Formularium
Kansrekening en statistiek

UNIVERSITEIT ANTWERPEN
2014-2015

S. VAN AERT
P. GOOS

Hoofdstuk 1: Wat is statistiek?

Hoofdstuk 2: Data en hun voorstelling

Hoofdstuk 3: Beschrijvende statistieken van steekproefgegevens

Mediaan: M_e
middelste element van de geordende data:

- $(\frac{n+1}{2})$-de element bij oneven aantal elementen n
- gemiddelde van $\frac{n}{2}$-de en $(\frac{n}{2}+1)$-ste element bij even aantal elementen n

Modus: M_o
waarneming met de grootste frequentie
(gegroepeerde gegevens: klassecentrum van de modale klasse)

Rekenkundig gemiddelde:
$\overline{x} = \frac{1}{n}\sum_{i=1}^{n} x_i$ (gegroepeerde gegevens: $\overline{x} = \frac{1}{n}\sum_{i=1}^{k} f_i x_i$)

Ordestatistieken, percentielen en kwantielen:
$i$de ordestatistiek of $i$de ordekengetal $x_{(i)}$: $i$de waarneming van geordende data;
$(100\times p)$de steekproefpercentiel of -kwantiel c_p $(0 < p < 1)$: reëel getal groter dan $100\times p\%$ van de waarnemingen, en kleiner dan $100\times(1-p)\%$ van de waarnemingen;
berekening: $c_p = x_{(q)} + f(x_{(q+1)} - x_{(q)})$ met $a = 1 + p(n-1)$, q het grootste geheel getal kleiner dan a, en $f = a - q$
Eerste kwartiel $= Q_1 = c_{0.25}$, tweede kwartiel $= Q_2 = c_{0.5}$, derde kwartiel $= Q_3 = c_{0.75}$

Spreidingsbreedte:
$R = x_{max} - x_{min}$

Interkwartielbreedte:
$Q = Q_3 - Q_1$

Gemiddelde absolute afwijking:
$MAD = \frac{\sum_{i=1}^{n}|x_i-\overline{x}|}{n}$

Steekproefvariantie:
$s^2 = \frac{\sum_{i=1}^{n}(x_i-\overline{x})^2}{n-1} = \frac{1}{n-1}(\sum_{i=1}^{n} x_i^2 - n\overline{x}^2) = \frac{1}{n-1}\left\{\sum_{i=1}^{n} x_i^2 - \frac{1}{n}(\sum_{i=1}^{n} x_i)^2\right\}$
(gegroepeerde gegevens: $s^2 = \frac{1}{n-1}\sum_{i=1}^{k} f_i(x_i-\overline{x})^2$)
Steekproefstandaarddeviatie $s = \sqrt{s^2}$

Populatievariantie:
$\sigma^2 = \frac{\sum_{i=1}^{N}(x_i-\mu)^2}{N}$

Variatiecoëfficiënt :
$VC = \frac{s}{\overline{x}}$

Pearsons scheefheidscoëfficiënt:
$S_P = \frac{3(\overline{x}-M_e)}{s}$

Covariantie:
Steekproefcovariantie $s_{XY} = \frac{1}{n-1}\sum_{i=1}^{n}(x_i-\overline{x})(y_i-\overline{y})$
Populatiecovariantie: $\sigma_{XY} = \frac{1}{N}\sum_{i=1}^{N}(x_i-\mu_X)(y_i-\mu_Y)$

Correlatiecoëfficiënt :
Steekproefcorrelatiecoëfficiënt: $r_{XY} = \frac{s_{XY}}{s_X s_Y} = \frac{\sum_{i=1}^{n}(x_i-\overline{x})(y_i-\overline{y})}{\sqrt{\sum_{i=1}^{n}(x_i-\overline{x})^2}\sqrt{\sum_{i=1}^{n}(y_i-\overline{y})^2}}$
Populatiecorrelatiecoëfficiënt : $\rho_{XY} = \frac{\sigma_{XY}}{\sigma_X\sigma_Y}$

Hoofdstuk 4: Kansrekenen

Axioma's:

- $P(G) \geq 0$
- $P(\Omega) = 1$
- voor mekaar uitsluitende G_i: $P(G_1 \cup G_2 \cup G_3 \cup \dots) = \sum_i P(G_i)$

Rekenregels:

- $P(\emptyset) = 0$
- Indien $G_1 \subseteq G_2$, dan is $P(G_2 \backslash G_1) = P(G_2) - P(G_1)$
- Voor een willekeurige gebeurtenis G geldt dat $0 \leq P(G) \leq 1$
- $P(G) + P(G^c) = 1$
- $P(G_1 \cup G_2) = P(G_1) + P(G_2) - P(G_1 \cap G_2)$ (optelregel)
- $P(G_1 \cup G_2 \cup G_3) = P(G_1) + P(G_2) + P(G_3) - P(G_1 \cap G_2) - P(G_1 \cap G_3) - P(G_2 \cap G_3) + P(G_1 \cap G_2 \cap G_3)$ (veralgemeende optelregel)

