TENTAMEN I SF1902 SANNOLIKHETSTEORI OCH STATISTIK
ONSDAGEN DEN 1:A JUNI 2016 KL 8.00–13.00.

Kursledare och examinator : Björn-Olof Skytt, tel 790 8649.

Tillåtna hjälpmedel: miniräknare, lathund till statistikfunktioner på Texas Instruments-räknare (TI-82 Stats och högre) utan egna tillägg, läroboken av Blom m.fl. utan egna tillägg, formelsamlingen BETA utan egna tillägg.

Resonemang och uträkningar skall vara så utförliga och väl motiverade att de är lätta att följa. Numeriska svar skall anges med minst två siffrors noggrannhet.

Uppgift 1

I ett stort bostadsbestånd har man infört elektroniska lås i portarna. I genomsnitt 3 av 10 nätter är det någon som inte kommer in i sin port p.g.a. elfel. Varje gång någon inte kommer in i porten p.g.a. elfel får nattjouren rycka in.

a) Beräkna sannolikheten att under en natt ingen behöver ringa till nattjouren p.g.a. att de inte kommer in i porten.

b) Beräkna sannolikheten att det under en månad som innehåller 30 nätter inkommer minst 10 samtal till nattjouren.

Uppgift 2

I ett stort bostadsområde är $\frac{1}{6}$ av lägenheterna ettor, $\frac{1}{3}$ tvåor, $\frac{2}{5}$ treor och $\frac{1}{10}$ fyror. Under en 10-årsperiod anses risken för vattenskada för en etta vara 10% , för en tvåa 5%, för en trea 8 % och för en fyra 15%. Anta att en vattenskada inträffat. Vad är då sannolikheten att det är en tvåa som har drabbats?

Uppgift 3

Den stokastiska variabeln X har täthetsfunktionen

$$f_X(x) = \begin{cases} cx^5 & \text{om } 2 \leq x \leq 8 \\ 0 & \text{annars} \end{cases}$$

Bestäm c och beräkna $P(6 < X < 12)$.

Uppgift 4

I ett stort bostadsområde är $\frac{1}{6}$ av lägenheterna ettor, $\frac{1}{3}$ tvåor, $\frac{2}{5}$ treor och $\frac{1}{10}$ fyror. Bestäm sannolikheten (approximativt) att 60 slumpvis utvalda lägenheter tillsammans har minst 135 rum.

Uppgift 5

Tabellen visar hur många personer som i Sifos väljarbarometer från april 2016 respektive april 2015 sade att de skulle ha röstat på alliansen, de rödgröna partierna respektive övriga partier om det vore riksdagsval den dag de tillfrågades om sina partisympatier.

	Alliansen	Rödgröna	Övriga	Totalt
April 2016	724	686	311	1721
April 2015	672	708	269	1649

Bestäm ett approximativt 95%-igt konfidensintervall för förändringen i stödet för de rödgröna partierna mellan de båda undersökningstillfällena. Kan man på signinfikansnivån 5 % se om det varit någon statistiskt säkerställd förändring i stödet för de rödgröna partierna? Ange tydligt vilka de uppställlda hypoteserna är. Din slutsats skall tydligt motiveras.

Uppgift 6

Tabellen visar hur många personer som i Sifos väljarbarometer från april 2016 respektive april 2015 sade att de skulle ha röstat på alliansen, de rödgröna partierna respektive övriga partier om det vore riksdagsval den dag de tillfrågades om sina partisympatier.

	Alliansen	Rödgröna	Övriga	Totalt
April 2016	724	686	311	1721
April 2015	672	708	269	1649

Undersök om det överhuvudtaget skett någon förändring i väljaropinionen mellan de två undersökningstillfällena, där alla tre grupperna (alliansen,de rödgröna partierna samt övriga partier) analyseras samtidigt. Använd signinfikansnivån 5 %. Ange tydligt de uppställda hypoteserna och motivera tydligt vilken slutsats du drar.

Uppgift 7

Någon hävdar att ett visst preparat S ökar blodtrycket. Tabellen anger värdena på blodtrycket för 16 försökspersoner omedelbart före och efter en viss tid efter behandling med S.

Pers nr	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16
Före	113	111	124	120	115	122	102	122	106	109	109	120	105	129	103	115
Efter	112	113	125	125	120	118	101	123	112	112	113	121	101	127	107	121

Vi kan inte anta att någonting här är utfall av Normalfördelade stok.var. Avgör på risknivån 10% om preparatet S har någon effekt på blodtrycket. Ange tydligt vilka de uppställlda hypoteserna är. Din slutsats skall tydligt motiveras.

