Ministerie van Defensie

# Faculteit Militaire Wetenschappen

| Gegevens student | |
|---|---|
| **Naam:** | |
| **Peoplesoftnummer:** | |
| **Klas:** | |
| **Handtekening:** | |

| Algemeen | | | |
|---|---|---|---|
| **Vak:** | Statistiek (deel 2) | **Vakcode:** | STA#2 |
| **Datum:** | 25 juli 2025 | **Tijdsduur:** | 9:00-12:00 |
| **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:**

- $X \sim$ 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 $p$.

- $X \sim$ Poisson$(\lambda \cdot t)$: tellen van aantal “gebeurtenissen” in een “interval” van tijd / ruimte.

  **Parameters:** het gemiddelde aantal gebeurtenissen $\lambda$ per meeteenheid (tijd / ruimte) en het aantal meeteenheden $t$.

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

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

  **Parameter:** het gemiddelde aantal gebeurtenissen $\lambda$ per meeteenheid (tijd / ruimte).

**Verwachtingswaarde en variantie van veelgebruikte kansverdelingen:**

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

**Veelgebruikte functies op de grafische rekenmachine**

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

**$z$-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 $\mu$ en standaardafwijking $\sigma$, 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 $\mu$ en standaardafwijking $\frac{\sigma}{\sqrt{n}}$.

## Statistiek deel 2:

### Betrouwbaarheidsintervallen voor het gemiddelde $\mu$ ($\sigma$ bekend)

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

$$z_{\alpha/2} = \text{InvNorm}(\text{opp} = 1-\alpha/2; \mu = 0; \sigma = 1)$$
$$[\overline{x} - z_{\alpha/2} \cdot \frac{\sigma}{\sqrt{n}}; \overline{x} + z_{\alpha/2} \cdot \frac{\sigma}{\sqrt{n}}]$$

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

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

### Betrouwbaarheidsintervallen voor het gemiddelde $\mu$ ($\sigma$ onbekend)

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

$$t = \text{InvT}(\text{opp} = 1-\alpha/2; \text{df} = n-1)$$
$$[\overline{x} - t \cdot \frac{s}{\sqrt{n}}; \overline{x} + t \cdot \frac{s}{\sqrt{n}}]$$

- **Minimale steekproefomvang voor** $100 \cdot (1-\alpha)$%**-BI als** $\mu$ **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) $\sigma$.*

- **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.

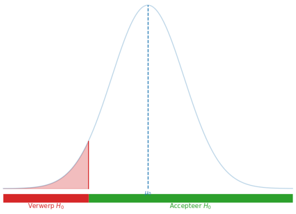
### Hypothesetoetsen

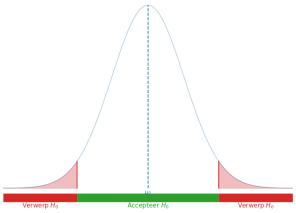
#### Stappenplan hypothesetoetsen

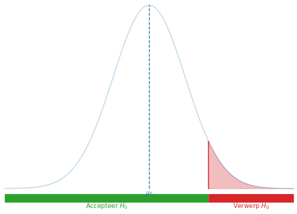
1. Definieer de nulhypothese $H_0$ en de alternatieve hypothese $H_1$.
2. Bepaal het significantieniveau $\alpha$ (kans op verwerpen van $H_0$ terwijl $H_0$ waar is $\rightarrow$ type-I fout)
3. Verzamel data voor de toetsingsgrootheid
4. Bereken de toetsingsgrootheid
   - Uitgaande van de nulhypothese $H_0$ 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** | **Tweezijdige toets** | **Rechtszijdige toets** |
|---|---|---|
| Kritiek gebied: | Kritiek gebied: | Kritiek gebied: |
| $(-\infty; g]$ | $(-\infty; g_1]$ en $[g_2; \infty)$ | $[g; \infty)$ |




