Ministerie van Defensie

Faculteit Militaire Wetenschappen

Gegevens student	
Naam:	
Peoplesoftnummer:	
Klas:	
Handtekening:	

Algemeen			
Vak:	Statistiek deel 2 (vierde kans)	**Vakcode:**	STA#2
Datum:	6 oktober 2025	**Tijdsduur:**	13:00-16:00
Examinator:	Dr. ir. D.A.M.P. Blom	**Aantal pagina's:**	4
Peer-review:	Dr. M.P. Roeling	**Aantal opgaven:**	4

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.
- Rond je antwoorden waar nodig af op vier decimalen.
- 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!

Formuleblad Statistiek (2024-2025)

Statistiek deel 1

Steekproefgemiddelde (gegeven een steekproef met n uitkomsten $x_1, x_2, \ldots, x_n$)

$$\overline{x} = \frac{\sum_i x_i}{n} = \frac{x_1 + x_2 + \ldots + x_n}{n}$$

Steekproefvariantie en steekproefstandaardafwijking:

$$s^2 = \frac{\sum_i (x_i - \overline{x})^2}{n-1} = \frac{(x_1 - \overline{x})^2 + (x_2 - \overline{x})^2 + \ldots + (x_n - \overline{x})^2}{n-1}$$

$$s = \sqrt{s^2} = \sqrt{\frac{(x_1 - \overline{x})^2 + (x_2 - \overline{x})^2 + \ldots + (x_n - \overline{x})^2}{n-1}}$$

Rekenregels kansrekening:

$$P(A \text{ of } B) = P(A) + P(B) - P(A \text{ en } B) \qquad \text{(optelregel)}$$

$$P(B) = 1 - P(\text{niet } B) \qquad \text{(complementregel)}$$

$$P(A \mid B) = \frac{P(A \text{ en } B)}{P(B)} \qquad \text{(conditionele kansen)}$$

Discrete en continue kansverdelingen:

	Discrete kansvariabelen	**Continue kansvariabelen**
Uitkomstenruimte:	Eindig / aftelbaar oneindig	Overaftelbaar oneindig
Toepassingen:	Tellen / categoriseren	Meten
Kansbegrip:	Kansfunctie $p(k) = P(X = k)$	Kansdichtheidsfunctie $f(x)$
CDF:	$F(k) = P(X \leq k) = \sum_{\ell : \ell \leq k} p(\ell)$	$F(x) = P(X \leq x) = \int_{-\infty}^{x} f(y)\, dy$
Verwachtingswaarde:	$E[X] = \sum_k k \cdot P(X = k)$	$E[X] = \int x \cdot f(x)\, dx$
Variantie:	$\text{Var}(X) = \sum_k (k - E[X])^2 \cdot P(X = k)$	$\text{Var}(X) = \int (x - E[X])^2 \cdot f(x)\, dx$
Standaardafwijking:	$\sigma(X) = \sqrt{\text{Var}(X)}$	$\sigma(X) = \sqrt{\text{Var}(X)}$

Speciale kansverdelingen:

- $X \sim \text{Binomiaal}(n, p)$: tellen van aantal successen bij onafhankelijke kansexperimenten met twee uitkomsten (Bernoulli-experimenten): succes / mislukking.

  Parameters: het aantal Bernoulli-experimenten n en de succeskans per experiment p.

- $X \sim \text{Poisson}(\lambda \cdot t)$: tellen van aantal "gebeurtenissen" in een "interval" van tijd / ruimte.

  Parameters: het gemiddelde aantal gebeurtenissen λ per meeteenheid (tijd / ruimte) en het aantal meeteenheden t.

  $\rightarrow$ *Voorbeeld: bij de meeteenheid van een dag bestaat een week uit $t = 7$ meeteenheden.*

- $T \sim \text{Exponentieel}(\lambda)$: meten van de tijd / ruimte tot de volgende gebeurtenis.

  Parameter: het gemiddelde aantal gebeurtenissen λ per meeteenheid (tijd / ruimte).

