6.53 Hva er galt?

a) En signifikanstest forkaster $H_0: \bar{x} = 500$

Galt. En hypotese handler om en parameter μ, ikke $\bar{x}$.

b) Testutviklere vil finne ut om deres studenter scorer bedre enn det nasjonale gjennomsnittet (forventningsverdien) 21.2.
De formulerer $H_0: \mu > 21.2$

Galt. Dette er påstanden de ønsker å undersøke, altså H_a

c) En studie oppsummeres med at resultatet er statistisk signifikant, og at p-verdien er 0.98.

Galt: En p-verdi er sannsynligheten for å observere minst like ekstreme observasjoner som du har observert, gitt H_0.
En høy p-verdi vil altså si at observasjonene er "vanlige" under H_0-betingelser, og da beholdes H_0. Da er det ikke signifikante sammenhenger.

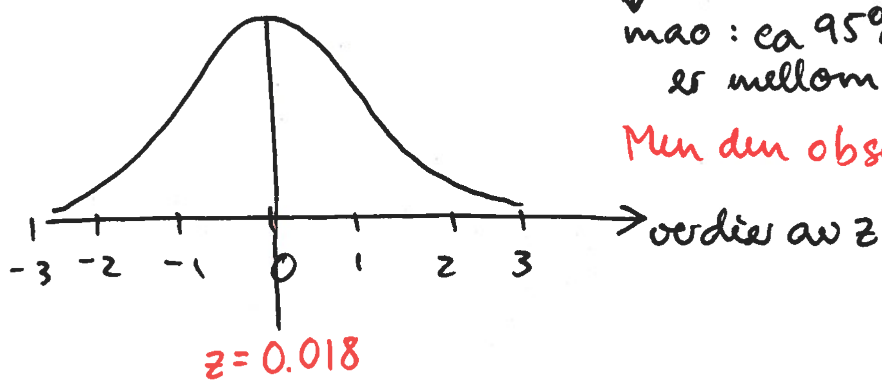
d) testobservatoren $z = 0.018$. Fordi dette er mindre enn $\alpha = 0.05$, forkastes H_0.

Galt. Testobservatoren viser verdien som må sammenlignes med andre verdier i den tilhørende fordelingen, ~~for~~ å få p-verdien, og så er det p-verdien som sammenlignes med den forhåndsvalgte α.

Her: Hvis H_0 er sann, er $z \sim N(0,1)$:

mao: ca 95% av verdiene er mellom −2 og 2.

Men den observerte $z = 0.018$



Da må vi slå opp i Table A for å finne

$$\text{p-verdi} = P(\text{Minst like ekstreme obs som det vi har obs} \mid H_0)$$
$$= P(z \geq 0.018 \mid z \sim N(0,1))$$
$$= 1 - P(z < 0.018) = 1 - 0.508 = \underline{0.492}$$

Dette tallet er større enn 0.05, og H_0 beholdes.

6.54

a) Ho : Andelen studenter med mobiltelefon, p, har ikke økt

$p \leq p_0$. $p \leq 0.88$

H_1 : Andelen studenter med mobiltelefon, p, har økt.

$p > p_0$. $p > 0.88$

b) Ho : Studenter som møter til tidlige klasser forventes å score like bra som de andre. Forventningsverdien : μ

$\mu \leq \mu_0 = 75$

H_1 : Studenter som møter til tidlige klasser forventes å score bedre :

$\mu > \mu_0 = 75$

c) Ho : Ingen bedring. Endring ≤ 0

H_1 : Bedring Endring > 0

Ensidige hypoteser

6.55

a) Forventet score : μ

Tosidig

Ho : Studenter med placement-test skiller seg ikke fra andre studenter : $\mu = \mu_0 = 77$

H_1 : Studenter med placement-test skiller seg fra de andre :

$\mu \neq \mu_0 = 77$

b) Forventet tid : μ ; Referanse : $\mu_0 = 20$

Ho : Mus med musikk er minst like raske : $\mu \geq \mu_0 = 20$

H_1 : Mus med musikk er tregere : $\mu < \mu_0$

Her er fasit og jeg uenig. Jeg mener at påstanden som testes er at de er tregere.

