The institute of cost and Management Accountants of Bangladesh
CMA December 2017 Examination
Foundation level
Subject: 003. Quantitative Techniques

Time: 1.5 hours | Full Marks: 50

Part-B: Business Statistics

1(a)	Define statistics? Discuss the uses of statistical methods in modern business organizations.
	Statistics is an important field of study because of its application in almost all walks of life. It is an important branch of mathematics. It is the analysis, interpretation, preservation, and presentation of data. The role of statistics in business management is pivotal. It comes in handy, especially for rating organization, financial markets, financial organizations, etc. **Uses of statistical methods in modern business organization** There are several ways in which statistics plays a role in business. **Management of Performance** – As mentioned earlier, the importance of statistics in management is significant; with the help of statistics, a manager can analyze the performance or the productivity of employees like the units produced or the task completed. The manager can use the data in sync with statistical techniques to improve the productivity of the workforce and multiply the production. **Alternative Scenarios** – The task or the function of a manager does not end after increasing the productivity of the employees. A manager has to participate with the other managers from different department for decision making. The decision can be on the choice of particular software, systems for customer automatic ordering systems, etc. **Data Collection** – The data that is collected for the purpose of processing with the statistical tools must be done in an ethical manner, otherwise, the result of the analysis will be false and non-beneficial. With the help of these data, comparisons can be drawn if the actual sales were less or more than the projected sales or the future capital requirement for the fulfilment of a huge order.

	Research And Development – The scope of statistics in business also extends to market research and product development. This is one of the most important functions of statistics, as a sample group is observed and their response to a product is tested and data collected. This data is essential in the determination of the launch of new products and the development of it.
1(b)	Define variable with its classification. Discuss different measurement scale used in statistics and mention which variable includes in which scale: (i) Height, (ii) IQ score, (iii) Economic status, (iv) Jersey number, (v) Income.
	Variable: The characteristic which varies over the units (or from unit to unit) is called variable. For example, family size, father's occupation of the students of Dhaka University. Classification of variable:- We can classify the variables according to the characteristics into two groups: • Qualitative variable • Quantitative variable **Qualitative variable**: There are many characteristics that cannot be expressed in any numerical form (e.g., color, sex, occupation etc.) but we can arrange them according to their quality or attribute. These types of characteristic are known as qualitative variable. For example, father's occupation of the students of Dhaka University. **Quantitative variable**: The characteristics of an unit or item that are expressed in numerical form or in numbers are called quantitative variable. For example, monthly family income of the students of Dhaka University. To measure a variable there are various ways. We classify the ways into four scales of measurement. **Nominal scale**: The measurement scale in which numbers are assigned to the categories or variable values for identification only is called a nominal scale. For example, religion, color, sex etc. **Ordinal scale**: The measurement scale in which numbers are assigned to the categories or variable values for identification as well as ranking is called an ordinal scale. For example, economic status. We can order the economic status in high, middle and low as follows:

	Interval scale: The measurement scale in which numbers are arranged to the variable values in such a way that the level of measurement is broken down on a scale of equal units and the zero value on the scale is not absolute zero is called an interval scale. For example, temperature, I. Q. score. Here, thermometer records temperature in terms of degrees and a 1 degree change (increase or decrease) in temperature implies the same amount of heat. **Ratio scale**: The measurement scale in which numbers are assigned to the variable values in such a way that the level of measurement is broken down on a scale of equal units and the zero value on the scale is absolutely zero is called a ratio scale. For example, height, weight etc. (i) Height: Ratio scale (ii) IQ score: Interval scale (iii) Economic status: Ordinal scale (iv) Jersey number: Nominal scale (v) Income: Ratio scale
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)	In an examination of 675 candidates the examiner submitted the following information:

Marks obtained	No. of candidates
Less than	
10%	7
20%	39
30%	95
40%	201
50%	381
60%	545
70%	631
80%	675

Calculate the mode and median of the percentage marks obtained.

