Tentamentsskrivning i **Matematisk Statistik med Metoder MVE490**

Tid: 9 januari 2019
Examinatorer: Erik Broman
Tentarond: Ivar Simonsson, tel. 0738027538
Hjälpmedel: Typgodkänd miniräknare, egenhändigt skriven formelsamling om två A4 fram och bak (dvs 4 sidor), samt utdelade tabeller.

Tentamen består av 8 frågor om sammanlagt 50 poäng. Preliminära betygsgränser är satta till:
betyg "3": minst 20 poäng
betyg "4": minst 30 poäng
betyg "5": minst 40 poäng.

OBS! Om inget annat uppges så skall alla lösningar vara väl redovisade och motiverade. Uppgifterna är ej ordnade efter svårighetsgrad

1. En kortlek består av totalt 52 kort varav 4 kort är kungar. Du blandar kortleken och utan att titta ger du det översta kortet till din kompis och nästa kort tar du själv.

   Vi kan anta att kortleken är perfekt blandad.

   (a) Låt A vara händelsen att din kompis får en kung och låt B vara händelsen att du får en kung. Beräkna $P(A \cap B)$. (2p)

   (b) Givet att du har fått en kung, vad är sannolikheten att din kompis också fick en kung? (2p)

   (c) Anta att du fortsätter dela ut alla korten till din kompis och dig själv. Vartannat kort går till din kompis och vartannat till dig själv, så att ni i slutändan har 26 kort var. Din kompis hävdar då att varje kort har samma sannolikhet att vara en kung: $4/52 = 1/13$. Hen hävdar sen att varje kort är oberoende av varandra så "vi gör upprepande, oberoende och likafördelade försök och därför är det totala antalet kungar vi får tillsammans binomialfördelat med parametrar $n = 52$ och $p = 1/13$". Detta stämmer uppenbarligen inte eftersom det är exakt fyra kungar totalt i kortleken och om ni delar ut alla kort så har ni såklart fyra kungar tillsammans. Vad i din kompis resonemang är det som är fel? (2p)

   Lösning:

   (a) Händelsen $A \cap B$ är händelsen att ni båda två får kungar. Din kompis får en kung med sannolikhet $4/52$, och när detta har hänt får du en kung med sannolikhet $3/51$. Vi får alltså att

$$P(A \cap B) = \frac{4}{52} \times \frac{3}{51} = \frac{1}{13 \times 17} = \frac{1}{221}.$$

Om vi inte vill resonera så mycket och tänka på vad händelserna betyder så kan vi använda betingning:

$$P(A \cap B) = P(B|A)P(A) = \frac{3}{51} \times \frac{4}{52} = \frac{1}{221}.$$

(b) Vi vill alltså beräkna $P(A|B)$. Vi använder oss av Bayes regel:

$$P(A|B) = \frac{P(B|A)P(A)}{P(B)}$$

där vi, precis som i (b)-uppgiften, vet att $P(B|A) = 3/51$ och att $P(A) = 1/13$. Vi har även att $P(B) = 1/13$ eftersom all kort i den blandade kortleken har lika stor sannolikhet att vara en kung. Vår alltså att

$$P(A|B) = \frac{3/51 \times 1/13}{1/13} = \frac{3}{51} = \frac{1}{17}.$$

Man kan också argumentera direkt att $P(A|B) = P(B|A)$ av symmetrisjäl.

(c) Din kompis hävdar att korten är oberoende av varandra, vilket inte är sant. I (a)-uppgiften argumenterade vi för att $P(B|A) = 3/51$, men vi har ju att $P(B) = 4/52 = 1/13$. Eftersom $P(B|A) \neq P(B)$ är inte A och B oberoende.

Man kan även argumentera för icke oberoende utan att blanda in sannolikheterna.

