Matematisk statistik F/Kf

Övningar 1999

47 sidor

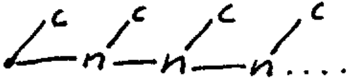
15:-

Övning 1

7 En hemdator är kopplad till en stordator via telefonmodem. Hemdatorn kommer att ringa upp stordatorn tills den får kontakt, då slutar den ringa.

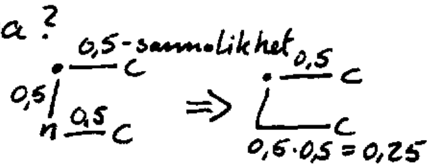
$c = \{\text{kontakt}\}$, $n = \{\text{ingen kontakt}\}$

(a) Konstruera träddiagram



(b) Är alla vägar lika troliga?

Nej.



(c) Utfallsrummet

$\Omega = \{c, nc, nnc, nnnc, \ldots\}$

(d) $A = \{\text{kontakt på högst 4 försök}\} = \{c, nc, nnc, nnnc\}$

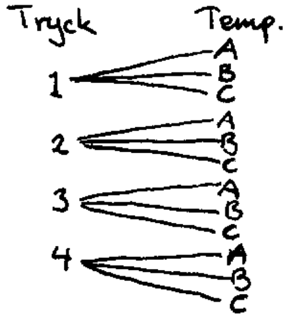
(e) 2 disjunkta händelser

$A = \{c\}$

$B = \{nc\}$ $\Rightarrow A \cap B = 0$

10 Man undersöker en gaslag genom att göra experiment vid 4 olika tryck och 3 olika temperaturer.

(a) Hur många experimentsituationer skall studeras?



$\Rightarrow 4 \cdot 3 = \underline{\underline{12}}$

(b) Varje exp. situation upprepas 5 ggr. $\Rightarrow 5 \cdot 12 = \underline{\underline{60}}$

(c) Vi testar 6 gaser $\Rightarrow 6 \cdot 60 = \underline{\underline{360}}$

15 Man testar 5 olika beläggningar som används för att skydda fiberoptikledare från kyla. Dessa görs i slumpmässig ordning.

(a) Hur många ordningar kan testen göras?

$5 \cdot 4 \cdot 3 \cdot 2 \cdot 1 = 5! = 120$

(b) Om 2 av beläggningarna har tillverkats av samma tillverkare, vad är då sannolikheten att dessa kommer testas i rad.

$P(\text{de 2 testas i rad}) = \dfrac{\text{Antalet ordningar där de 2 testas i rad}}{\text{Totala antalet ordningar}} =$

$= \dfrac{4! \cdot 2}{120} = \dfrac{48}{120} = 0{,}4$

<u>21</u> En firma anställer 10 programmerare, 8 systemanalyserare, 4 dataingenjörer, 3 statistiker. Man skall av dem välja ut ett team bestående av 3pr, 2Sa, 2Di, 1st.

(a) På hur många sätt kan detta team väljas?

Antalet sätt att välja 3 pr ur 10 $= \binom{10}{3} = \dfrac{10!}{3! \cdot 7!} = 120$

——— // ——— 2 sa ur 8 $= \binom{8}{2} = \dfrac{8!}{2!6!} = 28$

——— // ——— 2 di ur 4 $= \binom{4}{2} = \dfrac{4!}{2!2!} = 6$

——— // ——— 1 st ur 3 $= \binom{3}{1} = \dfrac{3!}{1!2!} = 3$

$\Rightarrow$ antalet sätt $= 120 \cdot 28 \cdot 6 \cdot 3 = 60480$

(b) Om kunden kräver att en viss di skall vara med. Hur många sätt?

Då skall vi välja 1 di ur de 3 återstående $= \binom{3}{1} = 3$

$\Rightarrow 120 \cdot 28 \cdot 3 \cdot 3 = 30240$

<u>24</u> Man sänder ut ett 128-bitars meddelande. Varje objekt av de 128 sänds antingen korrekt eller inkorrekt. Hur många olika meddelande kan uppkomma med exakt 2 fel?

= antalet sätt man kan välja 2 objekt ur 128 möjligheter.

$= \binom{128}{2} = \dfrac{128!}{2!\,126!} = 8128$

<u>26</u> En garagedörröppnare har 6 vippströmbrytare, var och en med 3 nivåer, upp, mitten, ner

(a) På hur många sätt kan de 6 vara inställda?

Antalet sätt

$3 \cdot 3 \cdot 3 \cdot 3 \cdot 3 \cdot 3 = 3^6 = 729$

(b) Vad är sannolikheten att en tjuv gissar rätt på första försöket?

$P(\text{gissa rätt}) = \frac{1}{729} \approx 0{,}0014$

(C) Hur många inställningar är möjliga med 2 stycken upp, 2 ner, 2 mitten

Antalet sätt $2 \text{ upp} = \binom{6}{2} = \frac{6!}{2!4!} = 15$

$2 \text{ mitten} = \binom{4}{2} = \frac{4!}{2!2!} = 6$

$2 \text{ ner} = \binom{2}{2} = \frac{2!}{0!2!} = 1$

sammanlagda inställningar: $15 \cdot 6 \cdot 1 = 90$

Övning 2.

2.3 4 blodgrupper: A, B, AB, 0

$P(A) = 0{,}41$, $P(B) = 0{,}09$, $P(AB) = 0{,}04$, $P(0) = 0{,}46$

P(Slumpvis vald individ har antigen A) $= P(A \cup AB) = P(A) + P(AB) =$

$= 0{,}41 + 0{,}04 = \underline{\underline{0{,}45}}$

P(Slumpvis vald individ har antigen B) $= P(B \cup AB) = P(B) + P(AB) =$

$= 0{,}09 + 0{,}04 = \underline{\underline{0{,}13}}$

P(Slumpvis vald individ varken antigen A eller B) $= P(0) = 0{,}46$

2.4 2 motorer är parallellkopplade  . Huvudmotorn har 95% tillförlitlighet. Back-upmotorn har 80% tillförlitlighet. Hela komponenten har 99% tillförlitlighet.

Låt A = {huvudmotorn fungerar} $\Rightarrow$ $P(A) = 0{,}95$

B = {back-upmotorn fungerar} $P(B) = 0{,}80$

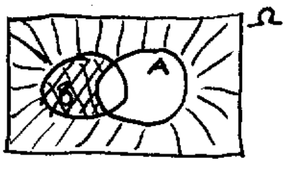
$P(A \cup B) = 0{,}99$

$P(A \cap B) = P(A) + P(B) - P(A \cup B) = 0{,}95 + 0{,}80 - 0{,}99 = \underline{\underline{0{,}76}}$

#: $P(A^c \cap B) = P(B) - P(A \cap B) = 0{,}80 - 0{,}76 = \underline{\underline{0{,}04}}$

$P(A \cap B^c) = P(A) - P(A \cap B) = 0{,}95 - 0{,}76 = \underline{\underline{0{,}19}}$

☼: $P(A^c \cap B^c) = 1 - P(A \cup B) = 1 - 0{,}99 = \underline{\underline{0{,}01}}$



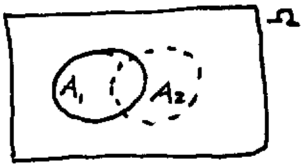
2.14 P(back-upen fungerar | huvudmotorn sönder) $= P(B|A^c) =$

$$= \frac{P(B\cap A^c)}{P(A^c)} = \frac{0{,}04}{1-P(A)} = \frac{0{,}04}{1-0{,}95} = \frac{0{,}04}{0{,}05} = 0{,}80$$

$P(B|A^c) = P(B)$ – (De 2 motorerna fungerar oberoende av varandra därför påverkar inte faktumet att huvudmotorn fungerar eller ej, sannolikheten att back-upen fungerar)

2.11 Bevisa additionsregeln

$$P(A_1 \cup A_2) = P(A_1) + P(A_2) - P(A_1\cap A_2)$$



$$A_1 = (A_1\cap A_2^c)\cup(A_1\cap A_2)$$
$$A_2 = (A_1^c\cap A_2)\cup(A_1\cap A_2)$$
$$A_1\cup A_2 = (A_1\cap A_2^c)\cup(A_1^c\cap A_2)\cup(A_1\cap A_2)$$
$$P(A_1\cup A_2) = P\big(\underbrace{(A_1\cap A_2^c)}\cup\underbrace{(A_1^c\cap A_2)}\cup\underbrace{(A_1\cap A_2)}\big) =$$

de är disjunkta

$$= P(A_1\cap A_2^c) + P(A_1^c\cap A_2) + P(A_1\cap A_2) = \underbrace{P(A_1\cap A_2) + P(A_1\cap A_2^c)} +$$
$$+ \underbrace{P(A_1^c\cap A_2) + P(A_1\cap A_2)} - P(A_1\cap A_2) =$$
$$= P((A_1\cap A_2)\cup(A_1\cap A_2^c)) + P((A_1^c\cap A_2)\cup(A_1\cap A_2)) - P(A_1\cap A_2) =$$
$$= P(A_1) + P(A_2) - P(A_1\cap A_2)$$ V.S.V.

2.23 Ett tecken på koppar i marken är en mynta med lila blommor. Antag för ett visst område att det är 30% chans att jorden har hög kopparhalt och 23% chans att myntan finns där. Vidare gäller att om kopparhalten är hög så är myntan närvarande med sannolikheten 70%.

(a) P(Mynta och hög koppar halt)?

(b) P(Hög kopparhalt, givet att myntan finns)?

Låt $K = \{$hög koppar halt$\}$, $M = \{$myntan finns$\}$

$P(K) = 0{,}30$, $P(M) = 0{,}23$, $P(M|K) = 0{,}70$

(a) $P(M\cap K)$? $\quad P(M|K) = \dfrac{P(M\cap K)}{P(K)} \Rightarrow P(M\cap K) = P(M|K)\cdot P(k) =$

$$= 0{,}70\cdot 0{,}30 = \underline{\underline{0{,}21}}$$

(b) $$P(K|M) = \frac{P(K\cap M)}{P(M)} = \frac{0{,}21}{0{,}23} \approx \underline{\underline{0{,}91}}$$

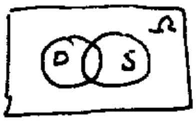
2.36 50% av alla datorchips är defekta. Inspektion försäkrar att endast 5% av chipsen som säljs lagligt är defekta. Olyckligtvis stjäls vissa chips innan inspektion. Om 1% av alla chips på marknaden är stulna, vad är då sannolikheten att ett chips är stulet, givet att det är defekt?

Lösning Låt $D=\{\text{defekt}\}$ $\quad S=\{\text{stulet, oinspekterat}\}$

$$P(S)=0{,}01 \quad , \; P(D|S)=0{,}5 \quad , \; P(D|S^c)=0{,}05$$

$$P(S|D)=\frac{P(S\cap D)}{P(D)} \; ?$$

$$P(D|S)=\frac{P(D\cap S)}{P(S)} \Rightarrow P(D\cap S)=P(D|S)P(S)=0{,}5\cdot 0{,}01=0{,}005$$



$$D=(D\cap S)\cup(D\cap S^c)$$

$$P(D)=P\big((D\cap S)\cup(D\cap S^c)\big)=P(D\cap S)+P(D\cap S^c)=$$
$$=P(D|S)P(S)+P(D|S^c)P(S^c)=0{,}005+0{,}05\cdot 0{,}99=0{,}0545$$
$$\Rightarrow P(S|D)=\frac{0{,}005}{0{,}0545}\approx 0{,}092$$

Övning 3

3.8 X = antal hål som måste borras.

X	1	2	3	4	5	6	7	8
$f(x)$	0,02	0,03	0,05	0,2	0,4	0,2	0,07	?
$F(x)$	0,02	0,05	0,1	0,3	0,7	0,9	0,97	1

(a) Vad är $f(8)$?

$$f(8)=P(X=8)$$
$$\sum_{x=1}^{8} f(x)=1 \Rightarrow f(8)=1-\sum_{x=1}^{7} f(x)=1-0{,}02-0{,}03-\ldots-0{,}07=0{,}03$$

(b) $F(x)=P(X\leq x)=\sum_{y=1}^{x} f(y)$

(c) $P(3\leq X\leq 5)=P(X\leq 5)-P(X\leq 2)=F(5)-F(2)=0{,}7-0{,}05=\underline{0{,}65}$

(d) $P(X\leq 4)=F(4)=0{,}3$ $\quad , \quad P(X<4)=P(X\leq 3)=0{,}1$

(e) $\left.\begin{array}{l} F(-3)=P(X\leq -3)=0 \\ F(10)=P(X\leq 10)=1 \end{array}\right\}$ antar ju bara värden mellan 1 & 8!!!

3.15 (a) $E[X] = \sum_{x=1}^{8} x f(x) = 1 \cdot 0{,}02 + 2 \cdot 0{,}03 + \ldots + 8 \cdot 0{,}03 = \underline{4{,}96}$

$E[X^2] = \sum_{x=1}^{8} x^2 f(x) = 1^2 0{,}02 + 2^2 \cdot 0{,}03 + \ldots + 8^2 0{,}03 = 26{,}34$

(b) $Var(X) = E[(X-\mu)^2] = E[X^2] - [E[X]]^2 = 26{,}34 - 4{,}96^2 = 1{,}7384$

Standardavvikelse $= \sigma = \sqrt{Var(X)} = \sqrt{1{,}7384} \approx 1{,}3185$

3.10 Sannolikheten att kunna logga in på en dator från avlägsen terminal är 0,7 X = antalet försök som krävs

(a) $f(1) = P(X=1) = 0{,}7$

$f(2) = P(X=2) = 0{,}3 \cdot 0{,}7 = 0{,}21$

$f(3) = P(X=3) = 0{,}3 \cdot 0{,}3 \cdot 0{,}7 = 0{,}063$

$f(4) = P(X=4) = 0{,}3 \cdot 0{,}3 \cdot 0{,}3 \cdot 0{,}7 = 0{,}0189$

(b) $f(x) = P(X=x) = 0{,}3^{x-1} \cdot 0{,}7 \quad , x = 1, 2, 3, \ldots$

(c) $f(6) = P(X=6) = 0{,}3^{6-1} \cdot 0{,}7 = 0{,}0017$

(d) $F(x) = P(X \leq x) = \sum_{y=1}^{x} f(y) = \sum_{y=1}^{x} 0{,}3^{y-1} \cdot 0{,}7 =$ $\left[\sum_{k=1}^{n} a r^{k-1} = \frac{a(1-r^n)}{1-r} \; , r \neq 1 \right]$

$= \frac{0{,}7(1-0{,}3^x)}{1-0{,}3} = 1 - 0{,}3^x$

(e) $P(X \leq 4) = F(4) = 1 - 0{,}3^4 = 1 - 0{,}0081 = 0{,}9919$

(f) $P(X \geq 5) = 1 - P(X \leq 4) = 0{,}0081$

3.24 Sannolikheten att finna olja då man borrar hål är $1/13$. Man borrar tills man hittar något. Låt X = antalet ggr. man måste borra

(a) $X \sim Geo(p = \frac{1}{13})$

(b) $f(x) = P(X=x) = \left(\frac{12}{13}\right)^{x-1} \left(\frac{1}{13}\right) \quad x_1 = 1, 2, \ldots$

(c) $m_X(t) = E[e^{tX}] = \frac{pe^t}{1-(1-p)e^t} = \frac{\left(\frac{1}{13}\right)e^t}{1-\frac{12}{13}e^t}$

(d) $E[X] = \frac{1}{p} = 13$

$Var(X) = \frac{1-p}{p^2} = \frac{12/13}{(1/13)^2} = 156 \Rightarrow \sigma = \sqrt{Var(X)} = \sqrt{156} \approx 12{,}5$

$E[X^2] = Var(X) + (E[X])^2 = 325$

(e) $F(x) = P(X \leq x) = 1 - (1-p)^x = 1 - \left(\frac{12}{13}\right)^x$

$\Rightarrow P(X \geq 2) = 1 - P(X \leq 1) = 1 - F(1) = 1 - \left(1 - \left(\frac{12}{13}\right)\right) = 12/13$

(f) 10 hål har borrats utan lycka. Vad är sannolikheten att minst 2 hål till måste borras.

$$P(X \geqslant 12 \mid X \geqslant 11) = \frac{P(X \geqslant 12 \cap X \geqslant 11)}{P(X \geqslant 11)} = \frac{P(X \geqslant 12)}{P(X \geqslant 11)} =$$

$$= \frac{1 - P(X \leq 11)}{1 - P(X \leq 10)} = \frac{1 - F(11)}{1 - F(10)} = \frac{(^{12}/_{13})^{11}}{(^{12}/_{13})^{10}} = \frac{12}{13}$$

3.32 En diskret s.v. X har momentg. funktion $m_X(t) = E[e^{tX}] = e^{2(e^t-1)}$

Sats 3.4.2. $\left.\frac{\partial^k m_X(t)}{\partial t^k}\right|_{t=0} = E[X^k]$

$$\Rightarrow E[X] = \left.\frac{\partial(e^{2(e^t-1)})}{\partial t}\right|_{t=0} = \left.(2e^t e^{2(e^t-1)})\right|_{t=0} = 2$$

$$E[X^2] = \left.\frac{\partial^2(e^{2(e^t-1)})}{\partial t^2}\right|_{t=0} = \left.(2e^t e^{2(e^t-1)} + 4e^{2t} e^{2(e^t-1)})\right|_{t=0} = 6$$

$$Var(X) = E[X^2] - (E[X])^2 = 6 - 2^2 = 2$$

$$\sigma = \sqrt{Var(X)} = \sqrt{2} = 1{,}4142$$

3.33 $X \sim Geo(p)$

(a) Visa att $P(X \text{ udda}) = \frac{p}{1-q^2}$, där $q = 1-p$

$$f(x) = P(X=x) = (1-p)^{x-1}p = q^{x-1}p \quad , \; x = 1, 2, \ldots$$

$$P(X \text{ udda}) = P(X=1) + P(X=3) + P(X=5) + \ldots = \sum_{m=1}^{\infty} P(X=2m-1) =$$

$$= \sum_{m=1}^{\infty} q^{2m-1-1} p = \sum_{m=1}^{\infty} q^{2(m-1)} p = \left\{ \sum_{k=1}^{n} ar^{k-1} = \frac{a(1-r^n)}{1-r} \; , r \neq 1 \right\} =$$

$$= \sum_{m=1}^{\infty} (q^2)^{m-1} p = \frac{P(1-(q^2)^{\infty})}{1-q^2} = \frac{P}{1-q^2}$$

(b) Visa att $P(X \text{ udda}) \neq \frac{1}{2}$

Antag $P(X \text{ udda}) = \frac{1}{2} \Leftrightarrow \frac{P}{1-q^2} = \frac{1}{2} \Leftrightarrow 2p = 1 - q^2$

Men $1 - q^2 = 1 - (1-p)^2 = 1 - (1 - 2p + p^2) = 2p + p^2 \neq 2p$

$\Rightarrow$ motsägelse $\Rightarrow P(\text{udda}) \neq \frac{1}{2}$

Övning 4.

$X \sim Poi(\lambda_1)$

$y \sim Poi(\lambda_2)$, X, y ober. $\Rightarrow X + y \sim Poi(\lambda_1 + \lambda_2)$

3.40 Man testar 2 olika kameror 15 par av fotografier (en av varje). En opartisk domare skall välja det bästa fotot från varje par. Låt X = antalet valda från 35 mm-kameran

(a) Om kamrorna är lika bra och domaren därför väljer slumpm. vad är då $E[X]$?

$X \sim Bin(n=15, p=0{,}5)$

$\Rightarrow E[X] = np = 15 \cdot 0{,}5 = \underline{\underline{7{,}5}}$

(b) Skulle du bli förvånad om han väljer ≥ 12 foton från 35 mm-kameran?

Om domaren väljer slumpm.

$\Rightarrow P(X \geq 12) = P(X=12) + P(X=13) + P(X=14) + P(X=15) =$

$= \sum_{x=12}^{15} P(X=x) = \sum_{x=12}^{15} \binom{15}{x} 0{,}5^x (1-0{,}5)^{15-x} = \sum_{x=12}^{15} \binom{15}{x} 0{,}5^{15} \approx 0{,}0176$

3.41 80% av alla skrivare fungerar utan justering. En viss säljare säljer 10 stycken under en månad.

X = antalet fungerande skrivare (av de 10)

$\Rightarrow X \sim Bin(n=10, p=0{,}8)$

$f(x) = P(X=x) = \binom{10}{x} 0{,}8^x 0{,}2^{10-x}$, $x = 0, 1, \ldots, 8$

(a) $P(X \geq 9) = P(X=9) + P(X=10) = \binom{10}{9} 0{,}8^9 0{,}2^1 + \binom{10}{10} 0{,}8^{10} \cdot 0{,}2^0 = 0{,}3758$

(b) Antag att 10 skrivare säljes/månad under 5 månader. Vad är slh att vi har minst 9 fungerande varje månad?

y = antalet månader med minst 9 fungerande

$\Rightarrow y \sim Bin(n=5, p=0{,}3758)$

$\Rightarrow f(y) = \binom{5}{y} 0{,}3758^y \cdot 0{,}6242^{5-y}$

$P(y=5) = \binom{5}{5} 0{,}3758^5 \cdot 0{,}6242^0 = 0{,}0075$

3.55 $X \sim$ Hypergeometrisk $(N=20, n=5, r=3)$

Möjliga värden för X? $E[X]$? $Var(X)$?

