TECHNISCHE HOGESCHOOL EINDHOVEN
Afdeling Algemene Wetenschappen
Onderafdeling der Wiskunde

Vraagstukken

behorende bij het college

TOEGEPASTE STATISTIEK

van

Prof. Dr. R. Doornbos

samengesteld door

J.Th.M. Wijnen

Voorjaar 1976

2231

Bijl. Mog

Technische Hogeschool Eindhoven

Onderafdeling der Wiskunde

Vraagstukken

behorende bij het college Toegepaste Statistiek
van prof. dr. R. Doornbos

Samengesteld door J. Th. M. Wijnen

Dictaatnr. 2.231 Prijs *f* 4,--

TECHNISCHE HOGESCHOOL EINDHOVEN

Onderafdeling der Wiskunde

VRAAGSTUKKEN

bij het college

TOEGEPASTE STATISTIEK

Voorjaar 1976

Inhoudsbeschrijving

Vraagstukken bij

TOEGEPASTE STATISTIEK

R. Doornbos en J.Th.M. Wijnen

voorjaar 1976

De vindplaats van de antwoorden onder "Antwoorden blz" is achterin het dictaat.
De nummering der bladzijden aldaar is gelijk aan die der vraagstukken

JdG, 24 Juli 2005

INHOUD.

1. Herhaling van Wiskunde 31-49.

1.1. Een man van 40 jaar koopt een lijfrente die 20 jaar later zal ingaan. Zijn vrouw is 38. Van alle mannen van 40 leeft 4/5 nog na 20 jaar en van alle vrouwen van 38 nog 9/10. Wat is de kans dat minstens één van beiden nog in leven is als men de lijfrente gaat uitkeren?

1.2. Uit een doos met N lootjes waaronder slechts 1 prijs, trekken achter elkaar n personen elk 1 lot. Als de prijs echter getrokken is, is het spel afgelopen. Moet het spel mèt of zonder teruglegging van het getrokken lot gebeuren opdat het eerlijk is voor alle n spelers? ($n \le N$.)

1.3. Hoe kan men met een onzuivere munt een boek onder 2 mensen verloten, zodat elk gelijke kans heeft?

1.4. In een cirkel wordt "at random" een koorde getrokken. Hoe groot is de kans dat deze koorde groter is dan de zijde van een ingeschreven gelijkzijdige driehoek? (Paradox van Bertrand; deze demonstreert dat men "random" nader moet definiëren.)

1.5. Een keuring van sigarettenmerken op nicotinegehalte werd uitgevoerd in 1958 en 1961. Naar grootte van de uitkomst gerangschikt werd gevonden:

in 1958: C, E, A, F, B, D
in 1961: C, B, H, D, G, A.

De merken A, B, C en D zijn dus in beide jaren onderzocht, de merken E, F, G en H slechts in één van beide jaren. De bepaling van het nicotinegehalte is met grote toevallige fouten behept en het is best mogelijk dat de gevonden volgorden zuiver toevallig zijn. Het valt echter op dat het merk C beide malen op de eerste plaats staat. Vraag: mag men nu op grond van deze resultaten concluderen ($\alpha = 0.05$) dat het merk C inderdaad een hoger nicotinegehalte bevat dan de overige merken?

1.6. Bewijs: $\sum_{i=0}^{n} \binom{n}{i}^2 = \binom{2n}{n}$.

1.7. Hoe groot is de kans dat uit een willekeurige groep van 25 mensen precies 3 paren op dezelfde dag jarig zijn? (De overige op verschillende dagen.)

1.8. Bereken voor een groep van 4 toevalscijfers de kans op:

a) 4 gelijke;

b) 3 gelijke;

c) 2 paren;

d) 1 paar;

e) alle verschillend.

1.9. In het Morse-alfabet worden letters aangegeven door een opeenvolging van punten en strepen, waarbij herhalingen zijn toegestaan. Hoeveel verschillende combinaties van punten en/of strepen kunt U van 5 tekens vormen?

1.10. Vaas 1 bevat 1 witte en 3 rode ballen, vaas 2 bevat 2 witte en 2 rode, vaas 3 bevat 4 witte en 3 rode ballen. Men kiest aselect een vaas en trekt er blindelings een bal uit. Deze bal is wit. Hoe groot is de kans dat deze uit vaas 1 afkomstig is?

1.11. Hoe groot is de kans dat een gezin van 4 kinderen bestaat uit 3 jongens en 1 meisje als gegeven is:

a) het gezin bevat tenminste 3 jongens;

b) de oudste 3 kinderen zijn jongens.

1.12. Bij een experiment kunnen 3 gebeurtenissen A, B en C optreden, die elkaar noch uitsluiten, noch een volledig stelsel vormen. Alle hebben dezelfde kans om op te treden. A is onafhankelijk van BC, maar B en C zijn afhankelijk: de voorwaardelijke kans op C onder voorwaarde B is 2× de onvoorwaardelijke kans op C. De kans op het gelijktijdig optreden van A, B en C is $\frac{1}{4}$. Bewijs dat uit het optreden van B volgt dat ook C optreedt en andersom.

1.13. Uit een vaas die 4 witte en 6 zwarte ballen bevat worden eerst 2 ballen aselect getrokken en opzij gelegd zonder naar de kleur te kijken. Daarna wordt een derde bal getrokken. Hoe groot is de kans dat

a) de derde bal wit is? Wat merkt U dus op?

b) de derde bal wit is, als gegeven is dat onder de opzij gelegde ballen minstens 1 witte is?

1.14. Wat is de kans dat in een worp met n zuivere dobbelstenen

a) geen "6" voorkomt?

b) juist één "6" voorkomt?

c) tenminste één "6" voorkomt?

d) geen "5" en geen "6" voorkomt?

e) geen "5" en juist één "6" voorkomt?

f) geen "5" en tenminste één "6" voorkomt?

g) tenminste één "5" en/of één "6" voorkomt?

h) tenminste één "5" en één "6" voorkomt?

i) juist één "5" en één "6" voorkomt?

1.15. Een stochastische grootheid $\underline{x}$ kan bij realisatie drie verschillende uitkomsten A, B en C geven met kansen:

$$P(\underline{x}=A) = p_A \,, \quad P(\underline{x}=B) = p_B \,, \quad P(\underline{x}=C) = p_C \,, \quad p_A + p_B + p_C = 1 \,.$$

a) Hoe groot is de voorwaardelijke kans op $\underline{x} = A$ wanneer gegeven is $x \neq C$?

b) Hoe groot is de kans dat van n realisaties er 2 een uitkomst B zullen opleveren en $(n-2)$ de uitkomst A ?

1.16. Bij de controle van een produktieproces meet men ieder uur aan een steekproef van n produkten een karakteristieke grootheid $\underline{x}$. Men heeft verder 2 grenzen G_1 en G_2 $(G_2 > G_1)$ vastgesteld en de volgende voorschriften gegeven:

Het produktieproces wordt onmiddellijk gestopt en bijgeregeld indien

a) tenminste één van de n waarnemingen groter is dan G_2, ŏf

b) tenminste twee van de n waarnemingen groter zijn dan G_1.

Indien één van de n waarnemingen groter is dan G_1 doch kleiner dan G_2 wordt direkt een twee steekproef van n stuks genomen. Het produktieproces wordt gestopt als in deze tweede steekproef minstens één waarneming groter is dan G_1.

Zij $p_1 = P(\underline{x} < G_1)$, $p_2 = P(G_1 < \underline{x} < G_2)$, $p_3 = P(\underline{x} > G_2)$.

Geef een formule met afleiding voor de kans dat de produktie bij een dergelijk steekproefonderzoek zal worden gestopt.

1.17. Hoe groot is de kans dat in een groep van 60 aselecte getallen tenminste één van de 10 cijfers ontbreekt

a) als het getal bestaat uit één cijfer,

b) als het getal bestaat uit twee cijfers?

1.18. Gegeven zijn n waarnemingsuitkomsten $x_1,\dots,x_n$ van een stochastische grootheid $\underline{x}$.

a) Toon aan dat $\sum_i (x_i - a)^2/(n-1)$ minimaal wordt voor $a = \bar{x}$.

b) Bewijs dat $\text{var}(a\underline{x}+b) = a^2 \text{ var } \underline{x}$.

c) Laat zien: als $y_i = x_i - d$ voor $i = 1,\dots,n$, dan is

$$s^2(y) = s^2(x) \;, \qquad (s^2(x) = \frac{1}{n-1}\sum_{i=1}^{n}(x_i-\bar{x})^2\,) .$$

1.19. Zij $\underline{s}^2 := \sum_{i=1}^{n} \frac{(\underline{x}_i - \bar{\underline{x}})^2}{n-1}$. Bewijs

a) $\mathcal{E}\underline{s}^2 = \sigma^2$, d.w.z. $\underline{s}^2$ is een zuivere schatter voor σ^2.

b) $\mathcal{E}\underline{s} < \sigma$, d.w.z. $\underline{s}$ is géén zuivere schatter voor σ.

1.20. Bewijs dat voor 2 stochastieken $\underline{x}$ en $\underline{y}$ geldt dat de standaardafwijking van de som hoogstens gelijk is aan de som der standaardafwijkingen:

$$\sigma(\underline{x}+\underline{y}) \le \sigma(\underline{x}) + \sigma(\underline{y}) \;.$$

Geldt dit ook voor de variantie?

1.21. Twee even grote partijen kogels worden bij elkaar gevoegd tot één grote gemengde partij. De diameters van de kogels hebben resp. dichtheden f_1 en f_2 met verwachtingen $\mu_1 = 2$ en $\mu_2 = 2.2$ en standaardafwijkingen $\sigma_1 = 0.2$ en $\sigma_2 = 0.3$.

a) Bereken μ en σ^2 van de gemengde partij.

b) Als f_1 en f_2 beide normaal zijn, is de gemengde partij dan ook normaal verdeeld?

1.22. Wanneer geldt voor 2 stochastieken $\underline{x}$ en $\underline{y}$

a) $\text{var}(\underline{x}+\underline{y}) = \text{var } \underline{x} + \text{var } \underline{y}$,

b) $\sigma(\underline{x}+\underline{y}) = \sigma(\underline{x}) + \sigma(\underline{y})$?

1.23. $\underline{x}_1,\ldots,\underline{x}_n$ zijn n onafhankelijke toevalsvariabelen, elk met dezelfde verdelingsfunctie $F(x)$. Bepaal de verdelingsfunctie van:

a) $\underline{y} = \max(\underline{x}_1,\ldots,\underline{x}_n)$,

b) $\underline{z} = \min(\underline{x}_1,\ldots,\underline{x}_n)$.

1.24. $\underline{x}$ en $\underline{y}$ zijn onafhankelijke standaardrechthoekige stochastieken. Gevraagd de verdeling van $\underline{x}+\underline{y}$. Teken deze. Bereken μ, σ^2 en $P(\underline{x}+\underline{y} < 1\frac{1}{2})$.

1.25. $\underline{x}$ is een stochastische variabele met een standaard homogene verdeling:

$$f(x) = 1 \quad \text{voor } 0 \leq x \leq 1\ ,$$
$$f(x) = 0 \quad \text{elders.}$$

$\{\underline{x}_i: i = 1,\ldots,n\}$ is een aselecte steekproef uit deze verdeling en $\{\underline{x}_{(i)}: i = 1,\ldots,n\}$ is deze zelfde steekproef gerangschikt naar opklimmende waarden.

a) Wat is de verdeling van $\underline{x}(n)$ en van $\underline{x}(n-1)$?

b) Bereken $\mathcal{E}\underline{x}_{(n)}$, $\mathcal{E}\underline{x}_{(n-1)}$, $\mathrm{var}(\underline{x}_{(n)})$ en $\mathrm{var}(\underline{x}_{(n-1)})$.

1.26. Bereken de verwachting en de variantie van het aantal jongens, $\underline{x}$, in een gezin van 5 kinderen onder de voorwaarden

a) dat het oudste kind een jongen is,

b) dat het gezin tenminste één jongen telt.

Kunt U het verschil tussen de uitkomsten a) en b) op eenvoudige wijze verklaren?

1.27. Gegeven:

$$f(x,y) = k \quad \text{voor } x > 0,\ y > 0,\ x+y < 1\ ,$$
$$f(x,y) = 0 \quad \text{elders.}$$

a) Bepaal k.

b) Wat zijn de marginale verdelingen van $\underline{x}$ en $\underline{y}$?

c) Bereken $\mathcal{E}\underline{x}$, $\mathcal{E}\underline{y}$, var $\underline{x}$, var $\underline{y}$ en $\mathrm{cov}(\underline{x},\underline{y})$.

d) Wat is de voorwaardelijke verdeling van $\underline{x}$ voor $\underline{y} > 1/3$?

1.28. Gegeven: $\underline{x} \simeq PS(\mu = 1.5)$.

a) Bereken $\mathcal{E}(\underline{x} \mid \underline{x} \geq 3)$.

b) Bereken $\mathrm{var}(\underline{x} \mid \underline{x} \geq 3)$.

1.29. Zij $\underline{y} = \sum_{i=1}^{i=\underline{n}} \underline{x}_i$, waarin $\{\underline{x}_i : i = 1,\ldots,\underline{n}\}$ een aselecte steekproef is, en

$$\mathcal{E}\underline{x}_i = \mu_0 \;, \quad \mathrm{var}\ \underline{x}_i = \sigma_0^2 \;,$$

$$\mathcal{E}\underline{n} = \mu_1 \;, \quad \mathrm{var}\ \underline{n} = \sigma_1^2 \;.$$

Bepaal $\mathcal{E}\underline{y}$ en var $\underline{y}$.

1.30. a) $\underline{x}$ en $\underline{y}$ zijn onafhankelijk, $\sigma_x = 5$, $\sigma_y = 3$. Stel $\underline{z} = \underline{x} - \underline{y}$. Bereken var $\underline{z}$, $\mathrm{cov}(\underline{x},\underline{z})$ en $\mathrm{cov}(\underline{y},\underline{z})$.

b) $\underline{x}_1,\ldots,\underline{x}_n$ zijn ongecorreleerde stochastische variabelen met dezelfde verdeling. Bewijs dat $\underline{x}_i - \bar{\underline{x}}$ en $\bar{\underline{x}}$ ongecorreleerd zijn.

c) $\underline{x}$ en $\underline{y}$ zijn onafhankelijk. Er worden n paren onafhankelijke waarnemingen (x_i,y_i) verricht. Men wil het gemiddelde $\mathcal{E}(\underline{x}\underline{y})$ van $\underline{x}\underline{y}$ schatten. Toon aan dat $\bar{\underline{x}}\bar{\underline{y}}$ hiertoe "beter" (d.w.z. kleinere variantie heeft) is dan $\frac{1}{n}\sum_{i=1}^{n} \underline{x}_i\underline{y}_i$.

1.31. Gegeven de getallen 1,2,...,N. Men trekt hieruit aselect een steekproef van n stuks: $\underline{x}_1,\ldots,\underline{x}_n$. Stel $\underline{y} = \sum_{i=1}^{n} \underline{x}_i$. Bereken $\mathcal{E}\underline{y}$ en var $\underline{y}$ indien

a) trekking geschiedt met teruglegging;

b) trekking geschiedt zonder teruglegging.

2. Bewerking van waarnemingen. Beschrijvende statistiek.

2.1. Gegeven de serie waarnemingen:

15.813; 15.705; 15.748; 15.801; 15.720; 15.743.

a) Bereken $\bar{x}$ en s.

b) Codeer de waarnemingen en bereken weer $\bar{x}$ en s.

c) Geef een schatting van s met behulp van de range R en tabel S.C. 8.3 (kolom 2).

d) Rond de waarnemingen toelaatbaar af en bereken $\bar{x}$ en s. (Vergelijk de resultaten met die onder a), b) en c)!)

2.2. Een serie van 7 waarnemingen gaf de volgende uitkomsten:

5.372; 5.249; 5.280; 5.333; 5.262; 5.301; 5.288.

a) Bereken $\bar{x}$ en s, zonodig na een doelmatige afronding.

b) Bepaal ook een schatting s_R van σ uit de spreidingsbreedte R.

c) In hoeverre berusten deze berekeningen op bepaalde onderstellingen omtrent de wijze waarop de uitkomsten zijn ontstaan?

2.3. Tabel S.C. 8.3 geeft $s = A_n\bar{R}$, A5 = 0.430. Ga dit na door bijv. 50 steekproefjes van 5 stuks uit een N(0,1)-verdeling te simuleren met behulp van tabel S.C. 8.6.

2.4. Van 60 hartpatiënten is de diastolische bloeddruk (in mg Hg) bepaald:

72	62	96	80	78	68	104	82	108	108	98	76
64	78	70	54	88	72	52	94	84	58	58	72
78	54	66	78	68	79	74	90	74	60	72	70
64	68	76	79	77	62	76	78	70	74	72	60
80	58	78	78	96	74	78	106	68	76	54	76

Construeer een frequentietabel en een histogram. Bereken gemiddelde en standaardafwijking.

2.5. Zij $\underline{x}$ = totaal aantal met 5 dobbelstenen geworpen ogen.

a) Bereken $\mathcal{E}\underline{x}$ en var $\underline{x}$.

b) Genereer zelf met behulp van tabel S.C. 8.4 (aselecte getallen) 50 $\underline{x}$ waarden.

c) Construeer een frequentietabel en een histogram.

d) Bereken $\bar{x}$ en s uit de frequentietabel en vergelijk de resultaten met die onder a).

2.6. Gegeven een steekproef van 10 waarnemingen uit een normale verdeling:

12.05; 12.71; 12.25; 12.40; 12.15; 12.94; 12.00; 12.40; 12.49; 12.33.

a) Geef een schatting s_R van σ uit de range R.

b) Rond de waarnemingen eventueel af en codeer ze.

c) Bereken $\bar{x}$ en s zowel uit de gecodeerde oorspronkelijke als uit de gecodeerde afgeronde waarnemingen. Rond $\bar{x}$ en s toelaatbaar af en vergelijk beide resultaten.

2.7. In onderstaande tabel staan 50 getallen die als volgt werden verkregen. Geworpen werd met 3 dobbelstenen, één zwarte en twee witte. De ogen werden opgeteld, waarbij de ogen van de zwarte dubbel werden geteld. De 50 scores waren:

12	18	11	6	12	12	16	17	10	17
14	18	20	11	14	16	16	15	21	21
13	15	13	14	23	14	10	7	7	16
17	14	17	10	15	23	6	13	18	9
17	14	13	17	12	17	16	15	20	5

a) Welke zijn verwachting en variantie van deze score?

b) Construeer een frequentietabel en bereken daaruit de schattingen $\bar{x}$, s^2 en s.

c) Geef ook een schatting s_R voor σ uit een gemiddelde range $\bar{R}$.

2.8. Onderstaande tabel geeft metingen van een belangrijke maat aan 30 bakeliten knoppen:

5.25	5.35	5.31	5.38	5.29	5.37
5.38	5.34	5.41	5.36	5.33	5.31
5.35	5.28	5.33	5.40	5.30	5.31
5.30	5.30	5.35	5.37	5.32	5.38
5.39	5.29	5.28	5.33	5.32	5.37

a) Maak een schatting van σ met behulp van de range.

b) Trek uit deze 30 getallen aselect een steekproef van 6 stuks zonder teruglegging. Geef aan hoe de trekking wordt uitgevoerd.

c) Bereken uit deze steekproef $\bar{x}$ en s en rond deze juist af.

2.9. Gegeven zijn de volgende meetresultaten:

518	508	554	555	536	544	578	530	590	542
560	574	598	567	492	502	532	564	554	556
538	528	579	550	528	548	562	536	530	590
510	534	538	535	572	562	524	540	572	546
544	538	544	540	506	534	548	530	525	522

a) Maak een frequentietabel en een histogram.

b) Bereken $\bar{x}$ en s uit deze frequentietabel.

3. De normale verdeling: toetsen en betrouwbaarheidsintervallen.

3.1. Gegeven: $\underline{x} \simeq 50 + 7\underline{u}$. Bereken:

a) $P(\underline{x} > 60)$;

b) $P(42 < \underline{x} < 63)$;

c) $P(\underline{x} \leq 40)$.

d) Bepaal x, zodat $P(\underline{x} < x) = 0.025$.

e) Bepaal x, zodat $P(|\underline{x}-50| < x) = 0.85$.

3.2. Van een normaal verdeelde grootheid $\underline{x}$ met $\mu = 2$ en $\sigma = 3$ worden vier onafhankelijke waarnemingen verricht. Hoe groot is de kans dat de grootste van deze vier waarnemingen ≥ 5 is? En hoe groot is de kans dat de kleinste ≥ 5 is?

3.3. In een magazijn van 3.20 m. hoog ligt een groot aantal platte schijven, waarvan de dikte normaal is verdeeld met gemiddelde 12 cm en standaardafwijking 2 cm. Men heeft de gewoonte 25 schijven op elkaar te stapelen. Hoe groot is de kans dat het mislukt?

3.4. Een machine vult pakjes boter waarvan het nominale gewicht 250 gram is. Het gemiddelde vulgewicht kan worden ingesteld. Neem aan dat het ware gewicht van de pakjes normaal verdeeld is met een standaardafwijking van 1 gram. Op geregelde tijden wordt een steekproef van 4 pakjes gewogen om te controleren of de machine-instelling nog goed is. Wanneer het gemiddelde van deze steekproef meer dan 1 gram van het nominale gewicht afwijkt wordt de machine bijgeregeld.

a) Wat is de kans dat bij 10 controles minstens één keer wordt bijgeregeld als de machine telkens goed staat ingesteld?

b) Wat is de kans dat op grond van 1 steekproef van 4 stuks in de juiste richting wordt bijgeregeld als de instelling $\frac{1}{2}$ gram te hoog is?

c) Om een afwijking beter te kunnen vaststellen wordt de steekproefgrootte opgevoerd tot 9 pakjes. De kans op ten onrechte bijregelen (dus bij juiste instelling) wordt gelijk gehouden aan de oorspronkelijke waarde (zie bij steekproefgrootte 4). Bij welk verschil tussen steekproefgemiddelde en nominale waarde moet nu worden ingegrepen?

d) Hoe groot wordt in het onder c) beschreven geval de kans om een afwijking van de instelling van ½ gram te ontdekken, zodat de instelling verbeterd kan worden?

e) Hoe groot zou de steekproef minstens moeten zijn om de onder b) en d) berekende kansen minstens gelijk te maken aan 95%, als de kans op ten onrechte bijregelen gelijk blijft?

3.5. $\underline{x} \simeq \mu + 3\underline{u}$, μ onbekend. Hoeveel onafhankelijke waarnemingen van $\underline{x}$ moet men minstens nemen opdat de breedte van een tweezijdig 95%-betrouwbaarheidsinterval voor μ hoogstens 2 is?

3.6. $x_1,\ldots,x_n$ vormen een steekproef van $\underline{x}$ met μ onbekend en $\sigma^2 = 4$. Geef een tweezijdig betrouwbaarheidsinterval voor μ ($\alpha = 0.1$):

a) indien $\underline{x}$ normaal verdeeld is;

b) indien niets omtrent de verdeling van $\underline{x}$ bekend is, behoudens de variantie.

3.7. In tabel S.C. 1.1 zien we $\varphi(2.0) = 0.0540$. Reken dit na. (Gebruik hiervoor S.C, pag. 79 en tabel 9.5.)

3.8. In een kasboek komen 120 posten voor, alle groter dan ƒ 1,--. Hoe groot zijn de kansen dat men in het totale bedrag een fout maakt groter dan ƒ 5,-- als men:

a) alle posten eerst op gehele guldens afrondt alvorens ze op te tellen;

b) alle bedragen minder dan 1 gulden afkapt en het eindbedrag corrigeert net $120 \times 0.495 =$ ƒ 59.40 ?

