Avd. Matematisk statistik

KTH Matematik

TENTAMEN I SF1902 SANNOLIKHETSTEORI OCH STATISTIK
MÅNDAGEN DEN 15:E AUGUSTI 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

Låt A och B beteckna två händelser. Givet är $P(A) = 3/5$, $P(B|A) = 2/3$ och $P(B|A^*) = 1/3$, där A^* betecknar komplementet till händelsen A.

Avgör om A och B är beroende händelser.

Uppgift 2

Ett spel går till på följande sätt: Man satsar 1 kr i en spelautomat. Om man vinner får man 3 gånger insatsen tillbaka, men om man förlorar så förlorar man sin satsade krona. Sannolikheten att vinna vid en spelomgång är 25%.

a) Beräkna väntevärdet och standardavvikelsen för nettovinst om man spelar en omgång. Med nettovinst menas ”utbetalt belopp minus insats”.

b) Beräkna med välmotiverad approximation sannolikheten att man går med nettovinst om man spelar 120 omgångar.

Uppgift 3

GK-Motors utreder med egna resurser 10 av de bilolyckor som drabbat modellen GK-Roadrunner. Dessa tio olycksbilar är slumpmassigt valda ur ett mycket större antal (≈ 1000) olycksbilar. Om man upptäcker att minst åtta av dessa 10 bilar har två eller fler skador hos kretskorten i den elektroniska servostyrningen, så kommer en hel årspoduktion att återkallas för åtgärd hos GKs återförsäljare. Man vet att sannolikheten $= 0.6$ för att antalet skador hos kretskorten i den elektroniska servostyrningen hos en bil ska vara högst ett.

Beräkna sannolikheten för att en hel årsproduktion av GK-Roadrunner kommer att återkallas.

Uppgift 4

I en studie undersöktes hår- och ögonfärg hos 6800 slumpmässigt utvalda tyska män. Resultatet redovisas i följande tabell.

	Mörkt hår	Ljust hår
Bruna ögon	726	131
Grå eller gröna ögon	2133	999
Blå ögon	996	1815

Utför ett test på signifikansnivån 0.1% av om det finns ett statistiskt säkerställt samband mellan ögonfärg och hårfärg hos tyska män. Ange tydligt de uppställda hypoteserna och motivera tydligt vilken slutsats som dras från testet. (10 p)

Uppgift 5

I den organiska kemin spelar smältpunktsbestämningar en stor roll för identifierig av en substans eller för bedömning av dennas renhet. För att undersöka om renheten hos två olika partier av hydrokinon skiljer sig åt tar man därför åtta prover från vardera partiet och bestämmer smältpunkten för vart och ett av dessa prov.

Smältpunkt								
Parti 1	174.0	173.5	173.0	173.5	171.5	172.4	173.5	173.5
Parti 2	173.0	173.0	172.0	173.0	171.0	172.0	171.0	172.0

Undersök på lämpligast sätt om om vi kan anta att partierna skiljer sig åt om vi antar att att variationen mellan smältpunkterna i respektive prov är normalfördelad med en varians som är densamma för de båda partierna. Välj signifikansnivån 5%. Ange tydligt vilka de uppställda hypoteserna är och vad slutsatsen är.

Uppgift 6

Nedan finns data från två olika gruvor A och B. Data visar silverhalten i ounces per ton malm hos respektive malmprov från de 9 malmproverna från gruva A och de 14 från gruva B.

Silverhalt														
Gruva A	32	41	33	28	18	36	19	29	51					
Gruva B	22	25	28	34	28	33	28	23	20	19	42	21	20	37

Data varierar på ett sådant sätt att de inte kan anses vara normalfördelade.Undersök på lämpligast sätt om om vi kan anta att gruvorna skiljer sig åt vad gäller silverhalten i malmen. Använd signifikansnivån 5% Ange tydligt vilka de uppställlda hypoteserna är. Din slutsats skall tydligt motiveras.

Uppgift 7

Gör en fullständig boxplott för de mätdata som kommer från gruva B i uppgift 6. Ange även kvartilavstånd och variationsbredd för dessa mätdata.

