The institute of cost and Management Accountants of Bangladesh

CMA December 2016 Examination

Foundation level

Subject: 003. Quantitative Techniques

Time: 1.5 hours Full Marks: 50

Part-B: Business Statistics

1(a)	"Without adequate understanding of Statistics, the investor in Social Science may frequently be like the blind man groping in a dark closet for a black cat that is not here." Comment on this statement.
	In the ancient times Statistics was regarded only as the science of Statecraft. Currently, it is not merely a by-product of the administrative set up of the State but it embraces all sciences—social, physical, and natural, and is finding numerous applications in various diversified fields such as agriculture, industry, sociology, biometry, planning, economics, business, management, psychometry, insurance, accountancy and auditing, and so on. It is rather impossible to think of any sphere of human activity where Statistics does not creep in. It will not be exaggeration to say that Statistics has assumed unprecedented dimensions these days and statistical thinking is becoming more and more indispensable every day for an able citizenship. In fact to a very striking degree, the modern culture has become a statistical culture and the subject of Statistics has acquired tremendous progress in the recent past so much so that an elementary knowledge of statistical methods has become a part of the general education in the curricula of many universities all over the world. The importance of Statistics is amply explained in the following words of Carrol D. Wright (1887), United States Commissioner of the Bureau of Labour : "To a very striking degree our culture has become a Statistical culture. Even a person who may never have heard of an index number is affected…by … of those index numbers which describe the cost of living. It is impossible to understand Psychology, Sociology, Economics, Finance or a Physical Science without some general idea of the meaning of an average, of variation, of concomitance, of sampling, of how to interpret charts and tables."
1(b)	Distinguish between with example: (i) Inclusive data and Exclusive data (ii) Sample and Population (iii) Parameter and statistic.

(i) Difference between Inclusive data and Exclusive data

Inclusive Method	**Exclusive Method**
(1) In the inclusive method, the upper limit of a class interval is included in the class itself.	(1) When the upper limit of the class is excluded from the class and is included in the next class, it is called exclusive method.
(2) It is suitable for discrete variables.	(2) It is suitable for continuous variables.
(3) In this, class interval does not overlap e.g.10-19, 20-29, 30-39,etc.	(3) In this, class interval overlaps e.g.10-20, 20-30, 30-40,etc.

(ii) Difference between Sample and Population

A **population** is the entire group that we want to draw conclusions about. A **sample** is a small but representative group that we will collect data from population. The size of the sample is always less than the total size of the population.

In research, a population doesn't always refer to people. It can mean a group containing elements of anything you want to study, such as objects, events, organizations, countries, species, organisms, etc.

(iii) Difference between Parameter and statistic

When you collect data from a population or a sample, there are various measurements and numbers you can calculate from the data. A **parameter** is a measure that describes the whole population. That is any characteristics of population. A **statistic** is a measure that describes the sample. That is any characteristics of sample usually known as statistic.

We can use estimation or hypothesis testing to estimate how likely it is that a sample statistic differs from the population parameter.

1(c) Form a good questionnaire on the basis of the opinion of the inhabitants nearer to the projected Padma bridge about the effects on the socio-economy of the southern area people and the overall economy of the country for the bridge which are under construction over the river Padma.

Questionnaire for households

1. Identify of respondents and demographic information	
1.1 Status of Respondent	
Name of Household	
Village/Ward	
Union	Medinimondol= 1, Kumarvhog=2, Haludia=3, Kolapara=4, Vaggokul=5, Rarikhal= 6, Kathalbari=7, Matborer Char= 8, Kutubpur=9, Naodoba= 10, East-Naodoba=11, outside area=12, Others (Mention specifically)=13.....
Upazila	Louhajanj / Sreenagar / Shibchor / Jajira
District	Munshigonj / Madaripur / Sariatpur
Division	Dhaka

Types of psycho-social vulnerability	Supporting on the level for vulnerability				
	Strongly Agree	Agree	Some What Agree	Some What disagree	Strongly disagree
1.1 My family members are suffering from physical problems because of Padma Multi-purpose Bridge	1	2	3	4	5
1.2 Social and cultural bondage have been broken down due to Padma Multi-purpose Bridge	1	2	3	4	5
1.3 Individual and Social networking have been broken due to Padma Multi-purpose Bridge	1	2	3	4	5
1.4 My household occupation has changed due to PMB	1	2	3	4	5
1.5 After River Bank erosion, social inequality has increased due to PMB	1	2	3	4	5
1.6 Padma Multi-purpose Bridge has created human problem	1	2	3	4	5
1.7 We feel helpless due to Padma Multi-purpose Bridge	1	2	3	4	5
1.8 We face tremendous challenges with our older, pregnant women, specially needy people, widow and children due to Padma Multi-purpose Bridge	1	2	3	4	5

