (linear or non-linear, positive or negative, strong or weak relationship).

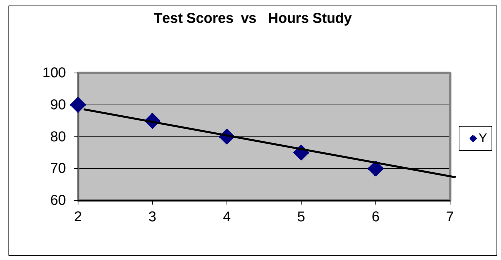
Why do we need a scattered diagram?

a) To see if data exhibit a **linear pattern** or not b) To see if linear pattern is **positive or negative**
c) To see how closely (**strongly**) the data are **clustered around** the **mean**
d) To detect any **outlier** (a point that is lying far away from the other data points).

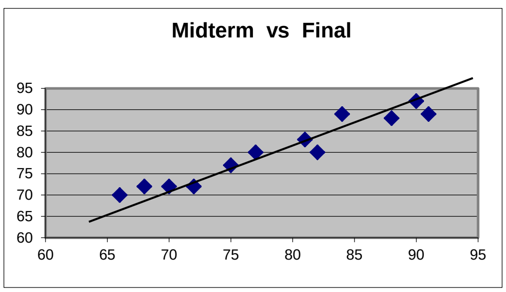
Different Possible shapes of a Scattered Diagram



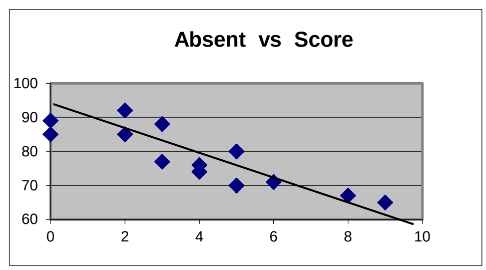
$r = 1$
Perfect Positive Linear Correlation



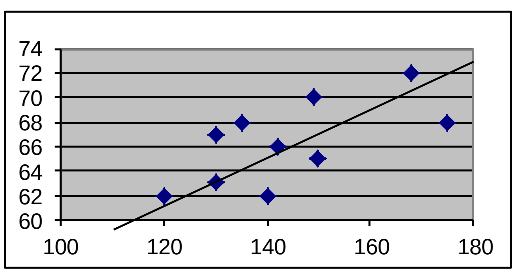
$r = -1$
Perfect Negative Linear Correlation



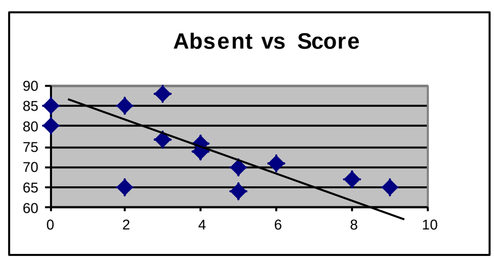
Strong Positive Linear Correlation



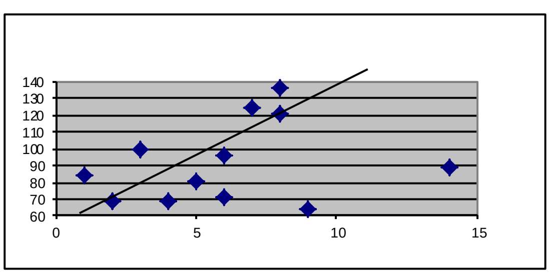
Strong Negative Linear Correlation



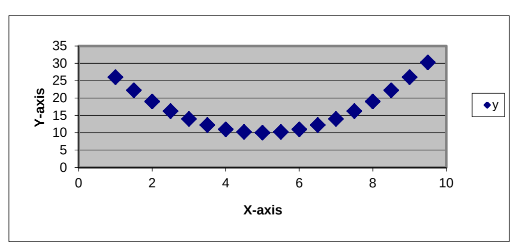
Positive Linear Correlation



Negative Linear Correlation



No Correlation

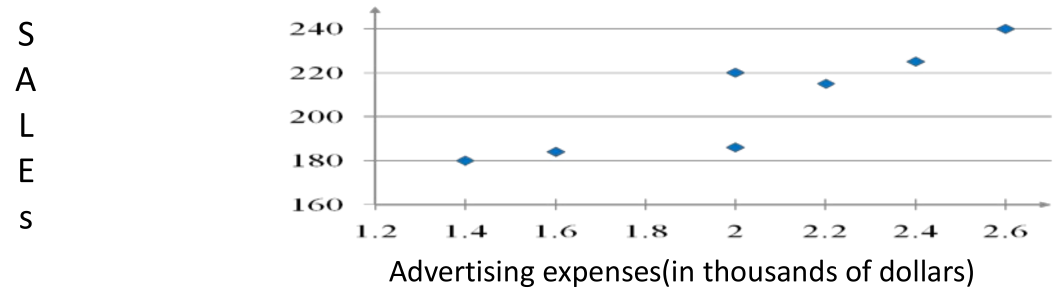


Non linear relationship

Very important: If pattern of data is not linear (looks like a **curve**) or it has an **outlier** or it show **no pattern** then **linear regression method** is **not valid** and **not applicable.**

