Föreläsning 9

Kap 10 Beskrivande statistik

Sannolikhetslära

väntevärde μ

varians σ^2

Statistik

medelvärde $\bar{x} = \frac{1}{n} \sum x_i$

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

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

+ populationsvariansen om hela populationen

$R(X) = \frac{D(X)}{E(X)}$ variationskoefficienten $100 \cdot \frac{s}{\bar{x}}$ uttrycks i procent

(endast om vi bara har icke-negativa data)

$F_X(\tilde{x}) = 0.5$ medianen = mittvärdet $= \tilde{x}$

$E[X, Y]$ kovariansen $c_{xy} = \frac{1}{n-1} \sum (x_i - \bar{x})(y_i - \bar{y})$

$\rho[X, Y]$ $r = \frac{c_{xy}}{s_x s_y}$ = korrelationscoefficienten

gör uppg 10.2
10.7

Presentation av data

Anta att vi har följande data

ogrupperade $\underset{x_1}{51}\ \underset{x_2}{49}\ 51\ 50\ 49\ 53\ 50\ 53\ 51\ \underset{x_{10}}{51}$

Gruppindelade data

Då kan vi gruppera dem och får frekvens, relativ frekvens $p_i = \frac{f_i}{n}$

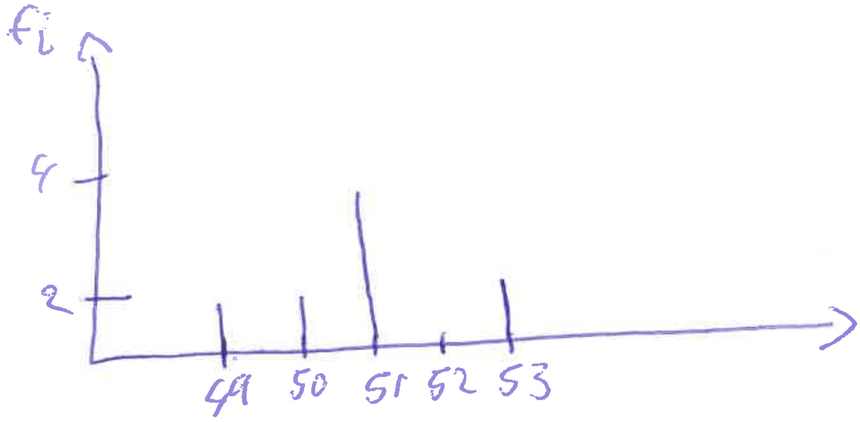
$y_1 = 49 \quad f_i = 2 \quad \frac{2}{10}$

$y_2 = 50 \quad f_i = 2 \quad \frac{2}{10}$

$y_3 = 51 \quad f_i = 4 \quad \frac{4}{10}$

$y_4 = 53 \quad f_i = 2 \quad \frac{4}{10}$

Vi presenterar dessa m.h.a stapeldiagram



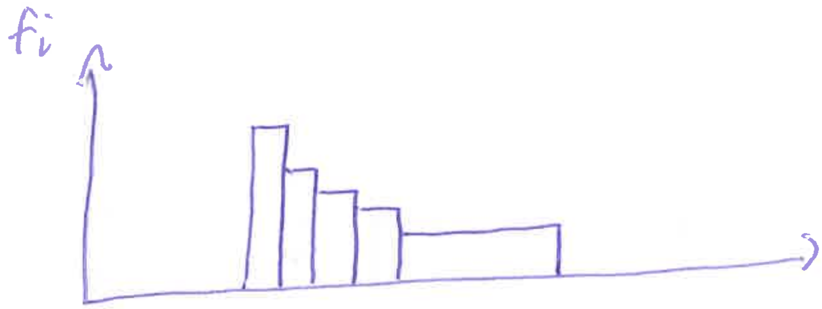
$$\bar{x} = \frac{1}{n}\sum f_i y_i \qquad s^2 = \frac{1}{n-1}\sum_{i=1}^{n} f_i (y_i - \bar{x})^2$$

Klassindelade data
Anta att vi har ~~många~~ data. Då delar vi in dessa i klasser. Där varje klass innehåller de data som ligger inom ett visst intervall. Visa ÖH sid 226-227