Solution:

Class of Marks (%)	No. of Candidates (f)	Cumulative no. of Candidates (F)
0 – 10	7	7
10 – 20	32	39
20 – 30	56	95
30 – 40	106	201
40 – 50	180	381
50 – 60	164	545
60 – 70	86	631
70 – 80	44	675
Total	675	

$$Mode = 40 + \frac{74}{74+16} = 40 + 8.2 = 48.2\%$$

$$Median = 40 + \frac{337.5-201}{180} * 10 = 40 + \frac{1365}{180} = 40 + 7.6 = 47.6\%$$

3(a)	Define mean deviation and standard deviation. Why standard deviation is the best measure of dispersion?
	Standard Deviation Standard deviation is the most common measure of variability and is frequently used to determine the volatility of markets, financial instruments, and investment returns. To calculate the standard deviation: 1. Find the mean, or average, of the data points by adding them and dividing the total by the number of data points. 2. Subtract the mean from each data point and square the difference of each result. 3. Find the mean those squared differences and then the square root of the mean. Squaring the differences between each point and the mean avoids the issue of negative differences for values below the mean, but it means the variance is no longer in the same unit of measure as the original data. Taking the square root of means the standard deviation returns to the original unit of measure and is easier to interpret and use in further calculations. **Mean Absolute Deviation** The average deviation, or mean absolute deviation, is calculated similarly to standard deviation, but it uses absolute values instead of squares to circumvent the issue of negative differences between the data points and their means. To calculate the average deviation: 1. Calculate the mean of all data points. 2. Calculate the difference between the mean and each data point. 3. Calculate the average of the absolute values of those differences. Standard deviation is often used to measure the volatility of returns from investment funds or strategies because it can help measure volatility. Higher volatility is generally associated with higher risk of losses, so investors want to see higher returns from funds that generate higher volatility. For example, a stock index fund should have relatively low standard deviation compared with a growth fund.

	The mean average, or mean absolute deviation, is considered the closest alternative to standard deviation. It is also used to gauge volatility in markets and financial instruments, but it is used less frequently than standard deviation. Generally, according to mathematicians, when a data set is of normal distribution — that is, there aren't many outliers — standard deviation is the preferable gauge of variability. But when there are large outliers, standard deviation will register higher levels of dispersion, or deviation from the center, than mean absolute deviation.
3(b)	Life of two models of refrigerators in a recent survey are:

Life in years	Number of refrigerators	
	Model A	Model B
0 – 2	5	2
2 – 4	16	7
4 – 6	13	12
6 – 8	7	19
8 – 10	5	9
10 - 12	4	1

What is the average life of each refrigerator? Which model of refrigerator is more stable and why?

Solution:

For Model A,

Life in years	f_1	Mid-value (x_i)	$f_1 x_i$	$f_1 x_i^2$
0 – 2	5	1	5	5
2 – 4	16	3	48	144
4 – 6	13	5	65	325
6 – 8	7	7	49	343
8 – 10	5	9	45	405
10 – 12	4	11	44	484
Total	50		256	1706

$$Mean = \frac{256}{50} = 5.12\ years$$

	$$Standard\ deviation = \sqrt{\frac{1706}{50} - 5.12^2} = \sqrt{34.12 - 26.21} = 2.81\ years$$ $$Coefficient\ of\ Variation = \frac{SD}{mean} * 100\% = \frac{2.81}{5.12} * 100\% = 54.9\%$$ For Model B,

Life in years	f_2	Mid-value (x_i)	f_2x_i	$f_2x_i^2$
0 – 2	2	1	2	2
2 – 4	7	3	21	63
4 – 6	12	5	60	300
6 – 8	19	7	133	931
8 – 10	9	9	81	729
10 – 12	1	11	11	121
Total	50		308	2146

$$Mean = \frac{308}{50} = 6.16\ years$$

$$Standard\ deviation = \sqrt{\frac{2146}{50} - 6.16^2} = \sqrt{42.92 - 37.95} = 2.23\ years$$

$$Coefficient\ of\ Variation = \frac{SD}{mean} * 100\% = \frac{2.23}{6.16} * 100\% = 36.2\%$$

Therefore, the average life of model A refrigerator is 5.12 years and model B refrigerator is 6.16 years. Since coefficient of variation of model A refrigerator is higher than model B, thus model B refrigerator is more stable than model A.

