

Plan:

- ① Hypotesetest: populasjonsforventning μ
- ② Hypotesetest: populasjonsandel p
- ③ Korrelasjonskoeffisient i datasett / utvalg

① Hypotesetest for populasjonsforventning μ

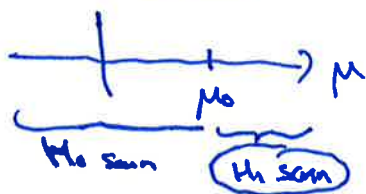
Antar: X_1, X_2, \dots, X_n uavhengige normalfordelte variabler med $E(x_i) = \mu$ og $\text{Var}(x_i) = \sigma^2$.

a) Hypoteser: $\begin{cases} H_0: \text{nullhypotese} \\ H_1: \text{alternativ hypotese} \end{cases}$

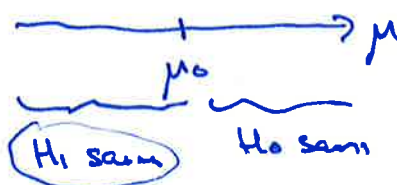
- komplementære
- bevisbyrden ligger på H_1

The types:

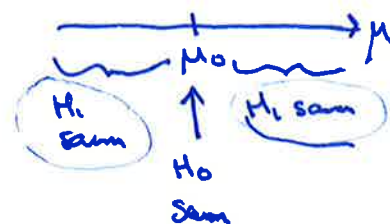
i) $H_0: \mu \leq \mu_0$
 $H_1: \mu > \mu_0$
høyresidig



ii) $H_0: \mu \geq \mu_0$
 $H_1: \mu < \mu_0$
venstresidig



iii) $H_0: \mu = \mu_0$
 $H_1: \mu \neq \mu_0$
tosidig



b) Mål: Beslutning $\left\{ \begin{array}{l} - \text{forkaste } H_0 \\ - \text{beholde } H_0 \end{array} \right.$

	<u>H_0 sann</u>	H_1 sann
beholde H_0	✓	Type II feil
forkaste H_0	Type I feil	✓

Signifikansnivå α :

max.
 $\alpha =$ Sannsynligheten for å gjøre type I feil

c) Testobservator: t-test / z-test

* Vi bruker $T = \frac{\bar{X} - \mu_0}{S/\sqrt{n}}$ når σ er ukjent
 (t-fordelt med $n-1$ frihetsgrader når $\mu = \mu_0$)

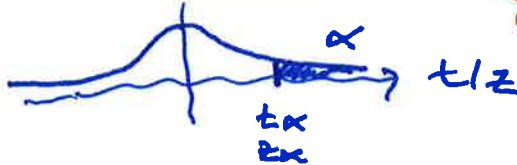
* Vi bruker $Z = \frac{\bar{X} - \mu_0}{\sigma/\sqrt{n}}$ når σ er kjent
 (stet normalfordelt når $\mu = \mu_0$)

d) Forkastningsområde:

Eks: z-test med $\alpha = 0.05$

$Z_{\alpha} = z_{0.05} = 1.645$

i) $H_0: \mu \leq \mu_0$
 $H_1: \mu > \mu_0$
 høyresidig



Hush:
 $P(Z > z_{\alpha}) = \alpha$
 \Leftrightarrow
 $P(Z \leq z_{\alpha}) = 1 - \alpha$
 (kalle)

Forkastningsområde:
 $Z > 1.645$

Forkastningsområde:

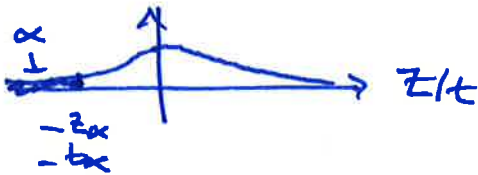
$T > t_{\alpha}$ eller $Z > z_{\alpha}$

$\left\{ \begin{array}{l} \text{Forkaster } H_0 \text{ om } T > t_{\alpha} / Z > z_{\alpha} \\ \text{Beholder } H_0 \text{ ellers} \end{array} \right.$

ii) Venstresidig

$$H_0: \mu \geq \mu_0$$

$$H_1: \mu < \mu_0$$

Kritisk verdi:

$$z_{1-\alpha} = -z_\alpha$$

$$t_{1-\alpha} = -t_\alpha$$

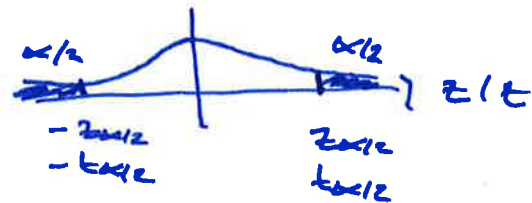
Forkastningsområde:

$$T < -t_\alpha \quad / \quad Z < -z_\alpha$$

iii) Tvåsidig:

$$H_0: \mu = \mu_0$$

$$H_1: \mu \neq \mu_0$$

Forkastningsområde:

$$|T| > t_{\alpha/2} \quad \text{eller}$$

$$|Z| > z_{\alpha/2}$$

Vi kan beskrive forkastningsområdet uka \bar{X} (i stedet for $Z(T)$)