Lösning:

$$\begin{array}{ccc} \max[0, n-(N-r)] & \leq X \leq & \min(n,r) \\ \shortparallel & & \shortparallel \\ \max[0, 5-(20-3))] & & \min(5,3) \\ \shortparallel & & \shortparallel \\ \max[0, -12] & & 3 \\ \shortparallel & & \\ 0 & & \end{array}$$

$\Rightarrow X \in [0,1,2,3]$

$$E[X] = n\left(\frac{r}{N}\right) = 5 \cdot \frac{3}{20} = 0{,}75$$

$$Var(X) = n\left(\frac{r}{N}\right)\left(\frac{N-r}{N}\right)\left(\frac{N-n}{N-1}\right) = 5 \cdot \frac{3}{20} \cdot \frac{17}{20} \cdot \frac{15}{19} = 0{,}5033$$

3.58 15 bilar tillverkas per timme. Under en given timme tillverkas 4 bilar med felaktiga dörrar. 3 bilar väljs slumpm. ut för inspektion.

Låt X = antalet inspekterade bilar med felaktiga dörrar

(a) $f(x)$?

Antalet sätt man kan välja 3 bilar av 15 $= \binom{15}{3}$

———— " ———— så att x har felaktiga dörrar $= \binom{4}{x}\binom{15-4}{3-x} = \binom{4}{x}\binom{11}{3-x}$

$$\Rightarrow f(x) = P(X=x) = \frac{\binom{4}{x}\binom{11}{3-x}}{\binom{15}{3}}, \quad x = 0,1,2,3$$

$\Rightarrow X \sim$ Hyperg. $(N=15, n=3, r=4)$

(b) $$E[X] = n\left(\frac{r}{N}\right) = 3 \cdot \frac{4}{15} = 0{,}8$$

$$Var(X) = n\left(\frac{r}{N}\right)\left(\frac{N-r}{N}\right)\left(\frac{N-n}{N-1}\right) = 3 \cdot \frac{4}{15} \cdot \frac{11}{15} \cdot \frac{12}{14} \approx 0{,}5029$$

(c) $$P(X \leq 1) = P(X=0) + P(X=1) = \frac{\binom{4}{0}\binom{11}{3}}{\binom{15}{3}} + \frac{\binom{4}{1}\binom{11}{2}}{\binom{15}{3}} = 0{,}8462$$

3.62 Ett visst kärnkraftverk avger en upptäckbar mängd radioaktiva gaser 2ggr./månad i genomsnitt.

(i) Vad är slh. för högst 4 utsläpp under 1 månad?

(ii) Vad är förv. antalet utsläpp under 3 mån.?

(iii) – räkna själva!

Lösning: Låt X_s = antalet utsläpp under S månader.

$X_1 \sim Poi(2)$, $X_2 \sim Poi(4)$ $X_s \sim Poi(2s)$

$E[X_s] = 2s$

(i) $P(X_1 \leq 4)$, $X_1 \sim Poi(2) \Rightarrow P(X_1 = x) = \frac{e^{-2}2^x}{x!}$

$$P(X_1 \leq 4) = \sum_{x=0}^{4} P(X_1 = x) = \sum_{x=0}^{4} \frac{e^{-2}2^x}{x!} = e^{-2} + 2e^{-2} + 2e^{-2} + \frac{8}{6}e^{-2} + \frac{16}{24}e^{-2} \approx 0{,}9473$$

(ii) $E[X_3] = 2 \cdot 3 = 6$

3.68 En Poissonf. S.V. antar värdena 0 och 1 med samma slh. Hitta värdet på Poissonparametern k.

Lösning: $X \sim Poi(k)$

$P(X=0) = P(X=1)$

$X \sim Poi(k) \Rightarrow P(X=x) = \frac{e^{-k}k^x}{x!}$, $x = 0, 1, \ldots \Rightarrow P(X=0) = \frac{e^{-k}k^0}{0!} = e^{-k}$

$P(X=1) = \frac{e^{-k}k^1}{1!} = ke^{-k}$

$P(X=0) = P(X=1) \Leftrightarrow e^{-k} = ke^{-k} \Rightarrow \underline{k=1}$

3.78 En rad pistoler avfyras efter varandra. Varje skott träffar målet med slh. 1/4. Vad är slh att 2:a träffen kommer före 7:e skottet?

Lösning: X = antalet träffar på de 6 första skotten

$X \sim Bin(n=6, p=1/4) \Rightarrow f(x) = P(X=x) = \binom{6}{x} 0{,}25^x 0{,}75^{6-x}$, $x = 0, 1, \ldots, 6$

P(2:a träffen före 7:e skottet) = P(minst 2 träffar på de 6 första skotten) = $P(X \geq 2) =$

$1 - P(X \leq 1) = 1 - P(X=0) - P(X=1) = 1 - \binom{6}{0} 0{,}25^0 0{,}75^6 - \binom{6}{1} 0{,}25^1 0{,}75^5 \approx 0{,}4661$

3.83 $f(x) = \frac{x^2}{14}$, $x = 1, 2, 3$ (a) Visa att f är täthetsf.

(i) $f(x) \geq 0$ ok.

(ii) $\sum_{x=1}^{3} f(x) = \sum_{x=1}^{3} \frac{x^2}{14} = \frac{1}{14} + \frac{2^2}{14} + \frac{3^2}{14} = 1$ ok. $\Rightarrow$ f täthetsf.

(b) $E[X] = \sum_{x=1}^{3} x f(x) = \sum_{x=1}^{3} x \cdot \frac{x^2}{14} = \frac{1^3}{14} + \frac{2^3}{14} + \frac{3^3}{14} = \frac{36}{14} \approx 2{,}57$

$E[X^2] = \sum_{x=1}^{3} x^2 f(x) = \sum_{x=1}^{3} x^2 \frac{x^2}{14} = \frac{98}{14} = 7$

(c) $m_X(t) = E[e^{tX}] = \sum_{x=1}^{3} e^{tx} f(x) = \sum_{x=1}^{3} e^{tx} \frac{x^2}{14} = \frac{e^t}{14} + e^{2t}\frac{4}{14} + e^{3t}\frac{9}{14} = \frac{1}{14}(e^t + 4e^{2t} + 9e^{3t})$

Övning 5

4.35 Glömskeegenskapen

X = livslängden hos en elektronisk komponent, $X \sim Exp(\lambda)$

Slh att komponenten fungerar vid tiden t_1+t_2, givet att den fungerar vid tiden t_1 = Slh att den fungerar vid t_2 tidsenheter, dvs. $P(X>t_1+t_2 \mid X>t_1)=P(X>t_2)$

4.4 X kont. s.v. med täthet $f(x)=\frac{1}{\ln 2}\cdot\frac{1}{x}$, $25\le x\le 50$

(a) Visa att f täthetsf.

(b) $P(30\le x\le 40)$?

Lösning: (a) (i) $f(x)\geq 0$, $\forall x$ ok.

(ii) $\int_{25}^{50} f(x)\,dx = \int_{25}^{50}\frac{1}{\ln 2}\frac{1}{x}\,dx = \frac{1}{\ln 2}\int_{25}^{50}\frac{1}{x}\,dx = \frac{1}{\ln 2}(\ln 50-\ln 25)=1$ ok.

(b) $P(30\le x\le 40)=\int_{30}^{40} f(x)\,dx = \frac{1}{\ln 2}(\ln 40-\ln 30)\approx 0{,}4150$

4.13

$$F(x)=P(X\le x)=\begin{cases} 0, & x<25 \\ 1, & x>50 \\ 25\le x\le 50: & P(X\le x)=P(25\le X\le x)= \\ & =\int_{25}^{x} f(x)\,dx=\frac{1}{\ln 2}(\ln x-\ln 25)=\frac{\ln(x/25)}{\ln 2}\end{cases}$$

$$F(x)=\begin{cases} 0, & x<25 \\ \ln(x/25)/\ln 2, & 25\le x\le 50 \\ 1, & x>50\end{cases}$$

(b) $P(30\le X\le 40)=P(X\le 40)-\underbrace{P(X<30)}_{P(X\le 30)}=F(40)-F(30)=$

$=\frac{\ln(1{,}6)-\ln(1{,}2)}{\ln 2}\approx 0{,}4150$

4.15 X kont. s.v. med täthet $f(x)=(1/6)x$, $2\le x\le 4$

(a) $E[X]$, (b) $E[X^2]$, (c) σ^2, σ?

(a) $E[X]=\int_{-\infty}^{\infty} x f(x)\,dx=\int_{2}^{4} x\frac{1}{6}x\,dx=\frac{1}{6}\int_{2}^{4}x^2\,dx=\frac{1}{18}(4^3-2^3)\approx 3{,}1111$

(b) $E[X^2]=\int_{-\infty}^{\infty} x^2 f(x)\,dx\int_{2}^{4}x^2\frac{1}{6}x\,dx=\frac{1}{24}(4^4-2^4)=10$

(c) $\sigma^2=Var(X)=E[X^2]-(E[X])^2=10-3{,}1111^2\approx 0{,}321 \Rightarrow \sigma=\sqrt{\sigma^2}\approx 0{,}567$

4.17 X kont. S.V.

$$f(x) = \frac{1}{10} e^{-x/10} \quad , x > 0$$

(a) Momentgenererande funktion?

(b) $E[X]$

(c) $Var(X)$? σ?

$$\left(\begin{array}{l} X \sim Exp(\lambda = \frac{1}{10}) \\ E[X] = \frac{1}{\lambda} = 10 \\ Var(X) = \frac{1}{\lambda^2} = 100 \end{array} \right)$$

{Men detta är kap 4.2 och Exp. är inte def. förrän i 4.3, alltså ska vi nog inte göra så....}

$$(a)\ m_X(t) = E[e^{tX}] = \int_{-\infty}^{\infty} e^{tx} f(x)\,dx = \frac{1}{10}\int_0^{\infty} e^{tx} e^{-x/10}\,dx = \frac{1}{10}\int_0^{\infty} e^{-x(-t+\frac{1}{10})}\,dx =$$

$$= \frac{1}{10(t-\frac{1}{10})}\left[e^{-x(-t+\frac{1}{10})}\right]_0^{\infty} \underset{\substack{\uparrow \\ t<\frac{1}{10}}}{=} \frac{1}{10t-1}(-1) = \frac{1}{1-10t}$$

$$E[X^k] = \left.\frac{\partial^k m_X(t)}{\partial t^k}\right|_{t=0}$$

$$E[X] = \left.\frac{\partial m_X(t)}{\partial t}\right|_{t=0} = \left.\left(\frac{10}{(1-10t)^2}\right)\right|_{t=0} = 10$$

$$(b)\ E[X^2] = \left.\frac{\partial^2 m_X(t)}{\partial t^2}\right|_{t=0} = \left.\left(\frac{200}{(1-10t)^3}\right)\right|_{t=0} = 200$$

$$(c)\ \sigma^2 = Var(X) = E[X^2] - (E[X])^2 = 200 - 10^2 = 100 \Rightarrow \sigma = 10$$

4.29 $X \sim Gamma(\alpha = 3, \beta = 4)$

(a) $f(x)$? (b) $m_X(t)$? (c) μ, σ^2, σ?

$$(a)\ f(x) = \frac{1}{\Gamma(\alpha)\beta^\alpha} x^{\alpha-1} e^{-x/\beta} = \frac{1}{\underbrace{\Gamma(3)}_{=2!} 4^3} x^{3-1} e^{-x/4} = \frac{1}{128} x^2 e^{-x/4} \quad , x > 0$$

$$(b)\ m_X(t) = (1-\beta t)^{-\alpha} = (1-4t)^{-3} \quad , \ t < \frac{1}{\beta} = \frac{1}{4}$$

$$(c)\ \mu = \alpha\beta = 3 \cdot 4 = 12$$

$$\sigma^2 = \alpha\beta^2 = 3 \cdot 4^2 = 48$$

$$\Rightarrow \sigma = \sqrt{48}$$

4.41 $X \sim N(\mu, \sigma^2)$, $\mu = 0$, $\sigma = 100$

(a) $P(-\underset{=0}{200} \leq X \leq 200)$?

$$\frac{X - 0}{100} \sim N(0,1)$$

$$P(-200 \leq X \leq 200) = P(X \leq 200) - P(X < -200) =$$

$$= P\left(\frac{X}{100} \leq 2\right) - P\left(\frac{X}{100} < -2\right) = \phi(2) - \underbrace{\phi(-2)}_{1-\phi(2)} = 0{,}9772 - 0{,}0228 =$$

$$= 0{,}9544$$

(b) $P(X < -250) + P(X > 250) = P(X < -250) + 1 - P(X \leq 250) =$

$$= P\left(\frac{X}{100} < -2{,}5\right) + 1 - P\left(\frac{X}{100} \leq 2{,}5\right) = \phi(-2{,}5) + 1 - \phi(2{,}5) =$$

$$= 0{,}0062 + 1 - 0{,}9938 = 0{,}0124$$

(c) $0{,}2 = P(X \leq -x) + P(X \geq x) = P(X \leq -x) + 1 - P(X < x) =$

$$= P\left(\frac{X}{100} \leq \frac{-x}{100}\right) + 1 - P\left(\frac{X}{100} < \frac{x}{100}\right) = \underbrace{\phi\left(\frac{-x}{100}\right)}_{=1-\phi\left(\frac{x}{100}\right)} + 1 - \phi\left(\frac{x}{100}\right) =$$

$$= 2\left(1 - \phi\left(\frac{x}{100}\right)\right)$$

$$\Rightarrow 1 - \phi\left(\frac{x}{100}\right) = 0{,}1$$

$$\Rightarrow \phi\left(\frac{x}{100}\right) = 0{,}9$$

tabell $\Rightarrow \phi(1{,}28) \approx 0{,}9 \Rightarrow \dfrac{x}{100} = 1{,}28$

$$\Rightarrow x = 128$$

(d) $m_X(t) = e^{\mu t + \sigma^2 t^2/2} = e^{0 \cdot t + 100\, t^2/2} = e^{5000 t^2}$

Kapitel 1:
TEORI & BETECKNINGAR

Ω - Utfallsrum

Låt A vara händelse i utfallsrummet ($A \subseteq \Omega$)

$\mathbf{P}(A)$ - sannolikheten att A inträffar

$$0 \le \mathbf{P}(A) \le 1$$

$\mathbf{P}(A)$ nära 1 betyder A har stor chans att inträffa
$\mathbf{P}(A)$ nära 0 betyder A har liten chans att inträffa

Exempel: Kast med tärning

$$\Omega = \{1,2,3,4,5,6\}$$

A=\{utfallet "sexa"\} $\Longrightarrow \mathbf{P}(A) = \frac{1}{6}$

Om alla ufall i Ω är lika troliga blir

$$\mathbf{P}(A) = \frac{n(A)}{n(\Omega)},$$

där $n(A)$ - antalet utfall som ger A
och $n(\Omega)$ - antalet utfall i Ω

Exempel: Kast med tärning

$$\Omega = \{1,2,3,4,5,6\}$$

A=\{utfallet "udda"\}=\{1,3,5\}

$$\mathbf{P}(A) = \frac{n(A)}{n(\Omega)} = \frac{3}{6}$$

2 händelser sägs vara $\boxed{\text{disjunkta}}$ om de ej kan inträffa samtidigt, dvs A och B disjunkta om $A \cap B = \emptyset$

tex A = utfallet udda
B = utfallet 6

Antalet sätt att ordna n element

$$n! = n \times (n-1) \times \ldots \times 2 \times 1$$

Antalet sätt att välja k ur n element då ordningen spelar ingen roll:

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

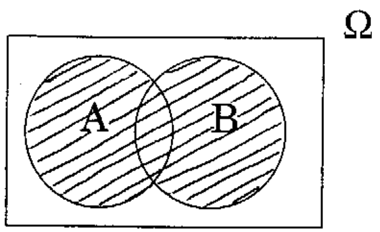
Antalet sätt att välja k ur n element då ordningen spelar roll:

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

Kapitel 2:
TEORI & BETECKNINGAR

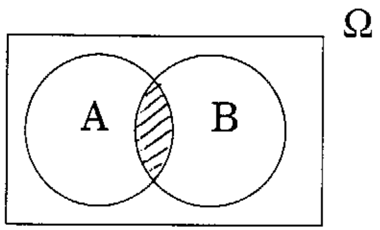
Ω - Utfallsrum
A och B - händelser i Ω

Union: $A \cup B$ (minst en av A och B)



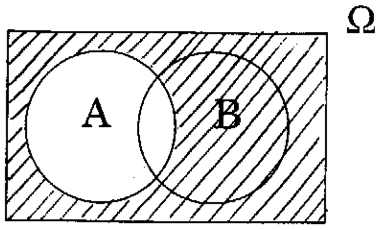
Figur 1: $A \bigcup B$

Snitt: $A \cap B$ (både A och B)



Figur 2: $A \bigcap B$

Komplement: A^C (inte A)



Figur 3: A^C

(a) $P(\Omega) = 1$

(b) $P(\emptyset) = 0$

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

(d) $P(A \cup B) = P(A) + P(B) - P(A \cap B)$
(additionsregeln)
$\Longrightarrow P(A \cap B) = P(A) + P(B) - P(A \cup B)$

(e) A, B disjunkta, dvs $P(A \cap B) = 0$,
$\Longrightarrow P(A \cup B) = P(A) + P(B)$

Sannolikhetsaxiomet

1. $P(\Omega) = 1$, då Ω är utfallsrum för experiment
2. $P(A) \geq 0$, för alla händelser A
3. $A_1, A_2, A_3, \ldots$ disjunkta:

$P(A_1 \cup A_2 \cup A_3 \ldots) = P(A_1) + P(A_2) + P(A_3) + \ldots$

Betingade händelser

$P(A$ inträffar givet att B inträffat)
$= P(A|B) = \frac{P(A \cap B)}{P(B)}$

$\Longrightarrow P(A \cap B) = P(A|B)P(B)$

Oberoende händelser

A, B oberoende omm $P(A \cap B) = P(A)P(B)$

A, B oberoende: $P(A|B) = P(A)$

Bayes formel

Låt $A_1, A_2, \ldots, A_n$ vara disjunkta händelser som bildar hela utfallsrummet, dvs

$$\bigcup_{i}^{n} A_i = \Omega$$

och B händelse som uppfyller $P(B) \neq 0$.
Då är

$$P(A_j|B) = \frac{P(B|A_j)P(A_j)}{\Sigma_{i=1}^{n} P(B|A_i)P(A_i)},$$

för $j = 1, 2, \ldots, n$

Kapitel 3.1-4:
DISKRETA FÖRDELNINGAR

En *stokastisk variabel* (s.v.) X är en variabel vars observerade värde bestäms av slumpen

En s.v. är **diskret** om den kan anta endast ett ändligt eller uppräkneligt antal värden

Exempel på diskreta stokastiska variabler:

(1) X = utfall vid kast med 6-sidig tärning

(2) X = antalet "krona" vid kast med 10 mynt

(3) X = antalet kort man måste dra ur "blandad" kortlek innan man får ett ess

Låt X vara en diskret s.v.

$$f(x) = P(X = x)$$

kallas för **täthetsfunktionen** för X (eller sannolikhetsfunktion)

f uppfyller villkoren:

$$\text{(i) } f(x) \geq 0$$

$$\text{(ii) } \sum_x f(x) = 1$$

$$F(x) = P(X \leq x) = \sum_{y \leq x} f(y)$$

kallas för **(kumulativa) fördelningsfunktionen** för X

Exempel: X = utfall vid kast med tärning

$$f(x) = \begin{cases} 1/6 & x \in \{1,2,3,4,5,6\} \\ 0 & \text{annars} \end{cases}$$

$$F(2) = P(X \leq 2) = P(X = 1) + P(X = 2) = 2/6$$

Väntevärde: (medelvärde)

$$\mu = E[X] = \sum_x x f(x)$$

$$E[g(X)] = \sum_x g(x) f(x)$$

Om X, Y diskreta s.v. och c konstant, gäller att:

1. $E[c] = c$
2. $E[cX] = cE[X]$
3. $E[X + Y] = E[X] + E[Y]$

Bevis av 1 & 2 (uppgift 3.19):

1. Låt $g(x) = c$. Då är
$E[c] = E[g(X)] = \Sigma_x g(x) f(x) = \Sigma_x c f(x)$
$= c \Sigma_x f(x) = c$

2. Låt $h(x) = cx$. Då är
$E[cX] = E[h(X)] = \Sigma_x h(x) f(x)$
$= \Sigma_x cx f(x) = c \Sigma_x x f(x) = cE[X]$

Varians:

$\sigma^2 = Var(X) = E[(X-\mu)^2] = E[X^2]-(E[X])^2$

Bevis av sista likheten:

$E[(X-\mu)^2] = E[X^2 - 2\mu X + \mu^2]$
$= E[X^2] - E[2\mu X] + E[\mu^2]$
$= E[X^2] - 2\mu E[X] + \mu^2$
$= E[X^2] - 2E[X]E[X] + (E[X])^2$
$= E[X^2] - (E[X])^2$

Om X, Y **oberoende** diskreta s.v. och c konstant, gäller att:

1. $Var(c) = 0$
2. $Var(cX) = c^2 Var(X)$
3. $Var(X+Y) = Var(X) + Var(Y)$

Bevis av 1 & 2 (uppgift 3.20):

1. $Var(c) = E[c^2] - (E[c])^2 = c^2 - c^2 = 0$
2. $Var(cX) = E[(cX)^2] - (E[cX])^2 = c^2E[X^2] - (cE[X])^2 = c^2E[X^2] - c^2(E[X])^2 = c^2(E[X^2] - (E[X])^2) = c^2 Var(X)$

$\sigma = \sqrt{\sigma^2} = \sqrt{Var(X)}$

kallas $\underline{\text{STANDARDAVVIKELSE}}$

Momentgenererande funktion:

$$m_X(t) = E[e^{tX}]$$

Sats:

$$\frac{d^k m_X(t)}{dt^k}\Big|_{t=0} = E[X^k]$$

Geometrisk fördelning

Antag att ett experiment lyckas med sannolikhet p. Vidare, antag att vi upprepar experimentet tills vi får lyckat utfall och låt X = antalet experiment som behöver utföras. Då sägs X vara *geometriskt fördelad* med parameter p (skrivs $X \sim Geo(p)$)

Det går att visa att

(i) $f(x) = P(X = x) = (1-p)^{x-1}p$,
 då $x = 1, 2, 3, \ldots$
(ii) $E[X] = 1/p$
(iii) $Var(X) = (1-p)/p^2$

Kapitel 3.5-8:
DISKRETA FÖRDELNINGAR

Binomialfördelningen

Antag att ett experiment lyckas med sannolikhet p (och misslyckas med slh $1-p$). Antag att vi utför n oberoende sådana experiment och låter

X = antalet lyckade försök (av de n).

