Indian School of Business

Certificate Programme in Business Analytics

Statistical Analysis (1): Estimation and Testing

Academic year 2018-19, Term 1

Sample Final Exam (40 points)

Instructions:

- This exam is a closed book.
- You can use laptop only for R.
- You have two hours for the exam.
- These are multiple choice questions with only one correct answer
- The paper is set for 40 marks.
- You are allowed to use last blank pages for your rough work.
- You must observe the Honor Code at all times.

All the best!

Student Name:______________________________

ID:______________________

Section:____________________

1. When constructing a confidence interval, which case is preferable?

   A. Large confidence level, small confidence interval
   B. Small confidence level, small confidence interval
   C. Small confidence level, large confidence interval
   D. Large confidence level, large confidence interval

2. The 68.3% confidence interval for a population mean (with known population standard deviation and sufficiently large sample size) has a margin of error of 20. If we quadruple the sample size, but have the same margin of error for the new confidence interval, the approximate confidence level associated with the new confidence interval would be

   A. 99%
   B. 90%
   C. 95%
   D. Cannot be determined without more information

Answer: (C) First, note that we are assuming that σ is known so we will deal with z-values here. A 68.3% confidence interval implies that $z_{\alpha/2} = 1$. Thus, the margin of error is 1*σ√n. Since quadrupling the sample size cuts the standard error in half, the Z multiplier should have gone up by the same amount for the margin of error to remain the same. So now $z_{\alpha/2} = 2$, which corresponds roughly to 95% confidence interval.

3. *The Economic Times* reported a survey that indicated 68% of readers agree with the statement "Auditors should obtain regulatory clearance before auditing listed companies." If the survey result is based on a random sample of 170 readers of *The Economic Times*, then we should conclude that

   A. 68% of all readers agree with the statement.
   B. We are 99% confident that more than half of all readers agree with the statement.
   C. Less than half of all readers agree with the statement.
   D. More than 60% of all readers agree with the statement.

Answer: (B) Here, p = 0.68 and n = 170. Thus, the 99% confidence interval for the population proportion is given by 0.68±2.58*√(0.68*0.32/170) = [0.59,0.77]. So we are 99% confident that this proportion is greater than 50%.

4. A retail store buys imported apparel from a cheap offshore supplier. The clothing is not of top quality, but the customers like the low price. The retailer would like to know the defective rate of apparels to within a margin of error of 0.05 (95.4% confidence). Regardless of the true defect rate, the smallest sample size that guarantees this margin for error is

   A. 25 items
   B. 100 items
   C. 200 items
   D. 400 items

Answer: (D) Using the formula, we have $n \geq (z_{\alpha/2}/DMOE)^2*(1/4)$. For 95.4% confidence, $z_{\alpha/2}=2$. DMOE = 0.05. Thus, the resulting answer is 400.

5. Let $\bar{X}_1$ and $\bar{X}_2$ be the means of two separate samples of size 25 each that are randomly, and independently selected from the same population. Assume that CLT conditions apply. Consider

the following expressions: a $= P(\mu - 0.2\sigma < \bar{X}_1 < \mu + \sigma)$, and b $= P(\mu - \sigma < \bar{X}_2 < \mu + 0.2\sigma)$. Which of the following statements is true?

A. a > b
B. a < b
C. a = b
D. Need more information

Answer: (C).The distribution of the sample means will be normal with mean μ and SE 0.2σ. As normal distribution is symmetric around mean $P(\mu - 0.2\,\sigma < \overline{X_1} < 0) = P(0 < \overline{X_2} < \mu + 0.2\sigma)$ and $P(0 < \overline{X_1} < \mu + \sigma) = P(\mu - \sigma < \overline{X_2} < 0)$. Therefore, a=b.

Use the following information to answer the next 3 questions (Q6, Q7, and Q8).

A tire manufacturer believes that the tread-life of its snow tires can be described by a normal model with a mean of 32,000 miles and standard deviation of 2500 miles.

6. Approximately what fraction of these tires can be expected to last between 30,000 and 35,000 miles?

A. 0.55
B. 0.67
C. 0.73
D. 0.82

Answer : (B) P ((30,000 – 32,000)/2500 < Z < (35,000 – 32,000)/2500)

= P (-0.8 < Z < 1.2) = P (Z< 1.2) – P (Z< -0.8) =pnorm(1.2) – pnorm(-0.8) = 0.67

7. Which of the following is the closest estimate of the IQR (Inter-Quartile range) of the tread-life?

A. 2100 miles
B. 2700 miles
C. 3300 miles
D. 3900 miles

Answer: (C) Interquartile range is between the 25th and 75th percentile. As the thread-life is normal, the z values are obtained using qnorm(0.25) and qnorm(0.75), which are -0.67 and 0.67 respectively. Interquartile range is 2 * z * σ = 2 * 0.67 * 2500 = 3350. The closest answer would be 3300 miles.

