107.254 Statistics and Probability Theory (VO) 2025W

1. Written Exam – Monday 26 January 2026

Solutions

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(3). No, because probabilities can *not* be negative.
b(1). Yes, for a unique *negative* a.
c(4). No, because probabilities can *not* be greater than 1.
d(2). Yes, for a unique *positive* a.

Solution(Q01): Evaluating $p(x)$ for all values of x gives

$$p(1) = a + \frac{1}{2}, \quad p(2) = 4a + \frac{3}{2}, \quad p(3) = 9a + \frac{5}{2}.$$

we know that probabilities sum to 1, which gives the following equation to solve for a

$$1 \overset{!}{=} p(1) + p(2) + p(3) = 14a + \frac{9}{2}$$

yielding $a = -1/4$. Substituting a into $p(x)$ gives

$$p(1) = \frac{1}{4}, \quad p(2) = \frac{1}{2}, \quad p(3) = \frac{1}{4}.$$

Those are all positive and sum to one, which means that $p(x)$ is a p.m.f. for the unique value $a = -\frac{1}{4}$.

(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(1). (i) $e^{-2.5}$, (ii) 5
b(4). (i) $e^{-2.5}$, (ii) 2.5
c(2). (i) e^{-150}, (ii) 300
d(3). (i) e^{-150}, (ii) 150

Solution(Q02): Convert the rate from cars per hour to cars per minute $\lambda_{\text{min}} = \frac{150\,\text{cars/h}}{60\,\text{min/h}} = 2.5\,\frac{\text{cars}}{\text{min}}$. Let X denote the random variable for the cars per minute, then

$$P(X = 0) = \frac{\lambda_{\text{min}}^0 e^{-\lambda_{\text{min}}}}{0!} = \mathbf{e^{-2.5}}.$$

With the expected number of cars per minute being $\mathbb{E}[X] = \lambda_{\text{min}} = 2.5$ we expect twice the number of cars in double the time (can also be computed explicitely) leading to $\mathbb{E}[X_{2\ \text{min}}] = 2\,\lambda_{\text{min}} = \mathbf{5}$.

(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(2). $\approx 34\%$
b(3). $\approx 50\%$

c(1). $\approx 68\%$
d(4). $\approx 95\%$

Solution(Q03): The variance is $s^2 = \mathrm{Var}(\overline{X}) = \frac{\sigma^2}{n} = \frac{5^2}{25} = 1$ which gives with $\mathbb{E}\,\overline{X} = \mu = 50$ that

$$P(49 \le \overline{X} \le 51) = P(\mu - 1\,s \le \overline{X} \le \mu + 1\,s) \approx 68\% \quad \text{(or roughly 2/3.)}$$

(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(1). $f_Y(y) = \frac{1}{2\sqrt{y}}$
b(2). $f_Y(y) = \sqrt{y}$
c(4). $f_Y(y) = y^{-1/2}$
d(3). $f_Y(y) = y^2$

Solution(Q04): The c.d.f. for $y \in (0,1)$ is $F_Y(y) = P(Y \le y) = P(X^2 \le y) = P(X \le \sqrt{y}) = \sqrt{y}$ where the last equality holds due to $X \sim \mathrm{U}(0,1)$. Differentiation of the c.d.f. gives the density

$$f_Y(y) = F_Y'(y) = \frac{\mathrm{d}}{\mathrm{d}y}\sqrt{y} = \frac{1}{2\sqrt{2}}.$$

(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(4). They reject H_0 at level $\alpha = 1\%$ but not at level $\alpha = 5\%$.
b(1). They *fail* to reject H_0 at level $\alpha = 1\%$.
c(3). They reject H_0 at level $\alpha = 1\%$.
d(2). They *fail* to reject H_0 at level $\alpha = 5\%$ but not at $\alpha = 1\%$.

Solution(Q05): The goal of answering if *more then* 80% would vote for Eve defines the alternative hypothesis as $H_1 : p > 80\%$ where p is the proportion of Eve voters. Since the observed proportion is $p_{\mathrm{obs}} = \frac{67}{100} = 67\% < 80\%$, we conclude that we fail to reject the null hypotheses $H_0 : p = 80\%$ for any (reasonable) significance level in a one-sided test.

