Ministerie van Defensie

Faculteit Militaire Wetenschappen

Gegevens student	
Naam:	
Peoplesoftnummer:	
Klas:	
Handtekening:	

Algemeen			
Vak:	Probability & Statistics (resit)	**Vakcode:**	P&S
Datum:	4 juli 2025	**Tijdsduur:**	10:00-13:00
Examinator:	Dr. ir. D.A.M.P. Blom	**Aantal pagina's:**	6
Peer-review:		**Aantal opgaven:**	6

Algemene instructies

- Alle antwoorden dienen gemotiveerd te worden. Indien u een deelopgave niet kunt oplossen en het antwoord in vervolgvragen nodig hebt, probeer uit te gaan van een redelijke fictieve waarde.
- U mag een grafische rekenmachine gebruiken zonder CAS (Computer Algebra Systeem).
- Antwoorden, in welke vorm dan ook, mogen de zaal niet verlaten.
- Vermeld op elk antwoordvel je naam, Peoplesoft-nummer en maak een nummering van je antwoordvellen.
- Iedere vorm van mobiele (potentiële) datadragers (telefoon, smartwatch, etc) of andere vormen om te frauderen (bv. communicatieapparatuur) zijn niet toegestaan gedurende de gehele duur van het tentamen en mogen ook niet in het lokaal meegebracht worden of zijn uitgeschakeld en ingeleverd.
- Schrijf leesbaar ter voorkoming van misverstanden bij de beoordeling van uw werk. Indien uw antwoord niet leesbaar is, wordt uw antwoord fout gerekend.
- Toiletbezoek tijdens het tentamen vindt enkel plaats na toestemming van de examinator.
- Lever bij het verlaten van de zaal, kladpapier, tentamenopgaven en andere tentamen-gerelateerde documenten in bij de examinator.

Cijferberekening / cesuur

- Het eindcijfer voor het vak Statistiek wordt voor 50% bepaald door dit tentamen.
- Het tentamen is opgebouwd uit 4 open vragen. Bij iedere (sub)vraag is het aantal te behalen punten tussen haakjes aangegeven. In totaal kunt u 100 punten verdienen.
- Het tentamencijfer wordt bepaald door het totaal aantal punten te delen door 10. Het tentamencijfer moet minimaal een 5,0 zijn om de cursus Statistiek met succes af te ronden.

Procedure na het tentamen

- De cijfers van dit tentamenonderdeel worden in principe binnen 10 werkdagen na de afname bekend gemaakt.
- Met vragen over de beoordeling kunt u tot 10 werkdagen na bekendmaking van de cijfers terecht bij de cursuscoördinator.

Veel succes!

Problem 1 (15 points) In the figure below, we are given a circuit consisting of four devices (blue squares) from left to right. This circuit operates only if there is a path of functional devices from left to right. The probability that each device is functional is given in each blue square.

1a [8pt] Assume that device failures are independent of each other. What is the probability that the circuit operates?

Solution

Let D_i define the event that device i is functional (hence $\overline{D_i}$ is the complement, i.e., the event that device i fails). Let S define the event that the system operates. In order for the circuit to operate, the following two conditions need to hold:

1. At least one of the following holds: Device 1 is functional, Device 2 is functional or Device 3 and 4 are functional.
2. Device 5 is functional (3pt)

In mathematical terms, we need to compute $P((D_1 \cup D_2 \cup (D_3 \cap D_4)) \cap D_5)$ (here, the symbol $\cup$ denotes ör", whereas $\cap$ denotes änd"). Since devices can fail independently of each other, we have

$$\begin{aligned}
P(S) &= P\left((D_1 \cup D_2 \cup (D_3 \cap D_4)) \cap D_5\right) \\
&= P\left(D_1 \cup D_2 \cup (D_3 \cap D_4)\right) \cdot P(D_5) \\
&= \left(1 - P(\overline{D_1} \cap \overline{D_2} \cap \overline{(D_3 \cap D_4)})\right) \cdot P(D_5) \\
&= (1 - P(\overline{D_1}) \cdot P(\overline{D_2}) \cdot P(\overline{D_3 \cap D_4})) \cdot P(D_5) \\
&= (1 - 0.03 \cdot 0.07 \cdot (1 - 0.95 \cdot 0.91)) \cdot 0.85 \\
&\approx 0.8498
\end{aligned}$$

(5pt)

1b [7pt] Using Bayes' Rule, determine the probability that Device 5 fails, given that the system of devices has failed.

