# Hertentamen 2010-2011

**Vraag 1**

Een kleine boekwinkel kent de volgende verkoopstatistiek

| | | | | | |
|---|---|---|---|---|---|
| Aantal boeken verkocht: | 0 | 1 | 2 | 3 | 4 of meer |
| Kans: | 0.09 | 0.14 | 0.44 | 0.27 | 0.06 |

Wat is de kans dat er op een dag minder dan 3 boeken worden verkocht?

a) 0.44
b) 0.75
c) 0.67
d) 0.33

**Vraag 2**

Een waarschijnlijkheidsvariabele (random variable) heeft de volgende kansverdeling (distribution):

| | | | | | |
|---|---|---|---|---|---|
| x | -10 | -5 | 0 | 2 | 3 |
| p(x) | 0.1 | 0.2 | 0.3 | 0.31 | 0.09 |

Welke kans heeft als antwoord 0.5?

a) $P(X \geq 0)$
b) $P(X < 0)$
c) $P(-5 \leq X < 1)$
d) $P(-5 < X \leq 0)$

**Vraag 3**

De kans op ten minste 4 successen in een binomiaal experiment met n=6 en p=0.6 is

a) 0.7667
b) 0.5163
c) 0.2540
d) 0.5443

*(antwoord: $P(X \geq 4) = 1 - P(X \leq 3) = 1 - 0.4557 = 0.5443$*

**Vraag 4**

X is een waarschijnlijkheidsvariabele (random variable) met de volgende kansverdeling:

| | | | | | | | |
|---|---|---|---|---|---|---|---|
| x | 0 | 1 | 2 | 3 | 4 | 5 | 6 |
| p(x) | 0.1 | 0.2 | 0.2 | 0.1 | 0.1 | 0.1 | 0.2 |

Wat is de standaard afwijking van X?

a) 4.200
b) 3.000
c) 2.0494

d) 2.978

*(Antwoord: mu = 3, sigma^2 =V(x) = 4.2, sigma = wortel(sigma^2)=wortel(4.2)=2.0494)*

**Vraag 5**

Laat het jaarinkomen onder pas afgestudeerden normaal verdeeld zijn met gemiddelde van 18000 euro en een standaard afwijking van 5000. Wat is de kans dat een willekeurig net afgestudeerde tussen de 16000 en 19000 euro verdient? P( 16000 < X < 19000)= ?

a) 0.2347
b) 0.5793
c) 0.3446
d) 0.4207

*(Antwoord: P( 16000 < X < 19000) = P(X<19000)-P(X<16000) = P(Z < (19000-18000)/5000) – P( Z< (16000-18000)/5000) = P(Z<0.2)- P(Z<-0.4) = 0.5793-0.3446 = 0.2347 )*

**Vraag 6**

Het aantal uren dat een Bedrijfskunde student per nacht slaapt is normaal verdeeld met een gemiddelde van 6 uren en met een variantie van 1,5 uur.
Aan vier Bedrijfskunde studenten wordt gevraagd hoeveel uren zij per nacht slapen. Wat is de kans dat het gemiddelde van hun antwoorden groter is dan 7?

A) 0.0516
B) 0.0918
C) 0.2061
D) 0.2514

**Vraag 7**

Een fabrikant van blikgroenten wil weten hoeveel het nettogewicht is van een blik met een brutogewicht van 500 gram. Hij heeft een steekproef genomen van 25 blikken, welke resulteerde in een steekproefgemiddelde van 358 gram. Van de populatie is bekend dat de standaarddeviatie van alle blikken 15 gram is. We kunnen aannemen dat het nettogewicht van blikgroenten normaal verdeeld is. Het 95% betrouwbaarheidsinterval van μ afgerond op één decimaal, is:

a) 351.8< μ < 364.2
b) 352.1< μ< 363.9

c) 353.1< $\mu$< 362.9
d) 349.7< $\mu$< 366.3

x +- Za/2 * sigma/wortelN =
358 + 1.96 * 15/5 = 363,9
358 – 1.96 * 15/5 = 352.1

**Vraag 8**

Het bedrag dat volwassenen gemiddeld per week in het café uitgeven is normaal verdeeld met eens standaarddeviatie van 9 euro. Bij een onderzoek worden 32 mensen gevraagd naar hun wekelijkse uitgaven in het café en het gemiddelde daarvan is 23 euro. Geef de ondergrens van het 90 % betrouwbaarheidsinterval (Engels: lower confidence limit)?