2. Familjen Wälitalo består av två föräldrar och fyra barn. När de firar jul tillsammans så köper de varsin julklapp och spelar sen en julklappslek. En i taget så lottas klapparna ut till en slumpmässig familjemedlem men ingen familjemedlem kan få fler än en klapp.

(a) På hur många sätt kan julklapparna delas ut? (2p)

(b) Äldsta dottern Ronja har dragit en rövare och köpt en klapp som hon helst själv vill ha. Hur stor är sannolikheten för detta? (2p)

(c) Föräldrarna Lovis och Mattis har lite högre budget är sina barn. Därför vill de helst att deras klappar hamnar hos barnen. Hur stor är sannolikheten för detta? (2p)

Lösning:

(a) Detta är ett typiskt exempel där fakultet används. Svaret är $6! = 6 \times 5 \times 4 \times 3 \times 2 \times 1 = 720$.

(b) Av symmetrisjäl har Ronja lika stor sannolikhet att få alla klappar, så även sin egen. Därför är sannolikheten att hon får sin egen lika med $1/6$.

Man kan även resonera som följer. Ronja kan bara få sin egen klapp på ett sätt. De andra klapparna kan sen delas ut på 5! sätt. Sannolikheten att Ronja får sin egen klapp är alltså

$$\frac{1 \times 5!}{6!} = \frac{1}{6}.$$

(c) För att detta ska ske så ska både Lovis och Mattis få klappar som barnen har köpt. Eftersom det finns 4 sådana klappar att välja på så finns det 4×3 sätt som detta kan ske på. Efter detta finns det 4! sätt att dela ut resten av klapparna till barnen. Sannolikheten att föräldrarnas klappar hamnar hos barnen är alltså

$$\frac{4 \times 3 \times 4!}{6!} = \frac{4 \times 3}{6 \times 5} = \frac{2}{5} = 0.4.$$

3. Sågföretaget AB sågar virke. Ett vanligt uppdrag de får är att såga brädor som är 3 meter långa. När dessa brädor sågas till är dock inte precisionen exakt. Vi kan anta att längden på en bräda är normalfördelad med väntevärde 3 meter och standardavvikelse 0.02 meter och vi kan anta att längden på brädorna är oberoende.

(a) Vad är sannolikheten att en sumpmässigt vald bräda är mellan 2.98 och 3.05 meter lång? (2p)

(b) 99% av alla brädor har en längd som är längre än c meter. Vad är värdet på c? (2p)

(c) Kalle behöver 10 brädor som är minst 2.97 meter långa. För att vara på den säkra sidan beställer han 12 stycken 3-metersbrädor. Vad är sannolikheten att Kalle får de brädor han behöver? (2p)

Lösning:

(a) Låt X vara längden på den slumpässigt valda brädan. Då har vi att $X \sim N(3, 0.02^2)$. Vi vet då att

$$Z = \frac{X - 3}{0.02}$$

är $N(0,1)$-fördelad. Vi får att

$$P(2.98 \leq X \leq 3.05) = P\left(\frac{2.98 - 3}{0.02} \leq \frac{X - 3}{0.02} \leq \frac{3.05 - 3}{0.02}\right)$$
$$= P(-1 \leq Z \leq 2.5) = P(Z \leq 2.5) - P(Z \leq -1) = 0.9938 - 0.1587 = 0.8351.$$

Här har vi använt normalfördelningstabellen.

(b) Vi söker alltså det värdet c som gör att $P(X > c) = 0.99$. Vi får att

$$\begin{aligned}0.99 = P(X > c) &= P\left(\frac{X-3}{0.02} > \frac{c-3}{0.02}\right)\\ &= P\left(Z > \frac{c-3}{0.02}\right) = 1 - P\left(Z \leq \frac{c-3}{0.02}\right)\end{aligned}$$

vilket ger att $0.01 = P\left(Z \leq \frac{c-3}{0.02}\right)$. Vi tittar i normalfördelningstabellen och ser at

$$\frac{c-3}{0.02} = -2.3263$$

vilket ger att $c = 2.9535$.