Voorwaardelijke kans:
$P(G_1|G_2) = \frac{P(G_1\cap G_2)}{P(G_2)}$

Vermenigvuldigingsregel:
$P(G_1 \cap G_2) = P(G_1|G_2)P(G_2)$
$P(G_1 \cap G_2 \cap G_3) = P(G_3|G_1 \cap G_2)P(G_2|G_1)P(G_1)$

Onafhankelijke gebeurtenissen:
G_1 is onafhankelijk van G_2 als $P(G_1) = P(G_1|G_2)$

Partitie:
Niet-lege G_i vormen een partitie van Ω als $G_1 \cup G_2 \cup G_3 \cup \cdots = \Omega$ en $G_i \cap G_j = \emptyset$ voor elke $i \neq j$

Stelling van de totale kans:
Voor G_0 en partitie $G_1, G_2, \ldots, G_k$ van Ω: $P(G_0) = \sum_{i=1}^{k} P(G_0|G_i)P(G_i)$

Kansregel van Bayes:
Voor G_0 en partitie $G_1, G_2, \ldots, G_k$ van Ω: $P(G_j|G_0) = \frac{P(G_0|G_j)P(G_j)}{\sum_{i=1}^{k} P(G_0|G_i)P(G_i)}$

Hoofdstuk 5: Univariate kansvariabelen

Kansvariabele of stochastische variabele:
Functie X die reëel getal associeert met uitkomst ω van experiment
Als G_x de gebeurtenis is waarvoor $X = x$, dan $p_X(x) = P(X = x) = P(G_x)$

Kansverdeling van discrete X:
Opsomming van $p_X(x_i) = P(X = x_i), \quad i = 1, 2, \ldots, k$

Kansdichtheid van continue X:
Niet-negatieve functie $f_X(x)$, gedefinieerd over de reële rechte, met
$P(X \in I) = \int_I f_X(x)dx$ voor alle intervallen I

(Cumulatieve) verdelingsfunctie:
$F_X(x) = P(X \leq x)$ voor elk reëel getal x
Discrete X: $F_X(x) = \sum_{x_i \leq x} p_X(x_i)$, continue X: $F_X(x) = \int_{-\infty}^{x} f_X(y)dy$

Hoofdstuk 6: Kengetallen van populaties en processen

Verwachte waarde:
Discrete X: $\mu_X = E(X) = \sum_{i=1}^{k} x_i p_X(x_i)$
Continue X:$\mu_X = E(X) = \int_{-\infty}^{+\infty} x f_X(x)dx$

Verwachte waarde van een functie van een kansvariabele :
$\mu_Y = E(Y) = E\{g(X)\} = \sum_x g(x)p_X(x)$

Verwachte waarde van een lineaire transformatie :
$Y = \sum_{i=1}^{k} a_i g_i(X)$, met constanten $a_1, a_2, \ldots, a_k$ en functies $g_1(X)$, $g_2(X), \ldots, g_k(X)$:
$\mu_Y = E(Y) = E\{\sum_{i=1}^{k} a_i g_i(X)\} = \sum_{i=1}^{k} a_i E\{g_i(X)\}$

Variantie:
$\sigma_X^2 = var(X) = E\{(X - \mu_X)^2\}$
Discrete X: $\sigma_X^2 = var(X) = \sum_{i=1}^{k} (x_i - \mu_X)^2 p_X(x_i)$
Continue X: $\sigma_X^2 = var(X) = \int_{-\infty}^{+\infty} (x - \mu_X)^2 f_X(x)dx$

Standaarddeviatie:
$\sigma_X = +\sqrt{\sigma_X^2}$

Variantie van een lineaire transformatie:
$var(aX + b) = a^2\sigma_X^2$

Gestandaardiseerde kansvariabele:
$Z = \frac{X - \mu_X}{\sigma_X}$ met $E(Z) = 0$ en $\sigma_Z^2 = \sigma_Z = 1$

Modus:
Waarde waarvoor kansverdeling of kansdichtheid maximale waarde aanneemt

Mediaan:
Discrete X: $\gamma_{0.5}$ waarvoor $F_X(\gamma_{0.5}) = P(X \leq \gamma_{0.5}) \geq \frac{1}{2}$ en $P(X \geq \gamma_{0.5}) \geq \frac{1}{2}$
Continue X: $\gamma_{0.5}$ waarvoor $F_X(\gamma_{0.5}) = \int_{-\infty}^{\gamma_{0.5}} f_X(x)dx = \int_{\gamma_{0.5}}^{+\infty} f_X(x)dx = \frac{1}{2}$

Kwantielen, percentielen en kwartielen:
p-de kwantielwaarde of $(100 \times p)$-ste percentiel γ_p van continue X:
$p = \int_{-\infty}^{\gamma_p} f_X(x)dx = F_X(\gamma_p)$
Eerste kwartiel=$\gamma_{0.25}$, tweede kwartiel=$\gamma_{0.5}$, derde kwartiel=$\gamma_{0.75}$