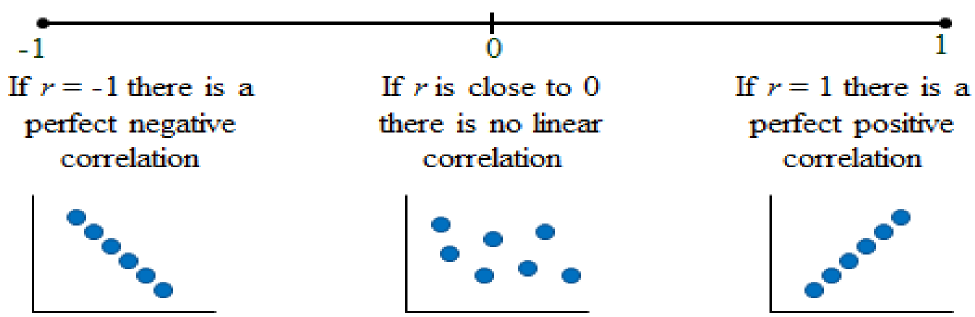
Example

Advertising expenses($1000), **X**	2.4	1.6	2.0	2.6	1.4	1.6	2.0	2.2
Company Sales ($1000), **Y**	225	184	220	240	180	184	286	215



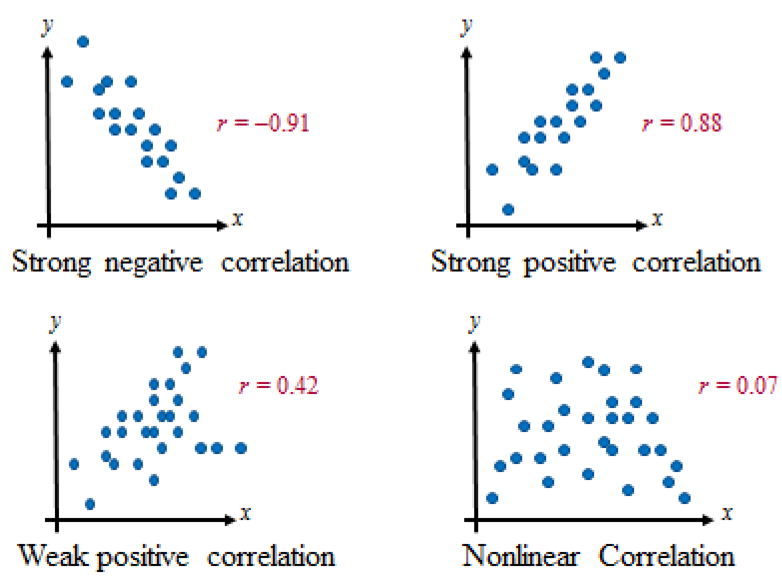
The pattern suggest a strong positive linear relationship.

2 Computing r = Correlation Coefficient (**the measurement of strength of relationship between 2 variables)** by formula given by $r = \dfrac{n\sum xy - \sum x \sum y}{\sqrt{n\sum x^2 - (\sum x)^2}\sqrt{n\sum y^2 - (\sum y)^2}} =$ or using Ti calculator and comment on its strength. The value of r is always between $-1 \le r \le 1$



In above example r ≈ 0.913 suggests a strong positive linear correlation. As the amount spent on advertising increases, the company sales also increase.

Linear Correlation Coefficient and scattered Diagram



4. Computing $\overline{x}, \overline{y}, s_x, s_y$,

5. Computing Slope ($\boldsymbol{a}$) and y-intercepts ($\boldsymbol{b}$) for the regression equation $\boldsymbol{y = a\,x + b}$ by formula given by

$$\boldsymbol{Slope} = \boldsymbol{a} = \frac{n\left(\sum xy\right) - \left(\sum x\right)\left(\sum y\right)}{n\left(\sum x^2\right) - \left(\sum x\right)^2}$$ and

$$\boldsymbol{y\text{ - }itc} = \boldsymbol{b} = \frac{\left(\sum y\right)\left(\sum x^2\right) - \left(\sum x\right)\left(\sum xy\right)}{n\left(\sum x^2\right) - \left(\sum x\right)^2}$$ or using Ti calculator

For above example $\boldsymbol{a} = 50.73$, $\boldsymbol{b} = 104.1$

5. Using $\boldsymbol{a} = 50.73$, $\boldsymbol{b} = 104.1$ and inputting them into regression equation $\boldsymbol{y = a\,x + b}$, $\boldsymbol{y = 50.72\,x + 104.1}$ then use this equation to **estimate or predict** one variable from the other.

Estimated values are labeled as y' (y -prime) and x' (x -prime).

For example if advertising expense is $2.5 thousand dollar, then the predicted company sale by regression equation will be $\boldsymbol{y' = 50.72}(2.5)\boldsymbol{x + 104.1 = 230.88}$ or $\boldsymbol{y' = \$230{,}880}$

Guideline for using the regression line:

1. If there is no significant linear correlation, do not use the regression equation.

2. When using the regression equation for prediction, **stay** within the range of the available sample data.

3. A Regression equation based on old data is not necessarily valid now.

Marginal Change (Slope): in a variable is the amount that it changes in y-variable when the x-variable increases by one unit.

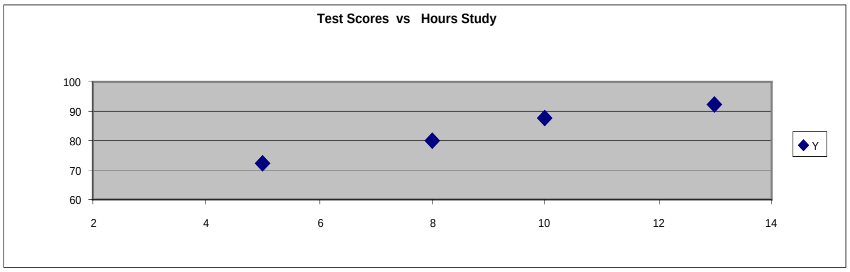
Outlier: is a point that is lying far away from the other data points.