Då är X **binomialfördelad** med parametrar n och p, vilket skrivs

$$X \sim Bin(n,p)$$

Exempel

(i) X=antalet "6:or" vid kast med 10 tärningar

$$\Longrightarrow X \sim Bin(n=10, p=\frac{1}{6}).$$

(ii) Tenta med 8 stycken 3-alternativsfrågor

X=antalet korrekta svar om man chansar

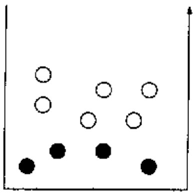
$$\Longrightarrow X \sim Bin(n=8, p=\frac{1}{3}).$$

Om $X \sim Bin(n,p)$ så gäller att:

(a) $f(x) = P(X=x) = \binom{n}{x} p^x(1-p)^{n-x}$,
$x = 0, \ldots, n$
(b) $E[X] = np$
(c) $Var(X) = np(1-p)$
(d) $m_X(t) = (1-p+pe^t)^n$.

Hypergeometrisk fördelning
Exempel

Urna med 6 vita och 4 svarta kulor. Vi drar 2 kulor



Figur 1: 6 vita, 4 svarta kulor

och noterar X=antalet svarta av dessa två.

$$f(x) = P(X=x) = \frac{\binom{4}{x}\binom{6}{2-x}}{\binom{10}{2}}, x = 0, 1, 2$$

Då är $X \sim Hypergeometrisk(10, 2, 4)$

Generellt Antag att vi har N element, varav r st är av Typ 1 och $N-r$ av Typ 2. Antag att vi drar n element ur de N och noterar

X=antalet element av Typ 1 (av de n dragna)

Då är

$$X \sim Hypergeometrisk(N, n, r)$$

och

(i) $f(x) = P(X=x) = \dfrac{\binom{r}{x}\binom{N-r}{n-x}}{\binom{N}{n}}$,

$max[0, n-(N-r)] \le x \le min[n, r]$

(ii) $E[X] = n(\frac{r}{N})$

(iii) $Var(X) = n(\frac{r}{N})(\frac{N-r}{N})(\frac{N-n}{N-1})$

Poissonfördelningen

X är Poissonfördelad med parameter k om

$$f(x) = P(X=x) = \frac{e^{-k}k^x}{x!}, x = 0, 1, \ldots; k > 0$$

Då är
(i) $m_X(t) = e^{k(e^t-1)}$
(ii) $E[X] = Var(X) = k$

Om n stor och p litet: $Bin(n,p) \approx Poisson(np)$

Poissonfördelningen brukar användas då vi beräknar antalet händelser i ett visst tidsintervall t ex inkomna telefonsamtal eller antalet jordbävningar under en viss tid

Kapitel 4.1-4:
KONTINUERLIGA FÖRDELNINGAR

En s.v. är **kontinuerlig** om den kan anta oändligt många värden, t ex alla reella tal i ett intervall och sannolikheten att anta ett specifikt värde är 0.

Om X kontinuerlig s.v. så har X täthetsfunktion f, definierad enligt

(i) $f(x) \geq 0$
(ii) $\int_{-\infty}^{\infty} f(x)dx = 1$
(iii) $P(a \leq X \leq b) = \int_a^b f(x)dx$

och fördelningsfunktion F enligt

$$F(x) = P(X \leq x) = \int_{-\infty}^{x} f(x)dx$$

(Obs! $P(X \leq x) = P(X < x)$)

Vidare definieras väntevärde enligt

$$\mu = E[X] = \int_{-\infty}^{\infty} xf(x)dx$$

Gammafördelningen

X är Gammafördelad med parametrar α och β ($X \sim Gamma(\alpha, \beta)$) om

$$f(x) = \frac{1}{\Gamma(\alpha)\beta^\alpha} x^{\alpha-1} e^{-x/\beta},$$

där $x, \alpha, \beta > 0$. Då är

(i) $m_X(t) = (1 - \beta t)^{-\alpha}$, $t < 1/\beta$
(ii) $E[X] = \alpha\beta$
(iii) $Var(X) = \alpha\beta^2$

Exponentialfördelningen

X är Exponentialfördelad med parameter λ ($X \sim Exp(\lambda)$) om

$$f(x) = \lambda e^{-\lambda x}, \text{ där } x, \lambda > 0$$

Då är $E[X] = 1/\lambda$, $Var(X) = 1/\lambda^2$ och $F(x) = 1 - e^{-\lambda x}$.

(Obs! Boken använder definitionen $X \sim Exp(\beta)$ om $f(x) = \frac{1}{\beta} e^{-x/\beta}$,
dvs Gammafördelning med $\alpha = 1$).

Normalfördelningen

X är Normalfördelad med väntevärde μ och varians σ^2 ($X \sim N(\mu, \sigma^2)$) om

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

Det går att visa att $m_X(t) = e^{\mu t + \sigma^2 t^2/2}$.

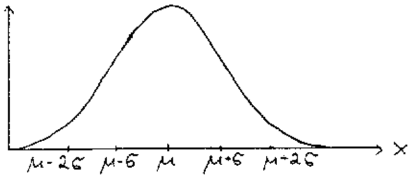
Om $X \sim N(0, 1)$ kallas X *standard normalfördelad.*

Om $X \sim N(\mu, \sigma^2)$ så gäller det att

$$X - \mu \sim N(0, \sigma^2)$$

och

$$\frac{X - \mu}{\sigma} \sim N(0, 1).$$



Om $X \sim N(0, 1)$ så är

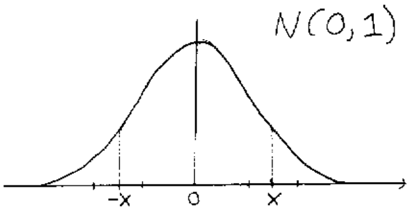
$$F(x) = P(X \leq x) = \phi(x),$$

där ϕ ges på tabellform, sid 730-731.

Om $X \sim N(\mu, \sigma^2)$ så får vi istället att

$$F(x) = P(X \leq x) = P(X - \mu \leq x - \mu)$$
$$= P(\tfrac{X-\mu}{\sigma} \leq \tfrac{x-\mu}{\sigma}) = \phi(\tfrac{x-\mu}{\sigma})$$

$$\phi(-x) = P(X \leq -x) = P(X \geq x)$$
$$= 1 - P(X < x) = 1 - \phi(x)$$



Kapitel 4.4-4.7:
KONTINUERLIGA FÖRDELNINGAR (forts)

Normalfördelningen

X är Normalfördelad med väntevärde μ och varians σ^2 ($X \sim N(\mu, \sigma^2)$) om

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

Om $X \sim N(0,1)$ kallas X *standard normalfördelad* och har fördelningsfunktion $\Phi(x) = P(X \leq x)$

Om $X \sim N(\mu, \sigma^2)$ så gäller det att

$$\frac{X-\mu}{\sigma} \sim N(0,1).$$

Om n stort gäller att

$$Bin(n,p) \approx N(np, np(1-p))$$

(Approximation acceptabel om

$$p \leq 0.5 \ \& \ np > 5$$

eller

$$p > 0.5 \ \& \ n(1-p) > 5)$$

Uppgift 4.42

X = glukosnivå hos diabetiker

$$X \sim N(\mu = 106, \sigma^2 = 8^2)$$

$$\Longrightarrow \frac{X-106}{8} \sim N(0,1)$$

(a) $P(90 \leq X \leq 122) = P(X \leq 122) - P(X < 90)$
$= P(X - 106 \leq 16) - P(X - 106 < -16)$
$= P(\frac{X-106}{8} \leq 2) - P(\frac{X-106}{8} < -2)$
$= \Phi(2) - \Phi(-2) = \Phi(2) - (1 - \Phi(2))$
$= 2\Phi(2) - 1 = 2 \times 0.9772 - 1 = 0.9544$

(b) $P(X \leq 120) = P(\frac{X-106}{8} \leq 1.75)$
$= \Phi(1.75) = 0.9599$

(c) Hitta x sådan att $P(X \leq x) = 0.25$

$0.25 = P(X \leq x) = P(\frac{X-106}{8} \leq \frac{x-106}{8}) = \Phi(\frac{x-106}{8})$
$\Phi(-0.675) \approx 0.25 \Longrightarrow \frac{x-106}{8} = -0.675$
$\Longrightarrow x = 8 \times (-0.675) + 106 = 100.6$

Weibullfördelningen

X är Weibullfördelad med parametrar α och β om

$$f(x) = \alpha\beta x^{\beta-1} e^{-\alpha x^\beta}, \quad x, \alpha, \beta > 0$$

Det går att visa att

$$\mu = E[X] = \alpha^{-1/\beta}\Gamma(1 + 1/\beta)$$

$$\sigma^2 = Var(X) = \alpha^{-2/\beta}\Gamma(1 + 2/\beta) - \mu^2$$

Denna fördelning brukar dyka upp i tillförlitlighetssystem

Tillförlitlighetsteori

Tillförlitlighetsfunktion: =överlevnadsfkn.

$$R(t) = P(\text{komponent fungerar vid tiden } t)$$

$$= 1 - P(\text{komponent går sönder innan tiden } t)$$

Riskfunktion: =Felintensitets fkn.

$$\rho(t) = f(t)/R(t)$$

Sats:

$$R(t) = exp[-\int_0^t \rho(x)dx]$$

Uppgift 4.60

X = livslängden (i 1000 körda mil)
$X \sim Weibull(\alpha = 0.04, \beta = 2)$

(a) $f(x) = \alpha\beta x^{\beta-1} e^{-\alpha x^\beta}$
$= 0.04 \times 2x^{2-1} e^{-0.04x^2} = 0.08xe^{-0.04x^2}$
$E[X] = \alpha^{-1/\beta}\Gamma(1 + 1/\beta) = 0.04^{-1/2}\Gamma(1 + 1/2)$
$= 5 \times 0.5\Gamma(0.5) = 2.5\sqrt(\pi) \approx 4.43$
$Var(X) = \alpha^{-2/\beta}\Gamma(1 + 2/\beta) - \mu^2$
$0.04^{-2/2}\Gamma(1 + 2/2) - (2.5\sqrt(\pi))^2$
$= 25\Gamma(2) - 6.25\pi = 25 - 6.25\pi \approx 5.37$

(b) Tillf=$R(t) = P(X \geq t) = \int_t^\infty f(x)dx$
$= \int_t^\infty 0.08xe^{-0.04x^2}dx$
$= [y = x^2, x = \sqrt(y), dx = \frac{1}{2\sqrt(y)}dy]$
$= \int_{t^2}^\infty 0.08\sqrt{y}e^{-0.04y}\frac{1}{2\sqrt(y)}dy = \int_{t^2}^\infty 0.04e^{-0.04y}dy$
$= e^{-0.04t^2}$

(c) $R(5) \approx 0.37$, $R(10) \approx 0.018$

(d) Riskf= $\rho(t) = \frac{f(t)}{R(t)} = \frac{0.08te^{-0.04t^2}}{e^{-0.04t^2}} = 0.08t$

(e) $\rho(5) = 0.4, = \rho(10) = 0.8$

(f) $P(X \leq 3) = 1 - P(X > 3) = 1 - R(3) \approx 0.30$

Kapitel 5:
TVÅ-DIMENSIONELLA FÖRDELNINGAR

Om X och Y är s.v. så kallas (X, Y) för en **2-dimensionell s.v.**

(X, Y) har täthetsf $f_{XY}(x, y) \geq 0$, som i diskreta fallet uppfyller

$$(i) f_{XY}(x, y) = P(X = x \text{ och } Y = y)$$

$$(ii) \sum_x \sum_y f_{XY}(x, y) = 1$$

och i kontinuerliga fallet

$$(i) \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} f_{XY}(x, y) dy dx = 1$$

$$(ii) P(a \leq X \leq b \text{ och } c \leq Y \leq d) = \int_a^b \int_c^d f_{XY}(x, y) dy dx$$

X har *marginaltäthet* $f_X(x)$ som uppfyller

$$f_X(x) = \sum_y f_{XY}(x, y) \text{ (diskreta fallet)}$$

$$f_X(x) = \int_{-\infty}^{\infty} f_{XY}(x, y) dy \text{ (kontinuerliga fallet)}$$

(motsvarande för Y)

$\boxed{X, Y \textbf{ oberoende} \text{ omm } f_{XY}(x, y) = f_X(x) f_Y(y)}$

UPPGIFT 5.5

$$\begin{cases} X = \#\text{ syntaxfel} \\ Y = \#\text{ logikfel} \end{cases}$$

då ett BASIC-program körs

x/y	0	1	2	3
0	0,400	0.100	0.020	0.005
1	0.300	0.040	0,010	0,004
2	0.040	0.010	0,009	0,003
3	0,009	0,008	0,007	0,003
4	0.008	0.007	0.005	0,002
5	0.005	0.002	0,002	0,001

(a) $P(\text{inget fel}) = P(X=0, Y=0) = 0{,}400$

(b) $P(\text{minst 1 syntaxfel, men högst 1 logikfel}) =$
$= P(X \geq 1, Y \leq 1) = 0{,}3 + 0{,}04 + 0{,}04 + 0{,}01 + 0{,}009$
$+ 0{,}008 + 0{,}008 + 0{,}007 + 0{,}005 + 0{,}002 = 0{,}429$

(c) X s marginaltäthet:

$P(X=0) = 0{,}4 + 0{,}1 + 0{,}02 + 0{,}005 = 0{,}525$
$P(X=1) = 0{,}3 + 0{,}04 + 0{,}01 + 0{,}004 = 0{,}354$
$P(X=2) = 0{,}04 + 0{,}01 + 0{,}009 + 0{,}003 = 0{,}062$
$P(X=3) = 0{,}009 + 0{,}008 + 0{,}007 + 0{,}003 = 0{,}027$
$P(X=4) = 0{,}008 + 0{,}007 + 0{,}005 + 0{,}002 = 0{,}022$
$P(X=5) = 0{,}005 + 0{,}002 + 0{,}002 + 0{,}001 = 0{,}01$

Y s marginaltäthet:

$P(Y=0) = 0{,}4 + 0{,}3 + 0{,}04 + 0{,}009 + 0{,}008 + 0{,}005 = 0{,}762$
$P(Y=1) = 0{,}1 + 0{,}04 + 0{,}01 + 0{,}008 + 0{,}007 + 0{,}002 = 0{,}167$
$P(Y=2) = 0{,}02 + 0{,}01 + 0{,}009 + 0{,}007 + 0{,}005 + 0{,}002 = 0{,}053$
$P(Y=3) = 0{,}005 + 0{,}004 + 0{,}003 + 0{,}003 + 0{,}002 + 0{,}001 = 0{,}018$

(d) $P(\text{minst 2 syntaxfel}) = P(X \geq 2) = P(X=2) + P(X=3) + P(X=4) + P(X=5$
$= 0{,}062 + 0{,}027 + 0{,}022 + 0{,}01 = 0{,}121$

(e) $P(\text{1 eller 2 logikfel}) = P(Y=1) + P(Y=2) = 0{,}167 + 0{,}053 =$
$= 0{,}22$

(f) X och Y oberoende?

Nej, ty

$P(X=1, Y=1) = 0.04 \neq 0.059118 = P(X=1)P(Y=1)$

Övning 6

$X \sim Bin(n=1000, p)$

$P(X \leq 500) = \sum_{x=0}^{500} P(X=x)$ – lättare att approx. med normal förd.

$P(X < 5) = P(X \leq 4)$?? $\Rightarrow P(X \leq 4{,}5)$

↑ ↑

kont. eller diskret? ⇒ ta värdet mitt emellan

4.54 En kemisk reaktion ger oftast en avkastning på 70% Man inför en ny metod som skall höja avkastningen. Förespråkare för den nya metoden hävdar att den ger bättre resultat än den gamla i 90% av fallen. Man testar nya metoden 60 ggr. och noterar

X = # ggr vi får avkastning > 70%

$p = P(\text{avkastning} > 70\%) \Rightarrow X \sim Bin(n=60, p)$

(a) $p = 0{,}9 \Rightarrow X \sim Bin(n=60, p=0{,}9)$

Regel: $Bin(n,p) \approx N(np, np(1-p))$

om (i) $p \leq 0{,}5$ & $np > 5$

eller (ii) $p > 0{,}5$ & $n(1-p) > 5$

Vårt fall: $p = 0{,}9 > 0{,}5$ & $n(1-p) = 60 \cdot 0{,}1 = 6 > 5$

Normalapprox. lämplig

(b) Om $p = 0{,}9$, vad är $E[X]$?

$X \sim Bin(n=60, p=0{,}9) \Rightarrow E[X] = np = 60 \cdot 0{,}9 = 54$

(c) Man hävdar att $p > 0{,}9$. Då ska i genomsnitt mer än 54 av 60 försök ge avkastning > 70%

Vi accepterar $p > 0{,}9$ om $X \geqslant 59$

Vad är slh att accept. $p > 0{,}9$ om p egentligen $= 0{,}9$

Ska räkna $P(X \geqslant 59)$ då $p = 0{,}9$

$X \sim Bin(n=60, p=0{,}9) \approx N(54, 5{,}4)$

I binomialfallet: $P(X \geqslant 59) = P(X > 58)$

Normalapprox. $P(X > 58{,}5) = 1 - P(X \leq 58{,}5) = 1 - P\left(\frac{X-54}{\sqrt{5{,}4}} \leq \frac{4{,}5}{\sqrt{5{,}4}}\right) =$

$= 1 - \phi\left(\frac{4{,}5}{\sqrt{5{,}4}}\right) \approx 1 - \phi(1{,}94) \approx 0{,}0262$

(d) Vad är slh att inte accept. $p > 0,9$ då $p = 0,95$?

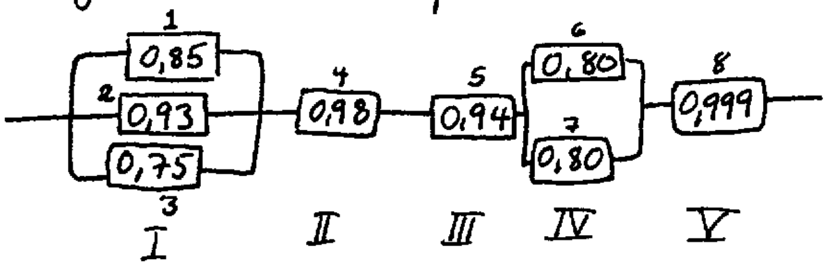
<u>Beräkna</u>: $P(X \leq 58)$ då $P = 0,95$

$\hookrightarrow P(X < 59)$

$$P = 0,95 \Rightarrow X \sim Bin(n = 60, p = 0,95) \approx N(57, 2,85)$$

$$P(X \leq 58,5) = P\left(\frac{X - 57}{\sqrt{2,85}} \leq \frac{1,5}{\sqrt{2,85}}\right) = \phi\left(\frac{1,5}{\sqrt{2,85}}\right) \approx \phi(0,89) = 0,8133$$

<u>4.65</u> System med 8 komponenter



(a) Beräkna tillförlitligheten hos de 2 parallellkopplade delarna I & IV

(b) Beräkna tillförlitligheten hos hela systemet.

<u>Lösning</u>: (a) $R(I) = P(\text{minst en av komp. } 1,2,3 \text{ fungerar}) =$

$= 1 - P(\text{ingen av komp. } 1,2,3 \text{ fungerar}) = 1 - \underbrace{P(k1 \text{ ur funk.})}_{=0,15}\underbrace{P(k2 \text{ ur funk})}_{=0,07}\underbrace{P(k3 \text{ ur funk})}_{=0,25} =$

$= 1 - 0,15 \cdot 0,07 \cdot 0,25 = 0,997375$

$R(IV) = P(\text{minst en av komp. } 6 \& 7 \text{ fungerar}) = 1 - P(\text{ingen av } 6,7 \text{ fungerar}) =$

$1 - 0,2 \cdot 0,2 = 0,96$

(b) $R(\text{systemet}) = R(I)R(II)R(III)R(IV)R(V) = 0,997375 \cdot 0,98 \cdot 0,94 \cdot 0,96 \cdot 0,999 = 0,89877$

<u>5.10</u> (X, Y) har täthet $f_{XY}(x,y) = \frac{x^3 y^3}{16}$, $0 \leq x \leq 2$, $0 \leq y \leq 2$

(a) Marginaltätheten?

(b) Oberoende?

(c) $P(X \leq 1)$?

(d) $P(X \leq 1)$ givet att $Y = 1$?

<u>Lösning</u>: (a) $f_X(x) = \int_0^2 \frac{x^3 y^3}{16} dy = \left[\frac{x^3 y^4}{64}\right]_0^2 = \frac{x^3}{4}$

$$f_Y(y) = \int_0^2 \frac{x^3 y^3}{16} dx = \frac{y^3}{4}$$

(b) Oberoende? Ja, ty $f_X(x) \cdot f_Y(y) = \frac{x^3}{4} \cdot \frac{y^3}{4} = \frac{x^3 y^3}{16} = f_{XY}(x,y)$

(c) $P(X \leq 1) = \int_0^1 f_X(x)\,dx = \int_0^1 \frac{x^3}{4}\,dx = \frac{1}{16}$

(d) Vet vi Y:s värde så är $P(X \leq 1) = \frac{1}{16}$ ty X, Y oberoende

5.9 $f_{XY}(x,y) = \frac{1}{x} \quad 0 < y < x < 1$

(a) Visa att detta är en 2-dim täthet för 2-dim S.V.

