MVE090 Matematisk statistik Z, 7.5 hp

Tentamen 17 oktober fm V

Tillåtna hjälpmedel: Valfri räknedosa utan lagrad information om kursen, Beta samt kursens formel- och tabellsamling.

Examinator: Tommy Norberg, ankn 3528 eller 0730 79 42 09.

Jour och telefonvakt: Tommy Norberg.

Maximalt antal tentamenspoäng är 30, av dessa krävs normalt 12 för godkänt betyg och 18 resp 24 för 4:a och 5:a. Lösningar till tentamensproblemen går att ladda ner från kurshemsidan.

Svar och lösningar till följande 8 uppgifter skall motiveras om ej annat sägs.

Obs att en bilaga med data bifogas tesen.

Uppgifter

1. Låt A och B vara händelser, sådana att $P(A) = 0.3$, $P(B) = 0.7$ och $P(B|A) = 0.5$. Bestäm $P(A|B)$. (3 p)

2. Låt A_1, A_2, A_3 vara oberoende och sådana att $P(A_i) = 0.25$, $i = 1, 2, 3$. Låt den stokastiska variabeln X vara 0, 1, 2 eller 3, beroende på hur många A-händelser som inträffar. Vilken fördelning har X? (3 p)

3. Olyckor av en viss typ tror man inträffar med intensiteten ca 1.33 st/månad. Bestäm approximativt

   (a) förväntad tid till nästa olycka,

   (b) sannolikheten att det under nästa år inträffar 8–24 olyckor.

   Det är viktigt att du i båda uppgifterna noga motiverar och anger förutsättningarna för dina beräkningar. Försök att inte räkna för noggrannt. (4 p)

4. Antag att X, Y är likformigt fördelad på triangeln i planet med hörnen i $(0, 0)$, $(1, 0)$ samt $(0, 1)$. Detta innebär att tätheten är

$$f(x, y) = c$$

   för $x, y \in (0, 1)$ med $x + y < 1$. Bestäm den betingade tätheten för Y givet $X = x$. (4 p)

5. Ett slumpmässigt urval av batterier utsattes för ett s.k accelererat livslängdstest. De arbetade då kontinuerligt i en förhöjd temperatur och med ett högt strömuttag. Man mätte antalet timmar batterierna fungerade tillfredsställande. Data kan sammanfattas enligt

$$n = 8 \quad \sum x = 206.00 \quad \sum x^2 = 5311.3670$$

   Tillverkarens krav på batteritypen är att under dessa omständigheter ska batterierna i medel fungera i minst 25 timmar. I vilken utsträckning kan man hävda att data bekräftar kravet? Antag normalfördelade livslängder. (4 p)

Databilaga till tentamen MVE090 111017

Följande tabell visar data från försöket med insektsmedel, som ska analyseras i uppgift 7.

i	x_i	y_i	$d_i = x_i - y_i$
1	8	5	3
2	7	6	1
3	8	6	2
4	8	7	1
5	12	11	1
6	14	14	0
7	8	6	2
8	12	11	1
9	13	10	3
10	9	7	2
11	12	9	3
12	9	8	1
13	7	5	2
14	14	12	2
15	9	12	−3
16	11	10	1
17	11	11	0
18	9	8	1
19	14	12	2
20	14	13	1
21	13	12	1
22	9	8	1
23	15	12	3
24	8	6	2
25	9	9	0
26	11	9	2
27	7	6	1
28	11	10	1
29	6	6	0
30	9	9	0

Här följer medelvärde och standardavvikelse för samtliga datamängder:

$$
\begin{array}{lll}
n = 30 & \bar{x} = 10.23 & s_x = 2.569 \\
 & \bar{y} = \ \ 9.00 & s_y = 2.613 \\
 & \bar{d} = \ \ 1.23 & s_d = 1.223
\end{array}
$$

LÖSNINGAR

MVE090 Matematisk statistik Z, 7.5 hp

Tentamen 17 oktober fm V

Tillåtna hjälpmedel: Valfri räknedosa utan lagrad information om kursen, Beta samt kursens formel- och tabellsamling.

Examinator: Tommy Norberg, ankn 3528 eller 0730 79 42 09.

Jour och telefonvakt: Tommy Norberg.

Maximalt antal tentamenspoäng är 30, av dessa krävs normalt 12 för godkänt betyg och 18 resp 24 för 4:a och 5:a. Lösningar till tentamensproblemen går att ladda ner från kurshemsidan.

Svar och lösningar till följande 8 uppgifter skall motiveras om ej annat sägs.

Obs att en bilaga med data bifogas tesen.

