Elementaire Statistiek

tuyaux.winak.be/index.php/Elementaire_Statistiek

Elementaire Statistiek

Richting	Informatica
Jaar	2BINF

Bespreking

Het vak werd in het academiejaar 2017-2018 gegeven door Valérie de Witte. De theorie is best wel taai om te studeren en de vragen die er rond worden gesteld zijn redelijk moeilijk. Oefeningen gaan iets vlotter dan de theorie en men kan op het examen oefeningen van hetzelfde kaliber verwachten. Sinds 2019-2020 wordt dit gegeven door Tim Verdonck met Valérie de Witte als assistent. eh

Puntenverdeling

8/20 THEORIE: multiple choice, zonder gis-correctie, maar men moet wel uitleg geven, anders geen punten.

8/20 OEFENINGEN

4/20 PROJECT in R

Examenvragen

Er wordt zeer strikt op gelet dat er geen bladen worden meegenomen, zelfs al je kladpapier moet afgegeven worden. Dus vragen het lokaal uitsmokkelen is onmogelijk en dus is er gier niet veel te zien :(

Academiejaar 2019 - 2020 (1ste zittijd)

Formularium en tabellen mocht je doorheen heel het examen gebruiken. Het examen staat op 16 punten, het project op de overige 4 punten. Indien niet anders vermeld gaan we ervan uit dat $\alpha=0.05$α=0.05.

1. (2 punten) Omcirkel en motiveer het juiste antwoord
    1. P(A)=14P(A)=14, P(B)=14P(B)=14 en P(A∪B)=1316P(A∪B)=1316
        1. A en B zijn onafhankelijk
        2. A en B zijn afhankelijk
        3. Onmogelijk te bepalen met de gegeven info
    2. Xi,...XnXi,...Xn zijn onderling onafhankelijk en identiek verdeelde stochastische variabelen met dichtheidsfunctie fx(x)=1π(1+x2)∀x∈Rfx(x)=1π(1+x2)∀x∈R dan geldt de centrale limietstelling als n→+∞n→+∞
        1. Waar
        2. Niet waar
    3. V, W, X, Y en Z zijn onafhankelijke standaard normaal verdeelde stochastische veranderlijken, dan is de stochastische veranderlijke U=2ZV2+W2+X2+Y2U=2ZV2+W2+X2+Y2
        1. Normaal verdeeld
        2. χ2χ2 verdeeld met 4 vrijheidsgraden
        3. F1,4F1,4 verdeeld
        4. Student verdeeld met 4 vrijheidsgraden
        5. Geen van bovenstaande
    4. Men bekomt met lineaire regressie het verband ŷ = 14.947x + 74.283. het steekproefgemiddelde is $\overline{yn}$=92.16y$\overline{n}$=92.16 voor de respons. Wat is het steekproefgemiddelde voor de verklarende variabele?
        1. $\overline{xn}$=1.196x$\overline{n}$=1.196
        2. $\overline{xn}$=11.136x$\overline{n}$=11.136
        3. $\overline{xn}$=1.442x$\overline{n}$=1.442
        4. $\overline{xn}$=1.040x$\overline{n}$=1.040
2. (2 punten) Stel we hebben n steekproefgegevens x1,...xnx1,...xn (realisaties van de toevalsveranderlijke X)
    1. Leg zo gedetailleerd mogelijk uit welke punten je zou plotten in een normaal kwantielplot om na te gaan of deze gegevens uit een normale verdeling komen met parameters μμ en σ2σ2
    2. Hoe verwacht je dat deze plot er uitziet indien de gegevens inderdaad uit een normale verdeling komen?
    3. Hoe kan je deze plot gebruiken om μμ en σσ te schatten?
3. (2 punten) Een landbouwer heeft 80 koeien. Een veearts zegt dat de kans dat een koe meer dan 3 vlekken heeft 43.75% is. Wat is de kans dat meer dan 50 koeien hoogstens 3 vlekken hebben? Gebruik onderstaande R output en duid de gekozen output duidelijk aan. Geef zoveel mogelijk cijfers na de komma.

```
> x = c(49, 49.5, 50, 50.5, 51)
> ppois(x, lambda=19.6875)
[1] 1 1 1 1 1
> ppois(x, lambda=45)
[1] 0.7531980 0.7531980 0.7962803 0.7962803 0.8342941
> pnorm(x, mean=45,sd=19.6875)
[1] 0.5805007 0.5903990 0.6002400 0.6100177 0.6197263
> pnorm(x, mean=45, sd=sqrt(19.6875))
[1] 0.8163382 0.8447528 0.8701018 0.8924308 0.9118518
> pnorm(x, mean=45, sd=19.6875*19.6875)
[1] 0.5041170 0.5046316 0.5051462 0.5056608 0.5061754
```

1. (3 punten) In onderstaande tabel zie je de keuze van 119 spelers bij de eerste beurt in schaar-steen-papier (en niet steen-papier-schaar zoals op het examen stond eh). Een speler heeft een voordeel indien je kan bewijzen dat de keuze in de eerste beurt niet gelijk verdeeld is over de 3 opties. Gebruik een test om te zien of de 3 mogelijke uitkomsten uniform verdeeld zijn op basis van de data (en significantieniveau $\alpha=0.05\alpha=0.05$). Welke keuze raad je aan om het spel mee te starten?

