Lektion 1 (10.4, 11.8, 11.9, PS-1)

Punktskattning:

Närmevärde av okänd parameter

- θ verkligt värde
- $\hat{\theta}$ approximation (skattning)
- $\hat{\Theta}$ stokastisk variabel (hur approximationen varierar)

Väntevärdesriktig skattning (VVR):

$$E(\hat{\theta}) = \theta$$

$$E(\hat{\sigma}^2) = \sigma^2$$

Effektiva skattningar:

$$V(\hat{\theta}_1) < V(\hat{\theta}_2)$$

$\Longrightarrow$ $\hat{\theta}_1$ är en mer effektiv skattning än $\hat{\theta}_2$

$$\left.\begin{aligned} \hat{\mu} &= \bar{x} \\ \hat{\sigma}^2 &= s^2 \end{aligned}\right\} \text{skattningar som ofta används}$$

- σ^2 : varians
- s^2 : stickprovsvarians

$$\sigma^2 = \frac{1}{n}\left(\sum_i^n (x_i - \mu)^2\right)$$

$$s^2 = \frac{1}{n-1}\left(\sum_i^n (x_i - \bar{x})^2\right) = \frac{1}{n-1}\left(\left(\sum_i^n x_i^2\right) - n\bar{x}^2\right)$$

repetition:

$$V(X) = E(X^2) - E(X)^2$$

$$E(aX+b) = aE(X) + b$$

$$V(aX+b) = a^2 V(X)$$

$\boxed{11.6}$

Stickprov: $x_1, \ldots, x_n$

$\bar{X} \sim N(\mu, \sigma)$, μ, σ är okända

2 olika skattningar:

$\boxed{1}$: $\hat{\mu}_1 = \bar{x} = \frac{1}{n} \sum_{1}^{n} x_i$

$\boxed{2}$: $\hat{\mu}_2 = \frac{x_1 + x_n}{2}$

a) visa att $\hat{\mu}_1$ o $\hat{\mu}_2$ är vvr

b) visa vilken skattning som är mest effektiv.

VVR

$\boxed{1}$ $E(\hat{\mu}_1) = E(\bar{X}) = E\left(\frac{1}{n} \sum_{1}^{n} \bar{X}_i\right) = \frac{1}{n} E\left(\sum \bar{X}_i\right) =$

$= \frac{1}{n} \cdot \sum_{i}^{n} \underbrace{E(\bar{X}_i)}_{=\mu} = \frac{1}{n} \cdot n \cdot \mu = \mu \Rightarrow$ vvr!

$\boxed{2}$ $E(\hat{\mu}_2) = E\left(\frac{\bar{X}_1 + \bar{X}_n}{2}\right) = \frac{1}{2} E(\bar{X}_1 + \bar{X}_n) = \frac{1}{2}\left(\underbrace{E(\bar{X}_1)}_{\mu} + \underbrace{E(\bar{X}_n)}_{\mu}\right) =$

$= \frac{1}{2}(2\mu) = \mu \Rightarrow$ vvr!

Mest effektiv?

$\boxed{1}$: $V(\hat{\mu}_1) = V(\bar{X}) = V\left(\frac{1}{n} \sum_{1}^{n} \bar{X}_i\right) =$

$= \frac{1}{n^2} V\left(\sum \bar{X}_i\right) = \frac{1}{n^2} \sum_{1}^{n} \underbrace{V(\bar{X}_i)}_{=\sigma^2} = \frac{1}{n^2} \cdot n\sigma^2 = \frac{\sigma^2}{n}$

$\boxed{2}$: $V(\hat{\mu}_2) = V\left(\frac{\bar{X}_1 + \bar{X}_n}{2}\right) = \frac{1}{4}\left(\underbrace{V(\bar{X}_1)}_{=\sigma^2} + \underbrace{V(\bar{X}_n)}_{=\sigma^2}\right) =$

$= \frac{2\sigma^2}{4} = \frac{\sigma^2}{2}$

$\frac{\sigma^2}{n} < \frac{\sigma^2}{2}$, $n > 2$

$\therefore n > 2 \rightarrow \hat{\mu}_1$ mer effektiv än $\hat{\mu}_2$.

$\boxed{10.4}$ två material:

- stickprov: $x_1, \dots x_{10}$
- stickprov: $y_1, \dots, y_5$

$\bar{x} = 5313$, $s_x = 5{,}2$, $\bar{y} = 5309$, $s_y = 3{,}0$

sökt: om alla 15 talen hade betraktats som samma material, vad hade man fått för medelvärde o standardavvikelse?

$$\bar{w} = x_1, \dots, x_n + y_1, \dots y_m = \frac{1}{n+m} \sum_{i=1}^{n+m} w_i = \frac{1}{n+m}\left(\sum_{i=1}^{n} w_i + \sum_{i=1}^{m} w_i\right) =$$

$$= \frac{1}{n+m}\left(\sum_{i=1}^{n} x_i + \sum_{i=1}^{m} y_i\right) = \frac{1}{n+m}(n\bar{x} + m\bar{y}) =$$

$$= \frac{n\bar{x} + m\bar{y}}{n+m} = \frac{10 \cdot 5313 + 5 \cdot 5309}{15} \approx 5311{,}67$$

$$s_x^2 = \frac{1}{n-1} \sum_{i=1}^{n} (x_i - \bar{x})^2 = \frac{1}{n-1}\left(\left(\sum x_i^2\right) - n\bar{x}^2\right)$$

så $\sum_{i=1}$

Lektion 2 (11.23, 11.10, 11.12, 11.14, 11.15)

Maximum - likelihood - funktion

Diskret: $L(\theta) = \prod_{i=1}^{n} P(X_i = x_i ; \theta)$

Kontinuerlig: $L(\theta) = \prod_{i=1}^{n} f_{X_i}(x_i ; \theta)$

} sannolikheten att alla händelser händer,

Punktskattning

- θ : verkligt värde
- $\hat{\theta}$: Approximation (skattning)
- $\hat{\Theta}$: s.v Hur skattningen varierar vid olika stickprov.