Verwachtingswaarde en variantie van veelgebruikte kansverdelingen:

Verdeling	Kans(dichtheids)functie	CDF	$E(X)$	Var(X)
		Discreet		
Uniform(a,b)	$p(k) = \frac{1}{b-a+1}$ $(k = a, a+1, \ldots, b)$	$F(k) = \begin{cases} 0 & x < a \\ \frac{k-a+1}{b-a+1} & a \leq k < b \\ 1 & k \geq b \end{cases}$	$\frac{a+b}{2}$	$\frac{(b-a+1)^2-1}{12}$
Binomiaal(n,p)	$p(k) = \binom{n}{k} p^k (1-p)^{n-k}$	$F(k) = \sum_{i=0}^{k} \binom{n}{i} p^i (1-p)^{n-i}$	np	$np(1-p)$
Poisson(λ)	$p(k) = e^{-\lambda} \cdot \frac{\lambda^k}{k!}$	$F(k) = \sum_{i=0}^{k} e^{-\lambda} \cdot \frac{\lambda^i}{i!}$	λ	λ
		Continuous		
Uniform(a,b)	$f(x) = \begin{cases} \frac{1}{b-a}, & a \leq x \leq b \\ 0, & \text{elders.} \end{cases}$	$F(x) = \begin{cases} 0, & x < a \\ \frac{x-a}{b-a}, & a \leq x < b \\ 1, & x \geq b \end{cases}$	$\frac{a+b}{2}$	$\frac{(b-a)^2}{12}$
Exponentieel(λ)	$f(x) = \lambda e^{-\lambda x}$, $x \geq 0$	$F(x) = \begin{cases} 1 - e^{-\lambda x}, & x \geq 0 \\ 0, & x < 0 \end{cases}$	$\frac{1}{\lambda}$	$\frac{1}{\lambda^2}$

Veelgebruikte functies op de grafische rekenmachine

Type vraag	TI-84 Plus	Casio
	Continue kansverdeling (willekeurig)	
$P(a \leq X \leq b)$	$\int_a^b f(x)\,dx$	$\int_a^b f(x)\,dx$
	$X \sim$ **Binomiaal**(n,p)	
$P(X = k)$ $P(X \leq k)$	binompdf(n,p,k) binomcdf(n,p,k)	BinomialPD(k,n,p) BinomialCD(k,n,p)
	$X \sim N(\mu, \sigma)$	
$P(a \leq X \leq b)$ Grenswaarde g zodat $P(X \leq g) = p$?	normalcdf(a,b,μ,σ) invNorm(p,μ,σ)	NormalCD(a,b,σ,μ) InvNormCD(tail=left, p,σ,μ)
	$X \sim$ **Poisson**(λ)	
$P(X = k)$ $P(X \leq k)$	poissonpdf(λ,k) poissoncdf(λ,k)	PoissonPD(k,λ) PoissonCD(k,λ)

z-score:

$$z = \frac{x - \mu}{\sigma}$$

Centrale limietstelling: Gegeven n kansvariabelen $X_1, X_2, \ldots, X_n$ die onderling onafhankelijk zijn en dezelfde kansverdeling hebben met een verwachtingswaarde μ en standaardafwijking σ, dan geldt (bij benadering) dat

- de som $\sum X = X_1 + X_2 + \ldots + X_n$ normaal verdeeld is met verwachtingswaarde $n \cdot \mu$ en standaardafwijking $\sqrt{n} \cdot \sigma$.
- het gemiddelde $\overline{X} = \frac{X_1 + X_2 + \ldots + X_n}{n}$ normaal verdeeld is met verwachtingswaarde μ en standaardafwijking $\frac{\sigma}{\sqrt{n}}$.

Statistiek deel 2:

Betrouwbaarheidsintervallen voor het gemiddelde μ (σ bekend)

- **$100 \cdot (1 - \alpha)$%-betrouwbaarheidsinterval (BI) voor μ:**

$$z_{\alpha/2} = \text{InvNorm}(\text{opp} = 1 - \alpha/2; \mu = 0; \sigma = 1)$$

$$[\overline{x} - z_{\alpha/2} \cdot \frac{\sigma}{\sqrt{n}}; \overline{x} + z_{\alpha/2} \cdot \frac{\sigma}{\sqrt{n}}]$$

- **Minimale steekproefomvang voor $100 \cdot (1 - \alpha)$%-BI als μ maximaal $\pm a$ mag afwijken:**

$$n \geq \left(\frac{z_{\alpha/2} \cdot \sigma}{a}\right)^2$$

Betrouwbaarheidsintervallen voor het gemiddelde μ (σ onbekend)