2(a)	What are the various measures of central tendency? Why are they called measures of central tendency?
	Usually mean, median, mode are the measures of central tendency. We can say arithmetic mean, geometric mean, harmonic mean, weighted mean are also measures of central tendency. A measure of central tendency is a single value that attempts to describe a set of data by identifying the central position within that set of data. For that reason they are called measures of central tendency.
2(b)	"Every average has its own peculiar characteristics. It is difficult to say which average is the best." Comment briefly.
	One may ask as to which of these three measures of central tendency the best is. There is no simple answer to this question. It is because these three measures are based upon different concepts. The arithmetic mean is the sum of the values divided by the total number of observations in the series. The median is the value of the middle observation that divides the series into two equal parts. Mode is the value around which the observations tend to concentrate. As such, the use of a particular measure will largely depend on the purpose of the study and the nature of the data; For example, when we are interested in knowing the consumers preferences for different brands of television sets or different kinds of advertising, the choice should go in favour of mode. The use of mean and median would not be proper. However, the median can sometimes be used in the case of qualitative data when such data can be arranged in an ascending or descending order. Let us take another example. Suppose we invite applications for a certain vacancy in our company. A large number of candidates apply for that post. We are now interested to know as to which age or age group has the largest concentration of applicants. Here, obviously the mode will be the most appropriate choice. The arithmetic mean may not be appropriate as it may be influenced by some extreme values. However, the mean happens to be the most commonly used measure of central tendency as will be evident from the discussion.

2(c)

Age Interval	Male(thousands) (Frequency-F1)	Female(thousands) (Frequency-F2)
18-19	121	481
20-24	2441	4184
25-29	5930	6952
30-34	6587	7193
35-44	11788	11893
45-54	9049	9022
55-64	8749	8171
65-74	5786	4654
75 &over	2581	1524

The above table shows the total number of marriages in the USA for male & female of different age groups during a certain year.

(i) Find the median ages of male & females for these marriages.

(ii) Why is the median a more suitable measure of central tendency than the mean in this case?

Age interval Class boundary	Male ('000) f1	Male ('000) F1	Female ('000) f2	Male ('000) F2
18.5-24.5	2562	2562	4665	4665
24.5-34.5	12517	15079	14145	18810
34.5-44.5	11788	26867	11893	30703
44.5-54.5	9049	35916	9022	39725
54.5-64.5	8749	44665	8171	47896
64.5-74.5	5786	50451	4654	52550
75+	2581	53032	1524	54074

(i) Median ages for male = $34.5 + \frac{26516-15079}{11788} * 10 = 34.5 + \left(\frac{11437}{11788}\right) * 10 = 44.2$

	Median ages for female = $34.5 + \frac{27037-18810}{11893} * 10 = 34.5 + \left(\frac{8227}{11893}\right) * 10 = 41.4$ (ii) If any of the class is open then the mid value cannot be calculated which is mandatory for calculating mean. Here last class is open 75+, no upper boundary. But still median can be calculated from this open end class. For that reason in this example median a more suitable measure of central tendency than the mean.
3(a)	What are the absolute & relative measures of dispersion? Explain.
	A measures of dispersion can be used in its absolute form, or in a relative form for comparisons. The measures of absolute dispersion are: 1. Range 2. Quartile deviation 3. Mean deviation 4. Standard deviation The measures of relative dispersion are: 1. Co-efficient of range 2. Coefficient of quartile deviation 3. Coefficient of mean deviation 4. Coefficient of variation
3(b)	When co-efficient of variation is more useful than the standard deviation?
	When we want to compare the variability of the two data sets, which may differ widely in either their averages or which are measured in different units, then the absolute measures of dispersion is not appropriate. In this situation, we usually calculate the relative measures of dispersion which are pure numbers, independent of units of measurement.
3(c)	A manufacturer produce two types of bulbs, A & B, respectively, the bulb have average mean life times of $\bar{Y}_A$ =1495 hours & $\bar{Y}_B$ =1875 hours & standard deviation of S_A=280 hours & S_B = 310 hours. Which bulb has the greater (a) absolute dispersion? (b) relative dispersions? Comment on which bulb is the better?