Pearson's scheefheidscoëfficiënt :
$SP^{pop} = \frac{3(\mu_X - \gamma_{0.5})}{\sigma_X} \in [-3, +3]$

Scheefheidscoëfficiënt:
scheefheidscoëfficiënt $= \frac{E\{(X - \mu_X)^3\}}{\sigma_X^3}$

Hoofdstuk 7: Belangrijke discrete kansverdelingen

Discreet uniforme verdeling:
$p_X(x) = \frac{1}{k}, \quad x = x_1, \ldots, x_k$

Bernoulli verdeling:
$p_X(x;\pi) = \pi^x(1-\pi)^{1-x}, \quad x = 0, 1$
$\mu_X = E(X) = \pi, \sigma_X^2 = var(X) = \pi(1-\pi)$

Binomiale verdeling:
$X \sim bin(n;\pi)$: $p_X(x;n,\pi) = \frac{n!}{x!(n-x)!}\pi^x(1-\pi)^{n-x}, \quad x = 0, 1, 2, \ldots, n$
$\mu_X = E(X) = n\pi, \sigma_X^2 = var(X) = n\pi(1-\pi)$

Poissonverdeling:
$X \sim \text{Poisson}(\lambda) : p_X(x;\lambda) = \frac{e^{-\lambda}\lambda^x}{x!}, \quad x = 0, 1, 2, \ldots$
$E(X) = \lambda, var(X) = \lambda$

Hoofdstuk 8: Belangrijke continue kansdichtheden

Continu uniforme dichtheid:

$$f_X(x;\alpha,\beta) = \begin{cases} \frac{1}{\beta-\alpha}, & \alpha \leq x \leq \beta, \\ 0, & \text{elders,} \end{cases} \qquad F_X(x;\alpha,\beta) = \begin{cases} 0, & x < \alpha, \\ \frac{x-\alpha}{\beta-\alpha}, & \alpha \leq x \leq \beta, \\ 1, & x > \beta \end{cases}$$

$\mu_X = E(X) = \frac{\alpha+\beta}{2}, \sigma_X^2 = var(X) = \frac{(\beta-\alpha)^2}{12}$

Normale dichtheid:

$X \sim N(\mu,\sigma^2) : f_X(x;\mu,\sigma) = \frac{1}{\sigma\sqrt{2\pi}}\, e^{-\frac{(x-\mu)^2}{2\sigma^2}}, \quad -\infty < x < +\infty$
$\mu_X = \mu, \sigma_X^2 = \sigma^2$

Standaardnormale verdeling $Z \sim N(0,1) : Z = \frac{X-\mu}{\sigma}$

$f_Z(z) = \frac{1}{\sqrt{2\pi}}\, e^{-\frac{z^2}{2}}, \quad -\infty < z < +\infty$

Stelling:
Een lineaire functie $Y = aX + b$ van een normaal verdeelde kansvariabele X met verwachte waarde μ en variantie σ^2 is een nieuwe normaal verdeelde kansvariabele met gemiddelde $E(Y) = a\mu + b$ en $var(Y) = a^2\sigma^2$

Hoofdstuk 9: Multivariate kansvariabelen

Gezamenlijke kansverdeling:
Discrete X, Y: $p_{XY}(x,y) = P(X = x, Y = y) = P\{(X = x) \cap (Y = y)\}$

Marginale kansverdeling:
$p_X(x) = \sum_y p_{XY}(x,y)$ en $p_Y(y) = \sum_x p_{XY}(x,y)$

Onafhankelijkheid:
X en Y zijn onafhankelijk indien $p_{XY}(x,y) = p_X(x)p_Y(y)$ voor elke (x,y)

Voorwaardelijke kansverdeling:
$p_{X|Y}(x|y) = \frac{p_{XY}(x,y)}{p_Y(y)}$ en $p_{Y|X}(y|x) = \frac{p_{XY}(x,y)}{p_X(x)}$

Hoofdstuk 10: Covariantie, correlatie en variantie van lineaire functies

Covariantie:
Voor discrete X,Y: $\sigma_{XY} = cov(X,Y) = E[(X-\mu_X)(Y-\mu_Y)]$
$= \sum_x\sum_y(x-\mu_X)(y-\mu_Y)p_{XY}(x,y) = E(XY) - \mu_X\mu_Y$

Verwachte waarde van functies van meerdere kansvariabelen:
voor discrete X, Y: $E[g(X,Y)] = \sum_{(x,y)\in D} g(x,y)p_{XY}(x,y)$

Correlatie:
$\rho_{XY} = corr(X,Y) = \frac{\sigma_{XY}}{\sigma_X\sigma_Y}$

Stelling:
Indien X en Y onafhankelijke kansvariabelen zijn, dan is $\sigma_{XY} = 0$ en $\rho_{XY} = 0$

