KTH Matematik

Avd. Matematisk statistik

TENTAMEN I SF1902 SANNOLIKHETSTEORI OCH STATISTIK,
MÅNDAGEN DEN 17:E AUGUSTI 2015 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, institutionens formelsamling utan egna tillägg samt 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.

Observera att en eventuellt godkänd kontrollskrivning ej tillgodoräknas på denna tentamen.

Uppgift 1

En bostadsrättsförening som har 24 stammar skall avgöra om man ska genomföra ett stambyte eller ej. För att avgöra detta bestämmer man sig för att slumpvis välja ut 8 stammar och genomföra stambytet om minst 3 av dessa stammar klassificeras som dåliga. Anta att det finns 9 stammar som skulle klassificeras som dåliga. Vad är då sannolikheten att stambytet kommer att ske?

Uppgift 2

I ett bostadsområde är 10% av lägenheterna ettor, 30% tvåor, 45% treor och 15% fyror. Antag att sannolikheten för att en etta skall drabbas av elfel under året är 7%, och att motsvarande sannolikheter för tvåor är 9%, för treor 12% och för fyror 13%.
Vad är då sannolikheten att en i området slumpvis vald lägenhet skall drabbas av elfel under året?

Uppgift 3

Vid en simuleringsstudie drar man 1200 oberoende slumptal,X_i.
Varje X_i är likformigt fördelat mellan 0 och 1. Dessa tal adderas.

$$Y = \sum_{i=1}^{1200} X_i$$

.
Beräkna $P(Y > 585)$.

Uppgift 4

Två förskolor, Bullerbyn och Skogsgläntan, finns i samma stad men är belägna nära en trafikled respektive i kanten av ett skogsområde. Kommunens miljöenhet tror att Bullerbyns läge kan göra att barnen där får förhöjda halter av bly i blodet. Miljöenheten valde slumpmässigt ut fem barn från vardera förskola och mätte deras halter av bly i blodet. Resultat:

Blykoncentration (ng/ml)					
Bullerbyn	0.93	0.63	1.21	1.30	0.58
Skogsgläntan	0.96	0.43	0.93	0.85	0.48

Undersök på lämpligast sätt om om miljöenhetens misstanke kan styrkas,när vi kan anta att variationen i blyhalt mellan barn på respektive förskola är normalfördelad med en varians som är densamma för de båda förskolorna.Välj signifikansnivån 5%.Ange tydligt vilka de uppställda hypoteserna är och vad slutsatsen är.

Uppgift 5

En elfirma har ansvaret att åtgärda elfelen i fyra områden:A, B ,C , och D. Vi antar att μ är väntevärdet för sammanlagda antalet elfel som inträffar i alla fyra områdena under en månad.
Under en slumpvis vald månad inträffar 7 elfel i område A, 2 elfel i område B, 3 elfel i område C, och 9 elfel i område D.
Bestäm utifrån dessa data ett konfidensintervall för μ med en konfidensgrad på approximativt 95 %. Ange vilka statistiska modellantaganden du gör samt motivera varför du får göra detta approximativa konfidensintervall.

Uppgift 6

För att undersöka kvaliteten på avgasreningen på nya bilar av tre olika märken, valde en motortidning ut 100 bilar av vardera märke och testade dem. Resultatet var att för bilmärkena A, B och C hade 19, 15 respektive 26 bilar ett mindre eller större fel på avgasreningen.

Undersök med ett lämpligt statistiskt test, på nivån 5%, om det kan vara så att bilmärkena har samma felfrekvens vad gäller avgasreningen. Ange tydligt vilka de uppställda hypoteserna och slutsatsen är.

Uppgift 7

För att undersöka effektiviteten i ett nytt motionsprogram väljer man ut en grupp om 10 personer vilka slumpmässigt delas i två grupper om 5 personer vardera. Personerna i den ena gruppen får träna efter det nya programmet medan de andra tränar efter ett äldre program med kända egenskaper. Som mått på konditionsförändringen väljer man kvoten mellan syreupptagningsförmågan efter motionsprogrammet och innan motionsprogrammet och får följande resultat:

Kvot mellan syreupptagningsförmågan efter och innan motionsprogrammet.

Nytt program	1.03	1.19	1.22	1.26	1.29
Gammalt program	0.98	1.01	1.02	1.10	1.17

Dessa kvoter kan inte antagas komma från någon Normalfördelning.Avgör på signifikansnivån 5% om motionsprogrammen kan anses skilja sig åt vad gäller förbättringen av syreupptagningsförmågan. Ange tydligt vilka de uppställda hypoteserna är och vad slutsatsen är.