Ekse: høyreidig Z -test
med $\alpha = 0.05$

anta: $\sigma = 10$
 $n = 25$
 $\mu_0 = 14$

$$Z > z_{\alpha} = 1.645$$

"

$$\frac{\bar{X} - \mu_0}{\sigma/\sqrt{n}} > 1.645$$

$$\bar{X} - \mu_0 > 1.645 \cdot \frac{10}{\sqrt{25}}$$

$$\bar{X} > \mu_0 + 1.645 \cdot \frac{10}{\sqrt{25}}$$

$$\bar{X} > 14 + 1.645 \cdot \frac{10}{\sqrt{25}}$$

$$Z > 1.645$$



$$\bar{X} > 17.29$$

Når vi skal utføre hypotesetesten og konkludere:

Alternativ 1: Hovedmetode

Bestem ut testobservator og sammenlikn med forkastningsområde.

Det er ofte flere mulige valg av testobservator.
For eksempel kan man bruke T eller \bar{X} i en t -test.

Alternativ 2: p-verdi

Definisjon: p-verdi til en hypotesetest:

sannsynligheten for å finne en verdi på testobservatoren som er mindre like elstær som den observerte verdien

Eksempel:
 $H_0: \mu \leq 14$
 $H_1: \mu > 14$
 (høyresidig test)

Z-test med $\sigma = 10$
 $n = 25$

$$\bar{X} = 18.3 \Rightarrow Z = \frac{18.3 - 14}{10 / \sqrt{25}} = 2.15$$

p-verdi:

$$p = P(Z \geq 2.15) = 1 - P(Z < 2.15)$$

$$= 1 - \Phi(2.15)$$

$$= \underline{\underline{0.016}} = 1.6\%$$

Konklusjon:

$p < \alpha$: Forkast H_0
 $p \geq \alpha$: Behold H_0

e) Utfør hypotesetesten: $\left\{ \begin{array}{c} Z \\ T \\ \bar{X} \end{array} \right\}$ rekn ut, basert på data fra utvalget, og sammenlign med forkastingsområdet

testobservator \rightarrow

Hvis vi har forkastingsområde: Forkast H_0
 Hvis vi ikke har: — | — : Behold H_0 .

Alternativ: $p < \alpha$: Forkast H_0
 $(p\text{-verdi})$ $p \geq \alpha$: Behold H_0

f) Styrkefunksjon:

Sannsynligheten for å forkaste H_0 , gitt en verdi av den ubekjente parameteren μ .

gamma " styrkefunksjon $\rightarrow \delta(\mu) = P(\text{forkaster } H_0 \mid \mu)$

Bes: høyresidig z-test på $\alpha = 0.05$. $\begin{cases} H_0: \mu \leq 14 \\ H_1: \mu > 14 \end{cases}$ $\sigma = 10$
 $n = 25$

Forkast H_0 : $Z > 1.645 \iff \bar{X} > 17.29$

Anta $\mu = 15$: $\delta(15) = P(\bar{X} > 17.29 \mid \mu = 15)$

(burde forkaste H_0)

$$= P\left(Z = \frac{\bar{X} - \mu}{\sigma/\sqrt{n}} > \frac{17.29 - \mu}{\sigma/\sqrt{n}} \mid \mu = 15\right)$$

$$= P\left(Z > \frac{17.29 - 15}{10/\sqrt{25}}\right) = P(Z > 1.145)$$

$$= 1 - \Phi(1.145) = 1 - 0.874 = \underline{\underline{0.126}}$$

"Godtakingsfeil" : 87.4%
Type II feil av tilfellene
 " $1 - \delta(\mu)$ "

Vi kan redusere godhetsskall (type II feil) uten å endre α ved å øke n .

