MVE090 Matematisk statistik Z, 7.5 hp

Tentamen 23 augusti 2011 em V

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

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

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.

Svar och lösningar skall motiveras om ej annat sägs i uppgiften.

Uppgifter

1. I en urna ligger 7 röda, 3 gröna och 4 blåa bollar. Ur den ska 3 st bollar dras. Beräkna med klassisk metodik sannolikheten att man drar en boll av varje färg. (3 p)

   Lösning: Den sökta sannolikheten är

$$\frac{\binom{7}{1}\binom{3}{1}\binom{4}{1}}{\binom{14}{3}} = \frac{7\cdot 3\cdot 4}{\frac{14\cdot 13\cdot 12}{3\cdot 2}} = \frac{3}{13} \approx 0.231$$

2. Enheter som produceras på ett löpande band är antingen felfria eller behäftade med fel. Av fem felaktiga enheter är i snitt ett allvarligt. Varje producerad enhet kontrolleras (ev i flera steg). Undersökningar har visat att ca 95% av enheterna med allvarliga fel och ca 90% av enheterna med mindre allvarliga fel hittas i det första kontrollsteget, men man kan i detta inte avgöra vilken typ av fel en funnen enhet har. för det krävs ytterligare undersökningar. Man vet att ca 90% av alla tillverkade enheter är felfria. En tillverkningsbatch består av 1000 enheter.

   (a) Ungefär hur många enheter återstår i batchen när de felaktiga enheterna som hittats i det första kontrollsteget plockats bort? (1 p)

   (b) Ungefär hur många av dessa är försedda med mindre allvarliga fel? (1 p)

   (c) Ungefär hur många av dessa är försedda med allvarliga fel? (1 p)

   (d) Av de utsorterade (alltså de med detekterade fel), ungefär hur stor är proportionen enheter med allvarliga fel? (1 p)

   Lösning: I en tillverkningsbatch kommer det att finnas ca $1000 \cdot 0.9 = 900$ felfria enheter. Av de återstående ca 100 enheterna är $\approx 100 \cdot \frac{4}{5} = 80$ behäftade med mindre allvarliga fel och $\approx 100 \cdot \frac{1}{5} = 20$ med allvarliga fel. Av de ca 80 enheterna med mindre allvarliga fel kommer $\approx 80 \cdot 0.90 = 72$ att hittas. Ca 8 enheter med mindre allvarliga fel kommer alltså att vara kvar i batchen. Av de ca 20 enheterna med allvarliga fel kommer $\approx 20 \cdot 0.95 = 19$ att hittas. Ca 1 enhet med allvarliga fel kommer följaktligen att finnas kvar i batchen. Så svaret på fråga (a) är 909 enheter, (b) 8 enheter, (c) 1 enhet. På fråga (d) är svaret $\frac{19}{19+72} = \frac{19}{101} \approx 0.209$, alltså knappt 21%.

3. Den geometriska fördelningen definieras av att sannolikheten för utfallet k är

$$f(k) = p(1-p)^{k-1} \quad \text{för} \quad k = 1, 2, \ldots$$

där $0 < p < 1$. Visa att väntevärdet i denna fördelning är $\mu = \frac{1}{p}$. (4 p)

<u>*Lösning:*</u> Lättast är nog att ta vägen via momentgenererande funktion mgf, som för den geometriska fördelningen är

$$m(t) = \frac{pe^t}{1-qe^t} \quad \text{för} \quad t < -\ln q$$

där $q = 1-p$. Enl en sats gäller ju att $\mu = m'(0)$. Så vi deriverar $m(t)$ och får

$$m'(t) = \frac{pe^t(1-qe^t) - pe^t(-qe^t)}{(1-qe^t)^2} = \frac{pe^t}{(1-qe^t)^2} = \frac{m(t)}{1-qe^t}$$

och det följer att

$$m'(0) = \frac{m(0)}{1-q} = \frac{1}{p}$$

vilket skulle visas.