(a) It is clearly seen from this problem that B types bulb has greater absolute measure of dispersion as standard deviation of S_A=280 hours & S_B = 310 hours.

(b) Coefficient of variation of A= $\frac{280}{1495} * 100 = 18.7\%$

And Coefficient of variation of B= $\frac{310}{1875} * 100 = 16.5\%$

Therefore, relative dispersion is greater in bulb type A. We can say that bulb type B is more stable and uniform.

4(a) Distinguish between correlation and regression analysis? What is scatter diagram? Explain the following values of '*r*' with the help of diagrams:

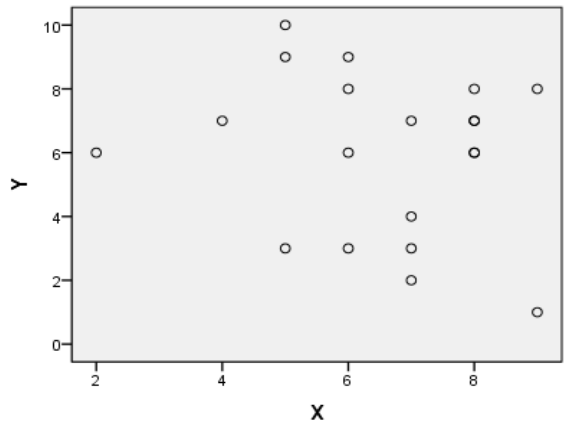
(i) $r = 0$, (ii) $r = -0.8$, (iii) $r = +0.9$.

Distinction between correlation and regression analysis:

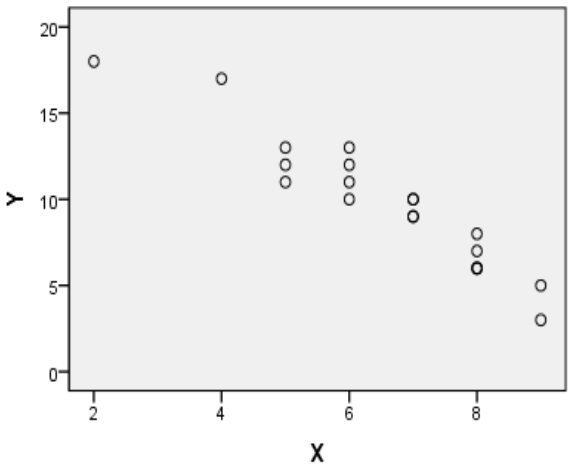
Basis for Comparison	**Correlation**	**Regression**
Meaning	Correlation is a statistical measure that determines the association between two variables.	Regression describes how to numerically relate an independent variable to the dependent variable.
Usage	To represent a linear relationship between variables.	To fit the best line and to estimate one variable based on another.
Dependent and independent variable	No difference	Both variables are different
Indicate	Correlation coefficient indicates the extent to which two variables move together	Regression indicates the impact of a change of unit on the estimated variable (y) in the known variable (x).
Objective	To find a numerical value expressing the relationship between variables.	To estimate values of random variables on the basis of the values of a fixed variables.

Scatter diagram: Scatter diagram is a graphical method of studying correlation. It is the simplest method of ascertaining the correlation between two variables. Let X and Y be two variables, each consisting the same number of values. If we plot the x values along x-axis and corresponding y values along y-axis, we shall get a number of dots on the graph paper. The diagram so obtained consisting all the dots is said to be scatter diagram.

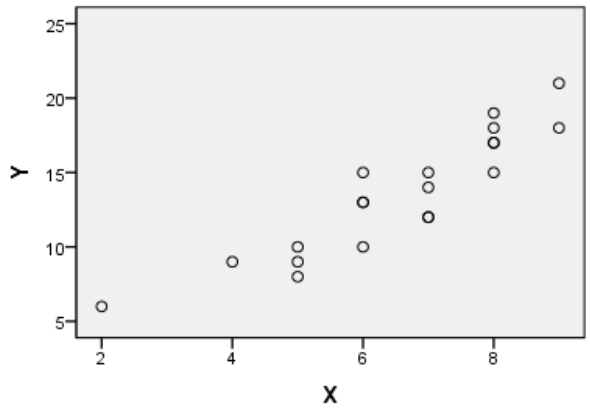
(i) $r = 0$



(ii) $r = -0.8$



(iii) $r = +0.9$



4(b) The following data gives the information on sales and advertising expenses for last 6 months of a particular furniture shop. The data were recorded as follows:

Advertising expense (million \$) x	2	4	5	3	8	6
Sales revenue (million \$), y	17	21	27	18	45	29

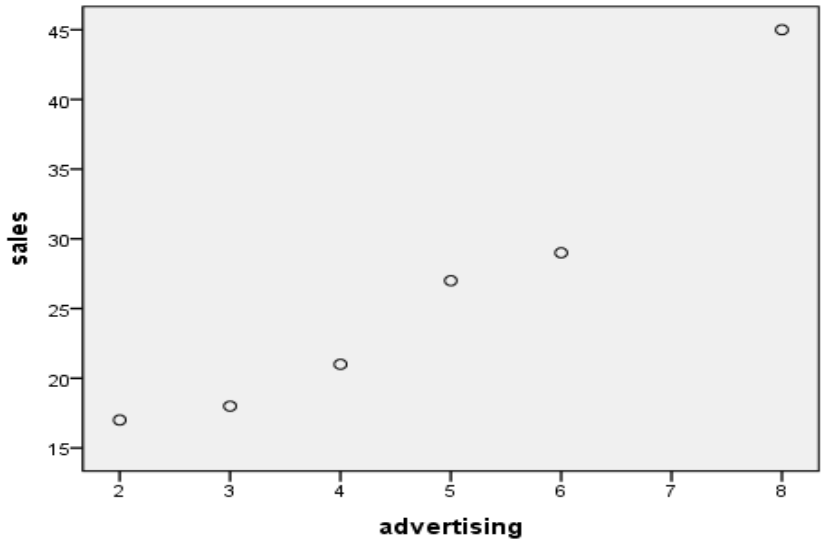
(i) Draw a scatter diagram.

(ii) Fit a linear regression model

(iii) Estimate the sales revenue if advertisement expense is 12 million.

Solution:

(i) A scatter plot of advertising expense and sales revenue is



	(ii) A fitted linear regression model is $$Sales\ revenue = 4.50 + 4.64 * advertising\ expense$$ (iii) If advertising expense is 12 million then the estimated sales revenue is $$Sales\ revenue = 4.50 + 4.64 * 12 = 60.18\ \ million.$$
5(a)	What is a time series? What are the components of time series? Explain briefly.
	A time series is a sequence of numerical data points in successive order. In investing, a time series tracks the movement of the chosen data points, such as a security's price, over a specified period of time with data points recorded at regular intervals. **Time series consist of four components:** **Secular Trends:** The secular trend is the main component of a time series which results from long term effects of socio-economic and political factors. This trend may show the growth or decline in a time series over a long period. This is the type of tendency which continues to persist for a very long period. Prices and export and import data, for example, reflect obviously increasing tendencies over time. **Seasonal Trends:** These are short term movements occurring in data due to seasonal factors. The short term is generally considered as a period in which changes occur in a time series with variations in weather or festivities. For example, it is commonly observed that the consumption of ice-cream during summer is generally high and hence an ice-cream dealer's sales would be higher in some months of the year while relatively lower during winter months. Employment, output, exports, etc., are subject to change due to variations in weather. These types of variations in a time series are isolated only when the series is provided biannually, quarterly or monthly. **Cyclic Movements:** These are long term oscillations occurring in a time series. These oscillations are mostly observed in economics data and the periods of such oscillations are generally extended from five to twelve years or more. These oscillations are associated with the

well-known business cycles. These cyclic movements can be studied provided a long series of measurements, free from irregular fluctuations, is available.

Irregular Fluctuations:

These are sudden changes occurring in a time series which are unlikely to be repeated. They are components of a time series which cannot be explained by trends, seasonal or cyclic movements. These variations are sometimes called residual or random components. These variations, though accidental in nature, can cause a continual change in the trends, seasonal and cyclical oscillations during the forthcoming period. Floods, fires, earthquakes, revolutions, epidemics, strikes etc., are the root causes of such irregularities.