Avd. Matematisk statistik

KTH Matematik

LÖSNINGSFÖRSLAG TENTAMEN I SF1902 MATEMATISK STATISTIK.
ONSDAGEN DEN 1:A JUNI 2016 KL 8.00–13.00

Uppgift 1

a) Om vi här inför den stokastiska varibeln X_i där X_i är antal inkommande samtal till nattjouren natt i så gäller att $X_i \in Po(\mu)$ där $\mu = 0.3$.
Alltså blir $P(X_i = 0)$=[se tab 5 i F.S.]=$P(X_i \leq 0) = 0.74082$.
Svar: $\underline{P(X_i = 0) = 0.74082.}$

b) Om vi inför den stokastiska varibeln Y där Y är antal inkommande samtal till nattjouren under månaden så gäller att

$$Y = \sum_{i=1}^{30} X_i \in Po(30 \cdot 0.3) = Po(9)$$

Sökt är $P(Y \geq 10)$=[se tab 5 i F.S.]=$1 - P(Y \leq 9) = 1 - 0.58741 = 0.41259$.
Svar: $\underline{P(Y \geq 10) = 0.41259.}$

Uppgift 2

Här har vi Bayes sats.
Låt A, B, C, D beteckna händelserna att en lägenhet har ett, två, tre resp. fyra rum.
Låt V beteckna händelsen att en vattenskada inträffat. Då blir

$$P(B|V) = \frac{P(B \cap V)}{P(V)} = \frac{P(V|B)P(B)}{P(V|A)P(A) + P(V|B)P(B) + P(V|C)P(C) + P(V|D)P(D)} =$$

$$= \frac{0.05 \cdot \frac{1}{3}}{0.10 \cdot \frac{1}{6} + 0.05 \cdot \frac{1}{3} + 0.08 \cdot \frac{2}{5} + 0.15 \cdot \frac{1}{10}} = \frac{50}{241} = 0.207$$

Svar: <u>Sannolikheten är 20.7% att det är en tvåa som har drabbats.</u>

Uppgift 3

Eftersom det alltid gäller att

$$\int_{-\infty}^{\infty} f_X(x)dx = 1$$

fås c ur ekvationen

$$\int_{2}^{8} c \cdot x^5 dx = 1$$

Vilket ger

$$\left[\frac{c \cdot x^6}{6}\right]_2^8 = 1$$

Nu fås

$$c \cdot \frac{8^6 - 2^6}{6} = 1$$

Alltså blir

$$c = \frac{6}{8^6 - 2^6} = \frac{1}{43680} = 2.289 \cdot 10^{-5}$$

Detta medför att

$$P(6 < X < 12) = \int_{6}^{8} c \cdot x^5 dx = \left[\frac{c \cdot x^6}{6}\right]_6^8 = \frac{1}{43680}\frac{8^6 - 6^6}{6} = \frac{215488}{262080} = 0.822$$

Svar: $\underline{c = 2.289 \cdot 10^{-5} \text{ och } P(6 < X < 12) = 0.822}$

Uppgift 4

Låt X_i vara den stokastiska variabeln *antal rum* för en på måfå observerad lägenhet. Vi får då

$$\mu = E[X_i] = \sum_{k=1}^{4} k \cdot p_{X_i}(k) = 1\frac{1}{6} + 2\frac{1}{3} + 3\frac{2}{5} + 4\frac{1}{10} = \frac{146}{60} \approx 2.433$$

och

$$\sigma = D(X_i) = \sqrt{V(X_i)}$$

där

$$V(X_i) = E(X_i^2) - E^2(X_i) = E(X_i^2) - (\frac{146}{60})^2$$

och

$$E(X_i^2) = \sum_{k=1}^{4} k^2 \cdot p_{X_i}(k) = 1^2\frac{1}{6} + 2^2\frac{1}{3} + 3^2\frac{2}{5} + 4^2\frac{1}{10} = \frac{402}{60}$$

Detta medför att

$$V(X_i) = \frac{402}{60} - (\frac{146}{60})^2 = \frac{2804}{3600} \approx 0.779$$

Vilket ger att

$$\sigma = D(X_i) = \sqrt{\frac{2804}{3600}} \approx 0.883$$

Låt nu

$$Y = \sum_{i=1}^{60} X_i$$

Vi antar nu att X_i:na är många och oberoende. Eftersom de dessutom är likafördelade gäller då enligt Centrala Gränsvärdessatsen C.G.S. att

$$Y \sim N(60\mu, \sigma\sqrt{60}) = N(60\frac{146}{60}, \sqrt{\frac{2804}{3600}}\sqrt{60}) = N(146, \sqrt{\frac{2804}{60}})$$

$$P(Y \geq 135) = P(\frac{Y-146}{\sqrt{\frac{2804}{60}}} \geq \frac{135-146}{\sqrt{\frac{2804}{60}}}) = [Z = \frac{Y-146}{\sqrt{\frac{2804}{60}}} \sim N(0,1)] = P(Z \geq \frac{135-146}{\sqrt{\frac{2804}{60}}}) =$$

$$= 1 - P(Z \leq \frac{135-146}{\sqrt{\frac{2804}{60}}}) = 1 - \Phi(\frac{-11}{\sqrt{\frac{2804}{60}}}) = 1 - [1 - \Phi(\frac{11}{\sqrt{\frac{2804}{60}}})] \approx \Phi(1.61) \approx 0.946$$

