Probability Quiz

STAT II (Spring 2025)

Section 1: Multiple Choice

Circle the correct answer.

Question 1. The expectation $\mathbb{E}[X]$ of a random variable X is:

a. The value of X that occurs most frequently.
b. ✓ A weighted average of possible values of X, weighted by their probabilities.
c. The variance of X.
d. The square root of the variance of X.

Question 2. Which of the following is **NOT** true about conditional expectation $\mathbb{E}[X \mid Y]$?

a. It is a random variable itself.
b. $\mathbb{E}[X \mid Y]$ can sometimes equal $\mathbb{E}[X]$.
c. ✓ $\mathbb{E}[X \mid Y]$ is always greater than $\mathbb{E}[X]$.
d. It provides the expected value of X given the value of Y.

Question 3. The variance $V(X)$ of a random variable is:

a. ✓ $\mathbb{E}[X^2] - (\mathbb{E}[X])^2$.
b. $(\mathbb{E}[X])^2 - \mathbb{E}[X^2]$.
c. $\mathbb{E}[X]\mathbb{E}[X]$.
d. $\mathbb{E}[X^2] + \mathbb{E}[X]$.

Question 4. What does $X \perp\!\!\!\perp Y$ signify?

a. X and Y are uncorrelated.
b. ✓ X and Y are independent.
c. X and Y have the same mean.
d. X and Y have the same variance.

Question 5. If X and Y are independent, which of the following is incorrect?

a. $\mathbb{E}[XY] = \mathbb{E}[X]\mathbb{E}[Y]$.

b. $\mathbb{E}[f(X)Y] = \mathbb{E}[f(X)]\mathbb{E}[Y]$ for any function $f(X)$.

c. $\mathbb{E}[X \mid Y] = \mathbb{E}[X]$.

d. ✓ $\text{Cov}[X, Y] \neq 0$.

Section 2: Short Answers

Answer briefly. Provide the necessary steps and formulas for your proofs.

Question 6. Explain the difference between variance and covariance, and describe what a positive covariance indicates about the relationship between two random variables.

Answer: Variance measures the dispersion of a single random variable around its mean, while covariance assesses how two random variables change together. A positive covariance indicates that when one variable increases, the other tends to increase as well, suggesting a positive linear relationship.

Question 7. Define the Law of Iterated Expectation and provide a brief example of its application.

Answer: The Law of Iterated Expectation (also known as the Law of Total Expectation) states that the expectation of a random variable can be computed by iteratively taking the expectation conditional on another variable. Mathematically, $\mathbb{E}[X] = \mathbb{E}[\mathbb{E}[X \mid Y]]$. For example, if X is the earnings of an individual and Y is their education level, then $\mathbb{E}[X]$ can be calculated by first finding the expected earnings given each education level and then taking the average of these expectations weighted by the distribution of Y.

Question 8. Explain the Central Limit Theorem (CLT) and why it is fundamental for hypothesis testing.

Answer: The Central Limit Theorem (CLT) states that the distribution of the sample mean of a sufficiently large number of independent, identically distributed variables approaches a normal distribution, regardless of the original distribution. This is fundamental for hypothesis testing because it allows reseachers to make inferences about population parameters using the normal distribution, facilitating the construction of confidence intervals and hypothesis tests even when the underlying data distribution is unknown.

Question 9. What is a p-value, and how do you interpret it in the context of a hypothesis test?

Answer: A p-value is the probability of obtaining a test statistic as extreme as, or more extreme than, the observed value under the null hypothesis. It is used to determine the statistical significance of the test results. In the context of a hypothesis test, if the p-value is below a predetermined significance level (e.g., $\alpha = 0.05$), we reject the null hypothesis, suggesting that the observed data is inconsistent with the null hypothesis.

Question 10. Consider the simple mean estimator for a population mean. Define the estimator $\bar{X}$ as the average of a sample of observations $X_1, X_2, \ldots, X_n$, where each X_i is an independent and identically distributed (i.i.d.) random variable with the population mean μ. Show that $\bar{X}$ is an unbiased estimator of μ.

Answer: The sample mean $\bar{X}$ is calculated as $\bar{X} = \frac{1}{n}\sum_{i=1}^{n} X_i$. The expected value of $\bar{X}$ is

$$\mathbb{E}[\bar{X}] = \mathbb{E}\left[\frac{1}{n}\sum_{i=1}^{n} X_i\right] = \frac{1}{n}\sum_{i=1}^{n}\mathbb{E}[X_i] = \frac{1}{n}\sum_{i=1}^{n}\mu = \mu.$$

Therefore, $\bar{X}$ is an unbiased estimator of the population mean μ.