c) Ho : Forventet areal, μ, er $880\ ft^2$

Ensidig

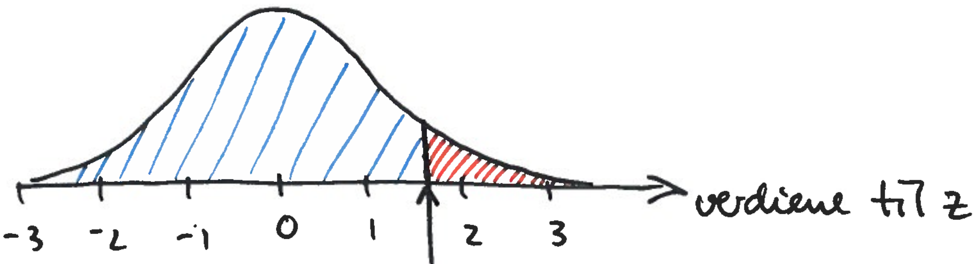
H_1 : $\mu < 880\ ft^2$

Prinsipp : H_1 er påstand.

Ho : Uskyldig inntil det motsatte er bevist.

6.58 $H_0 : \mu = \mu_0$

Man samler data, og bruker så testobservatoren z.
En z indikerer at test-observatoren ("statistic") er
N(0,1)-fordelt når H_0 er sann:



Vi får vite at $\boxed{z = 1.77}$

a) Hva er p-verdien hvis $H_a : \mu > \mu_0$?

P-verdien = P(minst like ekstreme observasjoner som det vi har observert | H_0)

OK: hvis H_0 er sann, bør verdiene til z være rundt 0, ikke for langt unna. H_a sier at verdiene til μ er store, altså er "enda mer ekstremt enn 1.77" det samme som enda større verdier enn 1.77:

$$\text{p-verdien} = P(z \geq 1.77) = 1 - P(z < 1.77) \underset{\uparrow}{=} 1 - 0.9616 = \underline{0.04}$$

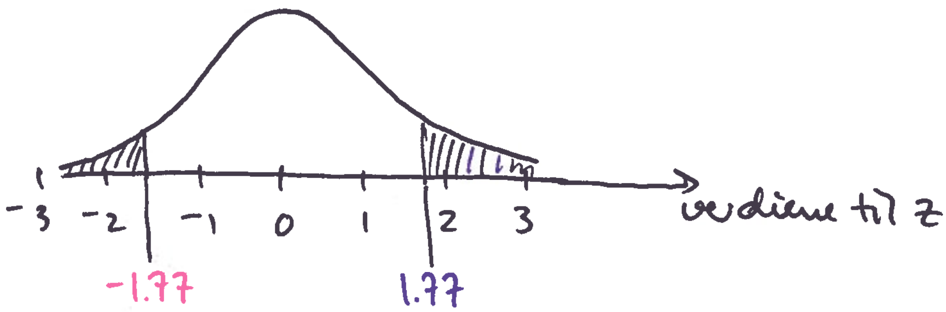
Table A

b) $H_a : \mu < \mu_0$

$$\text{p-verdien} : P(z \leq 1.77) = \underline{0.9616}$$

c) $H_a : \mu \neq \mu_0$

$$\text{p-verdien} = P(|z| \geq 1.77) = \underline{P(z \leq -1.77)} + \underline{P(z \geq 1.77)}$$
$$= 0.0384 + 0.0384 \approx \underline{0.08}$$



<u>6.73</u> Differanser mellom det drivstoff-forbruket produsentene oppgir, og det bileierne selv måler: D_i

Forventet forskjell: μ

Antar at standardavviket til D_i: $\sigma_D = 3.0$

$H_0: \mu = 0$

$H_1: \mu \neq 0$

Bruker gjennomsnittet for differansene som test-observator

$\bar{D} = \frac{1}{n}\sum_{i=1}^{n} D_i$

Fordelingen til D_i-ene $\overset{\text{tiln}}{\sim}$ Normal

σ kjent

$$\Rightarrow \bar{D} \sim N\left(\mu, \frac{\sigma_D}{\sqrt{n}}\right) \qquad \frac{3.0}{\sqrt{20}} \approx 0.67$$