Stelling:
$var(aX + bY + c) = a^2 var(X) + b^2 var(Y) + 2ab\, cov(X,Y)$

Hoofdstuk 11: Het schatten van populatieparameters

Zuivere of onvertekende schatter:
$E(\hat{\theta}) = \theta$
Vertekening: $V(\hat{\theta}) = |E(\hat{\theta}) - \theta|$

Relatieve efficiëntie:
relatieve efficiëntie van $\hat{\theta}_2$ t.o.v. $\hat{\theta}_1$: $var(\hat{\theta}_1)/var(\hat{\theta}_2)$

Gemiddelde gekwadrateerde afwijking:
$GGA(\hat{\theta}) = var(\hat{\theta}) + [V(\hat{\theta})]^2$

Verdeling steekproefgemiddelde:
$E(\overline{X}) = \mu$, $\sigma^2_{\overline{X}} = var(\overline{X}) = \frac{\sigma^2}{n}$ en $\sigma_{\overline{X}} = \frac{\sigma}{\sqrt{n}}$
$X \sim N(\mu, \sigma^2) \Rightarrow \overline{X} \sim N(\mu, \sigma^2/n)$
$X \not\sim N(\mu, \sigma^2)$ en $n \geq 30$: Centrale limietstelling $\Rightarrow \overline{X} \sim N(\mu, \sigma^2/n)$

Centrale limietstelling:
Indien $X_1, X_2, \ldots, X_n$ onafhankelijke kansvariabelen zijn met verwachte waarde $E(X_i) = \mu$ en variantie $var(X_i) = \sigma^2$, dan geldt onder heel algemene voorwaarden en voor een voldoende grote waarde n dat

1. de nieuwe kansvariabele $\overline{X} = \frac{\sum_{i=1}^n X_i}{n}$ benaderend normaal verdeeld is met gemiddelde μ en variantie $\frac{\sigma^2}{n}$,
2. en dus dat de nieuwe kansvariabele $\frac{\overline{X}-\mu}{\frac{\sigma}{\sqrt{n}}}$ benaderend standaardnormaal verdeeld is.

Verdeling steekproefproportie
Stelling:
Indien $X_1, X_2, \ldots, X_n$ onafhankelijke kansvariabelen zijn met als enige mogelijke uitkomsten 0 (faling) of 1 (succes), en met kans op succes gelijk aan π, dan geldt voor een voldoende grote waarde n ($n\pi > 5$ en $n(1-\pi) > 5$) dat

1. de steekproefproportie $\hat{P} = \frac{\sum_{i=1}^n X_i}{n}$ benaderend normaal verdeeld is met verwachte waarde π en variantie $\pi(1-\pi)/n$,
2. dus dat de nieuwe kansvariabele $\frac{\hat{P}-\pi}{\sqrt{\frac{\pi(1-\pi)}{n}}}$ benaderend standaardnormaal verdeeld is.

Verdeling steekproefvariantie S^2:
$S^2 = \frac{1}{n-1}\sum_{i=1}^n (X_i - \overline{X})^2$ en $E(S^2) = \sigma^2$
Stelling:
Indien $X_1, X_2, \ldots, X_n$ onafhankelijke, normaal verdeelde kansvariabelen zijn met variantie σ^2 dan geldt (voor elke n) dat $\frac{(n-1)S^2}{\sigma^2}$ χ^2-verdeeld is met $n-1$ vrijheidsgraden.

Chi-kwadraat verdeling:
$X \sim \chi^2_k \Leftrightarrow X = \sum_{i=1}^k X_i^2$ waarbij de X_i standaardnormaal verdeeld zijn en onafhankelijk
$f_X(x;k) = \frac{x^{\frac{k}{2}-1}e^{-x/2}}{\Gamma(\frac{k}{2})2^{\frac{k}{2}}}$, voor $x > 0$ en $E(X) = k$ en $var(X) = 2k$

Hoofdstuk 12: Intervalschatters

Student's t-verdeling:
$T \sim t_k \Leftrightarrow T = \frac{Z}{\sqrt{\frac{X}{k}}}$ waarbij $X \sim \chi^2_k$, $Z \sim N(0,1)$ en X en Z onafhankelijk
$f(t;k) = \frac{\Gamma\left(\frac{k+1}{2}\right)}{\Gamma\left(\frac{k}{2}\right)\sqrt{k\pi}}\left(1+\frac{t^2}{k}\right)^{-\frac{k+1}{2}}$ en $E(X) = 0$.