(i) $f_{XY}(x,y) \geqslant 0 \quad \forall x,y$ ok.

(ii) $\int_0^1 \int_y^1 \frac{1}{x}\,dx\,dy = \int_0^1 [-\ln y]\,dy = [y - y\ln y]_0^1 = 1$

(b) $P(X \leq 0{,}5\,,\, Y \leq 0{,}25) = \int_0^{0{,}25} \int_y^{0{,}5} \frac{1}{x}\,dx\,dy = 0{,}423$

(e) $f_X(x) = \int_0^x \frac{1}{x}\,dy = 1$

$f_Y(y) = \int_y^1 \frac{1}{x}\,dx = -\ln y$

(f) $P(X \leq 0{,}5) = \int_0^{0{,}5} f_X(x)\,dx = \int_0^{0{,}5} 1\,dx = 0{,}5$

(g) $P(Y \leq 0{,}25) = \int_0^{0{,}25} (-\ln y)\,dy = 0{,}597$

(h) $\left.\begin{array}{l} P(X \leq 0{,}5\,,\, Y \leq 0{,}25) = 0{,}423 \\ P(X \leq 0{,}5)P(Y \leq 0{,}25) = 0{,}5 \cdot 0{,}597 = 0{,}298 \end{array}\right\}$ ej samma!

$\Rightarrow X, Y$ ej oberoende!

Till övning 8.

Kapitel 7:
SKATTNING

En skattning $\hat{\theta}$ av parametern θ sägs vara
vvr = **väntevärdesriktig** (unbiased) omm $E[\hat{\theta}] = \theta$

Låt $X_1, X_2, \ldots, X_n$ vara ett stickprov av storlek n från s.v. X med väntevärde μ och varians σ^2. Då är (i) stickprovsmedelvärdet $\bar{X}$ en vvr-skattning av μ och (ii) stickprovsvariansen S^2 en vvr-skattning av σ^2

Bevis av (i):

$$
\begin{aligned}
E[\bar{X}] &= E[\tfrac{1}{n}(X_1 + X_2 + \ldots + X_n)] \\
&= \tfrac{1}{n}(E[X_1] + E[X_2] + \ldots + E[X_n]) \\
&= \tfrac{1}{n}(\mu + \mu + \ldots + \mu) = \tfrac{1}{n}n\mu = \mu
\end{aligned}
$$

$$\boxed{Var(\bar{X}) = \frac{\sigma^2}{n}}$$

Bevis:

$$
\begin{aligned}
Var(\bar{X}) &= Var(\tfrac{1}{n}(X_1 + X_2 + \ldots + X_n)) \\
&= \tfrac{1}{n^2}Var(X_1 + X_2 + \ldots + X_n) \\
&= \tfrac{1}{n^2}(Var(X_1) + Var(X_2) + \ldots + Var(X_n)) \\
&= \tfrac{1}{n^2}(\sigma^2 + \sigma^2 + \ldots + \sigma^2) = \tfrac{1}{n^2}n\sigma^2 = \frac{\sigma^2}{n}
\end{aligned}
$$

Skattningsmetoder
(1) MOMENTMETODEN

Skatta $E[X^k]$ med M_k, där $M_k = \frac{1}{n}\Sigma_{i=1}^{n} X_i^k$

(2) MAXIMIMETODEN

Skatta parametern θ med det värde på θ som maximerar *trolighetsfunktionen* $L(\theta) = \Pi_{i=1}^{n} f(X_i)$
(i praktiken maximerar man $lnL(\theta)$ istället)

EXEMPEL: $X \sim Exp(\lambda)$

Då skattar momentmetoden $E[X] = 1/\lambda$ med $\bar{X} = \frac{1}{n}\sum X_i \Rightarrow \lambda$ skattas med $1/\bar{X}$

Hur är det med maximimetoden?

$$L(\lambda) = \prod f(X_i) = \prod \lambda e^{-\lambda X_i} = \lambda^n e^{-\lambda \sum X_i}$$

$$\Rightarrow \ln L(\lambda) = n \ln(\lambda) - \lambda \sum X_i$$

$$\frac{\partial \ln L(\lambda)}{\partial \lambda} = \frac{n}{\lambda} - \sum X_i, \quad \frac{\partial \ln L(\lambda)}{\partial \lambda} = 0 \Rightarrow \lambda = \frac{n}{\sum X_i} = 1/\bar{X}$$

$\Rightarrow$ Maximimetoden skattar λ med $1/\bar{X}$

Om det för 2 s.v. X, Y gäller att $m_X(t) = m_Y(t)$ (för alla t i intervall kring 0) så har X och Y samma fördelning

Kapitel 7(forts)-8: KONFIDENSINTERVALL

Centrala Gränsvärdessatsen (CGS)

Om $X_1, X_2, \ldots, X_n$ stickprov från fördelning med väntevärde μ och varians σ^2, och $\underline{n \text{ är stort}}$ så gäller att

$$\bar{X} \approx N(\mu, \frac{\sigma^2}{n}) \text{ och } \frac{\bar{X} - \mu}{\sigma/\sqrt{n}} \approx N(0,1)$$

Om $\bar{X} \approx N(\mu, \frac{\sigma^2}{n})$, så

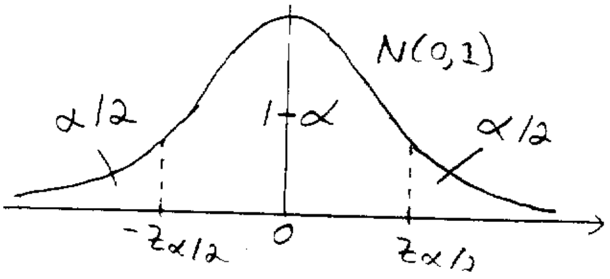
$$P(-1.96 \leq \frac{\bar{X} - \mu}{\sigma/\sqrt{n}} \leq 1.96) = 0.95$$

$$\leftrightarrow P(-1.96\frac{\sigma}{\sqrt{n}} \leq \bar{X} - \mu \leq 1.96\frac{\sigma}{\sqrt{n}}) = 0.95$$

$$\leftrightarrow P(\bar{X} - 1.96\frac{\sigma}{\sqrt{n}} \leq \mu \leq \bar{X} + 1.96\frac{\sigma}{\sqrt{n}}) = 0.95$$

$\bar{X} + / - 1.96\frac{\sigma}{\sqrt{n}}$ är ett 95% konfidensintervall för μ (alt $(\bar{X} - 1.96\frac{\sigma}{\sqrt{n}}, \bar{X} + 1.96\frac{\sigma}{\sqrt{n}})$)

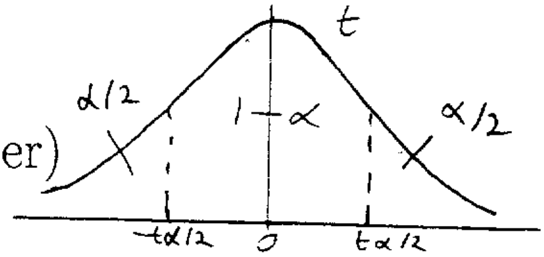
Generellt $\bar{X} + / - z_{\frac{\alpha}{2}}\frac{\sigma}{\sqrt{n}}$ är ett $100(1-\alpha)\%$ konfidensintervall för μ, där $z_{\frac{\alpha}{2}}$ uppfyller $\phi(-z_{\frac{\alpha}{2}}) = \frac{\alpha}{2}$ och $\phi(z_{\frac{\alpha}{2}}) = 1 - \frac{\alpha}{2}$



Om σ ej känd, utnyttja att om $X_1, X_2, \ldots, X_n$ stickprov från $N(\mu, \sigma^2)$-fördelning så är

$$\frac{\bar{X} - \mu}{s/\sqrt{n}} \sim t_{n-1}$$

(t-fördelad med $n - 1$ frihetsgrader)



Då blir $\bar{X} + / - t_{\frac{\alpha}{2},n-1}\frac{s}{\sqrt{n}}$ ett $100(1 - \alpha)\%$ konfidensintervall för μ

Konfidensintervall för σ^2

Om vi har stickprov från normalfördelningen så är

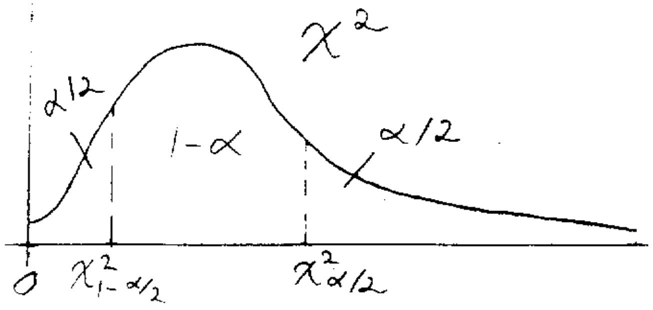
$$\frac{(n-1)s^2}{\sigma^2} \sim \chi^2_{n-1}$$

(χ^2-fördelad med $n - 1$ frihetsgrader)

$$P(\chi^2_{1-\frac{\alpha}{2},n-1} \leq \frac{(n-1)s^2}{\sigma^2} \leq \chi^2_{\frac{\alpha}{2},n-1}) = 1 - \alpha$$

$$\leftrightarrow P(\frac{(n-1)s^2}{\chi^2_{\frac{\alpha}{2},n-1}} \leq \sigma^2 \leq \frac{(n-1)s^2}{\chi^2_{1-\frac{\alpha}{2},n-1}}) = 1 - \alpha$$

vilket betyder att $(\frac{(n-1)s^2}{\chi^2_{\frac{\alpha}{2},n-1}}, \frac{(n-1)s^2}{\chi^2_{1-\frac{\alpha}{2},n-1}})$ är ett $100(1 - \alpha)\%$ konfidensintervall för σ^2



Kapitel 8 (forts):
HYPOTESTEST

Inledande Exempel

Antag att $X \sim N(\mu, \sigma^2 = 4)$, där μ okänd. Antag att μ tidigare har antagits vara 5, men att man inte längre tror att det är sant. För att bekräfta denna misstanke gör man 9 observationer på X

Resultat: 5, 7, 9, 8, 6, 4, 11, 7, 6 ej diskreta

Om $\mu = 5$ är $\frac{\bar{X}-5}{\sigma/\sqrt{n}} \sim N(0,1)$, vilket innebär att

$$P(-1.96 \leq \frac{\bar{X}-5}{\sigma/\sqrt{n}} \leq 1.96) = 0.95$$

Men vi har observerat

$$\frac{\bar{x}-5}{\sigma/\sqrt{n}} = \frac{7-5}{2/\sqrt{9}} = 3,$$

dvs det finns anledning att ej tro att $\mu = 5$

Hypotestest: Allmänt

Man vill testa en hypotes H_0 mot en alternativ H_1. Då finns 2 möjliga utfall: Antingen förkastar man H_0 och tror istället på H_1 eller så förkastar man ej H_0, vilket inte innebär att H_0 är sann utan bara att inte motsatsen kan bevisas. Den hypotes man vill påvisa sätts alltså som H_1. I exemplet ovan väljs därför $H_0 : \mu = 5$ och $H_1 : \mu \neq 5$

Då man testar hypotes H_0 mot H_1 kan 2 fel uppkomma, nämligen

(i) Typ I-fel: Förkasta H_0 då H_0 sann (α)- Slh att göra detta fel

(ii) Typ II-fel: Förkasta inte H_0 då H_1 sann (β)

Man vill konstruera testen så att dessa fel blir små

Både α och β kan ej minimeras simultant så man börjar med att fixera α , kallas testets *signifikansnivå*, dvs den högsta sannolikheten för Typ I-fel som man accepterar

$1-\beta$, dvs sannolikheten att förkasta H_0 då H_0 verkligen är falsk kallas testets *styrka*

Den minsta signifikansnivån man kan förkasta en hypotes på kallas för *p-värdet*

Exempel (forts)

Antag att $H_0 : \mu = 5$ testas mot $H_1 : \mu \neq 5$ på 5%-nivå. Då skall vi förkasta H_0 om

$$\frac{\bar{X} - 5}{\sigma / \sqrt{n}} < -1.96 \text{ eller } \frac{\bar{X} - 5}{\sigma / \sqrt{n}} > 1.96,$$

dvs vi kan förkasta H_0 på 5%-nivå

p-värde$= P(\frac{\bar{X}-5}{\sigma/\sqrt{n}} < -3) + P(\frac{\bar{X}-5}{\sigma/\sqrt{n}} > 3)$

$= \phi(-3) + 1 - \phi(3) = 2(1 - \phi(3))$

$= 2(1 - 0.9987) = 0.0026$

Om σ okänd, (använd S) använd t-test istället!

Icke-parametriska test

Teckentest för medianen

Låt $X_1, X_2, \ldots, X_n$ vara stickprov från fördelning med median M, dvs fördelning som uppfyller

$$P(X < M) = P(X > M) = 0.5$$

Hypotesen $H_0 : M = M_0$ testas m h a statistikan

$$Q_+ = \text{antalet observationer} > M_0$$

Under H_0 är

$$Q_+ \sim Bin(n, \frac{1}{2})$$

Alternativ $H_1 : M > M_0$: Förkasta för stora Q_+ (en-sidigt test)

Alternativ $H_1 : M < M_0$: Förkasta för små Q_+ (en-sidigt test)

Alternativ $H_1 : M \neq M_0$: Förkasta för små och stora Q_+ (två-sidigt test)

Wilcoxon test Läs själva!!

Övning 7

5.21 $f_{XY}(x,y) = \frac{1}{x} \quad 0<y<x<1$

$E[X]$? $E[Y]$? $E[XY]$? $Cov(X,Y)$?

$f_X(x) = \int_0^x f_{XY}(x,y)\,dy = 1$

$f_Y(y) = \int_y^1 f_{XY}(x,y)\,dx = -\ln y$

$E[X] = \int_0^1 x f_X(x)\,dx = \int_0^1 x\,dx = \frac{1}{2}$

$E[X] = \int_0^1 \int_0^x x f_{XY}(x,y)\,dy\,dx$

$E[Y] = \int_0^1 y f_Y(y)\,dy = \int_0^1 y(-\ln y)\,dy = \frac{1}{4}$

$E[XY] = \int_0^1 \int_0^x xy\, f_{XY}(x,y)\,dy\,dx = \frac{1}{6} \qquad \left(\int_0^1 \int_y^1 xy\, f_{XY}(x,y)\,dx\,dy \right)$

$Cov(X,Y) = E[XY] - E[X]E[Y] = \frac{1}{6} - \frac{1}{2}\cdot\frac{1}{4} = \frac{1}{24}$

5.33 $f_{XY}(x,y) = \frac{1}{x} \quad 0<y<x<1$

$E[X^2]$? , $E[Y^2]$? , $Var(X)$? $Var(Y)$? ρ_{XY}?

$E[X^2] = \int_0^1 x^2 f_X(x)\,dx = \int_0^1 x^2\,dx = \frac{1}{3}$

$E[Y^2] = \int_0^1 y^2 f_Y(y)\,dy = \int_0^1 y^2(-\ln y)\,dy = \frac{1}{9}$

$Var(X) = E[X^2] - (E[X])^2 = \frac{1}{3} - \left(\frac{1}{2}\right)^2 = \frac{1}{12}$

$Var(Y) = E[Y^2] - (E[Y])^2 = \frac{1}{9} - \left(\frac{1}{4}\right)^2 = \frac{7}{144}$

$\rho_{XY} = Corr(X,Y) = \frac{Cov(X,Y)}{\sqrt{Var(X)\,Var(Y)}} = \frac{1/24}{\sqrt{\frac{1}{12}\cdot\frac{7}{144}}} \simeq 0{,}655$

5.42 $f_{XY}(x,y) = \frac{1}{x} \quad 0<y<x<1$

(a) Hitta regressionskurvan för X och Y, dvs $\mu_{X|y}$ Är den linjär?

$f_{X|y}(x) = \frac{f_{XY}(x,y)}{f_Y(y)} = \frac{1/x}{-\ln y} = -\frac{1}{x\ln y}$

$\Rightarrow \mu_{X|y} = \int_y^1 x f_{X|y}(x)\,dx = \int_y^1 \frac{-x}{x\ln y}\,dx = \frac{y-1}{\ln y}$, ej linjär

(b) $\mu_{X|y} = 0{,}5 = \frac{0{,}5-1}{\ln 0{,}5} \approx 0{,}72$

(c) $f_{Y|x}(y) = \frac{f_{XY}(x,y)}{f_X(x)} = \frac{1/x}{1} = \frac{1}{x}$

$\Rightarrow \mu_{Y|x} = \int_0^x y\, f_{Y|x}(y)\, dy = \int_0^x y \frac{1}{x}\, dy = \frac{x}{2}$ <u>linjär</u>

(d) $\mu_{Y|x=0{,}75} = \frac{0{,}75}{2} = 0{,}375$

<u>6.6</u> X = antalet mikrogram partiklar /m³ luft

$X \sim N(\mu, \sigma^2)$ μ och σ^2 okända

Varje månad görs mätningar vid 5 24-timmars perioder

$\Rightarrow$ varje månad fås ett stickprov X_1, X_2, X_3, X_4, X_5 av storlek $n=5$ på X

(a) X_1:s fördelning? $X_1 \sim N(\mu, \sigma^2)$

(b) Givet en månad: $x_1 = 45$, $x_2 = 50$, $x_3 = 62$, $x_4 = 57$, $x_5 = 70$

$\sum x_i = 45+50+62+57+70 = 284$

$\sum x_i^2 = 45^2 + 50^2 + \ldots\ldots = 16518$

$\overline{x} = \frac{\sum x_i}{5} = \frac{284}{5} = 56{,}8$

$\max(x_i) = 70$, $\min(x_i) = 45$

(c) Är $X_5 - \mu$ en statistika? – Nej

Är $\frac{X_5 - \mu}{\sigma}$ en statistika? – Nej , μ, σ är okända.

<u>6.17</u> 2 stickprov

$$\begin{array}{ccc} & \textcircled{I} & \\ 1 & 3 & 2 \\ 2 & 5 & 4 \\ 4 & 3 & 3 \end{array} \qquad \begin{array}{cccc} & \textcircled{II} & & \\ 1 & 2 & 4 & 1 \\ 2 & 5 & 2 & 5 \\ 1 & 5 & 5 & 3 \end{array}$$

(a) $\overline{x}_I = \frac{1+3+2+\ldots+3}{9} = 3$, $\overline{x}_{II} = \frac{1+2+4+\ldots+3}{12} = 3$

$1223\underset{\underset{m_I}{\uparrow}}{3}3445$, $11122\underbrace{23}_{\underset{m_{II}}{\uparrow}}45555$ $m_{II} = \frac{2+3}{2} = 2{,}5$

(b) Variationsbredd = största − minsta

$\text{Var. bredd}_I = 5-1=4$

$\text{Var. bredd}_{II} = 5-1=4$

(c) $S_I^2 = \frac{1}{n-1}\left(\sum x_i^2 - n\bar{x}^2\right) = \frac{1}{8}\left(1^2+3^2+\dots+3^2-9\cdot 3^2\right) = \frac{1}{8}(93-81) = 1,5$

$\Rightarrow S_I = \sqrt{1,5} \approx 1,22$

$S_{II}^2 = \frac{1}{11}\left(1^2+2^2+\dots+3^2-12\cdot 3^2\right) = \frac{1}{11}(140-108) = 2,91$

$\Rightarrow S_{II} = \sqrt{2,91} \approx 1,71$

(d)

Det ser inte ut som om de kommer från samma förd. men vi har så små stickprov att det vore dumt att dra någon slutsats.

6.18 50 observationer $\sum_{i=1}^{50} x_i = 63707$, $\sum_{i=1}^{50} x_i^2 = 154924261$

(a) Skulle du bli förvånad om någon påstår att $E[X]=1270$?

$\bar{x} = \frac{\sum_{i=1}^{50} x_i}{n} = \frac{63707}{50} = 1274,14$

Nej, eftersom ↗

(b) $s^2 = \frac{1}{n-1}\left(\sum x_i^2 - n\bar{x}^2\right) = \frac{1}{50-1}\left(154924261 - 50\cdot 1274,14^2\right) = 1505155,592$

$\Rightarrow s = 1226,848$

Övning 8

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

För att visa att en funktion är normalförd. kan man visa att den har momentgenererande funktion som ser ut som:

$m_X(t) = e^{\mu t + \sigma^2 t^2/2}$

7.7 $X \sim$ Likformigt $(0,b)$

10 observationer: 10, 7, 11, 12, 8, 8, 9, 10, 9, 13

(a) Hitta vvr-skattning av $\mu (= E[X])$

$\bar{x} = \frac{10+7+11+\dots+13}{10} = 9,7$

(b) Vvr-skattning av σ^2

$s^2 = \frac{1}{n-1}\left(\sum x_i^2 - n\bar{x}^2\right) = \frac{1}{9}\left(10^2+7^2+\dots+13^2-10\cdot 9,7^2\right) = 3,57$

(c) Vvr-skattning av b

$E[X] = \frac{b}{2} \Rightarrow b = 2E[X]$

Skatta b med $\hat{b} = 2\bar{X}$, Vvr ty $E[\hat{b}] = 2E[\bar{X}] = 2E[X] = 2 \cdot \frac{b}{2} = b$

<u>Här</u>: $\hat{b} = 2 \cdot 9{,}7 = \underline{\underline{19{,}4}}$

(d) Skatta σ

$s = \sqrt{s^2} = \sqrt{3{,}57} \approx 1{,}89$

Är s vvr-skattning av σ?

