Introduction to Statistics

Probability and Inference Quiz – Answer Key

1. **(B) 0.12.** For independent events, $P(A \cap B) = P(A) \times P(B) = 0.3 \times 0.4 = 0.12$.

2. **(B) Sample means approach a normal distribution as sample size increases.** Regardless of the population distribution, the sampling distribution of the mean becomes approximately normal for large n (typically $n \geq 30$).

3. **(B) Rejecting a true null hypothesis.** Type I error (false positive) occurs when we incorrectly conclude there is an effect. Type II error is failing to reject a false null hypothesis.

4. **(B) Median.** The median is the middle value and is not affected by extreme values. The mean is pulled toward outliers, and range is directly determined by extremes.

5. **(C) 95% of similarly constructed intervals would contain the true parameter.** This is the frequentist interpretation—the procedure captures the parameter 95% of the time, not probability statements about any single interval.

6. **True.** The standard error (SE) is the standard deviation of a sampling distribution. For sample means: $SE = \sigma/\sqrt{n}$.

7. **False.** The p-value is the probability of observing data this extreme or more extreme, assuming the null hypothesis is true. It is not the probability that the null hypothesis is true.

8. **False.** Correlation measures linear association but does not establish causation. Causation requires controlled experiments, temporal precedence, and elimination of confounding variables.

9. **Descriptive Statistics:**

   - Summarizes and describes data characteristics
   - Measures: mean, median, mode, standard deviation, range, percentiles
   - Visualizations: histograms, box plots, scatter plots
   - Example: Calculating average test scores for a class, creating a frequency distribution of survey responses

   Inferential Statistics:

   - Makes generalizations about populations from samples
   - Techniques: hypothesis testing, confidence intervals, regression
   - Accounts for sampling variability and uncertainty
   - Example: Using a sample of 500 voters to estimate support for a candidate in the entire population

   Role of sampling:

   - Random sampling ensures representativeness

- Larger samples reduce sampling error
- Sampling distributions allow probability calculations
- Enables generalization from sample statistics to population parameters

10. Hypothesis Testing Steps:

(1) State hypotheses:

- Null hypothesis (H_0): No effect or no difference (status quo)
- Alternative hypothesis (H_a): Effect exists or difference present
- Example: H_0: $\mu = 100$ (mean IQ unchanged); H_a: $\mu \neq 100$ (mean IQ differs)

(2) Set significance level (α):

- Typically $\alpha = 0.05$ (5% chance of Type I error)
- Determines critical region for rejection

(3) Collect data and calculate test statistic:

- Example: Sample of n=36 students, $\bar{x} = 105$, s = 15
- t = $(\bar{x}$ - $\mu_0)/(s/\sqrt{n}) = (105 - 100)/(15/\sqrt{36}) = 5/2.5 = 2.0$

(4) Calculate p-value:

- Probability of observing test statistic this extreme under H_0
- For t = 2.0 with df = 35, two-tailed p ≈ 0.053

(5) Make decision:

- If p $\leq \alpha$: Reject H_0 (statistically significant)
- If p ¿ α: Fail to reject H_0 (not statistically significant)
- Example: p = 0.053 ¿ 0.05, so fail to reject H_0

(6) State conclusion in context:

- Example: "There is insufficient evidence at the 0.05 level to conclude that mean IQ differs from 100."