Schaar	14
Steen	66
Papier	39

1. (5 punten) Er worden metingen gedaan op het gewicht van zakken chips van 2 verschillende merken (Lazy en Crocks). Van elk merk worden 100 zakken genomen van 250 gram en geanalyseerd in R (zie code hieronder). Maak zoveel mogelijk gebruik van de data in de uitvoer en de figuren.
    1. Vul de ontbrekende woorden gemaarkeerd in de output met (1), (2) en (3) zo nauwkeurig mogelijk aan uit uitvoer 5 en 7. Geef hierbij telkens en berekening en zo nauwkeurig mogelijke benadering.
    2. Is de inhoud van Lazy gemiddeld meer dan van Crocks? ($\alpha=0.01\alpha=0.01$). Onderzoek ook of de voorwaarden voor de hypothesetest voldaan zijn.
    3. Maak een schets waarop je de testwaarde en p-waarde die horen bij de hypotheses in b aanduidt.
    4. Leg in woorden de betekenis van de p-waarde uit.

R output

Spoiler

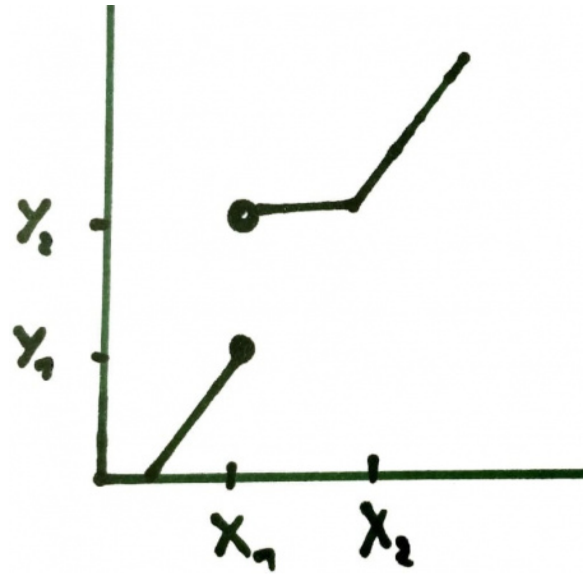
Academiejaar 2018 - 2019 (2de zittijd)

Theorie

1. $P(A)=0,3;P(\overline{A}\cup\overline{B})=0,5;P(\overline{B}|\overline{A})=0,5P(A)=0,3;P(\bar{A}\cup\bar{B})=0,5;P(\bar{B}|\bar{A})=0,5$ is:
    1. P(B) < P(A)
    2. P(B) > P(A)
    3. P(B) = P(A)
2. Zie grafiek in afbeelding.

1.
    1. x_1x_1 is kwantiel van y_1y_1
    2. x_1x_1 is kwantiel van y_2y_2
    3. y_1y_1 is kwantiel van x_1x_1
    4. x2−x12x2−x12 is kwantiel van y_2y_2

2. Zes patienten hebben last van slapeloosheid. Er wordt een test gedaan voor een slaappil. Gemiddeld hadden de patienten 1.35u meer slaap met een afwijking van [0,85; 1,85].
    1. Slaapmiddel is effectief
    2. Slaapmiddel is niet effectief
3. XX is student verdeelt met n vrijheidsgraden, dan is X2X2
    1. Student verdeeld met n+1 vrijheidsgraden
    2. normaal verdeeld
    3. χ2χ2 verdeeld
    4. F-verdeeld
    5. [Laatste optie vergeten]



Grafiek bij vraag 2 (verdelingsfunctie)

Oefeningen

1. In het restaurant van Ikea verkiest 40% van de bezoekers balletjes, 10% salade, 35% zalm en 15% soep. Van de personen die balletjes kiezen is 50% tevreden, van de salade 70%, van de zalm x%, van de soep y% (x en y waren gegeven, de exacte getallen zijn we vergeten)
    1. Wat is de kans dat een willekeurig persoon ontevreden was?
    2. Wat is de kans dat een willekeurig persoon ontevreden was, wetende dat hij zalm had?
    3. Wat is de kans dat twee personen ontevreden waren over hun soep?
    4. Wat is de kans dat voor 2 personen die soep aten 1 tevreden was en 1 niet?

Academiejaar 2016 - 2017 - Eerste zittijd

In het onderstaande bestand vind je enkele vragen die zijn voorgekomen, niet allemaal dus:

Media:ElemStat1617zit1.pdf

Academiejaar 2009 - 2010 - 2de zittijd

1. Formuleer de Centrale Limiet Stelling. Leg beknopt uit waarom deze stelling belangrijk is in de statistiek.
2. Geef in het kader van een algemene hypothesetest de definitie van
    1. Type-II fout
    2. p-waarde
    3. Teststatistiek
    4. Nullhypothese

3. Een continue stochastische veranderlijke X heet exponentieel verdeeld met parameter Lambda > 0 indien de dichtsheidsfunctie gelijk is aan f(x)=e−xλλf(x)=e−xλλ met x>=0x>=0 Bovendien is dan EX=λEX=λ. Men weet nu dat de lengte van de telefoongesprekken een exponentiële verdeling volgt, met een verwachtingswaarde van 3 minuten.
    1. Bereken de kans dat 1 telefoongesprek minder dan 4 minuten duurt.
    2. Bereken de kans dat je in een bepaalde week met 30 telefoongesprekken te voeren, langer dan 2uur getelefoneerd hebt.
4. Wat is de kans dat niemand in een groep van 25 personen dezelfde verjaardag heeft.
5. Stel dat 8 kinderen leren lezen aan de hand van Methode A en dat 10 andere kinderen leren lezen met methode B. Na verloop van tijd gaat men beide groepen een leesvaardigheidstest afnemen. Bij de A-group registreert men volgende uitslagen: 12,18,16,16,15,17,14,18 De gemiddelde score bij de B-groep is 16.8 met een variantie van 6.62. Mag men besluiten dat de score op de leesvaardigheidstest significant hoger is bij methode B (a = 5%)?