Tentamentsskrivning i **Matematisk Statistik med Metoder MVE490**

Tid: 22 augusti 2018
Examinatorer: Erik Broman.
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”: 20 till 29
betyg “4”: 30 till 39 poäng
betyg “5”: 40 eller fler poäng.

OBS! Alla lösningar skall vara väl redovisade och motiverade. Talen är ej ordnade efter svårighetsgrad

1. Takpannor-r-us tillverkar takpannor (!). Urban skall lägga om sitt tak och köper därför takpannor från ovan angivna företag. En takpanna är defekt med sannolikhet 0.007 och Urban köper in 500 stycken. Ange svaret i decimalform.

   (a) Varje paket med takpannor innehåller 10 stycken. Vad är sannolikheten att Urban får minst en defekt takpanna i sitt första paket? (2p)

   (b) Vad är sannolikheten att Urban får minst 8 men inte fler än 10 defekta takpannor (av de sammanlagt 500 han köper)? (3p)

   (c) Om Urban istället köper in 10000 takpannor, vad är sannolikheten att Urban får mellan 56 och 84 defekta takpannor? (3p)

   Lösning.

   (a) Låt X =antalet defekta takpannor så att $X \sim$Bin$(10, 0.007)$. Vi skall då beräkna

   $$\mathbb{P}(X \geq 1) = 1 - \mathbb{P}(X = 0) = 1 - (1 - 0.007)^{10} \approx 0.0678.$$

   (b) Här bör vi använda Poissonapproximation. Vi har nu att $X \sim$Poi(3.5) och ser att

   $$\begin{aligned}\mathbb{P}(8 \leq X \leq 10) &= \mathbb{P}(X = 8) + \mathbb{P}(X = 9) + \mathbb{P}(X = 10) \\ &= e^{-3.5}\left(\frac{3.5^8}{8!} + \frac{3.5^9}{9!} + \frac{3.5^{10}}{10!}\right) \approx 0.0257.\end{aligned}$$

(c) Vi använder nu normalapproximation. Vi har därför att $X \sim$N$(70, 69.51)$. Den sökta sannolikheten blir då

$$\begin{aligned}\mathbb{P}(56 \leq X \leq 84) &= \mathbb{P}\left(\frac{56-70}{\sqrt{69.15}} \leq \frac{X-70}{\sqrt{69.15}} \leq \frac{84-70}{\sqrt{69.15}}\right)\\ &\approx \mathbb{P}(-1.6836 \leq Z \leq 1.6836) = \mathbb{P}(Z \leq 1.6836) - \mathbb{P}(Z \leq -1.6836)\\ &= \mathbb{P}(Z \leq 1.6836) - \mathbb{P}(Z \geq 1.6836) = 2\mathbb{P}(Z \leq 1.6836) - 1\\ &\approx 2 * 0.954 - 1 \approx 0.908\end{aligned}$$

2. En vanlig kortlek består av 52 kort (13 hjärter, 13 spader, 13 klöver och 13 ruter). Blanca Vlasic får fem kort (inga byten).

(a) Vad är sannolikheten att Blanca får fem stycken hjärter? (2p)

(b) Vad är sannolikheten att Blanca får fyra hjärter och en spader? (2p)

(c) Vad är sannolikheten att Blanca får minst fyra hjärter? (2p)

Lösning.

(a) Sannolikheten att det första kortet är en hjärter är 13/52, sannolikheten att det andra kortet är en hjärter är då 12/51 osv. Svaret blir alltså

$$\frac{13 * 12 * 11 * 10 * 9}{52 * 51 * 50 * 49 * 48} = \frac{154440}{311875200} \approx 0.0004952$$

(b) Sannolikheten att Blanca får först fyra hjärter och sist en spader blir som ovan

$$\frac{13 * 12 * 11 * 10}{52 * 51 * 50 * 49}\frac{13}{48} = \frac{223080}{311875200} \approx 0.000715.$$

