



Avd. Matematisk statistik

KTH Matematik

TENTAMEN I SF1902 SANNOLIKHETSTEORI OCH STATISTIK
ONSDAGEN DEN 7:E JUNI 2017 KL 14.00–19.00.

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.

Uppgift 1

Antalet trafikolyckor per år på en viss väg kan anses vara oberoende och i genomsnitt 4 stycken.

- Beräkna sannolikheten att det inträffar fler än 8 olyckor den närmaste tvåårsperioden.
- Beräkna sannolikheten att det inträffar fler än 8 olyckor något (d.v.s. minst ett) av de kommande 5 åren.

Uppgift 2

Begreppet intelligens är operationellt definierat som resultatet av ett intelligenstest. För enkelhets skull antar vi att resultaten är oberoende och normalfördelade med samma standardavvikelse för alla individer i alla test. Man önskar undersöka om vana vid intelligenstest medför en förändring av testpoängen. För att pröva detta dras ett stickprov om 7 individer, vilka tidigare ej genomgått test. Dessa får genomgå prov på 6 test vardera, och nedan anges resultat på första och sista testet:

Individ nr	1	2	3	4	5	6	7
Första testet	108	142	121	138	112	102	91
Sista testet	106	146	122	140	117	104	94

Vi kan anta att resultaten från det första testet kommer från en normalfördelning $N(\mu_1, \sigma_1)$ och att resultaten från det sista testet kommer från en normalfördelning $N(\mu_1 + \Delta, \sigma_2)$. Avgör på risknivån 5% om det skett någon förändring i testpoäng. Ange tydligt vilka de uppställda hypoteserna är och vad slutsatsen är.

Uppgift 3

I samband med ett stort idrottsevenemang skall ett stort byggprojekt genomföras. Firma A bygger i två faser, där tiden i dagar för första fasen är $N(500, 50)$ och tiden i dagar för andra fasen är $N(400, 30)$ enligt planen. Fas 2 påbörjas direkt när fas 1 är avslutad. Kostnaden för första fasen är 8 miljoner kr/dag. Kostnaden för andra fasen är 10 miljoner kr/dag. Firma B har utvecklat en ny produktionsmetod med endast en fas där deras byggtid i dagar är $N(600, 60)$. Firma B tar dock 20 miljoner kr/dag. Alla fasers tider kan antas vara oberoende. Beräkna sannolikheten att det blir mer än 5 miljarder kr dyrare att anlita Firma B än firma A.

Uppgift 4

En enhet sänder ut signaler, 0:or och 1:or. Sannolikheten att en utsänd signal är en 1:a är p , att den är en 0:a är $q = 1 - p$ och de är 0:or och 1:or oberoende av varandra. Vad är sannolikheten att av 16 utsända signaler högst 10 stycken är en 1:a om $p = 0.6$?

Uppgift 5

Vid en tillverkningsprocess kontrolleras de tillverkade enheterna i en datorstyrd sensor. Härvid klassificeras defekta enheter som defekta med sannolikheten 0.9 och som korrekta med sannolikheten 0.1. Vidare klassificeras korrekta enheter som korrekta med sannolikheten 0.85 och som defekta med sannolikheten 0.15. Vad är den betingade sannolikheten att en enhet är defekt givet att den klassificerats som defekt, om processens felsannolikhet är 0.1?

Uppgift 6

Man misstänkte att ett roulettbord på ett kasino var manipulerat och genomförde ett test med 8000 försök. Om rouletten är korrekt skall röd, svart, grön(nollan) komma upp i proportionerna 18:18:1. Testresultatet gav röd: 3751, svart: 4018, grön: 231. Avgör med felrisken 1% om rouletten är korrekt. Ange tydligt vilka de uppställda hypoteserna är och vad slutsatsen är.

