2024 Fall 340 Midterm

● Graded

Student

MATEJ POPOVSKI

Total Points

32.5 / 40 pts

Question 1

MC1 **1.2** / 2 pts

– **0.4 pts** 1 incorrect

✔ – **0.8 pts** 2 incorrect

– **2 pts** all incorrect

– **1.6 pts** 4 incorrect

– **1.2 pts** 3 incorrect

– **0 pts** All correct

Question 2

MC2 **2** / 2 pts

✔ – **0 pts** Correct

– **2 pts** Incorrect/skipped

Question 3

MC3 **2** / 2 pts

✔ – **0 pts** Correct

– **2 pts** incorrect/skipped

Question 4

MC4 **2** / 2 pts

– **0 pts** A: Correct

✔ – **0 pts** not A: incorrect/ skipped

Question 5

MC5 **2** / 2 pts

5.1 **(no title)** **0.5** / 0.5 pts

✔ **– 0 pts** Correct

– 0.5 pts Click here to replace this description.

5.2 **(no title)** **0.5** / 0.5 pts

✔ **– 0 pts** Correct (dnorm)

– 0.5 pts incorrect

5.3 **(no title)** **0.5** / 0.5 pts

✔ **– 0 pts** Correct

– 0.5 pts Incorrect

5.4 **(no title)** **0.5** / 0.5 pts

✔ **– 0 pts** Correct

– 0.5 pts incorrect

Question 6

MC6 **2** / 2 pts

✔ **– 0 pts** b & e: Correct

– 0.4 pts circled a

– 0.4 pts didn't circle b

– 0.4 pts circled c

– 0.4 pts circled d

– 0.4 pts didn't circle e

– 2 pts Skipped (none circled)

Question 7

MC7 **2** / 2 pts

✔ **– 0 pts** Correct

– 2 pts incorrect / no answer

Question 8

MC8 **2** / 2 pts

✔ **– 0 pts** Correct

– 2 pts incorrect/skipped

Question 9

MC9 **1** / 2 pts

9.1 **(no title)** **0.5** / 0.5 pts

✔ **– 0 pts** Correct

– 0.5 pts incorrect

9.2 **(no title)** **0** / 0.5 pts

– 0 pts Correct

✔ **– 0.5 pts** incorrect

9.3 **(no title)** **0.5** / 0.5 pts

✔ **– 0 pts** Correct

– 0.5 pts incorrect

9.4 **(no title)** **0** / 0.5 pts

– 0 pts Correct

✔ **– 0.5 pts** incorrect

Question 10

MC10 **2** / 2 pts

✔ **– 0 pts** Correct

– 2 pts incorrect/skipped

Question 11

MC11 **1.2** / 2 pts

– 0 pts Correct

– 0.4 pts 4 of 5 selected

✔ **– 0.8 pts** 3 of 5 circled

– 1.2 pts 2 of 5 selected

– 1.6 pts 1 of 5 selected

Question 12

MC12 **1.6** / 2 pts

- **– 0 pts** Correct
- **– 0.4 pts** circled A
- **– 0.4 pts** circled B
- ✔ **– 0.4 pts** didn't circle C
- **– 0.4 pts** circled D
- **– 0.4 pts** didn't select E

Question 13

SA1 **4** / 4 pts

a: **FALSE; X is continuous**

- **– 1 pt** a. incorrectly answers **TRUE**
- **– 0.5 pts** answers **FALSE** but does not do justification or wrong justification
- **– 0.5 pts** Answers correctly but mentioned TRUE instead of FALSE mistakenly.
- **– 0.5 pts** Answers FALSE but with completely wrong justification (such as using a binomial distribution to explain)
- **– 0.25 pts** Justification almost correct, but wrote f(0)=f(1)=0, which is not true.__

---

b: **TRUE; F is nondecreasing**

- **– 1 pt** incorrectly answers **FALSE**
- **– 0.5 pts** justification is incorrect or not provided
- **– 0.25 pts** justification is partially correct (e.g., CDF doesn't grow linearly, if someone wrote it grows linearly, points are taken off.)

---

c: **TRUE; numerator of formula must = 0**

- **– 1 pt** incorrectly answers **FALSE**
- **– 0.5 pts** **TRUE** but justification is not correct
- **– 0.4 pts** States P(EnF)=0 or EnF={} but does not justify with conditional probability formula (Nabil: I considered identifying as mutually exclusive as a "correct explanation").