Logaritmlagar

- $\ln(xy) = \ln(x) + \ln(y) \Longrightarrow \ln\left(\prod_{i=1}^{n} x_i\right) = \ln(x_1 \cdot \ldots \cdot x_n) = \sum_{i=1}^{n} \ln x_i$
- $\ln x^y = y \ln x$
- $\ln\left(\frac{x}{y}\right) = \ln(x) - \ln(y)$
- $\ln e = 1$

$\boxed{11.15}$

$f_X(x) = \frac{x}{a} e^{-\frac{x^2}{2a}}, \quad x \geq 0$

stickprov: $x_1, \ldots, x_n$

Ange ML-skattningen av a.

(I) ställ upp maximum - likelihood - funktionen

$$L(a) = \prod_{i=1}^{n} f_X(x_i) = \frac{x_1}{a} \cdot e^{-\frac{x_1^2}{2a}} \cdot \ldots \cdot \frac{x_n}{a} \cdot e^{-\frac{x_n^2}{2a}} =$$

$$= \left(\frac{1}{a}\right)^n \cdot \prod_{i=1}^{n} x_i \cdot e^{-\frac{1}{2a}\sum_{i=1}^{n} x_i^2}$$

(II) logaritmera funktionen

$$l(a) = \ln(L(a)) = \ln\left(\left(\frac{1}{a}\right)^n \cdot \prod_{i=1}^{n} x_i \cdot e^{-\frac{1}{2a}\sum_{i=1}^{n} x_i^2}\right) =$$

$$= \ln\left(\frac{1}{a^n}\right) + \ln\left(\prod_{i=1}^{n} x_i\right) + \ln\left(e^{-\frac{1}{2a}\sum x_i^2}\right) =$$

$$= -n \cdot \ln a + \sum_{i=1}^{n} \ln x_i + \frac{-1}{2a} \sum_{i=1}^{n} x_i^2 \cdot \underbrace{\ln e}_{=1}$$

(III) Derivera m.a.p a:

$$\frac{dl}{da} = \frac{-n}{a} + \frac{1}{2a^2} \cdot \sum_{i=1}^{n} x_i^2 = 0 \iff$$

$$\iff \frac{1}{2a^2} \sum_{i=1}^{n} x_i^2 = \frac{n}{a} \iff \boxed{a = \frac{1}{2n} \sum_{i=1}^{n} x_i^2}$$

Den optimala skattningen blir då $\hat{a} = \dfrac{\sum_{i=1}^{n} x_i^2}{2n}$

(IV) Visa att det är en maximipunkt:

- teckentabell el. andraderivatan < 0

$\boxed{11.23}$ $X = 16$ $X \in Bin(25, p)$

a) Ange en skattning av P.

$$P = \frac{X}{n} = \frac{16}{25}$$

b) $\hat{\Theta} = ?$

$V(P^*)$ = hur vår skattning varierar
$V(X)$ = hur utfallet varierar.

$$V(P^*) = V\left(\frac{X}{n}\right) = \frac{1}{n^2}\underbrace{V(X)}_{=np(1-p)} = \frac{P(1-P)}{n}$$

$$\sigma = \sqrt{\frac{P(1-P)}{n}}$$

c) ange skattningens medelfel.

$$\sigma = \sqrt{\frac{P(1-P)}{n}} = \sqrt{\frac{\left(\frac{X}{n}\right)\left(1-\frac{X}{n}\right)}{n}} = \sqrt{\frac{\left(\frac{16}{25}\right)\left(1-\frac{16}{25}\right)}{25}} =$$

$$= 0{,}096$$

$\boxed{11.10}$ $P_X(k) = \theta(1-\theta)^{k-1}$ för $k = 1, 2, 3 \ldots$ där $0 \leq \theta \leq 1$

slumpmässigt stickprov 4, 5, 4, 6, 4, 1

a) $L(\theta) = ?$

$$L(\theta) = \prod_{i=1}^{n} \theta(1-\theta)^{x_i - 1} = \theta^n \cdot (1-\theta)^{\sum_{i=1}^{n} x_i - n}$$

b) ML-skattning av θ

$$l(\theta) = \ln(L(\theta)) = n \ln\theta + \left(\sum_{i=1}^{n} x_i - n\right)\left(\ln(1-\theta)\right)$$

$$\frac{dl}{d\theta} = \frac{n}{\theta} - \underbrace{\sum_{i=1} x_i}_{n\bar{x}} - n \cdot \frac{1}{1-\theta} =$$