(c) Låt Y vara antalet brädor i Kalles parti som är åtminstone 2.97 meter långa. Eftersom längderna på brädorna är oberoende så vet vi att $Y \sim Bin(12, p)$ där p är sannolikheten att en bräda är åtminstone 2.97 meter lång. Vi vill nu beräkna $P(Y \geq 10)$ men för att göra det så måste vi först beräkna $p = P(X \geq 2.97)$. Vi har att

$$\begin{aligned}p =& P(X \geq 2.97) = P\left(\frac{X-3}{0.02} \geq \frac{2.97-3}{0.02}\right)\\ =& P(Z \geq -1.5) = P(Z \leq 1.5).\end{aligned}$$

Vi slår upp detta i normalfördelningstabellen och får att $p = 0.9332$. Vi kan nu göra förljande beräkninar.

$$\begin{aligned}P(Y \geq 10) &= P(Y = 10) + P(Y = 11) + P(Y = 12)\\ &= \binom{12}{10} p^10(1-p)^2 + \binom{12}{11} p^11(1-p)^1 + \binom{12}{12} p^12(1-p)^0\\ &= 0.1475 + 0.3747 + 0.4362 = 0.9584\end{aligned}$$

4. Du har utfört ett experiment och gjort följande oberoende observationer:

3 2 6 6 2 2 5

(a) Ta fram medelvärde, median och typvärde av dina observationer. (2p)

(b) Anta att observationerna kommer från en binomialfördelning med $n = 10$ och okänt p. Innan du utfört experimentet, förklara varför summan av dina observationer också kan ses som en binomialfördelad slumpvariabel. Vad är värdet på parametern n i detta fall? (2p)

(c) Ta fram en punktskattning för p. (2p)

Lösning:

(a) Medelvärde:

$$\bar{x} = \frac{3+2+6+6+2+2+5}{7} = 3.7143.$$

Medianen är den mittersta observationen när vi ordnat dem efter storleksordning.

$$2 \quad 2 \quad 2 \quad 3 \quad 5 \quad 6 \quad 6$$

Vi ser att medianen är 3.

Typvärdet är det värdet som är mest förekommande. Vi har tre observationer av värdet 2, inget annat värde förekommer lika ofta, alltså är typvärdet 2.

0.5 poäng för varje rätt svar, 2 poäng om man har alla tre rätt.

(b) En binomialfördelad variabel kan ses som antalet lyckade försök ur en serie av n oberoende och likafördelade försök där varje försök har sannolikhet p att lyckas. Eftersom vi har 7 oberoende observationer från samma binomialfördelning kan vi se detta som $7 \times 10 = 70$ oberoende försök där vart och ett av försöken har samma sannolikhet p. Det totala antalet lyckade försök är alltså $Bin(70, p)$-fördelat och är lika med summan av våra 7 observationer från $Bin(10, p)$-fördelningen.

(c) Vi skattar sannolikheten p som en proportion, dvs med

$$\hat{p} = \frac{\text{antalet lyckade försök}}{\text{totalt antal försök}} = \frac{3+2+6+6+2+2+5}{70} = 0.37143.$$

5. Avgör om följande påståenden är sanna eller falska. Svara utan motivering! Varje rätt svar ger 1 poäng men varje fel svar ger -1 poäng. Inget svar ger 0 poäng. Totalt ger uppgiften som sämst 0 poäng.

(a) Om X är t-fördelad med 5 frihetsgrader och Y är t-fördelad med 8 frihetsgrader så är väntevärdet av X större än väntevärdet av Y.

(b) Om X är diskret likformigt fördelad på talen $\{0, 1, 2, \ldots, 10\}$ så gäller att $P(X = 5) = 1/10$.

(c) Om X är kontinuerligt likformigt fördelad på intervallet $[0, 10]$ så gäller att $P(X = 5) = 0$.