Då ska $E[s] = \sigma \Rightarrow (E[s])^2 = \sigma^2$

$Var(s) = E[s^2] - (E[s])^2 \Rightarrow (E[s])^2 = E[s^2] - Var(s) =$

$= \sigma^2 - Var(s) \neq \sigma^2 \Rightarrow \underline{s \text{ ej vvr}}$

<u>7.9</u> Antag att vi har k oberoende stickprov av storlek $n_1, n_2, \dots, n_k$ från samma förd. Varje stickprov ger en vvr-skattning av μ, nämligen $\bar{X}_1, \bar{X}_2, \dots, \bar{X}_k$

(a) Visa att $\frac{\bar{X}_1 + \bar{X}_2 + \dots + \bar{X}_k}{k}$ vvr för μ

<u>Bevis</u>: $E\left[\frac{\bar{X}_1 + \bar{X}_2 + \dots + \bar{X}_k}{k}\right] = \frac{1}{k} E\left[\bar{X}_1 + \bar{X}_2 + \dots + \bar{X}_k\right] = \frac{1}{k}\left(\overset{\mu}{E[\bar{X}_1]} + \overset{\mu}{E[\bar{X}_2]} + \dots + \overset{\mu}{E[\bar{X}_k]}\right)$

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

(b) 3 stickprov $\bar{x}_1 = 0{,}8$ $\bar{x}_2 = 0{,}95$ $\bar{x}_3 = 0{,}7$

$n_1 = 9$ $n_2 = 3$ $n_3 = 200$

$\frac{\bar{x}_1 + \bar{x}_2 + \bar{x}_3}{3} = \frac{0{,}8 + 0{,}95 + 0{,}7}{3} = 0{,}82$

vvr-skattning av μ

$\hat{\mu}_W = \frac{n_1\bar{X}_1 + n_2\bar{X}_2 + \dots + n_k\bar{X}_k}{n_1 + n_2 + \dots + n_k}$

vvr-skattning ty $E[\hat{\mu}_W] = \frac{1}{n_1 + n_2 + \dots + n_k} E\left[n_1\bar{X}_1 + n_2\bar{X}_2 + \dots + n_k\bar{X}_k\right] =$

$= \frac{1}{n_1 + n_2 + \dots + n_k}\left(n_1 \overset{\mu}{E[\bar{X}_1]} + n_2 \overset{\mu}{E[\bar{X}_2]} + \dots + n_k \overset{\mu}{E[\bar{X}_k]}\right) = \mu$

(d) $\hat{\mu}_W = \frac{9 \cdot 0{,}8 + 3 \cdot 0{,}95 + 200 \cdot 0{,}7}{9 + 3 + 200} = 0{,}708$

7.16 $X_1, X_2, \ldots, X_m$ stickprov från $Bin(n,p)$-fördelning där n känd och p okänd. Visa att momentmetodens skattning av p är $\hat{p} = \frac{\bar{X}}{n}$

Bevis $E[X] = np$

momentmetoden skattar $E[X^k]$ med $\frac{1}{m}\sum X_i^k$

$\Rightarrow E[X]$ skattas med $\frac{1}{m}\sum X_i = \bar{X}$

$\Rightarrow np$ skattas med $\bar{X}$

$\Rightarrow p$ skattas med $\frac{\bar{X}}{n}$

7.17 $X_1, X_2, \ldots, X_n$ stickprov från $Poisson(\lambda S)$

Hitta momentm. skattning av λS och λ

Lösning: $E[X] = \lambda S$

Moment m. skattar $E[X]$ med $\bar{X}$

$\Rightarrow \lambda S$ skattas med $\bar{X}$

$\Rightarrow \lambda$ skattas med $\frac{\bar{X}}{S}$

7.35 För att studera andelen defekta mikrochips som produceras, väljer man ut stickprov av storlek 5 vid 20 tillfällen och vid varje noterar $X = \#$ defekta chips av de 5

Antag $X \sim Bin(5, p)$ där p okänd.

Följande data ficks:

1	0	1	2	0
0	0	0	1	0

Hitta maximimetodens,

(ML-metodens) skattning av p.

Lösning: $X \sim Bin(5,p) \Rightarrow f(x) = \binom{5}{x} p^x (1-p)^{5-x}$, $x = 0, \ldots, 5$

$$L(p) = \prod_{i=1}^{10} f(x_i) = \prod_{i=1}^{10} \binom{5}{x_i} p^{x_i} (1-p)^{5-x_i} = \left(\prod_{i=1}^{10} \binom{5}{x_i}\right) p^{\Sigma x_i} (1-p)^{\Sigma(5-x_i)} =$$

$$= \left(\prod_{i=1}^{10} \binom{5}{x_i}\right) p^{\sum_1^{10} x_i} (1-p)^{50 - \sum_1^{10} x_i}$$

$$\ln L(p) = \sum_{i=1}^{10} \ln\binom{5}{x_i} + \sum_{i=1}^{10} x_i \ln p + \left(50 - \sum_{i=1}^{10} x_i\right) \ln(1-p)$$

$$\frac{\partial \ln L(p)}{\partial p} = \frac{\sum_{i=1}^{10} x_i}{p} - \frac{50 - \sum_{i=1}^{10} x_i}{1-p}$$

$$\frac{\partial \ln L(p)}{\partial p} = 0 \Rightarrow p = \frac{\sum x_i}{50}$$

$\Rightarrow P$ skattas med $\frac{\sum x_i}{50} = \frac{1+0+1+2+\ldots+0}{50} = \frac{5}{50} = 0,1$

7.38 (a) Låt X S.V. med momentgen. f. $m_X(t)$

Låt $Y=\alpha+\beta X$ och visa att $m_Y(t)=e^{\alpha t}\cdot m_X(\beta t)$

(b) Låt $X\sim N(\mu=10,\ \sigma^2=4)$ och hitta momentg. f. till $Y=8+3X$. Y's fördelning

(a) Bevis: $m_Y(t)=E[e^{tY}]=E[e^{t(\alpha+\beta X)}]=E[e^{\alpha t+t\beta X}]=$

$=E[\underset{\substack{\uparrow\\ \text{konst.}}}{e^{\alpha t}}e^{(\beta t)X}]=e^{\alpha t}E[e^{(\beta t)X}]=e^{\alpha t}m_X(\beta t)$

(b) $X\sim N(\mu,\sigma^2)\Rightarrow m_X(t)=e^{\mu t+\sigma^2t^2/2}$

$X\sim N(\mu=10,\sigma^2=4)\Rightarrow m_X(t)=e^{10t+2t^2}$

$m_Y(t)=e^{8t}m_X(3t)=e^{8t}e^{10(3t)+2(3t)^2}=e^{8t+30t+18t^2}=e^{38t+36t^2/2}$

$\Rightarrow Y\sim N(\mu=38,\sigma^2=36)$

7.39 $X_1,X_2,\ldots,X_n$ ober. S.V. med $m_{X_i}(t)$

Låt $a_0,a_1,\ldots,a_n$ reella tal och $Y=a_0+a_1X_1+a_2X_2+\ldots+a_nX_n$

Visa att $m_Y(t)=e^{a_0t}\prod_{i=1}^{n}m_{X_i}(a_it)$

$m_Y(t)=E[e^{tY}]=E[e^{t(a_0+a_1X_1+\ldots+a_nX_n)}]=e^{a_0t}E[e^{a_1tX_1}e^{a_2tX_2}\cdot\ldots\cdot e^{a_ntX_n}]=$

$\underset{\substack{\uparrow\\ \text{ober.}}}{=}e^{a_0t}\underbrace{E[e^{a_1tX_1}]}_{m_{X_1}(a_1t)}\underbrace{E[e^{a_2tX_2}]}_{m_{X_2}(a_2t)}\cdot\ldots\cdot E[e^{a_ntX_n}]=e^{a_0t}\prod_{i=1}^{n}m_{X_i}(a_it)$

7.41 $X_1,X_2,\ldots,X_n$ normalf. oberoende S.V., $X_i\sim N(\mu_i,\sigma_i^2)$

$a_0,a_1,\ldots,a_n$ reella tal och $Y=a_0+a_1X_1+\ldots+a_nX_n$

$X_i\sim N(\mu_i,\sigma_i^2)\Rightarrow m_{X_i}(t)=e^{\mu_it+\sigma_i^2t^2/2}$

$m_Y(t)=e^{a_0t}\prod_{i=1}^{n}m_{X_i}(a_it)=e^{a_0t}\prod_{i=1}^{n}e^{\mu_ia_it+\sigma_i^2a_i^2t^2/2}=e^{a_0t+\sum_{i=1}^{n}(a_i\mu_it+a_i^2\sigma_i^2t^2/2)}=$

$=e^{(a_0+\sum_{i=1}^{n}a_i\mu_i)t+(\sum_{i=1}^{n}a_i^2\sigma_i^2)t^2/2}$

$\Rightarrow Y\sim N(a_0+\sum_{i=1}^{n}a_i\mu_i,\ \sum_{i=1}^{n}a_i^2\sigma_i^2)$ $\Rightarrow$ Linjärkomb. av ober. normalf. S.V. är även de normalf.

Övning 9

Skattning $\hat{\theta}$ av θ vvr om $E[\hat{\theta}]=\theta$

Om $X_1,X_2,\ldots,X_n$ från förd. med μ,σ^2

$\bar{X}$ vvr av μ, s^2 vvr av σ^2, $Var(\bar{X})=\frac{\sigma^2}{n}$

7.49 Låt X = # neutroner som frigörs under fission av plutonium-239.

Stickprov: 3 2 2 2 2 3 ....
3
3
⋮

(a) Är X normalförd.? Nej!, ty X diskret.

(b) Skatta $\mu = E[X]$: $\bar{X} = \frac{3+2+2+\cdots}{40} = \frac{112}{40} = 2{,}8$

(c) Antag $\sigma = 0{,}5$, Konstruera 99% k.i. för μ.

CGS $\Rightarrow \frac{\bar{X}-\mu}{\sigma/\sqrt{n}} \approx N(0,1)$

Här: $\frac{2{,}8-\mu}{0{,}5/\sqrt{40}} \approx N(0,1)$

$$P\left(-2{,}575 \leq \frac{\bar{X}-\mu}{\sigma/\sqrt{n}} \leq 2{,}575\right) = 0{,}99$$

$$\Leftrightarrow P\left(\bar{X} - 2{,}575\frac{\sigma}{\sqrt{n}} \leq \mu \leq \bar{X} + 2{,}575\frac{\sigma}{\sqrt{n}}\right) = 0{,}99$$

$\Rightarrow \bar{X} \pm 2{,}575\frac{\sigma}{\sqrt{n}} = 2{,}8 \pm 2{,}575 \cdot \frac{0{,}5}{\sqrt{40}} = 2{,}8 \pm 0{,}2036$, 99% k.i. för μ

(d) Motsäger detta ett värde på $\mu = 3{,}0$?

Nej!, ($\mu = 3{,}0$ finns i intervallet)

7.61 Låt $Y_1, Y_2, \ldots, Y_n$ vara stickprov från $Poi(\lambda s/n)$

(a) Visa att $X = \sum_{i=1}^{n} Y_i \sim Poi(\lambda s)$

Bevis : Om $Z \sim Poi(\lambda) \Rightarrow m_Z(t) = e^{\lambda(e^t-1)}$

$Y_i \sim Poi(\lambda s/n) \Rightarrow m_{Y_i}(t) = e^{\lambda s/n (e^t-1)}$

$m_X(t) = E[e^{tX}] = E[e^{t\sum Y_i}] = E[e^{tY_1}]E[e^{tY_2}]\cdots E[e^{tY_n}] =$

$= \prod_{i=1}^{n} m_{Y_i}(t) = \prod_{i=1}^{n} e^{(\lambda s/n)(e^t-1)} = e^{\lambda s(e^t-1)} \Rightarrow X \sim Poi(\lambda s)$

(b) Använd CGS för att hitta approx. förd. för Y

CGS $\Rightarrow \bar{Y} \approx N\left(E[Y_i], \frac{Var(Y_i)}{n}\right) = N\left(\frac{\lambda s}{n}, \frac{\lambda s}{n^2}\right)$

(c) Visa att $X \approx N(\lambda s, \lambda s)$

$X = n\bar{Y} \Rightarrow E[X] = nE[\bar{Y}] = \lambda s$

$Var(X) = n^2 Var(\bar{Y}) = \lambda s$

$\Rightarrow X \approx N(\lambda s, \lambda s)$

8.1 Stickprov

1.48 1.26 .....
1.30
1.51
⋮

30 obs.

$\Sigma x_i = 45{,}34$, $\Sigma x_i^2 = 68{,}8984$

(b) vvr-skattning av σ^2?

$$s^2 = \frac{1}{n-1}\left(\Sigma x_i^2 - n\bar{x}^2\right) = \frac{1}{29}\left(68{,}8984 - 30\left(\frac{45{,}34}{30}\right)^2\right) = 0{,}0129154023$$

(c) Ge 95% k.i. för σ^2

$$\frac{(n-1)S^2}{\sigma^2} \sim \chi^2(n-1) = \chi^2(29)$$

$$\Rightarrow P\left(\chi^2_{1-\frac{\alpha}{2}}(29) \leq \frac{(n-1)S^2}{\sigma^2} \leq \chi^2_{\frac{\alpha}{2}}(29)\right) = 1-\alpha$$

$$\Rightarrow P\left(\chi^2_{0{,}975}(29) \leq \frac{(n-1)S^2}{\sigma^2} \leq \chi^2_{0{,}025}(29)\right) = 0{,}95$$

$$\Rightarrow P\left(\frac{(n-1)S^2}{\chi^2_{0{,}025}(29)} \leq \sigma^2 \leq \frac{(n-1)S^2}{\chi^2_{0{,}975}(29)}\right) = 0{,}95$$

$\left(\frac{(n-1)S^2}{\chi^2_{0{,}975}(29)}, \frac{(n-1)S^2}{\chi^2_{0{,}025}(29)}\right)$ är ett 95% k.i. för $\sigma^2 = \left(\frac{29s^2}{45{,}7}, \frac{29s^2}{16}\right) =$

$= (0{,}0082, 0{,}0234)$

(d) 95% k.i. för $\sigma = (0{,}091, 0{,}153)$

(e) Blir du förvånad om någon påstår att $\sigma > 0{,}2$?

Ja, 0,2 ligger utanför intervallet i (d).

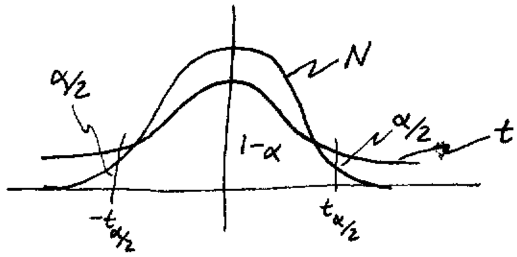
8.10

2,0 1,7 2,6 1,5 1,4
2,1 3,0 2,5 1,8 1,4

(a) $\bar{x} = 2$

$S^2 = 0{,}30$, $s = 0{,}55$

(b) 90% k.i. för μ



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

$$\frac{\bar{X}-\mu}{S/\sqrt{n}} \sim t(n-1) = t(9)$$

$$P\left(-t_{\alpha/2}(9) \leq \frac{\bar{X}-\mu}{S/\sqrt{n}} \leq t_{\alpha/2}(9)\right) = 1-\alpha \Rightarrow P\left(-t_{0{,}05}(9) \leq \frac{\bar{X}-\mu}{S/\sqrt{n}} \leq t_{0{,}05}(9)\right) = 0{,}90$$

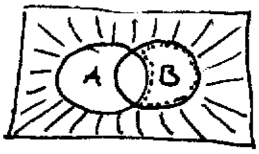
$\Rightarrow \bar{X} \pm t_{0{,}05}(9)\frac{S}{\sqrt{n}} = 2 \pm 1{,}833\,\frac{0{,}55}{\sqrt{10}} = 2 \pm 0{,}32$, 90% k.i. för μ

Tenta 980117.1

A, B oberoende händelser

Visa (i) A^c, B oberoende

(ii) A^c, B^c oberoende



Bevis A, B ober. $\Leftrightarrow P(A\cap B)=P(A)P(B)$

$P(A^c)=1-P(A)$, $P(B^c)=1-P(B)$

(i) Visa att $P(A^c\cap B)=P(A^c)P(B)$

$$P(A^c\cap B)=P(B)-P(A\cap B)\underset{\uparrow_{A,B\text{ ober.}}}{=}P(B)-P(A)P(B)=$$

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

$\Rightarrow A^c, B$ ober.

(ii) A, B ober. $\underset{(i)}{\Rightarrow}$ A^c, B ober.

B, A^c ober. $\Rightarrow$ B^c, A^c ober.

alt: Visa att $P(A^c\cap B^c)=P(A^c)P(B^c)$

$$P(A^c\cap B^c)=1-P(A\cup B)=1-(P(A)+P(B)-P(A\cap B))=$$

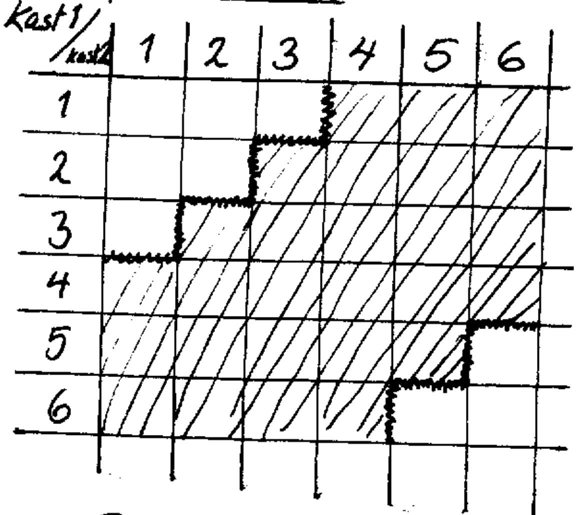
$$\overset{A,B\text{ ober.}}{=}1-P(A)-P(B)+P(A)P(B)=P(A^c)-P(B)(1-P(A))=$$

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

Tenta 990117.4

En symmetrisk tärning kastas 2 ggr.

Utfallsrum Ω



(a) $P(\text{Summan}>4, \text{ men } \leq 10)=\frac{27}{36}=\frac{3}{4}=0{,}75$

(b) $P(\text{minst en "4"})=P(\{\text{kast }1=4\}\cup\{\text{kast }2=4\})=$

$=\{P(A\cup B)=P(A)+P(B)-P(A\cap B)\}=P(\text{kast }1=4)+P(\text{kast }2=4)-P(\text{båda kasten }=4)=$

$=\frac{1}{6}+\frac{1}{6}-\frac{1}{36}=\frac{11}{36}$

(c) $P(\text{exakt ett av kasten}>4)=P(\text{kast}1>4\cap\text{kast}2\leq 4)+P(\text{exakt }1\leq 4\cap\text{kast }2>4)=$

$=\frac{2}{6}\cdot\frac{4}{6}+\frac{4}{6}\cdot\frac{2}{6}=\frac{16}{36}=\frac{4}{9}$

Övning 10.

8.22 Genomsnittsnivån på bakgrundsstrålning i USA är 0,3 rem/år Nu fruktar man att den ökade användningen av radioaktivt material har ökat denna nivå.

(a) Sätt upp lämplig noll-och mothypotes för att bekräfta detta.

$H_0: \mu = 0{,}3$

$H_1: \mu > 0{,}3$

(b) Förklara konsekvenserna av Typ I & II -fel

Typ I-fel = förk. H_0 då H_0 sann.

Typ II-fel = förk. ej H_0 då H_1 sann.

t.ex. Typ I ⇒ man tror att nivån ligger $>0{,}3$ när den bara ligger på 0,3, kanske sätter man då in onödiga åtgärder för att minska den.

Typ II ⇒ tror att bakg. strålningen bara är $=0{,}3$ när den i själva verket är högre.

8.27 Man tror att en majoritet av alla fel som uppstår i ett visst system beror på "oxidkretsen". Man gör 15 test och låter $X = \#$ fel pga oxidkretsen.

(a) $H_0: p = 0{,}5 \quad (p \leq 0{,}5)$

$H_1: p > 0{,}5$

(b) Om H_0 sann vad är då $E[X]$?

H_0 sann $\Rightarrow p = 0{,}5 \Rightarrow X \sim Bin(n=15, p=0{,}5)$

$\Rightarrow E[X] = np = 15 \cdot 0{,}5 = 7{,}5$

(c) Förkasta H_0 då $X \geq 11$. sign nivån α?

$\alpha = P(\text{förk. } H_0 \text{ då } H_0 \text{ sann}) = P(X \geq 11 \text{ då } p = 0{,}5) =$

$= \binom{15}{11} 0{,}5^{15} + \binom{15}{12} 0{,}5^{15} + \binom{15}{13} 0{,}5^{15} + \binom{15}{14} 0{,}5^{15} + \binom{15}{15} 0{,}5^{15} = 0{,}0592$

(d) Hitta β då $p = 0{,}6$, $0{,}7$, $0{,}8$, $0{,}9$

$p = 0{,}6 \Rightarrow X \sim Bin(15, 0{,}6)$

$\beta = P(\text{förk. ej } H_0 \text{ då } p = 0{,}6) = P(X \leq 10 \text{ då } p = 0{,}6) =$

$= \sum_{x=0}^{10} \binom{15}{x} 0{,}6^x 0{,}4^{15-x} = 0{,}7827$

$p = 0{,}7 \Rightarrow X \sim Bin(15, 0{,}7) \Rightarrow \beta = \sum_{x=0}^{10} \binom{15}{x} 0{,}7^x 0{,}3^{15-x} = 0{,}4845$

$p = 0{,}8 \Rightarrow \beta = 0{,}1642$, $p = 0{,}9 \Rightarrow \beta = 0{,}0127$

(e) Hitta styrkan

Styrkan $= P(\text{förk. } H_0 \text{ då } H_1 \text{ sann}) = 1-\beta$

$p=0,6 \Rightarrow 0,2173$

$p=0,7 \Rightarrow 0,5155$

$p=0,8 \Rightarrow 0,8358$

$p=0,9 \Rightarrow 0,9873$

8.33 Man tror att över 15% av alla smältugnar som används i USA är martinugnar. För att verifiera detta undersöks 40 ugnar.

