EXTRA VRAAGSTUKKEN

Oefening 1

Uit een boek van 52 speelkaarten wordt random een kaart getrokken. Toon aan dat:
a) de gebeurtenissen “de kaart is een schoppen” en “de kaart is rood” niet onafhankelijk zijn
b) de gebeurtenissen “de kaart is een schoppen” en “de kaart is een heer” onafhankelijk zijn

Oefening 2

Gegeven: $P(A)=1/2$, $P(B)=1/3$ en $P(A \cap B)=1/4$
Bereken: $P(AUB)$, $P(A^C \cap B)$, $P(A^C UB)$, $P(A^C UB^C)$

Antwoord:
In volgorde: 7/12, 1/12, 3/4, 3/4

Oefening 3

De stad Nicosia in Cyprus bestaat uit 75% Grieken en 25% Turken. Van de Grieken spreekt 20% Engels en van de Turken 10%. Een bezoeker vraagt aan een persoon de weg in het Engels en deze persoon antwoordt in het Engels. Wat is de kans dat deze persoon een Griek is?

Antwoord:
6/7

Oefening 4

U bevindt zich in een gezelschap van 15 personen. Bereken de kans berekenen dat minstens 2 van hen dezelfde verjaardag hebben. Negeer het bestaan van schrikkeljaren en neem aan dat de kans dat iemand op een bepaalde dag verjaart 1/365 is.

Antwoord:
De kans dat iedereen een verschillende verjaardag heeft is 0.747. (Tip: maak gebruik van de complementregel om deze kans te berekenen. Maak bovendien gebruik van de veralgemeende vermenigvuldigingsregel om de kans te berekenen dat iedereen op een andere dag verjaart.)

Oefening 5

De kansen bij de geboorte zijn 0.51 voor een jongen en 0.49 voor een meisje. Een familie heeft 6 kinderen.
a) Wat is de kans dat er tenminste 1 jongen is?
b) Wat is de kans dat er tenminste 1 meisje is?
c) Wat is de kans dat alle kinderen meisjes zijn?
d) Wat is de kans dat alle kinderen jongens zijn?
e) Wat is de kans dat alle kinderen van hetzelfde geslacht zijn?

Antwoord:
a) 0.9862
b) 0.9824
c) 0.0138

d) 0.0176
e) 0.0314

Oefening 6

Sinaasappelen worden verpakt in kisten van 400 stuks.

a) Als uit ondervinding blijkt dat er 30% slechte zijn, bereken dan de kans dat een kist tenminste 130 slechte sinaasappelen bevat.
b) Bereken in het geval (a) de kans dat als vier willekeurige kisten gecontroleerd worden, er juist twee kisten tenminste 130 slechte sinaasappelen bevatten.

Antwoord:

a) 0.15 (Het aantal slechte sinaasappelen op 400 stuks is binomiaal verdeeld.)
b) 0.097

Oefening 7

Een kweekproces van bepaalde planten levert een normaal verdeelde opbrengst van gemiddeld 3000 zaadjes op. De kans dat de opbrengst minder dan 2632 zaadjes of meer dan 3368 zaadjes oplevert is 50% (samen!). Bepaal de standaardafwijking.

Antwoord:

$\sigma = 545$

Oefening 8

Beschouw de dataset 'pollutiedataset.txt' Ga voor het gemiddelde μ van de variabele 'Mortality' na welk van de hypothesen

$$\mu = 1000$$
$$\mu \neq 1000$$

de voorkeur krijgt (significantieniveau α=0.05). Maak hiervoor gebruik van drie benaderingen (kritieke waarden, toetsingsgrootheid, p-waarde). Vergelijk de resultaten met deze verkregen op basis van het commando t.test in R.

Antwoord:

1) Het kritiek gebied bestaat uit de intervallen: $]-\infty,983.9272[$ en $]1016.073,+\infty[$
Het steekproefgemiddelde (=940.3487) valt in het kritiek gebied zodat de nulhypothese (μ=1000) verworpen wordt.
2) Het kritiek gebied bestaat uit de intervallen: $]-\infty,-2.000995[$ en $]2.000995,+\infty[$
De toetsingsgrootheid (=-7.426348) valt in het kritiek gebied zodat de nulhypothese verworpen wordt.
3) De p-waarde van de test is gelijk aan 5.126739×10^{-10} (veel kleiner dan 0.05) zodat de nulhypothese verworpen wordt.
4) commando t.test(Mortality,mu=1000,alternative="two.sided")

Oefening 9

Het ideale cholesterol gehalte voor vrouwen van 35 jaar bedraagt 175. Een arts kan gebruik maken van cholesterol gehaltes van 12 vrouwen uit Oost-Vlaanderen: 136, 232, 187, 168, 287, 189, 175, 160, 272, 145, 190, 210. Zijn er redenen om aan te nemen dat de gemiddelde waarde te hoog ligt?

