Examination paper for TMA4240 Statistics

Examination date : 26/11/2020

1A

Introduction: Let X and Y be independent and normal distributed stochastic variables. Assume that X has mean 0.7 and standard deviation 0.5, and that Y has mean -0.3 and standard deviation 0.5.

Exercise: Fill in the correct values for the following probabilities. Enter the answer with two decimal places.

$$P(X \geq 1) = 0.274$$
$$P(X \leq 1.5 | X \geq 1) = 0.8$$
$$P(2X - Y > 1) = 0.734$$

1B

Innledning: La X og Y være uavhengige og normalfordelte stokastiske variabler. Anta at X har forventningsverdi 0.6 og standardavvik 0.5, og at Y har forventningsverdi -0.3 og standardavvik 0.5.

Oppgave: Fyll inn riktige verdier for følgende tre sannsynligheter. Angi verdi med to siffer etter komma.

$$P(X \geq 0.8) = 0.34$$
$$P(X \leq 1.5 | X \geq 0.8) = 0.90$$
$$P(2X - Y > 1) = 0.67$$

1C

Introduction: Let X and Y be independent and normal distributed stochastic variables. Assume that X has mean 0.7 and standard deviation 0.5, and that Y has mean -0.3 and standard deviation 0.8.

Exercise: Fill in the correct values for the following probabilities.Enter the answer with two decimal places.

$$P(Y \geq 1) = 0.052$$
$$P(Y \leq 1.5 | Y \geq 1) = 0.76$$
$$P(X - 2Y > 1) = 0.57$$

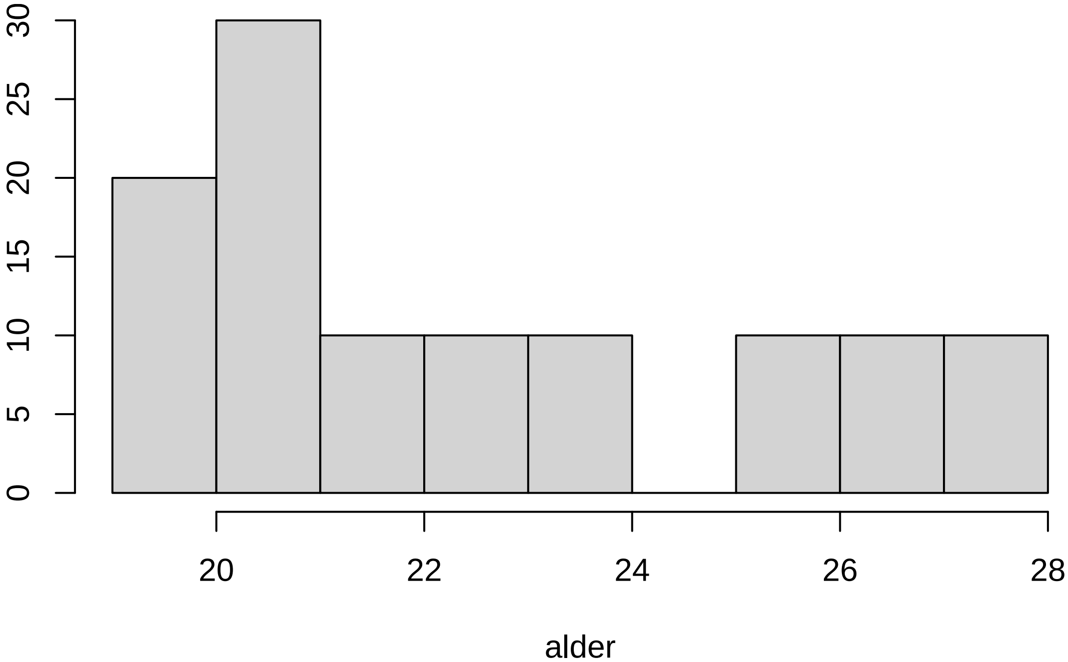
1D

Introduction: Let X and Y be independent and normal distributed stochastic variables. Assume that X has mean 0.7 and standard deviation 0.5, and that Y has mean -0.3 and standard deviation 0.8.

Exercise: Fill in the correct values for the following probabilities.Enter the answer with two decimal places.

$$P(Y \geq 0.9) = 0.06$$
$$P(Y \leq 1.5 | Y \geq 0.9) = 0.81$$
$$P(X - 2Y > 0.8) = 0.61$$

2A

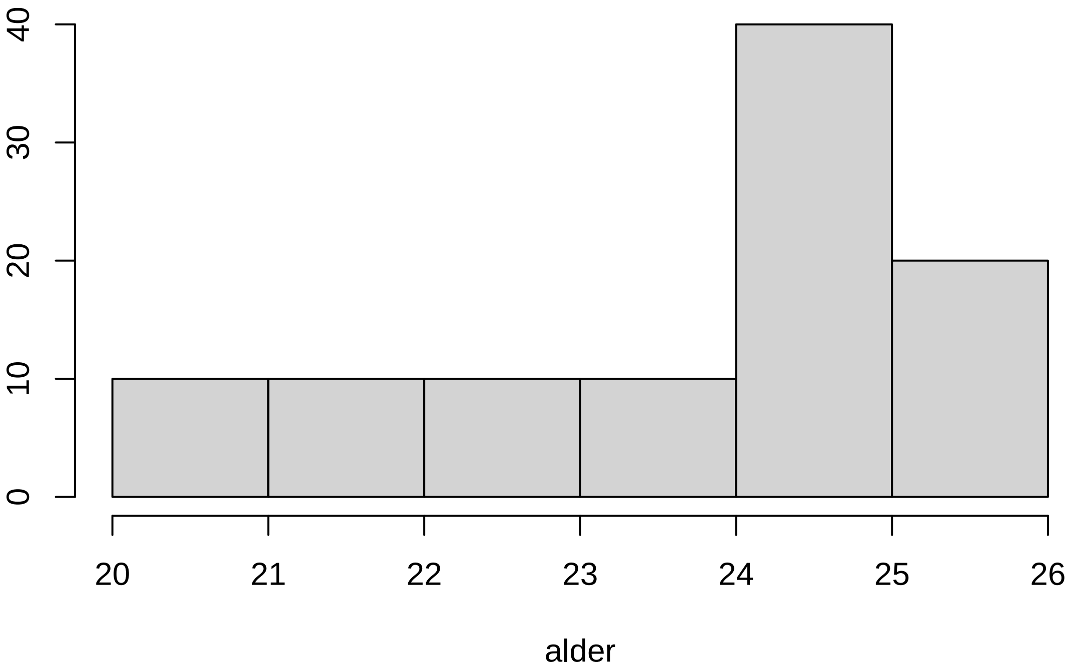


Innledning: Histogrammet ovenfor viser aldersfordelingen på personer som søker en bestemt stilling.

Oppgave: Hvilke av følgende utsagn er sanne?

- Den empiriske medianen er ca lik som gjennomsnitt
- Den empiriske medianen er større en gjennomsnitt
- **Den empiriske medianen er mindre en gjennomsnitt**
- Den empiriske medianen er mellom 24 og 25
- **Gjennomsnitt er mellom 22 and 23**

