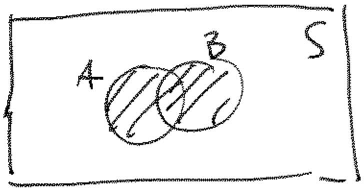
Händelser

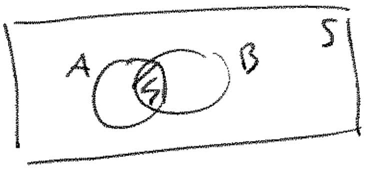
Ett <u>utfall</u> är en specifik observation av ett experiment. <u>Utfallsrummet</u> (S eller Ω) är mängden av alla utfall. En <u>händelse</u> (A, B, C, etc) är en samling utfall.

<u>Viktiga begrepp.</u>

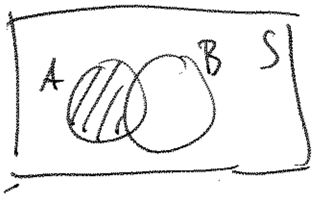
i) $A \cup B$ = "A union B" = "A eller B eller båda".



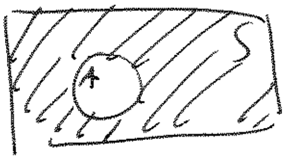
ii) $A \cap B$ = "A snitt B" = "A <u>och</u> B"



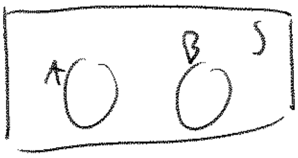
iii) $A \setminus B$ = "A men inte B"



iv) A^c = "A-komplement" = "A händer ej"



v) A och B kallas <u>disjunkta</u> om $A \cap B = \emptyset$, dvs om $A \cap B$ inte innehåller några utfall.



Räkneregler för sannolikheter

1) Om alla utfall i S är lika sannolika så gäller att

$$P(A) = \frac{\#\text{ utfall i } A}{\#\text{ utfall totalt}} = \frac{|A|}{|S|}$$

2) $P(A \cup B) = P(A) + P(B) - P(A \cap B)$

3) $P(A^c) = 1 - P(A)$

4) $0 \leq P(A) \leq 1$ för varje händelse $A \subseteq S$

5) $P(S) = 1$

Kombinatorik

Antal sätt man kan ordna n objekt: $n! = n \cdot (n-1) \cdot \dots \cdot 3 \cdot 2 \cdot 1$

Antal sätt att välja ut k objekt ur en samling av n objekt:

$$\binom{n}{k} = \frac{n!}{k!(n-k)!}$$

Betingade sannolikheter

Den betingade sannolikheten för A givet B betecknas med $P(A|B)$ och ges av

$$P(A|B) = \frac{P(A \cap B)}{P(B)}.$$

Detta tolkas som "om vi vet att B har hänt, vad är då sannolikheten att A händer?"

Bayes lag: $$P(A|B) = \frac{P(B|A)\,P(A)}{P(B)}$$

Vi säger att två händelser A och B är oberoende om

$P(A|B) = P(A)$ eller $P(B|A) = P(B)$

eller $P(A \cap B) = P(A)\,P(B)$.

Slumpvariabler

Diskreta slumpvariabler antar uppräkneligt antal värden, dvs värden som kan skrivas upp i en lista: $\{x_1, x_2, \ldots\}$

Kontinuerliga slumpvariabler antar värden inom ett intervall, t.ex. $[0,1]$, $[0,\infty)$, $(-\infty,\infty)$.

En diskreta slumpvariabels <u>fördelning</u> bestäms av dess <u>sannolikhetsfunktion</u> $P(X=k)$.

En sannolikhetsfunktion uppfyller alltid:

1. $P(X=k) \geq 0$ för alla k

2. $\sum P(X=k) = 1$ (summa över alla möjliga värden på k)

Några vanliga diskreta fördelningar:

<u>Likformig diskret fördelning</u>

Möjliga värden: $\{x_1, x_2, \ldots, x_n\}$

$P(X=x_k) = \frac{1}{n}$ för alla k

<u>Binomialfördelning</u>

Möjliga värden: $\{0, 1, 2, \ldots, n\}$

$P(X=k) = \binom{n}{k} p^k (1-p)^{n-k}$

Vi skriver $X \sim Bin(n,p)$ (n och p parametrar)

<u>Poissonfördelning</u>

Möjliga värden: $\{0, 1, 2, \ldots\}$

$P(X=k) = \frac{\lambda^k}{k!} e^{-\lambda}$

Vi skriver $X \sim Poi(\lambda)$, λ parameter

En kontinuerlig slumpvariabels <u>fördelning</u> bestäms av dess <u>täthetsfunktion</u> $f(x)$.

