8.8 Grafiske teknikkar

Definisjon 8.8. Ein kvantil i eit utval, $q(f)$ er ein verdi der $f \cdot 100\%$ av dataane er mindre eller lik $q(f)$

Eksempel

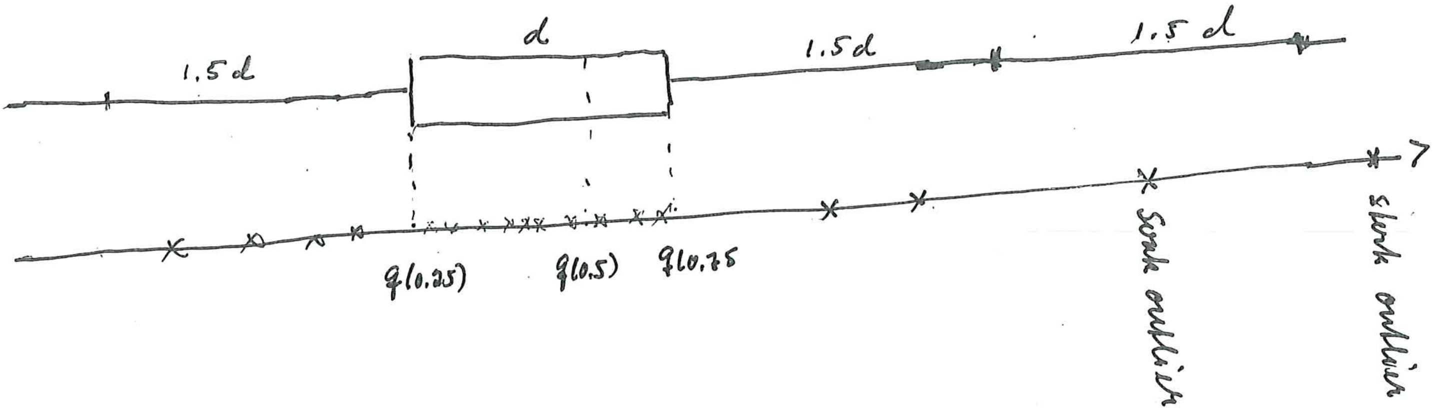
$q(0.25)$ = nedre kvartil (25% kvantil)

$q(0.50)$ = medianen (50% kvantil)

$q(0.75)$ = övre kvartil (75% kvantil)

Sjekk av outliers. Boksplott.

La $d = q(0.75) - q(0.25)$



Kvantilplott

Ordne dataane etter algebraisk storleik

$x_{(1)}, x_{(2)}, \ldots, x_{(n)}$

La $f_i = \# x_i \le x_{(i)} = \frac{i}{n}$. Eit kvantilplott er eit plott av $x_{(i)}$ mot f_i. Ofte nyttar ein verdien $f_i = \frac{i - \frac{3}{8}}{n + \frac{1}{4}}$ for f_i.

Normal-plott, Sannsyns-plott.

Sjekk for om ein kan anta normalfordeling.

La $X \sim N(\mu, \sigma^2)$

$$F_X(x) = P(X \le x) = P\left(\frac{X-\mu}{\sigma} \le \frac{x-\mu}{\sigma}\right) = \underbrace{\Phi}_{N(0,1)}\left(\frac{x-\mu}{\sigma}\right) \Rightarrow \Phi^{-1}(F_X(x)) = \frac{x-\mu}{\sigma}$$

La $X_1, X_2, \ldots, X_n$ vere eit tilfeldig utval frå ei normalfordeling. Ordna etter storleik kan verdiane (realiseringane) skrivast som

$$x_{(1)}, x_{(2)}, \ldots, x_{(n)}$$

Då er

$$F_X(x_{(i)}) \approx \frac{\# X \le x_{(i)}}{n} = \frac{i}{n} \approx \begin{cases} \frac{i-\frac{1}{2}}{n} \\ \frac{i-3/8}{n+\frac{1}{4}} \end{cases} = f_i$$

Eit plott av $\Phi^{-1}(f_i)$ mot $x_{(i)}$ skal tilnærma bli ei rett linje.

9 Estimering

Eksempel: Vil estimere talet på fiskar, N, i ein innsjö. Fangar 34 fisk, merkar dei og slepp dei ut att. Fangar deretter 50 stykker og registrerer kor mange som er merka.

La X vere talet på merka fisk.