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

a(1). $\mathbb{E}[(X - Y)(X + Y)] = \mathrm{Var}(X) - \mathrm{Var}(Y)$
b(4). $\mathbb{E}[(X + Y)^2] > \mathbb{E}\,X^2 + \mathbb{E}\,Y^2$
c(3). $\mathbb{E}[(X + Y)^2] = \mathrm{Var}(X) + \mathrm{Var}(Y)$
d(2). $\mathbb{E}[(X - Y)(X + Y)] > \mathbb{E}\,X^2$

Solution(Q06): With $\mathbb{E}\,X = \mathbb{E}\,Y = 0$ and the linearity of the expectation we get

$$\mathbb{E}[(X-Y)(X+Y)] = \mathbb{E}[X^2-Y^2] \overset{\text{lin}}{=} \mathbb{E}[X^2]-\mathbb{E}[Y^2] \overset{\text{given}}{=} \mathbb{E}(X-\mathbb{E}\,X)^2-\mathbb{E}(Y-\mathbb{E}\,Y)^2 \overset{\text{def}}{=} \mathrm{Var}(X)-\mathrm{Var}(Y).$$

Noting that X and Y can be dependent, consider the counter example $X \sim \mathcal{N}(0,1)$ and $Y = -X$. Then all other cases imply one of $0 > 1$ or $0 = 2$ or $0 \ge 2$, which are all *wrong* (the true case is $0 = 0$.)

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

a(1). $P(A \mid B) = P(B \mid A)P(A)$
b(4). $P(A \cap B) = P(A \mid B)P(B)$
c(2). $P(A \cup B) = P(A) + P(B) - P(A \cap B)$
d(3). $P(A \cup B) = P(A) + P(B \setminus A)$

Solution(Q07): The *wrong* identity can be identified through Bayes theorem (or through the conditional probability definition $P(B \mid A)P(A) = P(B \cap A) \overset{\text{gen}}{\neq} P(A \mid B)$ with "gen" indicating generality)

$$P(A \mid B) = \frac{P(B \mid A)P(A)}{P(B)} \neq P(B \mid A)P(A).$$

To "prove" the other statements *true* one could draw Vendiagrams. Alternatively, for the purpose of completenes of deriving that the remaining statement are true, we start with the definition of conditional probability $P(A \mid B) = \frac{P(A \cap B)}{P(B)}$ which gives

$$P(A \cap B) = P(A \mid B)P(B) \qquad \underline{\text{True}}.$$

Next, consider

$$P(A \cup B) = P(A \cup (B \backslash A)) = P(A) + P(B \backslash A) \qquad \underline{\text{True}}.$$

Now, using $P(B) = P((B \cap A) \cup (B \cap A^c)) = P(B \cap A) + P(B \cap A^c) = P(B \cap A) + P(B \backslash A)$ gives $P(B \backslash A) = P(B) - P(A \cap B)$. Substitution into the last identity gives

$$P(A \cup B) = P(A) + P(B \backslash A) = P(A) + P(B) - P(A \cap B) \qquad \underline{\text{True}}.$$

(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(2). $f_{\bar{X}}(x) = \frac{1}{\sqrt{2\pi}} \exp\left(-\frac{x^2}{2}\right)$
b(3). $f_{\bar{X}}(x) = \frac{1}{\sqrt{2\pi}} \exp\left(-\frac{(x-n)^2}{2}\right)$
c(1). $f_{\bar{X}}(x) = \sqrt{\frac{n}{2\pi}} \exp\left(-\frac{nx^2}{2}\right)$
d(4). $f_{\bar{X}}(x) = \frac{1}{\sqrt{2\pi n}} \exp\left(-\frac{x^2}{2n}\right)$

