EXAMEN STATISTIEK EN WISKUNDIGE DATA-ANALYSE (1ste zit '19-'20, reeks A)

Nr.:

Opleiding industrieel ingenieur
T. Van Hecke

FACULTEIT INGENIEURSWETENSCHAPPEN EN ARCHITECTUUR

Naam : Richting: /20

Schrijf netjes. Vul in op de opengelaten plaatsen.
Eenvoudig rekenmachine toegelaten. Geen gsm, smartphone, ...
Geef uitleg bij de open vragen. Veel succes!

UNIVERSITEIT GENT

1. Beschouw de variabele x die het gasverbruik (in m^3) voorstelt. Deze is N(b, 0.01 b) verdeeld. Men meet 100 keer het gasverbruik, onafhankelijk van elkaar. Benader zo goed mogelijk de kans dat in meer dan 10 gevallen het gasverbruik meer dan (0.015 b) afwijkt van b. /4

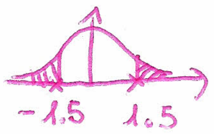
i : Bin$(100, p)$, met i = # gevallen waar $|x-b| > 0.015b$
en x is $N(b, 0.01b)$

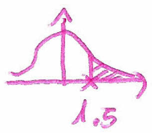
$p=?$ $p = P(|x-b|>0.015b) = P(x-b>0.015b) + P(x-b<-0.015b)$

$$\underset{x \text{ is } N(b,0.01b)}{=} P\Big(\underbrace{\frac{x-b}{0.01b}}_{z\in N(0,1)} > \frac{0.015b}{0.01b}\Big) + P\Big(\frac{x-b}{0.01b} < \frac{-0.015b}{0.01b}\Big)$$

wegens sym.

$$= P(z>1.5) + P(z<-1.5) \overset{\downarrow}{=} 2P(z>1.5) = 2(0.5-0.4332) = 0.1336$$

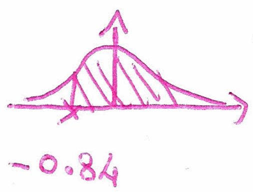




i is Bin$(100, p)$ met $np = 13.36 > 5$ en $n(1-p) = 86.64 > 5$
$\Rightarrow$ i is goed benaderd door de normale verdeling $\Rightarrow$
i is benaderd door y : $N(np, \sqrt{np(1-p)}) = N(13.36, \sqrt{11.57})$

Gevraagd: $P(i > 10) \Rightarrow$ benaderd door: $P(y \geq 10.5) = P\Big(\overbrace{\frac{y-13.36}{\sqrt{11.57}}}^{z\in N(0,1)} \geq \frac{10.5}{\sqrt{11.57}}\Big)$

$$= P(z \geq -0.84) = 0.5 + 0.2996 = 0.4996 \approx 80\%$$



2. Machines A en B produceren respectievelijk 10% en 90% van de totale productie aan lagers.

(a) Wat is het 90% betrouwbaarheidsinterval voor de kans dat een lager geproduceerd door machine A niet aan de kwaliteitseisen voldoet als je vastgesteld hebt dat bij een steekproef van 250 lagers geproduceerd door machine A er 10 defect waren. /6

p_A = kans op defect bij machine A $\quad \hat{p}_A = \frac{1}{25} = 0.04$; 90% BI voor p_A

De populatie is binomiaal verdeeld en $n\hat{p}_A = 250 \cdot \frac{1}{25} > 5$ en $n(1-\hat{p}_A) = 240 > 5$

$\Rightarrow$ de toetsingsgrootheid is $z_{ber} = \dfrac{p_A - \hat{p}_A}{\sqrt{\dfrac{\hat{p}_A(1-\hat{p}_A)}{250}}}$

$$1-\alpha = 0.90 = P(-z_{0.45} < z_{ber} < z_{0.45})$$
$$= P(-1.645 < z_{ber} < 1.645)$$