4. Låt U, V vara oberoende $\mathrm{N}(0,1)$-variabler. Definiera

$$\begin{aligned} X_1 &= \mu_1 + \sigma_1 U \\ X_2 &= \mu_2 + \sigma_2(\rho U + \sqrt{1-\rho^2}V) \end{aligned}$$

   (a) Beräkna $\mathrm{Cov}[X_1, X_2]$? (2 p)
   (b) Vilken korrelation har X_1 och X_2? (1 p)
   (c) Vilken bivariat fördelning har paret X_1, X_2? (1 p)

Deluppgift (c) skall ej motiveras.

<u>*Lösning:*</u> Vi ser direkt att $E[X_1] = \mu_1$ och $E[X_2] = \mu_2$, samt att $\mathrm{Var}[X_1] = \sigma_1^2$ och $\mathrm{Var}[X_2] = \sigma_2^2(\rho^2 + 1 - \rho^2) = \sigma_2^2$, så

   (a) $\mathrm{Cov}[X_1, X_2] = E\left[(X_1 - \mu_1)(X_2 - \mu_2)\right] = \sigma_1\sigma_2 E\left[U(\rho U + \sqrt{1-\rho^2}V)\right] = \sigma_1\sigma_2\rho$
   (b) Korrelationen mellan X_1 och X_2 är

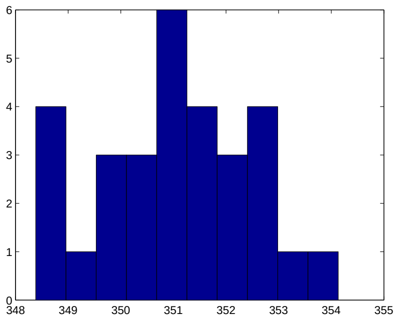
$$\frac{\mathrm{Cov}[X_1, X_2]}{\sqrt{\mathrm{Var}[X_1]\mathrm{Var}[X_2]}} = \frac{\sigma_1\sigma_2\rho}{\sigma_1\sigma_2} = \rho$$

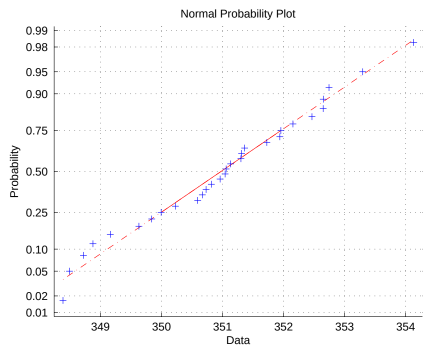
   (c) $\mathrm{N}(\mu_1, \mu_2; \sigma_1, \sigma_2, \rho)$

5. I en partisympatiundersökning ställde man frågan ”Vilket parti skulle du rösta på om det vore val idag?” till 1255 slumpvis utvalda röstberättigade. Låt oss förmoda att de 939 som svarade verkligen skulle rösta och att de övriga 316 inte skulle rösta. Av de 939 som svarade på frågan svarade 299 socialdemokraterna. Punkt- och intervallskatta proportionen socialdemokratiska väljare. (3 p)

<u>*Lösning:*</u> Skattningen av väljarandelen är $\hat{p}_{\mathrm{s}} = \frac{299}{939} \approx 0.318$, och den statistiska felmarginalen är $1.96\sqrt{\hat{p}_{\mathrm{s}}(1-\hat{p}_{\mathrm{s}})/n} \approx 0.030$.