---

d: **FALSE; P(E)P(F)>0, but P(EnF)=0**

- **– 1 pt** incorrectly answers **TRUE**
- **– 0.5 pts** justification is not good

- ✔ **– 0 pts** All correct

Question 14

SA2 **4** / 4 pts

a. **E(-3X)= -3(5)=-15**

✔ **– 0 pts** correct

– 1 pt incorrect

– 0.5 pts The process is correct but there is a typo in the answer.

---

b **SD(2Y) = 2(3)=6**

✔ **– 0 pts** Correct

– 1 pt incorrect

– 0.25 pts Answered correctly, but typo in the expression.

---

c **E(X+Y) = 5+(-2)=3**

✔ **– 0 pts** Click here to replace this description.

– 1 pt Click here to replace this description.

---

d. **Var(X-Y) = 2^2+3^2=13**

✔ **– 0 pts** Click here to replace this description.

– 1 pt incorrect

– 0.5 pts The process was correct but there was a typo/miscalculation in the answer.

– 0 pts correct

Question 15

SA3 **0.5** / 4 pts

a. $P(Pos|disease) = .88$

✔ **– 1 pt** completely incorrect

---

b. $P(disease|Pos) = \frac{\frac{3}{1000}(.88)}{\frac{3}{1000}(.88) + \frac{997}{1000}(.04)}$

– 2 pts completely incorrect

– 1.5 pts conditional formula only, no use of Bayes formula

– 0.5 pts one error in the formula

– 1 pt two errors in formula

– 1 pt partially filled out formula

– 1 pt incorrect denominator

✔ **– 1.5 pts** many errors in formula

Click here to replace this description.

– 0 pts c. $P(Neg) = \frac{3}{1000}(.12) + \frac{997}{1000}(.96)$

– 0 pts $P(Neg) = 1 - \left(\frac{3}{1000}(.88) + \frac{997}{1000}(.04)\right)$

✔ **– 1 pt** incorrect

– 1 pt Stated P(neg)=1-P(pos) but nothing more

– 1 pt skipped

– 0.5 pts expand P(neg) but don't put numbers in the formula

– 0.5 pts one error in the formula

– 0.75 pts Found P(Pos)

Question 16

SA4 **3** / 4 pts

a. H0: photos **is not** a raccoon; H1: photo **is** a raccoon

- **– 0.5 pts** H_0 incorrect
- **– 0.5 pts** H_1 incorrect

---

b. classify a non-raccoon as a raccoon

- **– 1 pt** incorrect

---

c. Strong evidence that the photo is a raccoon.

- **– 0 pts** Reject the null that the photo is not a raccoon.
- **– 0 pts** only 1.4% of non-raccoons look this "raccoon-like" or more
- **– 1 pt** interpretation incorrect

---

d. The type 1 error rate = 0.019 ("level" or "α")

- ✔ **– 1 pt** calls it Type 2 error rate, or power, or accuracy
- **– 1 pt** Click here to replace this description.

- **– 0 pts** correct

FA24 STAT340 Midterm in-class portion

First (given) name:

Write here: Matej

Last (family) name:

Write here: Popovski

Lecture section:

Circle one: **Bi's section** **Brian's section**

Instructions:

- Note some MC are "**choose ONE**" and some are "**choose ALL that apply**".
  - "Choose ONE" means there is exactly one right answer.
  - "Choose ALL that apply" means there is at least one right answer (i.e. the number of right answers is between 1 and all of them).
- Partial credit available only for SA and "choose ALL" MC questions.
- For SA, work must be shown for full credit.
- Please **do NOT simplify any expressions** after plugging in numbers.

MC1-3 (/6)	MC4-7 (/8)	MC8-12 (/10)	SA1 (/4)	SA2 (/4)	SA3 (/4)	SA4 (/4)	Total (/40)

Multiple Choice 2pts each

MC1

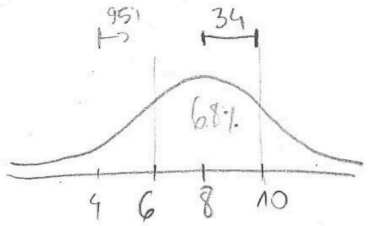
For which random variables X will $\mathrm{E}(X)$ increase if the parameter θ increases? Choose ***ALL*** that apply.

a. $X \sim \text{Binomial}(n=\theta,\, p)$
b. $X \sim \text{Poisson}(\lambda = \theta)$
c. $X \sim \text{Geometric}(p = \theta)$
d. $X \sim \text{Normal}(\mu,\, \sigma^2 = \theta)$
e. $X \sim \text{Uniform}(a,\, b = \theta)$

MC2

Random variable $X \sim \text{Normal}(8, 2^2)$. Which of the following is ***closest*** to $P(X > 10)$? Choose ***ONE***.