Coefficient of determination $= r^2 \times \% = \dfrac{\text{explained varation}}{\text{total variation}} =$ is the amount of variation in y that is explained by the regression line

Example 1.

	x = Hours Study/week	y = Test Score	x^2	y^2	$x\,y$
1	5	72	25	5184	360
2	10	88	100	7764	880
3	13	92	169	8464	1196
4	8	80	64	6400	640
	$\Sigma x = \mathbf{36}$	$\Sigma y = \mathbf{332}$	$\Sigma x^2 = \mathbf{358}$	$\Sigma y^2 = \mathbf{27792}$	$\Sigma xy = \mathbf{3076}$

1. Use the data and plot the data as a scattered diagram and **comment** on the pattern of the points.



Strong Positive

Linear Correlation

2. Compute the correlation coefficient and **comment** on that: ***a very strong positive linear correlation.***

$$r = \frac{n\sum xy - \sum x \sum y}{\sqrt{n\sum x^2 - \left(\sum x\right)^2}\sqrt{n\sum y^2 - \left(\sum y\right)^2}} = \frac{(4)(3076) - (36)(332)}{\sqrt{4(358) - (36)^2}\sqrt{4(27792) - (332)^2}} = \frac{12304 - 11952}{\sqrt{136}\sqrt{944}} = \frac{352}{358.307} = 0.9824$$

3. Compute the slope and y-intercept and write the equation of regression line.

$$\textbf{\textit{Slope}} = \boldsymbol{a} = \frac{n\left(\sum xy\right) - \left(\sum x\right)\left(\sum y\right)}{n\left(\sum x^2\right) - \left(\sum x\right)^2} = \frac{4(3076) - (36)(332)}{4(358) - (36)^2} = \frac{12304 - 11952}{1432 - 1296} = \frac{352}{136} = \mathbf{2.588 = 2.59}$$

$$\textbf{\textit{y - itc}} = \boldsymbol{b} = \frac{\left(\sum y\right)\left(\sum x^2\right) - \left(\sum x\right)\left(\sum xy\right)}{n\left(\sum x^2\right) - \left(\sum x\right)^2} = \frac{(332)(358) - (36)(3076)}{4(358) - (36)^2} = \frac{118856 - 110736}{1432 - 1296} = \frac{8120}{136} = \mathbf{59.71}$$

$$\boldsymbol{y = a\,x + b = 2.59\,x + 59.71}$$

4. Explain the slope based on the regression equation and the in relation of x and y variables.

   In general for every additional hour of study per week the score goes up by 2.59 points.

5. Compute average and standard deviation for both x and y variables.

$$\bar{x} = 36/4 = \mathbf{9}\ \ hrs \quad \boldsymbol{s_x} = 3.37 \qquad \bar{y} = 332/4 = \mathbf{83} \quad \boldsymbol{s_y} = 8.87$$

6. If one student studies 10 hours a week, use **Reg. Equ.** to estimate her test score. $x = 10\ hrs,\ \ y' = 85.61$

$$x = 10\ hrs,\ \ y' = 85.61$$

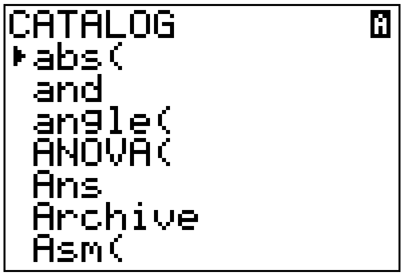
7. If one student has test score of 90, use **Reg. Equ.** to estimate number of hours he spends studying per week.

   and if $y = 90,\ \ x' = 11.69$ hrs

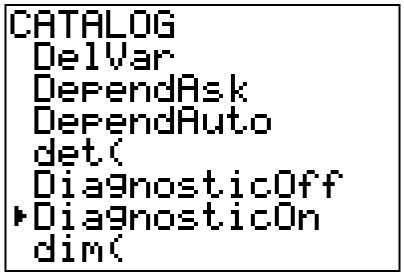
8. Compute the coefficient of determination $(r^2 \times 100)$ and **comment** on that: $(r^2 \times 100) = (.9824^2 \times 100) = 96.5\ \%$, 96.5% of variations in test score are explained by regression equation

