Analytics Primer

Quiz 1

Honor Pledge:

I certify that I have not received or given unauthorized aid in taking this exam.

Signed: _____

Printed Name: _____ KEY _____

Date: _____

There are 16 questions. The first 10 are worth 2 points each. The last 6 are worth 5 points each. There will be some True/False, Multiple Choice, and Short Answer questions.

An analyst conducted a study at a large university on the West coast. They are interested in calculating the average high school GPA of all applicants to the university in 2009 to determine if there is a relationship between high school GPA and the university's acceptance rate. The university provides the analyst with the high school GPA's of 250 randomly selected students in their database of students who were accepted to the university in 2009. The sample's average high school GPA was 2.95. Use this information to answer questions 1 through 3.

1. What is the parameter of interest in this study?
   a. 2.97.
   b. The average high school GPA of all applicants to the university in 2009.
   c. All applicants to the university in 2009.
   d. 250 students.

2. What type of data was collected by the analyst?
   a. Quantitative and Cross-Sectional.
   b. Quantitative and Time-Series.
   c. Qualitative and Cross-Sectional.
   d. Qualitative and Time-Series.

3. Are there any issues of bias in this data? If so, please explain.

   Undercoverage : the sample wasn't drawn
   from the population of interest
   but from a smaller subset.

4. What is the main difference between stratified random sampling and cluster sampling?

   Sample from every strata.
   Sample from some clusters.

5. There is no randomness in cluster sampling.
   a. True.
   b. False.

6. An outlier significantly affects the mean but not the median.
   a. True.
   b. False.

7. An analyst calculates the average of a right skewed distribution as 7.64.
Which of the following do we know to be true?
   a. The standard deviation is less than 7.64.
   b. The median is less than 7.64.
   c. The median is greater than 7.64.
   d. Both a and c, but not b.

8. List two characteristics of the Normal distribution.

   1. Bell-shaped   Unimodal   Defined by μ and σ

   2. Asymptotic to x-axis   Perfectly symmetric

9. You have a Normal distribution with a mean of 10 and a standard deviation
of 3. You calculate the z-score from a point on this distribution as $z = -1.67$.
The z-score of -1.67 means which of the following?
   a. The point was 1.67 standard deviations away from the mean.
   b. The point was greater than 10.
   c. The point was less than 10.
   d. Both a and c, but not b.

10. The Central Limit Theorem states that
   a. The distribution of the average of a large sample is approximately
Normal.
   b. The distribution of a large population is approximately Normal.
   c. The distribution of a large sample is approximately Normal.
   d. The distribution of the mean of a sample from a large population is
approximately Normal.

An employee for a credit card company is collecting data on common credit cards
owned by the American public. He studied both MasterCard and the company's new
card, ChiefCard. He found that 52% of people carry a MasterCard, 16% of people
carry a ChiefCard, and only 3% of people carry both. Use this information to answer
questions 11 and 12.

11. (5 points) What is the probability that any one random person carries a
MasterCard **or** a ChiefCard?

$$P(MC \text{ or } CC) = P(MC) + P(CC) - P(MC \text{ and } CC)$$

$$= 0.52 + 0.16 - 0.03$$

$$= 0.65$$

12. (5 points) The employee suspects that people who already have a MasterCard would not be very interested in the ChiefCard. What is the probability that someone carries a ChiefCard, given that they already carry a MasterCard?

$$P(CC \mid MC) = \frac{P(CC \text{ and } MC)}{P(MC)} = \frac{0.03}{0.52} = 0.058$$

A large company is evaluating their sales staff. The number of monthly successful sales calls (ones that a customer buys the product) per employee follows a Normal distribution with a mean of 18.54 and a standard deviation of 5.33. Use this information to answer questions 13 through 16.

13. (5 points) The company is going to put employees that make less that 10 successful sales calls a month through a training course. What is the probability that any **one** employee is selected to take this training course?

$$Z = \frac{X - \mu}{\sigma} = \frac{10 - 18.54}{5.33} = -1.6 \implies 0.0548$$

14. (5 points) The company is going to give a raise to the employees who make the top 25% of successful sales calls. How many successful sales does an employee need to get a raise?

$$Z = 0.67 = \frac{X - 18.54}{5.33} \implies X = 22.11$$

15. (5 points) What is the interquartile range of this data?

$$Z = 0.67 \implies X = 22.11$$

$$Z = -0.67 = \frac{X - 18.54}{5.33} \implies X = 14.97$$

$$IQR = 22.11 - 14.97 = 7.14$$

16. (5 points) The company is scared that too many people might be getting raises under their current plan, so they are adjusting it to give only raises to sales divisions. The sales staff is broken up into divisions of 64 people. They are only going to give raises to an entire division of sales staff if the division averages more than 20 successful sales calls per month. What is the probability that one sales staff division (64 people) average more than 20 successful sales calls per month.

$$Z_{\bar{x}} = \frac{\bar{X} - \mu}{\sigma/\sqrt{n}} = \frac{20 - 18.54}{5.33/\sqrt{64}}$$

$$= \frac{1.46}{0.666235} = 2.19 \implies 0.0143$$