- **$100 \cdot (1 - \alpha)$%-betrouwbaarheidsinterval (BI) voor μ:**

$$t = \text{InvT}(\text{opp} = 1 - \alpha/2; \text{df} = n - 1)$$

$$[\overline{x} - t \cdot \frac{s}{\sqrt{n}}; \overline{x} + t \cdot \frac{s}{\sqrt{n}}]$$

- **Minimale steekproefomvang voor** $100 \cdot (1 - \alpha)$%**-BI als** μ **maximaal $\pm a$ mag afwijken:**

$$\text{GR tabel (voor verschillende } n\text{):} \quad \frac{s}{\sqrt{n}} \cdot \text{InvT}(\text{opp} = 1 - \alpha/2; \text{df} = n - 1) \leq a$$

  NB: zodra $n \geq 30$, vallen de normale en de t-verdeling nagenoeg samen. Je mag dan rekenen met de schatting s in plaats van de daadwerkelijke (onbekende) σ.

- **Onderscheidend vermogen (toets met** $H_0 : \mu = \mu_0$ **vs.** $H_1 : \mu \neq \mu_0$**, en gegeven** $\mu = \mu_1$**)**

$$1 - \beta = P(\overline{X} \text{ neemt waarde aan in het kritieke gebied} \mid \mu = \mu_1)$$

Betrouwbaarheidsintervallen voor de binomiale succeskans p

Betrouwbaarheidsinterval voor p (Clopper-Pearson): Gegeven een binomiale verdeling met n Bernoulli-experimenten en onbekende p, en uitkomst k.

1. Bereken de succeskans p_1 zodat geldt $P(X \leq k) = \text{binomcdf}(n; p; k) = \alpha/2$
2. Bereken de succeskans p_2 zodat geldt $P(X \geq k) = 1 - \text{binomcdf}(n; p; k - 1) = \alpha/2$
3. De berekende waarden voor p_1 en p_2 zijn de grenzen van het Clopper-Pearson interval.

Hypothesetoetsen

Stappenplan hypothesetoetsen

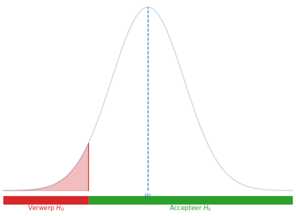
1. Definieer de nulhypothese H_0 en de alternatieve hypothese H_1.
2. Bepaal het significantieniveau α (kans op verwerpen van H_0 terwijl H_0 waar is → type-I fout)
3. Verzamel data voor de toetsingsgrootheid
4. Bereken de toetsingsgrootheid
   - Uitgaande van de nulhypothese H_0 maken we aannames over de kansverdeling van de toetsingsgrootheid!
5. Geef een conclusie (met behulp van het kritieke gebied / p-waarde) en vertaal deze terug naar de originele probleemcontext.

Drie typen hypothesetoetsen: linkszijdig, tweezijdig, rechtszijdig

Linkszijdige toets

Kritiek gebied:

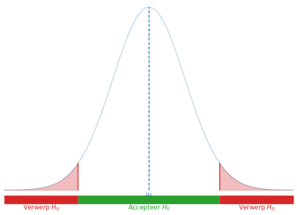
$(-\infty; g]$



Tweezijdige toets

Kritiek gebied:

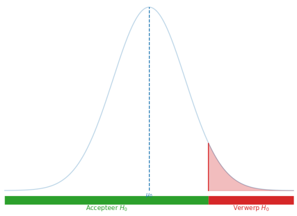
$(-\infty; g_1]$ en $[g_2; \infty)$



Rechtszijdige toets

Kritiek gebied:

