# Basic Statistics (Part – 2)

**Instructions:**

Please share your answers filled in-line in the word document. Submit code separately wherever applicable.

Please ensure you update all the details:

**Name: \_\_\_Jaidev Chhabria\_\_\_\_\_\_\_\_\_\_ Batch ID:** \_\_DSWEMON 290521\_\_\_\_\_\_\_\_\_

**Topic: Basic Statistics (Part 2)**

**Hints:**

1. For basic statistics, explanation of the solutions should be documented in black and white along with the codes.

Follow these guidelines:

* 1. Be thorough with the concepts of probability, Central Limit Theorem, and perform the calculations stepwise.
  2. For True/False problems, include an explanation.
  3. Codes for univariate analysis (histogram, box plot, bar plots etc.) and the distributions need to be attached wherever applicable.

1. All the codes should run without errors.

**Grading Guidelines:**

**Grading Guidelines:**

**1. An assignment submission is considered complete only when correct and executable code(s) are submitted along with the documentation explaining the method and results. Failing to submit either of those will be considered an invalid submission and will not be considered for evaluation.**

**2. Assignments submitted after the deadline will affect your grades.**

**Grading:**

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| **Ans** | **Date** |  |  | **Ans** | **Date** |
| Correct | On time | A | 100 |  |  |
| 80% & above | On time | B | 85 | Correct | Late |
| 50% & above | On time | C | 75 | 80% & above | Late |
| 50% & below | On time | D | 65 | 50% & above | Late |
|  |  | E | 55 | 50% & below |  |
| Copied/No Submission |  | F | 45 |  |  |

* **Grade A: (>= 90):** When all assignments are submitted on or before the given deadline.
* **Grade B: (>= 80 and < 90):** 
  + When assignments are submitted on time but less than 80% of problems are completed.

(OR)

* + All assignments are submitted after the deadline.
* **Grade C: (>= 70 and < 80):** 
  + When assignments are submitted on time but less than 50% of the problems are completed.

(OR)

* + Less than 80% of problems in the assignments are submitted after the deadline.
* **Grade D: (>= 60 and < 70):**
  + Assignments submitted after the deadline and with 50% or less problems.
* **Grade E: (>= 50 and < 60):** 
  + Less than 30% of problems in the assignments are submitted after the deadline.

(OR)

* + Less than 30% of problems in the assignments are submitted before the deadline.
* **Grade F: (< 50):** No submission (or) malpractice.

# Problem Statements:

1. For each of the following statements, indicate whether it is True/False. If false, explain why.
   1. The sample size of the survey should at least be a fixed percentage of the population size in order to produce representative results. –

Ans - False. It should not be fixed size but it should represent population.

* 1. The sampling frame is a list of every item that appears in a survey sample, including those that did not respond to questions.

Ans – False. The sampling frame work is a list of all the items in the target population from which the sample is selected

* 1. Larger surveys convey a more accurate impression of the population than smaller surveys.

I am not sure whether the question is referring to the size of survey (questions)

If it is related to samples to be larger, then answer is True as more responses will be much closer to the population mood compared to smaller sample survey.

1. *PC Magazine* asked all of its readers to participate in a survey of their satisfaction with different brands of electronics. In the 2004 survey, which was included in an issue of the magazine that year, more than 9000 readers rated the products on a scale from 1 to 10. The magazine reported that the average rating assigned by 225 readers to a Kodak compact digital camera was 7.5. For this product, identify the following:
2. The population 9000
3. The parameter of interest – Kodak compact digital camera
4. The sampling frame – PC Magazine readers in 2004
5. The sample size - 225
6. The sampling design – Responses on scale of 1 to 10
7. Any potential sources of bias or other problems with the survey or sample – It may include only those participants who may have used the product may reply.

Q3) Suppose we want to estimate the average weight of an adult male in Mexico. We draw a random sample of 2,000 men from a population of 3,000,000 men and weigh them. We find that the average person in our sample weighs 200 pounds and the standard deviation of the sample is 30 pounds. Calculate 94%, 98%, 96% confidence interval?