5(b) The following are the annual profit (million taka) in a business firm from 2000 to 2006:

Year:	2000	2001	2002	2003	2004	2005	2006	
Profit(Tk.):	12	9	15	19	26	15	30	

(i) Use the method of least squares to fit a straight line to the above data.

(ii) Estimate the profit for the year 2010 and comment on the estimate

(i) Using the straight line

$$Profit(y) = a + b * Year(x) + \varepsilon$$

The fitted line is $\widehat{Profit} = \hat{a} + \hat{b} * year = -5{,}490 + 2.75\ year$

Where a and b are estimated using least square method.

(ii) The estimated profit for the year 2010 is

$$\widehat{Profit} = -5{,}490 + 2.75 * year = -5{,}490 + 2.75 * 2010 = 37.5\ million\ taka$$

6(a) What is hypothesis? What steps are necessary to follow hypothesis testing as statistical tools?

Hypothesis:- A hypothesis is an assertion or statement about a population or equivalently about the probability distribution characterizing a population, which we want to verify on the basis of information contained in a sample. For examples:

(1) A physician may hypothesize that the recommended drug is effective in 90 percent cases.

(2) A sewing machine company claims that their new machine is superior to the one available in the market.

Seven steps procedures for testing a hypothesis:

Step 1: State the Null Hypothesis

The null hypothesis can be thought of as the opposite of the "guess" the research made. For example A recent article indicated the mean age of US *commercial* aircraft is 15 years. To conduct a statistical test regarding this estimate, the first step is to determine the null and the alternate hypothesis We state the Null hypothesis as:

The null hypothesis represents the current condition. it is written by Ho: μ =15

Step 2: State the Alternative Hypothesis

The reason we state the alternative hypothesis this way is that if the Null is rejected. We state the alternative hypothesis as:

The alternate hypothesis is that the statement is not true, that is written by H1: $\mu \neq 15$

Step 3: Select level of significance

Level of significance is the probability of rejecting the null hypothesis when it is true. The level of significance is denoted by the Greek letter alpha. It is also sometimes called the level of risk. There is no one level of significance that is applied to all tests. Traditionally 0.05 level is selected for the consumer research project, 0.01 for quality assurance and 0.10 for political polling.

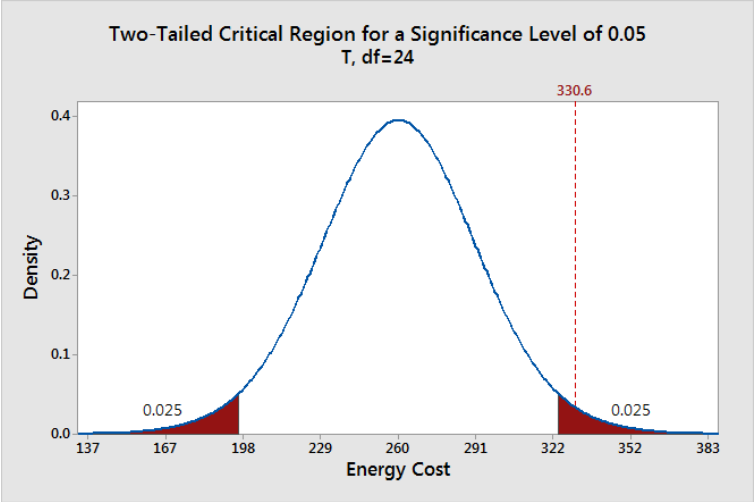
Step 4: Collect Data

Remember the importance of recognizing whether data is collected through experimental design or observational.

Step 5: Calculate a test statistic

A value, determined from sample information, used to determine whether to reject the null hypothesis. Every test statistic follows a particular distribution under null hypothesis. They usually follow Z, t, F and so on.

Step 6: Set decision rule

	A decision rule is a statement of conditions under which the null hypothesis rejected or the conditions under which it is not rejected. In this step, we are to define acceptance and rejection region of Ho. **Step 7: Based on steps 5 and 6, draw a conclusion about H0** The last step is to make decision on the basis of the calculated value of test statistic to find it we are to take random sample.
6(b)	Explain the level of significance with example.
	The significance level, also denoted as alpha or α, is the probability of rejecting the null hypothesis when it is true. For example, a significance level of 0.05 indicates a 5% risk of concluding that a difference exists when there is no actual difference.  In the graph above, the two shaded areas are equidistant from the null hypothesis value and each area has a probability of 0.025, for a total of 0.05. In statistics, we call these shaded areas the *critical region* for a two-tailed test. If the population mean is 260, we'd expect to obtain a sample mean that falls in the critical region 5% of the time. The critical region defines how far away our sample statistic must be from the null hypothesis value before we can say it is unusual enough to reject the null hypothesis.
6(c)	You are working as a purchase manager for a Company. The following information has been supplied to you by two manufacturers of electric bulbs.