Men, vi får inte glömma bort att spadern kan antingen vara första, andra, tredje, fjärde eller femte kortet som hon får. Svaret blir alltså

$$\frac{223080}{311875200} * 5 \approx 0.0036.$$

(c) För att Blanca skall få minst fyra hjärter tar vi svaret i a och lägger till tre gånger svaret i b. Då får vi

$$\frac{154440}{311875200} + \frac{223080}{311875200} * 3 = 0.0004952 + 0.0036 * 3 \approx 0.0112.$$

3. Låt X vara en slumpvariabel med sannolikhetsfunktion

$$\mathbb{P}(X = -1) = 0.5,\ \mathbb{P}(X = 2) = 0.3,\ \mathbb{P}(X = 5) = 0.2.$$

(a) Beräkna väntevärdet och variansen av X. (2p)

(b) Beräkna väntevärdet och variansen av X^2. (2p)

(c) Låt $X_1, X_2, \ldots, X_{10}$ vara oberoende och alla ha samma fördelning som X. Beräkna väntevärdet av $X_1 + X_2 + \cdots + X_{10}$. (2p)

Lösning:

(a) Vi har att

$$\mathbb{E}[X] = -1 * 0.5 + 2 * 0.3 + 5 * 0.2 = 1.1$$

och att

$$\mathbb{E}[X^2] = (-1)^2 * 0.5 + 2^2 * 0.3 + 5^2 * 0.2 = 6.7,$$

så att

$$\mathrm{Var}(X) = \mathbb{E}[X^2] - \mathbb{E}[X]^2 = 6.7 - 1.1^2 = 5.49.$$

(b) Vi har redan beräknat $\mathbb{E}[X^2]$ och vi har att

$$\mathbb{E}[X^4] = (-1)^4 * 0.5 + 2^4 * 0.3 + 5^4 * 0.2 = 130.3.$$

Detta ger att

$$\mathrm{Var}(X) = \mathbb{E}[X^4] - \mathbb{E}[X^2]^2 = 130.3 - 6.7^2 = 85.41.$$

(c) Vi använder linearitet och ser att

$$\mathbb{E}[X_1 + X_2 + \cdots + X_{10}] = \mathbb{E}[X_1] + \cdots + \mathbb{E}[X_{10}] = 10 * \mathbb{E}[X] = 11.$$

4. Armand har två väskor med stavar. I väska nummer ett har han tre gula och tre röda stavar och i väska nummer två har han fyra gula och två röda stavar.

Armand väljer först en väska och sedan drar han två stavar ur denna väska. Vi antar att väskorna och stavarna alla har samma sannolikhet att väljas/dras.

(a) Vad är sannolikheten att Armand får två gula stavar? (3p)

(b) Givet att Armand får två gula stavar, vad är sannolikheten att han valde väska nummer ett? (2p)

Lösning:

(a) Låt V1,V2 vara händelserna att Armand väljer väska nummer ett respektive väska nummer två. Låt sedan GG vara händelsen att han får två gula stavar. Vi har att

$$\mathbb{P}(GG) = \mathbb{P}(GG|V_1)\mathbb{P}(V_1) + \mathbb{P}(GG|V_2)\mathbb{P}(V_2) = \frac{3 * 2}{6 * 5} * \frac{1}{2} + \frac{4 * 3}{6 * 5} * \frac{1}{2} = \frac{18}{60} = \frac{3}{10}.$$

(b) Vi söker nu $\mathbb{P}(V_1|GG)$ och använder definitionen av betingad sannolikhet för att få

$$\mathbb{P}(V_1|GG) = \frac{\mathbb{P}(GG \cap V_1)}{\mathbb{P}(GG)} = \frac{6/60}{3/10} = \frac{1}{3}.$$

5. Du är intresserad av en proportion p och har samlat in ett stickprov från den relevanta populationen. Det framtagna konfidensintervallet för p blev $[0.25; 0.47]$ (signifikansnivå $\alpha = 0.05$).