Solution(Q08): With $\overline{X} \sim \mathcal{N}(0, \frac{1}{n})$ we know that $\mu = 0$ and $\sigma^2 = \frac{1}{n}$, leading to

$$f_{\overline{X}}(x) = \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left(-\frac{(x-\mu)^2}{2\sigma^2}\right) = \frac{1}{\sqrt{2\pi(1/n)}} \exp\left(-\frac{x^2}{2(1/n)}\right) = \sqrt{\frac{n}{2\pi}} \exp\left(-\frac{nx^2}{2}\right)$$

Note: The same works for only knowing that the standard normal density is something like $\exp(-z^2)$ *with some constants added throughout the formula. Then, consider the standardization of* $\overline{X}$ *as* $Z = \frac{\overline{X}-\mu}{\sigma/\sqrt{n}} = \sqrt{n}\,\overline{X}$ *and substitute* $z = \sqrt{n}x$*, uniquely identifying the correct solution.*

(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(3). $\beta \approx 13\%$
b(4). $\beta \approx 5\%$
c(2). $\beta \approx 68\%$
d(1). $\beta \approx 50\%$

Solution(Q09): The z-statistic with $\sigma = 10$ and sample size $n = 100$ for the sample mean $\bar{X} = \frac{1}{n}\sum_{i=1}^{n} X_i$ under the null $H_0 : \mu = 0$ is

$$Z = \frac{\bar{X} - 0}{\sigma/\sqrt{n}} = \frac{\bar{X}}{10/\sqrt{100}} = \bar{X} \sim \mathcal{N}(0, 1).$$

The critical value z_α to reject the null if $Z > z_\alpha$ is

$$z_\alpha = 0 + q_{1-\alpha}\frac{\sigma}{\sqrt{n}} = q_{95\%}\frac{10}{\sqrt{100}} = q_{95\%} = \Phi^{-1}(95\%) \overset{\text{given}}{\approx} 1.64.$$

If the true population mean is $\mu = \mu_1 = 1.64$, then the sample mean is distributed as $\overline{X} \sim \mathcal{N}(\mu_1, 1)$ and a Type II error is to *not* reject the null hypotheses given that the alternative is true. Therefore,

$$\beta = P(\overline{X} \leq z_\alpha) \approx P(\overline{X} \leq 1.64) = P(\overline{X} \leq \mu_1) = 50\%.$$

(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(1). 2
b(4). 6
c(3). 3
d(2). 1

Solution(Q10): Using the independence of X and Y we get the equation system

$$6 = \mathrm{Var}(2X - Y) \overset{\text{ind}}{=} 4\,\mathrm{Var}(X) + \mathrm{Var}(Y),$$
$$9 = \mathrm{Var}(X + 2Y) \overset{\text{ind}}{=} \mathrm{Var}(X) + 4\,\mathrm{Var}(Y).$$

Subtracting 4 times the second from the first equation gives

$$-30 = 6 - 4 \cdot 9 = 4\,\mathrm{Var}(X) + \mathrm{Var}(Y) - 4(\mathrm{Var}(X) + 4\,\mathrm{Var}(Y)) = \mathrm{Var}(Y) - 16\,\mathrm{Var}(Y) = -15\,\mathrm{Var}(Y)$$

which means that $\mathrm{Var}(Y) = \frac{-30}{-15} = 2$.

(11) Suppose $X \sim \mathrm{Exp}(\lambda)$ 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(2). $\frac{\ln(\lambda)}{2}$

b(1). $\frac{\ln(2)}{\lambda}$

c(3). $\frac{1}{\lambda}$

d(4). The median is not unique.

Solution(Q11): A median m is any solution to the equation $F_X(m) = \frac{1}{2}$. As such we get that

$$\begin{aligned} 1 - \exp(-\lambda m) = F_X(m) &= \tfrac{1}{2} \\ \exp(-\lambda m) &= \tfrac{1}{2} \\ \lambda m &= \ln(2). \end{aligned}$$

The solution for the median is unique and given as $m = \frac{\ln(2)}{\lambda}$.

(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(3). Reject the null hypothesis and conclude that study method and exam outcome are *not* associated.

b(2). Fail to reject the null hypothesis and conclude that study method and exam outcome are *not* associated.

c(4). The test is invalid because the contingency table is 2×2.

d(1). Reject the null hypothesis and conclude that study method and exam outcome are associated.