a. 21.12
b.22.32
c. 20.38
d. 19.88
blz. 326

**Vraag 9**

De inhoud van een pak melk is normaal verdeeld met een gemiddelde van 1 liter en een standaarddeviatie van 10 milliliter. Bepaal het aantal pakken dat getest moet worden om een schatting van de inhoud te maken binnen 5 milliliter en met een 90% betrouwbaarheidsniveau. (Engels: confidence level).

a. 11
b. 27
c. 7
d. 22

**Vraag 10**

Een onderzoekster heeft een steekproéf genomen onder de studenten van de FEB. Honderd studenten werd gevraagd hoeveel uur per week ze aan hun studie besteden. Het bestuur van de faculteit denkt dat de gemiddelde student gemiddeld 30 uur per week aan hun studie besteedt. Welke hypotheses moet de onderzoekster opstellen om te testen of er bewijs is dat de gemiddelde student minder dat 30 uur per week aan zijn studie besteed?

a. H0: $\mu = 30$ en H1: $\mu \neq 30$

b. H0: $\mu \geq 30$ en H1: $\mu < 30$

c. H0: $\mu = 30$ en H1: $\mu < 30$

d. H0: $\mu = 30$ en H1: $\mu \leq 30$

**Vraag 11**

De gemiddelde lengte van een Nederlandse man is 188cm. Er is een steekproef genomen met een omvang van n=25. Lengte van mannen is normaal verdeeld met $\sigma$ = 8. Het gemiddelde van de steekproef is 185. Bereken de standardized test statistic (*gestandaardiseerde test-statistiek)* en trek een conclusie.
H0: $\mu$=188
H1: $\mu \neq$ 188
$\alpha$ = .05

a) 1.875, deze waarde ligt niet tussen -1.645 en 1.645 dus H0 wordt verworpen
b) -1.875, deze waarde tussen -1.96 en 1.96 dus H0 wordt niet verworpen
c)-1.875, deze waarde tussen -1.96 en 1.96 dus H0 wordt verworpen
d)1.875, deze waarde ligt niet tussen -1.645 en 1.645 dus H0 wordt niet verworpen
*Uitwerking: z < -z_ α/2 = -1.96 of z > z_ α/2 = 1.96 , standardized test statistic: z = ( xgemiddeld – μ)/ (σ / √n) = (185 – 188)/ (8/ √25)= -1.875, H0 niet verworpen -1.96 < -1.875 < 1.96, blz 351*

**Vraag 12**

Bepaal de overschrijdingskans (p-value) van de volgende hypothese:
H0: $\mu$=60
H1: $\mu$<60
$\sigma$ =15, n=100, xgemiddeld = 58, $\alpha$ = .05
a) 0.0918
b) 0.0448
c) -1.3333
d) 0.1089

*Uitwerking: -z_.05 = -1.645, z = ( xgemiddeld – μ)/ (σ / √n) = (58 – 60)/ 15/ √100)= -2/1.5 = -1.3333, p-value = P(Z < -1.33333) = 0.0918,*

**Vraag 13**
Vervalt omdat hier abusievelijk niet het goede antwoord in de meerkeuze lijst is opgegeven. Iedereen heeft hier 3 punten voor gekregen.

**Vraag 14**

Een onderzoeker is geïnteresseerd in het gemiddelde aantal kerstkaarten dat Nederlanders versturen. Het gemiddelde van een steekproef van 16 mensen is 18 kaarten en de steekproef variantie is 5. Kan de onderzoeker concluderen dat het populatie gemiddelde groter is dan 17?

a. Ja dat kan op een 2.5 procent significantie niveau
b. Ja dat kan op een 1 procent significantie niveau
c. Ja dat kan op een 5 procent significantie niveau
d Nee dat kan op geen enkel significantie niveau

***Antwoord:***
H0: gemiddelde =17
H1: gemiddelde > 17

De t-statistiek is 1.7888.