Uppgifter

1. Låt A och B vara händelser, sådana att $P(A) = 0.3$, $P(B) = 0.7$ och $P(B|A) = 0.5$. Bestäm $P(A|B)$. (3 p)

   Lösning: $P(A|B) = \dfrac{P(A)P(B|A)}{P(B)} = \dfrac{0.3 \cdot 0.5}{0.7} \approx 0.214$

2. Låt A_1, A_2, A_3 vara oberoende och sådana att $P(A_i) = 0.25$, $i = 1, 2, 3$. Låt den stokastiska variabeln X vara 0, 1, 2 eller 3, beroende på hur många A-händelser som inträffar. Vilken fördelning har X? (3 p)

   Lösning: $X \sim \text{Bin}(n = 3, p = 0.25)$ ty man kan se X som antalet lyckade försök om man betraktar försök nr i som lyckat då A_i inträffar och misslyckat annars.

3. Olyckor av en viss typ tror man inträffar med intensiteten ca 1.33 st/månad. Bestäm approximativt

   (a) förväntad tid till nästa olycka,

   (b) sannolikheten att det under nästa år inträffar 8–24 olyckor.

   Det är viktigt att du i båda uppgifterna noga motiverar och anger förutsättningarna för dina beräkningar. Försök att inte räkna för noggrannt. (4 p)

   Lösning: Det är inte orimligt att tänka sig att olyckor inträffar enligt en Poissonprocess med fix intensitet λ. Här gäller att $\lambda \approx 1.33$ olyckor per månad. Då är tiden till nästa olycka $\exp(\lambda)$, så (a) förväntad tid till nästa olycka är $1/\lambda \approx 0.75$ månader. Antalet olyckor under nästa år är $\text{Poi}(12\lambda \approx 16)$. Väntevärdet är ≈ 16 och variansen är $\approx 16 = 4^2$. Centrala gränsvärdessatsen ger att antalet olyckor är approximativt normalfördelat med väntevärde 16 och standardavvikelse 4. Enligt en känd tumregel gäller (b) att sannolikheten för 8–24 olyckor är ca 0.95.

4. Antag att X, Y är likformigt fördelad på triangeln i planet med hörnen i $(0,0)$, $(1,0)$ samt $(0,1)$. Detta innebär att tätheten är

$$f(x, y) = c$$

för $x, y \in (0,1)$ med $x + y < 1$. Bestäm den betingade tätheten för Y givet $X = x$. (4 p)

Lösning: $f_X(x) = \int_0^{1-x} c\, dy = c(1-x)$ för $0 < x < 1$, så $f_{Y|X=x}(y) = \frac{f(x,y)}{f_X(x)} = \frac{1}{1-x}$ för $0 < y < 1-x$. Vi ser att $Y|X = x \sim \mathrm{U}(0,\, 1-x)$.

5. Ett slumpmässigt urval av batterier utsattes för ett s.k accelererat livslängdstest. De arbetade då kontinuerligt i en förhöjd temperatur och med ett högt strömuttag. Man mätte antalet timmar batterierna fungerade tillfredsställande. Data kan sammanfattas enligt

$$n = 8 \quad \sum x = 206.00 \quad \sum x^2 = 5311.3670$$

Tillverkarens krav på batteritypen är att under dessa omständigheter ska batterierna i medel fungera i minst 25 timmar. I vilken utsträckning kan man hävda att data bekräftar kravet? Antag normalfördelade livslängder. (4 p)

Lösning: Här gäller $n = 8$, $\bar{x} = \frac{206.00}{8} = 25.75$ och $s^2 = \frac{1}{7}\left(5311.3670 - \frac{206.00^2}{8}\right) = 0.9810 = 0.9905^2$. Jag väljer att testa nollhypotesen $H_0 : \mu = 25$ mot det ensidiga alternativet $H_1 : \mu > 25$. Uppmätt värde på teststatistikan är $t = \frac{25.75 - 25}{0.9905/\sqrt{8}} \approx 2.142$ och vi ser i en t-kvantiltabell att p-värdet ligger mellan 2.5% och 5%. Obs att antalet frihetsgrader är $n - 1 = 7$. Jag kan alltså förkasta H_0 på nivån 5%.