Merka	Ikkje merka
34	N - 34

$$P(X=x) = \frac{\binom{34}{x}\binom{N-34}{50-x}}{\binom{N}{50}} \quad \text{ɔ: hypergeometrisk. Kap 5.4}$$

$$E[X] = \frac{nk}{N} = \frac{50\cdot 34}{N} \quad \text{ɔ:} \quad N = \frac{50\cdot 34}{E[X]}$$

Eit punkt estimat for N er $\frac{50\cdot 34}{X}$

Med $x = 9$ får vi $N = \frac{50\cdot 34}{9} = \frac{1700}{9} = 189$

Vanlege punktestimatorar

parameter	Estimator	
	$\bar{X}$	$E[\bar{X}] = \mu$.
μ	$S^2 = \sum_{i=1}^{n} \frac{(X_i - \bar{X})^2}{n-1}$	$E[S^2] = \sigma^2$
σ^2		
p	$\hat{p} = \frac{X}{n}$, $X \sim B(n,p)$	$E[\hat{p}] = \frac{np}{n} = p$

Når vi brukar ein statistikk til å estimere ein parameter, kallar vi statistikken ein estimator.

Estimator: $\hat{\Theta}(X_1, \dots, X_n)$ er ein stokastisk variabel (Stok. variable)

Estimat: $\hat{\theta}(x_1, \dots, x_n)$ er eit tal (Obs. verdiar)

<u>Definisjon 9.1</u>. Ein statistikk, $\hat{\theta}$, seiast å vere ein forventingsrett (ikkje skeiv) estimator for θ dersom $E[\hat{\theta}] = \theta$.

Eksempel $\bar{X}$, S^2, $\hat{p}$.

<u>Definisjon 9.2</u>. Mellom alle forventingsrette estimatorar for θ, blir den med minst varians kalla den mest ~~effektive~~ effisiente (effektive).

<u>Eks</u>. For normalfordelinga er

$$E[\bar{X}] = \mu \qquad \text{der } \tilde{X}\text{ (medianen)} = \begin{cases} X_{(\frac{n+1}{2})}, & n \text{ odde} \\ \dfrac{X_{(\frac{n}{2})} + X_{(\frac{n}{2}+1)}}{2}, & n \text{ jamn} \end{cases}$$

$$E[\tilde{X}] = \mu$$

Men variansen til $\bar{X}$ er langt mindre enn variansen til $\tilde{X}$.

Utvals fordelinga til s^2

Vi har $E[S^2] = \sigma^2$. Kva med fordelinga til S^2

La X vere $N(\mu, \sigma^2)$ og la $Y = \left(\frac{X-\mu}{\sigma}\right)^2 = Z^2$

$$F_Y(y) = P(Y \le y) = P(Z^2 \le y) = P(-\sqrt{y} \le Z \le \sqrt{y})$$

$$= \Phi(\sqrt{y}) - \Phi(-\sqrt{y}) = F_Z(\sqrt{y}) - F_Z(-\sqrt{y})$$

$$\Rightarrow \quad f_Y(y) = f_Z(\sqrt{y}) \cdot \frac{1}{2\sqrt{y}} + f_Z(-\sqrt{y}) \cdot \frac{1}{2\sqrt{y}} = f_Z(\sqrt{y}) \cdot y^{-\frac{1}{2}}$$

$$= \frac{1}{\sqrt{2\pi}} y^{\frac{1}{2}-1} e^{-\frac{y}{2}}, \quad y > 0$$

d: gamma$(\frac{1}{2}, 2)$

sidan $\Gamma(\frac{1}{2}) = \sqrt{\pi}$, $M_Y(t) = \left(\frac{1}{1-\underbrace{2}_{\beta}t}\right)^{\overbrace{\frac{1}{2}}^{\alpha}}$

La $X_1, \dots, X_n \sim N(\mu, \sigma^2)$

Vi har vist at $\sum_{i=1}^{n} (X_i - \bar{X})^2 = \sum_{i=1}^{n} (X_i - \mu)^2 - n(\bar{X} - \mu)^2$

d.a. $$\sum_{i=1}^{n} \left(\frac{X_i - \mu}{\sigma^2}\right)^2 = \sum_{i=1}^{n} \frac{(X_i - \bar{X})^2}{\sigma^2} + \frac{(\bar{X} - \mu)^2}{\frac{\sigma^2}{n}}$$

$\wr$ gamma$(\frac{n}{2}, 2)$ $=$ $\chi^2(n)$ — Kjikvadrat-fordeling

gamma$(\frac{1}{2}, 2)$ — $\chi^2(1)$

Det er difor rimeleg at $\sum_{i=1}^{n} \frac{(X_i - \bar{X})^2}{\sigma^2}$ er $\chi^2(n-1)$

d: $\frac{n-1}{\sigma^2} S^2 \sim \chi^2(n-1)$ eller gamma$(\frac{n-1}{2}, 2)$

og $E[S^2] = \frac{\sigma^2}{n-1}(n-1) = \sigma^2$ og $Var[S^2] = \frac{\sigma^4}{(n-1)^2} \cdot 2(n-1) = \underline{\frac{2\sigma^4}{n-1}}$