107.254 Statistics and Probability Theory (VO) 2025W

1. Written Exam – Monday 26 January 2026

Questions

This is a multiple-choice exam. It consists of 20 multiple-choice problems. All the problems carry 5 points each. The maximum score is 100.

For each problem, four possible answers (a, b, c, d) are offered, and *exactly one answer is correct*. The answer that best completes the statement or answers the question should be chosen by ticking in the *Answers form*. There are no negative points for ticking a wrong answer. Ticking no answer or ticking more than one answer leads to the question being marked as incorrect. A pen with either blue or black ink has to be used.

A non-programmable calculator and a two-sided handwritten A4 formulae sheet may be used during the exam. The formulae sheet has to be submitted with the exam. Please note that a copy of a handwritten sheet is not a handwritten sheet and cannot be used in the exam. Computers, smartphones, tablets, notes, books, etc., as well as discussions and consultations are prohibited during the exam. It is mandatory to bring your student ID or an official photo ID.

The examination time is 90 minutes.

Good luck!

(1) Can the function

$$p(x) = \begin{cases} ax^2 + x - \frac{1}{2}, & x = 1, 2, 3 \\ 0, & \text{else} \end{cases}$$

with $a \in \mathbb{R}$ be the probability mass function for a discrete random variable?

a. No, because probabilities can *not* be negative.
b. Yes, for a unique *negative* a.
c. No, because probabilities can *not* be greater than 1.
d. Yes, for a unique *positive* a.

(2) Cars pass independently by a point on a busy road at an average rate of 150 per hour. Assume the number of cars follows a Poisson distribution with probability mass function

$$P(X = k) = \frac{\lambda^k e^{-\lambda}}{k!}, \quad k = 0, 1, 2, \ldots$$

(i) What is the probability that no car passes in one minute?
(ii) What is the expected number of cars passing in two minutes?

a. (i) $e^{-2.5}$, (ii) 5
b. (i) $e^{-2.5}$, (ii) 2.5
c. (i) e^{-150}, (ii) 300
d. (i) e^{-150}, (ii) 150

(3) A factory produces metal rods with lengths that are normally distributed, with a mean of 50 cm and a standard deviation of 5 cm. A quality inspector takes a random sample of $n = 25$ rods. What is the probability that the sample mean length is between 49 cm and 51 cm?

a. $\approx 34\%$
b. $\approx 50\%$
c. $\approx 68\%$
d. $\approx 95\%$

(4) Let $X \sim \mathrm{U}(0, 1)$ and define $Y = X^2$. What is the *density* $f_Y(y)$ of Y for $y \in (0, 1)$?

a. $f_Y(y) = \frac{1}{2\sqrt{y}}$
b. $f_Y(y) = \sqrt{y}$
c. $f_Y(y) = y^{-1/2}$
d. $f_Y(y) = y^2$

(5) Eve is a candidate in an election. Her team asked $n = 100$ voters if they would vote for Eve or not. Out of the $n = 100$ voters, 67 answered 'yes'. The team wants to answer if *more then* 80% would vote for Eve.

a. They reject H_0 at level $\alpha = 1\%$ but not at level $\alpha = 5\%$.
b. They *fail* to reject H_0 at level $\alpha = 1\%$.
c. They reject H_0 at level $\alpha = 1\%$.
d. They *fail* to reject H_0 at level $\alpha = 5\%$ but not at $\alpha = 1\%$.

(6) Which of the following holds **true** for *any* two random variables X and Y with zero first moments and finite variance?

a. $\mathbb{E}[(X-Y)(X+Y)] = \mathrm{Var}(X) - \mathrm{Var}(Y)$
b. $\mathbb{E}[(X+Y)^2] > \mathbb{E}\,X^2 + \mathbb{E}\,Y^2$
c. $\mathbb{E}[(X+Y)^2] = \mathrm{Var}(X) + \mathrm{Var}(Y)$
d. $\mathbb{E}[(X-Y)(X+Y)] > \mathbb{E}\,X^2$

(7) Consider two events A, B with non-zero probability. Which of the following identities is **false**?

a. $P(A \mid B) = P(B \mid A)P(A)$
b. $P(A \cap B) = P(A \mid B)P(B)$
c. $P(A \cup B) = P(A) + P(B) - P(A \cap B)$
d. $P(A \cup B) = P(A) + P(B \setminus A)$