Solution

From our answers for question 1a, we can derive the following probabilities:

- $P(S) = 0.8498 \rightarrow P(\overline{S}) = 1 - P(S) = 0.1502$ (1pt)
- $P(D_5) = 0.85 \rightarrow P(\overline{D_5}) = 1 - P(D_5) = 0.15$ (1pt)
- $P(S|\overline{D_5}) = 0$ (if device 5 fails, the system fails) $\rightarrow P(\overline{S}|\overline{D_5}) = 1$ (1pt)

Using Bayes' Rule, we can compute $P(\overline{D_5}|\overline{S})$:

$$P(\overline{D_5}|\overline{S}) = P(\overline{S}|\overline{D_5}) \cdot \frac{P(\overline{D_5})}{P(\overline{S})} = 1 \cdot \frac{0.15}{0.1502} \approx 0.9987$$

(4pt)

Problem 2 (15 points) A radar system tracks the position of aircraft in a two-dimensional space (on a pre-specified altitude), with the coordinate (X, Y) representing the aircraft's position in kilometers relative to an air base. Assume the joint probability density function (PDF) of (X, Y) is given by

$$f_{X,Y}(x, y) = \begin{cases} C(x + y), & \text{if } 0 \leq x \leq 2, 0 \leq x \leq 2 \\ 0, & \text{otherwise,} \end{cases}$$

where C is a constant.

2a [6pt] Determine the value of C such that $f_{X,Y}(x, y)$ satisfies the properties of a valid joint probability density function.

Solution

The joint probability density function $f_{X,Y}(x, y)$ is a valid joint probability density function if the following two conditions hold

1. First condition: $f_{X,Y}(x, y) \geq 0$ for all possible values of x and y, and (1pt)
2. Second condition: $\int_{-\infty}^{\infty} \int_{-\infty}^{\infty} f_{X,Y}(x, y)\, dx\, dy = 1$. (1pt)

The first condition holds if and only if $C(x + y)$ for $0 \leq x \leq 2$ and $0 \leq y \leq 2$, so we need $C \geq 0$. The second condition holds if

$$1 = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} f_{X,Y}(x, y)\, dx\, dy = \int_0^2 \int_0^2 C(x + y) dx dy = C \int_0^2 \left[\int_0^2 (x + y) dx \right] dy$$

$$= C \int_0^2 \left[\frac{1}{2} x^2 + xy \right]_0^2 dy = C \int_0^2 2 + 2y dy = C \left[2y + y^2 \right]_0^2 = C(4 + 4) = 8C$$

$$\Rightarrow C = \frac{1}{8}$$

(4pt)

2b [5pt] Compute the marginal PDFs $f_X(x)$ and $f_Y(y)$, for X and Y respectively.

Solution

– The marginal PDF of X is given by

$$\begin{aligned} f_X(x) = \int_{-\infty}^{\infty} f_{X,Y} dy &= \int_0^2 \frac{1}{8}(x+y)dy \\ &= [\frac{1}{8}xy + \frac{1}{16}y^2]_0^2 \\ &= \frac{1}{4}x + \frac{1}{4} \end{aligned}$$

(3pt)

– By symmetry of the joint probability density function, the marginal PDF of y is given by $f_Y(y) = \frac{1}{4}y + \frac{1}{4}$. (2pt)

2c [4pt] Are X and Y independent random variables? Justify your answer with a computation.