En täthetsfunktion uppfyller alltid:

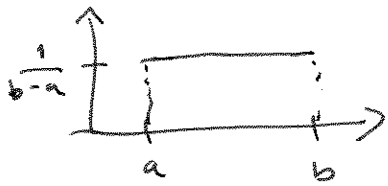
1: $f(x) \geq 0$ för alla x

2. $\int_{-\infty}^{\infty} f(x)\,dx = 1$

<u>Likformig</u> <u>kontinuerlig</u> <u>fördelning</u>

Möjliga värden i ett intervall $[a, b]$

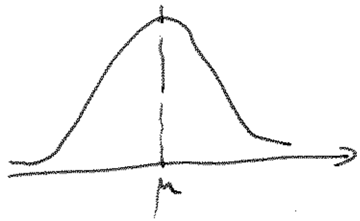
$f(x) = \frac{1}{b-a}$ för alla x



<u>Normalfördelning</u>

Möjliga värden i intervallet $(-\infty, \infty)$

$$f(x) = \frac{1}{\sqrt{2\pi\sigma^2}}\, e^{-\frac{(x-\mu)^2}{2\sigma^2}}$$



Vi skriver $X \sim N(\mu, \sigma^2)$, μ och σ^2 parametrar

~~Poissonförd~~

<u>Väntevärde och varians</u>

<u>Väntevärdet</u> av en slumpvariabel är det värdet vi kan förvänta oss i genomsnitt.

Om X diskret: $E(X) = \sum x P(X = x)$

Om X kontinuerlig: $E(X) = \int_{-\infty}^{\infty} x f(x) dx$

Två räkneregler:

1. $E(X + Y) = E(X) + E(Y)$
2. $E(aX) = aE(X)$ för en konstant a.

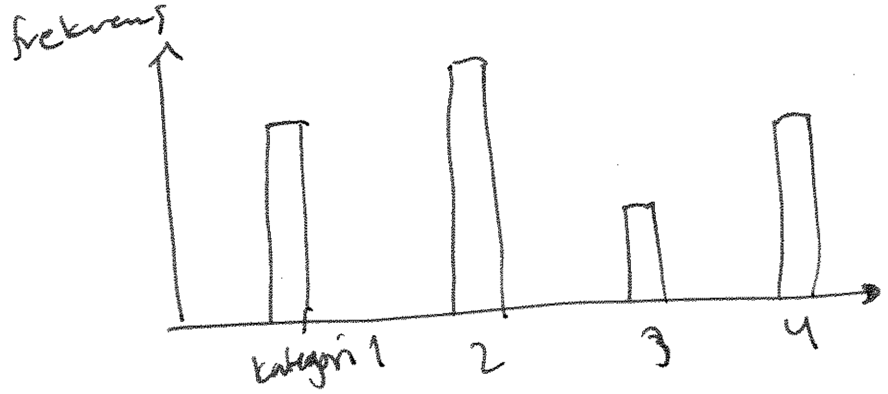
<u>Variansen</u> av en slumpvariabel är ett mått på hur mycket variation vi kan förvänta oss att se.

$$Var(X) = E\left[(X - E(X))^2\right]$$

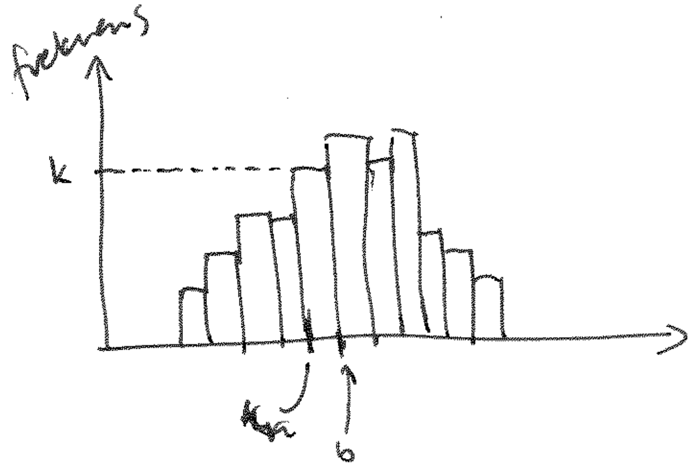
Räkneregel: $Var(X) = E(X^2) - E(X)^2$

<u>Beskrivande statistik</u>

<u>Stapeldiagram</u> är ett bra sätt att beskriva kategorisk (kvalitativ) eller diskret data.