3.9. Een bepaald type radiobuis heeft een levensduur van 200 uur met een standaardafwijking van 30 uur.

a) Hoeveel buizen moet men in voorraad hebben om met een betrouwbaarheid van 99.5% te kunnen rekenen op een totale levensduur van 2 jaar?

b) Wat zijn de onderstellingen en de statistische stellingen die bij de beantwoording van a) zijn toegepast?

3.10. Een biscuitfabriek fabriceert rollen die 40 biscuits behoren te bevatten. De biscuits wegen gemiddeld 3 gram met een standaardafwijking van 0.2 gram. De verpakking weegt gemiddeld 15 gram met een standaardafwijking van 0.5 gram (gewicht biscuit en gewicht verpakking worden onafhankelijk en normaal verdeeld ondersteld).

a) Indien een rol 132 gram weegt, is dan het vermoeden gerechtvaardigd dat er minder dan 40 biscuits inzitten?

b) Is de onderstelling dat de gewichten van biscuits en verpakking normaal verdeeld zijn essentieel?

3.11. De gemiddelde absorptie van een gas in absorptiekool is 5.37 eenheden, de standaardafwijking is 0.15 eenheden. Met een nieuwe partij absorptiekool worden 10 proeven gedaan. De gemiddelde absorptie is 5.23 eenheden. Is de kwaliteit van deze partij afwijkend?

3.12. In een fabriek staan 2 vulmachines A en B waarmee flessen worden gevuld met slaolie. Bij een juiste instelling van de machines is de gemiddelde inhoud van een fles 250 gram; de standaardafwijking van de gedoseerde hoeveelheid is 2.5 gram.
Om te controleren of de machines goed zijn ingesteld wordt van elk machine de inhoud van 4 flessen bepaald. Voor machine A is de gemiddelde inhoud van de 4 flessen 251.68 gram, voor machine B 252.68 gram.

a) Toets met een onbetrouwbaarheid van 5% of machine A goed is ingesteld. Doe hetzelfde voor machine B.

b) Toets of de instellingen van de machines A en B onderling verschillen ($\alpha = 0.05$).

c) Welke onderstellingen liggen ten grondslag aan de gebruikte toetsen?

3.13. Uit een baal katoen werd een steekproef genomen van 4000 draden om de vezellengte te bepalen. De gemiddelde lengte was 2.33 cm en standaardafwijking 0.4806 cm. Een andere steekproef van 200 draden werd volgens een andere methode genomen dan de eerste. Het gemiddelde van deze tweede steekproef was 2.54 cm.

a) Is er een significant verschil tussen de twee steekproefmethoden?

b) Is de afronding in de gegevens van dit vraagstuk juist?

c) De conclusie die U onder a) trekt is op grond van de verstrekte informatie niet gerechtvaardigd. Welke informatie ontbreekt?

3.14. Een fabriek vervaardigt staaldraad waarvan de treksterkte normaal verdeeld is met gemiddelde 140 en standaardafwijking 8 kg/mm^2. Het fabriekslaboratorium heeft een nieuw procédé ontwikkeld ter vervaardiging van staaldraad, waarvan men verwacht dat het de gemiddelde treksterkte vergroot. Er wordt een steekproef genomen van 64 volgens het nieuwe procédé vervaardigde proefstukken draad.

a) Formuleer de bijpassende nulhypothese H_0. Bij welke steekproefuitkomsten wordt H_0 verworpen met een onbetrouwbaarheid van 0.01?

Veronderstel nu verder dat het nieuwe procédé een gemiddelde treksterkte oplevert van 142 kg/mm^2, terwijl de standaardafwijking hetzelfde blijft.

b) Hoe groot is de kans dat H_0 niet verworpen wordt?
c) Bepaal de minimale steekproefgrootte zodat de kans genoemd in b) niet groter is dan 0.05 (α = 0.01).

3.15. "Dit blik verf is goed voor 8 tot 12 m^2."
Noemen we de oppervlakte die met één blik geverfd kan worden $\underline{x}$, dan kan men de slogan bijvoorbeeld interpreteren als $\underline{x} \simeq N(10,1)$.

a) Wat is de kans dat onder een partij van 100 blikken er tenminste één blik voorkomt waarmee nog geen 7 m^2 kan worden geverfd?
b) Wat zijn de nevenhypothesen die stilzwijgend in de formulering van het probleem en de vraagstelling liggen opgesloten en in hoeverre acht U deze aanvaardbaar?

3.16. In tabel S.C. 3.2 staan voor ν = 9 en α = 0.05 (tweezijdig) de getallen a_1 = 0.69 en a_2 = 1.83. Bereken deze getallen met behulp van tabel 3.1.

3.17. De output over 24 uren van een continu chemisch proces was gemiddeld over een lange periode 18678 lb, met een standaardafwijking van 555 lb. Na modificatie van enige procesgrootheden werden in een steekproef van 10 dagen de volgende opbrengsten in lb's gevonden:

19200, 18800, 18700, 19200, 19600, 19000, 19400, 19200, 19100, 19500.

a) Is er verandering opgetreden in de dagelijkse output? (α = 0.05.)
b) Bereken de standaardafwijking s van de steekproef.
c) Geef een 95% betrouwbaarheidsinterval voor σ en vergelijk dit met de waarde 555 lb.

3.18. Staaldraad heeft een trekvastheid van ongeveer 140 kg/mm^2. Niet alleen het gemiddelde doch ook de spreiding is van belang. Bij een keuring wordt volgens een standaardmethode aan 10 proefstukken de trekvastheid gemeten en een partij draad wordt afgekeurd wanneer $s > 10$ kg/mm^2.

a) Wat is bij een onbetrouwbaarheid van 5% de nulhypothese die hier in feite wordt getoetst?

b) Welke onderstellingen liggen aan deze toets ten grondslag?

c) Construeer met behulp van tabel S.C. 3.1 een kromme voor het onderscheidingsvermogen van deze toets.

d) Wat voor moeilijkheden zullen zich bij de bemonstering voordoen?

3.19. Aan 4 stalen kogels werd in een aantal richtingen de diameter gemeten. De uitkomsten in microns waren:

kogel 1: 5520, 5529, 5530, 5527
kogel 2: 5528, 5528, 5528, 5525, 5528, 5526
kogel 3: 5522, 5522, 5521, 5521, 5520
kogel 4: 5521, 5523, 5520.

Bepaal door samenvoegen van de varianties de gezamenlijke standaardafwijking van deze 4 waarnemingsreeksen. Voor welk kenmerk van deze stalen kogels is die standaardafwijking een maat?

3.20. Een steekproef van 8 stuks levert een variantie op van 2.59 en een steekproef van 16 stuks heeft variantie 4.92. Kan redelijkerwijze aangenomen worden dat deze steekproeven uit dezelfde normaal verdeelde populatie afkomstig zijn? ($\alpha = 0.05$.)

3.21. Gegeven is een steekproef van 10 waarnemingen uit een normale verdeling met onbekende μ en σ:

12.05; 12.71; 12.25; 12.40; 12.15; 12.94; 12.00; 12.40; 12.49; 12.33.

a) Geef een 95%-betrouwbaarheidsinterval voor μ en σ.

b) Indien gegeven was dat $\sigma = 0.3$, hoe wordt dan het betrouwbaarheidsinterval voor μ ?

3.22. Crooke's onderzoek naar het atoomgewicht van thallium gaf de volgende resultaten:

203.628; 203.632; 203.636; 203.638; 203.639; 203.642;
203.644; 203.649; 203.650; 203.666.

a) Bereken het 95%-betrouwbaarheidsinterval voor het atoomgewicht.
b) Hoeveel waarnemingen moeten extra worden gedaan om het atoomgewicht te kunnen bepalen met een nauwkeurigheid van ± 0.002 $(\alpha = 0.05)$.

3.23. Gegeven is de volgende serie van 10 waarnemingen:

210.68; 211.12; 210.72; 210.81; 210.91;
210.83; 211.13; 211.37; 211.65; 210.93.

a) Rond de waarnemingen op statistisch verantwoorde wijze af en bereken $\bar{x}$ en s.
b) Toets de hypothese H_0: $\mu = 210.90$ $(\alpha = 0.05)$:
   i) door de overschrijdingskans te berekenen,
   ii) m.b.v. het kritieke gebied,
   iii) m.b.v. een betrouwbaarheidsinterval.
c) Wat is het essentiële verschil tussen de 3 methoden van b)?

3.24. Gegeven twee steekproeven uit twee normale populaties met verschillende gemiddelden naar gelijke variantie:

I: 5.31; 5.34; 5.30; 5.31; 5.37; 5.35
II: 5.38; 5.39; 5.42; 5.40.

a) Geef een schatting s^2 van σ^2 met bijbehorend aantal vrijheidsgraden.
b) Toets of de gemiddelden inderdaad verschillen $(\alpha = 0.05)$.
c) Hoe groot zou men σ^2 schatten indien ook de gemiddelden gelijk waren?

3.25. Een analiste heeft een aantal titraties in duplo uitgevoerd met de volgende resultaten:

I:	2.12	1.68	1.55	2.57	2.40	1.97	1.19	2.88	1.74	1.10	2.34
II:	2.58	1.74	1.61	2.85	2.46	2.74	1.47	2.75	1.49	1.99	2.44.

Geef een schatting van de standaardafwijking van de titratiemethode, aannemende:

a) dat de tweede titratie, die doorgaans enige tijd na de eerste geschiedt, een systematisch verschil vertoont met de eerste titratie;

b) dat er geen systematisch verschil bestaat.

3.26. Past men de formule voor het samenvoegen van varianties toe op n paren waarnemingen (duplo-bepalingen) dan vindt men:

$$s^2 = \frac{1}{2n} \sum_{i=1}^{n} d_i^2 ,$$

waarin d_i het verschil is tussen de waarnemingen van het i-de paar. Bewijs dit.

3.27. Gegeven twee steekproeven van 7 waarnemingen uit 2 normale populaties met onbekende parameters μ_1, σ_1, resp. μ_2, σ_2:

I: 5.314; 5.347; 5.301; 5.319; 5.372; 5.361; 5.355

II: 5.382; 5.393; 5.420; 5.359; 5.390; 5.378; 5.402.

a) Rond de reeksen toelaatbaar af en codeer ze.

b) Bereken $\bar{x}_1$, s_1^2, $\bar{x}_2$ en s_2^2.

c) Geef zowel voor μ_1 als voor μ_2 een 95%-betrouwbaarheidsinterval.

d) Toets de hypothese $\sigma_1 = \sigma_2$ ($\alpha = 0.05$). Voeg daarna de varianties samen.

e) Toets de hypothese $\mu_1 = \mu_2$ ($\alpha = 0.05$).

f) Geef een tweezijdig betrouwbaarheidsinterval ($\alpha = 0.05$) voor $\mu_1 - \mu_2$ en voor σ_1/σ_2.

3.28. Uit een populatie werden 8 steekproeven getrokken van elk 5 stuks. De 8 steekproefgemiddelden waren:

.8745; .8753; .8740; .8744; .8746; .8741; .8751; .8748.

a) Toets de hypothese $\mu = .8750$ ($\alpha = 0.05$).

b) Toets de hypothese $\sigma = .0008$ ($\alpha = 0.05$).

3.29. Twee onafhankelijke steekproeven gaven de volgende uitkomsten:

$n_1 = 5$; $\bar{x}_1 = 10.3$; $s_1 = 1.7$;

$n_2 = 8$; $\bar{x}_2 = 12.8$; $s_2 = 3.2$.

a) Toets de hypothese $\mu_1 = \mu_2$ met de t-toets en een onbetrouwbaarheid van 5%.

b) Op welke onderstellingen is deze toets gebaseerd?

c) Construeer een betrouwbaarheidsinterval voor μ_1 met een linkeronbetrouwbaarheid van 1% en een rechteronbetrouwbaarheid van 5%. Wat is de betekenis van zo'n betrouwbaarheidsinterval?

d) Toets de hypothesen $\sigma_1 = \sigma_2$ en $\sigma_1 \leq \sigma_2$ met $\alpha = 0.05$.

3.30. Twee analisten A en B hebben ieder 10 bepalingen gedaan, waarvoor 20 monsters zijn gebruikt. De resultaten in procenten zijn:

A:	7	9	8	9	8	11	9	8	9	8
B:	11	7	10	10	9	10	10	9	11	11

a) Toets de hypothese $\mu_1 = \mu_2$ ($\alpha = 0.05$).

b) Welke onderstellingen liggen aan deze toets ten grondslag?

c) Toets de hypothese $\mu_1 = \mu_2$ als beide analisten hun bepalingen steeds aan hetzelfde monster doen (er zijn dus 10 monsters gebruikt voor de proef).

3.31. Om de nauwkeurigheid van twee analisten A en B te vergelijken laat men beiden een aantal metingen uitvoeren van het verzepingsgehalte van cocosolie. A vindt voor een serie monsters genomen uit één fles:

253.8; 255.4; 256.2; 256.1; 255.2; 255.4.

B vindt voor monsters uit dezelfde fles:

252.8; 258.9; 256.5; 255.7; 254.1; 255.6.

a) Vul de onvolledige schets van de situatie aan zodanig
   i) dat een éénzijdige toets de juiste is;
   ii) dat een tweezijdige toets de juiste is.

b) Voer beide toetsen uit ($\alpha = 0.05$) door na te gaan of de realisatie van de toetsingsgrootheid in de kritieke zone ligt.

c) Ga na of er reden is te concluderen dat het gemiddelde niveau van beide analisten verschilt ($\alpha = 0.05$; aanname: $\sigma_A^2 = \sigma_B^2$).

3.32. Onderstaande waarnemingen geven de totale jaarlijkse regenval in De Bilt in 8 opeenvolgende jaren:

950; 809; 792; 597; 818; 660; 752; 928.

a) Bereken $\bar{x}$, s^2 en s.

b) Geef een 95%-betrouwbaarheidsinterval voor de verwachting μ en de standaardafwijking σ.

c) Welke onderstellingen liggen aan deze intervallen ten grondslag?

3.33. Twee steekproeven uit verschillende populaties gaven de volgende resultaten:

$$n_1 = 40 \;;\quad \bar{x}_1 = 10.75 \;;\quad s_1 = 0.61$$
$$n_2 = 30 \;;\quad \bar{x}_2 = 10.37 \;;\quad s_2 = 0.85 \,.$$

Ondersteld wordt dat beide populaties gelijke varianties hebben.

a) Construeer een 95%-betrouwbaarheidsinterval voor μ_1 en μ_2 afzonderlijk.

b) Toets de hypothese $\mu_1 = \mu_2$ $(\alpha = 0.05)$.

c) Toets de bovengenoemde onderstelling betreffende de varianties $(\alpha = 0.10)$.

d) De betrouwbaarheidsintervallen onder a) overlappen elkaar. Hoe is dit te rijmen met Uw conclusie onder b) ?

3.34. Om twee spectrofotometers te vergelijken werden aan 16 monsters van een oplossing bepalingen gedaan (voor ieder instrument 8 monsters). De resultaten waren:

Instrument I: 4.17; 4.26; 4.20; 4.19; 4.17; 4.24; 4.21; 4.16
Instrument II: 4.15; 4.16; 4.14; 4.17; 4.13; 4.20; 4.15; 4.18.

a) Toets de hypothese $\sigma_1 = \sigma_2$ $(\alpha = 0.05)$.

b) Toets de hypothese $\mu_1 = \mu_2$ $(\alpha = 0.05)$.

c) Geef een 95%-betrouwbaarheidsinterval voor σ_1/σ_2.

3.35. Gegeven zijn twee aselecte steekproeven uit normale verdelingen, waarvan de parameters onbekend zijn:

x_{1i}: 11.3; 10.7; 9.3; 9.1; 9.4; 9.0
x_{2i}: 9.4; 10.7; 10.9; 11.6; 11.1; 13.0; 8.9; 11.6

$$\bar{x}_1 = 9.80 \;;\quad KS_1 = 4.60$$
$$\bar{x}_2 = 10.90 \;;\quad KS_2 = 11.72 \,.$$

a) Ga na of deze resultaten aanleiding geven tot de conclusie dat één of meer der parameters vermoedelijk verschillen door de betreffende toetsen uit te voeren en/of betrouwbaarheidsintervallen te construeren $(\alpha = 0.05)$. Argumenteer de keuze één- dan wel tweezijdig.

b) Beschrijf kort en duidelijk wat het verschil is in de verkregen informatie, als U

i) toetst H_0: $\theta_1 = \theta_2$ tegen H_{alt}: $\theta_1 \neq \theta_2$;

ii) een betrouwbaarheidsinterval opstelt voor $\theta_1 - \theta_2$.

c) Hoe luiden Uw conclusies als elk der gegeven waarnemingen het gemiddelde is van 5 waarnemingen?

3.36. Twaalf bepalingen van het soortelijk gewicht van een chemicalie gaven een gemiddelde waarde .683 en een standaardafwijking .0077. Bereken voor een betrouwbaarheid van 99% het aantal waarnemingen dat nodig is voor een nauwkeurigheid van de soortelijk gewicht bepaling van 2 decimalen.

3.37. De volgende gegevens hebben betrekking op het percentage $CaCO_3$ bepaald door duplo-titraties aan elk van 7 monsters van een mengsel:

Steekproefnr.	1	2	3	4	5	6	7
test 1	76.12	75.81	76.37	77.15	77.47	76.42	77.15
test 2	76.35	75.82	76.51	77.03	77.26	76.65	77.20

a) Is er verschil tussen de 1e en de 2e bepaling? ($\alpha = 0.05$.)

b) Er werd gesteld dat de verschillen tussen de twee bepalingen zo klein zijn, omdat de analist het resultaat van de eerste bepaling kent als hij de tweede doet. Daarom werden nog eens 5 monsters genomen en er werd voor gezorgd dat de analist het resultaat van de eerste bepaling niet kende bij het doen van de tweede. De percentages waren nu:

steekproefnr.	8	9	10	11	12
test 1	76.32	77.42	77.53	76.14	76.74
test 2	76.85	76.91	77.41	76.32	76.41

Ga na of de hierboven geuite onderstelling juist is ($\alpha = 0.05$).

3.38. Gegeven zijn de volgende waarnemingsuitkomsten:

x_{1i}: 0; 5; 2; 3; 8; 0

x_{2i}: 9; 3; 6; 7; 5

a) Bereken $\bar{x}_1, s_1^2, \bar{x}_2, s_2^2$.

b) Toets de hypothese $\sigma_1^2 = \sigma_2^2$ ($\alpha = 0.05$).

c) Stel dat de twee steekproeven afkomstig zijn uit populaties met dezelfde variantie σ^2. Geef een 95%-betrouwbaarheidsinterval voor σ^2.

d) Toets de hypothese $\mu_1 \geq \mu_2$ ($\alpha = 0.05$).

e) Wat zijn de onderstellingen bij de toegepaste toetsen?

3.39. Gegeven zijn twee series waarnemingen uit normale verdelingen (met onbekende μ en σ^2):

x_{1j}: 16.0; 15.2; 20.0; 19.8; 15.6; 15.3; 19.0
x_{2j}: 11.7; 10.5; 14.3; 13.6; 16.0.

In deze situatie zijn de volgende modellen mogelijk:

(1) $\underline{x}_{ij} = \mu + \underline{u}_{ij}\sigma$
(2) $\underline{x}_{ij} = \mu_i + \underline{u}_{ij}\sigma$
(3) $\underline{x}_{ij} = \mu + \underline{u}_{ij}\sigma_i$
(4) $\underline{x}_{ij} = \mu_i + \underline{u}_{ij}\sigma_i$.

a) Voer de noodzakelijke toetsen uit ($\alpha = 0.05$), geef daarbij de nulhypothese aan en vermeld of de toets één- of tweezijdig is.

b) Vat Uw conclusies samen in schattingen van de modelparameters.

3.40. a) In tabel S.C. 2.1 zien we voor $\nu = \infty$ de getallen 1.28, 1.64, etc. Waar staan die getallen nog meer en verklaar dit.

b) Er geldt: $t_\nu^2 = F_\nu^1$. Controleer dit voor bijvoorbeeld $\nu = 5$ met behulp van tabel S.C. 4.1. Kijk bij de juiste α. (Voor definitie t_ν, χ_ν^2, $F_{\nu_2}^{\nu_1}$: zie S.C. pag. 12.)

c) Er geldt: $F_\infty^\nu = \chi_\nu^2/\nu$. Controleer dit voor $\nu = 6$.

d) Controleer eveneens dat $F_\infty^1 = \chi_1^2 = u^2$.

Toetsen op normaliteit.

4.1. Tabel S.C. 8.4 bevat 25 kolommen × 50 rijen aselecte getallen van twee cijfers.

a) Wijs door loting aselect één rij en één kolom aan en beschrijf hoe U die loting heeft uitgevoerd.

b) Schrijf, te beginnen bij het onder a) aangewezen punt, 10 steekproeven uit die tabel op van 3 en 10 steekproeven van 5 van deze aselecte getallen.

c) Bepaal van deze steekproeven de medianen.

d) Toets voor beide groepen van 10 steekproeven of de verdeling van de mediaan bij benadering normaal is m.b.v. de grafische methode.

e) Bepaal weer voor beide groepen grafisch een schatting van $\mathcal{E}\underline{M}$ en $\sigma(\underline{M})$.

f) Bepaal ook de theoretische waarden van $\mathcal{E}\underline{M}$ en $\sigma(\underline{M})$ en vergelijk deze met de schattingen.

4.2. Van 60 hartpatiënten is het cholesterolgehalte van het bloed bepaald in mg%:

240	293	288	206	204	200	412	372	240	310	265	173
261	198	253	180	223	325	196	386	325	269	200	289
280	296	248	299	196	203	270	293	230	264	276	210
183	230	266	289	199	194	323	296	201	283	202	367
188	153	226	250	194	294	310	276	281	229	224	295

Toets of het cholesterolgehalte normaal verdeeld is

a) volgens de methode met klassegrenzen;

b) volgens de methode met klassemiddens.

c) volgens de methode van Shapiro en Wilk.

4.3. a) Zij $\underline{x}$ = som der ogen geworpen met 3 dobbelstenen. Genereer 12 realisaties van $\underline{x}$ en toets met de methode van Shapiro en Wilk of deze steekproef afkomstig is uit een normale verdeling ($\alpha = 0.05$).

b) Doe hetzelfde voor $\underline{y}$ = som der ogen geworpen met 5 dobbelstenen.

5. Hypergeometrische, binomiale en Poisson verdeling.

5.1. De staf van een advertentie-afdeling van een krantenbedrijf bestaat uit 30 personen, waarvan er 20 vroeger ergens anders werkten. Er wordt een steekproef genomen van 5 personen.

a) Wat is de kansverdeling van het aantal personen $\underline{x}$ in de steekproef dat altijd bij dit bedrijf heeft gewerkt?

b) Bereken $\mathcal{E}\underline{x}$ en var $\underline{x}$.

c) Wat is de kans dat alle 5 personen in de steekproef altijd bij dit bedrijf hebben gewerkt?

5.2. Een fabrikant van electromotoren koopt onderdelen in in partijen van 50 stuks. Uit iedere partij worden 5 exemplaren onderzocht op fouten. Als geen enkel fout exemplaar wordt gevonden wordt de partij geaccepteerd. Als er een of meer foute onderdelen in de steekproef zitten wordt de hele partij onderzocht. Veronderstel dat een bepaalde partij 6 foute exemplaren bevat, wat is dan de kans dat de hele partij wordt gecontroleerd?

5.3. Laat zien dat de hypergeometrische verdeling benaderd kan worden door de binomiale voor grote M en N, m.a.w. bewijs:

$$\lim_{\substack{M\to\infty \\ N\to\infty \\ M/N\to p}} \frac{\binom{M}{x}\binom{N-M}{n-x}}{\binom{N}{n}} = \binom{n}{x}p^x(1-p)^{n-x} .$$

5.4. Ongeveer 100.000 huisvrouwen gebruiken een bepaald produkt. Aan een steekproef van 200 huisvrouwen uit deze groep wordt het produkt met een gewijzigde samenstelling aangeboden. Van de ondervraagden vinden 124 het gewijzigde produkt beter dan het oude. Neem aan dat het aantal huisvrouwen (in de totale groep van 100.000) dat het nieuwe produkt beter vindt n is. Toets de hypothese H_0: $n \leq 50.000$ $(\alpha = 0.05)$.