2B

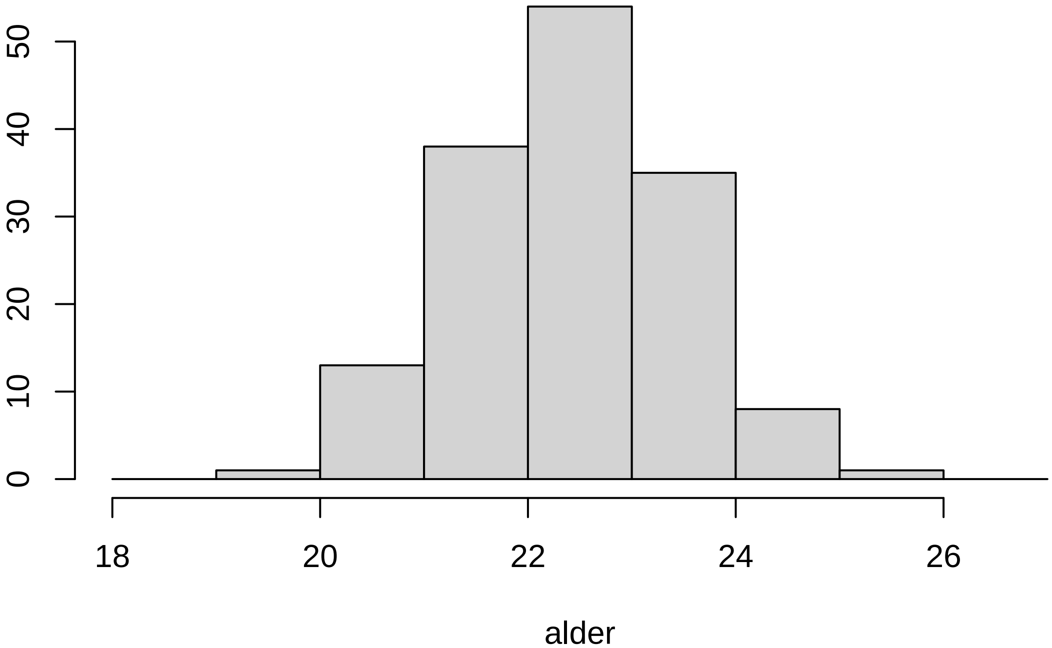


Innledning: Histogrammet ovenfor viser aldersfordelingen på personer som søker en bestemt stilling.

Oppgave: Hvilke av følgende utsagn er sanne?

- Den empiriske medianen er ca lik som gjennomsnitt
- **Den empiriske medianen er større en gjennomsnitt**
- Den empiriske medianen er mindre en gjennomsnitt
- **Den empiriske medianen er mellom 24 og 25**
- Gjennomsnitt er mellom 22 and 23

2C



Innledning: Histogrammet ovenfor viser aldersfordelingen på personer som søker en bestemt stilling.

Oppgave: Hvilke av følgende utsagn er sanne?

- **Den empiriske medianen er ca lik som gjennomsnitt**
- Den empiriske medianen er klart større en gjennomsnitt
- Den empiriske medianen er klart mindre en gjennomsnitt
- Den empiriske medianen er mellom 24 og 25
- **Gjennomsnitt er mellom 22 and 23**

3A

Innledning: La Y være Poisson fordelt med parameter $\lambda = 10$

Oppgave: Finn

- $P(Y = 7\ ) = 0.09$
- $P(Y \geq 8\ ) = 0.78$
- $P(Y < 10\ |Y \geq 8\ ) = 0.3$

3B

Innledning: La X være geometrisk fordelt med parameter $p = 0.3$. Dvs at X har sannsynlighet fordeling

$$P(X = x) = p(1 - p)^{x-1}$$

Oppgave: Finn

- $P(X = 5\ ) = 0.07$
- $P(X \geq 3\ ) = 0.49$
- $P(X < 5\ |X \geq 3\ ) = 0.51$

3C

Innledning: La X være geometrisk fordelt med parameter $p = 0.1$. Dvs at X har sannsynlighet fordeling

$$P(X = x) = p(1 - p)^{x-1}$$

Oppgave: Finn

- $P(X = 5\ ) = 0.07$
- $P(X \geq 3\ ) = 0.81$
- $P(X < 5\ |X \geq 3\ ) = 0.19$

3D

Innledning: La Y være Poisson fordelt med parameter $\lambda = 5$

Oppgave: Finn

- $P(Y = 1\ ) = 0.03$
- $P(Y \geq 2\ ) = 0.96$
- $P(Y < 4\ |Y \geq 2\ ) = 0.23$

%# Venndiagram

4A-D

The four version of the exercises all had the same possible choices. In some cases the left and right sides were inverted.

Introduction: Let A, B and C be three events in a sample space S

Exercise: Which of the following statements are always correct for three events? Hint:Draw a Venn diagram and use this to find which statements are correct.

- $(A \cap B) \cap C = A \cap (B \cap C)$ - Correct
- $(A \cap B) \cap C' = (A \cap C') \cap (B \cap C')$- Correct
- $A \setminus (B \cup C) = (A \cap B') \cap C'$- Correct
- $(A \cap B)' = A' \cap B'$
- $(A \cap B)' \cup C = (A \cap B) \cap C'$
- $(A \cup B) \cap C = A \cup (B \cap C)$

5A

Innledning: La X være en stokastisk variabel men sannsynlighet tetthet

$$f(x) = \begin{cases} 1 + x & \text{for } x \in (-1, 0) \\ 1 - x & \text{for } x \in (0, 1) \end{cases}$$

Oppgave:

- $P(X > 0.3\ ) = 0.24$
- $P(X <$-0.2 $) = 0.32$‘
- $P(X >$ -0.2$|X < 0.3\ ) = 0.58$

5B

Innledning: La X være en stokastisk variabel men sannsynlighet tetthet

$$f(x) = \begin{cases} \frac{1}{2}\exp(x) & \text{for } -\log 2 < x \leq 0 \\ \frac{1}{2}(x+1) & \text{for } 0 < x < 1 \end{cases}$$

Oppgave: Finn

- $P(X > 0.5\ ) = 0.44$
- $P(X < 0.2\ ) = 0.36$
- $P(X > 0.2|X < 0.5\ ) = 0.36$

5C

Innledning: La X være en stokastisk variabel men sannsynlighet tetthet

$$f(x) = \begin{cases} \frac{1}{2}\exp(x) & \text{for } -\log 2 < x \leq 0 \\ \frac{1}{2}(x+1) & \text{for } 0 < x < 1 \end{cases}$$

Oppgave: Finn

- $P(X > 0.3\ ) = 0.58$
- $P(X <$ -0.2 $) = 0.16$
- $P(X >$ -0.2$|X < 0.3\ ) = 0.62$

5D

Innledning: La X være en stokastisk variabel men sannsynlighet tetthet

$$f(x) = \begin{cases} 1 + x & \text{for } x \in (-1, 0) \\ 1 - x & \text{for } x \in (0, 1) \end{cases}$$

Oppgave: Finn

- $P(X > 0.4\ ) = 0.18$
- $P(X < 0.2\ ) = 0.68$
- $P(X > 0.2|X < 0.4\ ) = 0.17$

6A

X is a SV with distribution

$$f(x) = \begin{cases} 2x\exp(-x^2) & \text{for } x > 0 \\ 0 & \text{ellers} \end{cases}$$

The median is

$$m = \sqrt{log(2)} = 0.832$$

6B

X is a SV with distribution

$$f(x) = \begin{cases} \frac{1}{3}(4x+1) & \text{for } x \in (0,1) \\ 0 & \text{ellers} \end{cases}$$

Let m indicate the third quartile of X, then by definition:

$$0.75 = \int_{-\infty}^{m} f(x)dx =$$

and in this case the solution is $m = \frac{-4+\sqrt{304}}{16} = 0.84$

6C

X is a SV with distribution

$$f(x) = \begin{cases} 3x^2\exp(-x^3) & \text{for } x > 0 \\ 0 & \text{ellers} \end{cases}$$

Let m indicate the first quartile of X, is by definition:

$$0.25 = \int_{-\infty}^{m} f(x)dx$$

from this we get that

$$m = (-\log(0.75))^{1/3} = 0.660$$

6D

X is a SV with distribution

$$f(x) = \begin{cases} \frac{1}{3}(4x+1) & \text{for } x \in (0,1) \\ 0 & \text{ellers} \end{cases}$$

The median is $m = \frac{-2+\sqrt{52}}{8} = 0.65$

7A

Introduction: Assume that we have an urn with 20 balls: 8 red, 10 yellow and the rest blue. Assume further that we randomly draw 11 balls without replacement.

Exercise: If we do not take into account the order the balls are drawn, in how many ways can we draw:

- exactly 5 red balls? Enter the answer as an integer (51744)
- either exactly 5 red or exactly 5 yellow balls (including the cases where it is both exactly 5 red and exactly 5 yellow)? Enter the answer as an integer. (76440)

7B

Introduction: Assume that we have an urn with 20 balls: 9 red, 6 yellow and the rest blue. Assume further that we randomly draw 13 balls without replacement.

Exercise: If we do not take into account the order the balls are drawn, in how many ways can we draw

- exactly 5 red balls? Give the answer as an integer. (20790)
- either exactly 5 red or exactly 5 yellow balls (including the cases where it is both exactly 5 red and exactly 5 yellow)? Enter the answer as an integer. (31248)

7C

Introduction: Assume that we have an urn with 20 balls: 6 red, 6 yellow and the rest blue. Assume further that we randomly draw 15 balls without replacement.

