Business Data Analysis, H25, Practice Test 2

Name: _____ Student number_____

1. Financial fraud is a significant threat to online retail platforms. In a specific e-commerce sector, fraudulent transactions are primarily committed by organized external hacker groups or by internal employees abusing their privileges. Very rarely are they committed by accidental system errors. Despite strict security protocols, fraud attempts are frequent. In recent years, 62% of the fraud attempts were initiated by external hackers, and the remaining 38% were committed by internal employees. An external hacker has a 79% chance of escaping detection after committing a fraud. An internal employee has a 46% chance of escaping detection after committing a fraud. A specific fraud case has been detected and the perpetrator identified. What is the probability that this perpetrator is an internal employee?

2. Cold calling is a traditional marketing strategy used by sales teams to generate leads. A specific software company targets businesses in the financial sector. The sales pitch is highly specialized. Experience shows that a skilled salesperson closes a sale (secures a meeting) only if the prospect is immediately interested. The calls are challenging; still, 6 sales are closed weekly (on average) per team in this region. If the probability of a prospect rejecting the pitch is 75% (meaning a 25% success rate), compute the probabilities that a salesperson will close anywhere from 0 to 6 sales out of 6 calls made in a given day.

3. The battery life of a new model of portable point-of-sale (POS) terminal is a critical selling point for small business owners. On average, the battery lasts for 22 hours of continuous use. The standard deviation of the battery life is 3.3 hours, and the population battery lives are assumed to be normally distributed.

    - Compute the probability that a randomly selected POS terminal will last more than 18 hours but less than 25 hours.
    - Compute the cutoff values for the bottom 20% (defective/poor performance) and the top 10% (superior performance) of battery duration.

4. Quality control is essential for packaged consumer goods. A coffee roastery sells bags of beans labeled as $500g$. A sample of 50 bags taken from the production line had a mean weight of $500g$ with a standard deviation of $55g$.

   - Construct a 95% and a 98% confidence interval for the population average weight of the coffee bags.
   - Write a sentence commenting on the tension between confidence level and the precision of the confidence interval estimate.

5. Return on Ad Spend (ROAS) is a key metric for evaluating marketing campaigns. A marketing agency tested a new social media strategy for six different clients, yielding the following ROAS values (in percentage):

   35.1 48.2 25.5 60.0 41.8 30.9

   Assuming that the ROAS values are normally distributed, compute a 99% confidence interval for the population mean ROAS for clients using this strategy.

6. Customer satisfaction surveys often use Likert scales to gauge consumer sentiment. A retail chain considers a customer "Highly Satisfied" if they rate their experience a 5 or 6 on a 6-point scale. A sample of 80 customer surveys showed that 25 customers gave a rating of 5 or 6 (considered 'Highly Satisfied').

   - Construct a 98% confidence interval for the population proportion of Highly Satisfied customers.
   - Based on this confidence interval, can you claim that more than 25% of all customers are Highly Satisfied? Explain briefly.

7. The Average Order Value (AOV) is a primary KPI for online retailers. Historical data suggested the average order value for a specific category of electronics was \$60. A recent sample of 40 orders taken after a website redesign showed a mean value of \$71 with a standard deviation of \$12. Test the hypothesis $H_0 : \mu = 60$ versus $H_1 : \mu > 60$ (indicating the redesign increased spending). Report a p-value and draw a conclusion in the context of the problem.

8. Call center efficiency is measured by the Average Handling Time (AHT). Management aims for an AHT of 35 minutes or less for complex support tickets. The handling time (in minutes) was measured for six randomly selected support calls:

   37 32 36 31 38 33

   Assuming the population handling time is normally distributed, test the hypothesis $H_0 : \mu = 35$ versus $H_1 : \mu < 35$ (testing if the team is performing faster than the baseline). Report a range for the p-value and draw a conclusion in the context of the problem.

9. A digital marketing firm tracks the user engagement for two distinct services: Email Newsletters (E) and Mobile App Notifications (M). Historical data for a specific client segment shows the following probabilities:

   - The probability a user engages with the Newsletter is $P(E) = 0.60$.
   - The probability a user engages with the Mobile App is $P(M) = 0.42$.
   - The probability a user engages with BOTH is $P(E \cap M) = 0.24$.

   - Find the probability that a user engages with the Newsletter OR the Mobile App.
   - Find the probability that a user engages with the Newsletter only.
   - Find the probability that a user engages with the Mobile App GIVEN that they have already engaged with the Newsletter.
   - Find the probability that a user engages with the Mobile App and the Newsletter GIVEN that they have engaged with one or the other.

10. An investment analyst tracks the behavior of a particular tech stock. Let D be the event that the company **Declares a dividend**, and let S be the event that the **Stock price increases**. Based on historical data for this sector:

    - The probability that the company declares a dividend is $P(D) = 0.30$.
    - The probability that the stock price increases is $P(S) = 0.60$.
    - The probability that the company declares a dividend OR the stock price increases is $P(D \cup S) = 0.72$.

    - Calculate the probability that the company declares a dividend AND the stock price increases, $P(D \cap S)$.
    - Are the events D and S **mutually exclusive**? Justify your answer using the probability calculated above.
    - Are the events D and S **independent**? Justify your answer by performing the appropriate probability test.