LÖSNINGSFÖRSLAG TENTAMEN I SF1902 MATEMATISK STATISTIK.
MÅNDAGEN DEN 15:E AUGUSTI 2016 KL 8.00–13.00

Uppgift 1

Först bestäms

$$P(A \cap B) = P(B|A)\, P(A) = \frac{2}{3} \cdot \frac{3}{5} = \frac{2}{5}.$$

På samma sätt fås

$$P(A^* \cap B) = P(B|A^*)\, P(A^*) = \frac{1}{3} \cdot (1 - \frac{3}{5}) = \frac{1}{3} \cdot \frac{2}{5} = \frac{2}{15}.$$

Sålunda får vi att

$$P(B) = P(A \cap B) + P(A^* \cap B) = \frac{2}{5} + \frac{2}{15} = \frac{8}{15}.$$

Eftersom $P(A \cap B) = 2/5$, medan

$$P(A)\, P(B) = \frac{3}{5} \cdot \frac{8}{15} = \frac{8}{25} \neq \frac{2}{5},$$

så är händelserna A och B <u>beroende</u>.

Uppgift 2

Låt X = nettovinsten vid en spelomgång. X antar värdet 2 med sannolikheten $1/4$ och värdet -1 med sannolikheten $3/4$.
a)

$$E(X) = 2 \cdot \frac{1}{4} + (-1) \cdot \frac{3}{4} = -\frac{1}{4}.$$

$$V(X) = E(X^2) - (E(X))^2 = 2^2 \cdot \frac{1}{4} + (-1)^2 \frac{3}{4} - \left(-\frac{1}{4}\right)^2 = \frac{27}{16}.$$

$$D(X) = \sqrt{\frac{27}{16}} \approx 1.30.$$

b) Låt X_i = nettovinst omgång i, $i = 1, 2, \ldots, 120$. X_i:na antas oberoende. Centrala gränsvärdessatsen ger att $Y = \sum_{i=1}^{120} X_i$ är approximativt normalfördelad med $E(Y) = 120 \cdot E(X) = -30$ och $V(Y) = 120 \cdot V(X) = 202.5$, dvs $Y \in N(-30, \sqrt{202.5})$.

$$P(Y > 0) = P\left(\frac{Y + 30}{\sqrt{202.5}} > \frac{30}{\sqrt{202.5}}\right) \approx 1 - \Phi(2.11) \approx \underline{0.017}.$$

Uppgift 3

Låt oss införa den stokastiska variabeln X, där X är antalet undersökta bilar som har minst två skador hos de elektroniska kretskorten i den elektroniska servostyrningen. Sannolikheten att en viss bil har minst två skador hos de elektroniska kretskorten i den elektroniska servostyrningen är 0.4 eftersom sannolikheten för högst en skada är 0.6. Eftersom vi kan anta att sannolikheten för minst två skador är densamma för varje bil medför detta att $X \in Bin(10, 0.4)$

Vi söker nu $P(X \geq 8) = 1 - P(X \leq 7) =$[se t.ex tab 6 i F.S.]$= 1 - 0.98771 = 0.01229$

Svar: Sannolikheten för att en hel årsproduktion av GK-Roadrunner kommer att återkallas är 1.23 %.

Uppgift 4

Vi gör här ett homogenitetstest (avsnitt 14.3 i formelsamlingen).Nollhypotesen $H0$ är då att det inte finns ett samband mellan ögonfärg och hårfärg hos tyska män. Mothypotesen $H1$ är då att det finns ett sådant samband.

Vi gör här en tabell med observerade antal enligt

Observerade antal	Bruna ögon	Grå eller Gröna ögon	Blå ögon	Totalt
Mörkt hår	726	2133	996	3855
Ljust hår	131	999	1815	2945
Totalt	857	3132	2811	6800

Teststorheten blir

$$Q = \frac{(726 - \frac{857 \cdot 3855}{6800})^2}{\frac{857 \cdot 3855}{6800}} + \frac{(2133 - \frac{3132 \cdot 3855}{6800})^2}{\frac{2132 \cdot 3855}{6800}} + \frac{(996 - \frac{2811 \cdot 3855}{6800})^2}{\frac{2811 \cdot 3855}{6800}} +$$

$$+ \frac{(131 - \frac{857 \cdot 2945}{6800})^2}{\frac{857 \cdot 2945}{6800}} + \frac{(999 - \frac{3132 \cdot 2945}{6800})^2}{\frac{3132 \cdot 2945}{6800}} + \frac{(1815 - \frac{2811 \cdot 2945}{6800})^2}{\frac{2811 \cdot 2945}{6800}} = 957.6756$$