Exercise: If we do not take into account the order the balls are drawn, in how many ways can we draw:

- exactly 5 red balls? Give the answer as an integer. (6006)
- either exactly 5 red or exactly 5 yellow balls (including the cases where it is both exactly 5 red and exactly 5 yellow)? Enter the answer as an integer. (9996)

7D

Introduction: Assume we have an urn with 20 balls: 5 red, 7 yellow and the rest blue. Assume further that we draw 12 balls without replacement.

Exercise: If we do not take into account the order the balls are drawn, in how many ways can we draw

- exactly 5 red balls? Give the answer as an integer.(6435)
- either exactly 5 red or exactly 5 yellow balls (including the cases with both exactly 5 red balls and exactly 5 yellow balls)? Enter the answer as an integer (41883)

8A

Introduction: Let X_1 and X_2 be two dependent random variables with $E(X_1) = 0$,$E(X_2) = -1$,$\text{Var}(X_1) = 2$,$\text{Var}(X_2) = 2$and $\text{Cov}(X_1, X_2) = 1$. Assume further that we have a stochastic variable Y that is independent of X_1 and X_2 and with $E(Y) = 3$ and $\text{Var}(Y) = 2$

Let the stochastic variables Z_1 and Z_2 be defined as

$$Z_1 = X_2 + 2Y \quad \text{and} \quad Z_2 = 3X_1 + 2X_2 - 4Y.$$

Exercise: Find the mean and the variance of Z_1 and Z_2. Enter the answers as integers.

$$\begin{aligned} \text{E}[Z_1] &= 6 \\ \text{Var}[Z_1] &= 10 \\ \text{E}[Z_2] &= -14 \\ \text{Var}[Z_2] &= 70 \end{aligned}$$

8B

Introduction: Let X_1 and X_2 be two dependent random variables with $E(X_1) = 0$,$E(X_2) = -1$,$\text{Var}(X_1) = 3$,$\text{Var}(X_2) = 2$and $\text{Cov}(X_1, X_2) = 1$. Assume further that we have a stochastic variable Y that is independent of X_1 and X_2 and with $E(Y) = 3$ and $\text{Var}(Y) = 2$

Let the stochastic variables Z_1 and Z_2 be defined as

$$Z_1 = 3X_1 + Y \quad \text{and} \quad Z_2 = X_1 - 4X_2 + 2Y.$$

Exercise: Find the mean and the variance of Z_1 and Z_2. Enter the answers as integers.

$$\begin{aligned} \text{E}[Z_1] &= 3 \\ \text{Var}[Z_1] &= 29 \\ \text{E}[Z_2] &= 10 \\ \text{Var}[Z_2] &= 35 \end{aligned}$$

8C

Introduction: Let X_1 and X_2 be two dependent random variables with $E(X_1) = 0$,$E(X_2) = -2$,$\text{Var}(X_1) = 3$,$\text{Var}(X_2) = 2$and $\text{Cov}(X_1, X_2) = 1$. Assume further that we have a stochastic variable Y that is independent of X_1 and X_2 and with $E(Y) = 3$ and $\text{Var}(Y) = 2$

Let the stochastic variables Z_1 and Z_2 be defined as

$$Z_1 = X_2 + 2Y \quad \text{and} \quad Z_2 = 8X_1 + 2X_2 - Y.$$

Exercise: Find the mean and the variance of Z_1 and Z_2. Enter the answers as integers.

$$\begin{aligned} \text{E}[Z_1] &= 4 \\ \text{Var}[Z_1] &= 10 \\ \text{E}[Z_2] &= -7 \\ \text{Var}[Z_2] &= 234 \end{aligned}$$

9A

$$f_{Y_i}(y_i) = \begin{cases} \frac{\lambda^4 x_i^4}{6} y_i^3 e^{-\lambda x_i y_i} & \text{for } y_i \geq 0 \\ 0 & \text{otherwise} \end{cases}$$

Let the stochastic variable Z be defined as

$$Z = \sum_{i=1}^{n} x_i Y_i.$$

Exercise: Using the moment generating function, determine which of the following probability distributions is the correct distribution for Z.

- Chi-squared distribution with $4n$ degrees of freedom.
- **Gamma distribution with $\alpha = 4n$ and $\beta = \frac{1}{\lambda}$**
- T-distribution with $4n$ degrees of freedom.
- Gamma distribution with $\alpha = 4n$ and $\beta = \lambda$
- Chi-squared distribution with $8n$ degrees of freedom.
- Chi-squared distribution with $2n$ degrees of freedom.
- T-distribution with $2n$ degrees of freedom.

9B

$$f_{Y_i}(y_i) = \begin{cases} \frac{\theta^3 x_i^3}{2} y_i^2 e^{-\theta x_i y_i} & \text{for } y_i \geq 0 \\ 0 & \text{otherwise} \end{cases}$$

Let the stochastic variable Z be defined as

$$Z = \sum_{i=1}^{n} x_i Y_i.$$

Exercise: Using the moment generating function, determine which of the following probability distributions is the correct distribution for Z.

- Chi-squared distribution with $3n$ degrees of freedom.
- **Gamma distribution with $\alpha = 3n$ and $\beta = \frac{1}{\theta}$**
- T-distribution with $3n$ degrees of freedom.
- Gamma distribution with $\alpha = 3n$ and $\beta = \theta$
- Chi-squared distribution with $6n$ degrees of freedom.
- Chi-squared distribution with $3n/2$ degrees of freedom.
- T-distribution with $3n/2$ degrees of freedom.

9C

$$f_{Y_i}(y_i) = \begin{cases} \lambda^2 v_i^2 y_i e^{-\lambda v_i y_i} & \text{for } y_i \geq 0 \\ 0 & \text{otherwise} \end{cases}$$

Let the stochastic variable Z be defined as

$$Z = \sum_{i=1}^{n} v_i Y_i.$$

Exercise: Using the moment generating function, determine which of the following probability distributions is the correct distribution for Z.

- Chi-squared distribution with $2n$ degrees of freedom.
- **Gamma distribution with** $\alpha = 2n$ **and** $\beta = \frac{1}{\lambda}$
- T-distribution with $2n$ degrees of freedom.
- Gamma distribution with $\alpha = 2n$ and $\beta = \lambda$
- Chi-squared distribution with $4n$ degrees of freedom.
- Chi-squared distribution with $2n/2$ degrees of freedom.
- T-distribution with $2n/2$ degrees of freedom.

10A

From the pdf we can compute the cumulative distribution function for X which is $F_X(x) = x$. We have that:

$$F_Y(y) = P(Y < y) = P(X(1-X) < y) = P(X - X^2 - y < 0) = P(X^2 - X + y > 0)$$

We then need to find the roots of the equation:

$$X^2 - X + y = 0$$

which are

$$X = \frac{1 \pm \sqrt{1-4y}}{2}$$

The inequality of interest is veryfied for

$$X < \frac{1-\sqrt{1-4y}}{2} \text{ or } X > \frac{1+\sqrt{1-4y}}{2}$$

So, coming back to out cumulative distribution function we have that

$$\begin{aligned} F_Y(y) = P(Y < y) = \\ P(X < \frac{1-\sqrt{1-4y}}{2} \text{ or } X > \frac{1+\sqrt{1-4y}}{2}) = \\ P(X < \frac{1-\sqrt{1-4y}}{2}) + P(X > \frac{1+\sqrt{1-4y}}{2}) = \\ \frac{1-\sqrt{1-4y}}{2} + 1 - \frac{1+\sqrt{1-4y}}{2} = \\ 1 - \sqrt{1-4y} \end{aligned}$$