$[g; \infty)$



Kansverdeling (onder H_0)	**Linkszijdig**	**Tweezijdig**	**Rechtszijdig**
$N(\mu; \sigma)$	$g = \text{InvNorm}(\alpha; \mu; \sigma)$	$g_1 = \text{InvNorm}(\text{opp} = \frac{\alpha}{2}; \mu; \sigma)$ $g_2 = \text{InvNorm}(\text{opp} = 1 - \frac{\alpha}{2}; \mu; \sigma)$	$g = \text{InvNorm}(1 - \alpha; \mu; \sigma)$
$t(\text{df})$	$g = \text{InvT}(\alpha; \text{df})$	$g_1 = \text{InvT}(\text{opp} = \frac{\alpha}{2}; \text{df})$ $g_2 = \text{InvT}(\text{opp} = 1 - \frac{\alpha}{2}; \text{df})$	$g = \text{InvT}(1 - \alpha; \text{df})$
Grenzen die met de solver functie moeten worden opgelost:			
$\chi^2(\text{df})$ (chikwadraat)	$\chi^2\text{cdf}(0; g; \text{df}) = \alpha$	$\chi^2\text{cdf}(0; g_1; \text{df}) = \frac{\alpha}{2}$ $\chi^2\text{cdf}(g_2; 10^{99}; \text{df}) = \frac{\alpha}{2}$	$\chi^2\text{cdf}(g; 10^{99}; \text{df}) = \alpha$
$F(\text{df}_A; \text{df}_B)$	$\text{Fcdf}(0; g; \text{df}_A; \text{df}_B) = \alpha$	$\text{Fcdf}(0; g_1; \text{df}_A; \text{df}_B) = \frac{\alpha}{2}$ $\text{Fcdf}(g_2; 10^{99}; \text{df}_A; \text{df}_B) = \frac{\alpha}{2}$	$\text{Fcdf}(g; 10^{99}; \text{df}_A; \text{df}_B) = \alpha$

p-waardes uitrekenen (gegeven een theoretische en geobserveerde toetsingsgrootheid T en t)

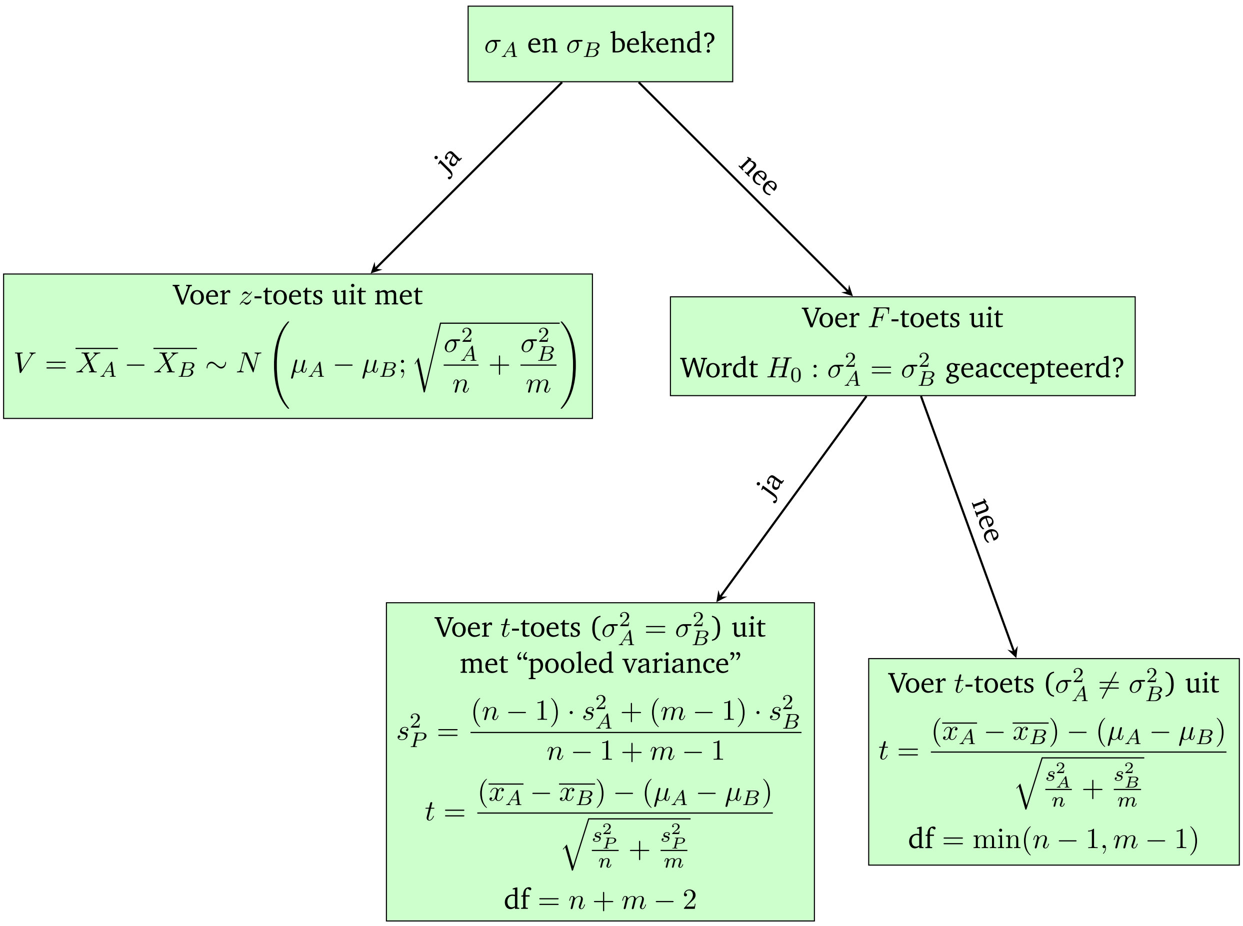
Kansverdeling (onder H_0)	**Linkszijdig ($P(T \leq t)$)**	**Rechtszijdig ($P(T \geq t)$)**
$N(\mu; \sigma)$	$p = \text{normalcdf}(-10^{99}; t; \mu; \sigma)$	$p = \text{normalcdf}(t; 10^{99}; \mu; \sigma)$
$t(\text{df})$	$p = \text{tcdf}(-10^{99}; t; \text{df})$	$p = \text{tcdf}(t; 10^{99}; \text{df})$
$\chi^2(\text{df})$	$p = \chi^2\text{cdf}(0; t; \text{df})$	$p = \chi^2\text{cdf}(t; 10^{99}; \text{df})$
$F(\text{df}_A; \text{df}_B)$	$p = \text{Fcdf}(0; t; \text{df}_A; \text{df}_B)$	$p = \text{Fcdf}(t; 10^{99}; \text{df}_A; \text{df}_B)$