Question 11. A study analyzes the relationship between individuals' access to the internet and their participation in anti-government protest. Let Y represent protest turnout (1 for protest, 0 for stay home), and X represent internet access (1 for access, 0 for no access). The joint probability distribution is given below:

$Y \setminus X$	**Access** ($X = 1$)	**No Access** ($X = 0$)
Protest ($Y = 1$)	0.4	0.2
Stay home ($Y = 0$)	0.1	0.3

1. Find $\Pr(Y = 1 \mid X = 1)$ and $\Pr(Y = 1 \mid X = 0)$.

   Answer: To find $\Pr(Y = 1 \mid X = 1)$ we can use the formula for conditional probability:

$$\Pr(Y = 1 \mid X = 1) = \frac{\Pr(Y = 1 \cap X = 1)}{\Pr(X = 1)} = \frac{0.4}{0.1 + 0.4} = 0.8$$

   Similarly, for $\Pr(Y = 1 \mid X = 0)$:

$$\Pr(Y = 1 \mid X = 0) = \frac{\Pr(Y = 1 \cap X = 0)}{\Pr(X = 0)} = \frac{0.2}{0.5} = 0.4$$

2. Are protest participation (X) and internet access (Y) independent? Justify mathematically.

   Answer: Two events X and Y are independent if $\Pr(Y = 1|X = 1) = \Pr(Y = 1|X = 0) = \Pr(Y = 1)$.

   Calculate $\Pr(Y = 1)$ as follows:

$$\Pr(Y = 1) = \Pr(Y = 1 \cap X = 1) + \Pr(Y = 1 \cap X = 0) = 0.4 + 0.2 = 0.6$$

   Since $\Pr(Y = 1 \mid X = 1) = 0.8 \neq 0.6$ and $\Pr(Y = 1 \mid X = 0) = 0.4 \neq 0.6$, protest participation (Y) and internet access (X) are not independent.

Question 12. You are studying the relationship between voter turnout (Y), campaign intensity in the district (Z), and political engagement (X) across citizens eligible to vote. The following frequency table summarizes the data you collected:

Political Engagement (X)	**Campaign Intensity** (Z)	**Voter Turnout** (Y)	**Frequency**
Low	Low	Yes	10
Low	Low	No	20
Low	High	Yes	15
Low	High	No	5
High	Low	Yes	25
High	Low	No	10
High	High	Yes	20
High	High	No	10

1. Calculate the joint probabilities $\Pr(Y, Z \mid X)$ for $X = \text{Low}$.

   Answer: Joint probabilities are:

$$\Pr(Y = \text{Yes}, Z = \text{Low} \mid X = \text{Low}) = \frac{10}{50} = 0.2$$
$$\Pr(Y = \text{No}, Z = \text{Low} \mid X = \text{Low}) = \frac{20}{50} = 0.4$$

$$\Pr(Y = \text{Yes}, Z = \text{High} \mid X = \text{Low}) = \frac{15}{50} = 0.3$$

$$\Pr(Y = \text{No}, Z = \text{High} \mid X = \text{Low}) = \frac{5}{50} = 0.1$$

2. Is Y independent of Z given X (i.e. is $Y \perp\!\!\!\perp Z \mid X$)? You can check this using the definition of conditional independence. *(Hint. Remember that it is enough for the conditional independence to not hold for at least one level of X)*

   Answer: For $Y \perp\!\!\!\perp Z \mid X = \text{Low}$, we need:

$$\Pr(Y, Z \mid X = \text{Low}) = \Pr(Y \mid X = \text{Low})\Pr(Z \mid X = \text{Low})$$

   For $Y = \text{Yes}$ and $Z = \text{Low}$ we have:

$$\Pr(Y = \text{Yes} \mid X = \text{Low})\Pr(Z = \text{Low} \mid X = \text{Low}) = 0.5 \times 0.6 = 0.3$$

   This is not equal to

$$\Pr(Y = \text{Yes}, Z = \text{Low} \mid X = \text{Low}) = \frac{10}{50} = 0.2$$

   Therefore, $Y \perp\!\!\!\perp Z \mid X$ does not hold for $X = \text{Low}$ and Y is not independent of Z given X.