(a) $H_0: p=0,15$

$H_1: p>0,15$

(b) Efter att ha samlat data fås $X=9$, p-värde? Ska H_0 förk.?

p-värde = den lägsta nivån på α vi kan förk. H_0 =

= P(få lika extremt eller extremare resultat om H_0 sann) =

$= P(\underset{=X>8}{X \geq 9} \text{ då } p=0,15)$

$P=0,15 \Rightarrow X \sim Bin(n=40, p=0,15) \approx N(6, 5,1)$

$P(X>8,5) = 1-P(X\leq 8,5) = 1-P\left(\frac{X-6}{\sqrt{5,1}} \leq \underbrace{\frac{8,5-6}{\sqrt{5,1}}}_{\approx 1,11}\right) = 1-\phi(1,11) = 1-0,8665 =$

$= 0,1335 \Rightarrow$ Förk. ej!

8.44 6,2 6,5 7.6 7.7 7.0
7.0 7,2 6.8 7,5 8,1
7,1 7.0 7,1 7,8 8,5 är 15 mätningar på pH-värdet i vatten efter en viss behandling. Finns det bevis på att behandlingen inte ger önskat resultat, dvs $\mu=7$?

Testa: $H_0: \mu=7$

$H_1: \mu \neq 7$ 2 sidigt test.

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

$\frac{\bar{X}-\mu}{S/\sqrt{n}} \sim t(n-1) = t(14)$

Om H_0 sann ($\mu=7$) så är $T = \frac{\bar{X}-7}{S/\sqrt{n}} \sim t(14)$

$\Rightarrow P(-1,761 \leq T \leq 1,761) = 0,9$

$\Rightarrow$ Vi kan förk. H_0 på 10%-nivå om $T<-1,761$ eller $>1,761$

Här: $T = \frac{\bar{x}-7}{s/\sqrt{n}} = \frac{7,2733-7}{0,6017/\sqrt{15}} = 1,759$

$\Rightarrow$ Vi kan ej förk. H_0 på 10% nivå

<u>8.52</u>

X_i	X_i-68	Tecken
65	−3	−
74	6	+
76	2	+
79	1	+
73	5	+
74	6	+
66	−2	−
70	2	+
72	4	+
66	−2	−
72	4	+
73	5	+
71	3	+
68	0	0 ←
67	−1	−
70	2	+
68	0	0 ←
69	1	+
73	5	+
74	6	+

20 obs. på "höjden"

Tror att medianen >68, stöder data detta antagande?

M = medianen

Testa $H_0: M=68$ mot $H_1: M>68$

Om H_0 sann $P(X_i<68)=P(X_i>68)=0,5$

Låt $Q_+ = \#$ '+'-tecken

Om H_0 sann $Q_+ \sim Bin(20, p=0,5)$

Förk. H_0 om Q_+ stor.

<u>Alt1</u> Klassa 0:or som '−'-tecken $\Rightarrow Q_+=14$

P-värde $= P(Q_+ \geq 14$ då H_0 sann$) = \sum_{x=14}^{20}\binom{20}{x}05^{20} = 0,058$

$\Rightarrow$ kan ej förk. H_0 på 5%-nivå men på 6%-nivå

<u>Alt2</u> Ta bort "0:or"

$\Rightarrow$ 18 obs. H_0 sann $Q_+ \sim Bin(18, 0,5)$

P-värde $= P(Q_+ \geq 14$ då H_0 sann$) = \sum_{x=14}^{18}\binom{18}{x}0,5^{18} = 0,016$

$\Rightarrow$ förk. H_0 på 5%-nivå

<u>Alt3</u> 2 "0-observationer"

Klassa en som "+" och en som "−".

$\Rightarrow Q_+=15$ dvs. $Q_+ \sim Bin(20, 0,5)$ då H_0 sann

$\Rightarrow$ P-värde $= P(Q_+ \geq 15$ då H_0 sann$) = \sum_{x=15}^{20}\binom{20}{x}0,5^{20} = 0,021$

$\Rightarrow$ Förk. H_0 på 5%-nivå

<u>Tenta 980117.2</u>

ξ, η 2 ober. kont. S.V. med täthetsf. f_ξ, f_η

Låt $f_{\xi+\eta}$ vara täthetsf. för $\xi+\eta$

Visa att $f_{\xi+\eta}(x) = \int_{-\infty}^{\infty} f_\xi(u) f_\eta(x-u)\,du$

<u>Bevis</u> $F_{\xi+\eta}(x) = P(\xi+\eta \leq x) = \iint_{u+v\leq x} f_{\xi,\eta}(u,v)\,dv\,du = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} f_{\xi,\eta}(u,v)\,d\,du =$

$= \int_{-\infty}^{\infty}\int_{-\infty}^{x-u} f_\xi(u) f_\eta(v)\,dv\,du = \int_{-\infty}^{\infty} f_\xi(u) \cdot \underbrace{\int_{-\infty}^{x-u} f_\eta(v)\,dv}_{F_\eta(x-u)}\,du = \int_{-\infty}^{\infty} f_\xi(u) F_\eta(x-u)\,du$

$\Rightarrow f_{\xi+\eta}(x) = \frac{d}{dx}\int_{-\infty}^{\infty} f_\xi(u) F_\eta(x-u)\,du = \int_{-\infty}^{\infty} \frac{d}{dx}\left(f_\xi(u) F_\eta(x-u)\right)du = \int_{-\infty}^{\infty} f_\xi(u) f_\eta(x-u)\,du$

Kapitel 9:
INFERENS BASERAD PÅ PROPORTIONER

Antag att vi har en viss sorts element som antingen är av "typ A" eller ej av typ A
Låt $p =$ verkliga andelen som är av typ A
Då kan p skattas genom följande förfarande:
Välj ut ett stickprov (av storlek n) av element och notera

$$X = \text{antal av dessa som är av typ A}$$

Vi kan då punktskatta p med

$$\hat{p} = \frac{X}{n}$$

Denna skattning är VVR, eftersom $X \sim Bin(n,p)$ vilket ger

$$E[\hat{p}] = E[\frac{X}{n}] = \frac{1}{n}E[X] = \frac{1}{n}np = p$$

Vidare är

$$Var(X) = \frac{1}{n^2}Var(X) = \frac{1}{n^2}np(1-p) = \frac{p(1-p)}{n}$$

$$\text{CGS} \Rightarrow \hat{p} \sim N(p, \tfrac{p(1-p)}{n}) \Rightarrow \frac{\hat{p}-p}{\sqrt{p(1-p)/n}} \sim N(0,1)$$

Därför är

$$\hat{p} \pm z_{\frac{\alpha}{2}}\sqrt{p(1-p)/n}$$

ett $100(1-\alpha)\%$ k.i. för p som (eftersom p okänd) approximeras med

$$\hat{p} \pm z_{\frac{\alpha}{2}}\sqrt{\hat{p}(1-\hat{p})/n}$$

Tittar vi istället på 2 proportioner p_1 och p_2, som skattas med $\hat{p_1} = \frac{X_1}{n_1}$ och $\hat{p_2} = \frac{X_2}{n_2}$, så skattas skillnaden $p_1 - p_2$ med

$$\hat{p_1} - \hat{p_2} = \frac{X_1}{n_1} - \frac{X_2}{n_2}$$

som har

$$E[\hat{p_1} - \hat{p_2}] = E[\hat{p_1}] - E[\hat{p_2}] = p_1 - p_2$$

och

$$Var(\hat{p_1} - \hat{p_2}) = Var(\hat{p_1}) + Var(\hat{p_2}) = \frac{p_1(1-p_1)}{n_1} + \frac{p_2(1-p}{n_2}$$

Därför fås $100(1-\alpha)\%$ k.i. för skillnaden genom

$$\hat{p_1} - \hat{p_2} \pm z_{\frac{\alpha}{2}}\sqrt{\hat{p_1}(1-\hat{p_1})/n_1 + \hat{p_2}(1-\hat{p_2})/n_2}$$

Vill man testa en hypotes $H_0 : p = p_0$ tittar man på

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

alt $H_1 : p > p_0$ - Förkasta H_0 för stora Z
alt $H_1 : p < p_0$ - Förkasta H_0 för små Z
alt $H_1 : p \neq p_0$ - Förkasta H_0 för stora och små Z

Testas istället $H_0 : p_1 - p_2 = (p_1 - p_2)_0$ tittar man på

$$Z = \frac{\hat{p_1} - \hat{p_2} - (p_1 - p_2)_0}{\sqrt{\hat{p_1}(1-\hat{p_1})/n_1 + \hat{p_2}(1-\hat{p_2})/n_2}}$$

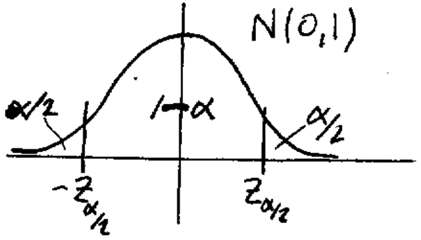
Obs Vid test av $H_0 : p_1 - p_2 = 0$, dvs $p_1 = p_2$ används istället

$$Z = \frac{\hat{p_1} - \hat{p_2}}{\sqrt{\hat{p}(1-\hat{p})(\frac{1}{n_1} + \frac{1}{n_2})}},$$

där

$$\hat{p} = \frac{n_1\hat{p_1} + n_2\hat{p_2}}{n_1 + n_2}$$

Övning 11



$\phi(z_{\alpha/2}) = 1-\frac{\alpha}{2}$

$\bar{X}_1,\dots,\bar{X}_n$

n stort $\Rightarrow \bar{X} \sim N(\mu, \frac{\sigma^2}{n})$

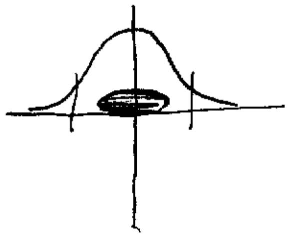
$\Rightarrow \frac{\bar{X}-\mu}{\sigma/\sqrt{n}} \sim N(0,1) \Rightarrow P(-z_{\alpha/2} \leq \frac{\bar{X}-\mu}{\sigma/\sqrt{n}} \leq z_{\alpha/2}) = 1-\alpha$

$\Leftrightarrow P(\bar{X} - z_{\alpha/2}\frac{\sigma}{\sqrt{n}} \leq \mu \leq \bar{X} + z_{\alpha/2}\frac{\sigma}{\sqrt{n}}) = 1-\alpha$

$P(\frac{\bar{X}-\mu}{\sigma/\sqrt{n}} \geq -z_{\alpha}) = 1-\alpha \Rightarrow P(\mu \leq \bar{X} + z_{\alpha}\frac{\sigma}{\sqrt{n}}) = 1-\alpha$

$P(\frac{\bar{X}-\mu}{\sigma/\sqrt{n}} \leq z_{\alpha}) = 1-\alpha \Rightarrow P(\mu \geq \bar{X} - z_{\alpha}\frac{\sigma}{\sqrt{n}}) = 1-\alpha$

$H_0 : \mu = \mu_0 \Rightarrow \frac{\bar{X}-\mu}{\sigma/\sqrt{n}} \sim N(0,1)$ om sann



$H_1 : \mu \neq \mu_0 \Rightarrow$

Om vi ej kan förkasta H_0 så betyder det inte att H_0 är sann!

9.2 I ett test visade det sig att 75 av 193 fel i ett elektriskt system berodde på mekaniska delar.

Låt p = andelen fel pga. mek. delar

(a) p skattas med $\hat{p} = \frac{75}{193} \approx 0,39$

(b) $\hat{p} \pm z_{\alpha/2}\sqrt{\hat{p}(1-\hat{p})/n}$ är ett $100(1-\alpha)\%$ k.i.

$\Rightarrow 95\%$ k.i. i $\hat{p} \pm z_{0,025}\sqrt{\frac{\hat{p}(1-\hat{p})}{n}} = 0,39 \pm 1,96\sqrt{\frac{0,39 \cdot 0,61}{193}} = 0,39 \pm 0,069$

(c) Hur stort stickprov som behövs för att $\hat{p} \pm 0,03$ ska ge 95% k.i.

$\Rightarrow 1,96\sqrt{\frac{\hat{p}(1-\hat{p})}{n}} = 0,03 \Rightarrow n = \frac{1,96^2 \cdot \hat{p}(1-\hat{p})}{0,03^2} = 1014,14 \Rightarrow \underline{\underline{n = 1015}}$

9.16 Man tror att mer än 80% av alla störningar beror på brus.

(a) Vill visa $p > 0,8$

Testa $H_0 : p = 0,8$ mot $H_1 : p > 0,8$

(b) Hitta kritiska punkten för $\alpha = 0{,}01$ - nivå test

H_0 sann $\Rightarrow$ $Z = \dfrac{\hat{p} - 0{,}8}{\sqrt{0{,}8 \cdot 0{,}2/n}} \sim N(0,1) \Rightarrow P(Z > 2{,}33) = 0{,}01$

Förk. H_0 då $Z > 2{,}33 \Leftrightarrow \dfrac{\hat{p} - 0{,}8}{\sqrt{0{,}8 \cdot 0{,}2/n}} > 2{,}33$

$\Leftrightarrow \hat{p} > 0{,}8 + 0{,}932/\sqrt{n}$

(c) $n = 150$, 133 pga brus

Förk. H_0 på 1% nivå om $\hat{p} > 0{,}8 + 0{,}932/\sqrt{150} = 0{,}876$

Här: $\hat{p} = \dfrac{133}{150} = 0{,}887 > 0{,}876$

$\Rightarrow$ Förk. H_0 på 1% - nivå

9.18 I 15 av 50 byggnadsprojekt i Dallas användes ett visst ämne.
15 av 60 i Boston.

p_1 = andelen i Dallas

p_2 = andelen i Boston

(a) Skatta $p_1 - p_2$

$\hat{p}_1 = \dfrac{15}{50} = 0{,}3$ är vvr-sk. av p_1

$\hat{p}_2 = \dfrac{15}{60} = 0{,}25$ är vvr-sk. av p_2

$\Rightarrow \hat{p}_1 - \hat{p}_2 = 0{,}3 - 0{,}25 = 0{,}05$ vvr-sk. av $p_1 - p_2$

(b) 95% k.i. för $p_1 - p_2$

$$\frac{\hat{p}_1 - \hat{p}_2 - (p_1 - p_2)}{\sqrt{\frac{\hat{p}_1(1-\hat{p}_1)}{n_1} + \frac{\hat{p}_2(1-\hat{p}_2)}{n_2}}} \sim N(0,1)$$

$\Rightarrow \hat{p}_1 - \hat{p}_2 \pm z_{\alpha/2}\sqrt{\dfrac{\hat{p}_1(1-\hat{p}_1)}{n_1} + \dfrac{\hat{p}_2(1-\hat{p}_2)}{n_2}}$ ett $100(1-\alpha)$% k.i. för $p_1 - p_2$

95% k.i. för $p_1 - p_2$

$$\hat{p}_1 - \hat{p}_2 \pm z_{0{,}025}\sqrt{\frac{\hat{p}_1(1-\hat{p}_1)}{n_1} + \frac{\hat{p}_2(1-\hat{p}_2)}{n_2}} = 0{,}3 - 0{,}25 \pm 1{,}96\sqrt{\frac{0{,}3 \cdot 0{,}7}{50} + \frac{0{,}25 \cdot 0{,}75}{60}} =$$

$= 0{,}05 \pm 0{,}168 \quad = \quad (-0{,}118 , 0{,}218)$

9.26 p_1 = andelen fel vid användning av karbon

p_2 = ———— // ———— laserteknik

Man byter till laserteknik om den reducerar andelen fel med mer än 0,02

(a) Sätt upp noll- och mothypotes för att stödja lasertekniken.

$H_0: p_1 - p_2 = 0{,}02$

$H_1: p_1 - p_2 > 0{,}02$

(b) Hitta kritiska värdet för $\alpha = 0{,}05$-test.

$$Z = \frac{\hat{p}_1 - \hat{p}_2 - 0{,}02}{\sqrt{\frac{\hat{p}_1(1-\hat{p}_1)}{n_1} + \frac{\hat{p}_2(1-\hat{p}_2)}{n_2}}} \sim N(0,1) \text{ om } H_0 \text{ sann}$$

Under H_0 $P(Z > 1{,}645) = 0{,}05 \Rightarrow$ Förk. H_0 på 0,05-nivå

om $Z > 1{,}645 \Leftrightarrow \hat{p}_1 - \hat{p}_2 > 0{,}02 + 1{,}645\sqrt{\frac{\hat{p}_1(1-\hat{p}_1)}{n_1} + \frac{\hat{p}_2(1-\hat{p}_2)}{n_2}}$

(c) 100 test med varje metod, 5 fel med karbon, 1 fel med laser.

$\hat{p}_1 = \frac{5}{100} = 0{,}05$, $\hat{p}_2 = \frac{1}{100} = 0{,}01 \Rightarrow \hat{p}_1 - \hat{p}_2 = 0{,}04$

Kritiska värdet $= 0{,}02 + 1{,}645\sqrt{\frac{0{,}05 \cdot 0{,}95}{100} + \frac{0{,}01 \cdot 0{,}99}{100}} \approx 0{,}059$

$\hat{p}_1 - \hat{p}_2 = 0{,}04 < 0{,}059 \Rightarrow$ kan ej förk. H_0 på 5%-nivå

Tenta 980117.5

I kommunikationssystem sänder man ett meddelande om 5 binära tecken (1:or och 0:or) "1" ingår alltid ett jämt antal ggr. Slh. för felaktig överföring av ett tecken är 0,01. Händelserna att olika tecken överförs felaktigt är oberoende. Mottagaren klassar meddelandet som felaktigt <u>omm.</u> det innehåller tecknet "1" udda antal ggr.

P(meddelande fel | klassas korrekt) =

$= \{$ Låt X = # felaktigt överförda tecken $\Rightarrow X \sim Bin(n=5, 0{,}01)$

$\Rightarrow P(X=x) = \binom{5}{x} 0{,}01^x 0{,}99^{5-x}$, $x = 0,1,\ldots,5$ $\} =$

$$= \frac{P(\overset{(1,2,3,4,5 \text{ fel})}{\text{medd. fel}} \cap \overset{(0,2,4 \text{ fel})}{\text{klassas korrekt}})}{P(\text{klassas korrekt})} = \frac{P(2 \text{ eller } 4 \text{ fel})}{P(0,2 \text{ eller } 4 \text{ fel}} =$$

$$= \frac{P(X=2) + P(X=4)}{P(X=0) + P(X=2) + P(X=4)} = 0{,}001$$

Övning 12

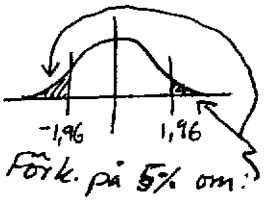
Tenta tips II!

Hypotesprövning : H_0 , H_1

Medicin, testa om blodtryckssänkande

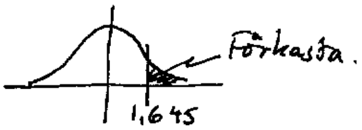
$X_1, \ldots, X_n$, σ känd

$H_0 : \mu = 0$
$H_1 : \mu \neq 0$ $\Rightarrow$



Rimligare:

$H_0 : \mu = 0$
$H_1 : \mu > 0$ $\Rightarrow$



∵ Ställ alltid upp noll- & mot-hypotes!

t.ex. $\bar{x}$, s^2 , $y = \alpha + \beta x$. Det finns formler att sätta in i.
Skriv inte: "det blev det i alla fall enl. min ABCQTX152899"
= "det sa miniräknaren, jag är för dum för att sätta in det i formler"

$\mu, \sigma^2, p, p_1 - p_2$, $X_1, \ldots, X_n$ från förd. μ, σ^2

$E[\bar{X}] = \mu$; $Var(\bar{X}) = \frac{\sigma^2}{n}$

2 stickprov $n_1, n_2, \mu_1, \mu_2, \sigma_1^2, \sigma_2^2$

$E[\bar{X}_1 - \bar{X}_2] = \mu_1 - \mu_2$

$Var(\bar{X}_1 - \bar{X}_2) = Var(\bar{X}_1) + Var(\bar{X}_2) = \frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}$

10.4 $\bar{X}_1$ - medel från stickprov av storlek $25 \sim N(8, 16)$
$\bar{X}_2$ - ——— 〃 ——— $36 \sim N(5, 9)$

Fördelning för $\bar{X}_1$? $\bar{X}_2$? $\frac{(\bar{X}_1 - 8)}{(4/5)}$? $\frac{(\bar{X}_2 - 5)}{(3/6)}$? $\bar{X}_1 - \bar{X}_2$?

$\frac{(\bar{X}_1 - \bar{X}_2) - (8 - 5)}{\sqrt{16/25 + 9/36}}$?