NB: Om met de p-waarde een conclusie te trekken uit een hypothesetoets vergelijken we de p-waarde met het significantieniveau α. Let op: bij tweezijdige toetsen neem je het minimum van de linkszijdige en rechtszijdige p-waarde en vergelijk je deze met $\alpha/2$!

Soorten toetsen

Soort toets	Toetsingsgrootheid	Kansverdeling (onder H_0)
Toetsen voor het gemiddelde $\mu \leq \mu_0$ of $\mu = \mu_0$ of $\mu \geq \mu_0$		
z-toets (σ bekend)	$\overline{X}$	$N(\mu_0; \frac{\sigma}{\sqrt{n}})$
t-toets (σ onbekend)	$T = \frac{\overline{X}-\mu_0}{\frac{s}{\sqrt{n}}}$	$t(\text{df} = n-1)$
Chikwadraattoetsen (χ^2)		
Onafhankelijkheid	$X^2 = \sum_{i,j} \frac{(O_{ij}-E_{ij})^2}{E_{ij}}$	$\chi^2(\text{df} = (\#\text{rijen-1}) \cdot (\#\text{kolommen-1}))$
Aanpassing (goodness-of-fit)	$X^2 = \sum_i \frac{(O_i-E_i)^2}{E_i}$	$\chi^2(\text{df} = (\#\text{categorieën-1}))$
Verschiltoetsen (op basis van twee populaties A en B)		
F-toets: $\sigma_A^2 = \sigma_B^2$	$F = \frac{S_A^2}{S_B^2}$	$F(\text{df}_A, \text{df}_B)$
z-toets	$V = \overline{X_A} - \overline{X_B}$	$N\left(\mu_A - \mu_B; \sqrt{\frac{\sigma_A^2}{n} + \frac{\sigma_B^2}{m}}\right)$
t-toets ($\sigma_A^2 = \sigma_B^2$)	$T = \frac{(\overline{X_A}-\overline{X_B})-(\mu_A-\mu_B)}{\sqrt{\frac{S_P^2}{n}+\frac{S_P^2}{m}}}$	$t(\text{df} = n+m-2)$
t-toets ($\sigma_A^2 \neq \sigma_B^2$)	$T = \frac{(\overline{X_A}-\overline{X_B})-(\mu_A-\mu_B)}{\sqrt{\frac{S_A^2}{n}+\frac{S_B^2}{m}}}$	$t(\text{df} = \min(n-1; m-1))$