TI-83/84

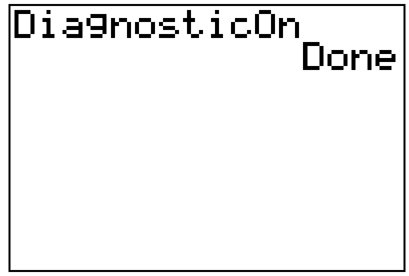
2n d → 0



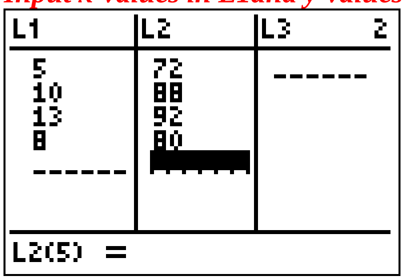
select Diagnostic on → enter



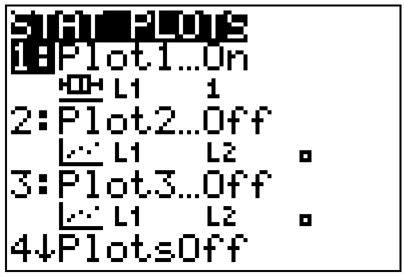
enter



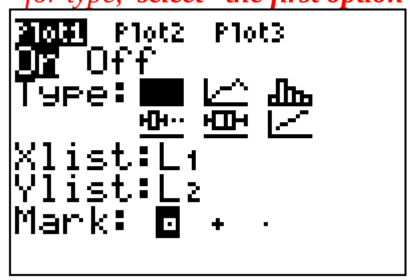
Input x-values in L1and y-values in L2



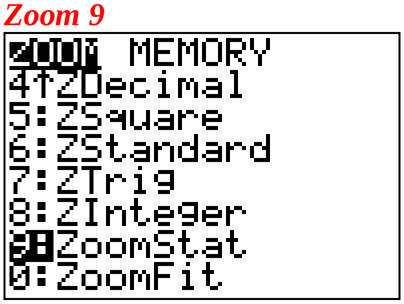
2nd STAT PlOTS



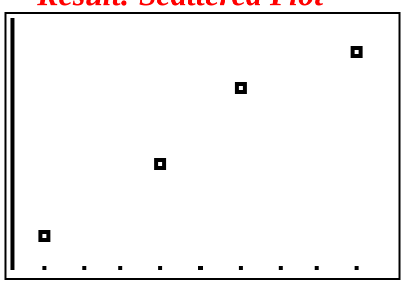
for type, select the first option



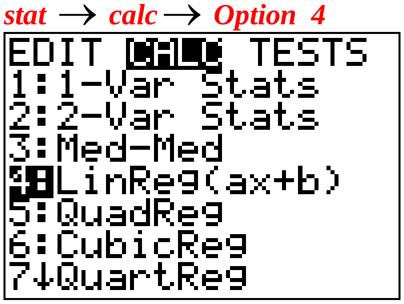
Zoom 9

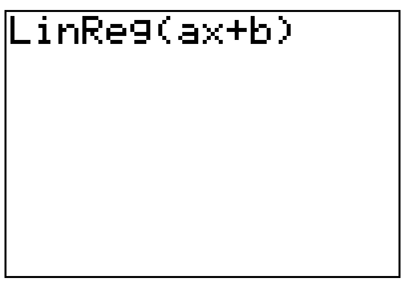


Result: Scattered Plot

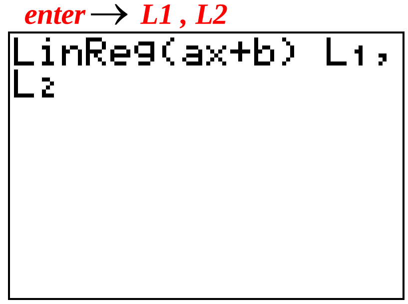


stat → calc → Option 4

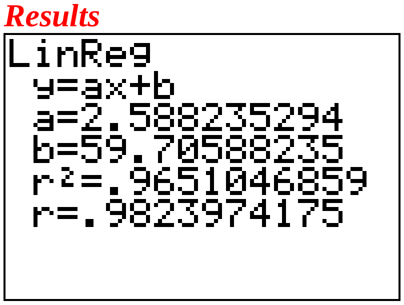




enter → L1 , L2



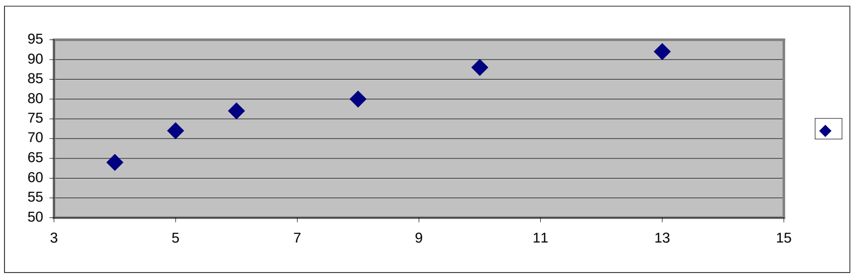
Results



More Practice

	x = Hours Study/week	y = Test Score	x^2	y^2	$x\,y$
1	5	72			
2	10	88			
3	13	92			
4	8	80			
5	6	77			
6	4	64			
	$\sum x = 46$	$\sum y = 473$	$\sum x^2 = 410$	$\sum y^2 = 37817$	$\sum xy = 3794$

1. Use the data and plot the data as a scattered diagram and **comment** on the pattern of the points.



Comment: A very strong positive linear correlation.

2. Compute the correlation coefficient and **comment** on that $r = 0.963$ Very strong…?

3. Compute the slope and y-intercept and write the equation of regression line. Slope = a = 2.92, y-itc = b = 56.41

$$y = a\,x + b = 2.92\,x + 56.41$$

4. Explain the slope based on the regression equation and the in relation of x and y variables.
In general for every additional hour of study per week the score goes up by 2.92 points.

5. Compute average and standard deviation for both x and y variables. $\bar{x} = 7.67$, $\bar{y} = 78.83$, $s_x = 3.386$, $s_y = 10.28$

6. If one student studies 6 hours a week, use **Reg. Equ.** to estimate her test score. $x = 6$, $y' = 73.93$

7. If one student has test score of 85, use **Reg. Equ.** to estimate number of hours he spends studying per week. $y = 85$, $x' = 9.79$

8. Compute the coefficient of determination $(r^2 \times 100)$ and **comment** on that.
$(r^2 \times 100) = (.962^2 \times 100) = 92\,\%$, *92% of variations in test score are explained by regression equation.*

Extra Practice*: Do Problems, A through F** *on* ***pages 11, 12, 13 *of practice problem* ***part 1***

The table below gives the height and shoe sizes of six randomly selected man.

