KTH Matematik

Avd. Matematisk statistik

TENTAMEN I SF1902 SANNOLIKHETSTEORI OCH STATISTIK,
TISDAGEN DEN 9:E JUNI 2015 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

Antag att i genomsnitt var tionde bil som passerar Häggvikavfarten på E6 kör för fort och att olika bilar håller oberoende hastigheter. En polis mäter hastigheten på 15 bilar.
a) Vad är sannolikheten att ingen bil kör för fort?
b) Vad är sannolikheten att minst 3 bilar kör för fort?

Uppgift 2

I faderskapsmål undersöks ofta blodgrupper hos moder, barn och den utpekade fadern. För enkelhets skull betraktar vi bara det s k AB0-systemet (i verkligheten undersöks ett flertal blodgruppssystem). För barn vars moder har blodgrupp 0 beror den av faderns blodgrupp betingade sannolikheten att barnet skall få blodgrupp A0 enligt:

faderns blodgrupp	sannolikheten att barnet skall få blodgrupp A0
A0	1/2
AA	1
AB	1/2
Övriga	0

Sannolikheten att en på måfå vald man skall ha blodgrupp A0, AA, AB respektive ”övriga”är 0.36, 0.08, 0.02 respektive 0.54.

Beräkna sannolikheten att fadern har AA när barnet har A0 och modern 0.

Uppgift 3

En ideell förening *Den Gröna Jorden* planerar en insamling för utökning av grönområden runt fastigheterna. Föreningen skickar därför ett brev till var och en av de 1000 medlemmarna där man ber om ett bidrag på 50 eller 100 kronor. Man vet från tidigare att 50 och 100 kronors bidrag är lika vanliga och att 20% av medlemmarna inte ger något bidrag.
Bestäm med hjälp av en lämplig och välmotiverad approximation, sannolikheten att föreningens sammanlagda insamling överstiger 58 000 kronor.

Uppgift 4

En icke tidigare känd obotlig sjukdom har börjat sprida sig. Man vill undersöka om ett nyutvecklat vaccin har någon effekt mot denna sjukdom. Man väljer därför slumpmässigt ut 5450 personer som befinner sig i riskområdet och vaccinerar hälften av dessa. Efter ett år finner man att 30 av de som har vaccinerats har drabbats av sjukdomen, medan 190 av de som inte har vaccinerats har drabbats av sjukdomen.
Avgör med ett väl motiverat konfidensintervall om vaccinet kan sägas ha effekt eller inte. Använd risknivån 1%. Ange tydligt vilka de uppställda hypoteserna är och vad slutsatsen är.

Uppgift 5

Vid tillverkning av en viss detalj kan två typer av fel uppstå, A och B. Dessa fel kan dock ej uppkomma bägge två på en och samma detalj.

I försäljningsspecifikationen för denna detalj står det angivet att fel A förekommer med sannolikheten 0.05, och fel B med sannolikheten 0.09. För att utröna om dessa värden fortfarande är aktuella så kontrollerades 150 enheter uttagna helt slumpmässigt. Av dessa befanns 12 st ha fel A, 21 st ha fel B, och resten var felfria.

Utför ett statistiskt test på 5% signifikansnivå för att avgöra om försäljningsspecifikationen är förenlig med data.

Uppgift 6

För att undersöka effekten av två olika rostskyddsmedel, M_1 och M_2, behandlades 10 järnstavar med medel M_1 och ytterliggare 10 stavar med medel M_2. På var och en av 10 olika platser grävdes därefter två stavar ned; en stav med varje behandling. Efter två månader togs alla stavar upp och graden av rost mättes. Följande mätvärden (i lämplig enhet) erhölls:

Plats:	1	2	3	4	5	6	7	8	9	10
Medel M_1:	19.15	23.35	20.10	16.70	31.85	17.70	22.77	21.71	34.06	18.71
Medel M_2:	21.96	27.70	22.93	19.02	32.26	17.35	27.39	25.43	33.43	18.14

Ange en lämplig statistisk modell baserad på normalfördelning som beskriver data, och undersök med hjälp av denna om det finns någon systematisk skillnad i grad av rost mellan de två rostskyddsmedlen. Välj signifikansnivån $\alpha = 0.05$.

Uppgift 7

En sjukhusingenjör vill undersöka två stålbitar som delvis innehåller radioaktivt material. Sjukhusingenjören mäter först under en timme på den ena biten och sedan under en timme på den andra, och registrerar då 270 respektive 180 sönderfall.
Undersök om det är någon skillnad mellan sönderfallsintensiterna hos de två stålbitarna på signifikansnivån 5%. Ange tydligt vilka de uppställda hypoteserna är och vad slutsatsen är.