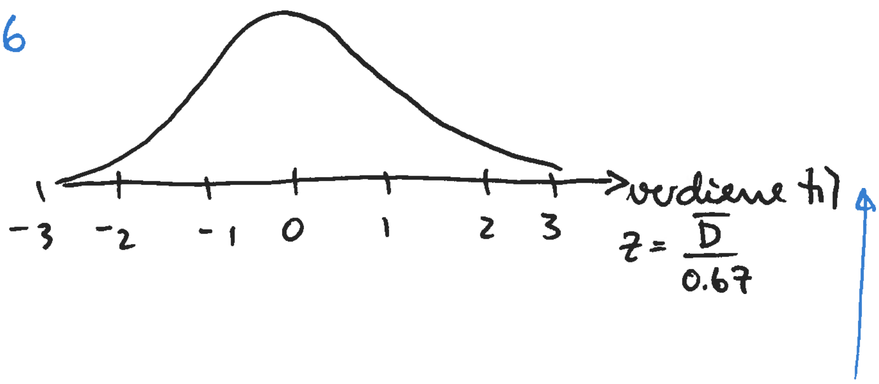
$$\downarrow$$

$$\frac{\bar{D} - \mu}{0.67} \sim N(0,1)$$

Hvis H_0 er sann: $\mu = 0$

$$z = \frac{\bar{D}}{0.67} \sim N(0,1)$$

$\bar{D} = 5 \Rightarrow z = \frac{5}{0.67} = 7.46$



p-verdi: $P(\text{minst like ekstreme obs som det vi har obs} \mid H_0)$

$$= P(z \geq 7.46 \mid z \sim N(0,1)) \approx 0$$

$\Rightarrow$ <u>Forkast H_0!</u>

6.83 signifikant på 1% nivå:

$$p\text{-verdi} = P\left(\begin{matrix}\text{minst like ekstreme obs} \\ \text{som det vi har obs}\end{matrix} \middle| H_0\right) \leq 0.01$$

Vi har observert forskjeller/sammenhenger/endringer som er så ekstreme at det er mindre enn 1% sannsynlighet for at slike observasjoner gjøres under H_0

↑

at slike forskjeller/sammenhenger/endringer oppstår ved en tilfeldighet, når det observeres egentlig ikke er noen forskjell/sammenheng/endring

, og da er det også mindre enn 5% sannsynlighet for det.

Obs: vi snakker alltid om sannsynligheten for observasjonene/å observere, aldri om ~~"sannsynligheten for at H_0 (eller H_1) er sann."~~

6.84 Signifikant på 5% nivå betyr ikke nødvendigvis signifikant på 1% nivå, bare at sh er ≤ 0.05

6.87 Ensidig hypotesetest

z-verdi 0.54

Bruk Table D for alternativet Ha "større enn"

? z^* 0.674 0.841

50% 60%

Confidence level C

→ p-verdi 0.5

Table A :

6.86 ? utgår sammen med 6.87

6.94 Review (gjennomgang) av forskningen på antidepressiva:

74 studier — 38 studier med positive resultater → 37 publiseres

74 studier — 36 studier med negative resultater → 22 publiseres ikke

36 studier med negative resultater → 11 publiseres, men konklusjonen "vris"

Dette kalles publiseringsbias/publiseringsskjevhet, og gir et feil bilde av effekten av antidepressiva.

Regler for publisering av RCT-studier: CONSORT statement skal forhindre dette.

6.95 En signifikanstest (hypotesetest) kan

a) <u>ikke</u> gi svar på om designet i studien er OK.

b) gi svar på om den observerte sammenhengen/effekten samsvarer med H_0 eller ikke.

c) <u>ikke</u> gi svar på om forskjellen er viktig.

Man kan ha forskjeller som ikke er statistisk signifikante, men der størrelsen på effekten ville vært interessant om den var det. Løsning: Ny studie med flere deltakere (eller en metaanalyse).

Man kan også ha forskjeller som er statistisk signifikante, men der størrelsen på effekten er så liten at den i enkelte situasjoner er uinteressant. (Kan være uinteressant på individnivå, men av betydning på populasjonsnivå, f. eks økonomisk)

6.106 Du utfører 1000 signifikanstester med $\alpha = 0.05$.

$\alpha = 0.05$ betyr

P(type I-feil) ≤ 0.05 i én test

Det er maksimalt 5% sannsynlighet for å observere noe som gjør at Ho forkastes, hvis Ho er sann.

Hvor mange signifikante tester forventer vi av de 1000:

Dette er en binomisk situasjon:

n uavhengige forsøk med
to mulige utfall: Forkast Ho / Behold Ho, og
P(Forkaste Ho) = 0.05 i hvert forsøk.

Da er X = antall signifikante tester av de 1000

$X \sim \text{bin}(1000, 0.05)$

Forventet antall signifikante tester:

$$E(X) = np = 1000 \cdot 0.05 = \underline{50}$$

OBS: Selv om Ho er sann i alle tilfellene!

jfr Big data

Gen-testing