Antwoord:
De grootte van de steekproef is kleiner dan 30. We dienen daarom eerst na te gaan of de gegevens normaal verdeeld zijn. De p-waarde van de Lillieforstest is gelijk aan 0.5551 zodat we normaliteit mogen veronderstellen en de T-toets mogen toepassen. Indien het significantieniveau van de test gelijk aan 0.05 gekozen wordt, zijn er geen redenen om aan te nemen dat de gemiddelde waarde te hoog ligt. De p-waarde van de test is gelijk aan 0.07624.

Oefening 10

Een veel voorkomend probleem bij grote winkels, is het voorzien van net voldoende kassiers op alle momenten. Het management moet in functie daarvan werknemers aannemen en uurroosters opstellen. Om beter inzicht te krijgen in het verloop van de klanten, meet men het aantal klanten dat binnenkomt in intervallen van 5 minuten. Het aantal klanten werd ingedeeld in 5 klassen: 0 of 1 klant, 2 of 3 klanten, 4 of 5 klanten, 6 of 7 klanten, 8 of meer klanten. We observeren het volgende:

Aantal klanten per interval van 5 minuten	Frequentie
[0,1]	10
[2,3]	22
[4,5]	40
[6,7]	38
≥8	18

Dit betekent dat het 10 keer voorkwam dat er 0 of 1 klant binnenkwam in een interval van 5 minuten, 22 keer dat er 2 of 3 klanten binnenkwamen,… Een veel gebruikte verdeling voor zulke problemen is de Poissonverdeling. Als we kunnen aannemen dat het verloop van de klanten een Poissonverdeling volgt, dan kan op basis van deze informatie een uurrooster uitgewerkt worden. Ga na of hier inderdaad sprake is van een Poissonverdeling, met λ=5. Neem α gelijk aan 0.05.

Antwoord:
We voeren een χ^2–test uit met Oi=[10,22,40,38,18] en theoretische frequenties Ei=[1280,2816,5120,4864,2304]. De toetsingsgrootheid χ is gelijk aan 7.76313. We vergelijken deze met de rechter kritieke waarde $\chi^2_{\alpha;4}$ = 9.487729. Vermits 7.7<9.4 aanvaarden we de nulhypothese (Poisson verdeeld met λ=5). De p-waarde van de test is gelijk aan 0.1.

Oefening 11

Men wenst na te gaan of het Cd gehalte verschilt tussen Schelde, Maas en Ijzer. In elke rivier neemt men 5 onafhankelijke waterstalen op verschillende plaatsen en bepaalt het Cd gehalte (in ppb). De data staan in onderstaande tabel.

Schelde	**Maas**	**Ijzer**
0.5	1.2	1.0
0.2	0.9	1.1
0.3	1.0	0.9
0.6	1.4	1.2
0.4	1.3	0.9

Ga na of er verschillen zijn en zo ja, welke rivieren significant van elkaar verschillen. Vergeet niet de voorwaarden van de door u gekozen test te onderzoeken.

Antwoord:
Uit een ANOVA analyse volgt dat er een hoogsignificant verschil is tussen de 3 rivieren. De toetsingsgrootheid f is gelijk aan 28.9. De p-waarde van de test is kleiner dan 0.0001. Er zijn geen aanwijzigingen dat er problemen zijn met de onderliggende voorwaarden (Bartlett test: p=0.67; Lilliefors test: p=0.95). Op basis van de Tukey methode kan men besluiten dat er geen significant verschil is tussen Ijzer en Maas (p=0.41), het cadmium gehalte het laagst is in de Schelde ten opzichte van de twee andere rivieren (verschil met Maas=0.76, p<0.0001; verschil met Ijzer=0.62, p=0.0002). De data kunnen grafisch weergegeven worden in een boxplot.

<u>Oefening 12</u>

Van de producten die in de laatste maanden door een bedrijf zijn vervaardigd, voldoet 11% niet aan de specificaties. Het bedrijf wijzigt het productieproces in een poging het percentage dat niet voldoet te verminderen. Tijdens een proefronde produceerde het gewijzigde productieproces op een totaal van 300 producten, 16 producten die niet voldeden aan de eisen. Tonen deze resultaten aan dat de wijziging effectief is? Onderbouw uw conclusies met een duidelijke formulering van uw aannames en met de resultaten van uw statistische berekeningen.

Antwoord:
We voeren een linkseenzijdige Z-toets uit voor proportie (H_0: $\pi=\pi_0$; Ha: $\pi<\pi_0$). Berekende toetsingsgrootheid z=-3.155. De berekende p-waarde is gelijk aan 0.0008024. Op basis van de p-waarde verwerpen we de nulhypothese en aanvaarden de alternatieve hypothese. Het gewijzigde productieproces is effectief.