	$x =$ Height	$y =$ Shoe Size	x^2	y^2	$x\,y$
1	72	9	***5184***	***81***	***648***
2	79	14			
3	74	11	***5476***	***121***	***814***
4	75	12			
5	78	13	***6084***	***169***	***1014***
6	71	8			
	$\Sigma x =$	$\Sigma y =$	$\Sigma x^2 =$	$\Sigma y^2 =$	$\Sigma xy =$

8. Use the data and plot the data as a scattered diagram and **comment** on the pattern of the points.

9. Compute the correlation coefficient and **comment** on that: ***a very strong positive linear correlation.***

$$r = \frac{n\sum xy - \sum x \sum y}{\sqrt{n\sum x^2 - \left(\sum x\right)^2}\sqrt{n\sum y^2 - \left(\sum y\right)^2}} =$$

$$\frac{(\qquad)-(\qquad)(\qquad)}{\sqrt{(\qquad)-(\qquad)^2}\sqrt{(\qquad)-(\quad)^2}} = \frac{\quad - \quad}{\sqrt{\quad}\sqrt{\quad}} = \frac{\quad}{\quad} = 0.977$$

10. Compute the slope and y-intercept and write the equation of regression line.

$$\textbf{\textit{Slope}} = \boldsymbol{a} = \frac{n\left(\sum xy\right)-\left(\sum x\right)\left(\sum y\right)}{n\left(\sum x^2\right)-\left(\sum x\right)^2} = \frac{(\qquad)-(\qquad)(\qquad)}{(\qquad)-(\qquad)^2} = \frac{\quad - \quad}{\quad - \quad} = 0.711$$

$$\textbf{\textit{y - itc}} = \boldsymbol{b} = \frac{\left(\sum y\right)\left(\sum x^2\right)-\left(\sum x\right)\left(\sum xy\right)}{n\left(\sum x^2\right)-\left(\sum x\right)^2} = \frac{(\qquad)(\qquad)-(\qquad)(\qquad)}{(\qquad)-(\qquad)^2} = \frac{\quad - \quad}{\quad - \quad} = -42.08$$

$$\boldsymbol{y = a\,x + b =} \qquad \boldsymbol{x} +$$

Use the repression equation to answer the following questions

Use, your own height and the linear equation to see if the predicted value matches your actual shoe size.

If man has shoe size 0f 10.5, what would be his predicted height?

Basic Probability

Probability of an **event** A $= P(A) = \frac{f}{N}$ = $\frac{\text{The Number of desired outcome Can Occur}}{\text{The Total Number Of Possible Outcomes}}$

$$0 \le P(A) \le 1$$

If the **probability** of occurrence of an event such as event A is between $0 \le P(A) < 5\%$ then its occurrence is called unusual.

Definition	Examples		
An **experiment** is a situation involving chance or probability that leads to results called outcomes.	**Tossing a coin.**	**Rolling a Die**	**Draw a card from a deck of card**
All possible outcomes of the experiment are called **sample space** Find **N = ?**	**sample space** **N = 2** outcomes (H,T)	**sample space** **N** = 6 outcomes (1,2,3,4,5,6)	**sample space** **N** = 52 outcomes
What is/are the **desired outcome** or **Outcomes**? Find **f = ?**.	(to be tail) **f = 1**	(an odd number 1,3,5) **f = 3**	(an face) **f = 12**
Probability is the measure of how likely an event is $= P(A) = \frac{f}{N}$	probability to be tail P(H) = 1/2 =50%	probability of an odd number P(odd number) = 3/6 =50%	probability of a face P(Ace) = 12/52 =23%

Three Types of probability;

a) **Classical or theoratical:** No experiment is needed and the sample space is always the same and probability of any event is already known like rolling a die (the probability of getting 5 is always 1/5)

b) **Empirical Probability:** Based on observation or data from a probability wxperiment and probability depends as where and when data have been collected. See example A on the same page)

c) Subjective: Intution, educated guesses or estimates (A doctor believes 70 % chance of recovevery)

Example A: Frequency distribution of annual income for U.S. families

Income	Frequency (1000s)
Under $10,000	5,216
$10,000–$14,999	4,507
$15,000–$24,999	10,040
$25,000–$34,999	9,828
$35,000–$49,999	12,841
$50,000–$74,999	14,204
$75,000 & over	12,961
	69,597

Part 1: Find the probability that a randomly selected person from this group makes $75,000 and over

1) Experiment: randomly selecting a person. **2)** Sample space $= N = 69,597$

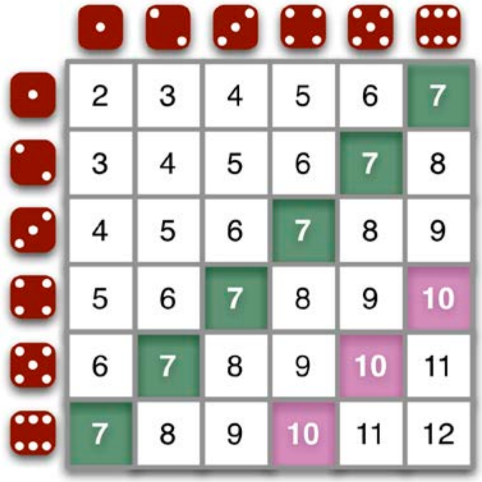
3) His/her income is $75,000 and over: $f = 12,961$ **4)** Prob ($75,000 and over) $= 12,961/69,597 = 18.63\ \%$

Part 2: Find the probability that a randomly selected person from this group makes $24,999 or less

1) Experiment: randomly selecting a person. **2)** Sample space $= N = 69,597$

3) His/her income is $75,000 and over: $f = 19,763$ **4)** Prob ($24,999 or less) $= 19,763/69,597 = 28.40\ \%$

Example B:
If we roll 2 dice, then there are 36 possible outcomes meaning that the **sample space is 36** or $= N = 36$



	1	2	3	4	5	6
1	2	3	4	5	6	7
2	3	4	5	6	7	8
3	4	5	6	7	8	9
4	5	6	7	8	9	10
5	6	7	8	9	10	11
6	7	8	9	10	11	12