Använd:

(i) Om $\bar{X}$ stickprovsmedel av storlek n från förd. med μ, σ^2
$\Rightarrow E[\bar{X}] = \mu$, $Var(\bar{X}) = \frac{\sigma^2}{n}$

(ii) Linj. komb. av ober. normalf. variabler är normalf.
$\Rightarrow X \sim N(\mu, \sigma^2) \Rightarrow \bar{X} \sim N(\mu, \frac{\sigma^2}{n})$

$$\bar{X}_1 \sim N(8, \tfrac{16}{25}) = N(8, 0{,}64) \Rightarrow \frac{\bar{X}_1 - 8}{\sqrt{16/25}} = \frac{\bar{X}_1 - 8}{4/5} \sim N(0,1)$$

$$\bar{X}_2 \sim N(5, \tfrac{9}{36}) = N(5, 0{,}25) \Rightarrow \frac{\bar{X}_2 - 5}{\sqrt{9/36}} = \frac{\bar{X}_2 - 5}{3/6} \sim N(0,1)$$

$$\mathcal{E}[\bar{X}_1-\bar{X}_2]=\mathcal{E}[\bar{X}_1]-\mathcal{E}[\bar{X}_2]=8-5=3$$

$$Var(\bar{X}_1-\bar{X}_2)=Var(\bar{X}_1)+Var(\bar{X}_2)=\frac{16}{25}+\frac{9}{36}=0{,}89$$

$$\Rightarrow \bar{X}_1-\bar{X}_2 \sim N\left(8-5,\ \frac{16}{25}+\frac{9}{36}\right)=N(3,0{,}89)$$

$$\Rightarrow \frac{(\bar{X}_1-\bar{X}_2)-(8-5)}{\sqrt{16/25+9/36}} \sim N(0,1)$$

10.14

Nytt	Gammalt
$n_1=61$	$n_2=61$
$\bar{x}_1=40$	$\bar{x}_2=29$
$s_1^2=24{,}9$	$s_2^2=22{,}7$

$\mu_1-\mu_2$

(a) Testa $H_0: \sigma_1^2=\sigma_2^2$ mot $H_1: \sigma_1^2\neq\sigma_2^2$ på $\alpha=0{,}2$-nivå

Om $\sigma_1^2=\sigma_2^2 \Rightarrow \frac{S_1^2}{S_2^2}\sim F(n_1-1,n_2-1)=F(60,60)$

Om $F\sim F(60,60) \Rightarrow \begin{cases} P(F\leq 1{,}395)=0{,}9 & (s.751) \\ P(F\geq 0{,}717)=0{,}1 & (s.746) \end{cases}$

$\Rightarrow P(0{,}717\leq F\leq 1{,}395)=0{,}8$ dvs om H_0 sann

$\Rightarrow P(0{,}717\leq \frac{S_1^2}{S_2^2}\leq 1{,}395)=0{,}8$

Så förk. H_0 på 0,2-nivå om $\frac{S_1^2}{S_2^2}<0{,}717$ eller $>1{,}395$

Här: $\frac{S_1^2}{S_2^2}=\frac{24{,}9}{22{,}7}\approx 1{,}1 \Rightarrow$ Kan ej förk. H_0 på 0,2-nivå.

(b) $S_p^2=\frac{(n_1-1)S_1^2+(n_2-1)S_2^2}{n_1+n_2-2}=23{,}8$

(c) $\bar{X}_1-\bar{X}_2\pm t_{\alpha/2}(n_1+n_2-2)\sqrt{S_p^2\left(\frac{1}{n_1}+\frac{1}{n_2}\right)}$ är ett $100(1-\alpha)\%$ k.i. för $\mu_1-\mu_2$

$\Rightarrow 95\%$ k.i. för $\mu_1-\mu_2$

$$\bar{X}_1-\bar{X}_2\pm \underset{\approx t_{0,025}(100)}{t_{0,025}(120)}\sqrt{S_p^2\left(\frac{1}{n_1}+\frac{1}{n_2}\right)}=40-29\pm 1{,}984\sqrt{23{,}8\left(\frac{1}{61}+\frac{1}{61}\right)}=$$

$=11\pm 1{,}76=(9{,}24,\ 12{,}76)$

10.21

Barn	Vuxna
$n_1=121$	$n_2=61$
$\bar{x}_1=2{,}6$	$\bar{x}_2=0{,}4$
$s_1^2=1{,}44$	$s_2^2=0{,}0121$

Man vill undersöka koncentrationen av Strontium-90 i benen hos barn och vuxna. Man tror att barn har en högre nivå.

(a) Sätt upp hypoteser: , μ_1 - hos barn, μ_2 - hos vuxna

Testa $H_0: \mu_1=\mu_2$ mot $H_1: \mu_1>\mu_2$

(b) Är det lämpligt att använda S_p^2?

Testa $H_0: \sigma_1^2=\sigma_2^2$ mot $H_1: \sigma_1^2 \neq \sigma_2^2$

om $\sigma_1^2=\sigma_2^2 \Rightarrow \frac{S_1^2}{S_2^2} \sim F(120,60)$

$\Rightarrow P(0,742 \leq \frac{S_1^2}{S_2^2} \leq 1,320)=0,8$

Vi får $\frac{S_1^2}{S_2^2} = \frac{1,44}{0,0121} = 119 \quad \Rightarrow$ Förk. på 0,2-nivå

Vidare om H_0 sann

$=P(0,68154 \leq \frac{S_1^2}{S_2^2} \leq 1,429)=0,9 \quad \Rightarrow$ Förk. på 0,1-nivå

(c) Testa $H_0: \mu_1=\mu_2$ $(\mu_1-\mu_2=0)$ mot $H_1: \mu_1 > \mu_2$ $(\mu_1-\mu_2>0)$

Antag att stickproven ej kommer från förd. med samma varians.

Använd:
$$\frac{(\bar{X}_1-\bar{X}_2)-(\mu_1-\mu_2)}{\sqrt{\frac{S_1^2}{n_1}+\frac{S_2^2}{n_2}}} \sim t_\gamma$$

Under H_0: $$T=\frac{(\bar{X}_1-\bar{X}_2)}{\sqrt{\frac{S_1^2}{n_1}+\frac{S_2^2}{n_2}}} \sim t_\gamma$$

$$\gamma=\frac{(S_1^2/n_1+S_2^2/n_2)^2}{\frac{(S_1^2/n_1)^2}{n_1-1}+\frac{(S_2^2/n_2)^2}{n_2-1}} \simeq 124 \quad \text{(använd 100)}$$

Vi får $$T=\frac{2,6-0,4}{\sqrt{\frac{1,44}{121}+\frac{0,0121}{61}}} \approx 20$$

Skall förkasta om $T=20$ är "orealistiskt" stort

Om H_0 sann $P(T \leq 3,391)=0,9995$ dvs. $P(T>3,391)=0,0005$

$\Rightarrow T=20>3,391 \Rightarrow$ Förk. H_0 på 0,0005-nivå

<u>10.34</u> 2 metoder för mätning av konc. av P_u^{239} (μ/ml)

Stickprov	Ny	Gammal	d
1	3,78	3,35	0,43
2	3,58	3,60	−0,02
3	3,77	3,41	0,36
4	3,82	3,69	0,13
5	3,67	3,48	0,19
6	3,66	3,50	0,16
7	3,48	3,33	0,15
8	3,63	3,64	−0,01
9	3,88	3,65	0,23
10	3,53	3,64	−0,11

Ger nya tekniken högre värden?

Testa: $H_0: \mu_n = \mu_G$ $(\mu_n - \mu_G = 0)$

mot $H_1: \mu_n > \mu_G$ $(\mu_n - \mu_G > 0)$

Har parvisa data, då bildar vi skillnaderna $d_1, d_2, \ldots, d_{10}$

Testa H_0 $\mu_d = 0$ mot $H_1: \mu_d > 0$

$\bar{d} = 0,151$ $s_d \approx 0,168$

$$\frac{\bar{d} - \mu_d}{s_d/\sqrt{n}} \sim t_{n-1} = t_9$$

så om H_0 sann $(\mu_d = 0)$ så $T = \frac{\bar{d}}{s_d/\sqrt{n}} \sim t_9$

$\Rightarrow P(T > 1,383) = 0,1$

$P(T > 1,833) = 0,05$

$P(T > 2,262) = 0,025$

$P(T > 2,821) = 0,01$

$P(T > 3,250) = 0,005$

Vi får $T = \frac{\bar{d}}{s_d/\sqrt{n}} = 2,847 > 2,821 \Rightarrow$ Förk. H_0 på 0,01-nivå.

<u>Tenta 980117.7.a,b,</u>

X = livslängd i timmar hos en viss typ av elkomponenter

$X \sim Exp(\lambda)$, λ okänd

6 observerade livslängder: 120, 118, 125, 130, 111, 119

(a) Punktskatta λ, $X \sim Exp(\lambda) \Rightarrow f(x) = \lambda e^{-\lambda x}$, $x > 0$

$F(x) = P(X \leq x) = 1 - e^{-\lambda x}$

$E[X] = 1/\lambda$, $Var(X) = 1/\lambda^2$

<u>Momentmetoden</u> skattar $E[X^k]$ med $\frac{1}{n}\sum_{i=1}^{n} X_i^k$

$\Rightarrow E[X] = \frac{1}{\lambda}$ med $\frac{1}{n}\sum X_i = \bar{X}$ $\Rightarrow$ λ skattas med $\frac{1}{\bar{X}}$

<u>Maximimetoden</u> skattar λ med det värde som maximerar

$$L(\lambda) = \prod_{i=1}^{n} f(X_i) = \prod_{i=1}^{n} \lambda e^{-\lambda X_i} = \lambda^n e^{-\lambda \sum X_i}$$

$\ln L(\lambda) = n \ln \lambda - \lambda \sum X_i$, $\frac{d \ln L(\lambda)}{d\lambda} = \frac{n}{\lambda} - \sum X_i$

$\frac{d \ln L(\lambda)}{d\lambda} = 0 \Rightarrow \lambda = \frac{n}{\sum X_i} = \frac{1}{\bar{X}}$, $\bar{x} = \frac{120 + 118 + \ldots + 119}{6} = 120,5$

Skatta λ med $\hat{\lambda} = \frac{1}{120,5} \approx 8,3 \cdot 10^{-3}$

(b) Låt $q_{120} = P(X \geq 120)$, skatta q_{120}! : $q_{120} = P(X \geq 120) = 1 - \overbrace{P(X < 120)}^{1 - e^{-\lambda \cdot 120}} = e^{-120\lambda}$

så skatta q_{120} med $e^{-120\hat{\lambda}} = e^{-120 \cdot \frac{1}{120,5}} \approx 0,37$

Övning 13

Tentatips III

Tryckfel i boken: s. 65 $X \sim Geo(p)$ minneslösheten

betyder: $P(X \geq x_1 + x_2 \mid X > x_1) = P(X \geq x_2)$

har visat:

95% k.i. för $\mu = (A, B)$

Vill testa $H_0 : \mu = 0$ mot $H_1 : \mu \neq 0$

Vi behöver inte utföra testet, om 0 ligger mellan A och B

Kan ej kasta H_0 på 0,05-nivå.

Ofta onödigt att både göra k.i. och hypotesprövning

Skriv ner formeln du använt!!!

11.11 Visa att (a) $\sum_{i=1}^{n} (x_i - \bar{x}) = 0$

(b) $\sum (x_i - \bar{x})(y_i - \bar{y}) = \sum (x_i - \bar{x}) y_i$

(c) $\sum (x_i - \bar{x})(y_i - \bar{y}) = \frac{(n \sum x_i y_i - \sum x_i \sum y_i)}{n}$

(a) $\sum (x_i - \bar{x}) = \sum x_i - \sum \bar{x} = \sum x_i - n\bar{x} = \sum x_i - n \cdot \frac{1}{n} \sum x_i = \sum x_i - \sum x_i = 0$

(b) $\sum (x_i - \bar{x})(y_i - \bar{y}) = \sum (x_i - \bar{x}) y_i - \sum (x_i - \bar{x}) \bar{y} = \sum (x_i - \bar{x}) y_i - \bar{y} \underbrace{\sum (x_i - \bar{x})}_{=0} =$

$= \sum (x_i - \bar{x}) y_i$

(c) $\sum (x_i - \bar{x})(y_i - \bar{y}) \overset{(b)}{\Rightarrow} \sum (x_i - \bar{x}) y_i = \sum x_i y_i - \sum \bar{x} y_i = \sum x_i y_i - \bar{x} \sum y_i =$

$= \sum x_i y_i - \frac{1}{n} \sum x_i \sum y_i = \frac{n \sum x_i y_i - \sum x_i \sum y_i}{n}$

11.23 $n = 10$; $\sum x_i = 16{,}75$; $\sum y_i = 170$; $\sum x_i^2 = 28{,}64$; $\sum y_i^2 = 2898$

$\sum x_i y_i = 285{,}626$ Skatta regressionslinjen $(y =)\ \mu_{Y|x} = \beta_0 + \beta_1 x$

β_1 skattas med $\hat{\beta}_1 = \frac{n \sum x_i y_i - \sum x_i \sum y_i}{n \sum x_i^2 - (\sum x_i)^2} = \frac{10 \cdot 285{,}625 - 16{,}75 \cdot 170}{10 \cdot 28{,}64 - 16{,}75^2} = 1{,}499$

och β_0 skattas med $\hat{\beta}_0 = \bar{y} - \hat{\beta}_1 \bar{x} = \frac{170}{10} - \hat{\beta}_1 \cdot \frac{16{,}75}{10} \simeq 14{,}489$

$\Rightarrow \mu_{Y|x} = 14{,}489 + 1{,}499x$

11.24 (a) $Var(Y_i) = \sigma^2$

(b) Standardav. för $\hat{\beta}_1$

(c) ——〃—— $\hat{\beta}_0$

(a) σ^2 skattas av $s^2 = \frac{1}{n-2} \sum_{i=1}^{n} (y_i - \hat{\beta}_0 - \hat{\beta}_1 x_i)^2 = \frac{1}{n-2} \left[\sum y_i^2 - \hat{\beta}_0 \sum y_i - \hat{\beta}_1 \sum x_i y_i \right] =$

$= 0{,}8360546506$

(b) $Var(\hat{\beta}_1) = \dfrac{\sigma^2}{\sum(x_i-\bar{x})^2}$

Skattas med $\dfrac{s^2}{\underbrace{\sum(x_i-\bar{x})^2}_{\sum x_i^2 - n\bar{x}^2}} = \dfrac{s^2}{0{,}58375} \approx 1{,}4322 \Rightarrow S_{\hat{\beta}_1} \approx 1{,}1968$

(c) $Var(\hat{\beta}_0) = \dfrac{\sigma^2 \sum x_i^2}{n\sum(x_i-\bar{x})^2}$

skattas med $\dfrac{s^2\sum x_i^2}{n\sum(x_i-\bar{x})^2} \approx 4{,}1019 \Rightarrow S_{\hat{\beta}_0} \approx 2{,}0253$

11.25 Testa $H_0 : \beta_1 = 0$ mot $H_1 : \beta_1 \neq 0$

$y = \beta_0 + \beta_1 x$

β_1 skattades av $\hat{\beta}_1 = 1{,}499$, där $\hat{\beta}_1 \sim N(\beta_1 \,;\, \sigma^2/\sum(x_i-\bar{x})^2)$

$$\Rightarrow \frac{\hat{\beta}_1 - \beta_1}{\sqrt{\sigma^2/\sum(x_i-\bar{x})^2}} \sim N(0,1)$$

$$\frac{\hat{\beta}_1 - \beta_1}{\sqrt{S^2/\sum(x_i-\bar{x})^2}} \sim t_{n-2} = t_8$$

Om H_0 sann: $T = \dfrac{\hat{\beta}_1}{\sqrt{S^2/\sum(x_i-\bar{x})^2}} \sim t_8$

Om $T \sim t_8$: $P(-3{,}355 \leq T \leq 3{,}355) = 0{,}99$

$\Rightarrow$ Förk. H_0 på 0,01-nivå om $T < -3{,}355$ el. $> 3{,}355$

$T = \dfrac{1{,}499}{1{,}1968} \approx 1{,}25 \Rightarrow$ Förk. ej H_0 på 0,01-nivå

11.51 x = andelen koppar i ett stickprov

y = Rockwellhårdhet

x	y
0,01	58,0
0,03	66.0
0,01	55,0
⋮	⋮

(b) Skatta korrelationen ρ

ρ skattas med $r = \dfrac{n\sum x_i y_i - \sum x_i \sum y_i}{\sqrt{(n\sum x_i^2 - (\sum x_i)^2)(n\sum y_i^2 - (\sum y_i)^2)}} \hookrightarrow \approx 0{,}586$

$n = 10$

$\sum x_i = 0{,}73$

$\sum y_i = 625{,}3$

$\sum x_i^2 = 0{,}0769$

$\sum y_i^2 = 39339{,}37$

$\sum x_i y_i = 47{,}04$

3.79 I videospel är spelarens uppgift att hitta en skatt gömd bakom en av 5 dörrar. Spelaren öppnar en dörr och får behålla skatten om den finns där, annars återvänder han till start och måste genom en FARLIG LABYRINT (!!) för att åter komma till dörrarna. Skatt hittas ⇒ spel slut

X = antalet försök som krävs.

$E[X]$? $P(X \geq 3)$? $P(X > 3)$?

Lösning: $P(X=x) = \left(\frac{4}{5}\right)^{x-1}\left(\frac{1}{5}\right)$, $x = 1, 2, \ldots \Rightarrow X \sim Geo\left(p = \frac{1}{5}\right)$

$X \sim Geo(p) \Rightarrow E[X] = \frac{1}{p} \Rightarrow E[X] = \frac{1}{(1/5)} = 5$

$P(X \leq 3) = P(X=1) + P(X=2) + P(X=3) = \left(\frac{1}{5}\right) + \left(\frac{4}{5}\right)\left(\frac{1}{5}\right) + \left(\frac{4}{5}\right)^2\left(\frac{1}{5}\right) = 0{,}488$

$P(X > 3) = 1 - P(X \leq 3) = 1 - 0{,}488 = 0{,}512$

Övning 14

om man vill jämföra $\mu_1 = \mu_2 = \mu_3$

$\mu_i = \mu_j$ 5%-nivå, om vi gör flera på varandra $\Rightarrow \{$ Vi kommer att förk. felaktigt!

istället Variansanalys.

13.5 (a) Mätningar av giftigt avfall på 3 platser

Plats 1	Plats 2	Plats 3
15	19	22
26	15	26
20	10	24
20	26	26
29	11	15
28	20	17
21	13	24
26	15	
	18	
$n_1 = 8$	$n_2 = 9$	$n_3 = 7$

$N = 24$

$k = 3$

Testa $H_0: \mu_1 = \mu_2 = \mu_3$ mot $H_1: \mu_i \neq \mu_j$, ngt i, j på 0,05-nivå

ANOVA-tabell

Variationskälla	Frihetsgrader	SS	MS	F
Mellan	$k-1 = 2$	$SS_{Tr} = 225{,}625$	$112{,}8125$	
Inom	$N-k = 21$	$SS_E = 478{,}875$	$22{,}8036$	
Total	$N-1 = 23$	$SS_{Tot} = 704{,}5$		

$\bar{Y}_{1.} = 23{,}125$, $\bar{Y}_{2.} = 16{,}33$, $\bar{Y}_{3.} = 22$, $\bar{Y}_{..} = 20{,}25$

$SS_{Tr} = \sum_{i=1}^{k} n_i(\bar{Y}_{i.} - \bar{Y}_{..})^2 = 225{,}625$, $SS_{Tot} = \sum_{i=1}^{k}\sum_{j=1}^{k}(Y_{ij} - \bar{Y}_{..})^2 = 704{,}5$

$SS_E = SS_{Tot} - SS_{Tr} = 478{,}875$

Om H_0 sann: $F = \frac{MS_{Tr}}{MS_E} \sim F(2,21)$

Förk. H_0 på 0,05-nivå om $F > 3,467$

Här: $F = 4,95 > 3,467 \Rightarrow$ Förk. H_0 på 0,05-nivå

15.1 En studie görs för att avgöra om allmänheten är för ett dammbygge. Man tror att 40% för, 30% bryr sig ej, 20% emot, 10% kan ej uttala sig. Man intervjuar 150 personer. Vad blir förv. antal i varje kategori om ovanst. siffror stämmer?

$E[\text{för}] = 0,4 \cdot 150 = 60$

$E[\text{neutral}] = 0,3 \cdot 150 = 45$

$E[\text{emot}] = 0,2 \cdot 150 = 30$

$E[\text{ej uttal.}] = 0,1 \cdot 150 = 15$

Fick: 42, 61, 33, 14

15.4 Testa $H_0: p_1 = 0,4\,,\ p_2 = 0,3\,,\ p_3 = 0,2\,,\ p_4 = 0,1$

Observerade O_i	Förv. E_i
42	60
61	45
33	30
14	15

$$Q = \sum_{i=1}^{4} \frac{(O_i - E_i)^2}{E_i} \overset{H_0}{\sim} \chi^2_{4-1} = \chi^2_3$$

Förk. H_0 om Q stor

Om $Q \sim \chi^2_3 \Rightarrow P(Q > 11,3) = 0,01 \Rightarrow$ Förk. H_0 på 0,01-nivå om $Q > 11,3$

$$Q = \frac{(42-60)^2}{60} + \frac{(61-45)^2}{45} + \frac{(33-30)^2}{30} + \frac{(14-15)^2}{15} = 11,46 > 11,3$$

$\Rightarrow$ Förk. H_0 på 0,01-nivå.

Fel i boken 11.9 (b) ska vara $-0,0032 \pm 25,4428x$

ex. Antag att 7.2, 5.3, 6.2, 6.0, 5.3 är 5 observationer från $N(\mu, \sigma^2)$-förd., där μ okänd och $\sigma^2 = 0,50$

(a) Vad blir p-värdet, då man testar $H_0: \mu = 5,5$ mot $H_1: \mu \neq 5,5$?