Solution

To show whether or not X and Y are independent random variables, we need to compute their covariance and check if it is equal to zero (independence) or not (not independence). We can compute the covariance as follows:

$$Cov(X,Y) = E[XY] - E[X]E[Y].$$

(1pt)

We first need to compute these quantities:

$$\begin{aligned} E[X] = \int_{-\infty}^{\infty} x f_X(x) dx &= \int_0^2 \frac{1}{4}x^2 + \frac{1}{4}x dx \\ &= [\frac{1}{12}x^3 + \frac{1}{8}x^2]_0^2 = \frac{7}{6} \end{aligned}$$

$$E[Y] = \frac{7}{6} \text{ (by symmetry of } X \text{ and } Y.)$$

$$
\begin{aligned}
E[XY] = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} xy f_{X,Y}(x,y)dx &= \int_0^2\int_0^2 \left(\frac{1}{8}x^2y + \frac{1}{8}xy^2\right)\, dx\, dy \\
&= \int_0^2 [\frac{1}{24}x^3y + \frac{1}{16}x^2y^2]_0^2 dy \\
&= \int_0^2 \frac{1}{3}y + \frac{1}{4}y^2 dy \\
&= [\frac{1}{6}y^2 + \frac{1}{12}y^3]_0^2 = \frac{4}{3}
\end{aligned}
$$

(2pt)

The covariance of X and Y is equal to $Cov(X,Y) = E[XY] - E[X]E[Y] = \frac{4}{3} - \frac{7}{6} \cdot \frac{7}{6} = -\frac{1}{36} \neq 0$, thus the random variables X and Y are not independent. (1pt)

Problem 3 (20 points) A set of n independent and identically distributed random variables $X_1, X_2, \ldots, X_n$ are drawn from a distribution with the probability density function:

$$f(x;\theta) = \frac{1}{\theta} e^{-\frac{x}{\theta}}, \quad x \geq 0, \quad \theta > 0$$

3a [3pt] Write down the likelihood function $L(x_1, x_2, \ldots, x_n; \theta)$ given a sample of realizations of $X_1, X_2, \ldots, X_n$.

Solution

The likelihood function can be written as follows:

$$\begin{aligned} L(x_1, x_2, \ldots, x_n; \theta) &= f(x_1;\theta) \cdot f(x_2;\theta) \ldots \cdot f(x_n;\theta) \\ &= \frac{1}{\theta} e^{-\frac{x_1}{\theta}} \cdot \frac{1}{\theta} e^{-\frac{x_2}{\theta}} \ldots \cdot \frac{1}{\theta} e^{-\frac{x_n}{\theta}} \\ &= (\frac{1}{\theta})^n e^{-\frac{\sum_{i=1}^n x_i}{\theta}} \end{aligned}$$

(3pt)

3b [7pt] Derive the maximum likelihood estimator (MLE) of the parameter θ.

Solution

The maximum likelihood estimator can be obtained by optimizing the likelihood function, i.e., taking the derivative of the likelihood function and setting it equal to zero:

$$\begin{aligned} \frac{dL(x_1, x_2, \cdot x_n; \theta)}{d\theta} &= -n \cdot (\frac{1}{\theta})^{n+1} \cdot e^{-\frac{\sum_{i=1}^n x_i}{\theta}} + (\frac{1}{\theta})^n \cdot \frac{\sum_{i=1}^n}{\theta^2} \cdot e^{-\frac{\sum_{i=1}^n x_i}{\theta}} \\ &= e^{-\frac{\sum_{i=1}^n x_i}{\theta}} \cdot (\frac{1}{\theta})^{n+2} \cdot \left(-n \cdot \theta + \sum_{i=1}^n x_i \right) = 0 \end{aligned}$$

(4pt)

Since the part outside of the brackets is strictly positive, it necessarily holds that

$$-n \cdot \theta + \sum_{i=1}^n x_i = 0 \Rightarrow \theta = \frac{\sum_{i=1}^n x_i}{n}$$

Hence, the maximum likelihood estimator of the parameter θ is equal to the sample mean $\frac{\sum_{i=1}^{n} x_i}{n}$. (2pt)

(1pt)

3c [5pt] Show that the sample mean $\overline{X} = \frac{(X_1+X_2+\ldots+X_n)}{n}$ is an unbiased estimator for the parameter θ.

