The institute of cost and Management Accountants of Bangladesh

CMA June 2016 Examination

Foundation level

Subject: 003. Quantitative Techniques

Time: 1.5 hours Full Marks: 50

Part-B: Business Statistics

1(a)	Define business statistics with its important applications. Describe the scope of business Statistics.
	Business statistics is a specialty area of statistics which are applied in the business setting. It can be used for quality assurance, financial analysis, production and operations, and many other business areas. Just as in general statistics, there are two categories: descriptive and inferential. **Descriptive statistics** are used to describe the total group of numbers. **Inferential statistics** infers relationships from the population of numbers. ***<u>Scope and importance of Statistics:</u>*** 1. *Statistics and planning:* Statistics in indispensable into planning in the modern age which is termed as "the age of planning". Almost all over the world the govt. are re-storing to planning for economic development. 2. *Statistics and economics:* Statistical data and techniques of statistical analysis have to immensely useful involving economical problem. Such as wages, price, time series analysis, demand analysis. 3. *Statistics and business:* Statistics is an irresponsible tool of production control. Business executive are relying more and more on statistical techniques for studying the much and desire of the valued customers. 4. *Statistics and industry:* In industry statistics is widely used inequality control. In production engineering to find out whether the product is confirming to the specifications or not. Statistical tools, such as inspection plan, control chart etc. 5. *Statistics and mathematics:* Statistics are intimately related recent advancements in statistical technique are the outcome of wide applications of mathematics. 6. *Statistics and modern science:* In medical science the statistical tools for

	collection, presentation and analysis of observed facts relating to causes and incidence of dieses and the result of application various drugs and medicine are of great importance. 7. *Statistics, psychology and education:* In education and physiology statistics has found wide application such as, determining or to determine the reliability and validity to a test, factor analysis etc. 8. Statistics *and war:* In war the theory of decision function can be a great assistance to the military and personal to plan "maximum destruction with minimum effort."
1(b)	What is frequency? What do you mean by a frequency distribution? How will you prepare a frequency distribution from raw data?
	Frequency: When summarizing large masses of data, it is often useful to distribute the data into classes or categories and to determine the number of individuals belonging to each class, called the class frequency. **Frequency Distribution:** A tabular arrangement of data by classes together with the corresponding class frequencies is called frequency distribution, or frequency table. **Construction of Frequency Distribution:-** *Step 1:* Determine the number of classes:- To determine number of class we use the formula- $$K = 1 + 3.322 \log_{10} N$$ Here, K is the number of class and N is the number of observations. *Step 2:* Determine the class interval or Width:- The classes all taken together must cover at least the distance from the lowest value in the raw data up to the highest value. Generally we can use the formula to determine number of class interval. $$i \geq \frac{H - L}{K}$$ Here, i is the class interval, H is the highest observed value, L is the lowest observed value and

	K is the number of classes. *Step 3:* Use Tally:- A bar (|) called tally mark is put against the number when it occurs. Having occurred four times, the fifth occurrence is represented by putting a cross tally (/) on the first four tallies. This technique facilitates the counting of the tally marks at the end. *Step 4:* Count the number of items in each class:- The number of observation in each class is called class frequency. This is obtained by counting the tallies. Step 5: Relative frequency:- Relative frequency refers to the ratio of the number of frequency of a certain class and total number of frequency existing in a frequency distribution. $\text{Re}\,lative\ \ frequency = \dfrac{Frequency\ \ of\ \ a\ certain\ class}{Total\ \ Frequency}$. Step 6: Cumulative frequency:- The total frequency of all values less than the upper class boundary of a given class interval is called the cumulative frequency upto and including that class interval.
1(c)	What is the necessity of graphs in business statistics? The following data show the monthly expenditure of a family:
	Necessity of graphs in business statistics:

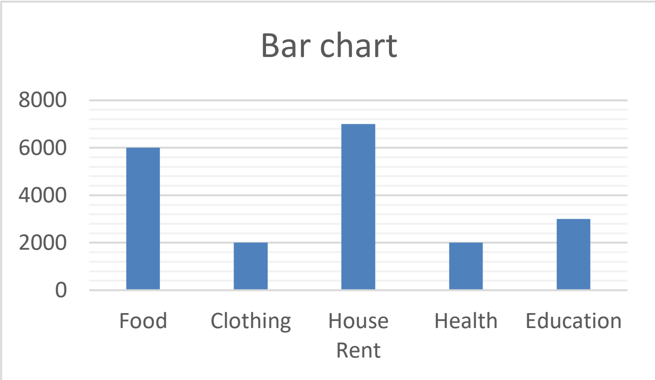
Item	**Food**	**Clothing**	**House Rent**	**Health**	**Education**
Expenditure (Tk.)	6000	2000	7000	2000	3000

Present the data by a suitable diagram.

Necessity of graphs in business statistics:

Walk into almost any business meeting and you'll see one of these talked about at some point. It's either a graph or a chart describing something about the business. It could be a chart showing the progress the team is making on a big project. Or it could be a graph showing the sales of the business and comparing it with the sales of the competition. Either way, these graphs and charts make the information much easier to digest and understand.

A graph or a chart may be defined as a visual presentation of data. For example, a utility company uses a column chart to help its customers see just how much energy they've used during the last billing cycle. A bakery may use a pie chart to show how many breads it sells when compared to its other products, such as cheesecakes and apple pies.



2(a) What do you mean by measures of central tendency? What are their measures? Among all measures of central tendency which one is the best and why?

A measure of central tendency is a summary statistic that represents the center point or typical value of a dataset. These measures indicate where most values in a distribution fall and are also referred to as the central location of a distribution. We can think of it as the tendency of data to cluster around a middle value. In statistics, the three most common measures of central tendency are the mean, median, and mode. Each of these measures calculates the location of the central point using a different method.

Choosing the best measure of central tendency depends on the type of data we have. When we have a symmetrical distribution for continuous data, the mean, median, and mode are equal. In this case, analysts tend to use the mean because it includes all of the data in the calculations. However, if we have a skewed distribution, the median is often the best measure of central tendency. When we have ordinal data, the median or mode is usually the best choice. For categorical data, we have to use the mode.

In cases where we are deciding between the mean and median as the better measure of central tendency, we are also determining which types of statistical hypothesis tests are appropriate for our data—if that is our ultimate goal.

2(b) Discuss when geometric mean is appropriate as a measure of central tendency. A sales man makes a trip by car for 6 days. Each day he travels 220 kilometers. His average

	speeds for 6 days are 48, 36, 42, 48, 30 and 36 kilometers/hour. What is average speed for the whole trip?
	When we need to find the average value of some data like rate, ratio, proportion or percentage, then instead of arithmetic mean we usually use geometric mean to find appropriate central value. For example if we find average growth of last 5 monthly sales of a salesman, then we need to use geometric mean. To find the average speed we usually use harmonic mean, The average speed is $HM = \frac{6}{\frac{1}{48}+\frac{1}{36}+\frac{1}{42}+\frac{1}{48}+\frac{1}{30}+\frac{1}{36}} = \frac{6 \times 5040}{778} = 38.87 \ kilometers/hour$
3(a)	What is dispersion? What are different measures of variation? Why standard deviation is considered to be the best measure?
	The degree to which numerical data tend to spread about an average value is called the dispersion or variation of the data. **Measures of Dispersion:-** 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

Standard deviation is considered to be the best measure of dispersion and is therefore, the most widely used measure of dispersion.

(i) It is based on all values and thus, provides information about the complete series. Because of this reason, a change in even one value affects the value of standard deviation.

(ii) It is independent of origin but not of scale.

(iii) It is useful in advance statistical calculations like comparison of variability in two data sets.

(iv) It can be used in testing of hypothesis.

(v) It is capable of further algebraic treatment.

3(b) Two companies lamps are given bellow:

Length of life (hrs)	700-900	900-1100	1100-1300	1300-1500	
Lamp of company A	10	16	26	8	
Lamp of company B	3	42	12	3	

(i) Which company's bulbs give a higher average life?

(ii) (ii)Which company' bulbs are the best?