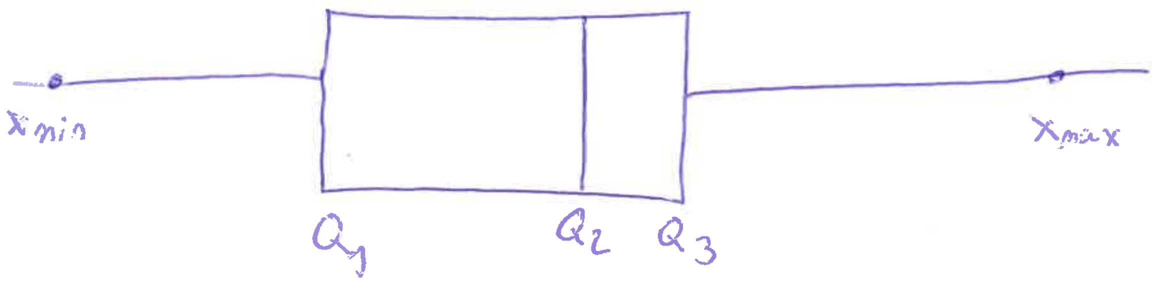
1) Vi räknar som om alla data i klassen har klassmittens värde
2) inga öppna klasser
3) klassbredden bör vara konstant

~~OBSERVERA~~

Vi ritar histogram där antalet data i klassen är proportionellt mot arean av varje rektangel