$$= \frac{n}{\theta} - \frac{1}{1-\theta} \cdot (n \cdot \bar{x} - n) =$$

$$= \frac{n(1-\theta) - \theta(n \cdot \bar{x} - n)}{\theta(1-\theta)} =$$

$\boxed{11.12}$ X = antalet anrop till telefonväxel , $x_1, \ldots, x_n$ anrop

$Po(\mu)$

a) $\hat{\mu} = ?$

$$P_X(k) = \frac{\mu^k}{k!} e^{-\mu}$$

$$L(\mu) = \prod_{i=1}^{n} \frac{\mu^{x_i}}{x_i!} e^{-\mu} = \left(\frac{1}{x_1! \ldots x_n!}\right) \cdot \mu^{\sum_{i=1}^{n} x_i} \cdot e^{-n\mu}$$

$$l(\mu) = \ln\left(\frac{1}{x_1! \ldots x_n!}\right) + \sum_{i=1}^{n} x_i \cdot \ln\mu + -n\mu \cdot \ln e$$

$$\frac{dl}{d\mu} = \frac{\sum_{i=1}^{n} x_i}{\mu} - n = 0 \iff$$

$$\iff \mu = \frac{\sum_{i=1}^{n} x_i}{n} = \bar{X}$$

b) så $E(\bar{X}) = \frac{1}{n}\sum_{i=1}^{n} E(X_i) = \frac{1}{n}\sum_{i=1}^{n} \mu =$

$$= \frac{1}{n} \cdot n \cdot \mu = \mu$$

$$V(\bar{X}) = V\left(\frac{1}{n}\sum_{i=1}^{n} x_i\right) = \frac{1}{n^2} V\left(\sum_{i=1}^{n} x_i\right) =$$

$$= \frac{1}{n^2}\sum_{i=1}^{n} \underbrace{V(x_i)}_{\mu} = \frac{1}{n^2} \cdot n \cdot \mu = \frac{\mu}{n}$$

$$\Longrightarrow \sigma = \sqrt{\frac{\mu}{n}}$$

c) 8 dagar, 115 82 108 106 118 87 99 92

skatta μ $n = 8$

$$\mu = \frac{\sum_{i=1}^{8} x_i}{n} = \frac{115+82+108+106+118+87+99+92}{8} =$$

$$= 100,875$$

d) $\sigma = \sqrt{\frac{\mu}{n}} = \sqrt{\frac{100,875}{8}} = 3,51$

$\boxed{11.14}$ $X_1, \dots, X_n$

täthetsfunktionen $f_X(x) = \theta x^{\theta-1}$ för $0 < x < 1$

ange ML-skattningen av θ

$$L(\theta) = \prod_{i=1}^{n} \theta x^{\theta-1} = \theta^n \cdot x^{(\sum_{i=1}^{n} \theta - 1)}$$

$$l(\theta) = n \ln \theta + \left(\sum_{i=1}^{n} \theta - 1\right) \cdot \ln x_i$$

$$\frac{dl}{d\theta} = \frac{n}{\theta} + \sum_{i=1}^{n} \ln x_i = 0 \iff$$

$$\iff \theta = -\frac{n}{\sum_{i=1}^{n} \ln x_i}$$

max?

$$\frac{d^2 l}{d\theta^2} = -\frac{n}{\theta^2} < 0 \implies \text{max!}$$

Lektion 3 (12.1, 12.4, 12.16, 12.18, 12.19)

Konfidensintervall:

$\hat{\mu} = \bar{x}$

- σ känt, $D = \sigma/\sqrt{n}$

$I_\mu = (\bar{x} - \lambda_{\alpha/2} D, \bar{x} + \lambda_{\alpha/2} D)$

- σ okänt, skattning: $d = s/\sqrt{n}$

$I_\mu = (\bar{x} - t_{\alpha/2}(t)d, \bar{x} + t_{\alpha/2}(t)d$

$\hat{\sigma} = s$

$I_\sigma = \left(s \cdot \sqrt{\frac{t}{\chi^2_{\alpha/2}(t)}}, s\sqrt{\frac{t}{\chi^2_{1-\alpha/2}(t)}}\right.$

$s^2 = \frac{1}{n-1} \sum (x_i) - \bar{x})^2$

$t = n-1$

$\boxed{12.11}$

Fyra mätningar: 8,24, 8.18, 8.15, 8.23

$\bar{X} \sim N(\mu + \Delta, \sigma)$

$\Delta = 0,1$, systematiskt fel

$\sigma = 0,05$

Bestäm konfidensintervall för μ med konfidensgraden 0,99.

$P(-a \leq \hat{\mu} \leq a) = 0,99$, vi vill att $\hat{\mu}$ ska ligga inom intervallet med 99% sannolikhet.

$\bar{X} \sim N(\mu + \Delta \; \sigma/\sqrt{n})$

$\frac{\bar{X} - \mu - \Delta}{\sigma/\sqrt{n}} \sim N(0,1)$

$\Rightarrow P\left(-a \leq \frac{\bar{x} - \mu - \Delta}{\sigma/\sqrt{n}} \leq a\right) = \lambda_a - \lambda_{-a} = \lambda_a - (1 - \lambda_a) = 2\lambda_a - 1 =$

$= 0,99 \Leftrightarrow \lambda_a = \frac{1,99}{2} = 0,995 \Rightarrow a = 2,575$

$\Rightarrow P\left(-2,575 \leq \frac{\bar{x} - \mu - \Delta}{\sigma/\sqrt{n}} \leq 2,575\right) = 0,99$ (lös nu ut μ)

$\mu = (\bar{x} - \Delta \pm 2,575 \cdot \sigma/\sqrt{n})$

Lektion 4 (12.21, 12.22, 12.25, PS-4, 12.34, 12.36)