Om H_0 är sann så är 957.6756 ett utfall från en stokastisk variabel som approximativt har en χ^2-fördelning med $(3-1)(2-1) = 2$ frihetsgrader. Eftersom $\chi^2_{0.001}(2) = 13.8 < 957.6756$ så kan H_0 förkastas på nivån 0.1%. Alternativt kan vi beräkna sannolikheten att en $\chi^2(2)$-variabel är större än eller lika med 957.6756 (`X2cdf` på en TI-räknare). Denna sannolikhet, dvs p-värdet för testet, är 0.000. Detta p-värde är så lågt att vi förkastar H_0 på risknivån 0.1%. Både teststorheten och p-värdet fås direkt med funktionen `X2-Test` på en TI-räknare.

Svar: Vi drar slutsatsen att sambandet mellan hårfärg ogh ögonfärg för tyska män är statistiskt säkerställd på risknivån 0.1% .

Uppgift 5

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

Låt $x_1, \ldots, x_8$ och $y_1, \ldots, y_8$ vara de uppmätta smältpunkterna i parti 1 respektive parti 2. Dessa är observationer av stokastiska variabler $X_1, \ldots, X_8$ respektive $Y_1, \ldots, Y_8$ (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 $n_X = n_Y = 8$, $\bar{x} = 173.11$, $\bar{y} = 172.12$, $s_X = 0.80256$, $s_Y = 0.83452$, $s = 0.81870$, $t_{0.025}(14) = 2.14$ får vi att $I_{\mu_Y - \mu_X} = 0.99 \pm 0.88$. Eftersom 0 inte ingår i konfidensintervallet kan man förkasta H_0 på signifikansnivån 5%. Detta innebär att man med 5% felrisk kan påstå att det föreligger systematisk skillnad mellan de två partierna vad gäller smältpunkten.

<u>Svar:Vi kan anta att partierna skiljer sig åt på risknivån 5 %.</u>

Uppgift 6

Här använder vi Wilcoxons rangsummetest. Vi har hypoteserna
H_0 : De två gruvornas silverhalt skiljer sig inte åt.
H_1 : De två gruvornas silverhalt skiljer sig åt.

Om vi ordnar data i storleksordning fås

Gruva A	18	19	28	29	32	33	36	41	51					
Gruva B	19	20	20	21	22	23	25	28	28	28	33	34	37	42

Vi får följande ranger

Gruva A	1	2.5	11.5	14	15	16.5	19	21	23					
Gruva B	2.5	4.5	4.5	6	7	8	9	11.5	11.5	11.5	16.5	18	20	22

Vi har då att rangsummorna blir $W_A = 123.5$, $W_B = 152.5$, $n_A = 9$ och $n_B = 14$. Detta ger oss att $U_A = 78.5$ och $U_B = 47.5$. Vi ska testa på nivån 5% och det kritiska värdet blir då $U_{0.05} = 31$. Eftersom U_B är mindre än U_A jämförs U_B med det kritiska värdet. Eftersom $U_B = 47.5 \geq U_{0.05} = 31$ förkastas inte H_0 på nivån 5%.
<u>Vi drar slutsatsen att vi inte kan säga att silverhalten skiljer sig åt mellan de två gruvorna</u>

Uppgift 7

$x_{\min} = 19, x_{\max} = 42$, så variationsbredden blir $42 - 19 = 23$.
Medianen blir $(x_7 + x_8)/2 = (25 + 28)/2 = 26.5$.

$$Q_1 : \; 0.25 \cdot 14 = 3.5; \; 0.25 \cdot 14 + 1 = 4.5 \Rightarrow Q_1 = x_4 = 21$$

$$Q_3 : \; 0.75 \cdot 14 = 10.5; \; 0.75 \cdot 14 + 1 = 11.5 \Rightarrow Q_3 = x_{11} = 33.$$

Kvartilavståndet blir alltså $Q_3 - Q_1 = 33 - 21 = 12$.