Uppgift 7

Man vill undersöka om kokta kräftor från företag A respektive företag B smakar lika gott, Man anlidade en panel av kräftätareexperter som fick smaka på kräftor från A respektive B. Eftersom smakerna varierade inte bara mellan fabrikaten, utan också mellan förpackningar fick panelen rangordna smakerna från flera förpackningar. Man hade tillgång till 10 förpackningar från företag A och 17 förpackningar från företag B. Panelen rangordnade de 27 förpackningarna efter smak i stigande ordning så här:

A A B B B A B A A B B B A B A A A A B B B B B B B B B

Kan man på risknivån 10% avgöra om det finns en systematisk skillnad i smak mellan de två företagens kokta kräftor? Ange tydligt vilka de uppställda hypoteserna är och vad slutsatsen är.



Avd. Matematisk statistik

KTH Matematik

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

Uppgift 1

Låt Y beteckna antalet inträffade olyckor följande tvåårsperiod. Då gäller att $Y \in Po(4 \cdot 2) = Po(8)$
Då fås $P(Y > 8) = 1 - P(Y \leq 8) = [\text{se tab 5}] = 1 - 0.59255 = 0.40745$

b) Låt $X_i =$ antal olyckor år i , $i = 1, 2, 3, 4, 5$.

$P(\text{fler än 8 olyckor minst ett år}) = 1 - P(\text{högst 8 olyckor varje år})$

$$\begin{aligned} &= 1 - P(X_1 \leq 8 \cap X_2 \leq 8 \cap X_3 \leq 8 \cap X_4 \leq 8 \cap X_5 \leq 8) = \left\{ X_i\text{:na oberoende, likafördelade} \right\} \\ &= 1 - P(X_1 \leq 8)^5 = \left\{ \text{formelsaml tab 7} \right\} = 1 - 0.97864^5 = 0.1023. \end{aligned}$$

Uppgift 2

Här har vi parvisa jämförelser. Bilda differenser z_i inom individer (sista–första).

Nya data. $-2, 4, 1, 2, 5, 3$

som är utfall av Z_i där Z_i :na oberoende och $N(\Delta, \sigma)$. $\bar{z} = \frac{1}{7}(-2+4+1+2+5+2+3) = 15/7 \approx 2.143$,
 $s_z^2 = \frac{1}{7-1}((-2)^2 + \dots + 3^2 - 15^2/7) \approx 5.14$ som ger $s_z \approx 2.27$.

Hypotesen $H_0 : \Delta = 0$ testas till exempel med konfidensmetoden. Ett 95 % konfidensintervall för Δ blir

$$I_\Delta = \bar{z} \pm t_{0.025}(7-1) \cdot \frac{s_z}{\sqrt{7}} = 2.143 \pm 2.45 \cdot \frac{2.27}{\sqrt{7}} = 2.143 \pm 2.10$$

Eftersom 0 ej tillhör intervallet kan H_0 förkastas på nivån 5%.

Svar: Att ingen förändring i testpoäng skett kan förkastas på risknivån 5%.

Uppgift 3

Låt X vara byggtiden i dagar för första fasen hos firma A och Y byggtiden i dagar för andra fasen hos firma A. Låt sedan Z vara byggtiden i dagar för firma B.

Vi söker $P(20 \cdot Z - (8 \cdot X + 10 \cdot Y) > 5000)$.

Vi har att $X \in N(500, 50)$, $Y \in N(400, 30)$, $Z \in N(600, 60)$. Sätt $W = 20 \cdot Z - (8 \cdot X + 10 \cdot Y)$.

Eftersom W är en linjärkombination av oberoende Normalfördelade stokastiska variabler är även W Normalfördelad.

$$E(W) = 20 \cdot 600 - 8 \cdot 500 - 10 \cdot 400 = 4000$$

$$V(W) = 20^2 \cdot 60^2 + 8^2 \cdot 50^2 + 10^2 \cdot 30^2 = 169 \cdot 10^4$$

$$D(W) = \sqrt{169 \cdot 10^4} = 13 \cdot 10^2$$

Nu har vi att $W \in N(4000, 1300)$

$$P(Z > 5000) = 1 - \Phi\left(\frac{5000 - 4000}{1300}\right) = 1 - \Phi\left(\frac{1}{13}\right) = 1 - \Phi(0.77) = 0.2206.$$

Svar: Sannolikheten att det blir mer än 5 miljarder dyrare att anlita firma B är ungefär 22%.