(d) P-värdet av ett hypotestest är alltid mindre än signifikansnivån.

(e) Om nollhypotesen är sann så är sannolikheten att man inte förkastar den lika med signifikansnivån av testet.

(f) Om A och B är disjunkta händelser så gäller att $P(A \cap B) = P(A)P(B)$.

(g) Du slår en tärning en gång. Om A är händelsen att du slår udda och B är händelsen att du slår en femma så gäller att $P(A \cap B) = P(B)$.

(h) Om $X_1, X_2, \ldots, X_n$ är ett stickprov från en normalfördelning med väntevärde μ så gäller att $\bar{X} = \mu$.

Lösning:

(a) Falskt. Väntevärdet för en t-fördelad variabel är alltid 0, oavsett antalet frihetsgrader.

(b) Falskt. Eftersom det är 11 möjliga värden så gäller att $P(X = 5) = 1/11$.

(c) Sant. Om X har en kontinuerlig fördelning, vilken som helst, så gäller att $P(X = a) = 0$ för alla värden a.

(d) Falskt. Det beror på utfallet av testet.

(e) Falskt. Om nollhypotesen är sann så är sannolikheten att man förkastar den lika med signifikansnivån av testet.

(f) Falskt. Detta gäller generellt inte.

(g) Sant. Om vi har slagit en femma så har vi slagit udda så att slå en femma *och* att slå udda är detsamma som att bara slå en femma, alltså $A \cap B = B$.

(h) Falskt. $\bar{X}$ är en slumpvariabel så vi kan inte säga att $\bar{X} = \mu$. Däremot vet vi att $E(\bar{X}) = \mu$.

6. Ivar gillar att jonglera och han har övat en hel del under jullovet. När han försökte jonglera 5 bollar innan lovet så tog det i genomsnitt 18 sekunder innan han tappade bollarna. För att testa om hans övande under jullovet har gjort honom till en bättre jonglör så provjonglerade han 5 bollar ett antal gånger tills han tappade bollarna. Nedan kan du se hur lång tid det tog innan tappade bollarna i de olika försöken.

29.7	22.7	13.7	26.3	19.2	33.8	5.9	21.4	32.9	15.1	13.8	10.7

(a) Ställ upp hypoteser för att testa om Ivar i genomsnitt klarar av att jonglera 5 bollar längre nu än innan jullovet. (1p)

(b) Ta fram p-värdet för testet. Du kan anta att observationerna kommer från en normalfördelning med varians 10. Kan vi på signifikansnivå $\alpha = 0.05$ förkasta nollhypotesen? (4p)

(c) Om det riktiga väntevärdet för jongleringstiderna hade varit 20 sekunder så hade sannolikheten för typ-II-fel varit 0.29. Vad är då testets styrka? (1p)

Lösning:

(a) Vi vill säkerställa att en förbättring har skett, dvs att Ivar i genomsnitt klarar av att jonglera längre nu än tidigare. Alternativhypotesen är alltså att den nuvarande genomsnittliga tiden μ är större än 18. Vi får alltså att

$$H_0 : \mu \leq 18 \quad \text{och} \quad H_1 : \mu > 18.$$

(b) Eftersom vi kan anta att observationerna kommer från en normalfördelning med känd varians ($\sigma^2 = 10$) så använder vi testvariabeln

$$Z = \frac{\bar{X} - \mu_0}{\sigma/\sqrt{n}}$$

som är $N(0,1)$-fördelad. Vår observerade testvariabeln blir

$$z = \frac{20.4333 - 18}{\sqrt{10}/\sqrt{12}} = 2.6656 \approx 2.67.$$

Vårt p-värde ges då av $P(Z > 2.67)$ där $Z \sim N(0,1)$. Vi tittar i $N(0,1)$-tabellen och ser att $P(0 \leq Z \leq 2.67) = 0.4962$. Vi får till slut att

$$\begin{aligned}\text{p-värde} =&P(Z > 2.67) = 1 - P(Z \leq 2.67)\\ =&1 - P(Z \leq 0) - P(0 \leq Z \leq 2.67)\\ =&1 - 0.5 - 0.4962 = 0.0038.\end{aligned}$$

Eftersom p-värdet är mindre än signifikansnivån $\alpha = 0.05$ så förkastar vi H_0. Ivar har blivit bättre på att jonglera tack var övningen under jullovet!