KTH Matematik

Avd. Matematisk statistik

LÖSNINGSFÖRSLAG TENTAMEN I SF1902 MATEMATISK STATISTIK.
TiSDAGEN DEN 9:E JUNI 2013 KL 14.00–19.00

Uppgift 1

Låt X vara antalet bilar som kör för fort. Då har vi att X$\in$ Bin(n, p), där $n = 15$ och $p = 0.1$.
a)$P(Y = 0) = 0.20589$
b)$P(Y \geq 3) = 1 - P(Y \leq 2) = 1 - 0.81954 = 0.18406$

Uppgift 2

Som utfallsrum Ω tar vi händelsen att modern har blodgrupp 0, eftersom det är fallet i det aktuella exemplet och det också är förutsättningen för de redovisade sannolikheterna.

Vi inför händelserna

$$\begin{aligned} A &= \{\text{barnet har blodgrupp A0}\}, \\ H_1 &= \{\text{fadern har blodgrupp A0}\}, \\ H_2 &= \{\text{fadern har blodgrupp AA}\}, \\ H_3 &= \{\text{fadern har blodgrupp AB}\}, \\ H_4 &= \{\text{fadern har en annan blodgrupp}\}, \end{aligned}$$

och vi söker $P(H_2|A)$. Enlig Bayes sats, alternativt definitionen av betingad sannolikhet samt satsen om total sannolikhet, gäller

$$\begin{aligned} P(H_2|A) &= \frac{P(H_2 \cap A)}{P(A)} = \frac{P(A|H_2)P(H_2)}{\sum_{i=1}^{4} P(A|H_i)P(H_i)} = \\ &= \frac{1 \cdot 0.08}{1/2 \cdot 0.36 + 1 \cdot 0.08 + 1/2 \cdot 0.02 + 0 \cdot 0.54} = \frac{0.08}{0.27} = \frac{8}{27} \approx 0.296. \end{aligned}$$

Uppgift 3

Inför beteckningen X_i för antalet kronor som medlem nr i bidrar med och Y för totala antalet kronor som insamlingen ger. Vi har då att

$$Y = \sum_{i=1}^{1000} X_i = X_1 + X_2 + \ldots + X_{1000}.$$

Den sökta sannolikheten är $P(Y > 58000)$.
Vi antar att X_i:na är oberoende och likafördelade. Y blir då en summa av många, oberoende och likafördelade variabler vilket leder till att CGS kan användas. $Y \sim N(n \cdot \mu, \sigma \cdot \sqrt{n})$, där $n = 1000$, $\mu = E(X_i)$ och $\sigma = D(X_i)$

$$\begin{aligned}
E(X_i) &= \sum_k kP(X_i = k) = 0 \cdot 0.2 + 50 \cdot 0.4 + 100 \cdot 0.4 = 60 \\
E\left(X_i^2\right) &= \sum_k k^2 P(X_i = k) = 0^2 \cdot 0.2 + 50^2 \cdot 0.4 + 100^2 \cdot 0.4 = 5000 \\
V(X_i) &= E\left(X_i^2\right) - E(X_i)^2 = 5000 - 60^2 = 1400 \\
D(X_i) &= \sqrt{V(X_i)} = \sqrt{1400}
\end{aligned}$$

Vi får då att $Y \sim N(1000 \cdot 60, \sqrt{1400} \cdot \sqrt{1000})$. Den sökta sannolikheten blir

$$P(Y > 58000) = P\left(\frac{Y - 1000 \cdot 60}{\sqrt{1400 \cdot 1000}} > \frac{58000 - 1000 \cdot 60}{\sqrt{1400 \cdot 1000}}\right) \approx 1 - \Phi(\frac{-2000}{\sqrt{1400 \cdot 1000}}) =$$

$$= 1 - [1 - \Phi(\frac{2000}{\sqrt{1400 \cdot 1000}}) \approx \approx \Phi(1.69) = 0.9545$$

Svar: Föreningens sammanlagda insamling överstiger 58000 kronor med sannolikheten 0.9545.

Uppgift 4

H_0 är här att vaccinet inte har någon effekt. H_1 är att vaccinet har effekt.
Vi antar här att X_1 är antalet som drabbas av sjukdomen som inte fått vaccinet och att X_2 är antalet som drabbas av sjukdomen som fått vaccinet. 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 $p_1 - p_2$, 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 $x_1/n_1 = 190/2725$ och p_2 med $x_2/n_2 = 30/2725$. Vi får då att $n_1 p^*_{1,obs}(1-p^*_{1,obs}) = 176.75 > 10$ och $n_2 p^*_{2,obs}(1-p^*_{2,obs}) = 29.66 > 10$. 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 risknivå på 1% och då fås $\lambda_{\alpha/2} = 2.5758$ och intervallet blir

$$5.87 \cdot 10^{-2} \pm 5.27 \cdot 10^{-3} \cdot 2.5758 = 5.87 \cdot 10^{-2} \pm 1.36 \cdot 10^{-2}.$$