(I) $I_{\mu_1 - \mu_2}$ då σ_1, σ_2 kända

$$\frac{\bar{X} - \bar{Y} - (\mu_1 - \mu_2)}{\sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}} \sim N(0,1)$$

(II) $I_{\mu_1 - \mu_2}$ då $\sigma_1 = \sigma_2 = \sigma$ (okänd men samma)

$$\underbrace{\frac{\bar{X} - \bar{Y} - (\mu_1 - \mu_2)}{S\sqrt{\frac{1}{n_1} + \frac{1}{n_2}}}}_{\text{leder till}} \sim t\underbrace{(n_1 + n_2 - 2)}_{a}, \quad S^2 = \frac{(n_1 - 1)S_1^2 + (n_2 - 1)S_2^2}{n_1 + n_2 - 2}$$

(III) I_σ då $\sigma_1 = \sigma_2 = \sigma$

$$\frac{(n_1 + n_2 - 2)}{\sigma^2} \sim \chi^2(n_1 + n_2 - 2)$$

(II) extra: $\mu_1 - \mu_2 \in \bar{X} - \bar{Y} \pm a \cdot S\sqrt{\frac{1}{n_1} + \frac{1}{n_2}}$

$\boxed{12.25}$

$X \sim N(\mu_x, \sigma_1)$

$Y \sim N(\mu_Y, \sigma_2)$

$\bar{x} = 148{,}2$, $s_x = 10$, $n_x = 50$

$\bar{Y} = 151{,}7$, $s_Y = 8{,}0$, $n_Y = 25$

Bestäm $I_{\mu_1 - \mu_2}$, $1 - \alpha = 0{,}95$

(I) ställ upp hjälpvariabel:

$\bar{X} \sim N(\mu_x, \sigma_1/\sqrt{n})$

$\bar{Y} \sim N(\mu_Y, \sigma_2/\sqrt{n})$

$\bar{Y} - \bar{X} \sim N\left(\mu_Y - \mu_X, \sqrt{\frac{\sigma_2^2}{n_y} + \frac{\sigma_1^2}{n_x}}\right)$

$$\Rightarrow \frac{\bar{Y} - \bar{X} - (\mu_Y - \mu_X)}{\sqrt{\frac{\sigma_2^2}{n_Y} + \frac{\sigma_1^2}{n_X}}}$$

då σ_1 o σ_2 är okända görs en skattning och vi får hjälpvariabeln:

$$\Rightarrow \frac{\bar{Y} - \bar{X} - (\mu_Y - \mu_X)}{S\sqrt{\frac{1}{n_x} + \frac{1}{n_y}}} \sim t(n_x + n_y - 2)$$

(II) Bestäm S o $t(n_x + n_y - 2)$:

$$S = \sqrt{\frac{(n_x - 1)s_x^2 + (n_y - 1)s_y^2}{n_x + n_y - 2}} = \sqrt{\frac{6436}{73}} \approx 9{,}39$$

$f = n_x + n_y - 2 = 73$ (frihetsgrad)

$\underline{\underline{t_{0,025}(73) = 1{,}99}}$

(III) Beräkna $I_{\mu_Y - \mu_X}$:

$$\mu_Y - \mu_X \in \bar{y} - \bar{x} \pm t_{0,025}(73) \cdot S\sqrt{\frac{1}{n_x} + \frac{1}{n_y}} \Leftrightarrow$$

$$\Leftrightarrow \mu_Y - \mu_X \in 151{,}7 - 148{,}2 \pm 1{,}99 \cdot 9{,}39\sqrt{\frac{1}{50} + \frac{1}{25}}$$

$$\Rightarrow \mu_Y - \mu_X \in 3{,}5 \pm 4{,}5$$

$$\Rightarrow \underline{\underline{I_{\mu_Y - \mu_X} = (-1, 8)}}$$

$\underline{\underline{P(-1 \leq \mu_Y - \mu_X \leq 8) = 95\%}}$

Lektion 5 (13.4, 13.5, 13.26, 13.8, 13.9, 13.11, 13.12, 13.21)

Hypotesprövning

$H_0 : \theta = \theta_0$

$H_1 : \theta \neq \theta_0 \quad <, >$

Fel typ I

H_0 förkastas fast H_0 sann

α = risk för fel I. $\Rightarrow$ vill att risken att vi slänger en hypotes som är sann ska vara så liten som möjligt.

Fel typ II

H_0 förkastas inte, fast H_0 falsk

β = risk för fel II $\Rightarrow$ vill att risken att vi sparar en hypotes som är falsk ska vara så liten som möjligt

Signifikansnivå

$\alpha = P(H_0$ förkastas trots H_0 är sann$)$

Dåligt $\Rightarrow$ α lågt

Styrka

För θ_1:

$h(\theta_1) = P(H_0$ förkastas då $\theta = \theta_1)$

Bra $\Rightarrow$ hög styrka

$\beta = 1 - h(\theta_1)$

Styrkefunktion

Beroende på vad θ är kommer testet vara olika "starkt".