Uppgift 4

Sätt X lika med antal utsända 1:or. Då är $X \in \text{Bin}(16, 0.6)$. Antal utsända 0:or, Y , är på samma sätt $\text{Bin}(16, 0.4)$. Vi får att

$$P(X \leq 10) = P(Y \geq 6) = 1 - P(Y \leq 5) = 1 - 0.3288 = \underline{0.6712}$$

med värdet hämtas ur binomialtabell.

Uppgift 5

Här har vi Bayes sats.

Låt KK beteckna händelsen att en enhet klassificeras som korrekt, KD händelsen att en enhet klassificeras som defekt, K händelsen att en enhet är korrekt, och D händelsen att en enhet är defekt.

Sökt är

$$\begin{aligned} P(D|KD) &= \frac{P(D \cap KD)}{P(KD)} = \frac{P(KD|D)P(D)}{P(KD|D)P(D) + P(KD|K)P(K)} = \\ &= \frac{0.9 \cdot 0.1}{0.9 \cdot 0.1 + 0.15 \cdot 0.9} = 0.40 \end{aligned}$$

Svar: Sannolikheten är 40% att en enhet som klassificerats som defekt verkligen är defekt.

Uppgift 6

Vi har typsituationen för χ^2 -test, nämligen test av given fördelning. Här med tre möjliga resultat: $A_1 \Leftrightarrow$ Rött, $A_2 \Leftrightarrow$ Svart, $A_3 \Leftrightarrow$ Grönt (=nollan)

D.v.s. $r = 3$. Beteckningar enligt formelsamlingen. Nollhypotesen H_0 att rouletten är korrekt uttrycks då som följer.

$$H_0 : p_1 = \frac{18}{37}, p_2 = \frac{18}{37} \text{ och } p_3 = \frac{1}{37}.$$

Eftersom $np_3 = 8000 \cdot \frac{1}{37} = 216.22 \geq 5$ så gäller att alla $np_j \geq 5$, vilket är ett krav för χ^2 -testet.

Vi har alltså $n = 8000$ observationer varav $x_1 = 3751$, $x_2 = 4018$ och $x_3 = 231$. Vi får därför

$$Q = \sum_{j=1}^r \frac{(x_j - np_j)^2}{np_j} = \frac{(3751 - 3891.89)^2}{3891.89} + \frac{(4018 - 3891.89)^2}{3891.89} + \frac{(231 - 216.22)^2}{216.22} = 10.20.$$

$\chi_\alpha^2(r-1) = \chi_{0.01}^2(2) = 9.21 < 10.20$, vilket gör att vi förkastar H_0 på 1% nivån.

Svar: Vi drar alltså slutsatsen att rouletten inte är korrekt.

Uppgift 7

Här använder vi Wilcoxon's rangsummetest. Låt A vara de från företag A och B de från företag B. Vi har hypoteserna

H_0 : Kräfterna från de båda företagen smakar lika bra.

H_1 : Kräfterna från de båda företagen smakar inte lika bra.

Vi får följande ranger

Företag A	1	2	6	8	9	13	15	16	17	18							
Företag B	3	4	5	7	10	11	12	14	19	20	21	22	23	24	25	26	27

Vi har då att rangsummorna blir $W_A = 105$, $W_B = 273$, $n_A = 10$ och $n_B = 17$ detta ger oss $U_A = 50$ och $U_B = 120$. Vi ska testa på nivån 10% och det kritiska värdet blir då $U_{0.10} = 51$. Eftersom U_A är mindre än U_B jämförs U_A med det kritiska värdet. Eftersom $U_A = 50 \leq U_{0.10} = 51$ förkastas H_0 på nivån 10%.

Svar: Vi drar slutsatsen att kräftorna från de båda företagen inte smakar lika bra (och att alltså kräftorna från företag B är de som smakar bäst).