5.5. In een aselecte steekproef van 300 flessen melk heeft men bij 80% een voldoende vetgehalte geconstateerd. Bepaal een tweezijdig 95%-betrouwbaarheidsinterval voor het percentage flessen melk met voldoende vetgehalte in de produktie.

5.6. De chef van een autogarage heeft het idee dat bij zekere 6-cylinder motor de kleppen van de voorste cylinder in minder goede staat verkeren dan die van de andere cylinders. Hij meent een en ander te kunnen verklaren. Uit de reparatielijsten over een bepaalde periode blijkt dat in 27 van de 115 onder handen genomen motoren de kleppen van de voorste cylinder inderdaad in slechtere conditie zijn dan die der overige. Toets H_0: de slijtage van de voorste kleppen is niet ernstiger dan die van de andere ($\alpha = 0.05$).
Opmerking. Voer de toets op de volgende wijze uit:
i) met behulp van een kritieke zone;
ii) met behulp van een overschrijdingskans;
iii) met behulp van een betrouwbaarheidsinterval.

5.7. Bij zware terreinproeven met vrachtwagens van 2 merken A en B werd het volgende resultaat verkregen. Van de 30 wagens van merk A doorstonden er 22 de proeven goed, terwijl 8 ervan in moeilijkheden geraakten; van de 30 wagens van merk B kwamen er 12 in moeilijkheden, de overige 18 doorstonden de proeven.

a) Ga na of één van beide merken beter geschikt is voor terreinwerk van het onderzochte soort dan het andere ($\alpha = 0.05$).
b) Geef een 95%-betrouwbaarheidsinterval voor $p_A - p_B$, als p_A (p_B) de kans is dat een vrachtwagen van het type A (B) bij zo'n terreinproef in moeilijkheden geraakt.

5.8. Leid af dat de binomiale verdeling tot een Poisson verdeling nadert voor $n \to \infty$, $p \to 0$ en $np = \mu$.

5.9. Een medicus beweert dat de kans op een jongensgeboorte groter is dan die op de geboorte van een meisje. Hij kwam tot deze conclusie omdat 51% van de pasgeboren babies uit zijn praktijk jongens waren. Hoeveen geboorten moeten dat op zijn minst geweest zijn om deze conclusie te rechtvaardigen? ($\alpha = 0.05$.)

5.10. Bij een onbetrouwbaarheid α worden de grenzen van het betrouwbaarheidsinterval voor de parameter p van een binomiale verdeling in het geval dat een normale benadering is toegestaan gegeven door

$$\frac{\hat{p} + \frac{u^2_{(\frac{1}{2}\alpha)}}{2n} \pm u_{(\frac{1}{2}\alpha)} \sqrt{\frac{\hat{p}\hat{q}}{n} + \frac{u^2_{(\frac{1}{2}\alpha)}}{4n^2}}}{1 + \frac{u^2_{(\frac{1}{2}\alpha)}}{n}} .$$

Leid dit af.

5.11. Iemand koopt een dobbelsteen en controleert deze op zuiverheid. Hiertoe werpt hij 600 keer en vindt 70 keer een zes. Is het nodig te reclameren?

5.12. Men wil het aantal vissen schatten dat zich in een vijver bevindt. Daartoe vangt en merkt men 50 vissen. Men laat deze vervolgens weer zwemmen. Na verloop van enige tijd vangt men 100 vissen. Hiervan blijken er 12 gemerkt te zijn. Geef grenzen aan, waartussen behoudens een onbetrouwbaarheid α het totaal aantal vissen zal liggen ($\alpha = 0.05$). Welke veronderstellingen moeten hierbij nog worden gemaakt?

5.13. Men wil een betrouwbaarheidsinterval vaststellen voor de fractie stemgerechtigden die stemt op een zekere partij. Bepaal de omvang van de aselecte steekproef die daartoe genomen zal moeten worden zodanig, dat de breedte van het interval 0.04 bedraagt ($\alpha = 0.05$). Merk op dat er geen schatting van de bedoelde fractie ter beschikking staat.

5.14. Bij een proefopzet zijn 10 series van 3 waarnemingen uitgevoerd. Men beweert dat de waarnemingen binnen één serie onderling onafhankelijk zijn. Doch in 8 van de 10 series heeft de tweede waarneming een hogere uitkomst gegeven dan de eerste en de derde. Is dit een reden om aan de onafhankelijkheid te twijfelen? Wat is de nulhypothese en wat zijn de onderstellingen die aan Uw toets ten grondslag liggen?

5.15. Bij de fabrikage van aardewerken borden worden bij gemiddeld 5% van de borden kleine bakfouten geconstateerd.

a) Bij een steekproef van 20 stuks genomen tijdens het produktieproces worden 3 foutieve exemplaren aangetroffen. Geeft dit aanleiding tot bijregeling van het produktieproces? ($\alpha = 0.05$.)

b) Bij een tweede steekproef, eveneens van 20 stuks, worden weer 3 borden gevonden met bakfouten (er is terecht of ten onrechte niet bijgeregeld na de eerste steekproef). Wat is nu Uw conclusie?

c) Bepaal zowel voor a) als voor b) het 95%-betrouwbaarheidsinterval voor p.

5.16. Partijen van 3000 bakstenen worden gezamenlijk gebakken door ze op een karretje door een tunneloven te rijden. Men heeft in het verleden 5% misbaksels geconstateerd en vindt nu op één karretje van 3000 stenen 200 misbaksels, dat is 6.67%.

a) Welke onderstelling(en) is (zijn) nodig om te kunnen toetsen of er met de produktie wat mis is?

b) Is het percentage misbaksels in deze partij significant hoger dan verwacht? (Neem aan dat aan de nodige onderstellingen is voldaan.)

5.17. Geef een (geargumenteerde) benadering voor de kans dat 12 onafhankelijke worpen met een dobbelsteen tezamen

a) minstens 42, en

b) precies 42

opleveren.

5.18. Wanneer $\underline{x}_1 \simeq PS(\mu_1)$, $\underline{x}_2 \simeq PS(\mu_2)$ onderling onafhankelijk zijn, dan is

$$P(\underline{x}_1 = x_1 \mid \underline{x}_1 + \underline{x}_2 = n) = \binom{n}{x_1} p^{x_1} (1-p)^{x_2}$$

met

$$p = \frac{\mu_1}{\mu_1 + \mu_2} .$$

Bewijs dit.

5.19. In een bepaalde afdeling van een fabriek komen gemiddeld per jaar 3 ernstige ongevallen voor. Men besluit ter verhoging van de veiligheid de in gebruik zijnde machines door nieuwe van een ander type te vervangen.

a) In het eerste jaar dat met de nieuwe machines wordt gewerkt is er 1 ernstig ongeval. Kan men, behoudens een onbetrouwbaarheid van 10%, concluderen dat de nieuwe machines veiliger zijn?

b) Ook in het tweede jaar is er slechts 1 ernstig ongeluk. Hoe luidt nu Uw conclusie?

5.20. Twee radio-aktieve preparaten gaven in één minuut voor een Geigerteller resp. 252 en 315 tellingen.

a) Is er verschil in aktiviteit?

b) Geef betrouwbaarheidsintervallen voor μ_1, μ_2 en het verschil $\mu_1 - \mu_2$ ($\alpha = 0.05$).

Opmerking. μ is de verwachting van het aantal tellingen per minuut.

5.21. Bewijs: als

$$\underline{x} \simeq PS(\mu_x) \ , \quad \underline{y} \simeq PS(\mu_y) \ ,$$

$\underline{x}$ en $\underline{y}$ onderling onafhankelijk, $\underline{z} = \underline{x} + \underline{y}$,

dan

$$\underline{z} \simeq PS(\mu_x + \mu_y) \ .$$

5.22. Gemiddeld passeren er 15 auto's per dag een benzinestation. De kans dat een auto tankt is 1/5. Bereken de kans dat op een zekere dag precies x auto's tanken. (Neem daarbij aan dat het aantal auto's dat per dag passeert een Poisson-verdeling volgt.)

5.23. Aan twee radio-aktieve preparaten A en B heeft men met een Geigerteller waargenomen:

bij A: 540 aanslagen in 1 minuut,
bij B: 390 aanslagen in ½ minuut.

Zijn λ_A en λ_B het verwachte aantal aanslagen per seconde voor de preparaten, geef dan een 99%-betrouwbaarheidsinterval voor het verschil $\lambda_A - \lambda_B$.

5.24. Bij de keuring van een grote partij massaprodukten is voorgeschreven dat een steekproef van 150 stuks uit iedere partij moet worden genomen en de partij moet worden goedgekeurd wanneer hierin 5 of minder defectieven worden gevonden, doch afgekeurd wanneer het aantal defectieven 6 of meer bedraagt. Men kan dit interpreteren als een toets op H_0: $p \le p_0$, $\alpha = 0.05$ en afkeuren wanneer H_0 wordt verworpen of als een toets of H_0': $p \ge p_1$, $\alpha = 0.05$ en goedkeuren, wanneer H_0' wordt verworpen.
Vraag: Wat zijn de waarden van p_0 en p_1?

5.25. Van twee machines is het aantal foutieve exemplaren geteld dat ze gedurende een bepaalde tijd hebben geproduceerd. Bij de eerste machine werden in 8 uur gemiddeld 37 fouten per uur waargenomen en bij de tweede machine in 6 uur gemiddeld 30 fouten per uur. We mogen aannemen dat het aantal fouten per uur voor beide machines Poisson-verdeeld is. Toets de hypothese dat beide Poisson-verdelingen hetzelfde gemiddelde hebben ($\alpha = 0.05$).

5.26. Een autoverhuurder bezit twee auto's die hij per dag verhuurt. Het aantal aanvragen per dag is een onafhankelijke trekking uit een Poisson-verdeling met $\mu = 1.5$. Een auto vergt jaarlijks ƒ 3500,-- aan afschrijving, onderhoud en garagekosten en brengt ƒ 25,-- op per dag dat zij wordt verhuurd.

a) Gemiddeld hoeveel dagen per jaar zijn beide auto's thuis?
b) Indien beide auto's even vaak worden gebruikt, hoeveel dagen per jaar wordt dan gemiddeld elke auto gebruikt?
c) Aan welk percentage van de aanvragen kan niet worden voldaan?
d) Hoeveel verdient de verhuurder gemiddeld per jaar door zijn autoverhuur?
e) Wat wordt zijn gemiddelde verdienste wanneer hij een derde auto in bedrijf zou nemen?

5.27. Twee machines A en B worden vergeleken op het aantal ernstige storingen dat bij deze machines voorkomt. Bij machine A zijn in het afgelopen jaar 8 grote storingen voorgekomen, bij B slechts 2.

a) Voer de toets uit door (benaderingen van) de overschrijdingskans te bepalen op 2 manieren:
   I met de procedure waarbij wordt benaderd met de normale verdeling;
   II met de procedure waarbij wordt gewerkt onder de voorwaarde: totaal aantal incidenten = 10.
   De onbetrouwbaarheidsdrempel voor beide toetsen is $\alpha = 0.10$.
b) Bepaal voor elk der beide toetsen het kritieke gebied.
   Opmerking. De punten $(x_A, x_B) = (7,2), (9,3), (11,4), (12,5)$ liggen in het kritieke gebied van toets I.
c) Geef zonder berekening aan, welke van de beide toetsen het grootste onderscheidingsvermogen bezit. Geef vervolgens de formule voor de berekening van het onderscheidingsvermogen als functie van μ_A en μ_B.

5.28. Een fabrikant beweert dat een partij van 60 produkten een fractie p foute produkten bevat. Ter controle neemt een afnemer een aselecte steekproef van 20 stuks en vindt er 3 foute produkten in. Wat is zijn conclusie t.a.v. de bewering van de fabrikant

a) als trekking gebeurt met teruglegging en $p = 1/5$;

b) als trekking gebeurt met teruglegging en $p = 1/30$;

c) als trekking gebeurt zonder teruglegging en $p = 1/20$.

(Onbetrouwbaarheid $\alpha = 0.05$.)

6. De χ^2-verdeling.

6.1. Aan drie radio-aktieve preparaten heeft men met een G.M.-teller waargenomen:
bij A_1: 560 aanslagen in 1 minuut;
bij A_2: 330 aanslagen in ½ minuut;
bij A_3: 910 aanslagen in 1½ minuut.
Het aantal getelde deeltjes per seconde volgt een Poisson-verdeling met parameter λ_i voor preparaat A_i.

a) Toets de hypothese H_0: $\lambda_i = \lambda$, $i = 1,2,3$ $(\alpha = 0.05)$.

b) Geef, als H_0 niet kan worden verworpen, de schattingen voor λ en voor $\mathrm{var}(\underline{\hat{\lambda}})$.

6.2. Aan vier radio-aktieve preparaten A_1, A_2, A_3 en A_4 heeft men met een Geigerteller waargenomen:
bij A_1: 540 aanslagen in 1 minuut;
bij A_2: 300 aanslagen in ½ minuut;
bij A_3: 200 aanslagen in 20 seconden;
bij A_4: 460 aanslagen in 40 seconden.

Toets de hypothese dat de 4 preparaten eenzelfde verwachte aantal getelde deeltjes per seconde hebben.
Geef in het geval dat deze hypothese niet wordt verworpen een schatting voor λ en voor $\mathrm{var}(\underline{\hat{\lambda}})$.

6.3. Uit 6 partijen van zeker produkt heeft men steekproeven genomen van resp. 100, 150, 100, 200, 150 en 200 stuks. De aantallen foute exemplaren hierin zijn resp. 4, 15, 8, 7, 9 en 11 stuks.

a) Toets de hypothese dat het fabricageproces als statistisch beheerst kan worden beschouwd $(\alpha = 0.05)$.

b) Toets de hypothese dat het percentage foute produkten bij dit fabricageproces kleiner dan of gelijk is aan 5% $(\alpha = 0.05)$.

6.4. Aan een bepaald tentamen deden 160 E- en 240 N-studenten mee. Er slaagden 60 E- en 80 N-studenten.

a) Zijn de E-studenten "beter" in dit onderdeel?

b) Toets eveneens de nulhypothese: er is geen verschil tussen beide groepen.

6.5. Een Wiskunde-tentamen gaf voor studenten van W- en T-afdeling de volgende uitslag:

	goed	voldoende	onvoldoende
W:	8	27	20
T:	5	9	6

Onderzoek of er een verschil bestaat in prestaties tussen studenten van beide afdelingen.

6.6. Gegeven het aantal eerstejaars studenten in Delft van 1945-1955 en het percentage W-studenten.

jaar :	1945	1946	1947	1948	1949	1950	1951	1952	1953	1954
aantal :	2355	1357	1092	980	908	870	660	642	665	773
% W-stud.:	20.8	20.0	21.8	21.9	21.3	22.6	22.9	24.6	23.0	22.5

Toets de hypothese dat het percentage W-studenten gedurende die jaren constant is gebleven.

6.7. Toets met behulp van onderstaande gegevens of er verband is tussen de voorkeur voor het automerk en het geslacht ($\alpha = 0.05$).

	merk		
	A	B	C
mannen	60	80	110
vrouwen	80	70	100

6.8. In een drietal verpleeghuizen werden de patiënten onderzocht op urineweginfecties. Gevonden werd

		infectie		
		negatief	dubieus	positief
Verpleeghuis	A	53	5	9
	B	49	11	14
	C	41	0	0

Zijn er significante verschillen tussen de verpleeghuizen wat betreft het vóórkomen van infecties?

6.9. De controle van een produkt betreft 2 gezichtspunten:

1e een controle op de afmetingen,

2e een controle op de afwerking.

Een produkt voldoet al dan niet aan de tolerantie-eisen en wat de afwerking betreft onderscheid men goed, voldoende en onvoldoende. De aantallen in de verschillende kategorieën zijn:

		afmetingen	
		voldoet	voldoet niet
afwerking	goed	324	57
	voldoende	225	50
	onvoldoende	127	37

Toets de hypothese dat de kwaliteit van de afwerking onafhankelijk is van de afmetingen.

6.10. Bij controle van 2352 ebonieten blokjes werden er 154 gevonden met foutieve afmetingen. Van deze 154 stuks waren er 47 bovendien poreus. Van de blokjes met goede afmetingen waren er 410 poreus. Ga na of er verband is tussen de afmetingen en de poreusheid.

6.11. Van een onderdeel worden 2 diameters gemaakt op een automatische draaibank met 6 koppen. Deze diameters moeten voldoen aan nauwe tolerantie-eisen. Onderstaande tabel geeft aantallen onderdelen met goede en foute afmetingen voor de verschillende koppen.

	kop	1	2	3	4	5	6
Diam. A	voldoende	537	582	574	556	568	570
	onvoldoende	63	18	26	44	32	30
Diam. B	voldoende	491	529	521	532	546	538
	onvoldoende	109	71	79	68	54	62

a) Wijs met de dobbelsteen een van de 6 koppen aan en vergelijk hiervoor de resultaten voor de beide diameters met een χ^2-toets.

b) Geef exact aan welke praktische conclusie uit het resultaat van deze ene toets mag worden getrokken en onder welke voorwaarden deze conclusie geldt.

c) Als men de X^2-waarden voor de 6 koppen optelt krijgt men $X^2 = 101.6$, $\nu = 6$, wat zeer significant is. Dit bewijst echter niet dat de ene diameter meer moeilijkheden bij de produktie geeft dan de ander! Waarom niet?

6.12. Simuleer 60 worpen met een dobbelsteen en toets of deze "dobbelsteen" zuiver is.

6.13. Tabel S.C. 8.4 wordt geacht te bestaan uit aselecte trekkingen uit de getallen 0 t/m 9. De tabel is onderverdeeld in blokken van 25 paren. Wijs aselect 2 blokken van 50 cijfers uit deze tabel aan en toets aan deze 100 cijfers of de frequentie waarmee de cijfers 0 t/m 9 in deze 100 cijfers voorkomen, met de bewering in overeenstemming is. Geef duidelijk aan hoe U de 2 blokken van 50 cijfers aanwijst.

6.14. Tabel S.C. 8.7 bevat aselecte trekkingen uit de standaard-exponentiële verdeling:

$$f(x) = e^{-x} , \quad 0 < x < \infty .$$

a) Construeer een frequentietabel voor de eerste kolom van deze tabel ($n = 50$).
b) Bereken $\bar{x}$ en s^2.
c) Bereken de verwachte frequenties op grond van het gegeven dat $f(x) = e^{-x}$.
d) Toets of deze verdeling bij de waarnemingen past.

6.15. Bij het kweken van bacterieën op een glazen plaat telt men het aantal bacterieën dat in verschillende vakjes van 1 cm^2 vóórkomt. Men vond:

aantal bacterieën per vakje:	0	1	2	3	4	5	6
aantal vakjes:	5	19	26	26	21	13	8

Geen enkel vakje bevatte meer dan 6 bacterieën.
Ga nu na of deze waarnemingen een Poisson-verdeling volgen.

6.16. Een machine produceert per uur 600 drukringen. Regelmatig wordt uit de produktie een steekproef van 20 stuks genomen om een bepaalde afmeting te controleren. Het resultaat van 50 steekproeven wordt hieronder gegeven:

aantal foute exemplaren per steekproef:	0	1	2	3	4	5	6
aantal steekproeven:	5	15	12	13	4	0	1

a) Ga na of het aantal foute exemplaren binomiaal verdeeld is.

b) Uit vroegere metingen is bekend dat deze machine gemiddeld 9% uitval levert. Is het resultaat van de 50 steekproeven hiermee in overeenstemming?

6.17. Hieronder staan de resultaten van efficiency-bepalingen van ammonia-oxidatie. De verwachte frequenties zijn berekend op grond van de onderstelling dat de verdeling normaal is.

Efficiency in %	Waargenomen frequentie	Verwachte frequentie
89.95 - 90.45	1	.66
90.45 - 90.95	2	1.76
90.95 - 91.45	4	5.04
91.45 - 91.95	15	12.35
91.95 - 92.45	20	25.45
92.45 - 92.95	47	44.20
92.95 - 93.45	63	64.61
93.45 - 93.95	78	79.53
93.95 - 94.45	88	82.50
94.45 - 94.95	69	72.48
94.95 - 95.45	59	53.22
95.45 - 95.95	35	33.16
95.95 - 96.45	10	17.29
96.45 - 96.95	8	7.71
96.95 - 97.45	4	2.87
97.45 - 97.95	0	.86
97.95 - 98.45	1	.30

a) Geef voor een willekeurige klasse aan hoe de verwachte frequentie is berekend.

b) Toets met de χ^2-toets of de waarnemingen afkomstig zijn uit een normaal verdeelde populatie.

6.18. Gegeven zijn de volgende afgeronde waarnemingsresultaten behorende bij een onbekende verdeling.

klasse	frequentie
121 ÷ 160	2
161 ÷ 200	1
201 ÷ 240	16
241 ÷ 280	27
281 ÷ 320	11
321 ÷ 360	9
361 ÷ 400	4

a) Toets de hypothese dat de betreffende verdeling een normale verdeling is met $\mu = 260.5$ en $\sigma^2 = 2500$.

b) Beschrijf overzichtelijk de uitvoering van de toets op de hypothese, dat de betreffende verdeling een normale verdeling is.

7. Regressie en correlatie.

7.1. Bewijs:

a) Als $\underline{y} \simeq a\underline{x} + b$, dan is $\rho^2(\underline{x},\underline{y}) = 1$.

b) Als $\underline{u} \simeq a\underline{x} + b$ en $\underline{v} \simeq c\underline{y} + d$, dan is $\rho^2(\underline{u},\underline{v}) = \rho^2(\underline{x},\underline{y})$, m.a.w. de correlatiecoëfficiënt is (op het teken na) invariant t.a.v. lineaire transformaties.

c) Als $\underline{x}$ en $\underline{y}$ onafhankelijk zijn, dan zijn ze ongecorreleerd, d.w.z. $\rho(\underline{x},\underline{y}) = 0$. Het omgekeerde geldt niet. Ga na!

7.2. Men werpt driemaal achtereen met een dobbelsteen. Bereken de correlatiecoëfficiënt $\rho(\underline{x},\underline{y})$ tussen de eerste worp ($\underline{x}$) en de som van de 3 worpen ($\underline{y}$).

7.3. Simuleer met behulp van aselecte getallen een serie van 30 x,y-paren en bereken hiervoor een schatting r van ρ. Toets vervolgens de hypothese: $\rho = 0$ ($\alpha = 0.05$).

7.4. Voor een serie van n waarnemingen (x_i,y_i) geldt het volgende model:

$$\underline{y}_i = \alpha_0 + \alpha_1 x_i + \underline{e}_i$$

met

$$\forall_i: \underline{e}_i \sim N(0,\sigma_0^2) \ ; \ \forall_{i\neq j}: \operatorname{cov}(\underline{e}_i,\underline{e}_j) = 0 \ .$$

a) Bepaal de kleinste kwadratenschatters a_0 en a_1 van α_0 resp. α_1.

b) Toon aan dat a_0 en a_1 zuivere schatters zijn.

7.5. Gegeven de leeftijd x en de bloeddruk y van 12 vrouwen, aselect gekozen uit een grote populatie:

x:	56	42	72	36	63	47	55	49	38	42	68	60
y:	147	125	160	118	149	128	150	145	115	140	152	155

a) Zet x en y tegen elkaar uit en trek op het oog zo goed mogelijk een rechte door deze punten en bepaal de vergelijking van die rechte.

b) Bepaal volgens de methode der kleinste kwadraten de regressielijn $\hat{y} = a + bx$ (codeer eerst).

c) Bewijs dat $\hat{\underline{y}}$ een zuivere schatter is voor $\mathcal{E}\underline{y}$.

d) Geef 95%-betrouwbaarheidsintervallen voor de regressie-coëfficiënten.

e) Geef een rechtséénzijdig betrouwbaarheidsinterval voor de bloeddruk van een 37-jarige vrouw ($1-\alpha = 0.95$).

f) Voor welke x is de bijbehorende, uit de regressielijn berekende $\hat{y}$ het meest nauwkeurig?