Solution:

For Lamp of company A,

Life in hours ('00)	f_1	Mid-value (x_i)	$f_1 x_i$	$f_1 x_i^2$
7 – 9	10	8	80	640
9 – 11	16	10	160	1600
11 – 13	26	12	312	3744
13 - 15	8	14	112	1568
Total	60		664	7552

$$Mean = \frac{664}{60} = 11.07\ ('00)\ hours = 1107\ hours$$

$$Standard\ deviation = \sqrt{\frac{7552}{60} - 11.07^2} = \sqrt{125.87 - 122.54}$$

$$= 1.82\ ('00) = 182\ hours$$

$$Coefficient\ of\ Variation = \frac{SD}{mean} * 100\% = \frac{182}{1107} * 100\% = 16.4\%$$

For Lamp of company B,

Life in hours ('00)	f_1	Mid-value (x_i)	$f_1 x_i$	$f_1 x_i^2$
7 – 9	3	8	24	192
9 – 11	42	10	420	4200
11 – 13	12	12	144	1728
13 - 15	3	14	42	588
Total	60		630	6708

$$Mean = \frac{630}{60} = 10.5\ ('00) hours = 1050\ hours$$

$$Standard\ deviation = \sqrt{\frac{6708}{60} - 10.5^2} = \sqrt{111.8 - 110.25} = 1.24\ ('00)$$

$$= 124\ hours$$

$$Coefficient\ of\ Variation = \frac{SD}{mean} * 100\% = \frac{124}{1050} * 100\% = 11.8\%$$

(i) The bulb of company A shows higher average life.

(ii) But the bulb of company B is more consistent and uniform than company A (as CV of B, 11.8% which is lower than CV of B, 16.4%). Therefore the bulb of company B is the best.

4(a)	When you will use co-efficient of variation instead of standard deviation to compare the better performance between the two groups.

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.

Coefficient of Variation:

$$CV = \frac{SD}{mean} \times 100\%$$

4(b)	Differentiate between correlation and regression.

Difference between correlation and regression:

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.

4(c) Prove that -1≤r≤1, where r is the correlation of co-efficient between two variables.

Proof:

Let us consider the following expression

$$\left(\frac{X_i - \bar{X}}{s_x} \pm \frac{Y_i - \bar{Y}}{s_y}\right)^2 \text{ which is always positive.}$$

In other words,

$$\left(\frac{X_i - \bar{X}}{s_x} \pm \frac{Y_i - \bar{Y}}{s_y}\right)^2 \geq 0$$

Performing the square and summing,

	$$\frac{\sum(X_i-\bar{X})^2}{s_x^2}+\frac{\sum(Y_i-\bar{Y})^2}{s_y^2}+\frac{2\sum(X_i-\bar{X})(Y_i-\bar{Y})}{s_x\,s_y}\geq 0$$ Or $$\frac{ns_x^2}{s_x^2}+\frac{ns_y^2}{s_y^2}+\frac{2nr_{xy}s_xs_y}{s_x\,s_y}\geq 0$$ From which we can write, $1\pm r_{xy}\geq 0$ When $1+r_{xy}\geq 0$, $r_{xy}\geq -1$. Again when $1-r_{xy}\geq 0$, $r_{xy}\leq +1$ Therefore, we can say, $-1\leq r_{xy}\leq +1$
4(d)	If the correlation between two variables, r is 0.90, find the co-efficient of determination & interpret the meaning of it.
	$coefficient\ of\ determination = r^2 = 0.9^2 = 0.81$ which means that 81% of the total variation of a dependent variable can be explained by the linear relationship between two variables.
4(e)	Define expected value of a discrete random variable x with probability distribution p(x). Given that, <table><tr><td>x</td><td>3</td><td>5</td><td>7</td><td>9</td></tr><tr><td>P(x)</td><td>0.20</td><td>0.35</td><td>0.16</td><td>0.29</td></tr></table>Find expected value of x.
	Solution: $E(X)=\sum x*P(x)=3*0.2+5*0.35+7*0.16+9*0.29=6.08$ Hence the expected value of x is 6.08
5(a)	Distinguish between budget & forecast.
	The key difference between a budget and a forecast is that a budget lays out the plan for what a business wants to achieve, while a forecast states its actual expectations for results, usually in a much more summarized format. In essence, a budget is a quantified expectation for what a business wants to achieve. Its characteristics are: • The budget is a detailed representation of the future results, financial position, and cash flows that management wants the business to achieve during a certain period of time.