   (a) Vad är värdet på punktskattningen $\hat{p}$? (1p)

   (b) Hur stort är stickprovet? (3p)

   (c) Om vi antar att $\hat{p}$ inte ändras, hur stort måste stickprovet vara för att längden på konfidensintervallet ska vara högst 0.16? (2p)

   Lösning:

   (a) Eftersom konfidensintervall för en proportion är symmetriska kring $\hat{p}$ vet vi att $\hat{p}$ ligger mitt emellan ℓ och u. Alltså: $\hat{p} = 0.36$.

   (b) Kom ihåg att konfidensintervallet ges av:

   $$\ell = \hat{p} - z_{\alpha/2}\sqrt{\frac{\hat{p}(1-\hat{p})}{n}}, \qquad u = \hat{p} + z_{\alpha/2}\sqrt{\frac{\hat{p}(1-\hat{p})}{n}}$$

   Eftersom vi har listat ut vad $\hat{p}$ är så vet vi allt förutom n. Vi vet t. ex. att $u = 0.47$, att $\hat{p} = 0.36$ och att $z_{\alpha/2} = 1.96$. Om vi sätter in dessa tal i formeln för u och löser ekvationen får vi att

   $$n = 1.96^2 * \frac{0.36 * 0.64}{0.11^2} = 73.14914 \approx 73. \tag{1}$$

   Samma härledning går att göra med formeln för ℓ istället för u, resultatet blir samma.

   (c) Här använder vi samma resonemang som i uppgift (b). Vi använder även ekvation (1) fast med $0.16/2 = 0.08$ istället för 0.11. Vi får alltså att

   $$n = 1.96^2 * \frac{0.36 * 0.64}{0.08^2} = 138.2976.$$

   Vårt stickprov måste alltså vara åtminstonde av storleken 139.

6. Ett renhållningsbolag hämtar sopor i ett lägenhetskomplex en gång i veckan. Bostadsföreningen och renhållningbolaget har en överenskommelse om att det ska i genomsnitt hämtas max 1100 kg sopor varje vecka. Renhållningsarbetarna misstänker dock att de faktiskt hämtar mer. Om så är fallet vill renhållningsbolaget ta mer betalt av bostadsföreningen. För att undersöka om det hämtas för mycket sopor så väger renhållningsarbetarna soporna som hämtas varje vecka under ett helt år (52 veckor på ett år). Stickprovsmedelvärdet blev $\bar{x} = 1128$ och stickprovsstandardavvikelsen blev $s = 126$.

   (a) Hjälp renhållningsarbetarna att utföra ett hypotestest, använd signifikansnivå $\alpha = 0.05$. Kan vi dra slutsatsen att de hämtar för mycket sopor i genomsnitt? (3p)

(b) Vilken typ av fel riskerar du att ha begått? (1p)

(c) Beräkna p-värdet för testet. Om du hade använt signifikansnivå 0.1 istället för 0.05, hade du då dragit samma slutsats i (a)-uppgiften? Svara med hjälp av p-värdet. (3p)

Lösning:

(a) Låt μ beteckna den genomsnittliga mängd sopor som hämtas från lägenhetskomplexet varje vecka. Hypoteserna blir då

$$H_0 : \mu = 1100 \quad \text{mot} \quad H_1 : \mu > 1100.$$

Ensidigt test eftersom vi bara är intresserade om det hämtas för mycket sopor, inte för lite. Stickprovet är stort nog för att vi ska kunna använda normalapproximationen för medelvärdet. Vi använder oss därför av teststatistikan

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

Vi stoppar in våra observerade värden $\bar{x} = 1128$, $s = 126$ och $n = 52$, samt $\mu_0 = 1100$. Vår observerade teststatistika blir då

$$z = \frac{1128 - 1100}{126/\sqrt{52}} = 1.602.$$

Vi jämför med normalfördelningens tabellvärden, dvs det kritiska värdet $z_{0.05} = 1.645$. Eftersom den observerade teststatistikan är mindre (närmre 0) än det kritiska värdet så kan vi inte förkasta H_0.