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

- $X \sim N(\mu, \sigma^2)$, σ^2 bekend: $Z = \frac{\overline{X}-\mu}{\sigma/\sqrt{n}} \sim N(0,1)$
  BI: $[\overline{X} - z_{\frac{\alpha}{2}} \cdot \frac{\sigma}{\sqrt{n}}; \overline{X} + z_{\frac{\alpha}{2}} \cdot \frac{\sigma}{\sqrt{n}}]$
- $X \sim N(\mu, \sigma^2)$, σ^2 onbekend: $T = \frac{\overline{X}-\mu}{S/\sqrt{n}} \sim t_{n-1}$
  BI: $[\overline{X} - t_{\frac{\alpha}{2};n-1} \cdot \frac{S}{\sqrt{n}}; \overline{X} + t_{\frac{\alpha}{2};n-1} \cdot \frac{S}{\sqrt{n}}]$
- $n \geq 30$, σ^2 bekend: $Z = \frac{\overline{X}-\mu}{\sigma/\sqrt{n}} \sim N(0,1)$
  BI: $[\overline{X} - z_{\frac{\alpha}{2}} \cdot \frac{\sigma}{\sqrt{n}}; \overline{X} + z_{\frac{\alpha}{2}} \cdot \frac{\sigma}{\sqrt{n}}]$
- $n \geq 30$, σ^2 onbekend: $Z = \frac{\overline{X}-\mu}{S/\sqrt{n}} \sim N(0,1)$
  BI: $[\overline{X} - z_{\frac{\alpha}{2}} \cdot \frac{S}{\sqrt{n}}; \overline{X} + z_{\frac{\alpha}{2}} \cdot \frac{S}{\sqrt{n}}]$

$(1-\alpha) \times 100\%$ **betrouwbaarheidsinterval voor σ^2:**

- $X \sim N(\mu, \sigma^2)$: $\frac{(n-1)S^2}{\sigma^2} \sim \chi^2_{n-1}$
  BI: $[\frac{(n-1)S^2}{\chi^2_{\alpha/2;n-1}}; \frac{(n-1)S^2}{\chi^2_{1-\alpha/2;n-1}}]$

$(1-\alpha) \times 100\%$ **betrouwbaarheidsinterval voor π:**

- $n\hat{p} > 5$ en $n(1-\hat{p}) > 5$: $Z = \frac{\hat{P}-\pi}{\sqrt{\pi(1-\pi)/n}} \sim N(0,1)$
  BI: $[\hat{P} - z_{\frac{\alpha}{2}} \cdot \sqrt{\frac{\hat{P}(1-\hat{P})}{n}}; \hat{P} + z_{\frac{\alpha}{2}} \cdot \sqrt{\frac{\hat{P}(1-\hat{P})}{n}}]$

Hoofdstuk 13: Het toetsen van hypothesen

Hoofdstuk 14: Hypothesetoetsen voor één populatie

Kwantieldiagram:
Puntenwolk van $(\mu + z_{1-cf_i}\sigma, x_i)$ met $cf_i = \frac{j-0.5}{n}$ of $cf_i^* = \frac{j-0.375}{n+0.25}$

Hoofdstuk 15: Hypothesetoetsen voor twee populaties

Fischer's F-verdeling
$X \sim F_{\nu_1,\nu_2} \Leftrightarrow X = \frac{X_1/\nu_1}{X_2/\nu_2}$ waarbij $X_1 \sim \chi^2_{\nu_1}, X_2 \sim \chi^2_{\nu_2}$ en X_1 en X_2 onafhankelijk

Hoofdstuk 16: Hypothesetoets voor meer dan twee populatiegemiddeldes

One-way ANOVA
$SST = SSE + SSA$
$SST = \sum_{i=1}^{g} \sum_{j=1}^{n_i} (X_{ij} - \overline{X})^2$
$SSE = \sum_{i=1}^{g} \sum_{j=1}^{n_i} (X_{ij} - \overline{X}_i)^2$
$SSA = \sum_{i=1}^{g} n_i(\overline{X}_i - \overline{X})^2$
$MSE = \frac{SSE}{n-g}$ en $E(MSE) = \sigma^2$
$MSA = \frac{SSA}{g-1}$ en $E(MSA) = \sigma^2 + \frac{\sum_{i=1}^{g} n_i(\mu_i-\mu)^2}{g-1}$

Stelling:
Indien de nulhypothese dat $\mu_1 = \mu_2 = \cdots = \mu_g$ juist is, en indien de g bestudeerde populaties normaal verdeeld zijn met eenzelfde variantie σ^2, dan zijn zowel $\frac{SSE}{\sigma^2}$ als $\frac{SSA}{\sigma^2}$ onafhankelijke χ^2-verdeelde kansvariabelen met respectievelijk $n-g$ en $g-1$ vrijheidsgraden en bijgevolg is $F = \frac{MSA}{MSE} \sim F_{g-1,n-g}$

Hoofdstuk 17: Lineaire regressie

Regressiemodel
$Y \sim N(\beta_0 + \beta_1 x, \sigma^2)$

Kleinste kwadraten methode
De kleinste kwadraten methode zoekt de recte, $Y = \hat{\beta}_0 + \hat{\beta}_1 x$, waarvoor de som $S^2 = \sum_{i=1}^{n}(y_i - \beta_0 - \beta_1 x_i)^2$ minimaal is.
$\hat{\beta}_0 = \overline{y} - \hat{\beta}_1 \overline{x}$
$\hat{\beta}_1 = \frac{\sum_{i=1}^{n}(y_i-\overline{y})(x_i-\overline{x})}{\sum_{i=1}^{n}(x_i-\overline{x})^2} = \frac{s_{XY}}{s_X^2} = r_{XY}\frac{s_Y}{s_X}$