The pdf is found by deriving $F_Y(y)$ wrt y

$$f_Y(y) = \frac{dF_Y(y)}{dy} = \frac{2}{\sqrt{1-4y}}$$

10B

From the pdf we can compute the cumulative distribution function for X which is $F_X(x) = \frac{1}{16}x^2$ We have that:

$$F_Y(y) = P(Y < y) = P(X^2 - 4 < y) = P(X^2 - 4 - y < 0)$$

We then need to find the roots of the equation:

$$X^2 - 4 - y = 0$$

which are

$$X = \sqrt{4+y}$$

Moreover we have that $X > 0$ so the inequality of interest is veryfied for

$$X < \sqrt{4+y}$$

So, coming back to out cumulative distribution function we have that

$$F_Y(y) = P(Y < y) = P(X < \sqrt{4+y}) = F_X(\sqrt{4+y}) = \frac{4+y}{16}$$

The pdf is found by deriving $F_Y(y)$ wrt y

$$f_Y(y) = \frac{dF_Y(y)}{dy} = \frac{1}{16} \text{ for } x \in (-4, 12)$$

10C

From the pdf we can compute the cumulative distribution function for X which is $F_X(x) = \frac{1}{4}x$ We have that:

$$F_Y(y) = P(Y < y) = P(X^2 - 4 < y) = P(X^2 - 4 - y < 0)$$

We then need to find the roots of the equation:

$$X^2 - 4 - y = 0$$

which are

$$X = \pm\sqrt{4+y}$$

Since we know also that $X > 0$, the inequality of interest is verified for

$$X < \sqrt{4+y}$$

So, coming back to out cumulative distribution function we have that

$$F_Y(y) = P(Y < y) = P(X < \sqrt{4+y}) = F_X(\sqrt{4+y}) = \frac{1}{4}\sqrt{4+y}$$

The pdf is found by deriving $F_Y(y)$ wrt y

$$f_Y(y) = \frac{dF_Y(y)}{dy} = \frac{1}{8\sqrt{4+y}}$$

10D

From the pdf we can compute the cumulative distribution function for X which is $F_X(x) = x^2$. We have that:

$$F_Y(y) = P(Y < y) = P(X(1-X) < y) = P(X - X^2 - y < 0) = P(X^2 - X + y > 0)$$

We then need to find the roots of the equation:

$$X^2 - X + y = 0$$

which are

$$X = \frac{1 \pm \sqrt{1-4y}}{2}$$

The inequality of interest is veryfied for

$$X < \frac{1-\sqrt{1-4y}}{2} \text{ or } X > \frac{1+\sqrt{1-4y}}{2}$$

So, coming back to out cumulative distribution function we have that

$$\begin{aligned} F_Y(y) = P(Y < y) = P(X < \frac{1-\sqrt{1-4y}}{2} \text{ or } X > \frac{1+\sqrt{1-4y}}{2}) = \\ = P(X < \frac{1-\sqrt{1-4y}}{2}) + P(X > \frac{1+\sqrt{1-4y}}{2}) = \\ = \left(\frac{1-\sqrt{1-4y}}{2}\right)^2 + 1 - \left(\frac{1+\sqrt{1-4y}}{2}\right)^2 = \\ 1 - \sqrt{1-4y} \end{aligned}$$

The pdf is found by deriving $F_Y(y)$ wrt y

$$f_Y(y) = \frac{dF_Y(y)}{dy} = \frac{2}{\sqrt{1-4y}}$$

11A

We assume that $X_1, X_2, \ldots, X_n$ are iid from the distribution

$$f(x) = \begin{cases} \theta e^{(x-\theta e^x)} & \text{for } x > 0 \\ 0 & \text{otherwise} \end{cases}$$

We derive the likelihood function as

$$\begin{aligned} L(\theta; X_1, \ldots, X_n) &= \prod_{i=1}^{n} f(x_i; \theta) \\ &= \prod_{i=1}^{n} \theta e^{(x_i - \theta e_i^x)} \end{aligned}$$

We then take the log

$$\begin{aligned} l(\theta; X_1, \ldots, X_n) &= \log L(\theta; X_1, \ldots, X_n) \\ &= \sum_{i=1}^{n} [\log \theta + (x_i - \theta e_i^x)] \\ &= n \log \theta + \sum_{i=1}^{n} x_i - \theta \sum_{i=1}^{n} e^{x_i} \end{aligned}$$

To find the MLE we need to set the derivative of $l(\theta; X_1, \ldots, X_n)$ wrt to θ to 0:

$$\begin{aligned} \frac{d\ l(\theta; X_1, \ldots, X_n)}{d\theta} &= 0 \\ \frac{n}{\theta} - \sum_{i=1}^{n} e^{x_i} &= 0 \\ \hat{\theta} &= \frac{n}{\sum_{i=1}^{n} e^{x_i}} \end{aligned}$$

11B

We assume that $X_1, X_2, \ldots, X_n$ are iid from the distribution

$$f(x) = \begin{cases} \frac{\theta}{x} e^{\theta \log x} & \text{for } 0 \leq x \leq 1 \\ 0 & \text{otherwise} \end{cases}$$

We derive the likelihood function as

$$\begin{aligned} L(\theta; X_1, \ldots, X_n) &= \prod_{i=1}^{n} f(x_i; \theta) \\ &= \prod_{i=1}^{n} \frac{\theta}{x_i} e^{\theta \log x_i} \end{aligned}$$

We then take the log

$$\begin{aligned} l(\theta; X_1, \ldots, X_n) &= \log L(\theta; X_1, \ldots, X_n) \\ &= n \log \theta - \sum \log x_i + \theta \sum \log x_i \end{aligned}$$

To find the MLE we need to set the derivative of $l(\theta; X_1, \dots, X_n)$ wrt to θ to 0:

$$\begin{aligned} \frac{d\ l(\theta; X_1, \dots, X_n)}{d\theta} &= 0 \\ \frac{n}{\theta} + \sum \log x_i &= 0 \\ \hat{\theta} &= -\frac{n}{\sum_{i=1}^{n} \log x_i} \end{aligned}$$

11C

We assume that $X_1, X_2, \dots, X_n$ are iid from the distribution

$$f(x) = \begin{cases} \frac{4}{\theta} x^3 e^{-x^4/\theta} & \text{for } x \geq 0 \\ 0 & \text{otherwise} \end{cases}$$

We derive the likelihood function as

$$\begin{aligned} L(\theta; X_1, \dots, X_n) &= \prod_{i=1}^{n} f(x_i; \theta) \\ &= \prod_{i=1}^{n} \frac{4}{\theta} x_i^3 e^{-x_i^4/\theta} \end{aligned}$$

We then take the log

$$\begin{aligned} l(\theta; X_1, \dots, X_n) &= \log L(\theta; X_1, \dots, X_n) \\ &= n \log 4 - n \log \theta + 3 \sum \log x_i - \frac{\sum x_i^4}{\theta} \end{aligned}$$

To find the MLE we need to set the derivative of $l(\theta; X_1, \dots, X_n)$ wrt to θ to 0:

$$\begin{aligned} \frac{d\ l(\theta; X_1, \dots, X_n)}{d\theta} &= 0 \\ -\frac{n}{\theta} + \frac{1}{\theta^2} \sum x_i^4 &= 0 \\ \hat{\theta} &= -\frac{\sum_{i=1}^{n} x_i^4}{n} \end{aligned}$$

11D

We assume that $X_1, X_2, \dots, X_n$ are iid from the distribution

$$f(x) = \begin{cases} \frac{3}{\theta x} (\log x)^2 e^{(\log x)^2/\theta} & \text{for } 0 \leq x \leq 1 \\ 0 & \text{otherwise} \end{cases}$$

We derive the likelihood function as

$$\begin{aligned} L(\theta; X_1, \dots, X_n) &= \prod_{i=1}^{n} f(x_i; \theta) \\ &= \prod_{i=1}^{n} \frac{3}{\theta x_i} (\log x_i)^2 e^{(\log x_i)^2/\theta} \end{aligned}$$

We then take the log

$$\begin{aligned} l(\theta; X_1, \ldots, X_n) &= \log L(\theta; X_1, \ldots, X_n) \\ &= n \log 3 - n \log \theta - \sum \log x_i + 2 \sum \log(\log x_i) + \frac{1}{\theta} \sum (\log x_i)^3 \end{aligned}$$

To find the MLE we need to set the derivative of $l(\theta; X_1, \ldots, X_n)$ wrt to θ to 0:

$$\begin{aligned} \frac{d\ l(\theta; X_1, \ldots, X_n)}{d\theta} &= 0 \\ -\frac{n}{\theta} - \frac{1}{\theta^2} \sum (\log x_i)^3 &= 0 \\ \hat{\theta} &= -\frac{\sum (\log x_i)^3}{n} \end{aligned}$$

12A

Let $Y_i, i = 1, 2$ be two discrete stochastic variables with distribution

$$P(Y_i = y_i) = \frac{(t_i\lambda)^{y_i}}{y_i!}\exp(-t_i\lambda) \text{ for } y_i = 0, 1, 2\ldots,$$

where $t_1 = 2$ and $t_2 = 5$.