Svar: Sannolikheten att antalet rum är minst 135 är ungefär 94.6%

Uppgift 5

H_0 är här att ingen statistiskt säkerställd förändring ägt rum i stödet för de rödgröna. H_1 är att vi har en statistiskt säkerställd förändring Vi antar här att X_1 är antalet som i april 2015 sade att de skulle röstat på de rödgröna och att X_2 är antalet som i april 2016 sade att de skulle röstat på de rödgröna . X_1 är då $Bin(n_1, p_1)$ och X_2 är $Bin(n_2, p_2)$. Vi vill nu bilda ett approximativt konfidensintervall för $p_2 - p_1$, men då måste X_1 och X_2 vara approximativt Normalfördelade. Villkoret för detta är att $np(1-p) > 10$ för X_1 och X_2. Vi skattar p_1 med $x_1/n_1 = 708/1649$ och p_2 med $x_2/n_2 = 686/1721$. Vi får då att $n_1 p^*_{1,obs}(1 - p^*_{1,obs}) = 404.02 > 10$ och $n_2 p^*_{2,obs}(1 - p^*_{2,obs}) = 412.56 > 10$. Villkoret är uppfyllt och vi kan bilda konfidensintervallet

$$I_{p_2-p_1} = \left(p^*_{2,obs} - p^*_{1,obs} \pm \sqrt{p^*_{1,obs}(1-p^*_{1,obs})/n_1 + (p^*_{2,obs}(1-p^*_{2,obs})/n_2} \cdot \lambda_{\alpha/2}\right).$$

Vi har en konfidensgrad på $1 - \alpha = 0.95$ och då fås $\lambda_{\alpha/2} = 1.9600$ och intervallet blir

$$-3.07 \cdot 10^{-2} \pm 17 \cdot 10^{-3} \cdot 1.96 = -3.07 \cdot 10^{-2} \pm 3.33 \cdot 10^{-2} = (-0.064, 0.0026).$$

0 tillhör intervallet och vi kan således inte förkasta H_0 på risknivån 5%.
Vi drar slutsatsen att förändringen ej är statistiskt säkerställd på risknivån 5% .

Uppgift 6

Vi gör här ett homogenitetstest (avsnitt 14.3 i formelsamlingen).Nollhypotesen $H0$ är då att fördelningen av stödet för de tre väljargrupperna är oförändrad mellan de båda undersökningstillfällena. Mothypotesen $H1$ är då att det skett en sådan förändring.

Vi gör här en tabell med observerade antal enligt

Observerade antal	Alliansen	Rödgröna	Övriga	Totalt
April 2016	724	686	311	1721
April 2015	672	708	269	1649
Totalt	1396	1394	580	3370

Teststorheten blir

$$Q = \frac{(724 - \frac{1721\cdot 1396}{3370})^2}{\frac{1721\cdot 1396}{3370}} + \frac{(686 - \frac{1721\cdot 1394}{3370})^2}{\frac{1721\cdot 1394}{3370}} + \frac{(311 - \frac{1721\cdot 580}{3370})^2}{\frac{1721\cdot 580}{3370}} +$$

$$+ \frac{(672 - \frac{1649\cdot 1396}{3370})^2}{\frac{1649\cdot 1396}{3370}} + \frac{(708 - \frac{1649\cdot 1394}{3370})^2}{\frac{1649\cdot 1394}{3370}} + \frac{(269 - \frac{1649\cdot 580}{3370})^2}{\frac{1649\cdot 580}{3370}} = 3.79$$

Om H_0 är sann så är 3.79 ett utfall från en stokastisk variabel som approximativt har en χ^2-fördelning med $(3-1)(2-1) = 2$ frihetsgrader. Eftersom $\chi^2_{0.05}(2) = 5.99 > 3.79$ så kan H_0 inte förkastas på nivån 5%. Alternativt kan vi beräkna sannolikheten att en $\chi^2(2)$-variabel är större än eller lika med 3.79 (`X2cdf` på en TI-räknare). Denna sannolikhet, dvs p-värdet för testet, är 0.15. Detta p-värde är inte så lågt att vi förkastar H_0 på risknivån 5%. Både teststorheten och p-värdet fås direkt med funktionen `X2-Test` på en TI-räknare.
Svar: Vi drar slutsatsen att någon förändring i stödet mellan de tre väljargrupperna ej är statistiskt säkerställd på risknivån 5% .

Uppgift 7

H_0 är här att preparatet S inte har någon effekt på blodtrycket. Här använder vi oss av teckentestet. Vi inför den stokastiska variabeln Z som är antalet personer som får ett högre blodtryck efter behandling med S. Om H_0 gäller ska $Z \in Bin(n, 0.5) = Bin(16, 0.5)$. Vi får alltså att

$$p\text{-värdet} = 2 \cdot P(Z \geq 11) = 2 \cdot P(Z \leq 5) = [\text{se tab 6}; x = 5, n = 16, p = 0.5] = 2 \cdot 0.1056 = 0.21012$$

d.v.s. 21.0%. Detta innebär att vi inte kan förkasta H_0 på risknivån 10%.
Vi drar slutsatsen att förändringen ej är statistiskt säkerställd på risknivån 10% .