Boxplott



Q_1 är första kvartilen ; 25% av mätdata ligger till vänster

Q_2 — andra — 50% — '' —

Q_3 tredje 75% — '' —

$$Q_2 = \tilde{x}$$

$Q_3 - Q_1$ kallas kvartilavståndet

(Q_1, Q_3) kalla kvartilintervallet

$x_{max} - x_{min}$ kallas variations~~avståndet~~ bredden

(x_{min}, x_{max}) variationsintervallet

Q_1 kallas också 25% percentilen

Q_2 50%

Q_3 75% percentilen

Hur ~~ta~~ man fram Q_1

Om vi har n mätdata borde det vara det mätdata x_k

där k ~~det k som~~ uppfyller ~~...~~ $\frac{k}{n} = 0.25$

Vi väljer ~~k~~ k till det heltal som uppfyller

$$0.25n \le k \le 0.25n + 1$$

ex $n = 11$

$0.25 \cdot 11 = \frac{11}{4} = 2.75$

$0.25 \cdot 11 + 1 = 3.75$

välj k till 3

$\Rightarrow x_3 = Q_1$

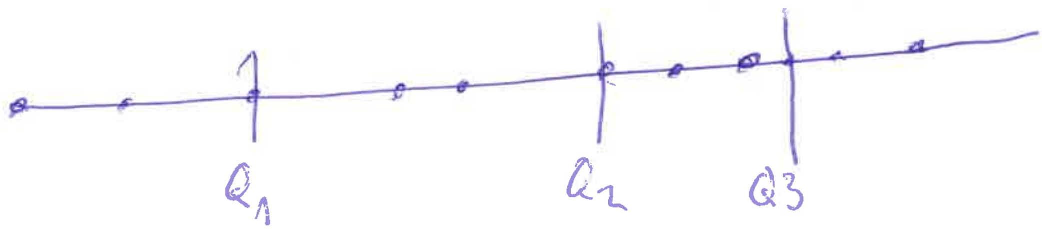
$0.5 \cdot 11 = \frac{11}{2} = 5.5$

$0.5 \cdot 11 + 1 = 6.5$

$\Rightarrow Q_2 = 6$

8.25 9.25

p.s.s $Q_3 = 9$



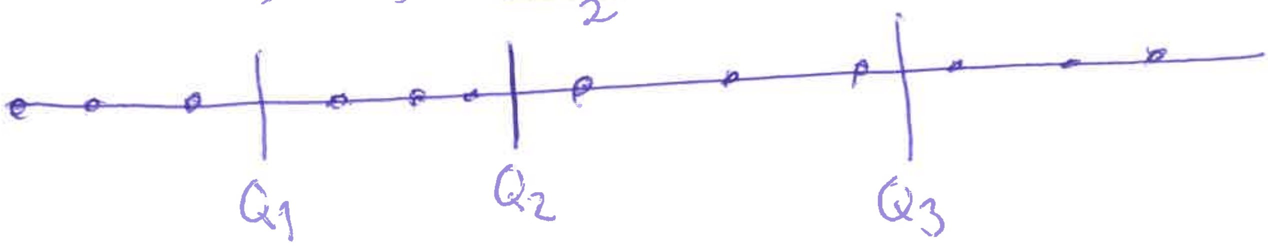
ex $n = 12$

$0.25 \cdot 12 = 3$

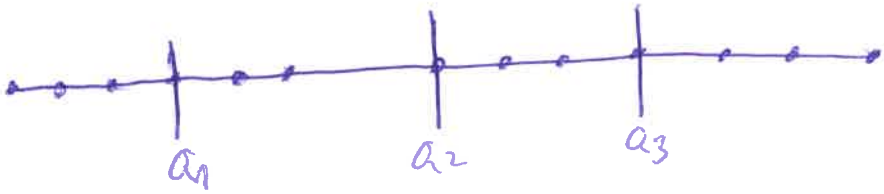
$0.25 \cdot 12 + 1 = 4$

k uppfyller både 3 och 4

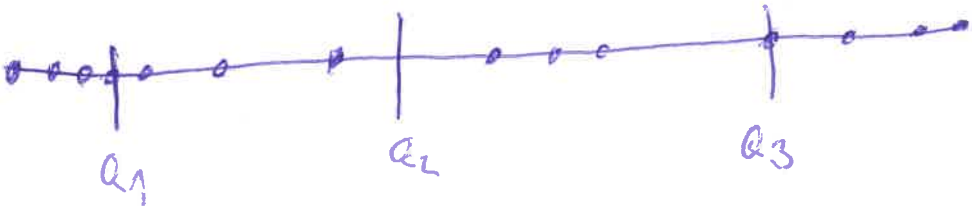
$\Rightarrow Q_3 = \frac{x_3 + x_4}{2}$



n=13



n=14



5% percentilen: Välj det x_h som uppfyller

$0.05n \le h \le 0.05n+1$

Räkna uppg 10.7

~~Linj regression~~ förs

~~Räkna uppg 11.7~~ förs

Föreläsning 9 sid 1

En skattning av ~~en parameter~~ θ kallas θ^*_{obs} och är ett utfall av den stokastiska variabeln θ^*

θ^* kallas även stickprovsvariabeln $\theta^* = \theta^*(X_1, X_2 \dots X_n)$

θ^*_{obs} — // — punktskattningen $\theta^*_{obs} = \theta^*(x_1, x_2 \dots x_n)$

Skattningen beror alltid av mätdata

ex1 $E(\bar{X}) = \mu$

$$\mu^*_{obs} = \bar{x}$$

$$\mu^* = \bar{X}$$

ex2 $E(\bar{X}) = \mu$

$$\mu^*_{obs} = \frac{5x_1 + 2x_2}{7}$$

$$\mu^* = \frac{5X_1 + 2X_2}{7}$$

ex3 $D(X) = \sigma$

$$\sigma^*_{obs} = \sqrt{\frac{1}{n-1}\sum(x_i - \bar{x})^2} = s$$

$$\sigma^* = S = \sqrt{\frac{1}{n-1}\sum(X_i - \bar{X})^2}$$

S^2 = stickprovsvariansen ~~[illegible]~~ se sid 251

Om vi inte vet vilken fördelning vi har skattas väntevärde med medelvärde och standardavvikelsen med stickprovsstandardavvikelsen

Bin-fördeln $X \in Bin(n,p)$

$$p^*_{obs} = \frac{X}{n}$$

Hyp-fördeln $X \in Hyp(N,n,p)$

$$p^*_{obs} = \frac{x}{n}$$

Po-fördeln $X \in Po(\mu)$

$$\mu^*_{obs} = \overline{X}$$

exp-fördeln $X \in exp(\lambda)$

väntevärde skattas med medelvärde

$$E[X] = \cancel{\int_0^\infty x \lambda e^{-\lambda x} dx} = \frac{1}{\lambda}$$

$$\frac{1}{\lambda^*_{obs}} = \overline{X} \qquad \lambda^*_{obs} = \frac{1}{\overline{X}}$$

ffg-fördeln $E[X] = \frac{1}{p}$ $\qquad p^*_{obs} = \frac{1}{\overline{X}}$

likformig-fördeln $U[a,b]$ knepigt 2 parametrar $\bar{x} = \frac{(a+b)}{2}$ $s^2 = \frac{(b-a)^2}{12}$

$N(\mu,\sigma)$ $\mu^*_{obs} = \bar{x}$ $\qquad \sigma^*_{obs} = s$

$$\boxed{\text{Räkna 11.14 som exempel} \Rightarrow \theta^*_{obs} = \frac{\overline{X}}{1-\overline{X}}}$$

Def En skattning är konsistent

$\Leftrightarrow$

$$\lim_{n \to \infty} P(|\theta^*_n - \theta| > \varepsilon) \to 0$$

n är stickprovsstorleken

11.14 $f_X(x) = \theta x^{\theta-1}$

$$E(X) = \int_0^1 x\theta x^{\theta-1}\,dx = \int_0^1 \theta x^{\theta}\,dx =$$

$$\theta \cdot \left[\frac{x^{\theta+1}}{\theta+1}\right]_0^1 = \theta\frac{1}{\theta+1}$$

$$\bar{X} = \frac{\theta^*_{obs}}{\theta^*_{obs}+1}$$

$$\bar{X}(\theta^*_{obs}+1) = \theta^*_{obs}$$

$$\bar{X}\cdot\theta^*_{obs} + \bar{X} = \theta^*_{obs}$$

$$\bar{X}\theta^*_{obs} - \theta^*_{obs} = -\bar{X}$$

$$\theta^*_{obs}(1-\bar{X}) = \bar{X}$$

$$\theta^*_{obs} = \frac{\bar{X}}{1-\bar{X}}$$

Maximum-likelihood-metoden

= största sannolikhetsmetoden

exp 11.10 Anta att vi vet att X_i:na är obero och $Po(\mu)$

Vi vill nu skatta μ.