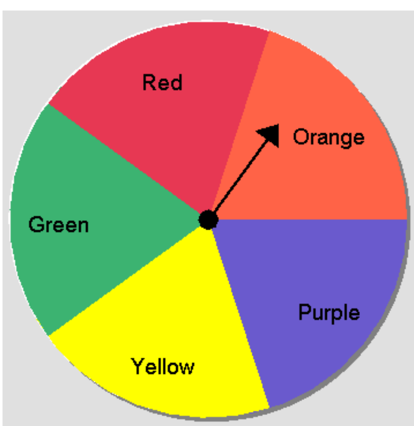
Solution:

a) find the probability that their total is 10 **Event** or desired outcomes: a total of
$10 \Rightarrow \{(4,6),(5,5),(6,4)\} \Rightarrow f = 3$ Prob (a sum of 10) $= 3/36 = 1/12 = 8.33\,\%$

b) find the probability that their total 10 or more **Event** or desired outcomes: to get a total 10 or more
$\Rightarrow \{(4,6),(5,5),(5,6),(6,4),(6,5),(6,6),\} \Rightarrow f = 6$ Prob (a total of 10 or more) $= 6/36 = 1/6 = 16.67\,\%$

c) find the probability that their total is 5 **Event** or desired outcomes: to get a total of 5
$\Rightarrow \{(1,4),(2,3),(3,2),(4,1)\} \Rightarrow f = 4$ Prob (a total of 5) $= 4/36 = 1/9 = 11.11\,\%$

Example C.



Example D
What is the probability that you spin the dial on the left spinner and you get yellow?
What is the probability that you spin the dial on the right spinner and you get lose turn?

Example E In a deck of 52 cards there are 13 diamonds and 12 faces, and 4 aces. If one card is drawn randomly find the probability that

a) it is a diamond b) it is a face c) It is a diamond and face

Solution:
a) P (diamond) $= 13/52 = 25\,\%$ b) P(face) $= 12/52 = 23.08\,\%$ c) P(diamond and face) $= 3/52 = 5.77\,\%$

Multiplication Rule (Keywords: and, both, all)

$$P(A \text{ and } B \text{ and } C \text{ and} \ldots) = P(A)P(B)P(C)\ldots$$

We use multiplication rule to find the probability that events A, B, C happen together
Hint:
When you make a selection out of a group by using multiplication rule be aware of **with** or **w/o** replacement effect.

In a deck of 52 cards there are 13 diamonds and 12 faces, and 4 aces.

If 2 cards are randomly drawn **w/o replacement,** what is the probability that both are ace?
P(both ace} $= \frac{4}{52} \bullet \frac{3}{51} = 0.0045$
If 2 cards are randomly drawn **with replacement,** what is the probability that both are ace?
P(both ace) $= \frac{4}{52} \bullet \frac{4}{52} = 0.0059$

Example 1. There are 13 diamonds and 12faces, and 4 aces in a deck of card. If 4 cards are randomly drawn **w/o replacement** then,

a) What is the probability that all 4 are diamond and how likelihood is this?

13 Diamo, 39 Others, 52 $\Rightarrow$ 12 Diamo, 39 Others, 51 $\Rightarrow$ 11 Diamo, 39 Others, 50 $\Rightarrow$ 10 Diamo, 39 Others, 49

$\frac{13}{52} \bullet \frac{12}{51} \bullet \frac{11}{50} \bullet \frac{10}{49} = .26\%$ very unlikely

b) What is the probability that all 4 are aces and how likelihood is this?

4 Aces, 48 Others, 52 $\Rightarrow$ 3 Aces, 48 Others, 51 $\Rightarrow$ 2 Aces, 48 Others, 50 $\Rightarrow$ 1 Aces, 48 Others, 49

$\frac{4}{52} \bullet \frac{3}{51} \bullet \frac{2}{50} \bullet \frac{1}{49} = .000369$ % very much unlikely

c) What is the probability that all 4 are faces and how likelihood is this?

12 Faces, 40 Others, 52 $\Rightarrow$ 11 Faces, 40 Others, 51 $\Rightarrow$ 10 Faces, 40 Others, 50 $\Rightarrow$ 9 Faces, 40 Others, 49

$\frac{12}{52} \bullet \frac{11}{51} \bullet \frac{10}{50} \bullet \frac{9}{49} = .001828 = .1828\%$ much unlikely

d) What is the probability that all 4 are non faces and how likelihood is this?

12 faces, 40 non-faces, 52 $\Rightarrow$ 12 faces, 39 non-faces, 51 $\Rightarrow$ 12 faces, 38 non-faces, 50 $\Rightarrow$ 12 faces, 37 non-faces, 49

$\frac{40}{52} \bullet \frac{39}{51} \bullet \frac{38}{50} \bullet \frac{37}{49} = .337575 = 33.76$ % It is likely

Example 1. In a bag of 2 green, 5 reds and 3 yellow ball, if we select 2 balls at random with replacement , then find

a) the probability that both are red $\frac{5}{10} \bullet \frac{5}{10} = \frac{25}{100} = 0.25$

b) the probability that none of the balls are yellow $\frac{7}{10} \bullet \frac{7}{10} = \frac{49}{100} = 0.49$

There are 10 men, and 8 women in a group. If two people are selected at random **without replacement**, then,

1.Write all the possibilities

2. Compute all the probabilities

Possibilities | **Probabilities**

10 M / 8 W / 18 $\Rightarrow$ 9 M / 8 W / 17

MM $\quad p(M) = 10/18 \cdot p(M) = 9/17 \quad \Rightarrow \quad P(MM) = \frac{10}{18} \cdot \frac{9}{17} = 29.4\%$

10 M / 8 W / 18 $\Rightarrow$ 9 M / 8 W / 17

MW $\quad p(M) = 10/18 \cdot p(W) = 8/17 \quad \Rightarrow \quad P(MW) = \frac{10}{18} \cdot \frac{8}{17} = 26.14\%$