6. Ett sätt att tolka ordet kvalitet är att hålla variationerna i tillverkningen nere. Vid ett tillfälle valde man slumpmässigt ut 15 tillverkade enheter och mätte en för den fortsatta tillverkningen viktig diameter. Härvid erhölls medelvärdet 7.36 cm och variansen 0.028. En kontroll visar att data är approximativt normalfördelade. Kan man med 99% säkerhet kan påstå att standardavvikelsen är högst 0.30 mm?(4 p)

Lösning: Beklagar tryckfelet. Uppgiften är alltså att kolla om vi med 99% säkerhet kan påstå att standardavvikelsen är högst 0.30 cm? Här gäller $n = 15$, $\bar{x} = 7.36$ och $s^2 = 0.028 \approx 0.1673^2$. Lättast är att beräkna ett 99% uppåt begränsat konfidensintervall för σ och undersöka om gränsen ≤ 0.30 eller ej. I en χ^2-kvantiltabell avläser vi att 99%-kvantilen är 4.660 (obs 14 frihetsgrader). Så den övre gränsen i intervallet är $\mathrm{ucl99} = \sqrt{\frac{14 \cdot 0.028}{4.660}} \approx 0.29 < 0.30$. Vi ser att svaret är ja på den ställda frågan.

7. Under Djungelkriget led landet Y:s soldater svårt av insektsbett. Ett nytt insektsmedel utvecklades och jämfördes med det ordinarie medlet genom att man på 30 slumpmässigt utvalda soldater smorde in en godtyckligt vald arm med det nya

medlet och sedan smorde in den återstående armen med standardmedlet. Efter ett antal timmar i djungeln undersöktes armarna. I bifogad tabell visas soldatnummer (i), antal bett på armen med standardmedlet (x_i), antal bett på armen med det nya medlet (y_i) samt differensen ($d_i = x_i - y_i$). Heltalsdata är ju verkligen inte normalfördelade, men låt oss anta att centrala gränsvärdessatsen är tillämpbar. Utforma och genomför ett lämpligt test för att avgöra om det nya insektsmedlet är bättre än standardmedlet. (3 p)

Lösning: Vi testar om medelvärdet av differenserna är signifikant större än noll på, eftersom inget annat är sagt, nivån 5%. (Vi ska alltså testa $H_0 : \mu_d = 0$ mot $H_1 : \mu_d > 0$). Data är $n = 30$, $\bar{d} = 1.23$ och $s_d = 1.223$. Observerat värde på teststatistikan är $t = \frac{1.23}{1.223/\sqrt{30}} \approx 5.509$. Detta värde är mycket större än 1%-kvantilen som är 2.462 (obs 29 frihetsgrader). Vi kan alltså med en säkerhet som är betydligt större än 99% hävda att det nya medlet är bättre.

8. Det är inte ovanligt att man i enkätundersökningar bryter ner resultaten för att t.ex påvisa regionala skillnader. I en partisympatiundersökning tillfrågades totalt 1682 slumpmässigt utvalda röstberättigade individer. Av dessa valde 231 att inte svara alls. Låt oss anta att dessa väljare inte tänker rösta, så att vi kan bortse ifrån deras sympatier. Av de återstående 1451 väljarna sympatiserade 7.8% med X-partiet.

   (a) Intervallskatta X-partiets väljarandel. Konfidensgrad: ca 95%. (2 p)

Av de 1451 valjarna var 917 st storstadsbor (8.9% av dessa sympatiserade med X-partiet) och 534 boende i övriga landet (5.8% av dessa sympatiserade med X-partiet). Det skiljer alltså 3.1%-enheter mellan storstad och övriga landet.

   (b) Testa på nivån 5% nollhypotesen ingen skillnad i väljarandelarna mot alternativet att väljarandelen är större i storstäderna än i övriga landet. (3 p)

Lösning: (a) $p = 0.078 \pm 1.96\sqrt{\frac{0.078(1-0.078)}{1451}} = 0.078 \pm 0.014 \in (0.064, 0.092)$.
(b) Vi ska testa $H_0 : p_s = p_l$ mot $H_1 : p_s > p_l$, där p_s respektive p_l är partiets väljarandel i storstad och övriga landet. Estimaten är $\hat{p}_s = 0.089$ och $\hat{p}_l = 0.058$ och under $H_0 : p_s = p_l$ estimeras den gemensamma proportionen av $\hat{p} = 0.078$. Teststatistikan är $z = \frac{0.089 - 0.058}{\sqrt{0.078(1-0.078)\left(\frac{1}{917} + \frac{1}{534}\right)}} \approx 2.12$. Sannolikheten för en minst lika stor observation är 0.0169, så H_0 kan förkastas på nivån 5%. Alternativt kan testet utföras genom att man beräknar en undre konfidensgräns för $p_s - p_l$ med konfidensen 95% och ser om denna är större än 0 eller ej. Då fås $p_s - p_l \geq$
$$0.031 - 1.645\sqrt{\frac{0.089(1-0.089)}{917} + \frac{0.058(1-0.058)}{534}} \approx 0.031 - 0.023 = 0.008 > 0.$$
Även denna metod ger att vi kan förkasta H_0 på nivån 5%.