7.6. Van 10 paren waarnemingen (x_i, y_i) is gegeven:

$$\Sigma x_i = 30 \qquad \Sigma x_i^2 = 120$$

$$\Sigma y_i = 32 \qquad \Sigma y_i^2 = 160 \qquad \Sigma x_i y_i = 125 .$$

We werken volgens het model $\underline{y} = \beta_0 + \beta_1(x-\bar{x}) + \underline{e}$ en $\underline{e} \sim N(0,\sigma^2)$.

a) Toets de hypothese: $\beta_1 = 1$.

b) Toets de hypothese: de lijn gaat door de oorsprong.

7.7. a) Pas een lineair model

$$\underline{y} = \alpha_0 + \alpha_1 x + \underline{e}$$

aan bij de volgende waarnemingen

x	1	2	3	4	5
y	3	2	4	6	5

Geef schattingen voor de parameters α_0 en α_1 en voor de restvariantie σ_0^2.

b) Geef een 95%-betrouwbaarheidsinterval voor α_1.

c) Geef zonder gebruik te maken van tabel S.C. 3.2 een 95%-betrouwbaarheidsinterval voor σ_0.

7.8. Gegeven is de volgende serie waarnemingen:

$$x_i = 0,\ 1,\ 1,\ 2,\ 3,\ 3,\ 4$$

$$y_i = 3,\ 5,\ 6,\ 7,\ 6,\ 7,\ 8 .$$

We beschouwen hierbij het model

$$\underline{y} = \alpha_0 + \alpha_1 x + \underline{e} .$$

a) Bereken de schattingen voor α_0, α_1 en σ_0^2.

b) Wat zijn de schattingen voor var $\underline{\hat{\alpha}}_0$, var $\underline{\hat{\alpha}}_1$ en cov($\underline{\hat{\alpha}}_0$,$\underline{\hat{\alpha}}_1$) ?

c) Geef het 95%-betrouwbaarheidsinterval voor α_1.

7.9. Men veronderstelt dat het aantal foute exemplaren in de produktie van een bepaalde machine afhangt van de snelheid (omw/min), waarmee deze machine werkt. Onderstaande gegevens zijn verkregen door na te gaan hoeveel fouten de machine in een interval van zekere lengte produceerde (de tijdsintervallen werden aselect gekozen).

Snelheid (gecodeerd) :	10	12	15	13	14	17	13	18
Aantal foute exemplaren:	3	4	8	5	6	7	6	9

a) Teken een grafiek.

b) Bereken de rechte lijn, die het beste past bij de waarnemingen. Teken deze lijn in de grafiek.

c) Toets of de helling van deze lijn significant van nul verschilt ($\alpha = 0.05$).

d) Welke onderstellingen liggen aan deze toets ten grondslag?

e) Ga na in hoeverre aan die onderstellingen is voldaan.

7.10. Gegeven is de volgende serie van paren waarnemingen:

x	5	15	23	25	10	35	13	30	16	18
y	7	18	12	21	5	20	15	11	10	6

a) Bereken voor het model $\underline{y} = \alpha + \beta x + \underline{e}$ de schattingen voor α en β.

b) Geef het 95%-betrouwbaarheidsinterval voor $\underline{\hat{y}}$ behorende bij x = 20.

c) Als b) voor x = 30.

d) Als b) voor x = 50.

7.11. De volgende gegevens hebben betrekking op de leeftijd in weken (x) en de hoogte in cm (y) van maïsplanten:

x	1	2	3	4	5
y	6	17	23	34	39

a) Het regressiemodel is: $\mathcal{E}\underline{y} = \alpha x$. Bepaal de kleinste kwadratenschatting van α.

b) Wat is de geschatte variantie van $\underline{a}$, de schatter van α ?

7.12. In 1965 zijn op 12 achtereenvolgende dagen in Den Helder o.a. de luchttemperatuur in °C (x) en de luchtdruk in mbar (y) gemeten.

x	y
11.8	1013.9
12.5	1019.6
9.9	1020.6
10.1	1020.4
12.0	1014.9
13.5	1008.5
13.2	1009.1
14.0	1005.9
13.8	1010.0
11.6	1018.3
11.0	1020.9
13.2	1018.0

a) Bereken de correlatiecoëfficiënt r.

b) Bereken de lijn $\hat{y} = a + bx$.

c) Bereken de lijn $\hat{x} = c + dy$. (y wordt nu beschouwd als de onafhankelijk variabele.)

d) Zet x en y in een figuur tegen elkaar uit en teken in deze figuur beide regressielijnen.

e) Wat zijn de coördinaten van het snijpunt van beide regressielijnen?

8. Variantie-analyse.

8.1. In een landbouwkundige proef werden 4 tarwesoorten vergeleken voor wat betreft hun opbrengst. Elke soort werd op 3 proefveldjes verbouwd. De opbrengsten (in kg) waren als volgt:

Tarwesoort	Opbrengst		
A	28	25	20
B	21	24	30
C	21	22	21
D	20	20	18

Toets of er verschillen zijn tussen de tarwesoorten ($\alpha = 0.05$).

8.2. In een varkensfokkerij wordt een experiment uitgevoerd om de invloed van 3 verschillende soorten voeding op de gewichtstoename na te gaan. Daartoe worden 15 varkens in 3 groepen verdeeld. Elke groep krijgt een ander soort voedsel. De gewichtstoenamen (in kg), gemeten over een vast tijdsinterval, zijn:

voeding A: 133, 144, 135, 149, 143
voeding B: 163, 148, 152, 146, 157
voeding C: 210, 233, 220, 226, 229

Analyseer deze gegevens en geef Uw conclusie.

8.3. Van 5 benzinemerken is het octaangetal bepaald. De verkregen gegevens luiden als volgt:

Merk	Octaangetal
A	87, 91, 92
B	95, 95, 97
C	89, 90, 95
D	85, 84, 85
E	84, 80, 87

a) Voer een variantie-analyse uit.
b) Toets of er verschillen zijn tussen de merken wat betreft het octaangetal.
c) Aan welke onderstellingen moet voldaan zijn voor de toets onder b) ?

8.4. In een industrieel proces werd gesuggereerd dat de temperatuur waarbij een bepaalde bewerking wordt uitgevoerd van belang is voor de opbrengst. Om dit na te gaan werd de temperatuur op 4 achtereenvolgende dagen constant gehouden, iedere dag op een ander niveau. Iedere dag werd 4 maal de opbrengst bepaald.

Dag	Opbrengst in kg/100ℓ.			
1	27	32	33	29
2	44	44	51	43
3	30	33	39	29
4	43	47	36	43

a) Ga na of er een temperatuureffect is.

b) Welke bezwaren kunt U formuleren tegen deze proefopzet?

8.5. In een experiment dat uitgevoerd wordt om na te gaan of er verschil is tussen drie pH-meters (A, B en C) van een zeker type zijn 9 metingen uitgevoerd aan een bepaalde (stabiele) oplossing. De waarnemingen zijn:

A	3.05	3.07	3.08
B	3.12	3.10	3.18
C	3.08	3.14	3.09

(Het experiment werd zodanig uitgevoerd dat de waarnemer niet op de hoogte was van het feit dat bij deze metingen telkens dezelfde oplossing werd gebruikt.)

a) Ga na of er aanleiding is te concluderen tot het bestaan van verschillen tussen de 3 meters.

b) Geef een 95%-betrouwbaarheidsinterval voor de variantie binnen de Ph-meters.

8.6. Monsters van een bepaalde kwaliteit steenkool werden toegezonden aan 5 verschillende laboratoria die hieraan ieder een aantal malen het zwavelgehalte bepaalden. De resultaten waren:

Lab	Zwavelgehalte in gew.%						
A	3.18	3.14	3.12	3.14	3.25	3.23	
B	3.14	3.13	3.27				
C	3.02	3.07	3.04				
D	3.12	3.11	3.08				
E	3.18	3.20	3.22	3.14	3.09	3.10	3.10

a) Bereken voor elk laboratorium het gemiddelde, de kwadratensom en de variantie. Bereken tevens één enkele schatting voor de variantie, σ_0^2, binnen laboratoria.

b) Geef een 95%-betrouwbaarheidsinterval voor de standaardafwijking σ_i $(i = 1,\ldots,5)$ voor ieder laboratorium apart en voor de standaardafwijking σ_0 voor de 5 laboratoria gezamenlijk.

c) Toets de hypothese H_0: $\sigma_i^2 = \sigma_0^2$, $i = 1,\ldots,5$, m.b.v. de toets van Bartlett $(\alpha = 0.05)$.

d) Toets of er verschillen zijn tussen de laboratoria wat betreft het gemiddelde zwavelgehalte $(\alpha = 0.05)$.

8.7. Bij een studie van manueel werk werden snelle bewegingen van de rechterarm onderzocht. De getallen in de volgende tabel zijn verhoudingen van maximum en gemiddelde snelheid van 90 bewegingen.

M_1 = horizontale beweging van rechts naar links;

M_2 = horizontale beweging naar het lichaam toe;

M_3 = verticale beweging.

Bewegingstype		
M_1	M_2	M_3
1.59	1.55	1.85
1.53	1.53	1.60
1.50	1.47	1.80

a) Codeer de waarnemingen.

b) Voer een variantie-analyse uit.

c) Toets de hypothese dat het bewegingstype geen effect heeft $(\alpha = 0.05)$.

d) Aan welke onderstellingen moet voor het toetsen onder c) zijn voldaan?

9. Foutenvoortplanting.

9.1. Zij $\underline{z} = \underline{x}\underline{y}$; $\underline{x}$ en $\underline{y}$ zijn 2 onafhankelijke stochastieken met procentuele fout resp. 100 V(x) en 100 V(y). Wat is de procentuele fout in $\underline{z}$? (Exacte afleiding.) Waarin gaat deze over indien V(x) en V(y) klein zijn?

9.2. Zij $\underline{z} = \underline{x}\underline{y}$; $\underline{x}$ en $\underline{y}$ zijn onafhankelijke stochastische variabelen, vrij van systematische fouten.

a) Bereken de systematische fout in $\underline{\hat{z}}_g$.

b) Bereken var $\underline{z}$.

c) Laat zien dat $\overline{\underline{x}}\,\overline{\underline{y}}$ een nauwkeurigere schatter is voor z_g dan $\frac{1}{n}\sum_{i=1}^{n} \underline{x}_i\underline{y}_i$.

9.3. Gegeven: $\underline{x} \sim PS(\mu)$.

a) Wat is in eerste benadering $\mathcal{E}\sqrt{\underline{x}}$, wat var$(\sqrt{\underline{x}})$?

b) Leid hieruit een 95%-betrouwbaarheidsinterval af voor $\sqrt{\mu}$ en voor μ.

c) Vergelijk dit betrouwbaarheidsinterval voor μ voor enkele waarden van x met het 95%-betrouwbaarheidsinterval volgens tabel S.C. 8.1 en ga na voor welke waarden van x het onder b) gevonden interval reeds redelijke uitkomsten geeft.

9.4. Zij $\underline{z} = (\underline{x}+\underline{y})^2$. Hierin zijn $\underline{x}$ en $\underline{y}$ onafhankelijke statistische grootheden vrij van systematische fouten.

a) Bereken var$(\underline{z})$ bij benadering.

b) Bereken var$(\underline{z})$ exact. Neem daarbij aan dat $\underline{x}$ en $\underline{y}$ normaal verdeeld zijn zodat $\mathcal{E}(\underline{x}-\mu_x)^4 = 3\sigma_{\underline{x}}^2$ (zie S.C. pag. 11).

c) Vergelijk de twee uitkomsten voor de situatie: $\mu_x = \mu_y$, $V(x) = V(y) = \frac{1}{4}$.

9.5. Zij $\underline{z} = (\underline{x}+\underline{y})^2$; $\underline{x}$ en $\underline{y}$ onafhankelijk en vrij van systematische fouten.

a) Bereken de systematische fout in $\underline{\hat{z}}_g$ bij één waarneming $(\underline{x},\underline{y})$, die wordt gemaakt als tweedegraads termen worden verwaarloosd.

Er worden n onafhankelijke waarnemingen $(\underline{x}_i,\underline{y}_i)$ gedaan.

b) z_g wordt geschat door $\frac{1}{n}\sum_i (\underline{x}_i+\underline{y}_i)^2$. Bereken de systematische fout in $\underline{\hat{z}}_g$.

c) z_g wordt geschat door $(\bar{\underline{x}} + \bar{\underline{y}})^2$. Bereken de systematische fout in $\hat{\underline{z}}_g$.
d) Welke van de schatters b) en c) is de nauwkeurigste? Waarom?

9.6. Zij $\underline{y} = \underline{x}^2$. Rechtstreekse waarneming van $\underline{y}$ is niet mogelijk. Wel kan $\underline{x}$ worden gemeten zonder systematische fout. Bovendien is $\underline{x} \sim N(\mu,\sigma^2)$.
Geef benadering en exacte waarde van var($\underline{y}$). Controleer het verschil in de uitkomsten rechtstreeks aan de hand van informatie over de χ^2-verdeling (S.C. pag. 11).

10. Steekproefsystemen. Controlekaarten.

10.1. a) Bereken en teken met behulp van de tabellen S.C. 6.1 en S.C. 6.2 de keuringskarakteristiek voor het volgende dubbele steekproefschema:

$$n_1 = 50,\ n_2 = 100,\ c_1 = 1,\ c_2 = c_3 = 3.$$

Geef de berekeningen weer in een overzichtelijke tabel.

b) Teken in dezelfde figuur de keuringskarakteristiek voor het enkelvoudige schema:

$$n = 75,\ c = 2.$$

c) Bereken en teken voor het dubbele steekproefsysteem de kromme die het gemiddelde aantal te verrichten keuringen weergeeft als functie van het percentage foute produkten p, en vergelijk het resultaat met de steekproefgrootte van het enkelvoudige schema.

10.2. Een dubbel steekproefsysteem werkt als volgt.
Neem eerst een steekproef van 10 stuks. De partij wordt goedgekeurd als deze steekproef geen fouten bevat. Bij 2 of meer fouten wordt de partij afgekeurd. Is het aantal foute produkten 1, dan wordt een tweede steekproef van 10 stuks genomen. Bevat de tweede steekproef 0 fouten dan wordt de partij alsnog goedgekeurd. Bij één of meer fouten in de tweede steekproef wordt de partij afgekeurd.
Veronderstel dat een partij 10% fouten bevat.

a) Wat is de kans dat deze partij wordt afgekeurd?
b) Wat is de verwachting van de totale steekproefgrootte?

10.3. Aan radiobuizen moet een of andere karakteristieke grootheid worden gemeten. De volgende tabel geeft de resultaten voor steekproeven van 10 stuks, die elk uur uit de produktie werden genomen.

steekpr.nr.	$\bar{x}$	s	steekpr.nr.	$\bar{x}$	s
1	2.73	.18	14	2.80	.19
2	2.63	.17	15	2.77	.17
3	2.75	.12	16	2.67	.17
4	2.82	.14	17	2.69	.11
5	2.72	.08	18	2.79	.10
6	2.75	.13	19	2.73	.13
7	2.66	.20	20	2.74	.08
8	2.67	.15	21	2.69	.15
9	2.78	.17	22	2.61	.10
10	2.84	.13	23	2.63	.17
11	2.69	.07	24	2.76	.23
12	2.74	.11	25	2.77	.09
13	2.65	.12			

a) Ga met behulp van controlekaarten na of gemiddelde en/of standaardafwijking verlopen.

b) Geef een schatting van het percentage buizen dat niet voldoet aan de specificatie 2.70 ± .22.

c) Als het gemiddelde ingesteld kan worden, maar de spreiding rond het gemiddelde niet verbeterd kan worden, bepaal dan nieuwe regelgrenzen voor het gemiddelde van steekproeven van 10 stuks. Wat wordt nu de schatting voor het percentage buizen dat voldoet aan de specificatie?

10.4. Een fabriek maakt o.a. zekeringen (10A). Gedurende 2 dagen werd ieder uur een steekproef van 3 stuks genomen uit de produktie. Gemeten werd bij welke stroomsterkte (in A) de zekeringen doorbrandden. De resultaten waren:

steekpr.nr.	stroomsterkte in A			steekpr.nr.	stroomsterkte in A		
1	10.2	10.1	10.3	9	10.0	9.8	9.8
2	9.7	9.9	10.4	10	9.8	9.7	10.0
3	10.6	10.1	9.9	11	10.1	10.1	10.1
4	10.1	9.8	10.3	12	10.3	10.2	10.3
5	9.8	10.0	10.2	13	10.0	10.2	10.0
6	10.2	10.1	10.0	14	10.0	10.1	10.2
7	9.5	10.1	9.7	15	10.1	10.4	10.1
8	9.9	9.9	9.7	16	10.5	10.2	10.4

a) Bereken regelgrenzen voor de mediaan M van deze steekproeven.

b) Ga na of het fabricageproces beheerst is.

11. Parametervrije methoden.

11.1. Een dierenhandelaar krijgt een offerte van een nieuw soort kattenbrood (B) dat iets duurder is dan het kattenbrood (A) dat hij tot nu toe steeds verkocht. Om te beslissen of hij soort B zal kopen neemt hij een proef met 30 katten, die hij tegelijkertijd een bakje A en een bakje B voorzet. Hij wil soort B alleen kopen als blijkt dat B beter is dan A. Bij de proef bleken 10 katten A te prefereren boven B, terwijl 20 katten meer B aten. Wat zal volgens de tekentoets de beslissing van de handelaar zijn indien hij een kans 0.05 wil riskeren om de offerte te accepteren als A niet slechter is dan B ?

11.2. De directie van een fabriek bestudeert twee offertes voor een gemeenschappelijke ziekteverzekering voor de employees van het bedrijf. Alvorens een definitieve beslissing te nemen wil de directie beide offertes ter bestudering voorleggen aan een aantal aselect gekozen personeelsleden. Aan hen wordt gevraagd hun voorkeur uit te spreken. Dit zijn de resultaten:

Employee	A	B	C	D	E	F	G	H	I	J	K	L
voorkeur voor offerte	1	2	1	2	2	2	2	1	1	2	2	1

Welke offerte heeft de voorkeur?

11.3. Dertig nieuwe medewerkers werden op basis van intelligentie en ervaring ingedeeld in 15 paren. Zij volgden daarna een cursus "data-processing" waarbij twee onderwijsmethoden werden toegepast. De oude methode A werd gebruikt voor een groep van 15 personen, waarvan elk aselect werd gekozen uit elk paar. De overige 15 personen volgden de cursus, waarbij een vermoedelijk betere onderwijsmethode B werd toegepast. Aan het einde van de cursus werd een tentamen afgenomen, waarbij men maximaal 100 punten kon scoren.
Toets de hypothese dat beide methoden even effectief zijn tegen het alternatief dat methode B inderdaad beter is, als de resultaten waren:

A	60	70	80	85	75	40	70	45	95	80	90	60	80	75	65
B	65	85	85	80	95	65	100	60	90	85	100	75	90	60	80

11.4. Monsters van twee soorten verf werden gedurende 3 maanden aan bepaalde weercondities blootgesteld. De scores voor de weerbestendigheid van de twee soorten verf waren:

A: 92 95 94 85 82

B: 78 74 81 69 88 75

Ga met behulp van de toets van Wilcoxon na of er verschil is tussen beide soorten verf.

11.5. Gebruik de toets van Wilcoxon om na te gaan of het inspuiten van insuline invloed heeft op het glycogeenpercentage in de spieren van dieren bij de volgende meetresultaten:

ingespoten dieren	.15	.13	.00	.07	.27	.24	.19	.04	.08	.20	.12
controledieren	.19	.18	.21	.30	.66	.42	.08	.12	.30	.27	

11.6. Een onafhankelijke instantie scoorde winkelpersoneel uit wijkwinkels en supermarkten op algemeen voorkomen, behulpzaamheid, produktkennis, enz. Deze scores waren:

supermarkt: 32, 60, 88, 16, 43, 70, 97, 23, 49, 74, 36

wijkwinkel: 24, 49, 73, 97, 97, 21, 44, 67, 90, 13, 29

Toets de hypothese dat beide populaties dezelfde verdeling hebben ($\alpha = 0.05$).

11.7. In een bedrijf werden 10 personeelsleden gescoord voor twee vaardigheden A en B. De resultaten waren:

	1	2	3	4	5	6	7	8	9	10
vaardigheid A	20	14	10	18	12	9	25	23	21	18
vaardigheid B	17	12	11	16	10	8	22	20	19	14

Toets m.b.v. de rangcorrelatietoets van Spearman of er verband is tussen de scores voor de twee vaardigheden.

11.8. Bij 15 vrouwtjesratten vindt men de volgende waarden voor het begingewicht (x) en de gewichtstoename (y) van de 28-ste tot aan de 84-ste dag met een proteïnerijk dieet (beide gewichten in gr):

Rat	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15
x	50	64	76	63	74	60	69	68	56	48	57	59	46	45	65
y	128	159	158	119	133	112	82	126	132	118	107	106	96	103	104

Toets de hypothese dat begingewicht en gewichtstoename onderling onafhankelijk zijn ($\alpha = 0.05$).

11.9. De rangcorrelatietoets van Spearman kan ook worden gebruikt om de hypothese te toetsen, dat in een waarnemingsreeks geen verloop optreedt. Dit komt neer op toetsing van de hypothese dat tussen de waarnemingen en de tijdstippen waarop deze betrekking hebben geen correlatie aanwezig is tegen het alternatief dat er een positieve (stijgend verloop) of negatieve (dalend verloop) correlatie bestaat.

Van een patiënt is op 16 achtereenvolgende dagen de systolische bloeddruk gemeten:

dag	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16
bloeddruk	113	115	114	117	118	120	124	125	126	131	128	134	141	143	147	148

Ga na of er verloop is ($\alpha = 0.05$).

11.10. Onderstaande tabel geeft de aantallen bloedplaatjes bij voldragen kinderen op de 1e, 3e en 5e levensdag. Toets met de methode der m rangschikkingen of de verschillen tussen de kinderen op toeval berusten ($\alpha = 0.05$).

kind	1e dag	2e dag	3e dag
1	195 860	150 000	152 600
2	219 000	212 000	244 800
3	235 600	232 600	228 800
4	148 980	183 600	161 280
5	173 900	186 200	173 950
6	149 300	173 800	152 800
7	171 440	169 860	179 580
8	284 480	206 450	231 840
9	280 000	244 800	248 160
10	161 000	169 740	194 000

11.11. Men heeft van een bepaalde stof 3 verschillende oplossingen gemaakt van sterktes van ongeveer 10%, 15% en 20%. Iedere oplossing wordt over 7 flesjes verdeeld, die aan 7 verschillende laboratoria worden toegezonden met het verzoek door middel van een nauwkeurig omschreven chemische analyse het gehalte van de stof in kwestie te bepalen. De 3 flesjes die ieder laboratorium kreeg toegezonden waren naar opklimmende sterkte gemerkt met A, B en C.

De volgende resultaten zijn verkregen:

Lab	A	B	C
1	10.23	15.17	21.10
2	10.29	15.52	21.18
3	10.22	15.20	21.07
4	10.33	15.31	21.12
5	10.42	15.40	21.82
6	10.15	15.03	21.05
7	10.20	15.19	21.01

Het is bekend dat de variantie van de waarnemingsuitkomsten bij toenemend gehalte niet constant is. De variantie bij hetzelfde gehalte op verschillende laboratoria mag als constant worden beschouwd. Gevraagd wordt na te gaan of er systematische verschillen tussen de laboratoria bestaan.