We re given two estimators

$$\widehat{\lambda} = \frac{t_1Y_1 + t_2Y_2}{t_1^2 + t_2^2} \text{ and } \widetilde{\lambda} = \frac{Y_1 + Y_2}{t_1 + t_2}.$$

We need to find the mean and the variance.

We start with $\widehat{\lambda}$

$$E(\widehat{\lambda}) = \frac{t_1E(Y_1) + t_2E(Y_2)}{t_1^2 + t_2^2} = \frac{22\lambda + 55\lambda}{4 + 25} = \lambda$$

$$\text{Var}(\widehat{\lambda}) = \frac{t_1^2\text{Var}(Y_1) + t_2^2\text{Var}(Y_2)}{(t_1^2 + t_2^2)^2} = \frac{4(2\lambda) + 25(5\lambda)}{29^2} = \frac{133}{29^2} = 0.158\lambda$$

Then $\widetilde{\lambda}$:

$$E(\widetilde{\lambda}) = \frac{E(Y_1) + E(Y_2)}{t_1 + t_2} = \frac{2\lambda + 5\lambda}{2 + 5} = \lambda$$

$$\text{Var}(\widetilde{\lambda}) = \frac{\text{Var}(Y_1) + \text{Var}(Y_2)}{(t_1 + t_2)^2} = \frac{7\lambda}{49} = 0.142\lambda$$

Both $\widehat{\lambda}$ and $\widetilde{\lambda}$ are unbiased. $\widetilde{\lambda}$ has smaller variance and therefore it is to be preferred.

12B

Let X and Y be two discrete stochastic variables with distribution respectively:

$$f_X(x;\lambda) = \begin{cases} \frac{1}{\lambda^2}x\exp(-x/\lambda) & \text{for } x > 0, \\ 0 & \text{otherwise,} \end{cases} \quad f_Y(y;\lambda) = \begin{cases} \frac{1}{4\lambda^2}y\exp(-y/2\lambda) & \text{for } y > 0, \\ 0 & \text{otherwise,} \end{cases}$$

We are given two estimators

$$\widehat{\lambda} = \frac{X}{2}, \text{ and } \widetilde{\lambda} = \frac{1}{2}(\frac{X}{2} + \frac{Y}{4})$$

We need to find the mean and the variance.

We start with $\widehat{\lambda}$

$$E(\widehat{\lambda}) = \frac{E(X)}{2} = \frac{2\lambda}{2} = \lambda$$

$$\text{Var}(\widehat{\lambda}) = \frac{\text{Var}(X)}{4} = \frac{1}{2}\lambda^2$$

Then $\widetilde{\lambda}$:

$$E(\widetilde{\lambda}) = \frac{1}{2}\left(\frac{E(X)}{2} + \frac{E(Y)}{4}\right) = \frac{1}{2}\left(\frac{2\lambda}{2} + \frac{4\lambda}{4}\right) = \lambda$$

$$\text{Var}(\widetilde{\lambda}) = \frac{1}{4}\left(\frac{\text{Var}(X)}{4} + \frac{\text{Var}(Y)}{16}\right) = \frac{1}{4}\left(\frac{2\lambda^2}{4} + \frac{8\lambda^2}{16}\right) = \frac{1}{4}\lambda^2$$

Both $\widehat{\lambda}$ and $\widetilde{\lambda}$ are unbiased. $\widetilde{\lambda}$ has smaller variance and therefore it is to be preferred.

13A

We have that $Y_i \sim N(\alpha x_i^2, \sigma^2 x_i)$, $i = 1, \ldots, n$ and that the MLE is

$$\hat{\lambda} = \frac{\sum x_i Y_i}{\sum x_i^3}$$

We want to find a 95% confidence interval for α.

We have that

$$E(\hat{\alpha}) = \sum \frac{x_i}{x_i^3} E(Y_i) = \sum \frac{x_i}{x_i^3} \alpha x_i^2 = \alpha$$

$$\text{Var}(\hat{\alpha}) = \left(\frac{1}{\sum x_i^3}\right)^2 \sum x_i^2 \text{Var}(Y_i) = \sum \left(\frac{1}{\sum x_i^3}\right)^2 \sum x_i^2 \sigma^2 x_i = \frac{\sigma^2}{\sum x_i^3}$$

Moreover, $\hat{\alpha}$ is normally distributed since it is a linear combination of normally distributed RV. We have then

$$Z = \frac{\hat{\alpha} - \alpha}{\sqrt{\frac{\sigma^2}{\sum x_i^3}}} \sim N(0, 1)$$

This we can use to set up a 95% confidence interval for α as

$$P(-z_{0.025} < Z < z_{0.025}) = 0.95$$

$$P(-1.96 < \frac{\hat{\alpha} - \alpha}{\sqrt{\frac{\sigma^2}{\sum x_i^3}}} < 1.96) = 0.95$$

$$P(\hat{\alpha} - \frac{1.96\sigma}{\sqrt{\sum x_i^3}} < \alpha < \hat{\alpha} + \frac{1.96\sigma}{\sqrt{\sum x_i^3}}) = 0.95$$

13B

We have that $Y_i \sim N(\beta \log x_i, \sigma^2 x_i^2)$, $i = 1, \ldots, n$ and that the MLE is

$$\hat{\beta} = \frac{\sum Y_i \log x_i / x_i^2}{\sum (\log x_i)^2 / x_i^2}$$

We want to find a 95% confidence interval for β.

We have that

$$E(\hat{\beta}) = \frac{\sum E(Y_i) \log x_i / x_i^2}{\sum (\log x_i)^2 / x_i^2} = \frac{\sum \beta \log x_i \log x_i / x_i^2}{\sum (\log x_i)^2 / x_i^2} = \beta$$

$$\text{Var}(\hat{\beta}) = \frac{\sigma^2}{\sum (\log x_i)^2 / x_i^2}$$

Moreover, $\hat{\beta}$ is normally distributed since it is a linear combination of normally distributed RV. We have then

$$Z = \frac{\hat{\beta} - \beta}{\sqrt{\frac{\sigma^2}{\sum (\log x_i)^2 / x_i^2}}} \sim N(0, 1)$$

This we can use to set up a 95% confidence interval for α as

$$P(-z_{0.025} < Z < z_{0.025}) = 0.95$$
$$P(-1.96 < \frac{\hat{\beta} - \beta}{\sqrt{\frac{\sigma^2}{\sum(\log x_i)^2/x_i^2}}} < 1.96) = 0.95$$
$$P(\hat{\beta} - \frac{1.96\sigma}{\sqrt{\sum(\log x_i)^2/x_i^2}} < \beta < \hat{\beta} + \frac{1.96\sigma}{\sqrt{\sum(\log x_i)^2/x_i^2}}) = 0.95$$

13C

We have that $Y_i \sim N(\theta x_i(1-x_i), \sigma^2 x_i)$, $i = 1, \ldots, n$ and that the MLE is

$$\hat{\theta} = \frac{\sum Y_i(1-x_i)}{\sum x_i(1-x_i)^2}$$

We want to find a 95% confidence interval for β.