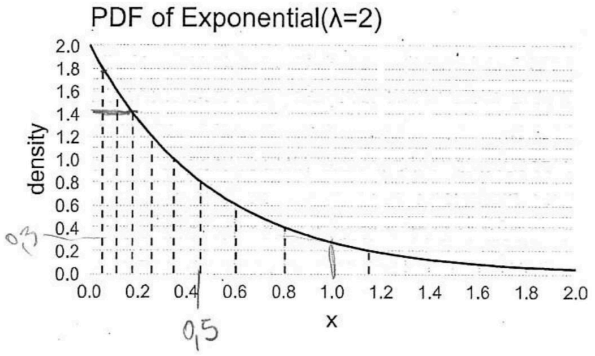
a. 0.05
b. 0.16
c. 0.32
d. 0.68
e. 0.95



50-34 = 16

MC3,4

For MC3,4, $X \sim \text{Exponential}(\lambda = 2)$ (PDF shown). Vertical dashed lines show the 10th, 20th, …, 90th percentiles.



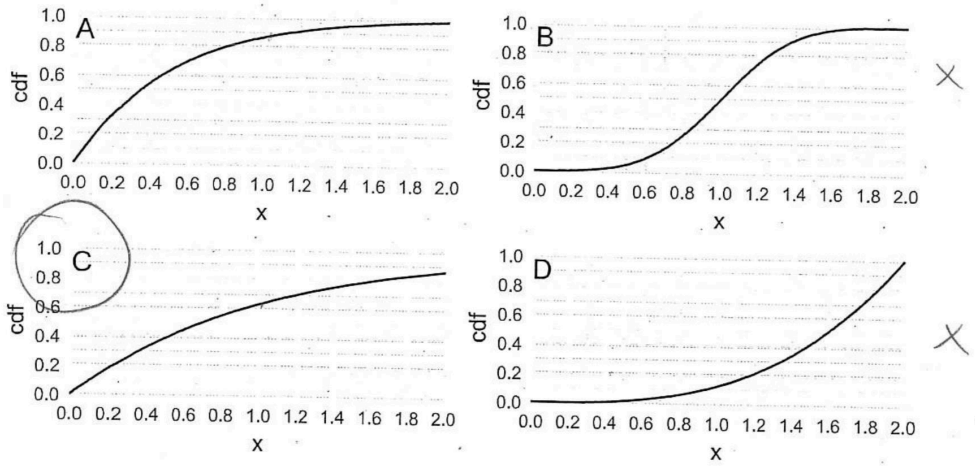
MC3

What is the approximate difference between the mean and the median of X? Choose ***ONE***.

a. 0
b. 0.15
c. 0.35
d. 0.5
e. 1

MC4

Which of the following is the CDF of X? Choose ***ONE***.



MC5

Consider a normal variable. Match each description below with one of the following R functions: `rnorm`, `pnorm`, `dnorm`, `qnorm`.

a. cumulative function: pnorm
b. density function: dnorm
c. inverse cumulative function: qnorm
d. simulation function: rnorm

MC6

Let X, Y be arbitrary random variables. Which of the following are NOT always true? Choose ***ALL*** that apply.

a. $\mathrm{E}(2X - 1) = 2\,\mathrm{E}(X) - 1$ ✓
b. $\mathrm{Var}(3X) \leq 3\,\mathrm{Var}(X)$ ×
c. $\mathrm{Var}(X + 3) \leq \mathrm{Var}(X)$ ✓
d. $\mathrm{E}(X + Y) = \mathrm{E}(X) + \mathrm{E}(Y)$
e. $\mathrm{Var}(X + Y) = \mathrm{Var}(X) + \mathrm{Var}(Y)$

MC7

You run a small MC simulation to estimate the probability of event E. You record the following sample of 1s and 0s where 1 indicates E occurred and 0 indicates E did not occur: $0, 0, 1, 0, 0, 1, 0, 1, 1, 0$. Which of the following intervals contains your estimate of $P(E)$? Choose ***ONE***.