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

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

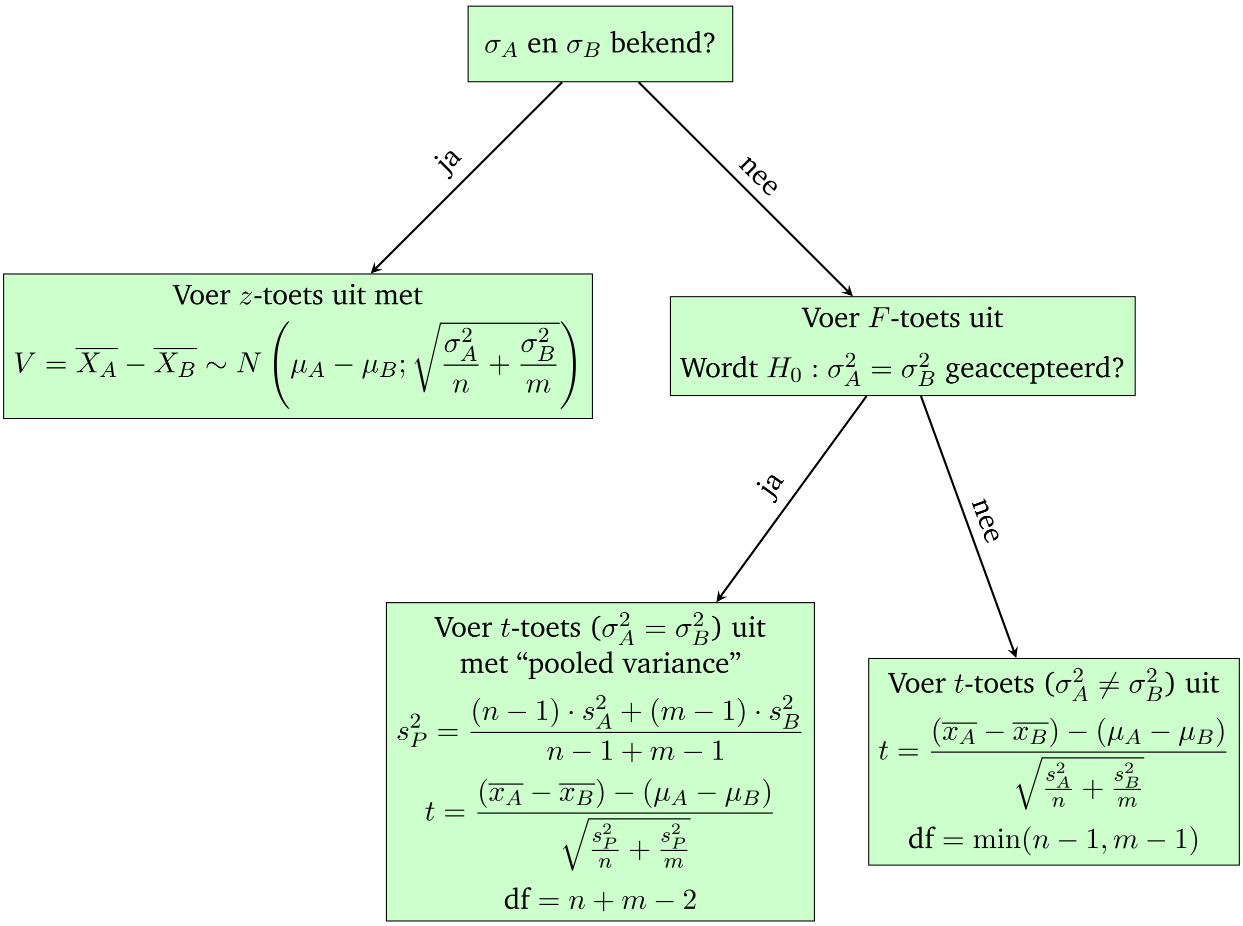
| **Kansverdeling (onder $H_0$)** | **Linkszijdig ($P(T \leq t)$)** | **Rechtszijdig ($P(T \geq t)$)** |
|---|---|---|
| $N(\mu;\sigma)$ | $p = \text{normalcdf}(-10^{99};t;\mu;\sigma)$ | $p = \text{normalcdf}(t;10^{99};\mu;\sigma)$ |
| $t(\text{df})$ | $p = \text{tcdf}(-10^{99};t;\text{df})$ | $p = \text{tcdf}(t;10^{99};\text{df})$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

*NB: Om met de $p$-waarde een conclusie te trekken uit een hypothesetoets vergelijken we de $p$-waarde met het significantieniveau $\alpha$. 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 ($\sigma$ bekend) | $\overline{X}$ | $N(\mu_0; \frac{\sigma}{\sqrt{n}})$ |
| $t$-toets ($\sigma$ onbekend) | $T = \frac{\overline{X}-\mu_0}{\frac{s}{\sqrt{n}}}$ | $t(\text{df} = n-1)$ |
| **Chikwadraattoetsen ($\chi^2$)** | | |
| 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 $\varepsilon$:**

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

**$100 \cdot (1 - \alpha)\%$-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)** Een onderzoeksteam van het Maritime Warfare Center (MWC) onderzoekt de signaalsterkte (in decibel) van sonarpulsen die worden gereflecteerd door een nieuw type onderzeeboot. Het huidige (oude) ontwerp heeft een gemiddelde gereflecteerde signaalsterkte van $\mu = \mu_0 = 65$ decibel.