$h(\theta) = P(H_0$ förkastas då θ är det sanna värdet$)$

P-värde

P = sannolikheten, givet H_0, att få ett visst värde/utfall

$P < \alpha \Rightarrow$ förkasta

$P \geq \alpha \Rightarrow$ förkasta ej

$\boxed{13.8}$

$\bar{X} \sim N(\mu ; 0.2)$

$H_0 : \mu = 4.0$

$H_1 : \mu > 4.0$

$\alpha = 0.05$

10 mätningar gav $\bar{x} = 4.1$

$0.05 = P(H_0 \text{ förkastas fast } H_0 \text{ sann})$

$\bar{X} \sim N(\mu ; 0.2/\sqrt{n})$

$$Z = \frac{\bar{X} - \mu}{0{,}2/\sqrt{n}} \sim N(0,1)$$

$P(z > a \mid H_0 \text{ sann}) = 0{,}05$

$P(z < a) = 0{,}95 = \Phi(a)$

$\Rightarrow a = 1{,}645$

$\uparrow$ tabell

$\therefore$ Sannolikheten att $z > 1{,}645$ är $0{,}05$ %, givet H_0 (sann). $\mu = 4$

Studera nu utfallet:

$\bar{x} = 4.1$, $n = 10$, $\mu = 4$

$$z = \frac{\bar{x} - \mu}{0{,}2/\sqrt{n}} = \frac{4.1 - 4}{0{,}2/\sqrt{10}} = 1{,}5811 < 1{,}645$$

$\longrightarrow$ H_0 kan inte förkastas!, (H_0 kan stämma)

$P(\bar{X} > b) = 0{,}05$, $\bar{X} \sim N(\mu_1 ; 0{,}2/\sqrt{n})$

"Sannolikheten att medelvärdet $\bar{x}$ är högre än b är 0,05"

<u>gör variabelbyte</u>:

$$P(Z > a) = 0{,}05 \Rightarrow z = \frac{\bar{X} - \mu}{0{,}2/\sqrt{n}} \sim N(0,1)$$

$\boxed{13.4}$

$H_0 : \theta = 1000$

Om $X_1 < a \Rightarrow$ förkasta H_0

$P(X < a) = \alpha$ där α är signifikationsnivån

a) Bestäm a som funktion av α

Exponentialfördelad:

$f(x) = \frac{1}{\theta} e^{-x/\theta}$, $x \geq 0$ (täthetsfunktionen)

$$F_X(t) = P(X \leq t) = \int_{-\infty}^{t} f_X(x)\,dx = \int_0^t \frac{1}{\theta} e^{-x/\theta}\,dx =$$

$$= 1 - e^{-t/\theta}$$

då $\alpha = P(X < a) = F_X(a) = 1 - e^{-a/\theta} \iff$

$\iff 1 - \alpha = e^{-a/\theta} \iff -a/\theta = \ln(1-\alpha) \iff$

$\iff a = -\theta \ln(1-\alpha)$ med $\theta = 1000$ fås:

$$a = -1000 \ln(1-\alpha)$$

b) $\alpha = 0{,}05$ ger $a = -1000 \ln(1 - 0{,}05) = 51{,}293$

så $X < 51{,}293$ då $X = 75$ här så förkastas inte H_0.

c) $X = 50 < 51{,}293 \Rightarrow H_0$ förkastas

d) $a = X = 45$ så fås:

$P(X < a) = F_X(a) = 1 - e^{a/\theta} = 1 - e^{45/1000} = 0{,}044003$

så den lägsta signifikationsnivå hypotesen $\theta = 1000$ kan förkastas på, givet observationen $X = 45$ är 4,4%

$\boxed{13.5}$

X = antalet spel till första vinst

$p_X(k) = p(1-p)^{k-1}$ för $k = 1, 2 \ldots$

$H_0 : p = 0{,}2$

$H_1 : p < 0{,}2$

$\alpha = 0{,}10$

vinner den 11:e spelomgången

P-värdesmetoden ska användas:

$P(X > k) = (1-p)^k$

Lektion 6 (PS-15, PS-16, PS-20, PS-21)

Stokastiska vektorer

$$\underline{Y} = \begin{pmatrix} Y_1 \\ \vdots \\ Y_n \end{pmatrix} = A\underline{X} + b$$

$$\begin{cases} E(\underline{Y}) = A\,E(\underline{X}) + b \\ C_y = A C_x A^T \end{cases}$$

$$\underline{Y} \sim N(E(\underline{Y}), C_y)$$

Linjärkombination

$$Z = a_1 X_1 + \dots + a_n X_n$$

$$var(Z) = C_Z = (a_1 \dots a_n) \begin{pmatrix} X_1 \\ \vdots \\ X_n \end{pmatrix} \begin{pmatrix} a_1 \\ \vdots \\ a_n \end{pmatrix}$$

Korrelation ρ

$$\rho(X, Y) = \frac{Cov(X, Y)}{\sqrt{var(X) var(Y)}}$$

$$C = \begin{pmatrix} \sigma_1^2 & \rho\sigma_1\sigma_2 \\ \rho\sigma_1\sigma_2 & \sigma_2^2 \end{pmatrix}$$

Om X & Y är oberoende $\Rightarrow \rho = 0$

$\Longrightarrow$ C blir en diagonalmatris

Om X & Y är likafördelade (och oberoende)

$$\Rightarrow C = \begin{pmatrix} \sigma^2 & & \\ & \ddots & \\ & & \sigma^2 \end{pmatrix} = \sigma^2 I$$

Flera dimensioner N (täthetsfunktionen)

$$f(\underline{Y}) = \frac{1}{\sqrt{2\pi}^n \sqrt{\det C}} \cdot e^{-\frac{1}{2}((y-\mu)^T C^{-1} (y-\mu))}$$