$\hat{\sigma}^2 = \frac{1}{n-2}\sum_{i=1}^{n} r_i^2$ met $r_i = y_i - \hat{\beta}_0 - \hat{\beta}_1 x_i$

Eigenschappen van de kleinste kwadraten schatters
$var\left(\hat{\beta}_0\right) = \sigma^2\left(\frac{1}{n} + \frac{\overline{x}^2}{(n-1)s_X^2}\right)$
$var\left(\hat{\beta}_1\right) = \frac{\sigma^2}{(n-1)s_X^2}$
$cov\left(\hat{\beta}_0, \hat{\beta}_1\right) = -\frac{\overline{x}\sigma^2}{(n-1)s_X^2}$
$E\left(\hat{\beta}_0\right) = \beta_0$
$E\left(\hat{\beta}_1\right) = \beta_1$

Verdelingen van de schattingen
$\hat{\beta}_0 \sim N\left(\beta_0, \sigma^2\left(\frac{1}{n} + \frac{\overline{x}^2}{(n-1)s_X^2}\right)\right)$
$\hat{\beta}_1 \sim N\left(\beta_1, \frac{\sigma^2}{(n-1)s_X^2}\right)$
$\frac{\hat{\beta}_1-\beta_1}{\frac{\sigma}{\sqrt{(n-1)s_X^2}}} \sim N(0,1)$
$\frac{\hat{\beta}_1-\beta_1}{\frac{\hat{\sigma}}{\sqrt{(n-1)s_X^2}}} \sim t_{n-2}$
$\frac{\hat{\beta}_0+\hat{\beta}_1 x_0-(\beta_0+\beta_1 x_0)}{\hat{\sigma}\sqrt{\frac{1}{n}+\frac{(x_0-\overline{x})^2}{(n-1)s_X^2}}} \sim t_{n-2}$
$\frac{Y_m-\hat{y}_m}{\hat{\sigma}\sqrt{1+\frac{1}{n}+\frac{(x_m-\overline{x})^2}{(n-1)s_X^2}}} \sim t_{n-2}$
Op basis van deze verdelingen kan men betrouwbaarheidsintervallen opstellen en hypothesetoetsen uitvoeren voor de parameters β_0, β_1, voor de gemiddelde respons bij gegeven x_0 en voor de individuele Y_m bij een gegeven x_m.

Toetsen voor 1 steekproef						
Onderwerp	Hypothese	Min. schaal	Voorwaarden	Steekproefvariabele	Toetsingsgrootheid onder H_0	Naam
populatie-gemiddelde μ	a) $H_0 : \mu = \mu_0$ $H_a : \mu \neq \mu_0$ b) $H_0 : \mu = \mu_0$ $H_a : \mu > \mu_0$ c) $H_0 : \mu = \mu_0$ $H_a : \mu < \mu_0$	interval	$n < 30$: i) $X \sim N(\mu, \sigma^2)$, σ^2 bekend; ii) $X \sim N(\mu, \sigma^2)$, σ^2 onbekend $n \geq 30$: iii) σ^2 bekend; iv) σ^2 onbekend	gemiddelde $\overline{X} = \frac{\sum_{i=1}^n X_i}{n}$ variantie $S^2 = \frac{\sum_{i=1}^n (X_i - \overline{X})^2}{n-1}$	i) $Z = \frac{\overline{X} - \mu_0}{\sigma/\sqrt{n}} \sim N(0,1)$ ii) $T = \frac{\overline{X} - \mu_0}{S/\sqrt{n}} \sim t_{n-1}$ iii) $Z = \frac{\overline{X} - \mu_0}{\sigma/\sqrt{n}} \sim N(0,1)$ iv) $T = \frac{\overline{X} - \mu_0}{S/\sqrt{n}} \sim t_{n-1}$	Z- of T-toets voor 1 gemiddelde
populatie-mediaan Me	a) $H_0 : Me = Me_0$ $H_a : Me \neq Me_0$ b) $H_0 : Me = Me_0$ $H_a : Me > Me_0$ c) $H_0 : Me = Me_0$ $H_a : Me < Me_0$	ordinaal		$Me_0^> = (\#\text{ data} > Me_0)$ $Me_0^< = (\#\text{ data} < Me_0)$	a) $S = Max(Me_0^>, Me_0^<)$ b) $S = Me_0^>$ c) $S = Me_0^<$ $S \sim bin(n, \frac{1}{2})$	Tekentoets
populatie-proportie π	a) $H_0 : \pi = \pi_0$ $H_a : \pi \neq \pi_0$ b) $H_0 : \pi = \pi_0$ $H_a : \pi > \pi_0$ c) $H_0 : \pi = \pi_0$ $H_a : \pi < \pi_0$	nominaal (0-1)	$n\hat{p} > 5$ en $n(1-\hat{p}) > 5$	proportie $\hat{P}$	$Z = \frac{\hat{P} - \pi_0}{\sqrt{\pi_0(1-\pi_0)/n}} \sim N(0,1)$	Z-toets voor proportie
populatie-variantie σ^2	a) $H_0 : \sigma^2 = \sigma_0^2$ $H_a : \sigma^2 \neq \sigma_0^2$ b) $H_0 : \sigma^2 = \sigma_0^2$ $H_a : \sigma^2 > \sigma_0^2$ c) $H_0 : \sigma^2 = \sigma_0^2$ $H_a : \sigma^2 < \sigma_0^2$	interval	$X \sim N(\mu, \sigma^2)$	variantie S^2	$\chi = \frac{(n-1)S^2}{\sigma_0^2} \sim \chi^2_{n-1}$	χ^2-toets voor variantie
populatie-kansverdeling	H_0 : pop. kansverdeling is ... (cel freq. E_i=...) H_a : niet H_0	kwalitatief	$\forall i : O_i \geq 1$ p=# geschatte parameters	cel freq. O_i	$\chi = \sum_{i=1}^k \frac{(O_i - E_i)^2}{E_i} \sim \chi^2_{k-1-p}$	χ^2-toets voor verdeling
normaliteit van populatie-kansdichtheid	H_0 : pop. normaal verdeeld (met rel. cum. verd. F_X) H_a : niet H_0	kwantitatief		correlatiecoëfficiënt Q-Q-diagram		Shapiro-Wilk toets voor normaliteit