Solution

We need to show that $E[\frac{\sum_{i=1}^{n} X_i}{n}] = \theta$. Notice that the given probability density function is equal to the density function of a exponential distribution with parameter $\lambda = \frac{1}{\theta}$. (2pt)

This means that each X_i has an expected value of $E[X_i] = \frac{1}{\lambda} = \theta$. (1pt)

Therefore, we have that

$$E[\frac{\sum_{i=1}^{n} X_i}{n}] = \frac{1}{n}\sum_{i=1}^{n} E[X_i] = \frac{1}{n}(\theta + \theta + \ldots + \theta) = \frac{1}{n} \cdot n\theta = \theta,$$

which finishes the proof. (2pt)

3d [5pt] Determine whether or not the maximum likelihood estimator obtained in (3a) is also an unbiased estimator for θ.

Solution

Notice that the maximum likelihood estimator we obtained earlier is the sample mean, and we showed in (3c) that it is an unbiased estimator.

Problem 4 (15 points) A defense contractor is testing the mean lifespan of a new type of battery. From a sample of $n = 12$ batteries, the following measurements (in hours) are found:

$$98, 103, 110, 87, 92, 108, 81, 123, 130, 90, 99, 101$$

Furthermore, the battery lifespan is assumed to follow a normal distribution.

4a [3pt] Determine the mean and standard deviation of this sample.

Solution

We can compute the sample mean and the sample variance as follows:

$$\overline{x} = \frac{x_1 + x_2 + \ldots + x_n}{n} = \frac{98 + 103 + \ldots + 101}{12} \approx 101.8333$$

(1pt)

$$\begin{aligned} s^2 &= \frac{1}{n-1} \sum_{i=1}^{n} (x_i - \overline{x})^2 \\ &= \frac{1}{12-1} \cdot \left((98 - 101.8333)^2 + (103 - 101.8333)^2 + \ldots + (101 - 101.8333)^2\right) \\ &\approx 205.6060 \end{aligned}$$

(1pt)

The sample standard deviation is then obtained by taking the square root of the sample variance:

$$s = \sqrt{205.6060} \approx 14.3390$$

(1pt)

4b [4pt] Determine based on this sample – without using the TESTS submenu of the graphing calculator – a 95% (two-sided) confidence interval for the parameter μ for the normal distribution.

Solution

Since the sample size $n = 12$ is smaller than 30, and the population standard deviation σ is unknown, we need to resort to the t-distribution with $n - 1 = 11$ degrees of freedom. (1pt)

We want a two-sided 95%-confidence interval for the population mean μ, so the significance level $\alpha = 0.05$. The confidence interval is of the form $[\overline{x} - t \cdot \frac{s}{\sqrt{n}}; \overline{x} + t \cdot \frac{s}{\sqrt{n}}]$, where

$$t = \text{InvT}(1 - \frac{\alpha}{2}; n-1) = \text{InvT}(0.975; 11) = 2.2010.$$

(1pt)

The confidence interval can thus be given by

$$\begin{aligned}
&[\overline{x} - t \cdot \frac{s}{\sqrt{n}}; \overline{x} + t \cdot \frac{s}{\sqrt{n}}] \\
&= [101.8333 - 2.2010 \cdot \frac{14.3390}{\sqrt{12}}; 101.8333 + 2.2010 \cdot \frac{14.3390}{\sqrt{12}}] \\
&\approx [92.7228; 110.9439]
\end{aligned}$$

(1pt)

4c [8pt] The contractor claims that the average lifespan of this new type of battery is at least 105 hours. State the null and alternative hypothesis and perform a suitable hypothesis test to verify this claim based on the given sample. Choose a significance level $\alpha = 0.02$.