Exempel, tenta

$X_1, X_2 \sim N(0,1)$ $\Rightarrow$ oberoende & likafördelade

$$\underline{Y} = \begin{pmatrix} Y_1 \\ Y_2 \end{pmatrix} = \begin{pmatrix} X_1 - 2X_2 \\ X_1 + cX_2 \end{pmatrix}$$

a) Ange $\underline{Y}$:s fördelning

b) Bestäm c så att Y_1, Y_2 är oberoende

$$\underline{Y} = \underbrace{\begin{pmatrix} 1 & -2 \\ 1 & c \end{pmatrix}}_{A} \begin{pmatrix} X_1 \\ X_2 \end{pmatrix}$$

$E(Y)$

$E(X) = \begin{pmatrix} 0 \\ 0 \end{pmatrix}$

$E(Y) = AE(X) = \begin{pmatrix} 0 \\ 0 \end{pmatrix}$

C_Y

$C_Y = A C_X A^T$

$X_1, X_2 \sim N(0,1)$ oberoende

$\Rightarrow C_X = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix} = 1 \cdot I = I$

$$C_Y = \begin{pmatrix} 1 & -2 \\ 1 & c \end{pmatrix} \begin{pmatrix} 1 & 1 \\ -2 & c \end{pmatrix} = \begin{pmatrix} 5 & 1-2c \\ 1-2c & 1+c^2 \end{pmatrix}$$

a) svar: $\underline{Y} \sim N\left(\begin{pmatrix} 0 \\ 0 \end{pmatrix}, \begin{pmatrix} 5 & 1-2c \\ 1-2c & 1+c^2 \end{pmatrix}\right)$

b) C_Y ska vara en diagonalmatris och det blir den om $1-2c = 0$:

$1 - 2c = 0 \Rightarrow c = \frac{1}{2}$

svar: $c = 1/2$

$\boxed{PS-15}$

$X_1, X_2 \sim N(0,1)$ oberoende o likafördelade

$$\underline{Y} = \begin{pmatrix} Y_1 \\ Y_2 \end{pmatrix} = \underbrace{\begin{pmatrix} 1 & 1 \\ 2 & -1 \end{pmatrix}}_{A} \begin{pmatrix} X_1 \\ X_2 \end{pmatrix}$$

$$C_Y = \begin{pmatrix} 1 & 1 \\ 2 & -1 \end{pmatrix} \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix} \begin{pmatrix} 1 & 2 \\ 1 & -1 \end{pmatrix} = \begin{pmatrix} 1 & 1 \\ 2 & -1 \end{pmatrix} \begin{pmatrix} 1 & 2 \\ 1 & -1 \end{pmatrix} =$$

$$= \begin{pmatrix} 2 & 1 \\ 1 & 5 \end{pmatrix}$$

<u>det C</u>

$$\begin{vmatrix} 2 & 1 \\ 1 & 5 \end{vmatrix} = 9$$

<u>E(Y)</u>

$$E(X) = \begin{pmatrix} 0 \\ 0 \end{pmatrix}$$

$$E(Y) = A \cdot E(X) = \begin{pmatrix} 0 \\ 0 \end{pmatrix}$$

<u>Bestäm täthetsfunktionen</u>

$$f(\underline{Y}) = \frac{1}{\sqrt{2\pi}\sqrt{\det C}} \cdot e^{-\frac{1}{2}((y-\mu)^T \cdot C^{-1}(y-\mu))}$$

$$\Rightarrow f(\underline{Y}) = \frac{1}{\sqrt{2\pi}^2 \sqrt{9}} \cdot e^{-\frac{1}{2}((y-0)^T \cdot \begin{pmatrix} 2 & 1 \\ 1 & 5 \end{pmatrix}^{-1}(y-\mu))} =$$

$$= \frac{1}{6\pi} e^{-1/2\left((y_1, y_2) \frac{1}{9} \begin{pmatrix} 5 & -1 \\ -1 & 2 \end{pmatrix} \begin{pmatrix} y_1 \\ y_2 \end{pmatrix}\right)} =$$

$$= \frac{1}{6\pi} e^{-1/18 \left(\begin{pmatrix} 5y_1 - y_2 \\ -y_1 + 2y_2 \end{pmatrix} \begin{pmatrix} y_1 \\ y_2 \end{pmatrix} \right)} =$$

$$= \frac{1}{6\pi} e^{-1/18((5y_1 - y_2)y_1 + (-y_1 + 2y_2)y_2)} =$$

$$= \frac{1}{6\pi} e^{-1/18(5y_1^2 - 2y_1y_2 + 2y_2^2)}$$

$\boxed{\text{PS-16}}$

$$\begin{pmatrix} X_1 \\ X_2 \\ X_3 \end{pmatrix} \sim N\left(\begin{pmatrix} 60 \\ 60 \\ 60 \end{pmatrix}, \begin{pmatrix} 100 & 80 & 20 \\ 80 & 100 & 10 \\ 20 & 10 & 80 \end{pmatrix} \right)$$

$$Y = (X_1 + X_2 + 2X_3)/4$$

Bestäm fördelningen för Y samt ett värde a

så $P(Y > a) = 90$

$$Y = \underbrace{\frac{1}{4}(1\ 1\ 2)}_{A} \begin{pmatrix} X_1 \\ X_2 \\ X_3 \end{pmatrix}$$