6. I syfte att kalibrera en mätmetod gjordes 30 oberoende mätningar på ett objekt med det kända mätvärdet 350 ppm.[1] Man erhöll medelvärdet 351 ppm och standadavvikelsen 1.45 ppm. Ett histogram och en normalfördelningsplot visas nedan. Är mätmetoden väntevärdesriktig (unbiased)? (4 p)





<u>*Lösning:*</u> Vi kan antingen testa $H_0 : \mu = 350$ mot $H_1 : \mu \neq 350$ på lämplig nivå (5% eller mindre) eller bestämma ett konfidensintervall för μ med motsvarande konfidensgrad. Men vi måste först fråga oss om det är ok att använda normalfördelningsmetoder. På den frågan är svaret ja, ty vi har 30 oberoende observationer och vi ser i histogrammet att deras fördelning nog inte är alltför osymmetrisk och av normalplotten att det faktiskt är rimligt att anta att data är normalfördelade. Centrala gränsvärdessatsen kan vi därför luta oss emot och hävda att medelvärdet av 30 oberoende sådana observationer utan tvekan följer en normalfördelning.

Testet har p-värde $< 1\%$, ty

$$\hat{t} = \frac{351 - 350}{1.45/\sqrt{30}} \approx 3.777 > 2.756 = t_{0.005}(29)$$

och ett 99% konfidensintervall är

$$\mu = 351 \pm 2.756 \cdot 1.45/\sqrt{30} = 351 \pm 0.73$$

Med maximalt 1% felrisk (minst 99% konfidens) kan vi alltså hävda att mätmetoden är skev (ej väntevärdesriktig).

7. J.f.r ovanstående uppgift. Bestäm en övre konfidensgräns för standardavvikelsen. Konfidensen ska vara ca 0.95. (3 p)

<u>*Lösning:*</u> Vi har redan konstaterat att det är ok att använda normalfördelningsmetodik. Vi hämtar den 5:te percentilen i $\chi^2(29)$-fördelningen ur tabell. Den är $\chi_{0.05} = 17.708$. Vi har alltså att sannoliikheten att händelsen

$$\frac{(n-1)s^2}{\sigma^2} \geq 17.708$$

inträffat är ca 0.95. Stoppa in erhållet värde på $(n-1)s^2 = 29 \cdot 1.45^2 \approx 60.973$, stuva om olikheten och konstatera att den övre gränsen i ett ca 95% konfidensintervall för σ är $\sqrt{60.973/17.708} \approx 1.86$.

[1] ppm betyder parts per million

8. J.f.r ovanstående två uppgifter. En månad senare gjorde man ytterligare 30 oberoende mätningar på samma objekt. Nu erhölls medelvärdet 350 ppm och standardavvikelsen 1.32 ppm. Det finns ingen anledning att misstänka att standardavvikelsen nu inte är samma som för en månad sedan, men vad gäller för medelvärdet? Har det ändrats? (5 p)

<u>*Lösning:*</u> Nu kan man inte förkasta $H_0 : \mu = 350$, men detta betyder inte att medelvärdet har ändrats! Vi måste göra en korrekt statistisk analys av differensen $\mu_1 - \mu_2$, alternativt utföra testet av $H_0 : \mu_1 = \mu_2$ mot alternativet $H_1 : \mu_1 \neq \mu_2$. Här betecknar μ_1 det förra medelvärdet och μ_2 det nuvarande. Data är (med självklara beteckningar) $n_1 = n_2 = 30$, $\bar{x}_1 = 351$, $s_1 = 1.45$ och $\bar{x}_2 = 350$ och $s_2 = 1.32$. Vi "poolar" (väger samman) standardavvikelserna till en gemensam skattning med $2 \cdot 29 = 58$ frihetsgrader:

$$s = \sqrt{\frac{1.45^2 + 1.32^2}{2}} \approx 1.387$$

(Att man kan ta medelvärdet av varianserna beror på att de har samma antal frihetsgrader.) Vi ska jämföra t-statistikan

$$\hat{t} = \frac{1}{1.387\sqrt{2/30}} \approx 2.792$$

med kvantiler i $t(58)$-fördelningen. I formelsamlingen finns ej dessa, men man inser att de inte kan skilja sig mycket ifrån de för $t(60)$-fördelningen. Vi ser att $\hat{t} > 2.66 = t_{0.005}(60)$, så vi kan förkasta H_0 på nivån 1%. Vill man känna sig riktigt säker här jämför man istället med $t_{0.005}(40) = 2.704 < \hat{t}$.