Toetsen voor 2 onafhankelijke steekproeven

Onderwerp	Hypothese	Min. schaal	Voorwaarden	Steekproefvariabele	Toetsingsgrootheid onder H_0	Naam
twee populatie-gemiddelden, onafhankelijk	a) $H_0: \mu_1 - \mu_2 = \Delta_0$ $H_a: \mu_1 - \mu_2 \neq \Delta_0$ b) $H_0: \mu_1 - \mu_2 = \Delta_0$ $H_a: \mu_1 - \mu_2 > \Delta_0$ c) $H_0: \mu_1 - \mu_2 = \Delta_0$ $H_a: \mu_1 - \mu_2 < \Delta_0$	interval	$n_1 < 30$ en/of $n_2 < 30$ i) $X_1, X_2 \sim N(\mu_i, \sigma_i^2)$, σ_1^2, σ_2^2 bekend ii) $X_1, X_2 \sim N(\mu_i, \sigma_i^2)$, σ_1^2, σ_2^2 onbekend, $\sigma_1^2 = \sigma_2^2$ iii) $X_1, X_2 \sim N(\mu_i, \sigma_i^2)$, σ_1^2, σ_2^2 onbekend, $\sigma_1^2 \neq \sigma_2^2$ $n_1 \geq 30$ en $n_2 \geq 30$ iv) σ_1^2, σ_2^2 bekend v) σ_1^2, σ_2^2 onbekend	gemiddelden $\overline{X}_1, \overline{X}_2$ varianties S_1^2, S_2^2	i) $Z = \frac{\overline{X}_1 - \overline{X}_2 - \Delta_0}{\sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}} \sim N(0,1)$ ii) $T = \frac{\overline{X}_1 - \overline{X}_2 - \Delta_0}{\sqrt{\frac{s_p^2}{n_1} + \frac{s_p^2}{n_2}}} \sim t_{n_1+n_2-2}$ $s_p^2 = \frac{(n_1-1)S_1^2 + (n_2-1)S_2^2}{n_1+n_2-2}$ iii) $T' = \frac{\overline{X}_1 - \overline{X}_2 - \Delta_0}{\sqrt{\frac{S_1^2}{n_1} + \frac{S_2^2}{n_2}}} \sim t_\nu$ $\nu \simeq \frac{\left(\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}\right)^2}{\frac{(s_1^2/n_1)^2}{n_1-1} + \frac{(s_2^2/n_2)^2}{n_2-1}}$ iv) $Z = \frac{\overline{X}_1 - \overline{X}_2 - \Delta_0}{\sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}} \sim N(0,1)$ v) $T = \frac{\overline{X}_1 - \overline{X}_2 - \Delta_0}{\sqrt{\frac{S_1^2}{n_1} + \frac{S_2^2}{n_2}}} \sim t_{n_1+n_2-2}$	Z-toets of T-toets voor 2 gemidd.
twee populatie-locaties, onafhankelijk	a) H_0 : Locatie 1 = Locatie 2 H_a : Locatie 1 $\neq$ Locatie 2 b) H_0 : Locatie 1 = Locatie 2 H_a : Locatie 1 > Locatie 2 c) H_0 : Locatie 1 = Locatie 2 H_a : Locatie 1 < Locatie 2	ordinaal	$n_1 \leq n_2$	rangnummers over beide steekproeven samen	$T_1 = \sum$ rangnrs steekpr. 1 $\sim$ Tabel	Wilcoxon rangsom
twee populatie-proporties, onafhankelijk	a) $H_0: \pi_1 - \pi_2 = 0$ $H_a: \pi_1 - \pi_2 \neq 0$ b) $H_0: \pi_1 - \pi_2 = 0$ $H_a: \pi_1 - \pi_2 > 0$ c) $H_0: \pi_1 - \pi_2 = 0$ $H_a: \pi_1 - \pi_2 < 0$	nominaal (0-1)	$n_1\hat{p}_1 > 5$, $n_1(1-\hat{p}_1) > 5$ $n_2\hat{p}_2 > 5$, $n_2(1-\hat{p}_2) > 5$	proporties $\hat{P}_1, \hat{P}_2$	$Z = \frac{\hat{P}_1 - \hat{P}_2}{\sqrt{\overline{P}(1-\overline{P})(\frac{1}{n_1} + \frac{1}{n_2})}} \sim N(0,1)$ met $\overline{P} = \frac{n_1\hat{P}_1 + n_2\hat{P}_2}{n_1+n_2}$	Z-toets voor 2 proporties
twee populatie-varianties, onafhankelijk	a) $H_0: \sigma_1^2 = \sigma_2^2$ $H_a: \sigma_1^2 \neq \sigma_2^2$ b) $H_0: \sigma_1^2 = \sigma_2^2$ $H_a: \sigma_1^2 > \sigma_2^2$	interval	$X_i \sim N(\mu_i, \sigma_i^2)$ $s_1^2 > s_2^2$	varianties S_1^2, S_2^2	$F = \frac{S_1^2}{S_2^2} \sim F_{n_1-1, n_2-1}$	F-toets voor 2 varianties