(c) Styrkan av ett test är sannolikheten att förkasta en falsk nollhypotes, vilket är lika med ett minus sannolikheten att begå ett typ-II-fel. Alltså:

$$\text{Styrka} = 1 - P(\text{typ-II-fel}) = 1 - 0.29 = 0.71.$$

7. Två oberoende stickprov är tagna från två normalfördelade populationer med samma standardavvikelse. Stickprov av storlek 12 från population 1 gav $\bar{x} = 5.729$ och $s_X^2 = 2.202$. Stickprov av storlek 10 från population 2 gav $\bar{y} = 7.086$ och $s_Y^2 = 2.775$.

(a) Låt μ_X och μ_Y beteckna väntevärdet av populationerna. Ta fram ett 95%-igt konfidensintervall för skillnaden $\mu_X - \mu_Y$. (4p)

(b) Tycker du att vi kan dra slutsatsen att det är en skillnad mellan μ_X och μ_Y? Förklara med hjälp av ditt framtagna konfidensintervall. (2p)

Lösning:

(a) Eftersom populationerna är normalfördelade och har samma varians (standardavvikelse) så ges vårt observerade konfidensintervall av

$$\ell = (\bar{x} - \bar{y}) - t_{\frac{\alpha}{2}, n_1+n_2-2}\sqrt{s_p^2\left(\frac{1}{n_1} + \frac{1}{n_2}\right)}$$
$$u = (\bar{x} - \bar{y}) + t_{\frac{\alpha}{2}, n_1+n_2-2}\sqrt{s_p^2\left(\frac{1}{n_1} + \frac{1}{n_2}\right)},$$

där

$$s_p^2 = \frac{(n_1-1)s_X^2 + (n_2-1)s_Y^2}{n_1+n_2-2}.$$

Vi har $n_1 = 12$ och $n_2 = 10$ och vi får s_X^2 och s_Y^2 givet i uppgiften, vilket ger att $s_p^2 = 2.4596$. Vidare tittar vi i t-tabellen med $12+10-2 = 20$ frihetsgrader och ser att $t_{0.015,20} = 2.086$. Vi får till slut att

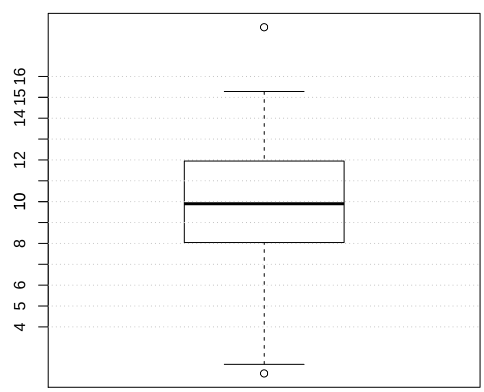
$$\ell = (5.729 - 7.086) - 2.086\sqrt{2.4596\left(\frac{1}{12}+\frac{1}{10}\right)} = -2.7582$$

$$u = (5.729 - 7.086) + 2.086\sqrt{2.4596\left(\frac{1}{12}+\frac{1}{10}\right)} = 0.0433.$$

Vårt observerade 95%-iga konfidensintervall är alltså $[-2.7582; 0.0433]$.

(b) Ingen skillnad mellan väntevärdena μ_X och μ_Y representeras av att $\mu_X - \mu_Y = 0$. Eftersom vårt framtagna konfidensintervall innehåller 0 så borde vi inte helt utesluta det inte är någon skillnad.