Het MWC wil toetsen of bij het nieuwe ontwerp de gemiddelde signaalsterkte significant verschilt van het oude ontwerp. In een steekproef van tien onafhankelijke metingen met het nieuwe ontwerp is de gereflecteerde signaalsterkte gemiddeld $63,75$ decibel met een standaardafwijking van $0,78$. Je mag aannemen dat de signaalsterktes normaal verdeeld zijn met onbekende standaardafwijking $\sigma$.

Gebruik voor de hypothesetoets een significantieniveau van $\alpha = 0,05$.

**1a [5pt]** Definieer de nulhypothese en de alternatieve hypothese van de hypothesetoets. Verklaar het gekozen type (tweezijdig, linkszijdig of rechtszijdig) van de toets.

**1b [9pt]** Voer de bijbehorende hypothesetoets uit met behulp van het kritieke gebied.

**1c [6pt]** Bereken de kans op een type-II fout $\beta$ als geldt dat het daadwerkelijke populatiegemiddelde van het nieuwe ontwerp gelijk is aan $\mu = 64,5$ decibel, en dat de standaardafwijking $\sigma$ gelijk is aan $0,8$.

**Opgave 2 (28 punten)** Bij een onderzoek binnen de Koninklijke Luchtmacht wordt gekeken naar de effectiviteit van de training van piloten van gevechtsvliegtuigen. Tijdens de training neemt een piloot plaats in een speciale centrifuge, waarbij hij of zij aan hoge G-krachten wordt blootgesteld. Hierbij wordt geoefend met de *Anti-G Straining Maneuver (AGSM)*, een speciale ademhalingstechniek die helpt om langer weerstand te kunnen bieden aan de hoge G-krachten.

De focus van het onderzoek ligt op het bepalen van de maximale G-belasting totdat een piloot verstoringen ervaart in het gezichtsveld, zoals tunnelvisie of tijdelijke blackouts. Hierbij zijn voor twee populaties steekproeven genomen, namelijk voor de populatie F-16 vliegers (populatie $A$) en voor de populatie F-35 vliegers (populatie $B$).

| | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|
| **G-belasting F-16 vliegers (in G)** | $7,4$ | $8,1$ | $8,3$ | $7,8$ | $7,9$ | $8,0$ | $7,7$ | $8,2$ | $7,6$ | $8,0$ |
| **G-belasting F-35 vliegers (in G)** | $8,3$ | $8,7$ | $8,9$ | $8,2$ | $8,5$ | $8,4$ | $8,8$ | $8,6$ | $8,3$ | $8,5$ |

Het doel van het onderzoek is om te toetsen of de gemiddelde maximale G-belasting significant hoger is voor F-35 vliegers. Voor beide populaties wordt aangenomen dat de maximale G-belasting van een willekeurige vlieger normaal verdeeld is met onbekende verwachtingswaarde en standaardafwijking.

**2a [10pt]** Bepaal voor beide populaties de steekproefgemiddeldes en de steekproefvarianties.

**2b [10pt]** Voer een $F$-toets uit en bepaal met behulp van de $p$-waarde of de varianties in de maximale G-belasting gelijk zijn voor beide populaties. Kies voor het significantieniveau $\alpha = 0,05$.

**2c [10pt]** Bepaal met behulp van een onafhankelijke $t$-toets of gemiddeld de maximale G-belasting significant hoger is voor F-35 vliegers. Kies opnieuw voor het significantieniveau $\alpha = 0,05$.

**Opgave 3 (20 punten)** Bij de sollicitatiegesprekken voor cadetten en adelborsten wordt door de aanstellingscommissie aan de aspirant-officier gevraagd naar zijn of haar primaire motivatie om officier te willen worden. De meest genoemde antwoorden van de aspiranten zijn ruwweg in te delen in drie categorieën: leiderschap, avontuur en dienen aan het vaderland.

Naar aanleiding van de sollicitatiegesprekken kunnen de antwoorden als volgt worden samengevat:

| | Cadetten | Adelborsten | Totaal |
|---|---|---|---|
| Leiderschap | 52 | 22 | 74 |
| Avontuur | 23 | 9 | 32 |
| Vaderland dienen | 38 | 28 | 66 |
| Totaal | 113 | 59 | 172 |

C-NLDA wil inzicht krijgen in of de primaire motivatie van cadetten significant afwijkt van die van adelborsten.

**3a [4pt]** Welk type hypothesetoets dienen we hiervoor uit te voeren? 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.

**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 gemiddelde $6$ 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.