Solution(Q12): The *expected counts* for the i^{th} row and the j^{th} column are

$$E_{ij} = \frac{(\text{row}_i \text{ total}) \cdot (\text{column}_j \text{ total})}{\text{grand total}}$$

leading to the *expected counts table* as

	Pass	Fail	Total
Group Study	$\frac{30 \cdot 24}{60} = 12$	$\frac{30 \cdot 36}{60} = 18$	30
Individual Study	$\frac{30 \cdot 24}{60} = 12$	$\frac{30 \cdot 36}{60} = 18$	30
Total	24	36	60

Denoting with O_{ij} the *observed counts* given in the assignement, the χ^2 statistic is

$$\chi^2_{\text{obs}} = \sum_{i,j=1}^{2} \frac{(O_{ij} - E_{ij})^2}{E_{ij}} = \frac{(18-12)^2}{12} + \frac{(6-12)^2}{12} + \frac{(12-18)^2}{18} + \frac{(24-18)^2}{18} = 10$$

To compair against the given table, we need the degrees of freedom

$$\text{df} = (\text{nr. rows} - 1) \cdot (\text{nr. columns} - 1) = (2-1) \cdot (2-1) = 1.$$

A lookup in the provided table gives the critical value as $\chi^2_{\text{df}}(\alpha) = \chi^2_1(5\%) = 3.84 < 10 = \chi^2_{\text{obs}}$.

With the observed statistic being bigger than the critical value, we *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(4). $\approx 66.7\%$
b(3). $= 0$
c(1). $= \frac{1}{2}$
d(2). $= 1$

Solution(Q13): The polynomial $x^3 - 3x^2 + 3x - 1 = (x-1)^3$, which is an odd function centered around 1. Similarly, we have for $X \sim \mathcal{N}(1, 4)$ that $P(X \leq 1) = P(X \geq 1) = \frac{1}{2}$. Therefore,

$$P(X^3 - 3X^2 + 3X - 1 \geq 0) = P((X-1)^3 \geq 0) = P(X \geq 1) = \frac{1}{2}.$$

Note: the probability $P(X = 1) = 0$, due to the continuouty of X.

(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(2). $\frac{3}{4} = 0.75$
b(3). $\frac{675}{4} = 168.75$
c(1). $\frac{45}{4} = 11.25$
d(4). 15

Solution(Q14): We start with the covariance estimate using the degrees of freedom $14 = \text{df} = n - 2$

$$\widehat{\text{Cov}}(X, Y) = \frac{1}{n-1} \sum_{i=1}^{n} (x_i - \overline{x})(y_i - \overline{y}) = \frac{S_{xy}}{n-1} = \frac{S_{xy}}{\text{df} + 1} = \frac{S_{xy}}{15}$$

The standard error of the slope $\hat{s}_\beta = 0.1$ with the residual standard error being $\hat{s} = 1.5$ gives

$$0.1 = \hat{s}_\beta = \frac{\hat{s}}{\sqrt{S_{xx}}} = \frac{1.5}{\sqrt{S_{xx}}} \qquad \Rightarrow \qquad S_{xx} = \left(\frac{1.5}{0.1}\right)^2 = 15^2$$