Vi får 5 mätdata $x_1 = 10$ $x_2 = 12$ $x_3 = 7$ $x_4 = 10$ $x_5 = 4$

Idén är nu att eftersom det blev dessa 5 utfall borde sannolikheten för att det skulle bli dessa 5 utfall vara hög

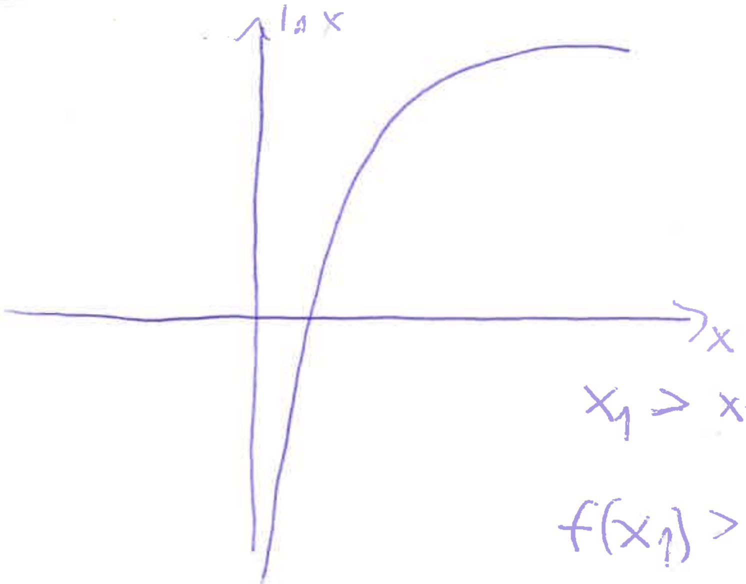
d.v.s $P(X_1 = 10 \cap X_2 = 12 \ldots) =$ ober $=$

$= p_X(10) \cdot p_X(12) \ldots =$ = borde vara så stor som möjligt ~~eftersom det som hu~~

$$= \frac{\mu^{10}}{10!} e^{-\mu} \cdot \frac{\mu^{12}}{12!} e^{-\mu} \ldots$$

Det μ som maximerar denna sannolikhet blir då maximum-likelihoodskattningen av μ.

Metod att ta fram μ



$x_1 > x_2 \Leftrightarrow \ln x_1 > \ln x_2$

$f(x_1) > f(x_2) \Leftrightarrow \ln f(x_1) > \ln f(x_2)$

där $\ln f(x)$ är störst är även $f(x)$ störst

ty ln-fknen är strängt växande och kontinuerlig

Vi har här

Likelihoodfunktionen $L(\mu) = P(X_1 = x_1 \cap X_2 = x_2 \ldots)$

$$= \frac{\mu^{x_1}}{x_1!} e^{-\mu} \cdot \frac{\mu^{x_2}}{x_2!} e^{-\mu} \ldots \frac{\mu^{x_n}}{x_n!} e^{-\mu}$$

$$\ln L(\mu) = \underline{\ln \mu^{(x_1 + \ldots x_n)}} + \ln e^{-n\mu} - \ln(x_1! \ldots x_n!)$$

$$\ln L(\mu) = \sum x_i \cdot \ln \mu - n\mu$$

$$\frac{d}{d\mu} \ln L(\mu) = \sum x_i \cdot \frac{1}{\mu} - n = 0$$

$$\mu = \frac{\sum x_i}{n} = \bar{x}$$

maximerar $L(\mu)$ =>

Maximum likelihoodskattningen av μ blir $\mu^*_{obs_{ML}} = \bar{x} = \frac{43}{5} = 8.6$

Vid diskret fördeln $L(\theta) = P_{X_1 \ldots X_n}(x_1, \ldots x_n; \theta)$

Vid kontinuerlig fördelning har vi

$L(\theta) = f_{X_1, X_2 \ldots X_n}(x_1, \ldots x_n; \theta) = f_{X_1}(x_1) \cdot f_{X_2}(x_2)$ om ober