Solution

The contractor makes a claim about the population parameter μ, so the null hypothesis and alternative hypothesis can be stated as

$$H_0 : \mu \geq 105 \text{ versus } H_1 : \mu < 105$$

(2pt)

The hypothesis test we need to perform is a one-sided t-test. Under assumption of the null hypothesis (in particular $\mu = 105$), the 95%-prediction interval (acceptable region) is then given as

$$[\mu - t \cdot \frac{s}{\sqrt{n}}; \infty),$$

where $t = \text{InvT}(1 - \alpha; n - 1) = \text{InvT}(0.98; 11) = 2.3281$. (2pt)

In other words, this interval is given by

$$[105 - 1.7959 \cdot \frac{14.3390}{\sqrt{12}}; \infty) = [95.3631; \infty).$$

(1pt)

The critical region therefore exists of all values of the sample mean smaller than 95.3631. Since the sample mean $\overline{x} = 101.8333$ does not lie in the critical region, (1pt) we fail to reject H_0. The data does not provide us enough evidence to reject the claim made by the contractor that the average lifespan is at least 105 hours. (2pt)

4d [5pt] Suppose the true mean of the battery lifespan is 100 hours. What is the minimum sample size required in order to have a type II error that is at most 0.1?

Solution

This question is disregarded, as it uses material not sufficiently covered in the course materials.

Problem 5 (20 points) A military logistics team tracks the time of day when supply deliveries are made to a forward operating base. Based on historical data, the team has identified the following expected distribution of deliveries across four time intervals in a 24-hour period:

- Night (0.00 – 6.00): 10% of deliveries
- Morning (6.00 – 12.00): 30% of deliveries
- Afternoon (12.00-18.00): 45% of deliveries
- Evening (18.00 – 24.00): 15% of deliveries

A total of 250 deliveries were made over the past week. The logistics team has recorded the following observed number of deliveries in each time interval:

Time Interval	Observed Deliveries
Night	17
Morning	83
Afternoon	112
Evening	38

5a [3pt] State the null and alternative hypotheses for the chi-square goodness-of-fit test.

Solution

With a chi-square goodness-of-fit test, we can state the null hypothesis and alternative hypothesis in this context as

H_0: the supply deliveries made to the forward operating base follow the expected distribution of deliveries. (2pt)

H_1: the supply deliveries made to the forward operating base do not follow the expected distribution of deliveries. (1pt)

5b [7pt] Calculate the expected frequencies for each time interval based on the given percentages and total number of deliveries.

Solution

The expected frequencies for each time interval can be calculated by taking percentages of 250 deliveries in accordance with the expected distribution: (3pt)

Time Interval	Observed Deliveries	Expected Deliveries
Night	17	$0.10 \cdot 250 = 25$
Morning	83	$0.30 \cdot 250 = 75$
Afternoon	112	$0.45 \cdot 250 = 112.5$
Evening	38	$0.15 \cdot 250 = 37.5$

(4pt)

5c [6pt] Compute the chi-square test statistic. Determine whether or not to reject the null hypothesis at a significance level of $\alpha = 0.05$.

Solution

We compute the chi-square test statistic χ^2 as

$$\begin{aligned}\chi^2 &= \sum \frac{(O-E)^2}{E} \\ &= \frac{(17-25)^2}{25} + \frac{(83-75)^2}{75} + \frac{(112-112.5)^2}{112.5} + \frac{(38-37.5)^2}{37.5} \\ &\approx 3.4222\end{aligned}$$ (3pt)

Since we have four different time intervals, this test statistic χ^2 approximately follows the χ^2-distribution with $n-1=3$ degrees of freedom. (1pt)

The p-value of the χ^2-test can be calculated as

$$p = \chi^2\text{cdf}(\text{lower} = 3.4222; \text{upper} = 10^{10}; \text{df} = 3) \approx 0.3310.$$ (1pt)

As the p-value is larger than the significance level α, we fail to reject the null hypothesis. (1pt)

5d [4pt] What can you conclude based on this chi-square test? Does the data suggest that the distribution of deliveries across the time intervals deviates significantly from the distribution claimed by the logistics team?