Toetsen voor 2 gepaarde steekproeven						
Onderwerp	Hypothese	Min. schaal	Voorwaarden	Steekproefvariabele	Toetsingsgrootheid onder H_0	Naam
2 populatie-gemiddelden, gepaard	a) $H_0: \delta = \delta_0$, $H_a: \delta \neq \delta_0$ b) $H_0: \delta = \delta_0$, $H_a: \delta > \delta_0$ c) $H_0: \delta = \delta_0$, $H_a: \delta < \delta_0$	interval	$n < 30$: i) $D \sim N(\delta, \sigma_D^2)$, σ_D^2 bekend; ii) $D \sim N(\delta, \sigma_D^2)$, σ_D^2 onbekend $n \geq 30$: iii) σ_D^2 bekend; iv) σ_D^2 onbekend	$D = X - Y$ $\overline{D}$ S_D	i) $Z = \frac{\overline{D}-\delta_0}{\sigma_D/\sqrt{n}} \sim N(0,1)$ ii) $T = \frac{\overline{D}-\delta_0}{S_D/\sqrt{n}} \sim t_{n-1}$ iii) $Z = \frac{\overline{D}-\delta_0}{\sigma_D/\sqrt{n}} \sim N(0,1)$ iv) $T = \frac{\overline{D}-\delta_0}{S_D/\sqrt{n}} \sim t_{n-1}$	T-toets of Z-toets voor 2 gemiddelden (gepaard)
2 populatie-locaties, gepaard	a) H_0 : Locatie 1 = Locatie 2, H_a : Locatie 1 $\neq$ Locatie 2 b) H_0 : Locatie 1 = Locatie 2, H_a : Locatie 1 > Locatie 2 c) H_0 : Locatie 1 = Locatie 2, H_a : Locatie 1 < Locatie 2	ordinaal	verwijder verschillen gelijk aan nul	rangnrs$(X_i - Y_i)$ $T_+ = \sum$ pos. rangnrs $T_- = \sum$ | neg. rangnrs |	a) $\text{Min}(T_+, T_-)$ b) T_- c) T_+ voor $n > 50$: $Z = \frac{T_+ - n(n+1)/4}{\sqrt{n(n+1)(2n+1)/24}} \sim N(0,1)$	Wilcoxon rangteken

Toets voor $g \geq 3$ onafhankelijke steekproeven						
Onderwerp	Hypothese	Min. schaal	Voorwaarden	Steekproefvariabele	Toetsingsgrootheid onder H_0	Naam
g populatie gemiddelden, onafhankelijk	$H_0: \mu_1 = \mu_2 = \ldots \mu_g$ H_a : niet alle μ_i gelijk	interval	$X_i \sim N(\mu_i, \sigma^2)$ $\sigma_i^2 = \sigma^2 \quad \forall i$	$MSA = \frac{\sum_{i=1}^{g} n_i(\overline{X}_i - \overline{X})^2}{g-1}$ $MSE = \frac{\sum_{i=1}^{g}\sum_{j=1}^{n_i}(X_{ij} - \overline{X}_i)^2}{n-g}$	$F = \frac{MSA}{MSE} \sim F_{g-1,n-g}$	One-way ANOVA