KTH Matematik

Avd. Matematisk statistik

LÖSNINGSFÖRSLAG TENTAMEN I SF1902 MATEMATISK STATISTIK.
MÅNDAGEN DEN 17:E AUGUSTI 2015 KL 8.00–13.00

Uppgift 1

Vi inför den stokastiska variabeln X, där $X \in Hyp(N, n, p) = Hyp(24, 8, \frac{9}{24})$.
Vi söker $P(X \geq 3) = 1 - [p_X(0) + p_X(1) + p_X(2)] =$

$$= 1 - [\frac{\binom{9}{0}\binom{15}{8}}{\binom{24}{8}} + \frac{\binom{9}{1}\binom{15}{7}}{\binom{24}{8}} + \frac{\binom{9}{2}\binom{15}{6}}{\binom{24}{8}}] = \underline{0.6675}$$

Uppgift 2

Här har vi lagen om total sannolikhet.Låt A,B,C,D beteckna händelserna att den slumpvis valda lägenheten är en etta,tvåa,trea,respektive fyra. Låt E beteckna händelsen att lägenheten drabbas av elfel under året.
$P(E) = P(E \mid A)P(A) + P(E \mid B)P(B) + P(E \mid C)P(C) + P(E \mid D)P(D) =$
$= 0.07 \cdot 0.10 + 0.09 \cdot 0.30 + 0.12 \cdot 0.45 + 0.13 \cdot 0.15 = \underline{0.1075}$

Uppgift 3

Slumptalen är många,oberoende och har alla samma fördelning. Då kan vi använda Centrala gränsvärdessatsen enligt §5 i formelsamlingen och ser då att $Y \sim N(n\mu, \sigma\sqrt{n})$.
Där n=1200 och där vi får fram μ och σ m.h.a. §4 eftersom X_i:na är likformigt fördelade.
$X_i \in U(0, 1)$ Då blir

$$\mu = E(X_i) = \frac{0+1}{2} = 0.5$$

och

$$\sigma = D(X_i) = \frac{1-0}{\sqrt{12}} = \frac{1}{\sqrt{12}}$$

Nu har vi att

$$Y \sim N(1200 \cdot 0.5, \frac{1}{\sqrt{12}} \cdot \sqrt{1200}) = N(600, 10)$$

$$P(Y > 585) = P(\frac{Y-600}{10} > \frac{585-600}{10}) = [p.g.a. C.G.S.] = 1 - \Phi(\frac{585-600}{10}) =$$

$=1 - \Phi(-1.50) = 1 - [1 - \Phi(1.50)] = \Phi(1.50) = \underline{0.9332}$

Uppgift 4

Vi har standardsituationen två oberoende stickprov från normalfördelningar med gemensam varians, dvs avsnitt 11.2 i formelsamlingen.

Låt $x_1, \ldots, x_5$ och $y_1, \ldots, y_5$ vara de uppmätta blyhalterna hos barnen från Bullerbyn respektive Skogsgläntan. Dessa är observationer av stokastiska variabler $X_1, \ldots, X_5$ respektive $Y_1, \ldots, Y_5$ (alla oberoende), med fördelningar $\mathrm{N}(\mu_X, \sigma)$ respektive $\mathrm{N}(\mu_Y, \sigma)$.
Detta är hypotesprövning med två oberoende stickprov. Vi ställer upp hypoteserna H_0 och H_1

$$H_0 : \mu_X = \mu_Y$$

$$H_1 : \mu_X \neq \mu_Y$$

Vi bildar konfidensintervallet

$$I_{\mu_X - \mu_Y} = \left(\bar{x} - \bar{y} \pm t_{\alpha/2}(n_X + n_Y - 2) \cdot s \cdot \sqrt{\frac{1}{n_X} + \frac{1}{n_Y}} \right)$$

där

$$s^2 = \frac{(n_X - 1)s_X^2 + (n_Y - 1)s_Y^2}{n_X + n_Y - 2}.$$

Med $n_X = n_Y = 5$, $\bar{x} = 0.93$, $\bar{y} = 0.73$, $s_X = 0.3270$, $s_Y = 0.2549$, $t_{0.025}(8) = 2.31$ får vi $I_{\mu_Y - \mu_X} = (0.2 \pm 0.428)$. Eftersom 0 ingår i konfidensintervallet kan man inte förkasta H_0 på signifikansnivån 5%. Detta innebär att man med 5% felrisk inte kan påstå att det föreligger systematisk skillnad mellan de två förskolorna vad gäller blyhalterna i blodet hos barnen .
Miljöenhetens misstanke kan inte styrkas på risknivån 5 %.