Bestäm $E(Y)$

$$E(X) = \begin{pmatrix} 60 \\ 60 \\ 60 \end{pmatrix}$$

$$E(Y) = A \cdot E(X) = \frac{1}{4}(1\ 1\ 2)\begin{pmatrix} 60 \\ 60 \\ 60 \end{pmatrix} = \frac{1}{4}\begin{pmatrix} 60 \\ 60 \\ 120 \end{pmatrix} = 60$$

Bestäm C_y

$$C_y = A\, C_x\, A^T = \underbrace{\frac{1}{4}(1\ 1\ 2)}_{1\times 3} \underbrace{\begin{pmatrix} 100 & 80 & 20 \\ 80 & 100 & 10 \\ 20 & 10 & 80 \end{pmatrix}}_{(3\times 3)} \frac{1}{4}\begin{pmatrix} 1 \\ 1 \\ 2 \end{pmatrix} =$$

$$= \frac{1}{16} \underbrace{(220, 200, 190)}_{1\times 3} \underbrace{\begin{pmatrix} 1 \\ 1 \\ 2 \end{pmatrix}}_{3\times 1} = \frac{1}{16} 800 = 50$$

$$\text{Var}(Y) = 50 \Rightarrow \sigma = \sqrt{50}$$

svar: $Y \sim N(60, \sqrt{50})$

$$Z = \frac{Y - \mu}{\sigma} = \frac{Y - 60}{\sqrt{50}}$$

$$P\left(\frac{Y-60}{\sqrt{50}} > \frac{a-\mu}{\sigma}\right) = 0{,}9 \Leftrightarrow P\left(Z > \frac{a-60}{\sqrt{50}}\right) = 0{,}9$$

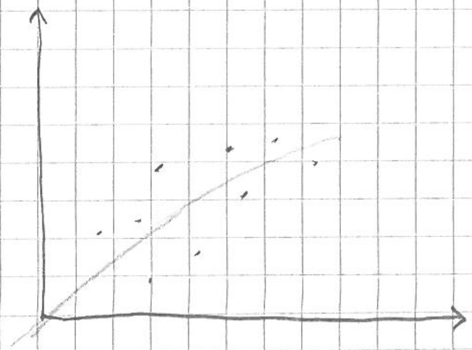
$$\Rightarrow P\left(Z < \frac{a-60}{\sqrt{50}}\right) = 0{,}1 \Rightarrow \Phi\left(\frac{a-60}{\sqrt{50}}\right) = 0{,}1$$

$$\frac{a-60}{\sqrt{50}} = \underbrace{\lambda_{0,1}}_{\text{tabellvärde}}$$

Lös sedan för a.

Lektion 7 (14.40-d, PS-26, PS-27, PS-25)

Regressionsanalys



$Y = \mu + \varepsilon$

Forma ett linjärt samband,

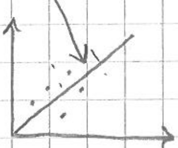
μ ligger på linjen

$\varepsilon = \varepsilon(0, \sigma) \Rightarrow$ avvikelsen från linjen

$$Y = \underbrace{\beta_0 + \beta_1 \cdot x}_{kx+m} + \varepsilon$$

Detta leder till att:

$$y \sim N(\underbrace{\beta_0 + \beta_1 x}, \sigma)$$



$$\underbrace{\beta_0 + \beta_1 x_1 + \beta_2 x_2}_{\text{Det förväntade, det som ska hända.}} + \underset{\uparrow \text{ avvikelsen}}{\varepsilon}$$

Generellt

$\underline{Y} = \underline{X}\beta + \varepsilon \quad , \varepsilon \sim N(0, \sigma)$

$E(\underline{Y}) = \underline{X}\beta$

$C_y = \sigma^2 I$

MK-skattning

$\det(\underline{X}^T \underline{X}) \neq 0$

$\hat{\beta} = (\underline{X}^T \underline{X})^{-1} \underline{X} y$

Specialfall av en variabel

$y = \alpha + \beta x + \varepsilon$

$\hat{\beta} = \frac{S_{xy}}{S_{xx}}$

$\hat{\alpha} = \bar{y} - \hat{\beta}\bar{x}$

$\hat{\sigma}^2 = S^2 = \frac{1}{n-2} \sum (y_i - \alpha - \beta x_i)^2$

Def. Kvadratsummor

n SS_{TOT} (Q_{TOT}) : Total variation hos y

$k-1$ SS_R (Q_{Regr}) : Variation som förklaras med observationer

$n-k-1$ SS_E (Q_{Res}) : Variation som ej förklaras av modellen.

Förklaringsgraden

$$R^2 = \frac{SS_R}{SS_{TOT}}$$ Bra: $R^2 \to 1$

$$\hat{\beta} \sim N(\beta, \sigma^2 (\underline{X}^T \underline{X})^{-1})$$ $\Rightarrow$ För hela vektorn kovariansmatris

$$(\underline{X}^T\underline{X})^{-1} = \begin{pmatrix} h_{00} & \cdots & \\ \vdots & \ddots & \\ \vdots & & h_{mm} \end{pmatrix}$$