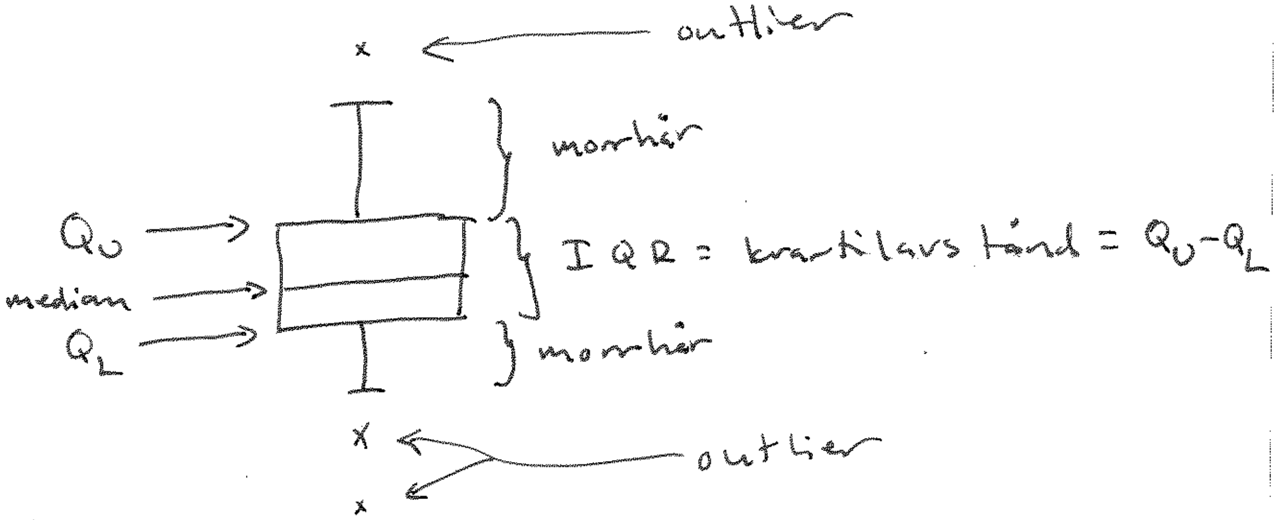


Histogram är ett bra sätt att beskriva kontinuerlig data:



Att stapeln mellan värdena a och b har höjd k betyder att vi har k observationer i intervallet [a, b].

Boxplots är ett annat bra sätt att beskriva kontinuerlig data:



~~Q_L = det värde som gör att~~

Q_L = den nedre kvartilen =

= det värde som är större än exakt 25% av datan

Q_U = den övre kvartilen =

= det värde som är större än exakt 75% av datan

median = det värde som är större än exakt 50% av datan

Punktskattningar

En punktskattning är vår ~~...~~ gissning av värdet på en viss parameter utifrån ett stickprov, $X_1, X_2, \ldots, X_n$

Punktskattning	Parameter
$\bar{X}$	μ = väntevärde
S^2	σ^2 = varians
$\hat{p}$	p = proportion

$$\bar{X} = \frac{1}{n}\sum_{k=1}^{n} X_k \,, \quad S^2 = \frac{1}{n-1}\sum_{k=1}^{n}(\bar{X}-X_k)^2$$

$$\hat{p} = \frac{\text{\# observationer med egenskapen}}{\text{\# observationer totalt}}$$

En punktskattare är väntevärdesriktig om väntevärdet av punktskattaren är lika med ~~...~~ parametern vi vill skatta.