Particulars	Company -A	Company-B
Mean life (in hours)	1300	1200
Standard deviation	80	90
Sample size	110	110

	Which brand of bulbs are you going to purchase if your desire to take a risk of 5%?
	Solution: Null hypothesis $H_0: \mu_A = \mu_B$ Versus alternative $H_1: \mu_A \neq \mu_B$ Test statistic $Z = \frac{\overline{x_A} - \overline{x_B}}{\sqrt{\frac{SD_A^2}{n_A} + \frac{SD_B^2}{n_B}}} = \frac{1300-1200}{\sqrt{\frac{80^2}{110} + \frac{90^2}{110}}} = \frac{100}{\sqrt{58.2+73.6}} = \frac{100}{11.48} = 8.71$ Since the calculated value of Z=8.71 is much greater than tabulated value of Z=1.96 at 5% risk, we may reject null hypothesis of equal life length. That is there exist significant difference between two mean lives of bulb. Therefore, I am going to purchase brand A company's bulb as its mean life is significantly higher than the mean life of brand B.
7(a)	Define normal distribution with their important properties. What type of errors are committed in testing hypothesis? What about the power of the test?
	Normal distribution: A continuous random variable *X* is said to have a normal distribution if its probability density function is given by $$f(x;\mu,\sigma^2) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{1}{2\sigma^2}(x-\mu)^2} \; ; \; -\infty < x < \infty$$ where the parameters μ and σ^2 satisfy $-\infty < \mu < \infty$, $\sigma^2 > 0$. The parameters μ and σ^2 are actually the mean and variance of the normal variable *X*. **Properties or characteristics of normal distribution:-** 1. The curve of the distribution is symmetrical about the point $x = \mu$ and it is bell shaped. 2. For normal distribution mean, median and mode are same, which is equal to μ. 3. For normal distribution, skewness and kurtosis are $\beta_1 = 0$ and $\beta_2 = 3$. 4. Linear combination of independent normal variates is also a normal variate. 5. The curve approaches nearer and nearer to the base but it never touches it, i.e., the curve is asymptotic to the base on either side. Hence the ranges are unlimited or infinite in both directions. 6. Under certain condition most of the distribution tends to normal distribution. 7. The area under the normal curve is distributed as follows: (a) 68.26% of the time, a normal random variable assumes a value within plus or minas 1 standard deviation of its mean. (b) 95.44% of the time, a normal random variable assumes a value within plus or minas 2 standard deviation of its mean.

	(c) 99.72% of the time, a normal random variable assumes a value within plus or minas 3 standard deviation of its mean. Usually two types of errors occur when hypothesis testing- **Type I error:-** The error of rejecting H_0 (accepting H_1) when H_0 is true is called type I error. The probability of type I error is denoted by α and it is called the level of significance. **Type II error:-** The error of accepting H_0 when H_0 is false (H_1 is true) is called type II error. The probability of type II error is denoted by β. **Power of the test:-** $1-\beta$, that is the probability of rejecting H_0 when H_0 is false (H_1 is true) is called the power of the test hypothesis H_0 against the alternative hypothesis H_1.
7(b)	An internet server claims that its users spend on the average 20 hours per week with a standard deviation of 3 hours on the information superhighway. To determine whether this is an overestimate, a competitor conducted a sample survey of 15 customers and found that the average time spent online was 22 hours per week. Do the data provide sufficient evidence to indicate that the average hours of use are less than that claimed by the first internet? Test at 5% level.
	Solution: Here given that $SD = 3,\ n = 15,\ \bar{x} = 22$ Null hypothesis $H_0: \mu \geq 20$ versus $H_1: \mu < 20$ Test statistic, $Z = \frac{\bar{x}-\mu_0}{SD/\sqrt{n}} = \frac{22-20}{3/\sqrt{15}} = 2.58$ Since calculated value of z = 2.58 is greater than tabulated value of – z = - 1.645 so we may accept the null hypothesis. That is average hours of internet use are not less than 20 hours per week.