11.12. Van 12 arbeiders werd de gemiddelde uurproduktie bepaald volgens 4 verschillende werkmethoden. Deze gemiddelde aantallen per uur waren:

	Methode			
Arbeider	A	B	C	D
1	40	50	48	44
2	53	63	48	41
3	46	40	55	61
4	45	64	54	44
5	48	64	42	48
6	65	56	64	46
7	63	59	64	56
8	69	52	64	56
9	69	42	58	44
10	59	61	53	62
11	50	64	53	55
12	53	53	43	55

Toets of er verschil is tussen de werkmethoden wat betreft de gemiddelde uurproduktie ($\alpha = 0.05$).

12. Gemengde opgaven.

12.1. Voor een aselecte steekproef uit de Nederlandse bevolking wordt nagegaan uit hoeveel personen het huishouden waarvan de ondervraagden deel uitmaken bestaat.

a) Is het gemiddelde aantal personen per huishouden uit deze steekproef berekend een zuivere schatting van het landelijk gemiddelde aantal
   i) indien de steekproef uit de gehele bevolking wordt getrokken;
   ii) indien de steekproef zich beperkt tot het mannelijke gedeelte der bevolking?

b) Hoe groot moet een steekproef uit de huishoudens zijn om de gemiddelde grootte van een huishouden met 95% betrouwbaarheid op ± 0,1 persoon nauwkeurig te kunnen schatten?

Motiveer de antwoorden!

Toelichting. Het onderzoek geldt privé-huishoudens. Pensions, weeshuizen, enz., worden dus niet meegeteld. Alleenstaanden vormen een huishouden van één persoon.

12.2. Bij een proef wordt de straling van een radio-aktieve bron gemeten met een Geigerteller, respectievelijk met 0,1,...,10 plaatjes aluminium tussen bron en teller. Hieronder volgen de resultaten van deze metingen verricht door 2 groepen studenten:

aantal plaatjes	0	1	2	3	4	5	6	7	8	9	10
groep 1	987	974	968	958	950	943	939	939	927	920	910
groep 2	964	951	952	938	939	932	920	920	907	903	908

De dikte van een aluminium plaatje is 1.08 mm.

Voor de getallen y_{ij} ($i = 1,2$; $j = 0,\ldots,10$) geldt: $y_{ij} = 10^3(3 - {}^{10}\log T_{ij})$.

Hierin is T_{ij} ; tijd in sec. benodigd voor 10000 aanslagen.

a) Laat met behulp van de wet van de voortplanting van fouten zien dat $\mathrm{var}(\underline{y}) = 18.9$.

b) Zij x_j ($j = 0,\ldots,10$) de totale dikte van de voorgeschakelde plaatjes. Voer voor beide groepen studenten een lineaire regressie-analyse uit volgens het model:

$$\underline{y}_{ij} = \beta_{0i} + \beta_{1i}(x_j - \bar{x}_.) + \underline{e}_{ij} \, .$$

c) Zijn $\hat{\sigma}_{01}^2$ en $\hat{\sigma}_{02}^2$ de beide geschatte restvarianties, toets dan H_0: $\sigma_{0i}^2 = 18.9$ $(i = 1,2)$, en geef een 95%-betrouwbaarheidsinterval voor de verhouding

$$\frac{\sigma_{01}^2}{\sigma_{02}^2}.$$

d) Toets de nulhypothesen

i) H_0: $\beta_{01} = \beta_{02}$;
ii) H_0: $\beta_{11} = \beta_{12}$.

e) Bij het model onder b) horen een aantal onderstellingen. Ga na in hoeverre daaraan is voldaan.

12.3. a) Uit 5 grote partijen neemt men steekproeven ter grootte van resp. 150, 300, 150, 200 en 100 stuks en vindt hierin resp. 4, 10, 5, 5 en 3 afgekeurde produkten. Toets de hypothese: de partijen hebben eenzelfde percentage foute produkten.

b) Twee groepen A en B van elk 100 personen hebben een bepaalde ziekte. Men heeft A een serum gegeven, B (de "controlegroep") echter niet. Na een zekere tijd waren 75 van A en 65 van B genezen. Toets de nulhypothese dat het serum geen effect heeft gehad m.b.v.
i) de χ^2-toets in een 2×2 tabel,
ii) de u-toets voor het verschil van fracties.
Gebruik in beide gevallen $\alpha = 0.05$.

c) Wat is het verband tussen de beide onder b) berekende toetsingsgrootheden? Licht dit toe.

d) Als U bij een χ^2-toets vindt $\chi_6^2 = 0.57$, wat is dan Uw conclusie?

12.4. Men heeft de samenstelling onderzocht van 320 gezinnen met 5 kinderen en de volgende frequentieverdeling gevonden:

jongens	5	4	3	2	1	0	
meisjes	0	1	2	3	4	5	totaal
frequentie	18	56	110	88	40	8	320

a) Toets H_0: $p_j = p_m = \frac{1}{2}$ (p_j = kans op een jongen).

b) Construeer een 95%-betrouwbaarheidsinterval voor p_j.

c) Welke onderstellingen liggen aan dit interval ten grondslag?

d) Toets de hypothese dat de frequentieverdeling een binomiale is met $p_j = \frac{1}{2}$.

e) Toets de hypothese dat de frequentieverdeling een binomiale is met een onbekende waarde van p_j.

12.5. Uit partijen produkten van 1000 of meer stuks worden steekproeven van 80 stuks genomen om het aantal foute exemplaren hierin te bepalen. Een serie van 80 steekproeven uit 80 verschillende partijen gaf het volgende resultaat:

$\underline{x}$ = aantal foute produkten	0	1	2	3	4	5	6	7	8	9	10
n = aantal steekproeven	13	20	18	13	7	4	1	0	2	1	1

a) Toets de hypothese dat het aantal foute exemplaren Poisson-verdeeld is.

b) Geef een 95%-betrouwbaarheidsinterval voor $\mathcal{E}\underline{x}$

i) in de onderstelling dat $\underline{x} \sim PS(\mu)$;

ii) voor het geval niet ondersteld wordt dat $\underline{x} \sim PS(\mu)$.

12.6. Pakjes margarine worden geproduceerd volgens een normale verdeling met

$$\mu = 252 \text{ gr}, \quad \sigma = 2 \text{ gr}.$$

Een automatische weegmachine splitst de pakjes in twee groepen:

groep A met een gewicht $\leq$ 250 gr,

groep B met een gewicht $>$ 250 gr.

a) Welk percentage van de produktie valt in groep A ?

b) Bereken voor elk der groepen het gemiddelde gewicht van de pakjes.

12.7. Het gewicht van "treets" (met chocolade omklede pinda's) is normaal verdeeld met

$$\mu = 2.0 \text{ gr}, \quad \sigma = 0.3 \text{ gr}.$$

Pakjes "treets" worden automatisch gevuld tot het totale gewicht $\geq$ 40 gr is.

a) Wat is de verdeling van het aantal "treets" in de pakjes?

b) Wat is het gemiddelde gewicht van een pakje "treets"?

12.8. Gegeven 10 waarnemingen:

x_{1j}:	60.0	59.5	58.1	59.9	58.0
x_{2j}:	60.8	60.1	58.4	59.6	59.6

Codeer als volgt:

$$u_{1j} = 10(x_{1j} - 59) \;; \quad u_{2j} = 10(x_{2j} - 59) \;.$$

Gegeven is verder

$$\sum_j u_{1j}^2 = 387 \;; \quad \sum_j u_{2j}^2 = 553 \;; \quad u_{1.} = 5 \;; \quad u_{2.} = 35 \;.$$

Geef een schatting van de restvariantie σ_0^2 (met bijbehorend aantal vrijheidsgraden en bijbehorend model) in de volgende gevallen:

a) de beide reeksen waarnemingen vormen samen één reeks onafhankelijke waarnemingen uit dezelfde populatie;

b) de beide reeksen waarnemingen komen uit populaties met verschillende μ doch dezelfde σ;

c) de waarnemingen vormen 5 duplo-metingen;

12.9. Gegeven zijn r onafhankelijke stochastische variabelen $\underline{x}_i$ met $\mathcal{E}\underline{x}_i = \mu$ en $\mathrm{var}(\underline{x}_i) = \sigma_i^2$.

Gevraagd een lineaire schatter voor μ (d.w.z. $\underline{\hat{\mu}} = \sum_{i=1}^{r} \beta_i \underline{x}_i$), die zuiver is en minimale variantie heeft.

12.10. In onderstaande tabel zijn de resultaten opgenomen van metingen aan de diameter van 16 in micaplaatjes geponste gaten. Deze metingen zijn verricht door optische vergroting op een "projectiekast".

Diameter van gaten in mica (in mm)

.870	.865	.845	.860
.855	.840	.855	.855
.860	.850	.850	.865
.860	.855	.850	.865

a) De waarnemingen zijn klaarblijkelijk op 5μ nauwkeurig opgegeven. Is deze afronding statistisch gezien toelaatbaar? (Motiveer Uw antwoord!)

b) Codeer de waarnemingen en bereken het gemiddelde en de variantie van deze (aselecte) steekproef.

c) Uit het verleden is bekend dat de diameter normaal verdeeld is. De gemiddelde diameter moet .855 mm zijn. Voldoet de partij waaruit de steekproef is genomen aan deze eis? (Motiveer; $\alpha = 0.05$.)

d) Welke onderstellingen liggen ten grondslag aan de onder c) gebruikte toets?

e) Geef een betrouwbaarheidsinterval voor σ^2 $((1-\alpha) = 0.95)$.

f) Uit een volgende partij micaplaatjes werd een aselecte steekproef genomen van 8 stuks. De meetresultaten waren:

.850	.860
.850	.850
.860	.870
.865	.860

Toets met een betrouwbaarheid van 95% of deze partij dezelfde gemiddelde diameter heeft als de eerste partij. Laat daarbij zien dat aan de nodige onderstellingen is voldaan.

12.11. a) Wijs door loting één van de 6 kolommen van de bij dit vraagstuk behorende tabel aan.

b) Maak voor deze kolom een frequentietabel en bereken gemiddelde en standaardafwijking.

c) Controleer de normaliteit door de cumulatieve frequentieverdeling uit te zetten op een lineaire waarschijnlijkheidsschaal en schat hieruit gemiddelde en standaardafwijking.

d) Wijs door loting uit de aangewezen kolom één van de 5 groepen van 10 waarnemingen aan en toets voor deze groep de normaliteit met de methode van Shapiro en Wilk. (Bereken zowel W als G.)

e) Maak een controlekaart voor gemiddelde en range door de aangewezen kolom te splitsen in 25 paren.

f) Voer een variantie-analyse uit op de eerste 5 rijen en de 6 kolommen van de tabel en toets of er verschil is tussen kolommen.

De tabel bevat gecodeerde garengewichten voor 6 posities van een automatische wikkelmachine.

Tabel behorende bij vraagstuk 12.11.

60	67	116	108	48	118
44	127	132	60	90	53
74	75	96	4	88	74
65	97	123	84	198	66
53	86	134	64	120	94
53	79	103	26	97	118
78	117	129	93	80	115
64	139	109	89	95	117
72	108	117	103	104	73
97	97	142	115	108	137
126	111	104	123	106	151
128	97	107	102	80	82
126	104	149	58	109	133
4	135	119	97	92	118
129	146	106	113	114	131
121	68	119	44	136	129
136	70	125	113	-2	75
79	73	144	118	115	129
135	20	108	57	112	149
125	119	127	132	125	98
89	95	142	114	131	121
69	135	152	126	108	90
76	118	79	145	136	126
130	84	120	97	173	106
109	123	82	75	123	104
80	118	131	63	166	116
92	150	79	63	125	191
85	190	64	96	137	124
116	136	131	103	111	77
141	137	101	155	124	153
77	130	128	125	124	114
53	86	127	57	114	112
47	93	134	145	125	165
123	90	107	102	102	121
88	78	94	98	127	82
55	128	146	99	139	85
61	92	10	97	131	106
119	115	104	128	108	168
96	118	129	104	144	128
111	123	52	135	110	147
82	156	88	105	130	114
105	69	96	137	75	154
116	106	104	109	55	143
95	94	69	129	89	146
123	114	151	136	109	125
155	102	112	128	100	165
122	68	96	120	90	79
93	118	141	153	138	135
107	110	110	126	106	117
97	114	134	117	22	128

12.12. Bij de omzetting van suiker m.b.v. een katalysator wordt om de 5 minuten het gehalte niet omgezette suiker bepaald en daaruit de reaktiesnelheid berekend voor iedere periode. De resultaten waren achtereenvolgens:

.00504; .00541; .00483; .00513; .00499; .00504; .00524;
.00529; .00534; .00544; .00549.

Er werd gesuggereerd dat de experimentele omstandigheden tijdens de periode waarin de laatste 5 bepalingen waren uitgevoerd, veranderd waren.

a) Ga met behulp van de t-toets na of dat inderdaad het geval was.
b) Welke essentiële onderstellingen liggen aan deze toets ten grondslag?
c) Is er reden aan één of meer van deze onderstellingen te twijfelen, en zo ja, waarom? (Aanwijzing: zet de waarnemingen grafisch uit.)

12.13. Zij $\underline{x}$ een stochastische variabele met een standaard homogene verdeling:

$$f(x) = 1 , \quad 0 \leq x \leq 1$$
$$f(x) = 0 , \quad \text{elders.}$$

a) Wat is de verdeling van de mediaan, $\underline{M}$, van een aselecte steekproef van 3 waarnemingen?
b) Wat is $\mathcal{E}\underline{M}$ en var $\underline{M}$?

12.14. Wanneer bij een aselecte steekproef uit de Nederlandse bevolking wordt nagegaan uit hoeveel personen het huishouden bestaat, waartoe de ondervraagden behoren, dan hebben grotere huishoudens een grotere kans in de steekproef te worden opgenomen dan de kleinere. Zij nu $P(n)$, $n = 1,2,\ldots$, de ware verdeling en $F(n)$, $n = 1,2,\ldots$, de door dergelijke steekproeven verkregen verdeling van het aantal personen, n, per huishouden.

a) Druk $F(n)$ uit in n en $P(n)$.
b) Hoe kan $\mathcal{E}\underline{n}$ worden bepaald uit $F(n)$?
c) Druk $P(n)$ uit in n en $F(n)$.
d) Gegeven een populatie met $P(n) = \frac{1}{4}$, $n = 1,2,3,4$. Trek met behulp vam tabel S.C. 8.4 een steekproef van 100 waarnemingen uit de hiermee corresponderende verdeling $F(n)$ en demonstreer hieraan de toepasbaarheid van de onder b) en c) gevonden methode.

12.15. Onderstaande tabel bevat 240 posten uit een kasboek, gesplitst in 3 deelpopulaties met resp. de kleine, middelgrote en grote uitgaven:

I: $x < 10$ gld.; $N_1 = 180$, $\mu_1 = 3.375$, $\sigma_1^2 = 5.682$

II: $10 \le x < 30$ gld.; $N_2 = 49$, $\mu_2 = 16.962$, $\sigma_2^2 = 34.418$

III: $30 \le x$ gld.; $N_3 = 11$, $\mu_3 = 60.727$, $\sigma_3^2 = 609.568$

I

2.12	2.55	.65	3.50	5.90	3.60
7.80	6.25	2.50	9.00	6.50	1.10
2.00	3.00	1.00	6.27	2.15	5.50
3.61	1.50	3.40	3.00	5.00	2.00
7.50	1.50	1.38	4.25	2.25	2.10
1.00	1.05	8.50	.60	4.55	5.50
4.50	1.00	1.00	2.60	2.25	6.00
4.50	.54	2.20	3.00	6.50	1.05
2.50	2.50	2.00	2.50	1.50	2.00
3.35	2.00	4.00	3.00	.55	2.50
.60	4.50	5.00	8.00	3.00	2.50
4.00	2.60	4.90	3.54	.50	.80
2.10	6.50	5.00	1.15	.50	8.50
4.90	2.00	3.30	4.25	1.00	2.00
7.50	8.00	2.15	1.50	3.60	8.75
4.85	3.56	9.20	1.25	.95	2.65
5.15	2.25	.75	.08	9.00	1.00
3.00	3.00	2.50	4.85	1.50	1.00
4.85	2.85	.35	1.00	6.00	.83
.90	4.60	.35	.80	6.30	4.63
2.55	4.50	.75	2.12	.25	6.00
5.86	2.95	6.00	2.25	.80	3.16
2.50	3.50	3.50	.45	9.23	4.50
3.75	7.25	4.75	6.45	1.17	4.12
1.25	3.95	4.25	.35	7.50	.30
2.54	1.25	6.75	9.50	2.50	2.25
2.50	2.50	3.00	6.00	2.25	7.40
6.75	1.00	9.75	1.80	6.61	2.50
.91	.85	2.40	.60	2.15	2.25
5.00	1.00	2.50	6.23	2.25	2.25

II

14.00	20.00	10.00	20.00	17.00
12.50	10.00	20.75	20.00	11.30
14.00	25.00	26.00	27.00	14.65
10.50	10.00	15.00	16.50	13.25
20.00	23.75	25.00	27.75	27.50
10.00	26.90	10.75	10.00	15.00
10.40	10.00	20.50	12.90	22.50
22.50	10.63	12.50	12.50	10.92
20.00	14.90	21.25	27.50	19.75
14.40	15.00	17.90	11.10	

III

47.50
35.00
105.50
30.00
55.00
50.00
50.00
65.00
80.00
100.00
50.00

a) Wat zijn de waarden van μ en σ^2 voor de gehele populatie? (N = 240.)

b) Men trekt uit de 3 deelpopulaties aselect steekproeven van resp. n_1, n_2 en n_3 stuks. Uit de steekproefgemiddelden $\bar{x}_1$, $\bar{x}_2$ en $\bar{x}_3$ berekent men een schatting van het totale bedrag der gehele populatie als volgt:

$$\hat{\underline{T}} = N_1\bar{\underline{x}}_1 + N_2\bar{\underline{x}}_2 + N_3\bar{\underline{x}}_3 .$$

Als de totale steekproefgrootte $n = n_1 + n_2 + n_3$ gegeven is, hoe moeten dan n_i, $i = 1,2,3$, gekozen worden zodat de schatting $\hat{\underline{T}}$ zo nauwkeurig mogelijk wordt? Wat is $\text{var}(\hat{\underline{T}})$ in dat geval?

c) Bereken de verhouding n'/n, als n' de grootte is van een aselecte steekproef uit de gehele populatie, zodanig dat de hieruit berekende schatting van T even nauwkeurig is als die onder b).

d) Trek met behulp van tabel S.C. 8.4 een aselecte steekproef van n = 25 posten volgens de onder b) gegeven methode en bepaal de 2-zijdige overschrijdingskans van de daaruit berekende $\hat{T}$.

12.16. Grammofoonplaten worden geperst tussen matrijzen. Zo nu en dan komt in de persmassa een zandkorrel voor waardoor op één van de matrijzen een kras ontstaat en deze door een nieuwe matrijs moet worden vervangen.
Zij p de kans dat bij een plaat een kras ontstaat. We nemen aan dat krassen onafhankelijk van elkaar optreden. Zij verder $\underline{x}$ het aantal platen zonder kras dat per matrijs kan worden geproduceerd.

a) Wat is de verdeling van $\underline{x}$?

b) Bereken $\mathcal{E}\underline{x}$ en var $\underline{x}$.

c) Beschouw alleen die matrijzen waarmee zonder ongelukken x_0 platen zijn gemaakt. Zij $\underline{y}$ het aantal platen (zonder kras) dat daarna nog met deze matrijzen kan worden gemaakt. Wat is de verdeling van $\underline{y}$? (M.a.w.: wat is de verdeling van $\underline{y} = \underline{x} - x_0$ onder de voorwaarde $\underline{x} \geq x_0$?)

d) Stel p = 0.02. De fabrikant ontvangt een order voor 3000 platen. Hoeveel matrijzen moet hij aanmaken, indien hij slechts een risico van 5% wil lopen dat hij met deze voorraad matrijzen de produktie van 3000 platen niet zal kunnen voltooien? (Maak gebruik van de centrale limietstelling!)

12.17. Oxide kathodes worden bespoten met een laag Barium-Strontium-carbonaat. Uit 12 partijen werd aan 5 kathodes het gewicht van de carbonaatlaag gemeten.
De variantie-analyse van de gegevens luidt:

Bron	KS	ν	GK	
Tussen partijen	81.73	11	7.43	$F^{11}_{48} = 2.90$
Binnen partijen	122.90	48	2.56	

De F-waarde heeft een overschrijdingskans van ongeveer 1% en is dus duidelijk significant.

a) Wat is de nulhypothese die met F wordt getoetst?

b) Wat zijn de bijbehorende onderstellingen?

12.18. a) $\underline{x} \sim N(0,\sigma^2)$. Bereken $\mathcal{E}|\underline{x}|$ en $\mathrm{var}|\underline{x}|$.

b) Uit een massaproduktie worden regelmatig steekproeven genomen van n stuks. Binnen één steekproef heeft ieder produkt een constante kans p om een bepaalde fout te tonen, doch tussen steekproeven is p een stochastische variabele $\underline{p}$. Zij $\underline{x}$ het aantal foute produkten in de steekproeven, dan geldt:

$$\mathrm{var}\,\underline{x} = n\,\mathcal{E}\,\underline{p}(1-\mathcal{E}\underline{p}) + (n^2-n)\,\mathrm{var}\,\underline{p}\ .$$

Bewijs dit.

12.19. Een partij wordt gekeurd met een steekproef van 200 stuks. Geëist wordt dat het risicopunt van de producent (p_{95}) hoogstens 3% mag zijn.
Wat volgt hieruit voor het risicopunt p_{10} van de afnemer?

12.20. In 1956 werden 76.753 rekruten gekeurd. Hiervan werden volgens het statistisch zakboekje afgekeurd:

1328 of 1.73% wegens neurosen of psychopatieën,
6628 of 8.64% wegens intellectstoornissen,
2091 of 2.72% wegens oogafwijkingen,
644 of 0.84% wegens hartafwijkingen.

Waarom kunnen deze percentages niet zonder meer worden geïnterpreteerd als zuivere schattingen van het voorkomen van deze afwijkingen onder de gekeurde rekruut?

12.21. Zij $\underline{x} \sim N(\mu = 10.0, \sigma^2 = 0.64)$. Bereken de kans dat

a) het gemiddelde van 3 waarnemingen groter is dan 10.6;
b) het gemiddelde van 6 waarnemingen groter is dan 10.6;
c) tenminste één van 3 waarnemingen groter is dan 10.6;
d) juist één van 3 waarnemingen groter is dan 10.6;
e) alle waarnemingen groter zijn dan 10.6.

Waarom zijn de uitkomsten a) en e) verschillend?

12.22. Zij $\underline{m}$ een stochastische variabele met

$$\underline{m} > 0 \ , \quad \mathcal{E}\underline{m} = \mu \ , \quad \text{var } \underline{m} = \sigma^2 \ .$$

Zij verder $\underline{x} \sim PS(\underline{m})$. Bereken $\mathcal{E}\underline{x}$, $\mathcal{E}\underline{x}^2$ en var $\underline{x}$.

12.23. De volgende gegevens zijn ontleend aan de resultaten der volkstellingen van 1920 en 1930.

Provincie	Percentage Roomskatholieken	
	1920	1930
Groningen	5.7	5.4
Friesland	7.0	7.0
Drente	6.2	6.1
Overijssel	27.6	28.4
Gelderland	36.1	36.6
Utrecht	31.7	31.0
Noord-Holland	27.2	27.2
Zuid-Holland	24.0	23.8
Zeeland	25.7	25.1
Noord-Brabant	89.1	88.6
Limburg	94.6	93.4
Het Rijk	35.6	36.4

Voor de provincies zijn de percentages in 1930 vrijwel over de gehele linie lager dan in 1920, voor het Rijk als geheel (zonder de overzeese gebiedsdelen) daarentegen is het percentage in 1930 hoger dan in 1920. Hoe verklaart U dit?