In de t-tabel zien we bij 15 vrijheidsgraden de volgende opties:

1% 2.602
2.5% 2.131
5% 1.753
10% 1.341

We kunnen dus concluderen dat we H0 op een 5% significantie niveau kunnen verwerpen.

**Vraag 15**

Een steekproef met een grootte van n = 25 wordt getrokken uit een normaal verdeelde populatie. Het steekproef gemiddelde is $\bar{x} = 10$ en de standaard deviatie van de steekproef bedraagt s = 3. Geef de bovengrens van het 95% betrouwbaarheidsinterval voor μ.

A) 13.72 *(er wordt uitgegaan van $s^2$)*
B) 11.03 *(als er wordt gekeken naar $t_{\alpha,n-1}$)*
C) 11.24
D) 11.55 *(als wordt gedaan: x + $t_{\alpha/2,n-1}$ * s/(sqrt(n)-1)*

*Antwoord:*
UCL = x + $t_{\alpha/2,n-1}$ * s/sqrt(n) = 10 + 2.064 * 3/5 = 11.24

**Open vraag 1 (25 punten)**

In december is in Stockholm de "De prijs van de Zweedse Rijksbank voor economie, ter nagedachtenis aan Alfred Nobel", oftewel de Nobelprijs voor de Economie, uitgereikt.

Er werd een collecte gehouden aan het begin van de avond. De donaties, omgerekend in euros, blijken normaal verdeeld met een standaardafwijking van €1,40. Aan het eind van de avond vragen we aan 7 mensen hoeveel geld ze hebben gedoneerd. Dit waren hun antwoorden:

€2,90 €5,20 €4,70 €6,80 €12,50 €9,30 €4,10

a. Bereken het 95%-betrouwbaarheidsinterval van de gemiddelde donatie. [7 punten]

*Gemiddelde: 2 punten*
*Berekening alpha: 1 punt*
*Juiste formule gebruikt: 2 punten*
*Juiste grenzen: 2 punten*

Steekproefgemiddelde: $\bar{x} = \dfrac{2{,}90+5{,}20+4{,}70+6{,}80+12{,}50+9{,}30+4{,}10}{7} = €6{,}50$

(p98)

Betrouwbaarheidsinterval is 95%, dus $\alpha = 0{,}05$.

Ondergrens betrouwbaarheidsinterval: $\bar{x} - z_{\alpha/2}\dfrac{\sigma}{\sqrt{n}} = 6{,}50 - 1{,}96\dfrac{1{,}40}{\sqrt{7}} \approx €5{,}46$

Bovengrens betrouwbaarheidsinterval: $\bar{x} + z_{\alpha/2}\dfrac{\sigma}{\sqrt{n}} = 6{,}50 + 1{,}96\dfrac{1{,}40}{\sqrt{7}} \approx €7{,}54$

Bij deze uitreiking was slechts een klein aantal oud-Nobelprijswinnaars aanwezig. Hun leeftijden waren op deze dag als volgt:

67 82 93 71 77

b. Bereken het gemiddelde en de standaardafwijking. [7 punten]

*Gemiddelde: 2 punten*
*Standaardafwijking:*
*Juiste formule: 3 punten*
*Correct antwoord: 2 punten*
*Indien gedeeld door n ipv n-1: 1 punt aftrek*
*Steekproefvariantie ipv standaardafwijking: 1 punt aftrek*

Steekproefgemiddelde: $\bar{x} = \dfrac{67+82+93+71+77}{5} = 78$

Steekproefstandaardafwijking:

$$s^2 = \frac{1}{n-1}\left[\sum_{i=1}^{n} x_i^2 - \frac{\left(\sum_{i=1}^{n} x_i\right)^2}{n}\right] = \frac{1}{5-1}\left[30832 - \frac{(390)^2}{5}\right] = \frac{1}{4}\left[30832 - \frac{152100}{5}\right] = \frac{412}{4}$$