8. In planning a marketing strategy, a local tire dealer wants to offer a refund to any customer whose tires fail to last a certain number of miles. However, the dealer does not want to take too big a risk. If the dealer is willing to give refunds to no more than 1 of every 25 customers, for what approximate mileage can he guarantee these tires to last?

A. 23000 miles
B. 27000 miles
C. 31000 miles
D. 35000 miles

Answer: (B) P (Z < z) < 1/25=0.04. z = qnorm(0.04) = -1.75. Now to get back to x, use $x = \mu + z * \sigma$. So, the approximate mileage that he should grant should be less than 32,000 – 1.75*2500, i.e., less than 27,625 miles.

9. What is the probability that the average value of 35 random rolls of a fair dice is greater than 4?

A. 0.06
B. **0.04**
C. 0.6
D. 0.4

Answer: (B) Let X be the random variable that denotes the outcome of the dice roll. Then the distribution of X is:

X=x	1	2	3	4	5	6
P(X=x)	1/6	1/6	1/6	1/6	1/6	1/6

mean = 1*(1/6) +2*(1/6) +3*(1/6) +4*(1/6) +5*(1/6) +6*(1/6) = 3.5

variance = $\sum (x_i - \mu)^2 P(X = x_i)$

$= \frac{(1-3.5)^2}{6} + \frac{(2-3.5)^2}{6} + \frac{(3-3.5)^2}{6} + \frac{(4-3.5)^2}{6} + \frac{(5-3.5)^2}{6} + \frac{(6-3.5)^2}{6} \approx 2.92.$

Std. Dev.= $\sqrt{2.92} = 1.71$

By CLT, the average value of 35 dice rolls is $N\left(3.5, \frac{1.71}{\sqrt{35}}\right)$. Thus, $\bar{X} \sim N(3.5, 0.29)$. So, $P(\bar{X} > 4) = P\left(Z > \frac{4-3.5}{0.29}\right) = P(Z > 1.72) = 1 - P(Z < 1.72) = 1 - pnorm(1.72) = 0.04$.

10. As the value of sample proportion increases from 0.1 to 0.9, for a given level of confidence, the Margin of Error of predicting the population proportion

A. Always increases
B. Always decreases
C. First increases, then decreases
D. First decreases, then increases

Answer (C) Note that the Margin of error of the sample proportion is approximated by $z_{\alpha/2}$p*(1-p)/√n. This increases as p increases from 0.1 to 0.5 and then decreases from 0.5 to 0.9. Consequently, the margin of error also follows the same behavior.

11. The average stock price for companies making up the S&P-500 is $30 and the standard deviation is $8.20.Assume that stock prices are normally distributed. Approximately how high does a stock price have to be to put a company in the top 10%?

A. $ 40.58
B. $ 10.25
C. $ 55
D. $ 30

Answer: (A) Let X be the stock price of a randomly chosen company. It is given that $X \sim N(\$30, \$8.20^2)$. We need the value of the stock price 'x' such that P(X > x) = 0.1. The z-score corresponding to 'x' will be pnorm(1-0.1) = pnorm(0.9) = 1.29. Thus, the required stock price is 30 +1.29*(8.2)=$ 40.578

12. The value of ATM transactions of a bank is normally distributed with mean Rs. 500 and standard deviation Rs. 80. Auditors of the bank decide to scrutinize a sample of 100 random transactions

0from the past year across the bank's ATM network. What is the probability that the total transaction amount for the random sample is in the interval [49000, 51000]?

A. 0.1128
B. 0.2211
C. 0.3224
D. **0.7888**

Answer: (D) The total transaction amount of the random sample being in the interval [49000, 51000] is the same as the sample mean of the random sample being in the interval [490, 510].

Each individual X_i has mean 500 and standard deviation 80. Then, the average transaction amount of 100 such individuals, $(\sum_{i=1}^{100} X_i)/100$ will have $N\left(500, (\frac{80}{\sqrt{100}})^2\right) = N(500, 8^2)$.
Then, $P(490 < \bar{X} < 510) = P\left(\frac{490-500}{8} < Z < \frac{510-500}{8}\right) = P(-1.25 < Z < 1.25)$ = 1-2pnorm(-1.25) = 0.7888.

13. Assume that male weights are normally distributed with a national average of 165.7 lbs. The probability of an individual chosen at random having a weight of greater than 190 lbs. is 25%. The probability that the mean weight of a sample of 40 men being greater than 190 lbs. is 2%. This rather large difference is due to the fact that

A. Standard deviation of samples is larger than the standard deviation of the population
B. The distribution of sample means is less variable than the distribution of individual data
C. Sampling error may have occurred during the research
D. The mean of the sample means varies from the population mean

Answer: (B) Since the distribution of the sample means has lower standard deviation (called standard error), the tails associated with that normal distribution are thinner and as a consequence, the probability is smaller.

14. Harsha's chocolate company sells two brands of chocolates - "Darko" and "Blanco". Annual sales of Darko follow a normal distribution with mean 107 and standard deviation 30 whereas annual sales of Blanco follow a normal distribution with mean 119 and standard deviation 40. (All numbers are in million INR). What is the probability that the sales of Darko will be greater than that of Blanco?