10 M / 8 W / 18 $\Rightarrow$ 10 M / 7 W / 17

WM $\quad p(W) = 8/18 \cdot p(M) = 10/17 \quad \Rightarrow \quad P(WM) = \frac{8}{18} \cdot \frac{10}{17} = 26.14\%$

10 M / 8 W / 18 $\Rightarrow$ 10 M / 7 W / 17

WW $\quad p(W) = 8/18 \cdot p(W) = 7/17 \quad \Rightarrow \quad P(WW) = \frac{8}{18} \cdot \frac{7}{17} = 18.3\%$ +

100%

then, find the following probabilities,

5. Both are men. **29.4 %**

6. Both are women. **18.3 %**

7. At least one (minimum one) woman. $P(MW) + P(WM) + P(WW) =$ 26.14 + 26.14 + 18.3 = **70.6 %**

8. At most one (maximum one) woman $P(MW) + P(WM) + P(MM) =$ 26.14 + 26.14 + 29.4 = **81.7 %**

9. One man and one woman. $P(MW) + P(WM) =$ 26.14 + 26.14 = **52.28 %**

B. In a box there are 10 Red and 8 Blue balls. If two balls are drawn at random **with replacement**, then

1.Write all the possibilities

2. Compute all the probabilities

Possibilities

Probabilities

10 R, 8 B (18)	$\Rightarrow$	10 R, 8 B (18)

RR $\quad p(R) = 10/18 \;\cdot\; p(R) = 10/18 \quad \Rightarrow \quad P(RR) = \frac{10}{18}\cdot\frac{10}{18} = 30.86\,\%$

10 R, 8 B (18)	$\Rightarrow$	10 R, 8 B (18)

RB $\quad p(R) = 10/18 \;\cdot\; p(B) = 8/18 \quad \Rightarrow \quad P(RB) = \frac{10}{18}\cdot\frac{8}{18} = 24.69\,\%$

10 B, 8 R (18)	$\Rightarrow$	10 R, 8 B (18)

BR $\quad p(B) = 8/18 \;\cdot\; p(R) = 10/18 \quad \Rightarrow \quad P(BR) = \frac{8}{18}\cdot\frac{10}{18} = 24.69\,\%$

10 R, 8 B (18)	$\Rightarrow$	10 R, 8 B (18)

BB $\quad p(B) = 8/18 \;\cdot\; p(B) = 8/18 \quad \Rightarrow \quad P(BB) = \frac{8}{18}\cdot\frac{8}{18} = 19.75\,\%$ +

100%

then, find the following probabilities,

5.Both are Red. $P(RR) =$ **30.86 %**

6.Both are Blue. $P(BB) =$ **19.75 %**

7. At least one Red. $P(RB)+P(BR)+P(RR) =$ 24.69 + 24.69 + 30.86 = **80.25 %**

8. At most one Red. $P(RB)+P(BR)+P(BB) =$ 24.69 + 24.69 + 19.75 = **69.4 %**

10. One Red and one Blue. $P(RB)+P(BR) =$ 24.69 + 24.69 = **49.38 %**

BP DATE:____________________, Name____________________________________

MULTIPLE CHOICE. Choose the one alternative that best completes the statement or answers the question.

Provide an appropriate response.

1) Which of the following cannot be the probability of an event? 1) ________

A) $\frac{\sqrt{5}}{3}$ B) −32 C) 0 D) 0.001

2) If A, B, C, and D, are the only possible outcomes of an experiment, find the probability of D using the table below. 2) ________

Outcome	A	B	C	D
Probability	1/7	1/7	1/7	

A) 4/7 B) 3/7 C) 1/7 D) 1/4

3) The probability that event A will occur is $P(A) = \frac{\text{Number of successful outcomes}}{\text{Number of unsuccessful outcomes}}$ 3) ________

A) True B) False

4) The probability that event A will occur is $P(A) = \frac{\text{Number of successful outcomes}}{\text{Total number of all possible outcomes}}$ 4) ________

A) False B) True

5) In terms of probability, a(n) ____________________ is any process with uncertain results that can be repeated. 5) ________

A) Experiment B) Event C) Sample space D) Outcome

6) A(n) _______________ of a probability experiment is the collection of all outcomes possible. 6) ________

A) Event set B) Prediction set C) Bernoulli space D) Sample space

7) True or False: An outcome is any collection of events from a probability experiment. 7) ________

A) False B) True

8) In a 1−pond bag of skittles the possible colors were red, green, yellow, orange, and purple. The probability of drawing a particular color from that bag is given below. Is this a probability model? Answer Yes or No. 8) ________

Color	Probability
Red	0.2299
Green	0.1908
Orange	0.2168
Yellow	0.1889
Purple	0.1816

A) Yes B) No

9) An unusual event is an event that has a 9) ________

A) Probability of 1 B) Low probability of occurrence

C) A negative probability D) Probability which exceeds 1

10) The table below represents a random sample of the number of deaths per 100 cases for a certain illness over time. If a person infected with this illness is randomly selected from all infected people, find the probability that the person lives 3–4 years after diagnosis. Express your answer as a simplified fraction and as a decimal. 10) ________

Years after Diagnosis	Number deaths
1–2	15
3–4	35
5–6	16
7–8	9
9–10	6
11–12	4
13–14	2
15+	13

A) $\frac{1}{35}$; 0.029 B) $\frac{35}{100}$; 0.35 C) $\frac{35}{65}$; 0.538 D) $\frac{7}{120}$; 0.058

11) A die is rolled. The set of equally likely outcomes is {1, 2, 3, 4, 5, 6}. Find the probability of getting a 2. 11) ________

