

①

$$\left. \begin{array}{l} a) A: X_1, \dots, X_n \sim N(\overset{\text{ukjent}}{\mu}, \sigma^2) \\ B: Y_1, \dots, Y_n \sim N(\mu, \sigma^2) \end{array} \right\} \begin{array}{l} \text{to utvalg A, B fra} \\ \text{samme fordeling} \end{array}$$

90% konfidensintervall basert på X-ene til μ

95% konfidensintervall basert på Y-ene til μ

Er den andre alltid lengre?

$$\left[\bar{X} - \frac{S_x}{\sqrt{n}} t_{0.05, n-1}, \bar{X} + \frac{S_x}{\sqrt{n}} t_{0.05, n-1} \right]$$

$$\left[\bar{Y} - \frac{S_y}{\sqrt{n}} t_{0.025, n-1}, \bar{Y} + \frac{S_y}{\sqrt{n}} t_{0.025, n-1} \right]$$

Siden empirisk standardavvik S_x og S_y kommer inn, kan en utvalg X med stor spredning ha en 90% CI som er lengre enn en 95% CI interval basert på en utvalg Y med liten variasjon.

Oppgave ②

a) Metode for minste kvadraters er en statistisk teknikk som brukes for å finne den beste linjen som passer til et datasett. Målet er å minimere summen av de kvadrerte avstandene (feilene, residualer) mellom de observerte datapunktene og estimerte verdiene fra lineære modeller.

- For hver datapunkt beregner vi residualen

$$y_i - (\beta_0 + \beta_1 x_i)$$

- Feilene kvadreres for å unngå at negative og positive feilene kansellerer hverandre ut, så ser man på

$$\sum_{i=1}^n (y_i - (\beta_0 + \beta_1 x_i))^2$$

- Vi finner verdiene for β_0 og β_1 som minimerer uttrykket, Dette gjøres vanligvis med å derivere feilsummen med hensyn til β_0 og β_1 , og sette dem lik 0 for å finne optimale verdiene.

Når en person er 0 år er maksimal hjertefrekvens forventet å være 197. Med hver enhet (år) i x endrer seg verdien med β_1 . Dvs hvert år synker forventet maks. hjertefrekvens med 0.66.

$$b) \quad H_0: \beta_1 = 0$$

$$H_1: \beta_1 \neq 0$$

Testobservator

$$T = \frac{\hat{\beta}_1}{\sqrt{S^2 / \sum (x_i - \bar{x})^2}}$$

$$\text{med } S^2 = \frac{\sum (y_i - \hat{\beta}_0 - \hat{\beta}_1 x_i)^2}{n-2} = \frac{5581}{50} = 111,62$$

$$\Rightarrow T = \frac{-0,66}{\sqrt{111,62 / 6969}} = -5,215$$

$$t_{\frac{\alpha}{2}, n-2} = t_{0,025, 50} = 2,009$$

$$\Rightarrow T \leq -t_{0,025, 50} \Rightarrow H_0 \text{ forkastes}$$

Det er ^{linear} sammenheng mellom x og y .

Oppgave 3

a) μ_j er forventet verdi for rute nr $j, j=1,2,3$

β_i er avvik ("effekt") for dag nummer i

$i=1,2,3,4,5$ der gjennomsnittlig avvik settes til 0

σ^2 er den felles variansen for hver kjøretur

$$\hat{\mu}_1 = \bar{Y}_{.1} = 25.8 \quad \hat{\mu}_2 = 27.0 \quad \hat{\mu}_3 = 30.2 \quad (\text{kolonne gj.snitt})$$

$$\hat{\beta}_i = \bar{Y}_{.i} - \bar{Y}_{..} = \bar{Y}_{.i} - 27.67$$

$$\Rightarrow (\hat{\beta}_i) = (-3.34, -0.34, 1.00, -0.67, 3.33)$$

Forventet kjøretid for rute 3 på en torsdag er

$$\mu_3 + \beta_4 \text{ som estimeres ved } 30.20 - 0.67 = 29.53$$

b) Komplet ANOVA tabell

	df	SS	MS	F
Treatment (Rute)	$(k-1)=2$	51.73	25.867	12.72
Block (Ukedag)	$(b-1)=4$	71.33	17.8325	8.77
Error	$(k-1)(b-1)=8$	16.27	2.033	
Total	$n-1=14$	139.33		

Ønsker å teste $H_0: \mu_1 = \mu_2 = \mu_3$

H_1 : ikke alle er like

Testobservator $F_{\text{ov}} = \frac{SSTR/2}{SSE/8} = 12.72$

Under H_0 er $F \sim \text{Fisher}(2, 8) \Rightarrow f_{0.01, 2, 8} = 8.65$

$\Rightarrow$ Forkast H_0

Dersom vi bruker vanlig ANOVA ville

$$F = \frac{SSTR/2}{(SSB+SSE)/12} = 3.543$$

Under H_0 er da $F \sim \text{Fisher}(2, 12) \Rightarrow f_{0.01, 2, 12} = 6.93$

$\Rightarrow F < 6.39 \Rightarrow$ Forkast ikke H_0

- c) Det ønskes en sammenligning to av μ_j -ene i form av simultane konfidensintervaller slik at

$$P(\bar{Y}_{.j} - \bar{Y}_{.k} - D\sqrt{MSE} < \mu_j - \mu_k < \bar{Y}_{.j} - \bar{Y}_{.k} + D\sqrt{MSE}, \\ \text{for } \underline{\text{alle}} \text{ par sit}) = 1 - \alpha$$

Med ord: Sannsynl. er $1 - \alpha$ for at alle konfidensintervaller for $\mu_j - \mu_k$ overlapper den riktige verdien

Med $D = 1.81$ blir $D\sqrt{MSE} = 1.81 \sqrt{2.03} \approx 2.58$

$\Rightarrow$

$$\mu_1 - \mu_2: 25.80 - 27.00 \pm 2.58 = (-3.78, 1.38)$$

$$\mu_1 - \mu_3: 25.80 - 30.20 \pm 2.58 = (-6.98, -1.82)$$

$$\mu_2 - \mu_3: 27.00 - 30.20 \pm 2.55 = (-5.78, -0.02)$$

Sannsynligheten er 95% for at alle konfidensintervallene inneholder/overdekker de tre riktige differansene mellom μ_j -ene, hvis forsøket gjentas under de samme betingelsene et stort antall ganger.