A. **0.4052**
B. 0.92
C. Sales of Darko will always be greater than that of Blanco
D. 0.67

Answer: (A) Let D denote the annual sales of Darko and B denote the annual sales of Blanco. Then, D~N(107, 30^2) and B~N(119, 40^2). We are interested in P(D>B). Define a new R.V, X= D-B. Then X~N(-12, 50^2), and P(D-B>0) = P(X>0) = P(Z> 12/50) = P(Z > 0.24) = 1-pnorm(0.24) = 1-0.5948 = 0.4052.

15. An online marketplace routinely selects a simple random sample of its visitors to estimate the average amount of time spent by them on the website. To reduce the margin of error of the estimate to half its current value, without changing the confidence level of the estimate, they should:

A. **Increase the sample size to four times its current value**
B. Reduce the sample size to half its current value
C. Increase the sample size to double its current value
D. Reduce the sample size to a fourth of its current value

Answer: (A) the margin of error is $z_{\alpha/2}*\sigma\sqrt{n}$. Quadrupling the sample size cuts the standard error in half, if the confidence level remains the same.

16. A simple random sample of 1000 persons is taken to estimate the percentage of democrats in a large population. It turns out that 543 of the people in the sample are Democrats. Based on this information, which of the following statements can you make?
    A. The 95% confidence interval for the sample percentage is [51.3%, 57.4%]
    B. There is a 95% chance that for the percentage of Democrats in the population is in the range [51.3%, 57.4%]
    C. At least 50% of the population are Democrats
    D. **The 95% confidence interval for the population percentage is [51.3%, 57.4%]**

Answer: (D) The 95% confidence interval for the percentage of Democrats in the population is $54.3\% \pm 1.96\sqrt{0.543(1-0.543)/1000} = [51.3\%, 57.4\%]$.

17. Terrahard Enterprises would like to estimate its share of the cloud storage market, i.e., proportion of internet users who use their services to store files on a cloud, with 98% confidence. If the margin of error tolerable to them is 1%, they should choose a sample size of approximately
    A. 10000
    B. **13600**
    C. 2500
    D. 1000

Answer: (B) DMOE = 1%, $\alpha/2 = 0.005$. Hence, $z_{\alpha/2} = 2.33$. Using the formula for the sample size calculations for proportions, we get $n \geq (1/4)*(2.33/0.01)^2 \cong 13572$. The only sample size from the alternatives that will give a margin of error less than DMOE is 13600.

18. A telephone company wants to estimate the mean number of minutes people in a city spend talking long distance with 95% confidence. From past records, it is known that the standard deviation of talk times is 12 minutes. What is the smallest sample size from the following alternatives that will give the length of the confidence interval to be 10 minutes or less?

    A. 28
    B. 11
    C. 24
    D. 19

Answer: (C) We have $\sigma = 12$ minutes. The confidence interval is $\left[\bar{x} - \frac{z*\sigma}{\sqrt{n}}, \bar{x} + \frac{z*\sigma}{\sqrt{n}}\right]$. Length of the CI = 2(z σ/ □n) .We want 'n' such that this length< 10. Using the z-value obtained from qnorm(0.975) we have, 2*(1.96*12/ □n) < 10, which gives n > 22.12

19. One public opinion poll uses a simple random sample of size 1500 drawn from a town with population of 25,000. Another poll uses a simple random sample of size 1500 from a town with a population of 250,000. The polls are trying to estimate the percentage of voters who favor single-payer health insurance. Other things being equal, which of the following statements is most accurate:
    A. The first poll is likely to be quite a bit more accurate than the second

B. The second poll is likely to be quite a bit more accurate than the first
C. The sample size of the second poll must be increased to 15,000 to get estimates as accurate as from the first poll
D. There is not likely to be much difference in accuracy between the two polls

Answer: (D) Variation in sample proportions depends only on the sample size and not on the population size. Hence, both polls are not likely to very different.

20. Suppose that the daily profit at a retail store is normally distributed with mean INR 0.2 million. Moreover, it has been found that profit is more than INR 0.5 million on 20% of the days. What is the APPROXIMATE fraction of days on which the store makes a loss?
    A. 0.1
    B. 0.2
    C. 0.3
    D. Need more information

Answer: (C) Let X be the random variable denoting daily profit. Then, we know that $P(X>0.5)=0.2$. We need to calculate the value of $P(X<0)$. Since 0.5 is at a distance of 0.3 from the mean of 0.2, whereas 0 is only at a distance of 0.2 from the mean, we know that the $P(X<0) > P(\text{Profits} > 0.5)$. The only alternative that matches this is (c).

Alternate Verification: $P(X>0.5)=0.2$. z-score corresponding to this is pnorm(0.8) = 0.84. Then $\sigma=(x-\mu)/z = (0.5-0.2)/0.84 = 0.36$. Hence, $P(X < 0) = P(Z < (0-0.2)/0.36) = P(Z < -0.56) = 0.29$, which is roughly equal to 0.3.