Uppgift 5

Vi antar att antalet elfel som inträffar i område A,B,C respektive D under månaden är X_A, X_B, X_C respektive X_D, och att $X_A \in Po(\mu_A)$,$X_B \in Po(\mu_B)$,$X_C \in Po(\mu_C)$ och $X_D \in Po(\mu_D)$. Dessutom antar vi att X_A, X_B, X_C och X_D är oberoende.
Då gäller att totala antalet elfel som vi kallar $Y = X_A + X_B + X_C + X_D \in Po(\mu_A + \mu_B + \mu_C + \mu_D) = Po(\mu)$. Villkoret för att få göra ett konfidensintervall för μ är att Poissonfördelningen kan approximeras till en Normalfördelning.Villkoret för detta är att $\mu \geq 15$ enligt §6 i Formelsamlingen. Vi har inte μ.
Men eftersom $E(Y) = \mu$ så skattar vi μ med $\mu^*_{obs} = x_A + x_B + x_C + x_D = 3 + 2 + 7 + 9 = 21 > 15$, vilket gör att villkoret är uppfyllt. Sedan bildar vi konfidensintervallet enligt(12.9) i Bloms bok.

$$\underline{I_\mu = \left(y \pm \lambda_{\alpha/2}\sqrt{y}\right) = \left(21 \pm 1.96\sqrt{21}\right) = (12.02, 29.88)}$$

Uppgift 6

Vi gör ett homogenitetstest (avsnitt 14.3 i formelsamlingen), och noterar att vi då för varje försöksserie (här: bilmärke) måste ha med utfall som täcker alla möjligheter. Inför alltså utfallen $j = 1$ och $j = 2$ svarande mot ”felfri avgasrening” och ”defekt avgasrening”.

Nollhypotesen H_0 är då att sannolikheten för utfall j, för en slumpvis vald bil, är densamma för alla tre bilmärkena A, B och C.

Mothypotesen H_1 är att H_0 inte gäller, dvs att åtminstone två av märkena har olika sannolikheter för utfallen.

Vi gör en tabell med observerade antal enligt

Observerade antal

	Felfria	Defekta	
Märke A	81	19	100
Märke B	85	15	100
Märke C	74	26	100
	240	60	300

Teststorheten blir

$$Q = \frac{(81 - \frac{100\cdot 240}{300})^2}{\frac{100\cdot 240}{300}} + \frac{(85 - \frac{100\cdot 240}{300})^2}{\frac{100\cdot 240}{300}} + \frac{(74 - \frac{100\cdot 240}{300})^2}{\frac{100\cdot 240}{300}} + \frac{(19 - \frac{100\cdot 60}{300})^2}{\frac{100\cdot 60}{300}} + \frac{(15 - \frac{100\cdot 60}{300})^2}{\frac{100\cdot 60}{300}} + \frac{(26 - \frac{100\cdot 60}{300})^2}{\frac{100\cdot 60}{300}} = \frac{(81-80)^2}{80} + \frac{(85-80)^2}{80} + \frac{(74-80)^2}{80} + \frac{(19-20)^2}{20} + \frac{(15-20)^2}{20} + \frac{(26-20)^2}{20} = \frac{31}{8} = 3.875.$$

Om H_0 är sann så är 3.875 ett utfall från en stokastisk variabel som approximativt har en χ^2-fördelning med $(3-1)(2-1) = 2$ frihetsgrader. Eftersom $\chi^2_{0.05}(2) = 5.99 > 3.875$ 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 3.875 (`X2cdf` på en TI-räknare). Denna sannolikhet, dvs p-värdet för testet, är 0.14. Detta p-värde är inte så lågt att vi förkastar H_0. Både teststorheten och p-värdet fås direkt med funktionen `X2-Test` på en TI-räknare.

<u>Data ger alltså inte belägg för att bilmärkena skiljer sig åt i det undersökta avseendet.
Detta innebär inte att man visat att bilmärkena är likvärdiga.</u>

Uppgift 7

Här använder vi Wilcoxons rangsummetest. Låt N vara det nya programmet och G det gamla programmet. Vi har hypoteserna
H_0 : Förbättringen av syreupptagningsförmågan är densamma i de två grupperna.
H_1 : Förbättringen av syreupptagningsförmågan är olika i de två grupperna.

Vi får följande ranger

Nytt program	4	7	8	9	10
Gammalt program	1	2	3	5	6

Vi har då att rangsummorna blir $W_N = 38$, $W_G = 17$, $n_N = 5$ och $n_G = 5$. Detta ger oss att $U_N = 23$ och $U_G = 2$. Vi ska testa på nivån 5% och det kritiska värdet blir då $U_{0.05} = 2$. Eftersom U_G är mindre än U_N jämförs U_G med det kritiska värdet. Eftersom $U_G = 2 \leq U_{0.05} = 2$ förkastas H_0 på nivån 5%.
Vi drar slutsatsen att det nya programmet förbättrar syreupptagningsförmågan mer än det gamla.