Plan:

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

1 Hypotesetest for populasjonsforventning μ

Anta: $X_1, X_2, \dots, X_n$ uavhengige normalfordelte variabler med $E(X_i) = \mu$ og $Var(X_i) = \sigma^2$.

a) **Hypoteser:**

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

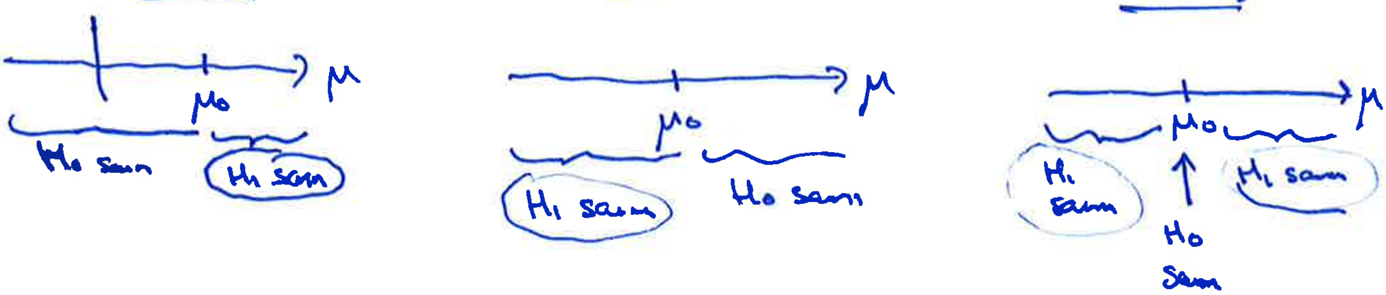
- komplementære
- bevisbyrden ligger på H_1

Tre typer:

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 $\begin{cases} - \text{forkaste } H_0 \\ - \text{beholde } H_0 \end{cases}$

	H_0 sann	H_1 sann
beholde H_0	V	Type II feil
forkaste H_0	Type I feil	V

Signifikansnivå α:

α = maks. sannsynlighet for å gjøre type I feil

c) Testobservator: t-test / z-test

* Vi bruker $T = \dfrac{\bar{X} - \mu_0}{S/\sqrt{n}}$ når σ er ukjent

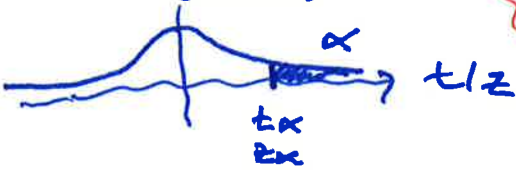
(t-fordelt med n-1 frihetsgrader når $\mu = \mu_0$)

* Vi bruker $Z = \dfrac{\bar{X} - \mu_0}{\sigma/\sqrt{n}}$ når σ er kjent

(std normalfordelt når $\mu = \mu_0$)

d) Forkastningsområde:

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



Husk:
z_α: $P(Z > z_\alpha) = \alpha$
⇕
$P(Z \leq z_\alpha) = 1 - \alpha$
tabell.

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

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

$\boxed{\begin{array}{c}\text{Forkastningsområde:} \\ Z > 1.645\end{array}}$

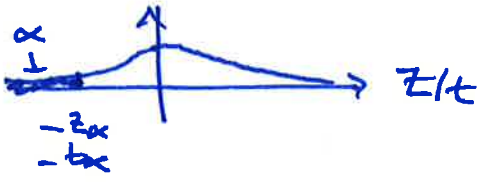
Forkastningsområde:

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

$\begin{cases} \text{Forkaster } H_0 \text{ om } T > t_\alpha \,/\, Z > z_\alpha \\ \text{Beholder } H_0 \text{ ellers} \end{cases}$

ii) Venstresidig:

$H_0: \mu \geq \mu_0$
$H_1: \mu < \mu_0$



Kritisk verdi:

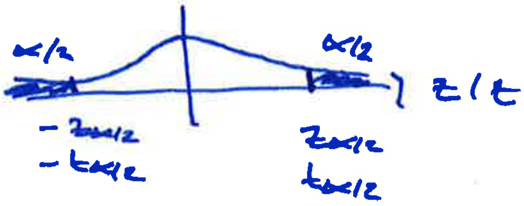
$z_{1-\alpha} = -z_\alpha$
$t_{1-\alpha} = -t_\alpha$

Forkastningsområde:

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

iii) Tosidig:

$H_0: \mu = \mu_0$
$H_1: \mu \neq \mu_0$



Forkastningsområde:

$|T| > t_{\alpha/2}$ eller
$|Z| > z_{\alpha/2}$

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

Eks: høyresidig Z-test med $\alpha = 0.05$: $Z > z_\alpha = 1.645$

antar: $\sigma = 10$, $n = 25$, $\mu_0 = 14$

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

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

$$\bar{X} > \mu_0 + 1.645 \cdot \frac{\sigma}{\sqrt{n}}$$

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

$$\boxed{Z > 1.645} \longleftrightarrow \boxed{\bar{X} > 17.29}$$

Når vi skal utføre hypotesetester og konkludere:

Alternativ 1: Hovedmetode

Regn 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

Def.: p-verdi til en hypotesetest: sannsynligheten for å finne en verdi på testobservatoren som er minst like ekstrem som den observerte verdien

Eks: $H_0: \mu \leq 14$
$H_1: \mu > 14$
(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: $p = P(Z \geq 2.15) = 1 - P(Z < 2.15)$
$= 1 - G(2.15)$
$= 0.016 = 1.6\%$

Konklusjon:

$$\boxed{\begin{array}{ll} p < \alpha: & \text{Forkast } H_0 \\ p \geq \alpha: & \text{Behold } H_0 \end{array}}$$

e) Utfør hypotesetesten: $\left\{ \begin{matrix} Z \\ T \\ \bar{X} \end{matrix} \right\}$ regn ut, basert på data fra utvalget, og sammenlikn med forkastningsområde

testobservator

Hvis vi havner i forkastn.område: Forkast Ho.
Hvis vi ikke havner: — " — : Behold Ho.