② Hypotesebtest for populasjonsandel p

Anta at X er binomial med parameter n, p , der p er ukjent.

$$\hat{p} = \frac{X}{n} \quad (\text{estimator for } p)$$

a) Hypotese:

$$\begin{aligned} \text{i) } H_0: p &\leq p_0 \\ H_1: p &> p_0 \end{aligned}$$

(høyresidig)

$$\begin{aligned} \text{ii) } H_0: p &\geq p_0 \\ H_1: p &< p_0 \end{aligned}$$

(venstresidig)

$$\begin{aligned} \text{iii) } H_0: p &= p_0 \\ H_1: p &\neq p_0 \end{aligned}$$

(tosidig)

b) Beslutning: — beholde H_0
— forkaste H_0

Signifikansnivå α :
maks sannsynlighet for type I feil som aksepteres

c) Testobservator:

$$Z = \frac{\hat{p} - p_0}{\sqrt{\frac{p_0(1-p_0)}{n}}}$$

— anta n er så stor at X er tilnærmet normal fordelt

$$\text{Var}(X) = np \cdot (1-p)$$

$$\text{Var}(\hat{p}) = \frac{\text{Var}(X)}{n^2} = \frac{np(1-p)}{n^2} = \frac{p(1-p)}{n}$$

d) Forlesthinsområde:

$$\text{i) } Z > z_{\alpha}$$



$$\text{ii) } Z < -z_{\alpha}$$



$$\text{iii) } |Z| > z_{\alpha/2}$$



Ekse: Andelen laks angrepet av lakeet
ligger på 30%. Vil undersøke om et
anlegg her

- i) mer enn 30% lakeet
- ii) ulik 30% lakeet

Av 25 laks som undersøkes, har 11
lakeet.

$$\begin{aligned} \text{i) } H_0: p &\leq 0.30 \\ H_1: p &> 0.30 = p_0 \\ &\text{(nøyresidus)} \end{aligned}$$

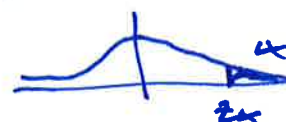
Er n stor nok?

$$\begin{aligned} n \cdot p(1-p) &\geq 5 \\ n \cdot p \cdot (1-p) &= 25 \cdot 0.3 \cdot 0.7 \\ &= 5.25 \quad \text{ok} \end{aligned}$$

$\alpha = 0.05$: signifikansnivå

$$Z = \frac{\hat{p} - p_0}{\sqrt{\frac{p_0(1-p_0)}{n}}} = \frac{n\hat{p} - np_0}{n\sqrt{\frac{p_0(1-p_0)}{n}}} = \frac{X - np_0}{\sqrt{np_0(1-p_0)}}$$

Forkastingsområde: $Z > z_{\alpha} = 1.645$

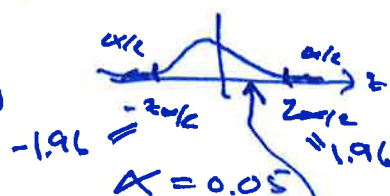


$$\boxed{Z > 1.645}: \frac{X - np_0}{\sqrt{np_0(1-p_0)}} > 1.645$$

$$\begin{aligned} X &> np_0 + 1.645 \cdot \sqrt{np_0(1-p_0)} \\ &= 7.5 + 1.645 \cdot 2.29 \end{aligned}$$

$$\boxed{X > 11.8} \rightarrow \text{Beholder } H_0$$

$$\begin{aligned} \text{ii) } H_0: p &= 0.30 \\ H_1: p &\neq 0.30 \\ &\text{(tosidig)} \end{aligned}$$



Fork. omr:

$$|Z| > 1.96$$

$$\rightarrow \text{Beholder } H_0 \quad \boxed{Z = 1.53}$$

③ Korrelasjonskoeffisient

Skal finne sammenhenger mellom to stokastiske variable X, Y .

Datasett fra utvalget:

X	Y
75	0.48
145	1.09
55	0.53
88	0.97
122	0.78

X = motorstørrelse
 Y = bensinförbruk

i) Utvalgs kovarians:

$$S_{XY} = \frac{1}{n-1} \sum_{i=1}^n (x_i - \bar{x})(y_i - \bar{y})$$

ii) Korrelasjonskoeff.:

$$R = \frac{S_{XY}}{S_X \cdot S_Y}$$

Tolkning av R:

- * $r > 0$ betyr positiv sammenheng
- $r < 0$ betyr negativ sammenheng

- * Vi har $-1 \leq r \leq 1$, og $|r|$ nært 1 betyr sterk sammenheng

På kalk:

1) C STAT

2) 75 INPUT 0.48 ↵
 145 " 1.09 "
 !
 122 " 0.97 "

$\bar{x} = 97$ [\bar{x}, \bar{y}] 13
 $\bar{y} = 0.77$ [\bar{x}, \bar{y}] SWAP
 $r = 0.783$ [\hat{x}, r] SWAP

Sterk positiv sammenheng.