	• The budget may only be updated once a year, depending on how frequently senior management wants to revise information. • The budget is compared to actual results to determine variances from expected performance. • The budget to actual comparison can trigger changes in performance-based compensation paid to employees. Conversely, a forecast is an estimate of what will actually be achieved. Its characteristics are: • The forecast is typically limited to major revenue and expense line items. There is usually no forecast for financial position, though cash flows may be forecasted. • The forecast is updated at regular intervals, perhaps monthly or quarterly. • The forecast may be used for short-term operational considerations, such as adjustments to staffing, inventory levels, and the production plan. • Changes in the forecast do not impact performance-based compensation paid to employees.
5(b)	Identify the components of a time series model.
	Components for Time Series Model: The various reasons or the forces which affect the values of an observation in a time series are the components of a time series. The four categories of the components of time series are • Trend • Seasonal Variations • Cyclic Variations • Random or Irregular movements
5(c)	What are the requirements of a good forecasting system?
	The requirements of a good forecasting system should satisfy the following criteria: • Time frame • Pattern of data • Cost/Economy of forecasting • Accuracy desired • Availability of data • Plausibility of operation and understanding • Durability

- Flexibility

5(d) Fit a straight line trend by the using the method of least square.

Years	2007	2008	2009	2010	2011	2012	2013	2014
Profits(millions)	101	100	105	112	114	120	124	134

Solution:

Using the straight line

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

The fitted line is $\widehat{Profit} = \hat{a} + \hat{b} * year = -9{,}412.2 + 4.74\, year$

Where a and b are estimated using least square method.

6(a) Define and explain.(i) 5% level of significance (ii) Type-1 and type-2 error (iii) Critical value (iv) Null and alternative hypothesis.

(i) **5% level of significance**: The researcher determines the significance level before conducting the experiment. The significance level 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.

(ii) **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.

(iii) **Critical value:** In hypothesis testing, a critical value is a point on the test distribution that is compared to the test statistic to determine whether to reject the null hypothesis. If the absolute value of your test statistic is greater than the critical value, you can declare statistical significance and reject the null hypothesis.

(iv) **Null and alternative hypothesis**: The null hypothesis is a general statement that states that there is no relationship between two phenomenons under consideration or that there is no association between two groups. An alternative hypothesis is a statement that describes that there is a relationship between two selected variables in a study.

6(b)	The nine items of a sample had the following values – 45, 47, 50, 52, 48, 47, 49, 53, 50. The sample mean is 49 and the sum of squares of deviation taken from mean is 52. Can this sample be regarded as taken from the population having 47 as mean? Also find 95% confidence interval. The table value of t for 8 d.f. at 5% level is 2.31.
	Solution: Here $mean\ \bar{x} = 49,\ \ \sum(x-\bar{x})^2 = 52\ \ then\ S^2 = \frac{\sum(x-\bar{x})^2}{n-1} = \frac{52}{8} = 6.2$ $S = \sqrt{6.2} = 2.49$ Null hypothesis, $H_0: \mu = 47$ versus $H_1: \mu \neq 47$ Test statistic, $t = \frac{\bar{x}-\mu_0}{\frac{S}{\sqrt{n}}} = \frac{49-47}{2.49/3} = \frac{6}{2.49} = 2.41$ Since calculated value of t=2.41 is greater than tabulated value of t=2.31, we may reject the null hypothesis. Thus we can conclude that the sample can't be regarded as taken from the population having mean 47. 95% confidence interval can be constructed as $\bar{x} \pm t_{8,0.05} * \frac{S}{\sqrt{n}} = 49 \pm 2.31 * \frac{2.49}{3} = 49 \pm 1.92 = (49-1.92\ , 49+1.92)$ $= (47.1\ , 50.9)$
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*. If *X* is a normal variable with parameters μ and σ^2, then $Z = \frac{X-\mu}{\sigma}$ is a standard normal variable with mean zero and variance one.

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.