12.24. De toepassing van een bepaald geneesmiddel blijft gemiddeld in één van de 10 gevallen zonder succes. Men neemt een nieuw geneesmiddel in gebruik en vindt daarbij onder 160 behandelde patiënten 6 gevallen zonder succes.

a) Wat is de nulhypothese, wat zijn de alternatieve hypothesen?
b) Toets de nulhypothese ($\alpha = 0.05$).
c) Wat zijn onder b) de onderstellingen?
d) Waarom zal de medische wereld de uitkomst van de proef nooit als afdoende bewijs ten gunste van het nieuwe geneesmiddel aanvaarden?

12.25. Tabel: totale jaarlijkse regenval in De Bilt van 1900-1959 (in mm).

1900	724	1910	818	1920	640	1930	911	1940	860	1950	950
	705		641		398		772		742		809
	625		1027		649		744		782		792
	926		719		841		511		669		597
	596		785		778		710		808		818
	776		891		899		917		799		660
	723		895		712		750		823		752
	636		782		875		727		752		928
	616		884		815		782		736		828
	831		742		626		839		668		536

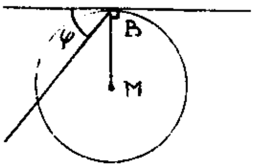
a) Maak frequentietabellen met intervalbreedten van 50, 75 en 100 mm.
b) Bereken voor de 3 tabellen gemiddelde en standaardafwijking.
c) Welke van de frequentietabellen geeft naar Uw mening de inhoud van de tabel het meest bevredigend weer?
d) Maak van de door U gekozen tabel een histogram.
e) Toets met de grafische methode en met de door U gekozen frequentietabel of de jaarlijkse regenval normaal verdeeld is.

Antwoorden

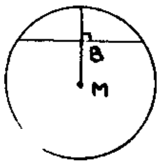
1.1. 49/50.

1.2. Zonder teruglegging.

1.3. Tweemaal werpen met munt: de ene wint bij MK, de ander bij KM.

1.4. 1^e 

Kies B op omtrek, trek raaklijn en straal in B. De hoek φ is rechthoekig verdeeld op $(0,\pi)$. P(koorde > zijde) = $= P(\pi/3 < \underline{\varphi} < 2\pi/3) = 1/3$.

2^e 

Trek een straal en kies B op die straal. De koorde trekken we door B loodrecht op de straal. Dan is P(koorde > zijde) = $P(\underline{BM} < \tfrac{1}{2}r) = \tfrac{1}{2}$.

Het antwoord hangt dus af van het "at random" trekken van de koorde.

1.5. P('n merk tweemaal op de eerste plaats) = $4(\frac{1}{6})^2 = \frac{1}{9} > \alpha$.

1.6. $\{(1+t)^n\}^2 = (1+t)^{2n} \Leftrightarrow \{\sum_{i=0}^{n} \binom{n}{i} t^i\}^2 = \sum_{i=0}^{2n} \binom{2n}{i} t^i$.

Beschouw de coëfficiënt van t^n.

linkerlid: $1 + \binom{n}{1}\binom{n}{n-1} + \ldots + \binom{n}{n-1}\binom{n}{1} + 1 = \sum_{i=0}^{n} \binom{n}{i}^2$.

rechterlid: $\binom{2n}{n}$.

1.7. $P = \binom{365}{22}\binom{22}{3}\frac{25!}{(2!)^3}/(365)^{25} = 0.0269$.

1.8. a) $P = 10^{-3}$ b) $P = \binom{10}{2}\binom{2}{1}\frac{4!}{3!1!}/10^4 = 0.036$

c) $P = \binom{10}{2}\frac{4!}{2!2!}/10^4 = 0.027$ d) $P = \binom{10}{3}\binom{3}{1}\frac{4!}{2!}/10^4 = 0.432$

e) $P = \binom{10}{4}4!/10^4 = 0.504$.

De som der kansen is 1.

1.9. $2^5 = 32$.

1.10. 7/37 (Bayes-theorema).

1.11. a) 4/5 b) $\frac{1}{2}$.

1.12. $P(ABC) = P(A)P(BC) = P(A)P(B)P(C|B) = P(A)P(B).2P(C) = \frac{1}{4} \to P(A) = P(B) =$
$= P(C) = \frac{1}{2}$

$P(C|B) = 2P(C) = 2.\frac{1}{2} = 1$

$P(B|C) = \frac{P(BC)}{P(C)} = \frac{P(ABC)}{P(A)P(C)} = \frac{\frac{1}{4}}{\frac{1}{2}.\frac{1}{2}} = 1.$

1.13. a) 2/5 b) 7/20.

1.14. a) $(5/6)^n$ b) $\frac{1}{6}\,n(5/6)^{n-1}$ c) $1 - (5/6)^n$

d) $(2/3)^n$ e) $\frac{1}{6}\,n(2/3)^{n-1}$ f) $(5/6)^n - (2/3)^n$

g) $1 - (2/3)^n$ h) $1 - 2(5/6)^n + (2/3)^n$ i) $2\binom{n}{2}(1/6)^2(4/6)^{n-2}$.

1.15. a) $P(A)/(P(A) + P(B))$ b) $\binom{n}{2}P^{n-2}(A)P^2(B)$.

1.16. P(stoppen na a)) $= 1 - (p_1 + p_2)^n$

P(stoppen na b)) $= (p_1 + p_2)^n - p_1^n - np_1^{n-1}p_2$

P(stoppen in 3^e situatie) $= np_1^{n-1}p_2(1 - p_1)^n = np_1^{n-1}p_2 - np_1^{2n-1}p_2$

Dan is de som:

P(stoppen) $= 1 - p_1^n - np_1^{2n-1}p_2$.

Andere oplossing: P(niet stoppen) $= p_1^n + np_2p_1^{n-1}p_1^n = p_1^n + np_2p_1^{2n-1}$.

1.17. a) $P = \sum_{i=1}^{9} (-1)^{i+1}\binom{10}{i}\left(\frac{10-i}{10}\right)^{60} \approx 10.\left(\frac{9}{10}\right)^{60} = 0.018$

b) $P = \sum_{i=1}^{9} (-1)^{i+1}\binom{10}{i}\left(\frac{10-i}{10}\right)^{120} \approx 10.\left(\frac{9}{10}\right)^{120} = 0.00003.$

1.18. a) $f(a) := \Sigma(x_i - a)^2/n-1$. Dan $f'(a) = 0 \Rightarrow a = \bar{x}$, $f''(\bar{x}) = \frac{2n}{n-1} > 0$.

1.19. a) $\mathcal{E}(n-1)\underline{s}^2 = \mathcal{E}\{\Sigma \underline{x}_i^2 - n\underline{\bar{x}}^2\} = n\mathcal{E}\underline{x}^2 - n\mathcal{E}\underline{\bar{x}}^2 = n[\{\sigma^2 + \mu^2\} - n\{\frac{\sigma^2}{n} + \mu^2\}] = (n-1)\sigma^2$

b) $\text{var } \underline{s} = \mathcal{E}\underline{s}^2 - (\mathcal{E}\underline{s})^2 > 0 \rightarrow (\mathcal{E}\underline{s})^2 < \mathcal{E}\underline{s}^2 = \sigma^2 \rightarrow \mathcal{E}\underline{s} < \sigma.$

1.20. Beschouw $\mathcal{E}[t(\underline{x} - \mathcal{E}\underline{x}) + (\underline{y} - \mathcal{E}\underline{y})]^2 = t^2\text{var } \underline{x} + 2t \text{ cov}(\underline{x},\underline{y}) + \text{var } \underline{y}.$
Het linkerlid is niet negatief dus

$$\text{cov}^2(\underline{x},\underline{y}) \leq \text{var } \underline{x} \,.\, \text{var } \underline{y} = \sigma_{\underline{x}}^2 \,.\, \sigma_{\underline{y}}^2 \,.$$

Dus

$$\sigma_{(\underline{x}+\underline{y})}^2 = \sigma_{\underline{x}}^2 + \sigma_{\underline{y}}^2 + 2 \text{ cov}(\underline{x},\underline{y}) \leq \sigma_{\underline{x}}^2 + \sigma_{\underline{y}}^2 + 2\sigma_{\underline{x}}\sigma_{\underline{y}} = (\sigma_{\underline{x}} + \sigma_{\underline{y}})^2 \,.$$

1.21. a) $f = \frac{1}{2}(f_1 + f_2)$

$$\left.\begin{array}{l} \mathcal{E}\underline{x} = \frac{1}{2}(\mu_1 + \mu_2) \\ \mathcal{E}\underline{x}^2 = \frac{1}{2}(\mu_1^2 + \sigma_1^2 + \mu_2^2 + \sigma_2^2) \end{array}\right\} \rightarrow \text{var } \underline{x} = \frac{1}{2}(\sigma_1^2 + \sigma_2^2) + \frac{1}{4}(\mu_1 - \mu_2)^2.$$

b) neen.

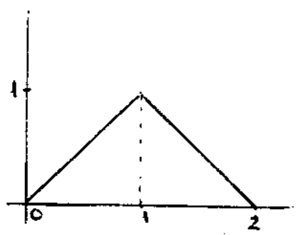
1.22. a) Als $\text{cov}(\underline{x},\underline{y}) = 0$

b) Als $\underline{y} = a\underline{x} + b$ $(a > 0)$, m.a.w. als $\rho = 1$.

1.23. a) $G(y) = P(\underline{y} < y) = F^n(y)$

b) $H(z) = P(\underline{z} < z) = 1 - \{1 - F(z)\}^n.$

1.24.



$z < 0$: $F(z) = f(z) = 0$
$0 \leq z \leq 1$: $F(z) = \frac{1}{2}z^2$, $f(z) = z$
$1 < z \leq 2$: $F(z) = -1 + 2z - \frac{1}{2}z^2$, $f(z) = 2 - z$
$z > 2$: $F(z) = 1$, $f(z) = 0$.

$\mu = 1$, $\sigma^2 = 1/6$, $P(\underline{x} + \underline{y} < 3/2) = 7/8$.

1.25. $\underline{y} := \underline{x}_{(n)}$, $z := \underline{x}_{(n-1)}$, $F(x) = x$.

a) $G(y) = P(\underline{y} < y) = F^n(y) = y^n$, $0 \leq y \leq 1$ dus $g(y) = ny^{n-1}$
$H(z) = P(\underline{z} < z) = nz^{n-1} - (n-1)z^n$, $0 \leq z \leq 1$ dus $h(z) = n(n-1)(z^{n-2} - z^{n-1})$

b) $\mathcal{E}\underline{y} = \frac{n}{n+1}$, $\mathcal{E}\underline{y}^2 = \frac{n}{n+2}$, $\text{var } \underline{y} = \frac{n}{(n+1)^2(n+2)}$

$\mathcal{E}\underline{z} = \frac{n-1}{n+1}$, $\mathcal{E}\underline{z}^2 = \frac{n(n-1)}{(n+1)(n+2)}$, $\text{var } \underline{z} = \frac{2(n-1)}{(n+1)^2(n+2)}$.

1.26. a) $\mathcal{E}\underline{x} = 3$, var $\underline{x} = 1$

b) $\mathcal{E}\underline{x} = \frac{80}{31}$, var $\underline{x} = \frac{1040}{961}$.

1.27. a) $k = 2$ b) $f_1(x) = 2(1 - x)$, $f_2(y) = 2(1 - y)$

c) $\mathcal{E}\underline{x} = \mathcal{E}\underline{y} = 1/3$, var $\underline{x}$ = var $\underline{y} = 1/18$, $\text{cov}(\underline{x},\underline{y}) = -1/36$

d) $f_1(x \mid y > 1/3) = 3 - \frac{9}{2}x$.

1.28. a) $\mathcal{E}(\underline{x} \mid \underline{x} \geq 3) = 3.47$ b) $\text{var}(\underline{x} \mid \underline{x} \geq 3) = 0.57$.

1.29. $\mathcal{E}\underline{y} = \mu_0\mu_1$, var $\underline{y} = \sigma_0^2\mu_1 + \mu_0^2\sigma_1^2$.

1.30. a) var $\underline{z} = 34$, $\text{cov}(\underline{x},\underline{z}) = 25$, $\text{cov}(\underline{y},\underline{z}) = -9$

c) var $\underline{xy} = \sigma_x^2\sigma_y^2 + \sigma_x^2\mu_y^2 + \sigma_y^2\mu_x^2$

$$\text{var } \frac{1}{n}\Sigma\underline{x}_i\underline{y}_i = \frac{1}{n} \text{ var } \underline{xy}$$

$$\text{var } \overline{\underline{x}}\,\overline{\underline{y}} = \frac{\sigma_x^2}{n}\cdot\frac{\sigma_y^2}{n} + \frac{\sigma_x^2}{n}\mu_y + \frac{\sigma_y^2}{n}\mu_x.$$

1.31. a) $\mathcal{E}\underline{y} = \frac{1}{2}n(N + 1)$, var $\underline{y} = \frac{n(N^2 - 1)}{12}$

b) $\mathcal{E}\underline{y} = \frac{1}{2}n(N + 1)$, var $\underline{y} = \frac{n(N - n)(N + 1)}{12}$.

2.1. a) $\bar{x} = 15.755$; $s = 0.043$.

c) $s = 0.395 * 0.108 = 0.043$

d) x_i afronden op 0.01; $\bar{x}$ en s afronden op 0.001.

2.2. a) $\bar{x} = 5.297$; $s = 0.042$

b) $s_R = 0.046$.

c) onderling onafhankelijke waarnemingen uit één populatie.

2.4. $\bar{x} = 75.3$; $s = 13.4$.

2.5. a) $\mathcal{E}\underline{x} = 17\frac{1}{2}$; var $\underline{x} = 175/12$.

2.6. a) $s_R = 0.31$

c) $\bar{x} = 12.37$; $s = 0.29$.

2.7. a) $\mathcal{E}\underline{x} = 14$; var $\underline{x} = 17\frac{1}{2}$

c) Uit ranges per kolom: $s_R = A_5 \times \bar{R} = 0.430 * 10.1 = 4.3$.

2.8. a) Uit ranges per kolom: $s_R = 0.037$.

2.9. b) $\bar{x} = 544.14$; $s = 23.23$.

3.1. a) 0.0766 b) 0.8418 c) 0.0766
d) 36.28 e) 10.08.

3.2. $P(\underline{x}_{(4)} \geq 5) = 0.4990$; $P(\underline{x}_{(1)} \geq 5) = 0.0006$.

3.3. $P = 0.0228$.

3.4. a) $P = 0.3662$
b) $P = 0.1587$
c) Ingrijpen als $\bar{x} > 250\frac{2}{3}$
d) $P = 0.3085$
e) $n \geq 54$.

3.5. $n \geq 35$.

3.6. a) $(\bar{x} - 3.29\sqrt{1/n},\ \bar{x} + 3.29\sqrt{1/n})$
b) Met Bienaymé- Chebyshev: $(\bar{x} - 6.32\sqrt{1/n},\ \bar{x} + 6.32\sqrt{1/n})$.

3.7. Met $\varphi(u) = \frac{1}{\sqrt{2\pi}} \exp(-\frac{1}{2}u^2)$.

3.8. a) Zij $\underline{e}_i$ de afrondingsfout. Gevraagd wordt $P\{|\Sigma\underline{e}_i| > 5\}$.
$\underline{e}_i$ doorloopt de waarden $-50,-49,\ldots,0,\ldots,50$ met kansen
$\frac{1}{200}, \frac{1}{100}, \ldots, \frac{1}{100}, \frac{1}{200}$. $\mathcal{E}\underline{e} = 0$, var $\underline{e} = 2[\frac{1}{200}(50)^2 + \frac{1}{100}\{(49)^2+\ldots+(1)^2\}] =$
$= 833.5$ (centen)^2. Dan $\mathcal{E}(\Sigma\underline{e}_i) = 0$, var$(\Sigma\underline{e}_i) = 120 * 0.08335 = 10.002$
(gulden)^2, $P = P\{|\underline{u}| > \frac{5}{\sqrt{10}} = 1.58\} = 0.1142$.
b) e_i doorloopt de waarden $0,1,\ldots,99$ met kansen $\frac{1}{100}$.
var $\underline{e} = 833.25$ (centen)^2, var$(\Sigma\underline{e}_i) = 9.999$ (gulden)^2
$P = 0.1142$.

3.9. a) $n \geq 92$
b) onderstelling: levensduren der lampen onderling onafhankelijk
stelling: centrale limietstelling.

3.10. a) $P(\underline{x} < 132) = 0.0137 < \frac{1}{2}\alpha$ $(\alpha = 0.05)$ dus vermoeden is gerechtvaardigd
b) Niet essentieel (centrale limietstelling).

3.11. $P(\underline{\bar{x}} < 5.23 \mid H_0) = 0.0016 < \frac{1}{2}\alpha$ $(\alpha = 0.05)$ dus H_0 verwerpen.

3.12. a) overschrijdingskans: $P(\underline{\bar{x}}_A > 251.68 \mid \mu_A = 250) = 0.0901 > \frac{1}{2}\alpha$

kritiek gebied: $\bar{x}_A = 251.68 \in Z = \{\bar{x} \mid \bar{x} < 247.55 \vee \bar{x} > 252.45\}$

betrouwbaarheidsinterval: $\mu = 250 \in (249.23, 254.13)$

Conclusie: $H_0 : \mu_A = 250$ niet verwerpen.

3.13. a) $P(\underline{\bar{x}} > 2.54) = P(\underline{u} > 6.18) << \frac{1}{2}\alpha$

b) De standaardafwijking mag worden afgerond op 0.001.

c) Gegeven moet nog zijn dat de tweede steekproef uit dezelfde baal afkomstig is.

3.14. a) H_0 wordt verworpen als $\bar{x} > 142.33$

b) $P = 0.6233$

c) $n \geq 253$.

3.15. a) $P = 0.1219$

b) Bedenk: 1) de aard van het oppervlak heeft grote invloed op de benodigde hoeveelheid verf.

2) het vakmanschap van de schilder speelt een grote rol.

3.16. $a_1 = \sqrt{\nu/\chi^2_\nu(\alpha = 0.025)}$; $a_2 = \sqrt{\nu/\chi^2_\nu(\alpha = 0.975)}$.

3.17. a) $P(\underline{\bar{x}} > 19170) = 0.0026 << \frac{1}{2}\alpha$

b) $s = 287$, $\nu = 9$

c) $\sigma = 555 \notin (198,525)$ dus ook de variantie is veranderd.

3.18. a) H_0: $\sigma \leq 7.3$

b) aselecte steekproef uit normale populatie

c) $\omega := P(\underline{x} \in Z \mid H_a) = P(\underline{s} > 10 \mid \sigma) = P(\underline{\chi}^2_9 > \frac{900}{\sigma^2})$

d) Om een aselecte steekproef te kunnen nemen moet de draad van de rol.

3.19. $s = 2.3$, $\nu = 14$; de standaardafwijking is een maat voor de onrondheid.

3.20. $F_8^{16} = 1.90 < F_8^{16}$ $(\frac{1}{2}\alpha = 0.025)$ dus H_0: $\sigma_1^2 = \sigma_2^2$ niet verwerpen.

3.21. a) Voor μ: (12.15,12.57)

Voor σ: (0.20,0.53)

b) (12.17,12.55).

3.22. a) (203.6347,203.6501)

b) $n \geq 117$.

3.23. a) $\bar{x} = 211.00$, $s = 0.30$, $\nu = 9$

b) i) $P(\underline{\bar{x}} > 211.00 \mid H_0) = P(\underline{t}_9 > 1.05) > \frac{1}{2}\alpha$ dus H_0 niet verwerpen

ii) $Z = \{\bar{x} \mid \bar{x} < 210.69 \vee \bar{x} > 211.11\}$

$\bar{x} \notin Z$ dus H_0 niet verwerpen

iii) $\mu = 210.90 \in (210.79,211.21)$ dus H_0 niet verwerpen.

c) Het betrouwbaarheidsinterval geeft bij de gegeven realisatie en bij de gegeven onbetrouwbaarheid alle parameterwaarden, die niet verworpen worden.

3.24. a) $s^2 = 584 * 10^{-6}$, $\nu = 8$

b) $P(\underline{\bar{x}}_2 - \underline{\bar{x}}_1 > 0.067 \mid \mu_1 = \mu_2) = P(\underline{t}_8 > 4.36) < 0.025$

c) $s^2 = 1734 * 10^{-6}$, $\nu = 9$.

3.25. a) $s_d^2 = 0.1248$, $s = \sqrt{\frac{1}{2}s_d^2} = 0.25$, $\nu = 10$ ($\mathcal{E}\underline{d}$ is niet bekend)

b) $s_d^2 = \frac{1}{n}\,\Sigma d_i^2 = 0.1685$, $s = 0.29$, $\nu = 11$ ($\mathcal{E}\underline{d} = 0$).

3.26. Voor één paar geldt: $s_i^2 = \dfrac{x_{1i}^2 + x_{2i}^2 - \frac{1}{2}(x_{1i} + x_{2i})^2}{2-1} = \frac{1}{2}(x_{1i} - x_{2i})^2 = \frac{1}{2}d_i^2$, $\nu_i = 1$.

Dus $s^2 = \dfrac{\Sigma\nu_i s_i^2}{\Sigma\nu_i} = \dfrac{\frac{1}{2}\Sigma d_i^2}{n} = \dfrac{1}{2n}\,\Sigma d_i^2$.

3.27. b) $\bar{x}_1 = 5.339$, $s_1^2 = 781 * 10^{-6}$, $\nu_1 = 6$

$\bar{x}_2 = 5.389$, $s_2^2 = 348 * 10^{-6}$, $\nu_2 = 6$

c) $\mu_1 \in (5.313,5.365)$

$\mu_2 \in (5.371,5.407)$

d) $F_6^6 = 2.24 < F_6^6$ ($\frac{1}{2}\alpha = 0.025$); $s^2 = 564 * 10^{-6}$, $\nu = 12$

e) $t_{12} = -3.90$ dus H_0: $\mu_1 = \mu_2$ verwerpen

f) $\mu_1 - \mu_2 \in (-0.078, -0.022)$

$\sigma_1/\sigma_2 \in (0.61, 3.56)$.

3.28. a) $t_7 = -2.49$ dus H_0 verwerpen

b) $\chi^2_7 = \frac{KS}{\sigma^2} = 11.25 < \chi^2_{7(\frac{1}{2}\alpha = 0.025)}$ dus H_0 niet verwerpen.

3.29. a) $t_{10} = -1.59$ (H_0 niet verwerpen)

b) onderling onafhankelijke aselecte steekproeven uit normaal verdeelde populaties met dezelfde variantie.

c) (7.5, 11.9)

Bij afwijkingen naar boven wil men eerder tot significantie besluiten dan bij afwijkingen naar beneden

d) H_0: $\sigma_1^2 = \sigma_2^2$, F_4^7 ($\frac{1}{2}\alpha = 0.025$) $= 9.07$

H_0: $\sigma_1^2 \leq \sigma_2^2$, F_4^7 ($\alpha = 0.05$) $= 6.09$

$F_4^7 = 3.54$; geen der nulhypothesen wordt verworpen.

3.30. a) $\bar{x}_A = 8.6$, $\bar{x}_B = 9.8$, $s_A^2 = 1.16$, $s_B^2 = 1.51$, $\nu_A = \nu_B = 9$, $s = 1.15$, $\nu = 18$

$t_{18} = -2.33$ dus H_0: $\mu_A = \mu_B$ verwerpen

b) - steekproeven aselect

- steekproeven onderling onafhankelijk

- populaties normaal verdeeld met dezelfde variantie

c) H_0: $\mathcal{E}\underline{d} = \mu_A - \mu_B = 0$, $\bar{d} = -1.2$, $s_d^2 = 3.07$, $\nu = 9$

$t_9 = -2.17$ dus H_0 niet verwerpen.