Beslisboom verschiltoetsen



Correlatie en regressie

Correlatiecoëfficiënt van Pearson:

$$r = \frac{\overline{xy} - \overline{x} \cdot \overline{y}}{\sqrt{(\overline{x^2} - \overline{x}^2) \cdot (\overline{y^2} - \overline{y}^2)}}$$

Correlatiecoëfficiënt van Spearman:

$$r_s = 1 - \frac{6 \cdot \sum_i d_i^2}{n^3 - n}$$

Coëfficiënten van de lineaire regressielijn $Y = a + b \cdot X$:

$$b = \frac{\overline{xy} - \overline{x} \cdot \overline{y}}{\overline{x^2} - (\overline{x})^2}$$

$$a = \overline{y} - b \cdot \overline{x}$$

Schatting van de variantie van de storingsterm ε:

$$s_\varepsilon^2 = \frac{\sum e_i^2}{n-2} = \frac{\sum \left(y_i - (a + b \cdot x_i)\right)^2}{n-2} = \frac{n}{n-2} \cdot \left(\overline{y^2} - a \cdot \overline{y} - b \cdot \overline{xy}\right)$$

$100 \cdot (1 - \alpha)\%$-betrouwbaarheidsinterval voor de gemiddelde Y bij een gegeven $X = x_0$:

$$t = \mathrm{InvT}(\text{opp} = 1 - \alpha/2; \text{df} = n - 2)$$

$$s_\mu = s_\varepsilon \cdot \sqrt{\frac{1}{n} \cdot \left(1 + \frac{(x_0 - \overline{x})^2}{\overline{x^2} - \overline{x}^2}\right)}$$

$$[a + b \cdot x_0 - t \cdot s_\mu; a + b \cdot x_0 + t \cdot s_\mu]$$

$100 \cdot (1 - \alpha)\%$-betrouwbaarheidsinterval voor Y bij een gegeven $X = x_0$:

$$t = \mathrm{InvT}(\text{opp} = 1 - \alpha/2; \text{df} = n - 2)$$

$$s_f = s_\varepsilon \cdot \sqrt{1 + \frac{1}{n} \cdot \left(1 + \frac{(x_0 - \overline{x})^2}{\overline{x^2} - \overline{x}^2}\right)}$$

$$[a + b \cdot x_0 - t \cdot s_f; a + b \cdot x_0 + t \cdot s_f]$$

Opgave 1 (20 punten) Om het hoofd te kunnen bieden aan toekomstige dreigingen, wordt onderzoek gedaan naar high-performance materialen voor het materieel van de toekomst. Onderdeel van dit onderzoek is een test van een nieuwe metaallegering, die beter bestand zou moeten zijn tegen zware explosies. Voor deze test worden een aantal samples gemaakt van de legering waarvan de treksterkte wordt gemeten (in N/mm^2). De waargenomen aantallen zijn als volgt:

$$931, 978, 954, 962, 984, 976, 923, 940, 988, 937$$

Neem aan dat de treksterkte een normale verdeling volgt.

1a [5pt] Bereken van de gemeten waarden het steekproefgemiddelde en de steekproefstandaardafwijking.

1b [7pt] Bereken een $90\,\%$-betrouwbaarheidsinterval voor de gemiddelde treksterkte μ van de nieuwe metaallegering, op grond van de bovengenoemde steekproefresultaten. Laat hierbij duidelijk je berekeningen zien en maak geen gebruik van de optie TESTS/Interval op de grafische rekenmachine. Rond het interval af op gehele getallen zodanig dat de betrouwbaarheid gewaarborgd blijft.

1c [6pt] De onderzoekers beweren dat de nieuwe metaallegering een gemiddelde treksterkte heeft van minstens $940 N/mm^2$. Toets deze bewering met een geschikte hypothesetoets. Bepaal de toetsuitslag aan de hand van het kritieke gebied op basis van bovenstaande steekproef. Kies in dit geval voor een significantieniveau van $\alpha = 0.05$.