4(a)	Distinguish 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.

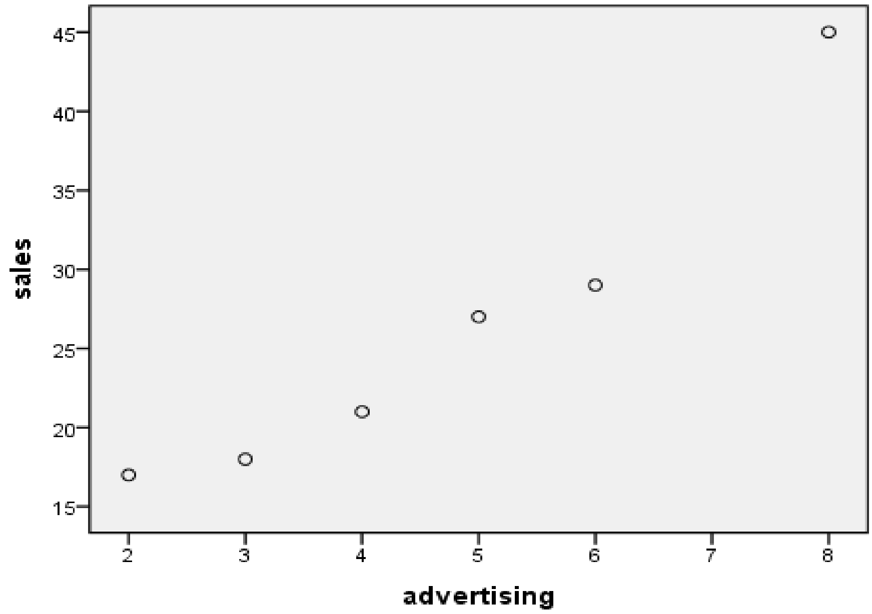
4(b) The following data gives the information on sales and advertising expenses for last 6 months of a particular furniture shop. The data were recorded as follows:

Advertising expenses (million $)	2	4	5	3	8	6
Sales revenue (million $)	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 20 lac taka.

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 20 million then the estimated sales revenue is

$$Sales\ revenue = 4.50 + 4.64 * 20 = 97.3\ \ million\ (\$).$$

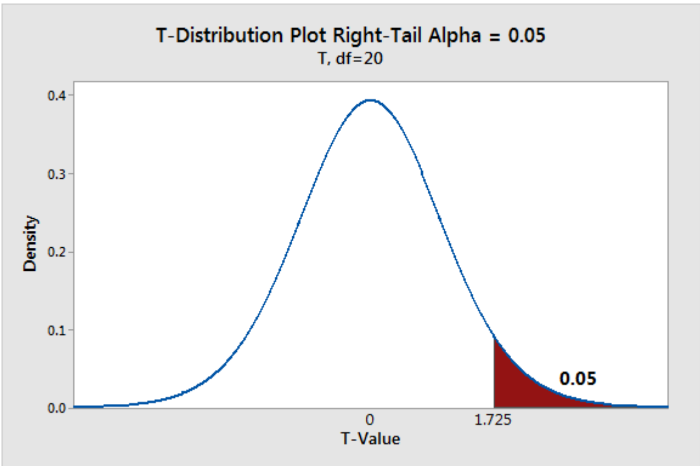
5(a) What is a time series? What are the components of time series? Explain briefly.