$s = \sqrt{s^2} = \sqrt{103} \approx 10{,}149$

Berekening kan ook via de formule $s^2 = \dfrac{\sum_{i=1}^{n}(x_i - \bar{x})^2}{n-1}$.

c. Toets de hypothese $H_0$: $\mu = 70$ tegen de alternatieve hypothese $H_1$: $\mu > 70$, op een significantieniveau van 0,05. [11 punten]

*Test statistic:*

*Juiste formule: 2 punten*
*Correct ingevuld: 2 punten*
*Noemen vrijheidsgraden: 1 punt*
*Aantal vrijheidsgraden correct: 1 punt*

*Gebruik alpha 5% (en niet gedeeld door 2): 1 punt*
*Kritieke waarde correct: 2 punten*
*Juiste conclusie: 2 punten*

*Indien formulering "accepteren $H_0$": 1 punt aftrek*

De test statistic is $t = \dfrac{\bar{x} - \mu}{s/\sqrt{n}} = \dfrac{78-70}{\sqrt{103}/\sqrt{5}} = \dfrac{8}{\sqrt{20{,}6}} \approx \dfrac{8}{4{,}539} \approx 1{,}763$

met v = n-1 = 4 vrijheidsgraden.

De kritieke t-waarde is $t_\alpha = t_{0,05} = 2{,}132$.

Aangezien $t < t_{0,05}$ verwerpen we $H_0$ niet.

**Open vraag 1 (30 punten)**

**a. 10 punten**

Een onderzoeker is geïnteresseerd in de vraag of we op basis van de volgende observaties die uit twee nonnormale populaties getrokken zijn, op een 5% significantie niveau kunnen stellen dat de locatie van populatie 1 zich aan de linkerkant van de locatie van populatie 2 bevindt.

| | | | | | | |
|---|---|---|---|---|---|---|
| Steekproef 1: | 12 | 15 | 14 | 13 | 21 | |
| Steekproef 2: | 16 | 12 | 17 | 12 | 12 | 15 |

Stel de hypothese op en gebruik een non-parametrische toets om de bewering van de onderzoeker te testen.

*Antwoord:*
*- Hypotheses (1)*
*- Ranken (3)*
*- Rank sum uitrekenen (2)*
*- Kritieke waarde opzoeken (2)*
*- Vergelijken + conclusie (2)*

Hypotheses:
$H_0$: De twee populatie locaties zijn hetzelfde
$H_1$: De locatie van populatie 1 bevindt zich links van de locatie van populatie 2.

rank the observaties

| Sample 1 | Rank | Sample 2 | rank |
|---|---|---|---|
| 12 | 2.5 | 16 | 9 |
| 15 | 7.5 | 12 | 2.5 |
| 14 | 6 | 17 | 10 |
| 13 | 5 | 12 | 2.5 |
| 21 | 11 | 12 | 2.5 |
| | | 15 | 7.5 |
| | **32 = $T_1$** | | **34 = $T_2$** |

Let op: de score 12 komt 4 keer voor en is de kleinste score. Dus elke 12 krijgt een rank van (1+2+3+4)/4= 2.5. 15 komt 2 keer voor. Dus krijgt 7.5

Stap 2: Vergelijk de rank $T = T_1 = 32$ met de rank die je in tabel 19.3 vindt bij $n_1=5$ en $n_2=6$.
$T_L = 20$; $T_U = 40$.

Omdat 32 niet kleiner is dan 20 kunnen we de nulhypothese **niet** verwerpen.

**B.** 10 punten

Een sap fabrikant is geïnteresseerd in de voorkeur van mensen voor bepaalde sappen. De fabrikant produceert 3 smaken en heeft aan 200 mensen gevraagd wat hun favoriete smaak is. De uitkomst is als volgt:

| | |
|---|---|
| Kiwi-appel: | 100 |
| Bosbes-mango: | 70 |
| Aardbei-banaan: | 30 |

Voordat de resultaten van de steekproef bekend zijn gemaakt veronderstelde de directeur van het bedrijf dat 50 procent van de consumenten kiwi-appel sap als favoriet zal kiest, 40% bosbes-mango sap zal kiezen en 10 procent aardbei-banaan zal kiezen.