(8) Let $X_1, \ldots, X_n$ be a random sample from a standard normal distribution. What is the density of the sample mean

$$\bar{X} = \frac{1}{n}\sum_{i=1}^{n} X_i?$$

a. $f_{\bar{X}}(x) = \frac{1}{\sqrt{2\pi}} \exp\Big(-\frac{x^2}{2}\Big)$
b. $f_{\bar{X}}(x) = \frac{1}{\sqrt{2\pi}} \exp\Big(-\frac{(x-n)^2}{2}\Big)$
c. $f_{\bar{X}}(x) = \sqrt{\frac{n}{2\pi}} \exp\Big(-\frac{nx^2}{2}\Big)$
d. $f_{\bar{X}}(x) = \frac{1}{\sqrt{2\pi n}} \exp\Big(-\frac{x^2}{2n}\Big)$

(9) Consider a z-test with null hypothesis $H_0 : \mu = 0$ versus the alternative hypothesis $H_1 : \mu > 0$ given an iid random sample $X_1, \ldots, X_{100}$ such that $X_1 \sim \mathcal{N}(\mu, \sigma^2)$ where $\sigma = 10$.

Assuming that the true population mean is $\mu_1 = 1.64$, what is the Type II error probability β of the test at significance level $\alpha = 5\%$?

The following standard normal quantiles might be of help:

p	90%	92.5%	95%	97.5%
$\Phi^{-1}(p)$	1.28	1.44	1.64	1.96

a. $\beta \approx 13\%$
b. $\beta \approx 5\%$
c. $\beta \approx 68\%$
d. $\beta \approx 50\%$

(10) Let X and Y be two independent random variables. Suppose we know that $\mathrm{Var}(2X - Y) = 6$ and $\mathrm{Var}(X + 2Y) = 9$. What is $\mathrm{Var}(Y)$?

a. 2
b. 6
c. 3
d. 1

(11) Suppose $X \sim \text{Exp}(\lambda)$ for $\lambda > 0$. The cumulative distribution function of X is then given as

$$F_X(x) = \begin{cases} 1 - \exp(-\lambda x), & 0 \leq x \\ 0, & \text{else} \end{cases}.$$

What is the *median* of X?

a. $\frac{\ln(\lambda)}{2}$
b. $\frac{\ln(2)}{\lambda}$
c. $\frac{1}{\lambda}$
d. The median is not unique.

(12) The following contingency table summarizes the study method and exam outcome for a group of students:

	Pass	Fail	Total
Group Study	18	12	30
Individual Study	6	24	30
Total	24	36	60

A chi-squared test for independence is conducted at the 5% significance level.

A portion of the chi-squared critical value table is given below:

df	0.10	0.075	0.05	0.025	0.01
1	2.71	3.17	3.84	5.02	6.63
2	4.61	5.18	5.99	7.38	9.21
3	6.25	6.90	7.81	9.35	11.34

Which of the following conclusions is **most** appropriate?

a. Reject the null hypothesis and conclude that study method and exam outcome are *not* associated.
b. Fail to reject the null hypothesis and conclude that study method and exam outcome are *not* associated.
c. The test is invalid because the contingency table is 2×2.
d. Reject the null hypothesis and conclude that study method and exam outcome are associated.

(13) For $X \sim \mathcal{N}(1, 4)$, the probability $P(X^3 - 3X^2 + 3X - 1 \geq 0)$ is

a. $\approx 66.7\%$
b. $= 0$
c. $= \frac{1}{2}$
d. $= 1$

(14) We fit a simple linear regression model by regressing Y on X, and obtain the following output:

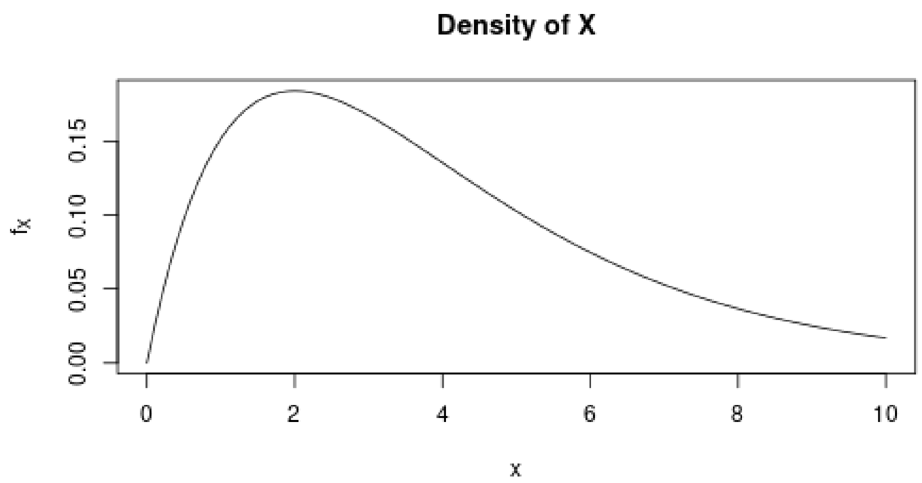
```
Coefficients:
                  Estimate   Std. Error   t value    Pr(>|t|)
(Intercept)         0.3200       0.7600     0.421    0.680
x                   0.7500       0.1000     7.500    2.88e-06 ***

Residual standard error:  1.5 on 14 degrees of freedom
```

What is the observed sample covariance between Y and X?