8. Figur 1 visar en boxplot över ett dataset bestående av 100 observationer.

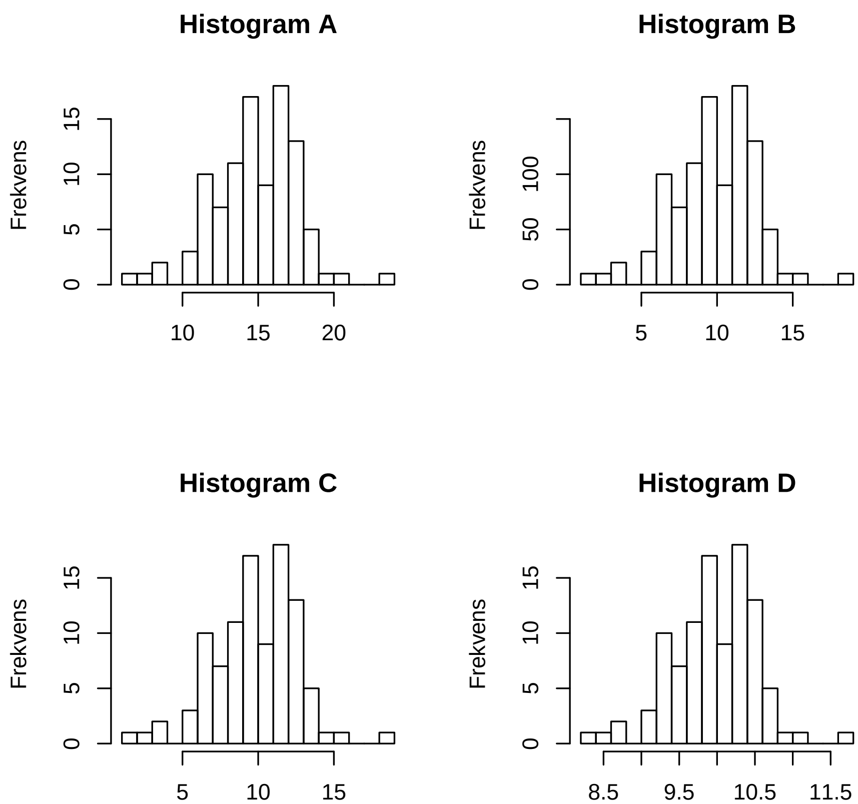


Figur 1: En boxplot över ett dataset.

(a) Ta fram medianen och kvartilavståndet för datasetet. (2p)

(b) I Figur 2 kan du se fyra olika histogram. Exakt ett av dessa är ett histogram över samma dataset som gett upphov till boxploten i Figur 1. Vilket? (4p)

Tips: Använd uteslutningsmetoden, dvs argumentera vilka av histogrammen i Figur 2 som *inte* kan höra ihop med boxploten i Figur 1.



Figur 2: Fyra olika histogram.

Lösning:

(a) Medianen ges av den mittersta horisontella linjen och ser ut att vara strax under 10. Kvartilavståndet ges av den övre kvartilen minus den nedre kvartilen. Dessa kvartiler ges av den övre respektive nedre gränsen av lådan, vilket ser ut att vara ungefär 12 och 8. Alltså har vi: $IQR = Q_U - Q_L \approx 12 - 8 = 4$.

(b) Vi utesluter histogram A eftersom dess median är betydligt större än 10. Vi utesluter histogram B eftersom frekvenserna är alldeles för höga, dvs dess dataset innehåller betydligt fler observationer än 100.

Vi utesluter histogram D eftersom dess varians är för liten; alla observationer ligger mellan Q_L och Q_U från boxploten. Vi drar således slutsatsen att histogram C har skapats från samma dataset som gett upphov till boxploten i Figur 1.