$E(\hat{\beta}) = \beta$

$C\beta = \sigma^2 (\underline{X}^T \underline{X}^{-1})$ y skall vara oberoende, denna skall vara en diagonalmatris

$$\hat{\beta}_i \sim N(\beta_i, \sigma\sqrt{h_{ii}})$$

$$\frac{\hat{\beta}_i - \beta_i}{\sigma\sqrt{h_{ii}}} \sim N(0,1)$$

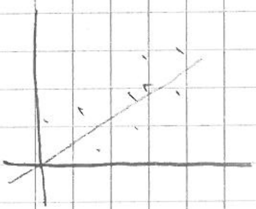
Lektion 8 (PS-30, PS-31, PS-38, PS-36)

$$\bar{Y} = \bar{X}\beta + \varepsilon \qquad (\bar{X}^T\bar{X})^{-1} = \begin{pmatrix} h_{00} & \cdots & \\ \vdots & \ddots & \\ & & h_{mm} \end{pmatrix}$$

$$\frac{\hat{\beta}_i - \beta_i}{\sigma\sqrt{n_{ii}}} \sim N(0,1)$$

$$\frac{\hat{\beta}_i - \beta_i}{s\sqrt{n_{ii}}} \qquad \left[\ s^2 = \frac{SS_E}{n-k-1}\right.$$

då σ är okänt



$$y = \beta_0 + \beta_1 \bar{X}$$

$$\frac{\hat{\beta}_1 - \beta_1}{s\sqrt{n_{11}}} \sim t(n-k-1)$$

$$\beta_1 \in \hat{\beta}_1 \pm t_{0,975}(n-k-1)\, s\sqrt{n_{11}}$$

$$\beta_1 \in \hat{\beta}_1 \pm t_{0,975}(n-k-1)\, d(\hat{\beta}_1)$$

Konfidensintervall för $E(\bar{Y})$

$$E(\bar{Y}_0) = \mu_0 \qquad u = \begin{pmatrix} 1 \\ \vdots \\ x_n \end{pmatrix}$$

$$\underbrace{\hat{\mu}_0 = u'\hat{\beta}}_{\text{En skattning av väntevärdet}}$$

$$u'\hat{\beta} \sim N\left(u'\beta,\ \sigma\sqrt{u(\bar{X}'\bar{X})^{-1}u}\right)$$

$$\frac{\hat{\mu}_0 - \mu_0}{s\sqrt{u'(\bar{X}'\bar{X})^{-1}u}} = \frac{u'\hat{\beta} - u'\beta}{s\sqrt{u'(\bar{X}'\bar{X})^{-1}u}} \sim t(n-k-1)$$

Prediktionsintervall för $\bar{Y}_0$

$$\left.\frac{Y_0 - u'\hat{\beta}}{s\sqrt{1+u'(\bar{X}'\bar{X})^{-1}u}} \sim t(n-k-1)\ \right\}$$

En enskild mätning, inom vilket intervall kan den specifika mängden hamna?

Hypotesläsning

$H_0: \beta_1 = \dots = \beta_n = 0$

H_1: minst ett $\beta_i = 0$

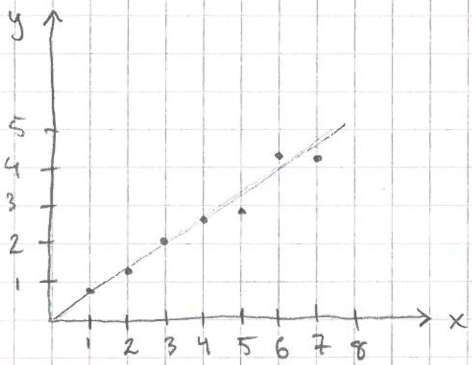
$$V = \frac{SS_R/k}{SS_E/(n-k-1)} \quad . \; V \sim F(k, n-k-1) \} \text{ F-fördelad}$$

$\boxed{14.4}$

inställning x
inställning y

x:	1.0	2.0	3.0	4.0	5.0	6.0	7.0	$\bar{x} = 4.0$
y:	0.9	1.4	2.2	2.7	3.2	4.3	4.2	$\bar{y} = 2.7$

a) Rita upp värdena i ett koordinatsystem y beror linjärt av x:



Beräkunar:

- $S_{xx} = \sum_{i=1}^{7} (x_i - \bar{x})^2 = (1-4)^2 + (2-4)^2 + (3-4)^2 + (4-4)^2 + (5-4)^2 + (6-4)^2 + (7-4)^2 =$
  $= 9+4+1+1+4+9 = 28$
- $S_{yy} = \sum_{i=1}^{7} (y_i - \bar{y})^2 = (0{,}9-2.7)^2 + (1.4-2.7)^2 + (2.2-2.7)^2 + (2.7-2.7)^2 + (3.2-2.7)^2 + (4.3-2.7)^2 + (4.2-2.7)^2 =$
  $= 10.24$
- $S_{xy} = \sum_{i=1}^{7} (x_i - \bar{x})(y_i - \bar{y}) = 16.7$

b) Regressionsanalysens parametrar

$\hat{\beta} = \frac{S_{xy}}{S_{xx}} = \frac{16.7}{28} = 0.5964$ $\qquad S^2 = \frac{1}{n-2} (S_{yy} - \hat{\beta} \cdot S_{xy}) = 0.559$

$\hat{\alpha} = 2.7 - 0.5964 \cdot 4 = 0.3144$ $\qquad \sigma = 0.2361$

c) väntevärdet 2.5 för dimension

$\therefore \; 2.5 = E(Y(x)) = \alpha + \beta x \iff x = \frac{2.5 - \alpha}{\beta} = \frac{2.5 - 0{,}3144}{0.5964} = 3.66$