Solution

In 5c, we determined that we fail to reject the null hypothesis based on the given data. (1pt)

The data therefore suggests (not shows or proves!) that the distribution of deliveries across the time intervals does not deviate significantly from the expected distribution claimed by the logistics team. (3pt)

Problem 6 (15 points) A military research team is evaluating the relationship between the weight of a battle tank and the time it takes to deploy the vehicle in combat situations (in hours). The team collects the following data on 10 different vehicles:

Vehicle weight (in tons)	Deployment time (hours)
5	4.2
6	4.8
7	5.1
8	5.3
9	5.9
10	6.0
11	6.7
12	6.8
13	7.2
14	7.4

The research team believes that there is a linear relationship between the weight of the vehicle and the deployment time.

6a [6pt] Calculate the least-squares estimates $\hat{\beta}_0$ and $\hat{\beta}_1$ for the slope and intercept of the linear regression line $Y = \beta_0 + \beta_1 X$, where Y is the deployment time, and X is the vehicle weight.

Solution

We start by generating a table based on the data, the sample sums and means can then be used to calculate the least-squares estimated $\hat{\beta}_0$ and $\hat{\beta}_1$:

X	Y	XY	X^2	Y^2
5	4.2	21.0	25	17.64
6	4.8	28.8	36	23.04
7	5.1	35.7	49	26.01
8	5.3	42.4	64	28.09
9	5.9	53.1	81	34.81
10	6.0	60.0	100	36.0
11	6.7	73.7	121	44.89
12	6.8	81.6	144	46.24
13	7.2	93.6	169	51.84
14	7.4	103.6	196	54.76
$\sum X = 95$	$\sum Y = 59.4$	$\sum XY = 593.5$	$\sum X^2 = 985$	$\sum Y^2 = 363.32$
$\overline{X} = 9.5$	$\overline{Y} = 5.94$	$\overline{XY} = 59.35$	$\overline{X^2} = 98.5$	$\overline{Y^2} = 36.332$

Using the standard formulas for the least-squares estimates, we get (3pt)

$$\hat{\beta}_1 = \frac{\sum_{i=1}^n x_i y_i - n \cdot \overline{x} \cdot \overline{y}}{\sum_{i=1}^n x_i^2 - n \cdot \overline{x}^2} = \frac{593.5 - 10 \cdot 9.5 \cdot 5.94}{985 - 10 \cdot 9.5^2} \approx 0.3539$$

$$\hat{\beta}_0 = \overline{y} - \hat{\beta}_1 \cdot \overline{x} = 5.94 - 0.3539 \cdot 9.5 \approx 2.5780$$

(2pt)

The equation of the regression line is thus equal to $Y = 2.5780 + 0.3539X$. (1pt)

6b [4pt] Interpret the slope β_1 of the regression model in the context of this exercise. What does the slope suggest about the relationship between vehicle weight and deployment time?

Solution

In the context of this question, the slope β_1 of the regression tells us by how much more time it takes to deploy a vehicle in combat situations if the weight of the vehicle increases (by a unit of weight). (2pt)

Since the least-squares estimate $\hat{\beta}_1$ of the slope is positive, this suggests that the deployment time increases whenever the vehicle weight increases. (2pt)

6c [5pt] Calculate the correlation coefficient $R(X, Y)$ between the vehicle weight and deployment time. Based on the value of the correlation coefficient, what can you conclude about the strength and direction of the relationship between the two variables?

Solution

The correlation coefficient $R(X, Y)$ can be computed as follows:

$$\begin{aligned} R(X,Y) = \frac{Cov(X,Y)}{\sqrt{s_X^2 \cdot s_Y^2}} &= \frac{\sum_{i=1}^{n} x_i y_i - n \cdot \overline{x} \cdot \overline{y}}{\sqrt{(\sum_{i=1}^{n} x_i^2 - n \cdot \overline{x}^2) \cdot (\sum_{i=1}^{n} y_i^2 - n \cdot \overline{y}^2)}} \\ &= \frac{593.5 - 10 \cdot 9.5 \cdot 5.94}{\sqrt{(985 - 10 \cdot 9.5^2) \cdot (363.32 - 10 \cdot 5.94^2)}} \\ &\approx 0.9929 \end{aligned}$$

(3pt)

The value of the correlation coefficient is very close to 1, which suggests a very strong positive correlation between vehicle weight and deployment time in a combat situation. (2pt)