(b) Vad blir p-värdet, då man testar $H_0: \mu = 5,5$ mot $H_1: \mu > 5,5$?

p-värdet = den lägsta nivå man kan förkasta H_0 på

P(få ett lika extremt eller extremare utfall då H_0 sann)

$$\bar{X} \sim N(\mu, \frac{\sigma^2}{n}) = N(\mu, 0.1) \Rightarrow \frac{\bar{X}-\mu}{\sqrt{0,1}} \sim N(0,1)$$

Om H_0 sann $(\mu = 5,5) \Rightarrow Z = \frac{\bar{X} - 5,5}{\sqrt{0,1}} \sim N(0,1)$

$\bar{x} = 6,0 \Rightarrow z = \frac{6,0-5,5}{\sqrt{0,1}} \approx 1,58$

(a) p-värdet $= P_{H_0}(Z \leq -1,58) + P_{H_0}(Z \geq 1,58) = \underbrace{\Phi(-1,58)}_{1-\Phi(1,58)} + 1 - \Phi(1,58) =$

$= 2 - 2\underbrace{\Phi(1,58)}_{=0,9429} = 0,1142$

(b) p-värdet $= P_{H_0}(Z \geq 1,58) = 1 - \Phi(1,58) = 0,0571$

ex Låt 5.1, 7.2, 8.3, 4.1, 6.0, 12.1, 1.2, 6.5, 5.3, 5.0 vara stickprov från förd. med okänt väntev. μ och varians $\sigma^2 = 8,0$
Skatta μ och ge 99% k.i.

μ skattas med $\bar{X}$

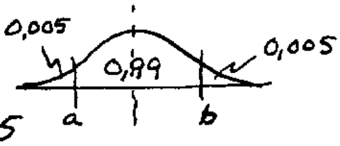
$\bar{X}$ vvr ty $E[\bar{X}] = E[\frac{1}{n}\sum X_i] = \frac{1}{n}\sum \underbrace{E[X_i]}_{\mu} = \frac{1}{n}\sum \mu = \frac{n}{n}\mu = \mu$

$Var(\bar{X}) = \frac{1}{n^2}\sum_1^n \underbrace{Var(X_i)}_{\sigma^2} = \frac{1}{n^2} \cdot n\,\sigma^2 = \frac{\sigma^2}{n}$

Om n stort CGS $\Rightarrow \bar{X} \sim N(\mu, \frac{\sigma^2}{n}) \Rightarrow \frac{\bar{X}-\mu}{\sigma/\sqrt{n}} \sim N(0,1)$

Om $Z \sim N(0,1)$ $P(a \leq Z \leq b) = 0,99$

$P(Z > b) = 0,005 \Rightarrow \Phi(b) = P(Z \leq b) = 0,995$



$\Phi(2,575) = 0,995 \Rightarrow b = 2,575 \Rightarrow a = -2,575$

$Z \sim N(0,1) \Rightarrow P(-2,575 \leq Z \leq 2,575) = 0,99$

$0,99 = P(-2,575 \leq \frac{\bar{X}-\mu}{\sigma/\sqrt{n}} \leq 2,575) = P(\bar{X} - 2,575\frac{\sigma}{\sqrt{n}} \leq \mu \leq \bar{X} + 2,575\frac{\sigma}{\sqrt{n}})$

$\Rightarrow (\bar{X} - 2,575\frac{\sigma}{\sqrt{n}}, \bar{X} + 2,575\frac{\sigma}{\sqrt{n}})$ alt. $\bar{X} \pm 2,575\frac{\sigma}{\sqrt{n}}$ är ett 99% k.i. för μ.

$\sigma^2 = 8$, $n = 10$, $\bar{x} = 6,08 \Rightarrow 6,08 \pm 2,575\frac{\sqrt{8}}{\sqrt{10}} = 6,08 \pm 2,31 = (3,77, 8,39,$

ex

Välj utan att titta en boll från varje urna och byt plats på dem. Dra en boll från urna1. Vad är S1h att den är vit?
Låt D=dragna bollens färg, F1 = flyttade bollen från urna1, F2=flyttade bollen fr. urna2

$P(D=vit) = P(D=vit|F1=vit, F2=vit)P(F1=vit, F2=vit) + P(D=vit|F1=vit,$
$, F2=svart)P(F1=vit, F2=svart) + P(D=vit|F1=svart, F2=vit)P(F1=svart,$
$, F2=vit) + P(D=vit|F1=svart, F2=svart)P(F1=svart, F2=svart) =$
$= \frac{1}{4}\cdot\frac{1}{4}\cdot\frac{1}{2} + 0\cdot\frac{1}{4}\cdot\frac{1}{2} + \frac{2}{4}\cdot\frac{3}{4}\cdot\frac{1}{2} + \frac{1}{4}\cdot\frac{3}{4}\cdot\frac{1}{2} = \frac{10}{32} = \frac{5}{16} = 0,3125$

Kapitel 5 (fortsättning): TVÅDIMENSIONELLA FÖRDELNINGAR

(X, Y) 2-dim s.v. med täthetsf $f_{XY}(x,y) \geq 0$

Väntevärde

$$E[h(X,Y)] = \sum_x \sum_y h(x,y) f_{XY}(x,y)$$

(diskreta fallet)

$$E[h(X,Y)] = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} h(x,y) f_{XY}(x,y) dy dx$$

(kontinuerliga fallet)

$$\Longrightarrow E[XY] = \sum_x \sum_y xy f_{XY}(x,y)$$

och

$$E[X+Y] = \sum_x \sum_y (x+y) f_{XY}(x,y)$$

i diskreta fallet

Kovarians

$Cov(X,Y) = E[(X-\mu_X)(Y-\mu_Y)] = E[XY]-E[X]E[Y$

Om X, Y oberoende så är $Cov(X,Y) = 0$ (eftersom då är $E[XY] = E[X]E[Y]$)

Bevis (diskreta fallet):

X, Y oberoende

$\Longrightarrow E[XY] = \Sigma_x \Sigma_y xy f_{XY}(x,y)$

$= \Sigma_x \Sigma_y xy f_X(x) f_Y(y) = \Sigma_x x f_X(x) \Sigma_y y f_Y(y)$

$= \Sigma_x x f_X(x) E[Y] = E[Y] \Sigma_x x f_X(x) = E[Y]E[X]$

Korrelation

$$\rho_{XY} = Corr(X,Y) = \frac{Cov(X,Y)}{\sqrt{Var(X)Var(Y)}}$$

$-1 \leq \rho_{XY} \leq 1$

ρ_{XY} mäter graden av linjärt beroende mellan X, Y

$|\rho_{XY}| = 1$ om och endast om det finns reella tal a, b så att $Y = a + bX$

Betingade tätheter

Betingade tätheten för X givet $Y = y$ ges av

$$f_{X|y}(x) = \frac{f_{XY}(x,y)}{f_Y(y)}, \; f_Y(y) > 0$$

Tätheten för Y givet $X = x$

$$f_{Y|x}(y) = \frac{f_{XY}(x,y)}{f_X(x)}, \; f_X(x) > 0$$

Grafen för väntevärdet av X givet $Y = y$, betecknas $\mu_{X|y}$, kallas *regressionslinjen* för X m a p Y

Kapitel 6: Beskrivande statistik

Låt $X_1, X_2, \ldots, X_n$ vara ett stickprov på en s.v. X

En **statistika** är en s.v. som kan fås från ett stickprov, t ex $\Sigma_{i=1}^n X_i$, $\Sigma_{i=1}^n X_i^2$ och $max_i X_i$

Stickprovsmedelvärde

$$\bar{X} = \frac{X_1 + X_2 + \ldots + X_n}{n} = \frac{1}{n} \sum_{i=1}^{n} X_i$$

Stickprovsvarians

$$S^2 = \frac{1}{n-1} \sum_{i=1}^{n} (X_i - \bar{X})^2 = \frac{1}{n-1} (\sum_{i=1}^{n} X_i^2 - n\bar{X}^2)$$

Stickprovsstandardavvikelse

$$S = \sqrt{S^2} = \sqrt{\frac{1}{n-1} \sum_{i=1}^{n} (X_i - \bar{X})^2}$$

Median

Exempel på grafisk beskrivning

Observerat stickprov: 25, 19, 22, 25, 24, 23, 31, 28, 24, 16, 23, 22, 26, 25, 28, 26, 25, 19, 21, 23

0 |

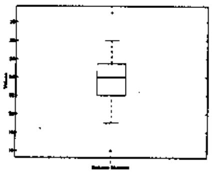
1 | 969

2 | 52543843265865 13

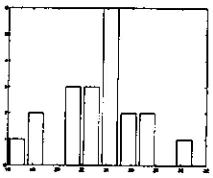
3 | 1

4 |

Figur 1: STAM-OCH-BLAD DIAGRAM



Figur 2: BOXPLOT



Figur 3: HISTOGRAM

Kapitel 10:
JÄMFÖRELSER AV 2 VÄNTEVÄRDEN (OCH 2 VARIANSER)

Låt $\bar{X}_1$ och $\bar{X}_2$ vara stickprovsmedelvärden från fördelninga med väntevärden μ_1 resp μ_2 och varianser σ_1^2 resp σ_2^2

Om n_1 och n_2 stora så är (CGS)

$$\frac{(\bar{X}_1 - \bar{X}_2) - (\mu_1 - \mu_2)}{\sqrt{\sigma_1^2/n_1 + \sigma_2^2/n_2}} \sim N(0,1)$$

då $\sigma_1^2 \neq \sigma_2^2$ och

$$\frac{(\bar{X}_1 - \bar{X}_2) - (\mu_1 - \mu_2)}{\sqrt{\sigma^2(1/n_1 + 1/n_2)}} \sim N(0,1)$$

då $\sigma_1^2 = \sigma_2^2 = \sigma^2$

Om varianserna okända utnyttjas istället att

$$\frac{(\bar{X}_1 - \bar{X}_2) - (\mu_1 - \mu_2)}{\sqrt{s_1^2/n_1 + s_2^2/n_2}} \sim t_\gamma$$

okänt # frihetsgrader

då $\sigma_1^2 \neq \sigma_2^2$ och

$$\frac{(\bar{X}_1 - \bar{X}_2) - (\mu_1 - \mu_2)}{\sqrt{s_p^2(1/n_1 + 1/n_2)}} \sim t_{n_1+n_2-2}$$

Känt # frihetsgrader

Man brukar för enkelhets skull anta att varianserna är lika, såvida man inte kan visa att de är olika.

då $\sigma_1^2 = \sigma_2^2 = \sigma^2$, där

$$s_p^2 = \frac{(n_1-1)s_1^2 + (n_2-1)s_2^2}{n_1+n_2-2}$$

γ kan väljas som

$$\gamma = \frac{(s_1^2/n_1 + s_2^2/n_2)^2}{\frac{(s_1^2/n_1)^2}{n_1-1} + \frac{(s_2^2/n_2)^2}{n_2-1}},$$

som avrundas nedåt till närmaste heltal

Om vi har parade data mellan stickproven, dvs observation i i stickprov 1 och 2 är observerade under samma förutsättningar (lika många observationer i varje stickprov), bildas istället

$$D_i = X_i - Y_i$$

och hypotesen $H_0 : \mu_D = 0$ används för att testa $\mu_X = \mu_Y$

Test av $\sigma_1^2 = \sigma_2^2$

Utnyttja att om $\sigma_1^2 = \sigma_2^2$ så är

$$S_1^2/S_2^2 \sim F(n_1 - 1, n_2 - 1),$$

dvs F-fördelad med $n_1 - 1$ och $n_2 - 1$ frihetsgrader

Kapitel 11:
LINJÄR REGRESSION

Antag att X och Y är 2 s.v. som är linjärt beroende. Antag att vi har n observationer $y_1, \ldots, y_n$ på Y och för varje Y-observation en X-observation $(x_1, \ldots, x_n)$. Vi vill nu skatta en **regressionslinje**

$$y = \beta_0 + \beta_1 x$$

Vi vill att skillnaderna

$$e_i = y_i - (\beta_0 + \beta_1 x_i),\ i = 1, \ldots, n$$

ska vara så små som möjligt ($e_1, \ldots, e_n$ kallas *residualer*)

Vi skattar därför β_0 och β_1 med de värden som minimerar

$$SSE = \sum_{i=1}^{n} (y_1 - \beta_0 - \beta_1 x_i)^2$$

$$\begin{cases} \frac{\delta SSE}{d\beta_0} = 0 \\ \frac{\delta SSE}{d\beta_1} = 0 \end{cases} \Rightarrow \begin{cases} \beta_1 = \frac{n\sum x_i y_i - \sum x_i \sum y_i}{n\sum x_i^2 - (\sum x_1)^2} \\ \beta_0 = \bar{y} - \beta_1 \bar{x} \end{cases},$$

dvs skatta regressionslinjen med

$$\hat{y} = \hat{\beta}_0 + \hat{\beta}_1 x,$$

där

$$\hat{\beta}_1 = \frac{n\sum x_i y_i - \sum x_i \sum y_i}{n\sum x_i^2 - (\sum x_i)^2}$$

och

$$\hat{\beta}_0 = \bar{y} - \hat{\beta}_1 \bar{x}$$

Andra böcker har:

$y = \alpha + \beta(x - \bar{x})$

$\beta = \hat{\beta}_1$

$\hat{\alpha} = \bar{y}$

ger samma linje

Uppgift 11.2:

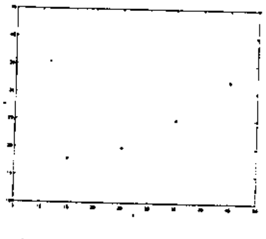
(a)

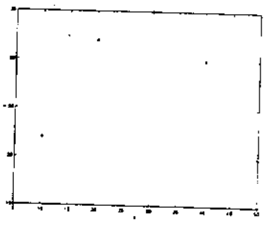
x	5	15	25	35	45	50
y	10	18	20	25	32	45

(b)

x	5	10	20	30	40	50
y	15	22	32	35	30	15

Är linjär regression lämplig?





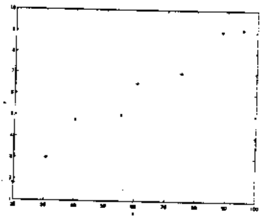
(a) Ja, ganska

(b) Nej, (pröva kvadratisk regression)

Uppgift 11.7:

Man har studerat förhållandet mellan inkomst X (\$1000/år) och energikonsumption Y (10^8 Btu/år)

y	1.8	3.0	4.8	5.0	6.5	7.0	9.0	9.1
x	20.0	30.5	40.0	55.1	60.3	74.9	88.4	95.2



Hyfsat linjärt, skatta regressionslinjen

$$\hat{\beta}_1 = \frac{8 \times 3173.17 - 464.4 \times 46.2}{8 \times 32089.6 - 464.4^2} = 0.0957$$

$$\hat{\beta}_0 = \frac{46.2}{8} - \hat{\beta}_1 \frac{464.4}{8} = 0.2173$$

Vi får

$$\hat{y} = 0.2173 + 0.0957x (= \mu_{Y|x})$$

Om x=50 förväntas energikonsumption $= 0.2173 + 0.0957 \times 50 \approx 5$

Om inkomsten ökar med 2 enheter, förväntas energiökning $= (0.2173 + 0.0957(x + 2)) - (0.2173 + 0.0957x) = 2 \times 0.0957 \approx 0.19$

Om $Y_i \sim N(\beta_0 + \beta_1 x_i, \sigma^2)$

$$\hat{\beta}_1 \sim N(\beta_1, \sigma^2 / \Sigma(x_i - \bar{x})^2)$$

$$\hat{\beta}_0 \sim N(\beta_0, \sigma^2 \Sigma x_i^2 / (n\,\Sigma(x_i - \bar{x})^2))$$

σ^2 skattas med

$$S^2 = \sum_{i=1}^{n} (y_1 - \beta_0 - \beta_1 x_i)^2 / (n-2)$$

Konfidensintervall och hypotesprövning kan utföras mha t_{n-2} fördelning

Korrelation

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

skattas med

$$r = \frac{n\sum x_i y_i - \sum x_i \sum y_i}{\sqrt{[n\sum x_i^2 - (\sum x_i)^2][n\sum y_i^2 - (\sum y_i)^2]}}$$

Om $\rho = 0$ så är

$$\frac{R\sqrt{n-2}}{\sqrt{1-R^2}} \sim t_{n-2}$$

Kapitel 13:
ENKEL VARIANSANALYS

Antag att vi har k olika behandlingar och vill testa om det är någon skillnaden mellan dessa, dvs testa

$$H_0 : \mu_1 = \mu_2 = \ldots = \mu_n$$

mot

$$H_1 : \mu_i \neq \mu_j, \text{ för ngt } i, j$$

Låt

$$Y_{11}, Y_{12}, \ldots, Y_{1n_1}$$

vara stickprov av storlek n_1, med stickprovsmedelvärde $\bar{Y}_{1.}$, från behandling 1,

$$Y_{21}, Y_{22}, \ldots, Y_{2n_2}$$

stickprov av storlek n_2, med stickprovsmedelvärde $\bar{Y}_{2.}$, från behandling 2, osv

Variationen inom behandling i är

$$\sum_{j=1}^{n_i} (Y_{ij} - \bar{Y}_{i.})^2$$

och vi får en total "inom-behandlings"-variation

$$SS_E = \sum_{i=1}^{k} \sum_{j=1}^{n_i} (Y_{ij} - \bar{Y}_{i.})^2$$

Variationen mellan behandlingarna fås av

$$SS_{Tr} = \sum_{i=1}^{k} n_i(\bar{Y}_{i.} - \bar{Y}_{..})^2$$

Totala variationen är

$$SS_{Tot} = SS_{Tr} + SS_E$$

Om H_0 sann borde större delen av variationen vara "inom-behandling"-variation medan variationen mellan behandlingar bör vara liten

Låt $N = \Sigma_{i=1}^{k} n_i$, $MS_{Tr} = \frac{SS_{Tr}}{k-1}$ och $MS_E = \frac{SS_E}{N-k}$

Om behandlingarna är normalfördelade med samma varians gäller under H_0 att

$$F = \frac{MS_{Tr}}{MS_E} \sim F(k-1, N-k)$$

och vi förkastar H_0 för stora värden på F

ANOVA-tabell:

Variationskälla	frihetsg	SS	MS	F
Treatment	$k-1$	SS_{Tr}	MS_{Tr}	$\frac{MS_{Tr}}{MS_E}$
Error	$N-k$	SS_E	MS_E	
Total	$N-1$	SS_{Tot}		

Kapitel 15:
KVALITATIVA DATA

Antag att vi utför n stycken experiment, där vart och ett kan resultera i k möjliga utfall med sannolikheterna $p_1, p_2, \ldots, p_k$ (där $\Sigma_{i=1}^{k} p_i = 1$)

Om $X_1, X_2, \ldots, X_k$ betecknar antalet som resulterar i utfall $1, 2, \ldots, k$ så får vi en **multinomial fördelning** och

$$E[X_i] = np_i$$

Om $(X_1, X_2, \ldots, X_k)$ är en multinomialfördelad s.v. och n stort så

$$\sum_{i=1}^{k} \frac{(X_i - np_i)^2}{np_i} \overset{\text{approx}}{\sim} \chi^2_{k-1},$$

eller enklare

$$\sum_{i=1}^{k} \frac{(O_i - E_i)^2}{E_i} \overset{\text{approx}}{\sim} \chi^2_{k-1},$$

där O står för observerat värde och E för förväntat

Exempel

Vid 60 kast med tärning ficks 15 1:or, 10 2:or, 8 3:or, 10 4:or, 5 5:or samt 12 6:or. Testa H_0 : Tärningen symmetrisk. Om H_0 sann:

$$Q = \sum_{i=1}^{6} \frac{(O_i - 10)^2}{10} \overset{\text{approx}}{\sim} \chi^2_5$$

$Q = 5.8$, så vi kan ej förkasta på rimlig nivå

TEST AV OBEROENDE: Uppgift 15.10

	Luftkvalitet			
Temperatur	Dålig	Medel	Bra	
Under medel	$n_{11} = 1$	$n_{12} = 3$	$n_{13} = 24$	$n_{1.} = 28$
Medel	$n_{21} = 12$	$n_{22} = 28$	$n_{23} = 76$	$n_{2.} = 116$
Över medel	$n_{31} = 12$	$n_{32} = 14$	$n_{33} = 30$	$n_{3.} = 56$
	$n_{.1} = 25$	$n_{.2} = 45$	$n_{.3} = 130$	$n = 200$

200 slumpmässigt utvalda dagar

Testa H_0 : Temperatur, luftkvalitet oberoende
mot H_1 : Temperatur, luftkvalitet beroende

Låt p_{ij}=P(hamna i rad i och kolumn j)
Om H_0 sann: $p_{ij} = p_{i.}p_{.j}$, vilket betyder att
$E_{ij} = E[\text{antal obs i } (i,j)] = 200p_{i.}p_{.j}$
E_{ij} skattas med

$$\hat{E}_{ij} = 200\frac{n_{i.}}{200}\frac{n_{.j}}{200} = \frac{n_{i.}n_{.j}}{200},$$

vilket ger $\hat{E}_{11} = 3.5$, $\hat{E}_{12} = 6.3$, $\hat{E}_{13} = 18.2$, $\hat{E}_{21} = 14.5$, $\hat{E}_{22} = 26.1$, $\hat{E}_{23} = 75.4$, $\hat{E}_{31} = 7$, $\hat{E}_{32} = 12.6$, $\hat{E}_{33} = 36.4$

Om H_0 sann är ($r = 3$, $c = 3$)

$$Q = \sum_{i=1}^{r}\sum_{j=1}^{c} \frac{(n_{ij} - \hat{E}_{ij})^2}{\hat{E}_{ij}} \overset{\text{approx}}{\sim} \chi^2_{(r-1)(c-1)} = \chi^2_4$$

$Q = 10.79$, så förkasta på 0.05-, men ej 0.025-nivå