a. $(0.1, 0.3]$ ×
b. $(0.3, 0.5]$ ✓
c. $(0.5, 0.7]$ ×
d. $(0.7, 0.9]$ ×
e. Not enough information to determine. ×

4/10 times occured
0,4 should be inside the interval

MC8

Which of the following is true if we get a p-value of 0.02? Choose ***ONE***.

a. There is a 2% chance that the null hypothesis is true given the data we've observed. ✓?
b. There is a 2% chance that the alternative hypothesis is true given the data we've observed. ×
c. There is a 2% chance of seeing data like this (or more extreme) if the null hypothesis is true. ✓
d. There is a 2% chance of seeing data like this (or more extreme) if the alternative hypothesis is true. ×
e. There is a 2% chance we reject the null hypothesis. ×

MC9

Suppose you are a quality assurance agent at a factory checking for defects. For each product, define H_0 as "there is no defect". You pass or fail products depending on if you think there's a defect or not. Label each of the following outcomes with the following labels: true positive, true negative, type I error, type II error.

a. You passed a defective product: Type 2
b. You passed a good product: TP
c. You failed a good product: Type 1
d. You failed a defective product: TN

MC10

My test statistic T follows a discrete distribution, taking only integer values. Which of the following is true? Choose ***ONE***.

a. A different standard α is commonly used for discrete distributions instead of the usual 0.05. ×
b. The test will have significantly greater power compared to similar tests using continuous statistics. ×
c. In some cases, it's possible that the p-value could be greater than 1. ×
d. It may be hard to get an exact α-level rejection rule for an arbitrary α due to discreteness.
e. The CDF of T is undefined because it is discrete. ×

MC11

Which of the following statements are false? Choose ***ALL*** that apply.

a. 1-tail tests always have higher power than 2-tail tests. ✓
b. When the null hypothesis is rejected, it's because it's not true. ×
c. The null and alternative hypotheses are stated using sample statistics.
d. The larger the p-value, the stronger the evidence against the null. ✓
e. If your result is not significant, you should increase α and try again.

1 sided test is more powerful than 2

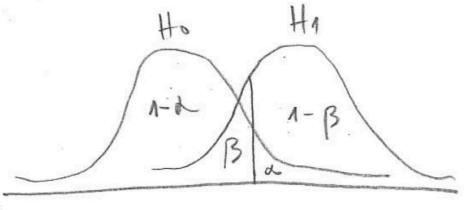
1 sided test is more likely to correctly detect and reject H_0

↳ Yes, I technically can, but I SHOULD NOT.

MC12

You decide to use a lower $\alpha = 0.01$ instead of the usual 0.05. Which of the following will also tend to decrease? Choose ***ALL*** that apply.

a. The computed test statistic. ×
b. The computed p-value. ×
c. The rate of type I errors.
d. The rate of type II errors.
e. The power of the test.



$\alpha \uparrow$ $\beta \downarrow$ Power $\uparrow$

$P(E \cup F) = P$

Short Answer 4pts each

SA1

For each of the following statements, state if it's true or false and ***briefly explain why***! Full credit only possible with correct explanation. (Note each part is a separate question with no relation to other parts).

a. For $X \sim N(0,1)$, $P(X = 0) > P(X = 1)$ because $f(0) > f(1)$.
b. For a random variable X with CDF $F(x)$, if $a \leq b$ then $F(a) \leq F(b)$.
c. If E, F are events with non-zero probability and $P(E \mid F) = 0$, then $P(E \text{ or } F) = P(E) + P(F)$
d. If E, F are events with non-zero probability and $P(E \mid F) = 0$, then $P(E \text{ and } F) = P(E) \cdot P(F)$

a) In normal distribution, by definition, we cannot check for a discrete value, because normal dist. is continious. So $P(X = \text{any number } n) = 0$. Therefore, $P(X=0) > P(X=1)$ is a false statement because $0=0$. Checking the density does not participate in the correctness of this rule.

b) Yes, the statement is correct, because cumulative functions (CDF) are allways growing functions, meaning as x increases $F(x)$ MUST increase too.

c) If given $P(E|F) = 0 \Rightarrow E$ and F are mutually exclusive, meaning both events cannot happen at the same time.