We have that

$$E(\hat{\theta}) = \frac{\sum E(Y_i)(1-x_i)}{\sum x_i(1-x_i)^2} = \frac{\sum \theta x_i(1-x_i)^2}{\sum x_i(1-x_i)} = \theta$$
$$\text{Var}(\hat{\theta}) = \frac{\sum \text{Var}(Y_i)(1-x_i)^2}{(\sum x_i(1-x_i)^2)^2} = \frac{\sigma^2}{\sum x_i(1-x_i)^2}$$

Moreover, $\hat{\theta}$ is normally distributed since it is a linear combination of normally distributed RV. We have then

$$Z = \frac{\hat{\theta} - \theta}{\sqrt{\frac{\sigma^2}{\sum x_i(1-x_i)^2}}} \sim N(0,1)$$

This we can use to set up a 95% confidence interval for α as

$$P(-z_{0.025} < Z < z_{0.025}) = 0.95$$
$$P(-1.96 < \frac{\hat{\theta} - \theta}{\sqrt{\frac{\sigma^2}{\sum x_i(1-x_i)^2}}} < 1.96) = 0.95$$
$$P(\hat{\theta} - \frac{1.96\sigma}{\sqrt{\sum x_i(1-x_i)^2}} < \theta < \hat{\theta} + \frac{1.96\sigma}{\sqrt{\sum x_i(1-x_i)^2}}) = 0.95$$

14A

Introduction: A producer of washing machines claims that the average lifespan, μ, of his washing machines is 5 years. A group of clients suspects that this is not true and that, in fact, the lifespan is shorter than what the producer claims.

Exercise: Which null hypotheses , H_0, and alternativ hypotheses, H_1, should the client use in this situation?

- $H_0 \neq 5$ og $H_1 = 5$
- $H_0 = 5$ og $H_1 \neq 5$
- $H_0 = 5$ og $H_1 < 5$ - correct
- $H_0 = 5$ og $H_1 > 5$
- $H_0 < 5$ og $H_1 = 5$
- $H_0 > 5$ og $H_1 = 5$

14B

Introduction: We know that the average weight of foxes in Trøndelag has been $\mu = 5$ kg. We suspect that, lately, the average weight has changed.

Exercise: Which null hypotheses ,H_0 , and alternativ hypotheses, H_1, should one use in this situation?

- $H_0 \neq 5$ og $H_1 = 5$
- $H_0 = 5$ **og** $H_1 \neq 5$ - correct
- $H_0 = 5$ og $H_1 < 5$
- $H_0 = 5$ og $H_1 > 5$
- $H_0 < 5$ og $H_1 = 5$
- $H_0 > 5$ og $H_1 = 5$

14C

Introduction: Ola grows tomatoes. In recent years, he has picked on average $\mu = 5$ kg of tomatoes per year from his garden. He has been on a cultivation course this year and he now claims that his production has increased.

Exercise: Which null hypotheses ,H_0 , and alternativ hypotheses, H_1, should one use in this situation?

- $H_0 \neq 5$ og $H_1 = 5$
- $H_0 = 5$ og $H_1 \neq 5$
- $H_0 = 5$ og $H_1 < 5$
- $H_0 = 5$ **og** $H_1 > 5$ - correct
- $H_0 < 5$ og $H_1 = 5$
- $H_0 > 5$ og $H_1 = 5$

15A

We have taht $X_1, \dots, X_n$ are iid with distribution

$$f(x_i|\theta) = \begin{cases} \frac{1}{2\theta^3} x_i^2 e^{-\frac{x_i}{\theta}} & \text{for } x \geq 0 \\ 0 & \text{ellers} \end{cases}$$

The MLE for θ is

$$\hat{\theta} = \frac{1}{3n} \sum X_i$$

We want to test

$$H_0 : \theta = 1$$
$$H_1 : \theta < 1$$

using a significance level $\alpha = 0.1$

Exercise A.

We have that

$$E(\hat{\theta}) = \theta$$
$$\text{Var}(\hat{\theta}) = \frac{1}{9n^2} \sum \text{Var}(X_i) = \frac{3n\theta^2}{9n^2} = \frac{\theta^2}{3n}$$

Moreover, since n is large we can rely on the central limit theorem. This implies that

$$Z = \frac{\hat{\theta} - \theta}{\frac{\theta}{\sqrt{3n}}} \approx N(0, 1)$$

Under H_0 we have that

$$Z = \frac{\hat{\theta} - 1}{\frac{1}{\sqrt{3n}}} \approx N(0, 1)$$

So the decision rule is defined as

$$\begin{aligned} 0.1 &= P(\text{Reject } H_0|H_0) \\ &= P(Z < z_\alpha|H_0) \\ &= P(\frac{\hat{\theta} - 1}{\frac{1}{\sqrt{3n}}} < z_\alpha|H_0) \\ &= P(\hat{\theta} < -\frac{-1.28}{\sqrt{3n}} + 1) \end{aligned}$$

Exercise B

We want the power of our test, when the true value of θ is 0.9, to be at least 0.85

$$\begin{aligned} 0.85 \leq P(\text{Reject} H_0|H_1) &= P(\hat{\theta} < \frac{-1.28}{\sqrt{3n}} + 1|\theta = 0.9) \\ &= P\left(\frac{\hat{\theta} - 0.9}{\frac{0.9}{\sqrt{3n}}} < \frac{\frac{-1.28}{\sqrt{3n}} + 1 - 0.9}{\frac{0.9}{\sqrt{3n}}}|\theta = 0.9\right) \\ &= P(Z < \frac{\frac{-1.28}{\sqrt{3n}} + 0.1}{\frac{0.9}{\sqrt{3n}}}|\theta = 0.9) \end{aligned}$$

We need therefore

$$\frac{\frac{-1.28}{\sqrt{3n}}+0.1}{\frac{0.9}{\sqrt{3n}}} \geq z_{0.15}$$
$$\frac{-1.28}{0.9}+0.1\frac{\sqrt{3n}}{0.9} \geq 1.036$$
$$3n \geq 22.12^2$$
$$n \geq 163.09$$

15B

We have taht $X_1,\ldots,X_n$ are iid with distribution

$$f(x_i;\theta)=\frac{\theta^{x_i}}{x_i!}e^{-\theta} \text{ for } x_i=0,1,2,\ldots,$$

The MLE for θ is

$$\widehat{\theta}=\frac{1}{n}\sum_{i=1}^{n}X_i.$$

We want to test

$$H_0:\theta=1$$
$$H_1:\theta>1$$

using a significance level $\alpha=0.05$

Exercise A.

We have that

$$E(\hat{\theta})=\theta$$
$$\text{Var}(\hat{\theta})=\frac{1}{n^2}\sum_{i=1}^{n}\text{Var}(X_i)=\frac{\theta}{n}$$

Moreover, since n is large we can rely on the central limit theorem. This implies that

$$Z=\frac{\hat{\theta}-\theta}{\sqrt{\frac{\theta}{n}}}\approx N(0,1)$$

Under H_0 we have that

$$Z=\frac{\hat{\theta}-1}{\sqrt{\frac{1}{n}}}\approx N(0,1)$$

So the decision rule is defined as

$$\begin{aligned}0.05 &= P(\text{Reject } H_0|H_0)\\ &= P(Z>z_\alpha|H_0)\\ &= P(\frac{\hat{\theta}-1}{\sqrt{\frac{1}{n}}}>z_\alpha|H_0)\\ &= P(\hat{\theta}>\frac{1.645}{\sqrt{n}}+1)\end{aligned}$$

Exercise B

We want the power of our test, when the true value of θ is 1.05, to be at least 0.8

$$\begin{aligned}
0.8 \leq P(\text{Reject} H_0|H_1) &= P(\hat{\theta} > \frac{1.645}{\sqrt{n}} + 1|\theta = 1.05) \\
&= P(\frac{\hat{\theta} - 1.05}{\sqrt{\frac{1.05}{n}}} > \frac{\frac{1.645}{\sqrt{n}} + 1 - 1.05}{\sqrt{\frac{1.05}{n}}}|\theta = 1.05) \\
&= P(Z > \frac{\frac{1.645}{\sqrt{n}} - 0.05}{\sqrt{\frac{1.05}{n}}}|\theta = 1.05)
\end{aligned}$$

We need therefore

$$\begin{aligned}
\frac{\frac{1.645}{\sqrt{n}} - 0.05}{\sqrt{\frac{1.05}{n}}} &< z_{0.8} \\
\frac{1.645}{\sqrt{1.05}} - 0.05\frac{\sqrt{n}}{\sqrt{1.05}} &< -0.84 \\
n &> 2511.5
\end{aligned}$$

15C

We have that $X_1, \dots, X_n$ are iid with distribution

$$f(x_i|\theta) = f(x_i;\theta) = \frac{1}{\theta}\left(1 - \frac{1}{\theta}\right)^{x_i} \quad \text{for} \quad x_i = 0, 1, 2, \dots,$$

The MLE for θ is

$$\widehat{\theta} = 1 + \frac{1}{n}\sum_{i=1}^{n} X_i.$$

We want to test

$$\begin{aligned}
H_0 &: \theta = 2 \\
H_1 &: \theta > 2
\end{aligned}$$

using a significance level $\alpha = 0.05$

Exercise A.