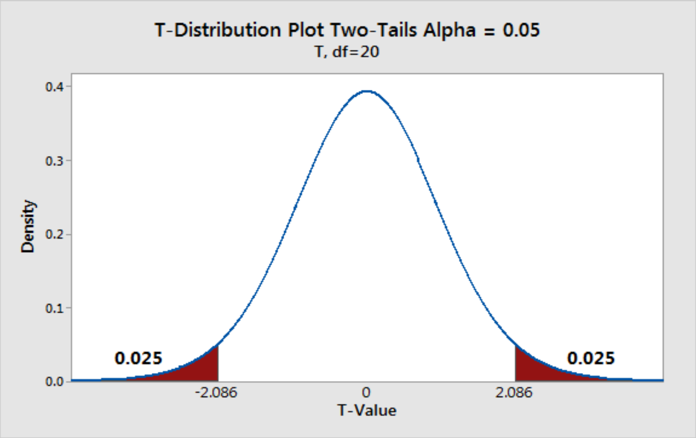
A time series is a sequence of numerical data points in successive order. In investing, a time series tracks the movement of the chosen data points, such as a security's price, over a specified period of time with data points recorded at regular intervals.

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 (in million) 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)	Define critical value with examples and distinguish between one tail test and two tail test graphically.
	Critical values are essentially cut-off values that define regions where the test statistic is unlikely to lie; for example, a region where the critical value is exceeded with probability α if the null hypothesis is true. The null hypothesis is rejected if the test statistic lies within this region which is often referred to as the rejection region. **One-tailed test:-** A test of any statistical hypothesis where the alternative is one-sided such as $H_0 : \mu = \mu_0$ $H_1 : \mu > \mu_0$ or perhaps, $H_0 : \mu = \mu_0$ $H_1 : \mu < \mu_0$ is called a one-tailed test.  **Two-tailed test:-** A test of any statistical hypothesis where the alternative is two-sided such as $H_0 : \mu = \mu_0$ $H_1 : \mu \neq \mu_0$

	is called a two-tailed test. 
6(b)	Explain the following: (i) Random variable (ii) random experiment (iii) mutually exclusive events and (iv) equally likely events.
	(i) **Random variable** A random variable is a variable whose value is unknown or a function that assigns values to each of an experiment's outcomes. Random variables are often designated by letters and can be classified as discrete, which are variables that have specific values, or continuous, which are variables that can have any values within a continuous range. Random variables are often used in econometric or regression analysis to determine statistical relationships among one another. (ii) R**andom experiment** An experiment or trial is any procedure that can be infinitely repeated and has a well-defined set of possible outcomes, known as the sample space. An experiment is said to be random if it has more than one possible outcome, and deterministic if it has only one. A random experiment that has exactly two (mutually exclusive) possible outcomes is known as a Bernoulli trial (iii) **Mutually exclusive events** If two events are mutually exclusive, it means that they cannot occur at the same time. For example, the two possible outcomes of a coin flip are mutually exclusive; when you flip a coin, it cannot land both heads and tails simultaneously.

	(iv) **Equally likely events** The events which have the same chance of occurring Probability. For example getting a 3 on the toss of a die and getting a 5 on the toss of a die are equally likely events, since the probabilities of each event are equal.
6(c)	On the average, one in 400 items is defective. If the items are packed in boxes of 100, what is probability that any given box will contain (i) No defectives; (ii) Less than two defectives; (iii) One or more defectives.
	Solution: Here $p = \frac{1}{400} = 0.0025,\ \ n = 100\ \ and\ \lambda = np = 100 * 0.0025 = 0.25$ Thus we need to use Poisson distribution (i) $P(X = 0) = \frac{e^{-0.25}0.25^0}{0!} = 0.779$ (ii) $P(X < 2) = P(X = 0) + P(X = 1) = \frac{e^{-0.25}0.25^0}{0!} + \frac{e^{-0.25}0.25^1}{1!}$ $= 0.779 + 0.195 = 0.974$ (iii) $P(X \geq 1) = 1 - P(X = 0) = 1 - \frac{e^{-0.25}0.25^0}{0!} = 1 - 0.779 = 0.221$
7(a)	Define normal distribution with their important properties. What types of error 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 in testing hypothesis: **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 of 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 of significance.
	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.