0 tillhör inte intervallet och vi kan således förkasta H_0 på risknivån 1%.
<u>Vi drar slutsatsen att vaccinet har effekt.</u>

Uppgift 5

Vi har typsituationen för χ^2-test med tre möjliga resultat:
$A_1 \Leftrightarrow$ Felet A inträffar, $A_2 \Leftrightarrow$ felet B inträffar,$A_3 \Leftrightarrow$ inget av felen inträffar.
D.v.s. $r = 3$. Beteckningar enligt formelsamlingen. Hypotesen är

$$H_0 : p_1 = 0.05,\ p_2 = 0.09 \ \text{ och } \ p_3 = 1 - p_1 - p_2 = 1 - 0.05 - 0.09 = 0.86.$$

Vi har $n = 150$ observationer varav $x_1 = 12$, $x_2 = 17$ och $x_3 = 150 - 12 - 21 = 117$. Vi får därför

$$Q = \sum_{j=1}^{r} \frac{(x_j - np_j)^2}{np_j} = \frac{(12 - 150 \cdot 0.05)^2}{150 \cdot 0.05} + \frac{(21 - 150 \cdot 0.09)^2}{150 \cdot 0.09} + \frac{(117 - 150 \cdot 0.86)^2}{150 \cdot 0.86} = 7.98.$$

$\chi^2_\alpha(r-1) = \chi^2_{0.05}(2) = 5.99 < 7.98$, vilket gör att vi förkastar H_0 på 5% nivån.
<u>Vi drar alltså slutsatsen att försäljningsspecifikationen inte är förenlig med data.</u>

Eftersom $np_1 = 150 \cdot 0.05 = 7.5 \geq 5$ så gäller att alla $np_j \geq 5$, vilket är ett krav för χ^2-testet.

Uppgift 6

Stickprov i par. De parvisa skillnaderna

Stålstav nr:	1	2	3	4	5	6	7	8	9	10
y_i: (Medel M_2-Medel M_1):	2.81	4.35	2.83	2.32	0.41	-0.35	4.62	3.72	-0.63	-0.57

är utfall av $N(\Delta, \sigma)$-fördelade stokastiska variabler, där Δ skattas med $\overline{y} = 1.951$ och σ med $s = 2.064006083$.
Hypoteserna blir
$H_0 : \Delta = 0$
$H_1 : \Delta \neq 0$

Vi gör konfidensintervall för Δ, med risknivån $\alpha = 0.05$, då standardavvikelsen är okänd och får att

$$\begin{aligned} I_\Delta &= (\overline{y} - t_{0.025}(9) \cdot \frac{s}{\sqrt{10}}, \overline{y} + t_{0.025}(9) \cdot \frac{s}{\sqrt{10}}) \\ &= (1.951 - 2.26 \cdot \frac{2.064006083}{\sqrt{10}}, 1.951 + 2.26 \cdot \frac{2.064006083}{\sqrt{10}}) \\ &= (0.476, 3.426) \end{aligned}$$

eller $I_\Delta = 1.951 \pm 1.475$
Eftersom 0 inte ligger i intervallet kan H_0 förkastas.
Vi drar slutsatsen att det finns en systematisk skillnad i grad av rost mellan de två rostskyddsmedlen.

Uppgift 7

Här antar vi att antalet sönderfall i en stålbit är Poissonfördelat. Vi antar att X_i är antalet uppmätta sönderfall hos stålbit i, och att $X_1 \in Po(\mu_1)$ och att $X_2 \in Po(\mu_2)$. Nollhypotesen blir i detta fall att det inte är någon skillnad i radioaktivitet mellan stålbitarna.
Således har vi $H_0 : \mu_1 = \mu_2$. $H_1 : \mu_1 \neq \mu_2$.
Här använder vi χ^2-test:

$$Q = \sum_{j=1}^{2} \frac{(x_j - \mu^*_{obs})^2}{\mu^*_{obs}}$$

Eftersom $E\left(X_i\right) = \mu$ när $X_i \in Po(\mu_i)$ så skattar vi μ med

$$\bar{x} = \frac{x_1 + x_2}{2} = \frac{180 + 270}{2} = 225$$

Detta ger

$$Q = \frac{(180 - 225)^2}{225} + \frac{(270 - 225)^2}{225} = 18$$

$$\chi^2_\alpha(r - 1) = \chi^2_{0.05}(2 - 1) = 3.84$$

Eftersom $Q > \chi^2_{0.05}(1)$ så förkastar vi H_0 på risknivån 5%.
Vi drar slutsatsen att det är skillnad i radioaktivitet mellan stavarna.