1. What are the chances that *X* > *μ* ?
2. ¼
3. ½
4. ¾
5. 1

Ans – B. There is a 50% change that X might be > average

1. A book publisher monitors the size of shipments of its textbooks to university bookstores. For a sample of texts used at various schools, the 95% confidence interval for the size of the shipment was 250 ± 45 books. Which, if any, of the following interpretations of this interval are correct?
2. All shipments are between 205 and 295 books. False (sampling range was from 205 to 295 not the whole population)
3. 95% of shipments are between 205 and 295 books. False (sampling range was from 205 to 295 not the whole population)
4. The procedure that produced this interval generates ranges that hold the population mean for 95% of samples. True
5. If we get another sample, then we can be 95% sure that the mean of this second sample is between 205 and 295. May be incorrect as 205 to 295 range was for one sample, it may be different for another sample. So hard to say it will fit in the given range.
6. We can be 95% confident that the range 160 to 340 holds the population mean. No
7. Which is shorter: a 95% *z*-interval or a 95% *t*-interval for *μ* if we know that σ =s?
8. The z-interval is shorter
9. The t-interval is shorter
10. Both are equal
11. We cannot say
12. How many randomly selected employers (minimum number) must we contact in order to guarantee a margin of error of no more than 4% (at 95% confidence)?

A. 600

B. 400

C. 550

D. 1000

Questions 8 and 9 are based on the following: To prepare a report on the economy, analysts need to estimate the percentage of businesses that plan to hire additional employees in the next 60 days.

1. Suppose we want the above margin of error to be based on a 98% confidence level. What sample size (minimum) must we now use?

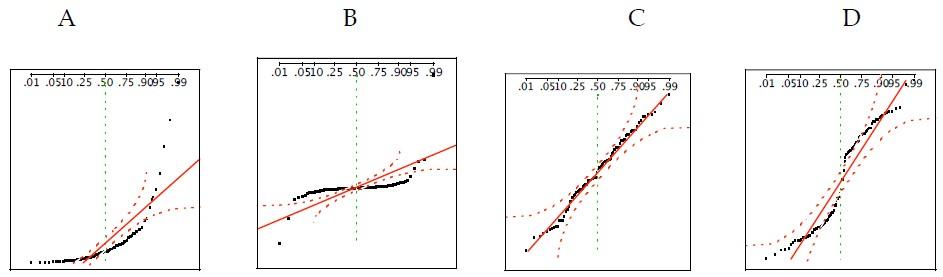
A. 1000

B. 757

C. 848

D. 543

1. Examine the following normal Quantile plots carefully. Which of these plots indicates that the data?
2. Are nearly normal? C
3. Have a bimodal distribution? (One way to recognize a bimodal shape is a “gap” in the spacing of adjacent data values.) D
4. Are skewed (i.e. not symmetric)? A
5. Have outliers on both sides of the center? B



1. For each of the following statements, indicate whether it is True/False. If false, explain why.

The manager of a warehouse monitors the volume of shipments made by the delivery team. The automated tracking system tracks every package as it moves through the facility. A sample of 25 packages is selected and weighed every day. Based on current contracts with customers, the weights should have μ = 22 lbs. and σ = 5 lbs.

* Before using a normal model for the sampling distribution of the average package weights, the manager must confirm that weights of individual packages are normally distributed. – False that will be done post analysis
* The standard error of the daily average SE(𝑥̅) = 1 = True

1. An educational startup that helps MBA aspirants write their essays is targeting individuals who have taken GMAT in 2012 and have expressed interest in applying to FT top 20 b-schools. There are 40000 such individuals with an average GMAT score of 720 and a standard deviation of 120. The scores are distributed between 650 and 790 with a very long and thin tail towards the higher end resulting in substantial skewness. Which of the following is likely to be true for randomly chosen samples of aspirants?
2. The standard deviation of the scores within any sample will be 120. – False, it may differ by sample to sample
3. The standard deviation of the mean of across several samples will be 120. - True
4. The mean score in any sample will be 720. – False, the mean score of all samples should be similar to population mean
5. The average of the mean across several samples will be 720. - True
6. The standard deviation of the mean across several samples will be 0.60 – False, it will be 120