We have that

$$\begin{aligned}
E(\hat{\theta}) &= \theta \\
\text{Var}(\hat{\theta}) &= \frac{\theta(\theta - 1)}{n}
\end{aligned}$$

Moreover, since n is large we can rely on the central limit theorem. This implies that

$$Z = \frac{\hat{\theta} - \theta}{\sqrt{\frac{\theta(\theta-1)}{n}}} \approx N(0, 1)$$

Under H_0 we have that

$$Z = \frac{\hat{\theta} - 2}{\sqrt{\frac{2}{n}}} \approx N(0, 1)$$

So the decision rule is defined as

$$
\begin{aligned}
0.05 &= P(\text{Reject } H_0|H_0) \\
&= P(Z > z_\alpha|H_0) \\
&= P(\frac{\hat{\theta}-2}{\sqrt{\frac{2}{n}}} > z_\alpha|H_0) \\
&= P(\hat{\theta} > 1.645\sqrt{\frac{2}{n}} + 2)
\end{aligned}
$$

Exercise B

We want the power of our test, when the true value of θ is 0.9, to be at least 2.05

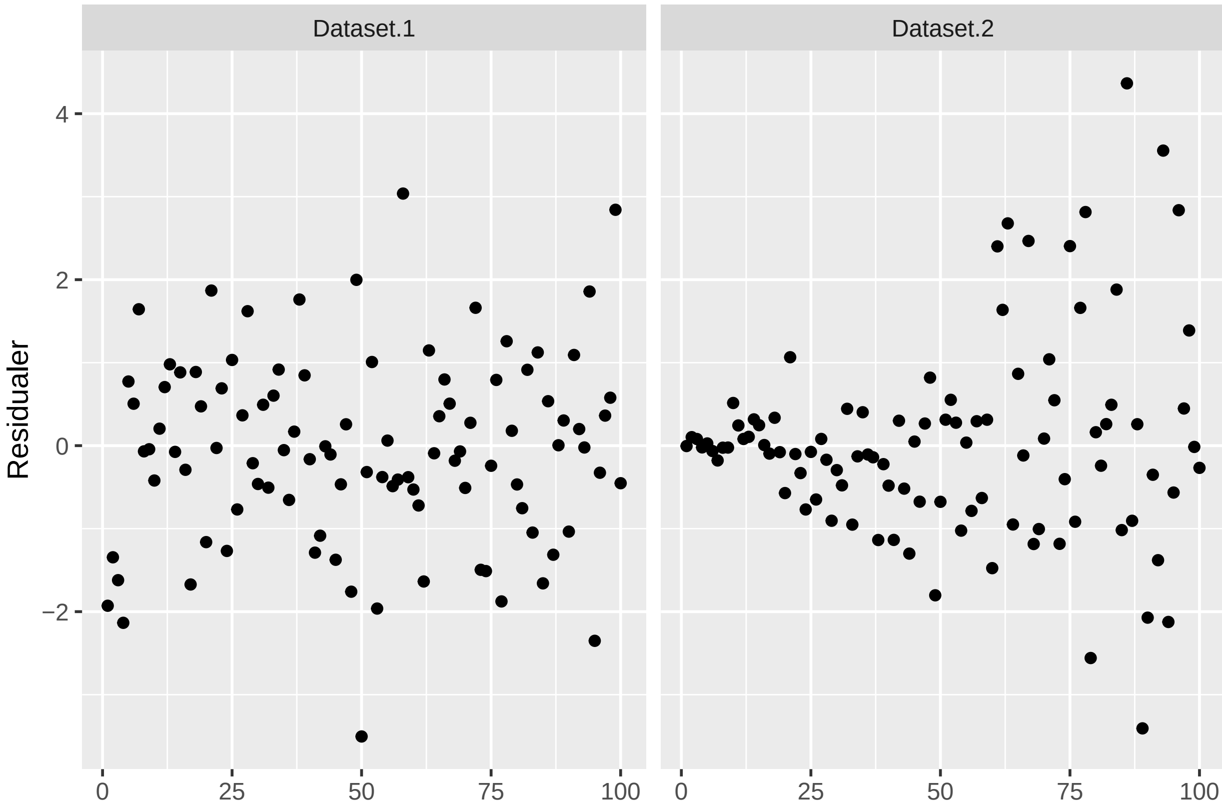
$$
\begin{aligned}
0.9 \leq P(\text{Reject } H_0|H_1) &= P(\hat{\theta} > 1.645\sqrt{\frac{2}{n}} + 2|\theta = 2.05) \\
&= P\left(\frac{\hat{\theta}-2.05}{\sqrt{\frac{2.1525}{n}}} > \frac{1.645\sqrt{\frac{2}{n}} + 2 - 2.05}{\sqrt{\frac{2.1525}{n}}}|\theta = 2.05\right) \\
&= P(Z > 1.645\sqrt{\frac{2}{2.1525}} - 0.05\sqrt{\frac{n}{2.1525}}\bigg|\theta = 2.05)
\end{aligned}
$$

We need therefore

$$
\begin{aligned}
1.645\sqrt{\frac{2}{2.1525}} - 0.05\sqrt{\frac{n}{2.1525}} &< z_{0.9} \\
1.645\sqrt{\frac{2}{2.1525}} - 0.05\sqrt{\frac{n}{2.1525}} &< -1.28 \\
n &> 2.1525\left(\frac{1}{0.05}\left(1.645\sqrt{\frac{2}{2.1525}} + 1.28\right)\right)^2 \\
n &> 7070.52
\end{aligned}
$$

16A

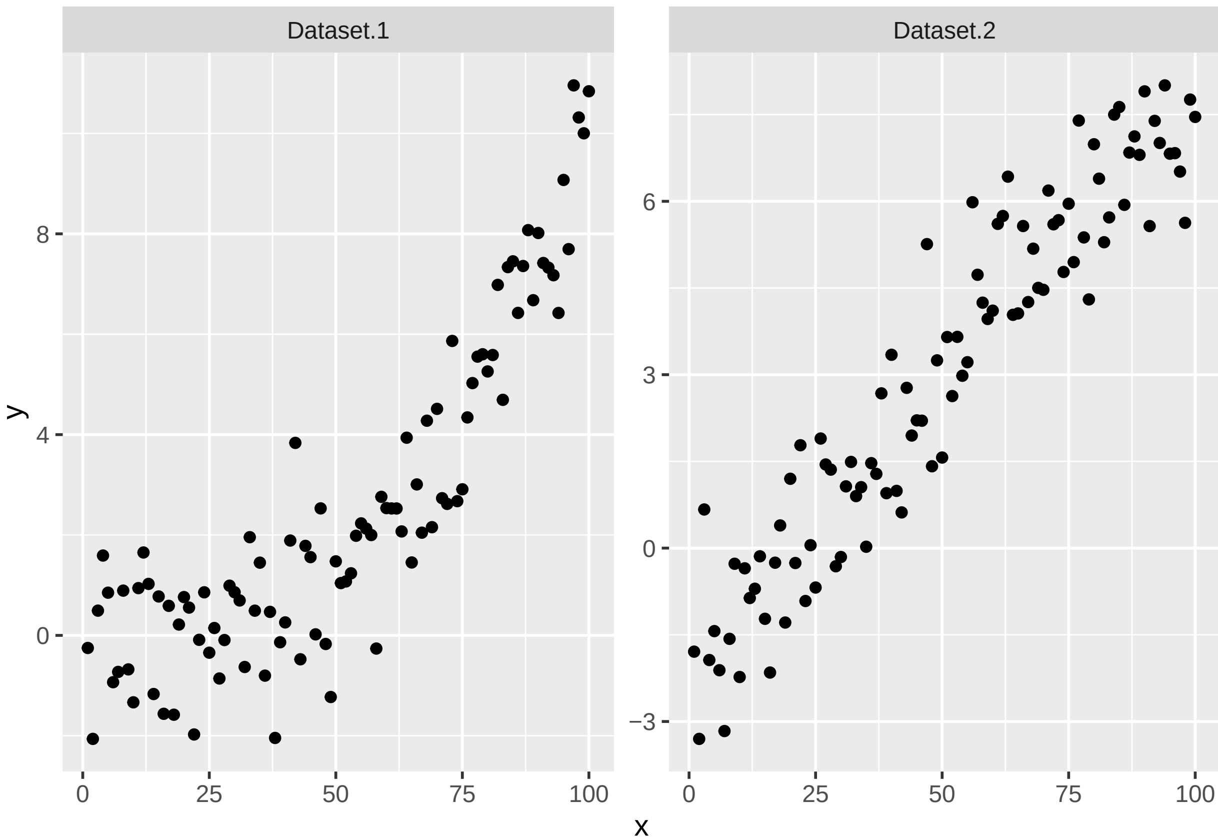
The residual plot is



While the plot related to dataset1 does not show any visible pattern, the residual plot for dataset 2 clearly shows that the variance is not constant wrt to x, something that is in contrast with the assumptions behind the linear model.