A) 0 B) $\frac{1}{6}$ C) 2 D) $\frac{1}{3}$

12) A fair coin is tossed two times in succession. The set of equally likely outcomes is {HH, HT, TH, TT}. Find the probability of getting the same outcome on each toss. 12) ________

A) $\frac{3}{4}$ B) $\frac{1}{4}$ C) 1 D) $\frac{1}{2}$

13) A single die is rolled twice. The set of 36 equally likely outcomes is {(1, 1), (1, 2), (1, 3), (1, 4), (1, 5), (1, 6), (2, 1), (2, 2), (2, 3), (2, 4), (2, 5), (2, 6), (3, 1), (3, 2), (3, 3), (3, 4), (3, 5), (3, 6), (4, 1), (4, 2), (4, 3), (4, 4), (4, 5), (4, 6), (5, 1), (5, 2), (5, 3), (5, 4), (5, 5), (5, 6), (6, 1), (6, 2), (6, 3), (6, 4), (6, 5), (6, 6)}. Find the probability of getting two numbers whose sum is greater than 10. 13) ________

A) $\frac{1}{18}$ B) 3 C) $\frac{5}{18}$ D) $\frac{1}{12}$

14) A single die is rolled twice. The set of 36 equally likely outcomes is {(1, 1), (1, 2), (1, 3), (1, 4), (1, 5), (1, 6), (2, 1), (2, 2), (2, 3), (2, 4), (2, 5), (2, 6), (3, 1), (3, 2), (3, 3), (3, 4), (3, 5), (3, 6), (4, 1), (4, 2), (4, 3), (4, 4), (4, 5), (4, 6), (5, 1), (5, 2), (5, 3), (5, 4), (5, 5), (5, 6), (6, 1), (6, 2), (6, 3), (6, 4), (6, 5), (6, 6)}. Find the probability of getting two numbers whose sum is less than 13. 14) ________

A) $\frac{1}{2}$ B) $\frac{1}{4}$ C) 1 D) 0

15) Three fair coins are tossed in the air and land on a table. The up side of each coin is noted. How many elements are there in the sample space? 15) ________

A) 4 B) 6 C) 8 D) 3

16) In a survey of college students, 880 said that they have cheated on an exam and 1721 said that they have not. If one college student is selected at random, find the probability that the student has cheated on an exam. 16) ________

A) $\frac{880}{2601}$ B) $\frac{2601}{880}$ C) $\frac{1721}{2601}$ D) $\frac{2601}{1721}$

Answer Key
Testname: UNTITLED1

1) B
2) A
3) B
4) B
5) A
6) D
7) B
8) A
9) B
10) B
11) B
12) D
13) D
14) C
15) C
16) A

Ungrouped Data

$$\mu = \frac{\sum x}{N} \quad or \quad \bar{x} == \frac{\sum x}{n} \qquad s = \sqrt{\frac{\sum (x-\bar{x})^2}{n-1}} \quad , \quad s = \sqrt{\frac{n\sum x^2 - (\sum x)^2}{n(n-1)}}$$

Variance (σ^2 , s^2): Variance is the **square of standard deviation.** **To Estimate s**: $s = Range/4$

TI-83/84 *Inputting data in **L1** (stat → Option 1 → enter)*
then *stat → calc → Option 1 → enter → 2nd → 1 → enter*

Grouped Data (Freq. Table)

$$\bar{x} = \frac{\sum (f \times m)}{\sum f} \qquad s = \sqrt{\frac{n\sum (f \times m^2) - (\sum f \times m)^2}{n(n-1)}}$$

TI-83/84 *Inputting **midpoints** in **L1** and **frequency** in **L2***
then *stat → calc → Option 1 → enter → **L1 , L2** → enter*

Empirical Rules: If the **box-plot is centered** then we can apply the **three** following empirical rules.

$68\% = \bar{x} \pm s$ ⇒ **68 %** of data are within 1 s of the mean ($\bar{x}$)
$95\% = \bar{x} \pm 2s$ ⇒ **95 %** of data are within 2 s of the mean ($\bar{x}$)
$99.7\% = \bar{x} \pm 3s$ ⇒ **99.7 %** of data are within 3 s of the mean ($\bar{x}$)

Z -Score $Z = \frac{x-\bar{x}}{s}$ *or* $Z = \frac{x-\mu}{\sigma}$

------------- -2 ---------------------- 0 -------------------- 2 -------------------------

Unusual Values: $Z < -2$ **Ordinary Values:** $-2 \le Z \le 2$ **Unusual Values:** $Z > 2$

$$Correrlation\ \ Coefficient = r = \frac{n\sum xy - \sum x \sum y}{\sqrt{n\sum x^2 - (\sum x)^2}\sqrt{n\sum y^2 - (\sum y)^2}} \qquad -1 \le r \le 1$$

Regression Equation: $\boldsymbol{y = a\,x + b}$ $\boldsymbol{a}$ = Slope , $\boldsymbol{b}$ = y intercept

$$\boldsymbol{Slope = a} = \frac{n(\sum xy) - (\sum x)(\sum y)}{n(\sum x^2) - (\sum x)^2} \qquad \boldsymbol{y\text{ - }itc = b} = \frac{(\sum y)(\sum x^2) - (\sum x)(\sum xy)}{n(\sum x^2) - (\sum x)^2}$$

TI-83/84 *2nd → 0 → **select** Diagnostic on → enter → enter **then** Inputting **x-values** in **L1** and **y-values** in **L2***
then *stat → calc → Option 4 → enter → **L1 , L2** → enter*
Using the regression equation to estimate or predict y and x that are shown by y' and x'

Multiplication Rule $P(A\ and\ B\ and\ C\ and ...) = P(A)P(B)P(C)...$