Fördelning för $\bar{X}$

Sats: Om $X_1, X_2, \ldots, X_n$ är ett stickprov från en $N(\mu, \sigma^2)$-fördelning så gäller att $\bar{X}$ är $N(\mu, \frac{\sigma^2}{n})$-fördelad

Sats: (CGS) Om $X_1, X_2, \ldots, X_n$ är ett stickprov från någon fördelning med väntevärde μ och varians σ^2 så är $\bar{X} \sim N(\mu, \frac{\sigma^2}{n})$ om n är stort $(n \geq 30)$

Intervallskattning — konfidensintervall

Ett $(1-\alpha)\%$-igt KI för μ ges av:

Om σ^2 känd och n stort eller populationen normalförd.

$$L = \bar{X} - z_{\frac{\alpha}{2}} \cdot \frac{\sigma}{\sqrt{n}}$$

$$U = \bar{X} + z_{\frac{\alpha}{2}} \cdot \frac{\sigma}{\sqrt{n}}$$

Om σ^2 okänd:

Om n stort

$$L = \bar{X} - z_{\frac{\alpha}{2}} \frac{\sigma}{\sqrt{n}}$$

$$U = \bar{X} + z_{\frac{\alpha}{2}} \frac{\sigma}{\sqrt{n}}$$

Om n litet men pop. normalf.

$$L = \bar{X} - z_{\frac{\alpha}{2}} \cdot \frac{\sigma}{\sqrt{n}}$$

$$U = \bar{X} + z_{\frac{\alpha}{2}} \cdot \frac{\sigma}{\sqrt{n}}$$

Ett $(1-\alpha)\%$-igt KI för en proportion p ges av

$$L = \hat{p} - z_{\frac{\alpha}{2}} \cdot \sqrt{\frac{\hat{p}(1-\hat{p})}{n}}$$

$$U = \hat{p} + z_{\frac{\alpha}{2}} \cdot \sqrt{\frac{\hat{p}(1-\hat{p})}{n}}$$

om n är stort. $(n\hat{p} \geq 15$ och $n(1-\hat{p}) \geq 15)$

Hypotestest

Hypoteser:

Väntevärdestest

$$\left.\begin{array}{l} H_0: \mu = \mu_0 \\ H_1: \mu \neq \mu_0 \end{array}\right\} \text{dubbel-sidigt}$$

$$\left.\begin{array}{l} H_0: \mu \geq \mu_0 \\ H_1: \mu < \mu_0 \\ \\ H_0: \mu \leq \mu_0 \\ H_1: \mu > \mu_0 \end{array}\right\} \text{enkel-sidigt}$$

Om vi känner till σ^2 använder vi teststatistikan

$$\frac{\bar{X} - \mu_0}{\sigma^2/\sqrt{n}}$$

som är N(0,1)-fördelad om n är stort ($n \geq 30$) eller populationen normalfördelad

Om σ^2 är okänd använder vi

$$\frac{\bar{X} - \mu_0}{s/\sqrt{n}}$$

som är N(0,1)-fördelad om n är stort och t_{n-1}-fördelad om n är litet och populationen normalfördelad.

Proportionstest

$$\left\{\begin{array}{l} H_0: p = p_0 \\ H_1: p \neq p_0 \end{array}\right.$$

$$\left\{\begin{array}{l} H_0: p \geq p_0 \\ H_1: p < p_0 \\ \\ H_0: p \leq p_0 \\ H_1: p > p_0 \end{array}\right.$$

Vi använder teststatistikan

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

som är N(0,1)-fördelad om n är stort,

$np_0 \geq 15$, $n(1-p_0) \geq 15$

P-värde

Om z är den observerade teststatistikan och Z är en $N(0,1)$-fördelad variabel så ges p-värdet av

$$p\text{-värde} = P(|Z| \geq z)$$ om testet är tvåsidigt

$$p\text{-värde} = P(Z \geq z)$$ om $H_1: \mu > \mu_0$

$$p\text{-värde} = P(Z \leq z)$$ om $H_1: \mu < \mu_0$

p-värde $< \alpha \iff z$ i den kritiska regionen.

Typer av fel

Ett typ-I-fel är att förkasta en sann nollhypotes.

Ett typ-II-fel är att inte förkasta en falsk nollhypotes.

$$\alpha = P(\text{typ-I-fel})$$

$$\beta = P(\text{typ-II-fel})$$

styrkan av testet $= 1 - \beta = P(\text{förkasta falsk nollhypotes})$