$P(E|F) = \frac{P(E,F)}{P(F)} \to 0$ ($P(E,F) \to 0$, $P(F) \neq 0$)

$P(E \text{ or } F) = P(E) + P(F)$ is true because $P(E \cup F) = P(E) + P(F) - P(E \cap F)$, where $P(E \cap F)$ = zero

d) Same explanation for mutual exclusiveness.

$P(E \text{ and } F) = P(E) \cdot P(F)$ is False because $P(E) \cdot P(F) \neq 0$ and above I explained that $P(E,F) = 0$

SA2

$Var = sd^2$

Variables X and Y are independent. Their expected values and variances are:

$$\mu_X = 5,\ \sigma_X^2 = 2^2, \quad \mu_Y = -2,\ \sigma_Y^2 = 3^2.$$

$Var = 4$ $\quad$ $Var = 9$

a. What is $\mathrm{E}(-3X)$?
b. What is $\mathrm{SD}(2Y)$? (i.e. the standard deviation of $2Y$?)
c. What is $\mathrm{E}(X+Y)$?
d. What is $\mathrm{Var}(X-Y)$?

a) $E(-3X) = -3E(x) = -3 \cdot 5$

b) $SD(2y) = 2SD(y) = 2 \cdot 3$

c) $E(x+y) = E(x) + E(y) = 5 + (-2)$

d) $Var(x-y) = 1^2 Var(x) + (-1)^2 Var(y) = Var(x) + Var(y) =$
$= 2^2 + 3^2$

$\beta = 88$

SA3 1,2,1pts

The prevalence of a disease in the general population is $3/1000$. A test for the disease has a false positive rate of 4% and a power of 88%. Find each of the following. Please **do NOT simplify** your expressions!

a. If a patient has the disease, what is the probability they will get a positive test result?
b. If a patient gets a positive test result, what is the probability they have the disease?
c. What proportion of test takers overall would get a negative test result?

D = disease

$P(D) = 0{,}003$ or $3/1000$

$P(\text{Posit} \mid \text{No } D) = 4/100 \Rightarrow P(\text{Pos} \mid D) = 96/100$

$P(\text{Neg} \mid D) =$

$\text{Power} = \frac{88}{100} = P(\text{Neg} \mid \text{No } D) = 88/100$

$\Rightarrow P(\text{Neg} \mid D) = 12/100$

a) $P(\text{Pos} \mid D) = 1 - P(\text{Pos} \mid \text{No } D) = 96/100$

b) $P(D \mid \text{Pos}) = \frac{P(\text{Pos} \mid D) \cdot P(D)}{P(\text{Pos})} = \frac{P(\text{Pos} \mid D) \cdot P(D)}{P(\text{Pos} \mid D) \cdot P(D) + P(\text{Pos} \mid \text{No } D) \cdot P(\text{No } D)}$

$= \frac{96/100 \cdot 3/1000}{(96/100 \cdot 3/1000) + (4/100 \cdot \frac{997}{1000})}$

c) $P(\text{Neg} \mid D) + P(\text{Neg} \mid \text{No } D) =$

$= \frac{12}{100} +$

H_0: The ML performs good.
H_1: The ML performs bad.

SA4

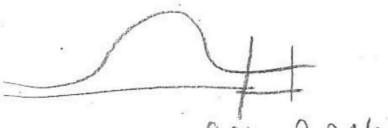
A data scientist trains a machine learner to identify photos of raccoon. The intention is that the learner assumes each photo is something else unless there is enough evidence to identify it as a raccoon.

a. State formally the null and alternative hypotheses.
b. What would a type I error be in this case?
c. The algorithm is run on a photo and gives a p-value of 0.014. How would you interpret this p-value?
d. 1000 photos of various animals that are not raccoons are fed into the classifier. If 19 of them are identified as raccoons, what does this tell you about the test? Be specific and quantitative!

a) H_0: There is a racoon in the photo.
H_1: There is NO a raccoon in the photo.

b) Type 1 error would be: There is no racoon in the photo, but she identifies that there is.

c) P-val = 0,014 < 0,05.
We reject the H_0 hypothesis.
We have enough evidence to reject the H_0.



d) We should know how many photos actually had a racoon in them so we can conclude whether the classified did a good job.
From what is given I can say that:
the error rate is 19/1000 = 0,019.
This p value does not differ a lot from the last test ~ 0,014.
Our $\alpha = 0,05$, and we would still reject the H_0 hypothesis.
We reject H_0, which means the classifier performs good (well)