<u>Oefening 13</u>

Geef, in de situatie van de vorige opgave, een 95% betrouwbaarheidsinterval voor de fractie ‘onvoldoende’ producten in het gewijzigde proces. Neem nu π_0=0.11 als de oude fractie en stel de fractie voor het gewijzigde proces gelijk aan π. Geef nu een 95% betrouwbaarheidsinterval voor π- π_0.

Antwoord:
[0.027 ; 0.078]
Het betrouwbaarheidsinterval voor π- π_0 wordt gevonden door 0.11 van het vorige interval af te trekken: [-0.083 ; -0.032]. Met andere woorden: het is 95% betrouwbaar dat het percentage ‘onvoldoende’ producten van het nieuwe proces 3.12% tot 8.21% lager ligt dan die van het oude proces.

<u>Oefening 14</u>

Het ontwerp van controleknoppen en instrumenten heeft grote invloed op de mate van het gemak waarmee mensen ze kunnen gebruiken. Dit effect werd onderzocht door aan 25 rechtshandige studenten te vragen met de rechterhand te draaien aan een knop die via een schroefbeweging een wijzer in werking bracht. Er waren twee identieke instrumenten, het ene met een rechtse schroef (de knop draait in wijzerrichting) en het andere met een linkse schroef (de knop draait in tegenwijzerzin). De data in de file ‘schroefbeweging.txt’ geeft de tijden (in seconden) die nodig waren om de wijzer over een vaste afstand te draaien. Het project hoopte aan te tonen dat rechtshandigen het gemakkelijker vinden rechtsdraaiende knoppen te gebruiken.

Formuleer geschikte hypothesen. Voer een toets uit van uw hypothesen. Geef de overschrijdingskans en doe verslag van uw bevindingen.

Antwoord:
We voeren een T-toets uit voor gemiddelden (gepaard) met hypothesen H_0: $\delta=0$; Ha: $\delta<0$ waarbij δ de verwachting is van 'rechtse knop – linkse knop'. De berekende toetsingsgrootheid t=-2.90. Met df=24 zien we dat de p-waarde gelijk is aan 0.003896. Dit is een sterk bewijs dat het verwachte verschil negatief is, d.w.z. dat de gemiddelde tijd voor rechtsdraaiende schroeven minder is dan die voor linksdraaiende schroeven. Aangezien de steekproef kleiner is dan 30 moeten we de normaal verdeeldheid van de verschilvariabele nagaan. Hiervoor kunnen we een Lillieforstoets uitvoeren. De p-waarde van de toets is gelijk aan 0.9573 wat een sterk bewijs is dat de verschilvariabele normaal verdeeld is.

<u>Oefening 15</u>

Pas de tekentoets toe op de data in de file 'schroefbeweging.txt' (zie vorige opgave) om te beoordelen of de proefpersonen met de rechtsdraaiende schroef een taak significant sneller kunnen uitvoeren dan met de linksdraaiende schroef.

Antwoord:
We maken gebruik van de Wilcoxon rangtekentoets voor 2 gepaarde steekproeven. De uitkomst van de toetsingsgrootheid W^{+}=V=56.5. De kans op een kleinere waarde is 0.003929, zodat we opnieuw zeer sterke aanwijzigingen hebben dat de gemiddelde tijd voor rechtsdraaiende schroeven minder is dan die voor linksdraaiende schroeven.

<u>Oefening 16</u>

Bij een onderzoek naar de effecten van Cd op de conditie van zeehonden wordt bij 10 zeehonden in de Waddenzee een staal van vetweefsel genomen en hun conditie bepaald (gewicht/lichaamslengte). Is er een verband en hoe kan de conditie voorspeld worden aan de hand van het Cd gehalte in het vetweefsel? Voorspel wat de gemiddelde conditie is van zeehonden met 12 ppb Cd in hun vetweefsel. Geef ook het bijbehorende betrouwbaarheidsinterval. Voorspel ook de conditie van een recent gevangen zeehond die ontsnapte voordat de lichaamsmetingen werden uitgevoerd en met 20 ppb in zijn vetweefsel. De volgende data werden verzameld:

ind	1	2	3	4	5	6	7	8	9	10
conditie	1.2	1.8	2.9	0.9	1.6	2.1	1.8	2.0	1.4	1.7
Cd	0.4	0.9	2.1	0.2	0.8	1.5	1.2	1.3	0.6	1.0

Antwoord:
We passen lineaire regressie toe op de gegevens. Er bestaat een significant verband tussen het Cd gehalte en de conditie van de zeehonden (t=16.1, p<0.0001). De geschatte regressierechte kan als volgt geschreven worden: conditie = 0.78 + 0.96 Cd. De gemiddelde conditie van zeehonden met een Cd gehalte van 12 ppb is 12.35 met een 95% betrouwbaarheidsinterval van [10.82 13.89]. De voorspelde conditie wanneer het Cd gehalte gelijk is aan 20 is gelijk aan 20.06 met een 95% betrouwbaarheidsinterval van [17.42 22.70].