3.31. a) Een éénzijdige toets wordt toegepast als van <u>tevoren</u> het vermoeden bestaat dat analist B minder nauwkeurig werkt dan analist A

b) $\frac{s_A^2}{s_B^2} = \frac{4.36}{0.743} = 5.87$

éénzijdig: $Z = \{F_5^5 \mid F_5^5 > 5.05\}$ dus H_0 verwerpen

tweezijdig: $Z = \{F_5^5 \mid F_5^5 > 7.15\}$ dus H_0 niet verwerpen

c) $t_{10} = -.27$ dus $\mu_A = \mu_B$ niet verwerpen.

3.32. a) $\bar{x} = 789$, $s = 120$, $\nu = 7$

b) $\sigma \in (79,245)$, $\mu \in (689,889)$

c) aselecte steekproef uit normale populatie

3.33. a) $\mu_1 \in (10.56,10.94)$

$\mu_2 \in (10.05,10.69)$

b) $t_{68} = 2.19$ dus H_0: $\mu_1 = \mu_2$ verwerpen

c) $F_{39}^{29} = 1.94$ dus H_0: $\sigma_1^2 = \sigma_2^2$ verwerpen ($\alpha = 0.10$)

d) $P(t_{\nu_1+\nu_2}\, \underline{s}\sqrt{\frac{1}{n_1} + \frac{1}{n_2}} < t_{\nu_1}\, \underline{s}_1\sqrt{\frac{1}{n_1}} + t_{\nu_2}\, \underline{s}_2\sqrt{\frac{1}{n_2}}) > 0.$

3.34. a) $\dfrac{s_1^2}{s_2^2} = 2.47 < F_7^7(\tfrac{1}{2}\alpha)$

b) $t_{14} = 2.49$

c) $(0.70,3.51)$.

3.35. a) $\dfrac{s_2^2}{s_1^2} = 1.82 < F_5^7(\tfrac{1}{2}\alpha)$; $t_{12} = -1.75$

b) Een betrouwbaarheidsinterval bevat alle parameterwaarden $\theta_1 - \theta_2$ die bij een gegeven realisatie niet worden verworpen.

c) De conclusies blijven dezelfde. Als de waarnemingen gemiddelden zijn van 5 stuks dan is:

$$s_1^2 = \frac{\hat{\sigma}_1^2}{5}\,,\quad s_2^2 = \frac{\hat{\sigma}_2^2}{5}\,,\quad s^2 = \frac{\hat{\sigma}^2}{5}\,,\quad \nu_1 = 5,\ \nu_2 = 7,\ \nu = 12\ .$$

Dan is

$$F_5^7 = \frac{\hat{\sigma}_1^2}{\hat{\sigma}_2^2} = \frac{5s_1^2}{5s_2^2} = \frac{s_1^2}{s_2^2}$$

en

$$t_{12} = \frac{\bar{x}_1 - \bar{x}_2}{\hat{\sigma}\sqrt{\frac{1}{5n_1} + \frac{1}{5n_2}}} = \frac{\bar{x}_1 - \bar{x}_2}{\frac{\hat{\sigma}}{\sqrt{5}}\sqrt{\frac{1}{n_1} + \frac{1}{n_2}}} = \frac{\bar{x}_1 - \bar{x}_2}{s\sqrt{\frac{1}{n_1} + \frac{1}{n_2}}}\ .$$

3.36. $t_\nu\ (\tfrac{1}{2}\alpha = 0.005) = \dfrac{s}{\sqrt{\nu + 1}} \leq 0.005 \Rightarrow n \geq 20.$

3.37. a) $t_6 = -0.74$

b) $F_6^4 = 6.00$.

3.38. a) $\bar{x}_1 = 3$, $\bar{x}_2 = 6$, $s_1^2 = 9.6$, $\nu_1 = 5$, $s_2^2 = 5$, $\nu_2 = 4$

b) $F_4^5 = 1.92$

c) $(3.6, 25.3)$

d) $t_9 = -1.80$

e) onderling onafhankelijke aselecte steekproeven uit normale populaties (met dezelfde variantie bij d)).

3.39. a) H_0: $\sigma_1^2 = \sigma_2^2$; $F_4^6 = 1.05$

H_a: $\sigma_1^2 \neq \sigma_2^2$

H_0: $\mu_1 = \mu_2$; $t_{10} = 3.15$

H_a: $\mu_1 \neq \mu_2$

b) $\hat{\mu}_1 = 17.27$; $\hat{\mu}_2 = 13.22$; $\hat{\sigma}^2 = 4.816$; $\hat{\sigma} = 2.19$; $\nu = 10$.

3.40. a) $\lim\limits_{\nu\to\infty} \underline{t}_\nu = \underline{u}$

b) t_5^2 $(\frac{1}{2}\alpha = 0.025) = (2.57)^2 = 6.60$

F_5^1 $(\alpha = 0.05) = 6.61$

c) F_∞^6 $(\alpha = 0.05) = 2.10$; $\frac{1}{6}\chi_6^2$ $(\alpha = 0.05) = \frac{1}{6} * 12.6 = 2.10$

d) F_∞^1 $(\alpha = 0.05) = 3.84$; χ_1^2 $(\alpha = 0.05) = 3.84$; u^2 $(\frac{1}{2}\alpha = 0.025) = 3.84$.

4.1. f) $\mathcal{E}\underline{M}_3 = \mathcal{E}\underline{M}_5 = 4\frac{1}{2}$; var $\underline{M}_3 = 5.082$; var $\underline{M}_5 = 3.6564$.

5.1. a) $P(\underline{x} = x) = \binom{10}{x}\binom{20}{5-x}/\binom{30}{5}$, $x = 0,\ldots,5$

b) $\underline{x} = 5/3$; var $\underline{x} = 250/261$

c) $P(\underline{x} = 5) = 2/1131 = 0.0018$.

5.2. $P = 0.49$.

5.3. Aanwijzing:

$$\frac{\binom{M}{x}\binom{N-M}{n-x}}{\binom{N}{n}} = \binom{n}{x}\left(\frac{M}{N}\right)^x\left(1-\frac{M}{N}\right)^{n-x}\frac{\{(1-\frac{x-1}{M})*\ldots*1\}*\{(1-\frac{n-x-1}{N-M}*\ldots*1\}}{(1-\frac{n-1}{N})*\ldots*1}.$$

5.4. $u = (124 - 100)/\sqrt{50} = 3.39$.

5.5. $(0.755, 0.845)$.

5.6. i) $Z = \{\hat{p} \mid \hat{p} > 0.224\}$

ii) $P(\underline{\hat{p}} > \hat{p} \mid H_0) = 0.025$

iii) $(0.179, 1]$.

H_0 wordt verworpen.

5.7. a) $u = -\sqrt{6/5}$

b) $(-0.37, 0.11)$.

5.8. Aanwijzing:

$$\binom{n}{x}p^x q^{n-x} = \frac{1}{x!}(np)^x.1.(1-\frac{1}{n})\ldots(1-\frac{x-1}{n})\{(1-p)^{-\frac{1}{p}}\}^{-np}(1-p)^{-x}.$$

5.9. $n > 9604$.

5.10. Aanwijzing: op de grenzen van het interval geldt

$$\left(\frac{\hat{p}-p}{\sqrt{pq/n}}\right)^2 = u^2_{(\frac{1}{2}\alpha)}.$$

5.11. $u = (70 - 100)/\sqrt{500/6} = -3.29$.

5.12. Binomiale benadering met $p = \frac{50}{N}$, $n = 100$ geeft (253,715).

5.13. $n > 2401$ (σ^2 is maximaal voor $p = \frac{1}{2}$).

5.14. H_0: series zijn onderling onafhankelijk

x_i := aantal malen dat i^e waarneming het hoogste is

$\forall i$: $\underline{x}_i \sim BN\ (p = 1/3,\ n = 10)$

H_0 wordt verworpen, want

$$P(\underline{x}_i \geq 8 \vee \underline{x}_i \leq 2,\ i = 1,2,3) \leq 6P(\underline{x}_2 \geq 8) = 6 * 0.0033 = 0.0198 < \alpha\ .$$

5.15. a) $P(\underline{x} \geq 3 \mid H_0) = 0.0755 > \alpha$, waarbij onder H_0 $\underline{x} \sim BN\ (p = 0.05,\ n = 20)$

b) $P(\underline{x} \geq 6 \mid H_0) = 0.0166 < \alpha$, waarbij onder H_0 $\underline{x} \sim BN\ (p = 0.05,\ n = 40)$ dus in benadering $\underline{x} \sim PS\ (\mu = 2)$,

c) (0.03,1] resp. (0.065,1].

5.16. a) - aselecte steekproef
- binomiale verdeling

b) $P(\hat{\underline{p}} > 0.0667) = P(\underline{u} > 4.20) < \alpha$.

5.17. Normale benadering:

a) $P(\underline{x} > 41\frac{1}{2}) = 0.5319$

b) $P(41\frac{1}{2} < \underline{x} < 42\frac{1}{2}) = 0.0638$.

5.18. Aanwijzing:

$$P(\underline{x}_1 = x_1 \mid \underline{x}_1 + \underline{x}_2 = n) = \frac{P(\underline{x}_1 = x_1)P(\underline{x}_2 = n - x_1)}{\sum_{x=0}^{n} \{P(\underline{x}_1 = x)P(\underline{x}_2 = n - x)\}}\ .$$

5.19. a) $P(\underline{x} \leq 1 \mid \mu = 3) = 0.1991 > \alpha$

b) $P(\underline{x} \leq 2 \mid \mu = 6) = 0.0620 < \alpha$.

5.20. a) $u = -63/\sqrt{567} = -2.65$

b) $\mu_1 \in (221,283)$; $\mu_2 \in (280,350)$; $\mu_1 - \mu_2 \in (-110,-16)$.

5.21. Aanwijzing:

$$P(\underline{z} = z) = P(\underline{x} + \underline{y} = z) = \sum_{x=0}^{z} \{P(\underline{x} = x)P(\underline{y} = z - x)\} .$$

5.22. $P(\underline{x} = x) = \dfrac{e^{-3}3^x}{x!}$.

5.23. (-6.0,-2.0).

5.24. $p_0 = 0.017$; $p_1 = 0.070$.

5.25. $u = \dfrac{37 - 30}{\sqrt{\frac{34}{8} + \frac{34}{6}}} = 2.22.$

5.26. a) 81.4 dagen
b) 222.5 dagen.
c) 18.7%
d) ƒ 4125,--
e) ƒ 2368,--.

5.27. a) I. $P(\underline{u} > \dfrac{x_A - x_B}{\sqrt{x_A + x_B}}) = P(\underline{u} > 1.90) = 0.0287 < \tfrac{1}{2}\alpha.$

II. $P(\underline{x}_B \leq \underline{x}_A \mid \underline{x}_A + \underline{x}_B = 10) = 0.0547 > \tfrac{1}{2}\alpha.$

b) I. $Z_I = \{(x_A,x_B) \mid |\dfrac{x_A - x_B}{\sqrt{x_A + x_B}}| > u(\tfrac{1}{2}\alpha)\}$, $u(\tfrac{1}{2}\alpha) = 1.645.$

II. $Z_{II} = \{(x_A, n - x_A) | P(\underline{x} \leq x_A \mid \underline{x} \sim BN(p = \tfrac{1}{2}, n) | \leq \tfrac{1}{2}\alpha\}$, $\tfrac{1}{2}\alpha = 0.05.$

c) $Z_I \supset Z_{II}$ dus $\omega_I = P((\underline{x}_A,\underline{x}_B) \in Z_I \mid H_a \text{ juist}) > \omega_{II}$.

I heeft dus het grootste onderscheidingsvermogen.

$$\omega_I = P\{(\underline{x}_A,\underline{x}_B) \in Z_I \mid \underline{x}_A \sim PS(\mu_A),\ \underline{x}_B \sim PS(\mu_B)\} .$$

5.28. a) $P(\underline{x} \geq 3 \mid \underline{x} \sim BN(p = 1/5,\ n = 20)) = 0.7939 > \alpha$

b) $P(\underline{x} \geq 3 \mid \underline{x} \sim PS(\quad = 2/3)) = 0.0302 < \alpha$

c) $P(\underline{x} \geq 3 \mid \underline{x} \sim HG(N = 60,\ n = 3,\ n = 20)) = 0.033 < \alpha.$

6.1. a) $X^2 = 5\frac{7}{9}$, $\nu = 2$

b) $\hat{\lambda} = 10$, vâr $\underset{\sim}{\hat{\lambda}} = 1/18$.

6.2. $X^2 = 15$, $\nu = 3$.

6.3. a) $X^2 = 7.98$, $\nu = 5$

b) $u = 1.38$.

6.4. a) $u = 0.86$, éénzijdig toetsen

b) $u = 0.86$, tweezijdig toetsen

of $X^2 = 0.73$, $\nu = 1$.

Opmerking: $X^2 = u^2$.

6.5. $X^2 = 1.09$, $\nu = 2$.

6.6. $X^2 = 8.65$, $\nu = 9$.

6.7. $X^2 = 4$, $\nu = 2$.

6.8. $X^2 = 18.01$, $\nu = 4$.

6.9. $X^2 = 4.68$, $\nu = 2$.

6.10. $X^2 = 12.99$, $\nu = 1$.

6.11. a) kop 1: $X^2 = 14.36$, $\nu = 1$

kop 2: $X^2 = 34.09$, $\nu = 1$

kop 3: $X^2 = 29.32$, $\nu = 1$

kop 4: $X^2 = 5.67$, $\nu = 1$

kop 5: $X^2 = 6.06$, $\nu = 1$

kop 6: $X^2 = 12.05$, $\nu = 1$

c) Men toetst dan H_0: $p_{A_i} = p_{B_i}$, $i = 1,\dots,6$.

Men wil echter toetsen H_0: $p_A = p_B$. De resultaten van de 6 koppen moeten dan worden samengevoegd tot een 2 * 2-tabel ($X^2 = 88.7$, $\nu = 1$).

6.12. $E_i = \mathcal{E}\underline{f}_i = 10$.

6.13. $E_i = \mathcal{E}\underline{f}_i = 10$.

6.14. b) $\bar{x} = 0.9685$, $s^2 = 0.5279$

c,d) $X^2 = 2.54$, $\nu = 4$ (4 klassen worden samengevoegd).

6.15. $X^2 = 1.01$, $\nu = 5$ (2 klassen worden samengevoegd).

6.16. a) $X^2 = 2.41$, $\nu = 3$ (4 klassen worden samengevoegd)

b) $u = 1.10$.

6.17. a) Bereken $\bar{x}$ en s.

b_i := bovengrens klasse i, k := aantal klassen.

v_i := ondergrens klasse i.

Bereken:

$$u_{1i} = \frac{v_i - \bar{x}}{s}, \quad u_{2i} = \frac{b_i - \bar{x}}{s}, \qquad i = 1,\ldots,k .$$

Bereken:

$$p_1 = P(\underline{u} < u_{21})$$

$$p_i = P(u_{1i} < \underline{u} < u_{2i}), \qquad i = 2,\ldots,k-1$$

$$p_k = P(\underline{u} > u_{1k}) .$$

Bereken:

$$\mathcal{E}\underline{f}_i = p_i * \Sigma f_i .$$

b) $X^2 = 6.48$, $\nu = 9$ (de eerste 3 en de laatste 4 klassen worden samengevoegd).

6.18. a) $X^2 = 9.03$, $\nu = 4$ (de eerste 2 en de laatste 2 klassen worden samengevoegd).

b) Nu moeten μ en σ worden geschat uit de waarnemingen; de toetsingsgrootheid heeft dan 2 vrijheidsgraden minder dan onder a).

7.1. a) $\rho^2(\underline{x},y) = \dfrac{\text{cov}^2(\underline{x}, a\underline{x} + b)}{\text{var}(\underline{x})\text{var}(a\underline{x}+b)} = \dfrac{(a\ \text{var}(\underline{x}))^2}{\text{var}(\underline{x})a^2\text{var}(\underline{x})} = 1$.

b) $\rho^2(\underline{u},\underline{v}) = \dfrac{\text{cov}^2(a\underline{x}+b, c\underline{y}+d)}{\text{var}(a\underline{x}+b)\text{var}(c\underline{x}+d)} = \dfrac{\{ac\ \text{cov}(\underline{x},\underline{y})\}^2}{a^2\text{var}(\underline{x})c^2\text{var}(\underline{y})} = \rho^2(\underline{x},\underline{y})$.

c) $\underline{x},\underline{y}$ onafhankelijk $\Rightarrow \mathcal{E}\underline{xy} = \iint xyf(x,y)dxdy = \iint xyf_1(x)f_2(y)dxdy$

$$= \int xf_1(x)dx \quad \int yf_2(y)dy = \mathcal{E}\underline{x}\mathcal{E}\underline{y}$$

$\rho(\underline{x},\underline{y}) = 0 \not\Rightarrow \underline{x},\underline{y}$ O.O.

B.v. $y = |\underline{x}|$, $P(\underline{x} = -1) = P(\underline{x} = 0) = P(\underline{x} = 1) = 1/3$

$$\left.\begin{aligned} &\text{Dan is } \mathcal{E}\underline{xy} = \mathcal{E}\underline{x}|\underline{x}| = 1/3\{-1+0+1\} = 0 \\ &\mathcal{E}\underline{x} = 0 \\ &\mathcal{E}\underline{y} = \mathcal{E}|\underline{x}| = 1/3\ \{1+0+1\} = 2/3 \end{aligned}\right\} \Rightarrow \mathcal{E}\underline{xy} - \mathcal{E}\underline{x}\mathcal{E}\underline{y} = 0$$

7.2. $\rho(\underline{x},\underline{y}) = 1/3\ \sqrt{3}$

7.3. a) Aanwijzing: minimaliseer $F(\alpha_0,\alpha_1) := \sum_{i=1}^{n}(y_i - \alpha_0 - \alpha_1 x_i)^2$

oplossing: $a_1 = \dfrac{n\Sigma x_i y_i - \Sigma x_i \Sigma y_i}{n\Sigma x_i^2 - (\Sigma x_i)^2}$, $a_0 = \bar{y} - a_1\bar{x}$

$$\left.\begin{aligned} &\text{b) } \mathcal{E}\underline{a}_1 = \mathcal{E}\frac{n\Sigma x_i \underline{y}_i - \Sigma x_i \Sigma \underline{y}_i}{nKS_x} = \frac{1}{nKS_x}\{n\Sigma(x_i\mathcal{E}\underline{y}_i) - \Sigma x_i \Sigma(\mathcal{E}\underline{y}_i)\} \\ &\mathcal{E}\underline{y}_i = \alpha_0 + \alpha_1 x_i \end{aligned}\right\} \Rightarrow \mathcal{E}\underline{a}_1 = \alpha_1$$

$$\mathcal{E}\,\underline{a}_0 = \mathcal{E}(\underline{y} - \underline{a}_1\bar{x}) = \mathcal{E}\bar{\underline{y}} - \bar{x}\mathcal{E}\underline{a}_1 = \alpha_0 + \alpha_1\bar{x} - \alpha_1\bar{x} = \alpha_0$$

7.5. b) $a = 81$, $b = 1.14$

c) $\mathcal{E}\hat{\underline{y}} = \mathcal{E}(\underline{a} + \underline{b}x) = \alpha + \beta x = \mathcal{E}\underline{y}$

d) $\alpha \in (60,102)$, $\beta \in (0.74,\ 1.54)$

e) $\hat{y} = 123{,}2$, $\hat{\text{var}}\ \hat{y} = \hat{\text{var}}\ \underline{a} + x^2\hat{\text{var}}\ \underline{b} + 2x\ \hat{\text{cov}}(\underline{a},\underline{b}) = 11.567$ $(\nu = 10)$

$\hat{y} \in (1.17,\infty)$

f) $x = \bar{x}$

7.6. a) $b_1 = 29/30$, $\hat{\text{var}}\ \underline{b}_1 = 0.123$ $(\nu = 8)$ $t = -0.10$

b) $H_0 : \beta_0 - \beta_1\bar{x} = 0$, $b_0 - b_1\bar{x} = 0.3$, $\hat{\text{var}}(\underline{b}_0 - \underline{b}_1\bar{x}) = 1.48$ $(\nu = 8)$

$t = 0.25$

7.7. a) $a_1 = 0.8$, $a_0 = 1.6$, $\hat{\sigma}_0^2 = 1.2$ $(\nu = 3)$

b) $\hat{\text{var}}\ \underline{a}_1 = 0.12$ $(\nu = 3)$ $\alpha_1 \in (-0.3, 1.9)$

c) $P\{.216 < \frac{KS_r}{\sigma_0^2} < 9.35\} = 0.95$ (tabel S.C. 3.1)

$KS_r = 3.6$

$\Big\} \Rightarrow \sigma_0^2 \in (0.62, 4.08)$

7.8. a) $a_1 = 1$, $a_0 = 4$, $\hat{\sigma}_0^2 = 0.8$ $(\nu = 5)$

b) $\hat{\text{var}}\ \underline{a}_0 = 8/21$, $\hat{\text{var}}\ \underline{a}_1 = 1/15$, $\hat{\text{cov}}(\underline{a}_0,\underline{a}_1) = -2/15$

c) $\alpha_1 \in (0.34, 1.66)$

7.9. b) $\hat{y} = -3.9 + 0.71\ x$

c) $\hat{\sigma}_0^2 = 0.65$ $(\nu = 6)$, $\hat{\text{var}}\ \underline{b}_1 = 0.0136$, $t = 5.90$, $H_0 : \beta_1 = 0$ verwerpen

d) (i) waarnemingen o.o.

(ii) $\underline{e}_i \sim N(0,\sigma_0^2)$

(iii) geen meetfout in snelheidsbepaling

e) Het aantal foute exemplaren is Poisson verdeeld dus var $\underline{e}_i = \alpha + \beta x_i$

7.10. a) $b = 0.357$, $a = 5.72$

b) $\hat{\sigma}_0^2 = 25.59$ $(\nu = 8)$, $\hat{\text{var}}\ \hat{\underline{y}} = 2.59$ $\mathcal{E}\hat{\underline{y}} \in (9.14, 16.58)$

c) $\hat{\text{var}}\ \hat{\underline{y}} = 6.59$, $\mathcal{E}\hat{\underline{y}} \in (10.50, 22.36)$

d) $\hat{\text{var}}\ \hat{\underline{y}} = 34.58$, $\mathcal{E}\hat{\underline{y}} \in (9.99, 37.15)$

7.11. a) Minimaliseer $F(\alpha) := \Sigma(y_i - \alpha x_i)^2 \Rightarrow a = \frac{\Sigma x_i y_i}{\Sigma x_i^2} = 8$

b) $\hat{\sigma}_0^{\,2} = 11/4$ ($\nu = 4$), var $\underline{a} = \dfrac{\sigma_0^{\,2}}{\Sigma x_i^{\,2}}$, $\hat{\text{var}}$ $\underline{a} = 1/20$

7.12. a) $r = -0.81$

b) $a = 1053.58$, $b = -3.157$

c) $c = 223.647$, $d = -0.208$

e) $(\bar{x},\bar{y}) = (12.217,\ 1015.008)$

8.1.

Bron	KS	ν	GK	F
Tussen soorten	63	3	21	2.15
Binnen soorten	78	8	$9\frac{3}{4}$	

$H_0 : \mu_i = \mu$ niet verwerpen

8.2.

Bron	KS	ν	GK	F
Tussen voedingen	19942.93	2	9971.5	173.7
Binnen voedingen	688.80	12	57.4	

$H_0 : \mu_i = \mu$ verwerpen

8.3. a)

Bron	KS	ν	GK
Tussen merken	$294\frac{4}{15}$	4	73.57
Binnen merken	$62\frac{2}{3}$	10	6.27

b) F = 11.7 dus H0 : $\mu_i = \mu$ verwerpen.

c) De waarnemingen zijn onderling onafhankelijke trekkingen uit normaal verdeelde populaties met dezelfde variantie.

8.4. a)

Bron	KS	ν	GK	F
Tussen dagen	646.19	3	215.4	13.8
Binnen dagen	187.25	12	15.6	

$H_0 : \mu_i = \mu$ verwerpen

b) Omdat de temperatuur met de dagen varieert is het niet duidelijk of het significante verschil veroorzaakt wordt door het temperatuureffekt of door het feit dat de opbrengst op verschillende dagen is bepaald.