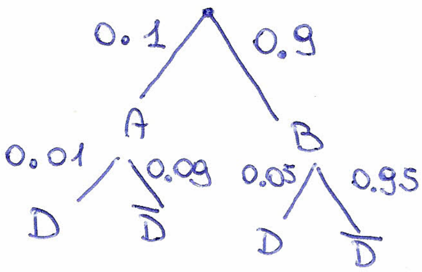
$\Rightarrow$ 90% BI voor p_A: $\left[\frac{1}{25} - \frac{1.645}{5}\sqrt{\frac{24}{25\cdot(250)}}, \frac{1}{25} + \frac{1.645}{5}\sqrt{\frac{24}{25(250)}}\right]$

$= [0.0196, 0.0604]$

(b) Stel dat je weet dat de kans dat een lager geproduceerd met B niet aan de kwaliteitseisen voldoet, gelijk is aan 0.05 en dat die kans voor A gelijk is aan 0.01, dit alles los van de steekproeven. Een willekeurige lager wordt getest en blijkt niet aan de kwaliteitseisen te voldoen. Wat is de kans dat deze lager geproduceerd werd door machine A?

D = kans op defect ; $P(A) = 0.1$; $P(B) = 0.9$

$P(D/A) = 0.01$; $P(D/B) = 0.05$



$$P(A/D) = \frac{P(A \wedge D)}{P(D)} = \frac{P(A)P(D/A)}{P(A)P(D/A) + P(B)P(D/B)}$$

$$= \frac{0.1(0.01)}{0.1(0.01) + 0.9(0.05)}$$

$$= \frac{0.1}{0.1 + 0.5} = \frac{1}{46} \approx 0.0217$$

3. Vul in onderstaande tabel met een hoofdletter de juiste oplossing aan. Er is telkens exact 1 juiste oplossing. Duid bij elke vraag een antwoord aan want standard setting wordt toegepast bij de evaluatie. /10

vraag	1	2	3	4	5	6	7	8
antwoord	A	B	B	A	B	C	C	C

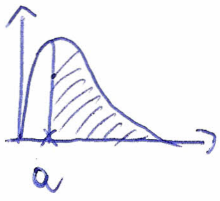
(1) Als $x : F(9, 10 \text{ d.f.})$, wat is a indien $P(x > a) = 90\%$?

A. 0.413

B. 2.42

C. 2.35

D. 0.426



$a = F_{0.9}(9,10) = \dfrac{1}{F_{0.1}(10,9)} = \dfrac{1}{2.42} = 0.413$

(2) Bij loononderhandelingen van voetbalspelers zijn volgende gegevens een basis: naast een handvol spelers die bijna 3 miljoen euro verdienen, verdienen de meeste spelers tussen de 100 000 euro en 150 000 euro. Met welke maat zullen de niet-topspelers het best een hoog loon onderhandelen bij de ploeg?

A. mediaan

B. gemiddelde

C. variantie

D. het maakt niet uit: mediaan, variantie of gemiddelde

d

(3) Als de momentenfunctie van een discrete variabele i gelijk is aan $M(t) = \dfrac{1}{1-3t}$, dan is

A. $\mu = 1$ en $\sigma^2 = 9$

B. $\mu = 3$ en $\sigma^2 = 9$

C. $\mu = 1$ en $\sigma^2 = 18$

D. $\mu = 3$ en $\sigma^2 = 18$

$\dfrac{dM(t)}{dt} = \dfrac{-(-3)}{(1-3t)^2} \Rightarrow \mu = \left(\dfrac{dM(t)}{dt}\right)_{t=0} = 3$

$\dfrac{d^2M(t)}{dt^2} = \dfrac{3(-2)(-3)}{(1-3t)^3} ; \left(\dfrac{d^2M(t)}{dt}\right)_{t=0} = 18 \Rightarrow \sigma^2 = 18 - 3^2 = 9$

(4) Welke uitspraak is waar bij de lineaire regressie $y = \beta_0 + \beta_1 x$ op basis van meetpunten (x_i, y_i) met $i \in \{1, 2, \ldots, n\}$?

A. De schattingen voor β_0 en β_1 zorgen ervoor dat SSE minimaal wordt.

B. De schattingen voor β_0 en β_1 zorgen ervoor dat SST minimaal wordt.

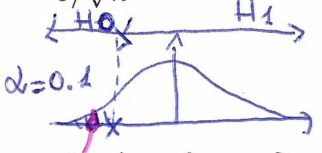
C. De schattingen voor β_0 en β_1 zorgen ervoor dat SSR minimaal wordt.