Vi kan konkludere at μ_1 er ulik μ_3 og μ_2 er ulik μ_3 (i testen med 5% nivå).

a) Vi ser at H_0 i 2) svarer til $H_0: C = 0$
med den givte definitionen av C

Nå er

$$\begin{aligned} E(\hat{m}_i) &= E(\bar{Y}_{i.}) = E\left(\frac{1}{3} \sum_{j=1}^3 Y_{ij}\right) \\ &= \frac{1}{3} \sum_{j=1}^3 (\mu_j + \beta_i) = \frac{1}{3} \sum_{j=1}^3 \mu_j + \beta_i = \\ &= \mu + \beta_i \Rightarrow \hat{m}_i = \bar{Y}_{i.} \text{ er forventningsrett} \end{aligned}$$

$$\begin{aligned} \Rightarrow \hat{C} &= \sum_{i=1}^5 c_i \hat{m}_i = \sum_{i=1}^5 c_i \bar{Y}_{i.} \\ &= -\frac{1}{4} (24.33 + 27.33 + 28.67 + 27.00) + 31.00 \\ &= 4.1675 \end{aligned}$$

$$\Rightarrow SS_C = \frac{4.1675^2}{\frac{1}{3} \left(\left(\frac{1}{4}\right)^2 + \left(\frac{1}{4}\right)^2 + \left(\frac{1}{4}\right)^2 + \left(\frac{1}{4}\right)^2 + 1^2 \right)} = 41.68$$

$$\Rightarrow F = \frac{SS_C / 1}{SSE / 8} = \frac{41.68}{2.033} = 20.50$$

Under H_0 er $F \sim \text{Fisher}(1, 8) \Rightarrow f_{0.05, 1, 8} = 5.32$

$\Rightarrow F > 5.32$ Vi forkaster H_0

Oppgave 4

a) $D_1, \dots, D_9$ er uavhengige og normalfordelte $N(\mu_D, \sigma_D^2)$

$$H_0: \mu_D = 0$$

$$H_1: \mu_D > 0$$

$$T = \frac{\bar{D}}{S_D / \sqrt{9}} = \frac{3\bar{D}}{S_D} \sim t \text{ fordelt under } H_0$$

$$\text{Her er } \bar{D} = \frac{36}{9} = 4$$

$$\begin{aligned} S_D^2 &= \frac{1}{8} \sum_{i=1}^9 (D_i - \bar{D})^2 = \frac{1}{8} (\sum D_i^2 - 9\bar{D}^2) \\ &= \frac{1}{8} (458 - 9 \cdot 4^2) = 39.25 \end{aligned}$$

$$\Rightarrow T = \frac{3 \cdot 4}{\sqrt{39.25}} = 1.915$$

Sammenlign med $t_{0.05, 8} = 1.860$

da forkast H_0 på 5% nivå.

b) La $p = P(X_i > \mu_i)$

Vi tester $H_0: p = \frac{1}{2}$ mot $H_1: p > \frac{1}{2}$

La U = antall ungdommer med $X_i > \mu_i$

Under H_0 er $U \sim \text{Bin}(n, \frac{1}{2})$

I dataene er $n = 9$ $U = 7$

$$\Rightarrow p\text{-verdi} = 2 \min(P(U \leq \text{observert verdi}),$$

$$P(U \geq \text{observert verdi}))$$

$$= 2 \cdot P(U \geq 7) = 2(1 - P(U \leq 6))$$

$$= 2 \cdot (1 - 0.910) = 2 \cdot 0.09 = 0.18$$

$\uparrow$
tabell $n = 9$
 $p = 0.5$

$\Rightarrow$ Vi forkaster H_0 ikke med tegn testen

c) Her antas det igjen at parene (x_i, y_i) er uavhengige. Man baserer seg igjen på $D_1, \dots, D_n$ og antar at disse har en kontinuertlig og symmetrisk fordeling, around 0 sentral på

Test observatoren er summen av rangene for de positive d_i -ene

Ordre dem først:

rank	①	③	③	③	⑤.5	⑤.5	⑦	⑧	⑨
obs.	2	-3	3	3	-7	7	8	11	12

$$\Rightarrow W_+ = 1 + 3 + 3 + 5.5 + 7 + 8 + 9 = 36.5$$

$$W_- = 3 + 5.5 = 8.5$$

$$\text{Sum} = \frac{9 \cdot 10}{2} = \underline{\underline{45}}$$

$$H_0: \mu_D = 0$$

$$H_1: \mu_D > 0$$

$$2 \cdot \min(0.05, 0.05) = 0.1$$

$$\Rightarrow n = 9$$

$$W_+ = 36.5$$

$$P(W_+ \geq 36.5) = 1 - P(W_+ \leq 36) = 1 - 0.951 = 0.049$$

$\Rightarrow$ forkaster H_0 på 5% nivå