1d [2pt] Leg in eigen woorden uit wat de toetsuitslag betekent voor de gemiddelde treksterkte van het nieuwe materiaal.

Opgave 2 (20 punten) Een Oekraïens militaire inlichtingencentrum heeft het vermoeden dat er een bepaald patroon zit in het aantal Shahed-drones dat gebruikt wordt in Russische drone-aanvallen. Om dit te onderzoeken, hebben ze de gegevens verzameld over het aantal drones dat gebruikt werd in 171 aanvallen.

Aantal drones	**Frequentie**
1	18
2	23
3	34
4	27
5	31
≥ 6	38

2a [10pt] Toets of de verdeling van het aantal drones significant afwijkt van een uniforme verdeling over deze zes categorieën. Bepaal de toetsuitslag aan de hand van de p-waarde en kies als significantieniveau $\alpha = 0.02$.

2b [5pt] Voer de toets opnieuw uit, maar bepaal nu de toetsuitslag aan de hand van het kritieke gebied.

2c [5pt] Verklaar de toetsuitslag aan de hand van de frequenties in de bovenstaande tabel en de verwachte frequenties volgens een uniforme verdeling.

Opgave 3 (30 punten) Het Ministerie van Defensie wil onderzoeken of er een verband is tussen het OPCO waar een militair deel van uitmaakt en de mate waarin hij/zij tevreden is met zijn/haar huidige gevechtstenue. De tevredenheid met de uitrusting wordt gemeten met een enquête en kan twee waarden aannemen: “tevreden” en “niet tevreden”.

	Tevreden	**Niet tevreden**
CLAS	250	150
CLRS	200	200
CZSK	220	180
KMAR	180	220

3a [4pt] Welk type hypothesetoets dienen we uit te voeren om aan te tonen of er een verband is tussen OPCO en mate van tevredenheid? Formuleer de nulhypothese H_0 en de alternatieve hypothese H_1 van de bijbehorende hypothesetoets.

3b [11pt] Bepaal de p-waarde van de desbetreffende hypothesetoets.

3c [5pt] Geef de conclusie van de hypothesetoets (op basis van een significantieniveau $\alpha = 0.05$) met behulp van de berekende p-waarde.

3d [10pt] Naast de vraag of de mate van tevredenheid afhangt van het OPCO, is het ministerie ook geïnteresseerd in het percentage tevreden militairen. Bepaal een $95\,\%$-betrouwbaarheidsinterval voor het percentage tevreden militairen (over alle OPCO’s samen).

Opgave 4 (30 punten) In een gezamenlijk project van sportinstructeurs van de KMA en de geneeskundige dienst wordt het verband onderzocht tussen de gemiddelde slaaptijd van een cadet (in uren) en de hersteltijd na een zware inspanning (in uren). Voor de hersteltijd wordt gekeken naar een spierpijn-index, en een cadet is volledig hersteld wanneer de spierpijn-index onder een gegeven drempelwaarde komt.

Van negen cadetten wordt gemeten hoe lang ze gemiddeld hebben geslapen en hoe lang hun hersteltijd is na een zware inspanning.

Slaaptijd (in uren)	4.0	4.5	5.0	5.5	6.0	6.5	7.0	7.5	8.0
Hersteltijd (in uren)	66	61	63	62	65	64	57	59	60

4a [2pt] Als we een regressie-analyse willen uitvoeren, welke variabele zou dan de afhankelijke variabele Y zijn en welke variabele zou de onafhankelijke variabele X zijn?

4b [5pt] Teken het bijbehorende spreidingsdiagram op basis van je antwoord op vraag a.

4c [8pt] Bereken Pearson's correlatiecoëfficiënt $r(x, y)$. Wat kun je concluderen over de samenhang van de twee variabelen?

4d [7pt] Bereken de regressielijn $Y = a + b \cdot X$ door berekening van de coëfficiënten a en b. Bepaal aan de hand van de regressielijn een statistisch verantwoorde voorspelling voor de hersteltijd van een cadet die 6 uur en 45 minuten heeft geslapen.

4e [8pt] Bereken een 90%-voorspellingsinterval voor de hersteltijd van een willekeurige cadet die 6 uur en 45 minuten heeft geslapen.