UiO : Universitetet i Oslo

KANDIDAT

15536

PRØVE

STK1000 1 Innføring i anvendt statistikk

Emnekode	STK1000
Vurderingsform	Hjemmeeksamen
Starttid	25.11.2020 14:00
Sluttid	25.11.2020 18:30
Sensurfrist	--
PDF opprettet	09.11.2021 23:19

1a $\mu = 124,5$

$\sigma = 4,8$

$$P(x > 128) = 1 - P(x < 128)$$
$$= 1 - P\left(Z < \frac{128 - 124.5}{4.8}\right)$$
$$= 1 - P(Z < 0,729)$$
$$= 1 - 0,7673$$
$$= \underline{\underline{0,233}}$$

SVAR: Omtrent 23% sjanse at 7-åring er over 128 cm

1b Velger genser slik at 90% passer.
Min. 10% er for høye.
gir $z = 1.28$

$$z = \frac{x - \mu}{\sigma}$$

$$x = z \cdot \sigma + \mu$$

$$= 1,28 \cdot 4,8 + 124,5$$

$$\approx 130,6$$

SVAR: Velger genser str 134.

2 a Vi har ikke forkunnskap som tilsier ensidig test. k er konsentrasjon i væske

$H_0: k = 0.6$ $\quad$ $H_a: k \neq 0.6$

signifikansnivå $\alpha = 0,05$

2b $\sigma = 0,022$ Vi kjenner til konsentrasjonens sanne σ, og vet at måleresultat er N-fordelt.

test observator:

$\bar{x} = \frac{0.58 + 0.56}{2} = 0.57$

$\bar{x}$ er gj.-resultat

μ_0 er verdi til nullhypotese

$$z = \frac{\bar{x} - \mu_0}{\sigma / \sqrt{n}}$$

$$= \frac{0.57 - 0.6}{0,022 / \sqrt{n}}$$

$$= -1.928$$

p-verdi: areal til venstre for $z = -1,928$ er 0.0268. To-sidig test gir

$p = 2 \cdot 0,0268 = 0,0536$

Konklusjon:

p-verdi $> \alpha$, vi beholder H_0 til fordel for H_a. på 5% signifikansnivå (og 1%).
En-sidig test kunne gitt annen konklusjon v/ siginifikansnivå 5% fordi p-verdi da ville vært 0,027.

2c 95% KI gir $z^* = 1.96$

$$0.6 \pm z^* \cdot \sigma$$
$$= 0.6 \pm 1.96 \cdot 0{,}022$$
$$= (0.56\ ,\ 0.64)$$

Vi er 95% sikre på at forventningsverdi til en prøve ligger i intervallet.
Det stemmer med konklusjonen i oppg b fordi 0.6 ligger i intervallet.

3a modell: $y_i = \beta_0 + \beta_1 x_i + \epsilon$

estimat for...

$\beta_0 : b_0 = 7{,}4225$

$\beta_1 : b_1 = 9{,}7636$

$\sigma : s = 4{,}601$

b_0 estimerer tannlengden om β_1 ikke har effekt. (altså er b_0 konstanten)

b_1 estimerer økning i tannlengden y ved en enhets økning av x. (altså er b_1 stigningstallet)

s estimerer standardavvik til residualene (residual standard error, fra R)

$R^2 \approx 0.64$ betyr at 64% av forklart variasjon i tannlengde kommer av regresjonsmodellen. $0 < R^2 < 1$, og 0.64 er et godt grunnlag for at responsvariabelen kommer av forklaringsvariabelen.

3b $H_0: \beta_1 = 0$ $H_a: \beta_1 \neq 0$

Hypotesetesten sjekker om β_1 er signifikant, noe den må være for at modellen er riktig. Signif. nivå: 0,05 (standard)

Observator:

$$t = \frac{\beta_1 - 0}{SE_{b_1}} = \frac{b_1 - 0}{SE_{b_1}} = \frac{9,7636}{0,9525} = 10,25$$

med $SE_{b_1} = 0,9525$

$df = n-p-1$
$= 60-2 = 58$
grader av frihet

p-verdi fra R:
$1,23 \cdot 10^{-14}$

Konklusjon om signifikansen til stigningstallet:

p-verdi tilnærmet 0 gir grunnlag for å forkaste H_0 til fordel for H_a.
Den er signifikant på alle rimelige nivå (spesielt $\alpha = 0,05$)
Altså er $\beta_1 \neq 0$ og C-vitamin er relevant for tannlengden.

3e estimert modell:

tannlengde(C-dose) = 7,42 + 9,76 · Cdose
tannlengde(2) = 7,42 + 9,76 · 2
$= \underline{\underline{26,95}}$

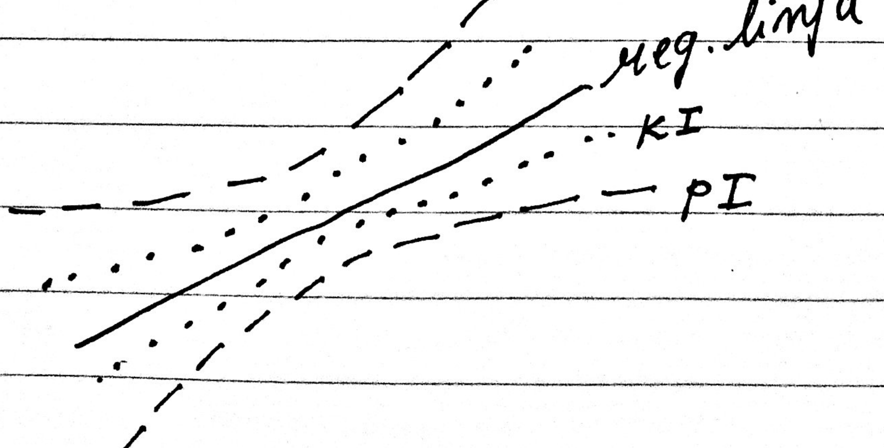
3d 95% KI v/ 2mg dose:
(lwr, upr)
= (25.0 , 28.9)

95% PI v/ 2mg :
(lwr, upr)
= (17.5 , 36.4)

<u>Konfidensintervall</u> er et predikert område hvor linja til regresjonsuttrykket kan havne. Den tar ikke høyde for en eventuell "neste" enkel observasjon.

<u>Prediksjonsintervall</u> er et predikert område, hvor vi kan tenke at en "neste" enkel observasjon har 95% sjanse for å havne i.
Intervallet tar høyde for ϵ_i i modellen, altså usikkerheten til linja, og enkle utfall (variasjonen til linja.)

KI er derfor mindre enn PI, og skal alltid få plass i PI.



4a logistisk modell:

$$\log\left(\frac{p}{1-p}\right) = \beta_0 + \beta_1 x_{i1} + \beta_2 x_{i2} + \beta_3 x_{i3}$$

"where p is a binomial proportion"
(fra kap 14 side 6)

4b 95% KI for odds-ratio e^{b_1}:

$$e^{b_1 \pm z^* SE_{b_1}}$$ med z^* for 95% KI = 1.96

$$\left.\begin{aligned} SE_{b_1} &= 0{,}09426 \\ b_1 &= 0{,}2476 \end{aligned}\right\}\text{fra R}$$

$$= e^{0{,}2476 \pm 1.96 \cdot 0{,}0943}$$

$$\doteq (1.065 \;,\; 1.541)$$

Diverse Z - verdier brukt i oppgaven, funnet med nettsiden
https://www.mathsisfun.com/data/standard-normal-distribution-table.html