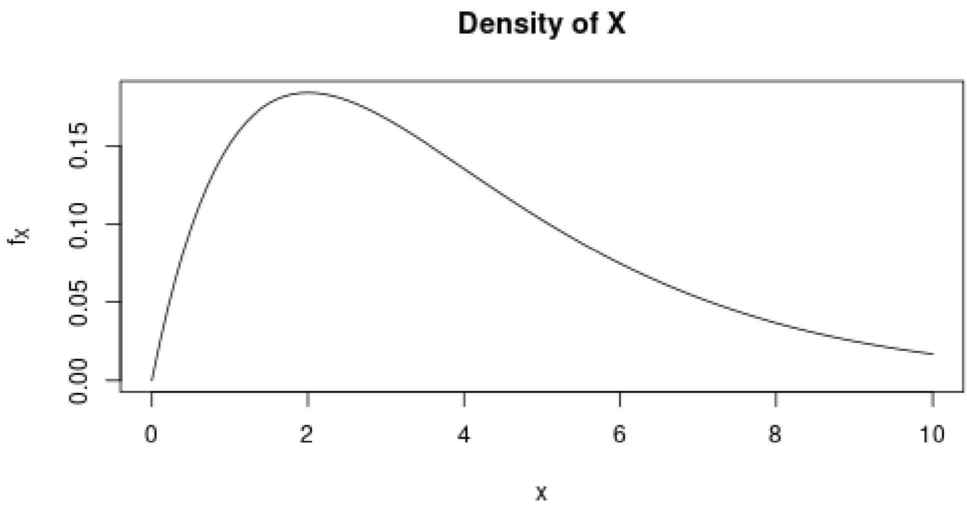
Similar, with the slope $\hat{\beta} = 0.75$ we get

$$0.75 = \hat{\beta} = \frac{S_{xy}}{S_{xx}} = \frac{S_{xy}}{15^2} \qquad \Rightarrow \qquad S_{xy} = 0.75 \cdot 15^2$$

Substituting back into the covariance estimate leads to the answer as

$$\widehat{\text{Cov}}(X,Y) = \frac{S_{xy}}{15} = \frac{0.75 \cdot 15^2}{15} = 0.75 \cdot 15 = \frac{45}{4} = 11.25$$

(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(1). $X \sim \chi^2(4)$, that is χ^2 with 4 degrees of freedom.
b(4). $X \sim \text{Binomial}(10, 1/2)$, that is binomial for 10 trials with success probability 1/2.
c(3). $X \sim \mathcal{N}(0, 1)$, that is standard normal.
d(2). $X \sim \text{Exp}(1)$, that is Exponential with rate 1.

Solution(Q15): The normal density is symmetric around its mean (not an option), and the Binomial does not have a density since it is descrete (not an option.) The Exponential density looks like "exponential decay", that is of the form e^{-x} (not an option), which only leaves $\chi^2(4)$ (which it is!)

(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(3). $H_0 : p_{\text{before}} \leq p_{\text{after}},\ H_1 : p_{\text{before}} > p_{\text{after}},\quad z \approx -0.29$
b(2). $H_0 : p_{\text{before}} = p_{\text{after}},\ H_1 : p_{\text{before}} > p_{\text{after}},\quad z \approx 0.29$
c(1). $H_0 : p_{\text{before}} = p_{\text{after}},\ H_1 : p_{\text{before}} < p_{\text{after}},\quad z \approx -4.65$
d(4). $H_0 : p_{\text{before}} = p_{\text{after}},\ H_1 : p_{\text{before}} \neq p_{\text{after}},\quad z \approx 0$

Solution(Q16): The one-sided test hypothesis for testing increased customer recognition are

$$H_0 : p_{\text{before}} = p_{\text{after}} \qquad \text{(no increase)}$$
$$H_1 : p_{\text{before}} < p_{\text{after}} \qquad \text{(recognition increased)}$$

Note: This already determines the correct answer!
For completenes, the z-test statistic is with $p_{\text{before}} = \frac{50}{250} = \frac{1}{5}$ and $p_{\text{after}} = \frac{80}{200} = \frac{2}{5}$ and the pooled proportion $\hat{p} = \frac{50+80}{250+200} = \frac{13}{45}$ given as

$$z = \frac{p_{\text{before}} - p_{\text{after}}}{\sqrt{\hat{p}(1-\hat{p})\left(\frac{1}{250} + \frac{1}{200}\right)}} = \frac{(1/5) - (2/5)}{\sqrt{\frac{13}{45}\frac{32}{45}\left(\frac{1}{250} + \frac{1}{200}\right)}} \approx -4.65$$

(17) Consider the following R code

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

What does it do?

a(3). 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(1). It approximates the probability that 6 tosses of a fair coin produce none or only 1 head.
c(4). The code has a bug, it throws an error.
d(2). It approximates the probability that a fair dice roles a 6.