Hint: Use the formula for $\widehat{\text{Cov}}(X, Y)$ *and resolve the unknown using both formulae for the intercept estimator* $\hat{\beta}_1$ *and its standard error* $s_{\hat{\beta}_1}$.

a. $\frac{3}{4} = 0.75$
b. $\frac{675}{4} = 168.75$
c. $\frac{45}{4} = 11.25$
d. 15

(15) The following graph plots the density f_X of a random variable X;



The random variable is distributed according to one of the following distributions. Which one is it?

a. $X \sim \chi^2(4)$, that is χ^2 with 4 degrees of freedom.
b. $X \sim \text{Binomial}(10, 1/2)$, that is binomial for 10 trials with success probability 1/2.
c. $X \sim \mathcal{N}(0, 1)$, that is standard normal.
d. $X \sim \text{Exp}(1)$, that is Exponential with rate 1.

(16) Marketing Research wants to determine if an advertising campaign for a new energy drink, "BoostUp," increased customer recognition.

A random sample of 250 residents of a major city were asked if they knew about "BoostUp" before the advertising campaign. In this survey, 50 respondents had heard of the drink. After the advertising campaign, a second random sample of 200 residents were asked the same question using the same protocol. In this case, 80 respondents had heard of "BoostUp."

(i) Formulate the null and alternative hypotheses for testing whether customer recognition *increased*.
(ii) Compute the value of the test statistic for a one-sided z-test for two population proportions.

a. $H_0 : p_{\text{before}} \leq p_{\text{after}},\ H_1 : p_{\text{before}} > p_{\text{after}}, \quad z \approx -0.29$
b. $H_0 : p_{\text{before}} = p_{\text{after}},\ H_1 : p_{\text{before}} > p_{\text{after}}, \quad z \approx 0.29$
c. $H_0 : p_{\text{before}} = p_{\text{after}},\ H_1 : p_{\text{before}} < p_{\text{after}}, \quad z \approx -4.65$
d. $H_0 : p_{\text{before}} = p_{\text{after}},\ H_1 : p_{\text{before}} \neq p_{\text{after}}, \quad z \approx 0$

(17) Consider the following `R` code

```
n <- 6
p <- 1 / 2
mean(rbinom(10000, n, p) < 2)
```

What does it do?

a. It approximates the expected value $\mathbb{E}[X \mid X < 2]$ for X Binomial with success probability $p = \frac{1}{2}$ for $n = 6$ trials.
b. It approximates the probability that 6 tosses of a fair coin produce none or only 1 head.
c. The code has a bug, it throws an error.
d. It approximates the probability that a fair dice roles a 6.

(18) Which of the `R` code snippets computes the p-value of a t-test for testing $H_0 : \mu = 0$ vs. $H_1 : \mu \neq 0$ given observations

```
x <- c(7.7, -11.4, -3.7, 14.1, 0.2, 9.9, 2.6, 13.3, -7.0)
```

a.
```
t.obs <- mean(x) / (sd(x) / 3)
2 * (1 - pt(abs(t.obs), 8))
```
b.
```
t.obs <- mean(x) / sd(x)
2 * (1 - pt(t.obs, 9))
```
c.
```
t.obs <- mean(x) / sd(x)
1 - pt(abs(t.obs), 9)
```
d.
```
t.obs <- mean(x) / (sd(x) / 3)
1 - pt(abs(t.obs), 8)
```

(19) Given a random sample $X_1, \ldots, X_n$ of size n from a distribution with expectation 2 and variance 9, let

$$\bar{X} = \frac{1}{n} \sum_{i=1}^{n} X_i$$

be the sample mean. Which of the following R-functions approximates the value of $P(\bar{X} \leq 3)$ based on the Central Limit Theorem?

a. `dnorm(3, 2, 3 / sqrt(n))`

b. `pnorm(3 / sqrt(n))`

c. `pnorm(sqrt(n) / 3)`

d. `dnorm(n / 9)`

(20) Eggs from a farm have normally distributed weights with a standard deviation of 8 g. What is the approximate expected weight if 97.5% of eggs weigh more than 34 g?

a. ≈ 30

b. ≈ 58

c. ≈ 50

d. ≈ 16