16B

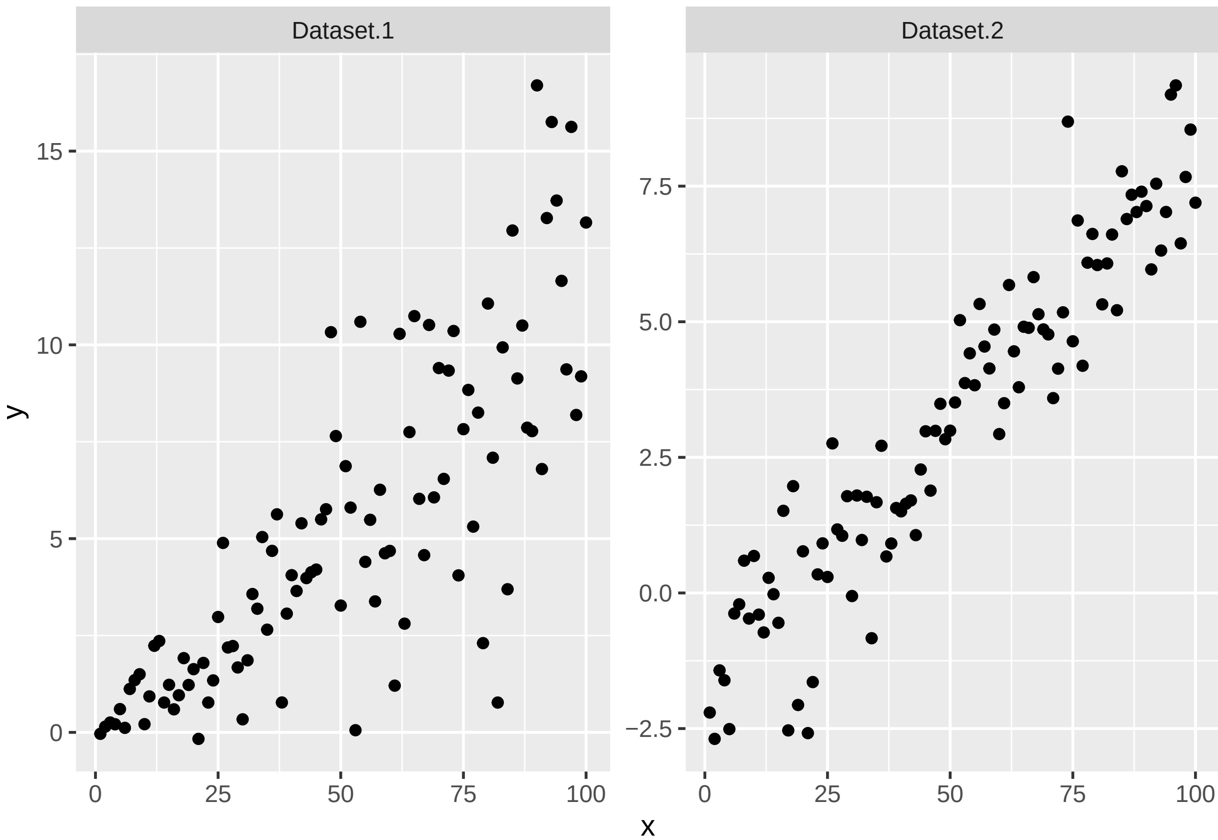
The scatterplot is



The scatterplot related to dataset 1 shows a non linear relatioship between x and y. The simple linear model assumes, on the contrary, a linear relatioship between the two variables. The scatterplot for dataset 2 seems to respect such assumption.

16 C

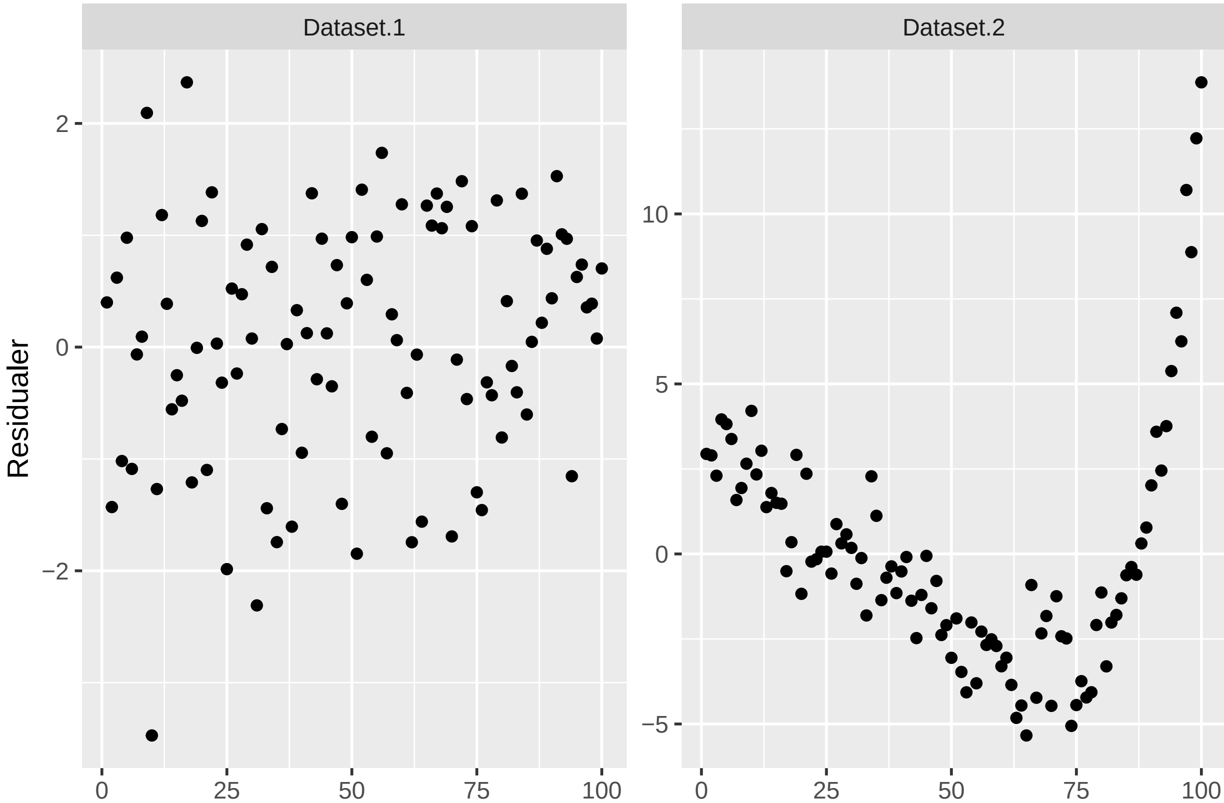
The scatter plot is



Both scatterplot could agree with the assumption of a linear relationship between x and Y. The scatterplot related to dataset 1 shows a clear tendency for the variance of $Y_i|x_i$ to increase with the value of x_i, this is contrast with the assumption of the linear model that assumes constant variance for the error term. This asssumption seems to be respected by the dataset 2.

16 D

The residual is



While the residuals for dataset1 show no visible pattern, they are centered around 0 and appear to have constant variance, the residuals for dataset2 show a clear pattern that points to a non-linear relatioship between the variables x and Y.