8.5. a)

Bron	KS	ν	GK	F
Tussen meters	$66.9 * 10^{-4}$	2	$33.4 * 10^{-4}$	3.34
Binnen meters	$60 * 10^{-4}$	6	$10 * 10^{-4}$	

$H_0 : \mu_i = \mu$ niet verwerpen.

b) $\sigma_0^2 \in (4.10 * 10^{-4}, 48.4 * 10^{-4})$.

8.6. Kodeer : $y_{ij} = (x_{ij} - 3.00) * 100$

a)

	$\bar{y}$	KS_y	ν	$s_y^2 = s_x^2.10^4$
A	17 2/3	141 1/3	5	28.27
B	18	122	2	61.00
C	4 1/3	12 2/3	2	6.33
D	10 1/3	8 2/3	2	4.33
E	14 5/7	169 3/7	6	28.24

$$s^2 = \frac{\Sigma \nu_i s_i^2}{\Sigma \nu_i} = 26.7 * 10^{-4}, \quad \nu = 17$$

b) $\sigma_A . 10^2 \in (3.30,\ 13.04)$ $\quad \sigma_D . 10^2 \in (1.08,\ 13.07)$

$\sigma_B . 10^2 \in (4.06,\ 49.05)$ $\quad \sigma_E . 10^2 \in (3.40,\ 11.69)$

$\sigma_C . 10^2 \in (1.31,\ 15.80)$ $\quad \sigma . 10^2 \in (3.88,\ 7.75)$

c) $X^2 = 3.69,\ \nu = 4.$

d)

Bron	$KS*10^4$	ν	$GK*10^4$	F
Tussen lab's	452.9	4	113.2	4.24
Binnen lab's	454.1	17	26.7	

$H_0 : \mu_i = \mu$ verwerpen.

8.7. a) Kodeer $y_{ij} = (x_{ij} - 1.47) * 100$

b)

Bron	$KS * 10^4$	ν	$GK * 10^4$
Tussen typen	990 8/9	2	495 4/9
Binnen typen	426 2/3	6	71 1/9

c) $F = 6.97$; $H_0 : \mu_i = \mu$ verwerpen

d) (i) onafhankelijke, aselecte steekproeven.

(ii) normale verdelingen.

(iii) gelijke varianties.

9.1. $\mathcal{E}\underline{z} = \mu_x\mu_y$, $\mathcal{E}\underline{z}^2 = \mu_x^2\mu_y^2 + \mu_x^2\sigma_y^2 + \mu_y^2\sigma_x^2 + \sigma_x^2\sigma_y^2$

$$\text{var } \underline{z} = \mu_x^2\sigma_y^2 + \mu_y^2\sigma_x^2 + \sigma_x^2\sigma_y^2$$

$$V^2_{(\underline{z})} = \frac{\text{var } \underline{z}}{(\mathcal{E}\underline{z})^2} = V^2_{(\underline{y})} + V^2_{(\underline{x})} + V_{(\underline{x})}V_{(\underline{y})} \approx V^2_{(\underline{x})} + V^2_{(\underline{y})}$$

9.2. a) $\mathcal{E}\underline{z} = \mu_x\mu_y = z_g$

b) $\text{var } \underline{z} = \mu_x^2\sigma_y^2 + \mu_y^2\sigma_x^2 + \sigma_x^2\sigma_y^2$

c) $$\text{var } (\bar{\underline{x}}\bar{\underline{y}}) = \mu_x^2 . \frac{\sigma_y^2}{n} + \mu_y^2 . \frac{\sigma_x^2}{n} + \frac{\sigma_x^2}{n} . \frac{\sigma_y^2}{n} = \frac{\mu_x^2\sigma_y^2 + \mu_y^2\sigma_x^2}{n} + \frac{\sigma_x^2\sigma_y^2}{n^2}$$

$$\text{var } (\tfrac{1}{n}\Sigma\underline{x}_i\underline{y}_i) = \frac{1}{n} \text{ var } \underline{x}_i\underline{y}_i = \frac{\mu_x^2\sigma_y^2 + \mu_y^2\sigma_x^2}{n} + \frac{\sigma_x^2\sigma_y^2}{n}$$

dus $\text{var } (\bar{\underline{x}}\bar{\underline{y}}) < \text{var } (\frac{1}{n}\Sigma\underline{x}_i\underline{y}_i)$

9.3. a) $\mathcal{E}(\sqrt{\underline{x}}) \approx \sqrt{\mu} + \frac{1}{2}\mu(-\frac{1}{4\mu\sqrt{\mu}}) = \sqrt{\mu} - \frac{1}{8\sqrt{\mu}} \approx \sqrt{\mu}$ als μ voldoende groot is.

$$\text{var } (\sqrt{\underline{x}}) \approx (\frac{1}{2\sqrt{\mu}})^2 . \mu = \frac{1}{4}$$

b) Normale benadering geeft: $\sqrt{x} - 1.96\sqrt{\frac{1}{4}} < \sqrt{\mu} < \sqrt{x} + 1.96\sqrt{\frac{1}{4}}$

$$\Rightarrow \sqrt{x} - 0.98 < \sqrt{\mu} < \sqrt{x} + 0.98$$

$$\Leftrightarrow x - 1.96\sqrt{x} + 0.96 < \mu < x + 1.96\sqrt{x} + 0.96$$

c) M.b.v. tabel S.C. 8.1 wordt de linkergrens gevonden uit $P(\underline{x} < x - 1) = 0.975$, de rechtergrens uit $P(\underline{x} < x) = 0.025$.

x	S.C. 8.1	uit b)
5	(1.623, 11.668)	(1.58, 10.34)
8	(3.454, 15.763)	(3.42, 14.50)
10	(4.795, 18.390)	(4.76, 17.16)
12	(6.201, 20.962)	(6.19, 19.77)
16	(9.144, 25.984)	(9.14, 24.82)
25	(16.178, 36.905)	(16.18, 35.78)
30	(20.241, 42.827)	(20.24, 41.72)

9.4. var $\underline{z} \approx (f_x')^2 \sigma_x^2 + (f_y')^2 \sigma_y^2 = 4(\mu_x + \mu_y)^2(\sigma_x^2 + \sigma_y^2)$

b) $\mathcal{E}\underline{z} = (\mu_x + \mu_y)^2 + \sigma_x^2 + \sigma_y^2$

$\mathcal{E}\underline{z}^2 = \mathcal{E}\{(\underline{x} - \mu_x) + (\underline{y} - \mu_y) - (\mu_x + \mu_y)\}^4 =$

$= (\mu_x + \mu_y)^4 + 2(\mu_x + \mu_y)^2(\sigma_x^2 + \sigma_y^2) + (\sigma_x^2 + \sigma_y^2)^2.$

wegens $\mathcal{E}(\underline{x} - \mu_x)^3 = \mathcal{E}(\underline{y} - \mu_y)^3 = 0$, $\mathcal{E}(\underline{x} - \mu_x)^4 = 3\sigma_x^4$ en $\mathcal{E}(\underline{y} - \mu_y)^4 = 3\sigma_y^4$.

Dan is var $\underline{z} = 4(\mu_x + \mu_y)^2(\sigma_x^2 + \sigma_y^2) + 2(\sigma_x^2 + \sigma_y^2)^2$

c) $\left.\begin{array}{l} \mu_x = \mu_y = \mu \\ V(\underline{x}) = V(\underline{y}) = \frac{1}{4} \end{array}\right\} \Rightarrow \sigma_x^2 = \sigma_y^2 = \frac{1}{16}\mu^2$

ad a) var $\underline{z} \approx 2\mu^4$ ad b) var $\underline{z} = 2\mu^4 + \frac{1}{32} \cdot \mu^4$

9.5. a) $\mathcal{E}\underline{z} = (\mu_x + \mu_y)^2 + \sigma_x^2 + \sigma_y^2 = z_g + \sigma_x^2 + \sigma_y^2$ dus $\partial_t(z) = \sigma_x^2 + \sigma_y^2$

b) $\mathcal{E}\hat{\underline{z}}_g = z_g + \sigma_x^2 + \sigma_y^2$ dus $\partial_t(\hat{\underline{z}}_g) = \sigma_x^2 + \sigma_y^2$

c) $\mathcal{E}\hat{\underline{z}}_g = z_g + \frac{1}{n}(\sigma_x^2 + \sigma_y^2)$ dus $\partial_t(\hat{\underline{z}}_g) = \frac{1}{n}(\sigma_x^2 + \sigma_y^2)$

d) De schatting onder c) heeft de kleinste systematische fout.

ad b) var $\hat{\underline{z}}_g \approx \frac{1}{n} \cdot 4(\mu_x + \mu_y)^2(\sigma_x^2 + \sigma_y^2)$

ad c) var $\hat{\underline{z}}_g \approx 4(\mu_x + \mu_y)^2(\frac{\sigma_x^2}{n} + \frac{\sigma_y^2}{n})$

De schatters zijn bij benadering even nauwkeurig.
Bij exacte beschouwing blijkt de schatter onder c) nauwkeuriger te zijn (vgl. opg. 9.4)

9.6. Benadering : var $\underline{y} \approx 4\mu_x^2\sigma_x^2$

Exact : var $\underline{y} = 4\mu_x^2\sigma_x^2 + 2\sigma_x^4$

$\underline{x} \sim N(0,1) \Rightarrow \underline{x}^2 \simeq \underline{\chi}_1^2$ dus $\mathcal{E}\underline{x}^2 = 1$, var $\underline{x}^2 = 2$ (S.C. pag. 11).

10.1. a) Aanwijzing : $P_g(p) = P(\underline{x}_1 \leq 1|p) + P(\underline{x}_1 = 2|p)P(\underline{x}_2 \leq 1|p) + P(\underline{x}_1 = 3|p)P(\underline{x}_2 = 0|$

b) Aanwijzing : $P_g(p) = P(\underline{x} \leq 2|p)$

c) Aanwijzing : $\mathcal{E}(\underline{n}|p) = n_1 + n_2 . P(\underline{x}_1 = 2 \vee \underline{x}_1 = 3|p)$

10.2. a) $1 - P_g = 0.5162$

b) $\mathcal{E}\underline{n} = 13.874$

10.3. a) $\bar{\bar{x}} = 2.7232$ $\bar{s} = 0.138$

grenzen voor $\bar{x}$: 95%: 2.636, 2.810
99%: 2.592, 2.854

grenzen voor s: 95%: 0.0763, 0.1997
99%: 0.0454, 0.2306

b) 11.56% voldoet niet.

c) grenzen voor $\bar{x}$: 95%: 2.613, 2.787
99%: 2.569, 2.831

11.18% voldoet niet.

10.4. a) $\bar{R} = 0.319$, grenzen voor M: 95%: 9.81, 10.31
99%: 9.67, 10.45 .

b) Het proces lijkt beheerst.

11.1. H_0: A ≤ B wordt verworpen (éénzijdig toetsen).

11.2. H_0: geen voorkeur, wordt niet verworpen.

11.3. H_0: A minstens zo goed als B wordt verworpen (éénzijdig toetsen).

11.4. Min $(W_A, W_B) = 4$; er is verschil tussen de soorten verf.

11.5. Min $(W_I, W_C) = 48$; inspuiten van insuline heeft invloed.

11.6. Min $(W_W, W_S) = 119$; hypothese niet verwerpen.

11.7. $S = 2\frac{1}{2}$; er is een sterk verband tussen de vaardigheden ($r_s = 0.985$).

11.8. $S = 356$; hypothese wordt niet verworpen.

11.9. $S = 4$; er is een stijgend verloop van de bloeddruk ($r_s = 0.994$).

11.10. $S = 594\frac{1}{2}$; de verschillen tussen de kinderen berusten niet alleen op toeval.

11.11. $S = 224$; er is verschil tussen laboratoria.

11.12. $S = 14\frac{1}{2}$; er is geen signifikant verschil tussen werkmethoden.

12.1. a) i) Neen, grotere huishoudens hebben grotere kans om in de steekproef te worden opgenomen.

ii) Neen, huishoudens zonder mannen, die gemiddeld kleiner zijn, komen dan niet in de steekproef.

b) $2\sigma/\sqrt{n} < 0.1 \leftrightarrow n > 400\sigma^2$.

Voor $\sigma \approx \frac{1}{6}R$, $R = 15$ geeft dat: $n > 2500$.

12.2. a) $\underline{T}$ heeft een Γ-verdeling met $\mathcal{E}\underline{T} = N/\lambda$, var $\underline{T} = N/\lambda^2$.

Dan is

$$\text{var}\ \underline{y} \approx \{f'(\mathcal{E}\underline{T})\}^2 \text{var}\ \underline{T} = 0.189.10^6 \cdot \frac{\text{var}\ \underline{T}}{(\mathcal{E}\underline{T})^2} = 0.189.10^6 \cdot \frac{1}{N}.$$

Voor $N = 10^4$ wordt var $\underline{y} = 18.9$.

b) $b_{01} = 946.8$ $\qquad b_{02} = 930.4$

$b_{11} = -6.51$ $\qquad b_{12} = -5.57$

$s_1^2 = 12.17$ $(\nu_1 = 9)$ $\qquad s_2^2 = 20.72$ $(\nu_2 = 9)$.

c) $X^2 = \dfrac{\nu_1 s_1^2}{\sigma_{01}^2} = 5.79$, $\nu_1 = 9$, niet significant.

$X^2 = \dfrac{\nu_2 s_2^2}{\sigma_{02}^2} = 9.87$, $\nu_2 = 9$, niet significant.

$\dfrac{\sigma_{01}^2}{\sigma_{02}^2} \in (0.15, 2.36)$.

d) i) var $\underline{b}_{0i} = \dfrac{\sigma_0^2}{n} = 2.1$, $|u| = 8.0$ (zeer significant)

ii) var $\underline{b}_{1i} = \dfrac{\sigma_0^2}{KS_x} = 0.147$, $|u| = 1.73$ (H_0 niet verwerpen).

e) i) waarnemingen o.o.

ii) $\underline{e}_{ij} \sim N(0, \sigma_0^2)$.

iii) fout in x_i verwaarloosbaar.

Aan de eis van normaliteit is niet voldaan. De centrale limietstelling echter rechtvaardigt hier de toepassing van de gebruikte toetsingsmethode.

12.3. a) $\hat{p} = 0.03,\ X^2 = 0.40,\ \nu = 4$

b) i) $X^2 = 2.38,\ \nu = 1$

ii) $\hat{p} = 0.7,\ u = 1.54$

c) $X_1^2 = u^2;\ \sum_{i=1}^{2} \{\frac{(x_i - n_i\hat{p})^2}{n_i\hat{p}} + \frac{((n_i - x_i) - n_i\hat{q})^2}{n_i\hat{q}}\} = \frac{(\hat{p}_1 - \hat{p}_2)^2}{\hat{p}\hat{q}\Sigma \frac{1}{n_i}}$.

d) $P(\chi_6^2 < 0.57) < 0.005$.

De realisatie $X^2 = 0.57$ is dus zeer onwaarschijnlijk: de aanpassing is te mooi!

12.4. a) $\hat{p}_j = 0.5375,\ u = 3$

b) $p_j \in (0.5131, 0.5619)$

c) aselecte steekproef uit binomiale verdeling

d) $X^2 = 11.96,\ \nu = 5$

e) $X^2 = 1.93,\ \nu = 4$.

12.5. a) $\hat{\mu} = 2.3,\ X^2 = 5.82,\ \nu = 4$

b) i) $\underline{x}_i \sim PS(\mu) \Rightarrow \Sigma\underline{x}_i \sim PS(80\mu)$

Bij benadering is dan $\bar{\underline{x}} \sim N(\hat{\mu} = 2.3,\ \hat{\sigma}^2 = \frac{2.3}{80})$. Dan: $\mu \in (1.97, 2.63)$.

ii) $\underline{x}_i \sim BN(p,\ n = 80)$.

Met normale benadering volgt $\mu \in (1.97, 2.63)$.

12.6. a) 15.87%

b) $$\mathcal{E}(\underline{x} \mid \underline{x} \in A) = \frac{\int_{-\infty}^{250} xf(x)dx}{\int_{-\infty}^{250} f(x)dx} = 248.95.$$

$\mathcal{E}(\underline{x} \mid \underline{x} \in B) = 252.58$.

12.7. a)
$$\left.\begin{aligned} P(\underline{n} > n) &= P(\sum_{i=1}^{n} \underline{x}_i < 40) = P(\underline{u} < \frac{40 - 2n}{0{,}3\sqrt{n}} =: u_n) \\ P(\underline{n} > n-1) &= P(\sum_{i=1}^{n-1} \underline{x}_i < 40) = P(\underline{u} < u_{n-1}) \end{aligned}\right\} \Rightarrow$$

$$\Rightarrow P(\underline{n} = n) = P(\underline{n} > n-1) - P(\underline{n} > n) = P(\underline{u} < u_{n-1}) - P(\underline{u} < u_n)\ .$$

n	$u_n = \frac{40 - 2n}{0,3\sqrt{n}}$	$P(\underline{u} < u_n)$	$P(\underline{n} = n)$
17	4.85	1.0000	
18	3.14	0.9992	0.0008
19	1.53	0.9370	0.0622
20	0	0.5000	0.4370
21	-1.45	0.0735	0.4265
22	-2.84	0.0023	0.0712
23	-4.17	0.0000	0.0023
			1.0000

b) $\left.\begin{array}{l} \mathcal{E}\underline{n} = \sum\limits_{n=18}^{23} nP(\underline{n} = n) = 20.512 \\ \mu = 2 \end{array}\right\} \Rightarrow \mu_g = 41.024$.

12.8. a) Model: $\underline{x}_{ij} = \mu + \sigma_0 \underline{u}_{ij}$, $\hat{\sigma}_0^2 = 0.867$, $\nu = 9$

b) Model: $\underline{x}_{ij} = \mu_i + \sigma_0 \underline{u}_{ij}$, $\hat{\sigma}_0^2 = 0.862$, $\nu = 8$

c) Model: $\underline{x}_{ij} = \mu_j + \sigma_0 \underline{u}_{ij}$, $\hat{\sigma}_0^2 = 0.374$, $\nu = 4$.

12.9. Zie syllabus 10.4.

12.10. a) Afronden op 5µ is te grof.

b) $\bar{x} = 0.856$, $s = 0.0081$, $\nu = 15$.

c) $t = 0.6$, $\nu = 15$.

d) De waarnemingen vormen een aselecte steekproef uit een normale verdeling (met onbekende variantie).

e) $\sigma^2 \in (0.36*10^{-4}, 156*10^{-4})$.

f) H_0: $\sigma_1^2 = \sigma_2^2$ wordt niet verworpen; $F_7^{15} = 1.14$; $s^2 = 0.62 * 10^{-4}$, $\nu = 22$, $t = -0.58$, $\nu = 22$.

12.11. f)

Bron	KS	ν	GK	F
Tussen kolommen (posities)	14632.3	5	2926.5	2.84
Binnen kolommen (posities)	24763.6	24	1031.8	

12.12. a) $t = -2.97$, $\nu = 9$

b) De waarnemingen vormen onderling onafhankelijke aselecte steekproeven uit normaal verdeelde populaties met dezelfde variantie

c) Er is reden tot twijfel aan de vereiste onafhankelijkheid, omdat vanaf de vijfde periode de reaktiesnelheid toeneemt met de tijd.

12.13. a) $f(M) = 6M - 6M^2$, $0 \le M \le 1$

$f(M) = 0$, elders

b) $\mathcal{E}\underline{M} = \frac{1}{2}$, $\mathcal{E}\underline{M}^2 = 3/10$, var $\underline{M} = 1/20$.

12.14. a) $F(n) = \dfrac{nP(n)}{\Sigma nP(n)}$

b) $\Sigma P(n) = \Sigma\{\dfrac{F(n)}{n} . \Sigma nP(n)\} = 1$ dus $\mathcal{E}\underline{n} = \Sigma nP(n) = \dfrac{1}{\Sigma \dfrac{F(n)}{n}}$

c) $P(n) = \dfrac{F(n)/n}{\Sigma F(n)/n}$

d) Aanwijzing: $\sum\limits_{n=1}^{4} n = 10$ dus stel $(0,1,2,3) \to 4$

$(4,5,6) \to 3$

$(7,8) \to 2$

$(9) \to 1$

$\Sigma nP(n) = \frac{1}{4}.10 = 2\frac{1}{2}$.

12.15. a) $\mu = 8.778$, $\sigma^2 = 198.487$

b) $n_i = \dfrac{nN_i\sigma_i}{\Sigma N_i\sigma_i}$, var $\underline{\hat{T}} = \frac{1}{n}(\Sigma N_i\sigma_i)^2$

c) $\dfrac{n'}{n} = \dfrac{N^2\sigma^2}{(\Sigma N_i\sigma_i)^2} = 11.7$

d) $n = 25 \Rightarrow n_1 = 11$, $n_2 = n_3 = 7$, var $\underline{\hat{T}} = 39054.85$, $T = 2106.635$.

12.16. a) $P(\underline{x} = x) = pq^x$, $x = 0,1,2,\ldots;(q = 1 - p)$

b) $\mathcal{E}\underline{x} = \dfrac{q}{p}$, var $\underline{x} = \dfrac{q}{p^2}$

c) $P(\underline{y} = y \mid x > x_0) = pq^y$, $y = 0,1,2,\ldots$

d) $n \ge 76$.

12.17. a) H_0: $\mu_i = \mu$, $i = 1,\ldots,12$ met μ_i := gemiddeld gewicht van de karbonaatlaag van partij i.

De oorzaak van de verschillen kan behalve bij de partijen ook liggen bij de verschillen in spuittijden en/of kathoden.

b) Onderling onafhankelijke aselecte steekproeven uit normale verdelingen met dezelfde varianties.

12.18. a) $\mathcal{E}|\underline{x}| = \sigma\sqrt{\frac{2}{\pi}}$, var $|\underline{x}| = \frac{\pi - 2}{\pi}\sigma^2$

b) $P(\underline{x} = x) = \int \binom{n}{x} p^x (1-p)^{n-x} f(p)dp$

$$\mathcal{E}\underline{x} = \sum_{x=0}^{n} xP(\underline{x} = x) = \int \{\sum_{x=0}^{n} x\binom{n}{x}p^x(1-p)^{n-x}\} f(p)dp = n\mathcal{E}\underline{p}$$

$$\mathcal{E}\underline{x}^2 = \sum_{x=0}^{n} x^2P(\underline{x} = x) = \int \{\sum_{x=0}^{n} x^2\binom{n}{x}p^x(1-p)^{n-x}\} f(p)dp = n\mathcal{E}\underline{p} - n\mathcal{E}\underline{p}^2 + n^2\mathcal{E}\underline{p}^2.$$

12.19. $p_{10} < 0.071$.

12.20. 1) Verschillende afwijkingen kunnen tegelijkertijd bij één rekruut optreden.

2) De gegevens kunnen door simulatie zijn vervalst.

3) De afwijkingen zijn moeilijk definieerbaar en er kunnen derhalve beoordelingsverschillen tussen artsen voorkomen.

12.21. a) 0.0968,

b) 0.0329,

c) 0.5374,

d) 0.4066,

e) 0.0116.

12.22. $\mathcal{E}\underline{x} = \mu$, $\mathcal{E}\underline{x}^2 = \mu + \mu^2 + \sigma^2$, var $\underline{x} = \mu + \sigma^2$.

12.23. In de overwegend katholieke provincies Noord-Brabant en Limburg is de bevolking relatief sterker toegenomen dan in de overige provincies.

12.24. a) H_0: $\mu \geq \mu_0 = 16$, H_a: $\mu < 16$

b) $u = -2.64$ (< -1.645)

c) aselecte steekproef uit binomiale verdeling

d) om allerlei neveninvloeden te elimineren, zal men een controlegroep willen die, het oude geneesmiddel krijgt.

12.25. b) $\bar{x} \approx 758$, $s \approx 120$.