Alternativ: (p-verdi) $p < \alpha$: Forkast Ho
$p \geq \alpha$: Behold Ho

f) Styrkefunksjon:

Sannsynlighet for å forkaste Ho, gitt en verdi av den ukjente parameteren μ.

gamma " styrkefn. $\longrightarrow \gamma(\mu) = p(\text{forkaster Ho} \mid \mu)$

Eks: 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 Ho: $Z > 1.645 \Leftrightarrow \bar{X} > 17.29$

Anta $\mu = 15$: (burde forkaste Ho)

$$\gamma(15) = p(\bar{X} > 17.29 \mid \mu = 15)$$
$$= p\left(Z = \frac{\bar{X} - \mu}{\sigma / \sqrt{n}} > \frac{17.29 - \mu}{\sigma / \sqrt{n}} \,\middle|\, \mu = 15\right)$$
$$= p\left(Z > \frac{17.29 - 15}{10 / \sqrt{25}}\right) = p(Z > 1.145)$$
$$= 1 - G(1.145) = 1 - 0.874$$
$$= \underline{\underline{0.126}}$$

"Godtakingsfeil": 87.4% av tilfellene
Type II feil $1 \overset{"}{-} \gamma(\mu)$

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

② Hypotesetest for populasjonsandel p

Anta at X er binomisk med parametre n, p, der p er ukjent.

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

a) Hypoteser:

i) $H_0: p \leq p_0$, $H_1: p > p_0$ (høyresidig)

ii) $H_0: p \geq p_0$, $H_1: p < p_0$ (venstresidig)

iii) $H_0: p = p_0$, $H_1: p \neq p_0$ (tosidig)

b) Beslutning:
- beholde H_0
- forkaste H_0

Signifikansnivå α: maks sannsynlighet for type I feil som aksepteres

c) Testobservator:

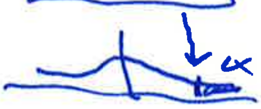
$$\boxed{Z = \frac{\hat{p} - p_0}{\sqrt{\frac{p_0(1-p_0)}{n}}}}$$

- antar n er så stor at X er tilnærmet normalfordelt
- $\mathrm{Var}(X) = np\cdot(1-p)$

$$\mathrm{Var}(\hat{p}) = \frac{\cancel{n}p(1-p)}{n^{\cancel{2}}} = \frac{p(1-p)}{n}, \qquad \hat{p} = \frac{X}{n}$$

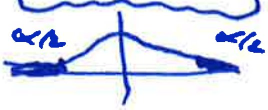
d) Forkastningsområde:

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



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

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



Eks: Andelen laks angrepet av lakselus ligger på 30%. Vil undersøke om et anlegg har

i) mer enn 30% lakselus

ii) ulik 30% lakselus

Av 25 laks som undersøkes, har 11 lakselus.

Er n stor nok?

$n \cdot p(1-p) \geq 5$

$n \cdot p \cdot (1-p) = 25 \cdot 0.3 \cdot 0.7 = 5.25$ ok.

i) $H_0: p \leq 0.30$

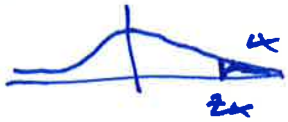
$H_1: p > 0.30 = p_0$

(høyresidig)

$\alpha = 0.05$: 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: $Z > z_\alpha = 1.645$



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

$X > np_0 + 1.645 \cdot \sqrt{np_0(1-p_0)}$

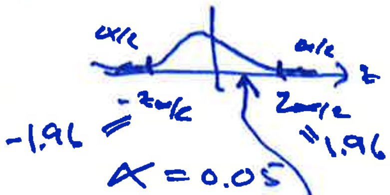
$= 7.5 + 1.645 \cdot 2.29$

$\boxed{X > 11.3}$ → Beholder H_0

ii) $H_0: p = 0.30$

$H_1: p \neq 0.30$

(tosidig)



Fork. omr:

$|Z| > 1.96$

→ Beholder H_0 ($Z = 1.53$)

③ Korrelasjonskoeffisient

Skal finne sammenhengen mellom to stokastiske variable X, Y.

Datasett fra utvalg:

x	y
75	0.48
145	1.09
55	0.53
88	0.97
122	0.78

x = motorstørrelse
y = bensinforbruk

i) **Utvalgskovarians:**

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

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) $\boxed{\text{C STAT}}$

2) 75 $\boxed{\text{INPUT}}$ 0.48 $\boxed{\Sigma+}$
145 " 1.09 "
⋮
122 " 0.78 "

$\bar{x} = 97$ $\boxed{\bar{x}, \bar{y}}$

$\bar{y} = 0.77$ $\boxed{\bar{x}, \bar{y}}$ $\boxed{\text{SWAP}}$

$r = 0.783$ $\boxed{\hat{x}, r}$ $\boxed{\text{SWAP}}$

Sterk positiv sammenheng.