Kunnen we op een 5 % significantie niveau stellen dat de daadwerkelijke voorkeur van de consument afwijkt van de veronderstelling van de directeur? Stel de hypotheses op en voer de test uit.

*Antwoord:*

*- Hypotheses (1)*

*- $e_i$ uitrekenen (3)*

*- Chi-square goodness of fit statistic uitrekenen (3)*

*- Kritieke waarde opzoeken (2)*

*- Conclusie (1)*

Hypotheses:

$H_0$: $p_1 = 0.5$, $p_2 = 0.4$, $p_3 = 0.1$

$H_1$: in ieder geval 1 $p_i$ is niet gelijk aan de gespecificeerde waarde

De verwachte waarde als de nulhypothese waar zou zijn is $e_i = np_i$

$e_1 = 200*0.5=100$

$e_2 = 200*0.4=80$

$e_3 = 200*0.1=20$

chi-square goodness of fit test statistiek

$$\chi^2 = \sum_{i=1}^{k} \frac{(f_i - e_i)^2}{e_i} = \frac{(100-100)^2}{100} + \frac{(70-80)^2}{80} + \frac{(30-20)^2}{20} = 6.25$$

De kritieke waarde is $\chi^2$ met $\nu = k-1$ vrijheidsgraden

$\chi^2_{\alpha,k-1} = \chi^2_{0.05,2} = 5.99$

Omdat $\chi^2 > \chi^2_{\alpha,k-1}$ wordt $H_0$ verworpen.

c. 10 punten

Een docent statistiek is geïnteresseerd in de prestatie van mannen en vrouwen bij een statistiek tentamen. In de onderstaande tabel is weergegeven hoeveel voldoendes en onvoldoendes er in beide groepen vielen.

| | Voldoende | Onvoldoende | |
|---|---|---|---|
| Mannen | 50 | 70 | [redacted] |
| Vrouwen | 25 | 15 | [redacted] |
| | [redacted] | [redacted] | [redacted] |

Test of geslacht en al dan niet een voldoende halen afhankelijke variabelen zijn. Stel de hypothese op en voer de test uit op een 5% significantie niveau.

*Antwoord:*
- *Hypotheses (1)*
- *Uitrekenen van onafhankelijke kansen (2)*
- *Uitrekenen van $e_i$ (kansen vermenigvuldigen met sample size) (2)*
  *(Of: formule 591 (4))*
- *Chi-square uitrekenen = 2 punten*
- *Rejection chi-squrea opzoeken = 2 punt*
- *Conclusie trekken = 1*

Hypotheses:
$H_0$: the twee variabelen zijn onafhankelijk
$H_1$: de twee variabelen zijn afhankelijk

P(man) = 0.75
P(vrouw) = 0.25
P(voldoende) = 0.46875
P(onvoldoende) = 0.53125

P(A en B) = P(A)*P(B) (onder onafhankelijkheid)

Kansen:

| | voldoende | onvoldoende |
|---|---|---|
| Mannen | 0.352 | 0.398 |
| Vrouwen | 0.117 | 0.133 |

Vermenigvuldig de kansen met n:

| | voldoende | onvoldoende |
|---|---|---|
| Mannen | 56.32 | 63.68 |
| Vrouwen | 18.72 | 21.28 |

$$\chi^2 = \sum_{i=1}^{k} \frac{(f_i - e_i)^2}{e_i} = \frac{(50-56.32)^2}{56.32} + \frac{(70-63.68)^2}{63.68} + \frac{(25-18.72)^2}{18.72} + \frac{(15-21.28)^2}{21.28} = 5.3$$

Vrijheidsgraden v = (r-1)(c-1) = 1

Kritieke waarde: $\chi^2_{\alpha,v} = \chi^2_{0.05,1} = 3.84$

Omdat
$\chi^2 > \chi^2_{\alpha,v}$ **verwerpen** we $H_0$.