D. De schattingen voor β_0 en β_1 zorgen ervoor dat $\sum_{i=1}^{n}(x_i - \bar{x})^2$ minimaal wordt.

(5) Bij het testen van $H_0 : \mu = 10$ (met $H_1 : \mu < 10$) op basis van een steekproef met grootte $n = 21$ en $t_{ber} = \dfrac{\overline{x} - \mu}{s/\sqrt{n}}$, verwerpen we de nulhypothese met 90% betrouwbaarheid als

links

A. $t_{ber} > 1.32$
B. $t_{ber} < -1.32$
C. $t_{ber} < -1.72$
D. $t_{ber} > 1.72$

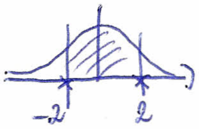


$-t_{0.9}(20\,df) = -1.32$

t_{ber}

(6) Wat is het percentage aan steekproefwaarden dat je kan verwachten dat hoogstens $2\,\sigma$ van μ zal gelegen zijn bij gegevens uit een normale verdeling?

A. 25%
B. 75%
C. 95.44%
D. hangt af van de grootte van de steekproef



$z \in N(0,1)$

$P(|x-\mu| < 2\sigma) = P\left(\left|\frac{x-\mu}{\sigma}\right| < 2\right)$

$P(|z| < 2) = 2(0.4772)$

$= 0.9544$

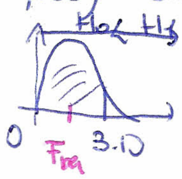
(7) Bij one-way anova waarbij de gemiddeldes van 4 groepen met elkaar vergeleken wordt (6 steekproefwaarden per groep) en berekend werd dat $F_{ber} = 2.9$, zal met 95% betrouwbaarheid aanvaard worden dat $\mu_1 = \mu_2 = \mu_3 = \mu_4$ (H_0)

A. omdat $F_{ber} \in$ aanvaardingsgebied $= [0,\ 3.16]$.
B. omdat $F_{ber} \in$ aanvaardingsgebied $= [0.07,\ 3.86]$.
C. omdat $F_{ber} \in$ aanvaardingsgebied $= [0,\ 3.10]$.
D. omdat $F_{ber} \in$ aanvaardingsgebied $= [0.07,\ 4.08]$.

$k = 4$

$m = 6$

$F(3, 20) = 3.10$



(8) Beschouw de onafhankelijke variabelen:

$$u : \chi^2(9 \text{ d.f.}) \qquad v : N(0,1) \qquad w : N(2,\ 3)$$

Zoek a zodat $P(9\,v^2 + w^2 + 9\,u > 4\,w + a) = 75\%$

A. $a = 98.60$
B. $a = 7.58$
C. $a = 64.22$
D. $a = 49.10$

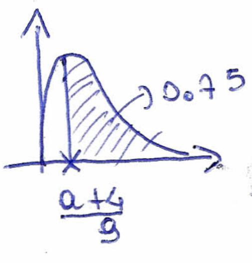
w is $N(2,3) \Rightarrow \frac{w-2}{3}$ is $N(0,1) \Rightarrow \frac{(w-2)^2}{9}$ is $\chi^2(1\ df)$

v is $N(0,1) \Rightarrow v^2$ is $\chi^2(1\,df)$

$\Rightarrow u + v^2 + \frac{(w-2)^2}{9}$ is $\chi^2(\underbrace{1+1+9}_{11}\ df)$

$0.75 = P(9v^2 + w^2 + 9u > 4w + a) = P\left(\frac{9v^2 + w^2 - 4w + 4 + 9u}{9} > \frac{a+4}{9}\right)$

$\Leftrightarrow 0.75 = P\left(v^2 + \frac{(w-2)^2}{9} + u > \frac{a+4}{9}\right)$



$\Rightarrow \frac{a+4}{9} = \chi^2_{0.25}(11\,df) = 7.58$

$\Leftrightarrow a = 64.22$