Solution(Q17): The expression `rbinom(10000, n, p)` draws random samples from the Binomial distribution. Specifically, it generates independent and identically distributed (iid) samples $X_i \sim \text{Binomial}(n, p)$ for $i = 1, \ldots, 10\,000$, where $n = 6$ and $p = \frac{1}{2}$. This setup corresponds to 6 tosses of a fair coin.

Then, `rbinom(10000, n, p) < 2` converts the Binomial samples into a logical vector that contains `True` if $X_i < 2$ (i.e., if the result is either 0 or 1 heads) and `False` otherwise. Taking the mean of a logical vector is identical to computing the proportion of `True` entries in the vector (by summing 0 for `False` and 1 for `True`, and then dividing by the vector's length.)

Therefore, this code snippet approximates the probability that 6 tosses of a fair coin produce none or only 1 head, which is $\frac{7}{64} \approx 0.109$, by computing the proportion of events $X_i < 2$ based on 10 000 repetitions.

(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(1).
```
t.obs <- mean(x) / (sd(x) / 3)
2 * (1 - pt(abs(t.obs), 8))
```
b(2).
```
t.obs <- mean(x) / sd(x)
2 * (1 - pt(t.obs, 9))
```
c(3).
```
t.obs <- mean(x) / sd(x)
1 - pt(abs(t.obs), 9)
```
d(4).
```
t.obs <- mean(x) / (sd(x) / 3)
1 - pt(abs(t.obs), 8)
```

Solution(Q18): Counting the entries in `x` gives the sample size $n = 9$. Therefore, the degree of freedom for the t-test is `df` $= n - 1 = 8$. And, given a two-sided test, the p-value is double the tail-probability $2(1 - P(T \leq |t_{\text{obs}}|))$ = `2 * (1 - pt(abs(t.obs), 8))` for uniquely identifying the right answer.

Alternatively, one can look for the correct observed t-statistic $t_{\text{obs}} = \frac{\overline{X}-\mu}{s/\sqrt{n}} = \frac{\overline{X}}{s/3}$ = `mean(x) / (sd(x) / 3)` and then the answer with the two-sided p-value.

(19) Given a random sample $X_1, \dots, X_n$ 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(3). `dnorm(3, 2, 3 / sqrt(n))`
b(4). `pnorm(3 / sqrt(n))`
c(1). `pnorm(sqrt(n) / 3)`
d(2). `dnorm(n / 9)`

Solution(Q19): The function `pnorm(x, mu = 0, sd = 1)` computed the probability $P(X \leq x)$ for $X \sim \mathcal{N}(\mu, \sigma^2)$ (for `mu` $= \mu$ and `sd` $= \sigma$). Then, by the CLT we have

$$\bar{X} \approx \mathcal{N}\Big(2, \frac{9}{n}\Big) \qquad \Rightarrow \qquad \frac{\bar{X} - 2}{3/\sqrt{n}} \approx \mathcal{N}(0, 1).$$

In other words, we get that $P(\bar{X} \leq 3) = P(\bar{X} - 2 \leq 1) = P\Big(\frac{\bar{X}-2}{3/\sqrt{n}} \leq \frac{\sqrt{n}}{3}\Big)$ = `pnorm(sqrt(n) / 3)`.

(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(4). ≈ 30
b(3). ≈ 58
c(1). ≈ 50
d(2). ≈ 16

Solution(Q20): For $X \sim \mathcal{N}(\mu, \sigma^2)$ holds $P(\mu - 2\sigma \leq X \leq \mu + 2\sigma) \approx 95\% = (100 - 2 \cdot 2.5)\%$. This leads to the one sided approximate probability $P(X \leq \mu - 2\sigma) \approx (100 - 2.5)\% = 97.5\%$. We have given that $\mu - 2\sigma \approx 34\,\text{g}$ and $\sigma = 8\,\text{g}$, which leads to $\mu \approx 34\,\text{g} + 2 \cdot 8\,\text{g} = 50\,\text{g}$.