På signifikansnivå $\alpha = 0.05$ kan vi alltså **inte** dra slutsatsen att renhållningsarbetarna hämtar mer än 1100 kg sopor genomsnitt.

(b) Vi riskerar att ha gjort ett typ-II-fel, dvs vi riskerar att inte ha förkastat en falsk nollhypotes.

(c) Den observerade teststatistikan var 1.602. Från normalfördelningstabellen får vi då fram p-värdet: 1-0.9474 = 0.0526. P-värdet är alltså väldigt nära signifikansnivån 0.05. Om signifikansnivån hade varit 0.1 så hade vi förkastat H_0 eftersom $0.1 > 0.0526$.

7. Sveriges och Norges utbidningsdepartement vill undersöka mattekunskaperna hos de bägge ländernas högstadieelever. Finns det någon skillnad i kunskapsnivån? För att testa detta skapas ett prov och man väljer slumpmässigt ut 64 svenska och 54 norska högstadieklasser som får göra provet. Totalt gjorde 1768 svenska och 1356 norska elever provet. Vi betecknar de svenska testresultaten med $x_1, x_2, \ldots$ och de norska tetresultaten med $y_1, y_2, \ldots$. Resultatet kan ses i tabellen nedan.

Sverige	Norge
$\sum x_i = 10862.8$	$\sum y_i = 7920.6$
$\sum x_i^2 = 70792.76$	$\sum y_i^2 = 49325.96$

Formulera hypoteser och gör ett hypotestest på signifikansnivå $\alpha = 0.05$. Är det någon skillnad på högstadieelevernas mattekunskaper i de bägge länderna? (4p)

Lösning: Om vi låter μ_X och μ_Y beteckna de riktiga väntevärdena så borde hypteserna se ut så här:

$$H_0 : \mu_X - \mu_Y = 0 \quad \text{mot} \quad H_1 : \mu_X - \mu_Y \neq 0.$$

Alternativt:

$$H_0 : \mu_X = \mu_Y \quad \text{mot} \quad H_1 : \mu_X \neq \mu_Y.$$

Stickprovsstorlekarna $n_x = 1768$ och $n_y = 1356$ gott och väl stora nog för att vi ska kunna använda oss av normalfördelningen. Vi använder oss därför av teststatistikan

$$Z = \frac{\bar{X} - \bar{Y}}{\sigma_{\bar{X}-\bar{Y}}}.$$

Vi har $\bar{x} = 10862.8/1768 = 6.144$ och $\bar{y} = 7920.6/1356 = 5.841$ och vi skattar $\sigma_{\bar{X}-\bar{Y}}$ med

$$\sqrt{\frac{s_X^2}{n_X} + \frac{s_Y^2}{n_Y}}.$$

Vi tar fram $s_X^2 = 2.292$ och $s_Y^2 = 2.258$ från tabellvärdena och får då att $\sigma_{\bar{X}-\bar{Y}} = 0.0544$. Vår observerade teststatistika blir alltså

$$\frac{6.144 - 5.841}{0.0544} = 5.569.$$

Normalfördelningstabellen ger oss de kritiska värdena ± 1.96 och vår teststatistika ligger utanför dessa så vi förkastar nollhypotesen. Vi drar slutsatsen att det är en skillnad i mattekunskaperna hos högstadieelever i Sverige och Norge.

8. I Figur 1 nedan kan du se tre olika histogram från tre olika stickprov. 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. Tänk igenom varje svar ordentligt!

   (a) Medelvärdet av stickprov 3 är märkbart större än medelvärdet av stickprov 1.

   (b) Variansen i stickprov 3 är märkbart större än variansen i stickprov 1.

   (c) Kvartilavståndet i stickprov 3 är märkbart större än kvartilavståndet i stickprov 1.

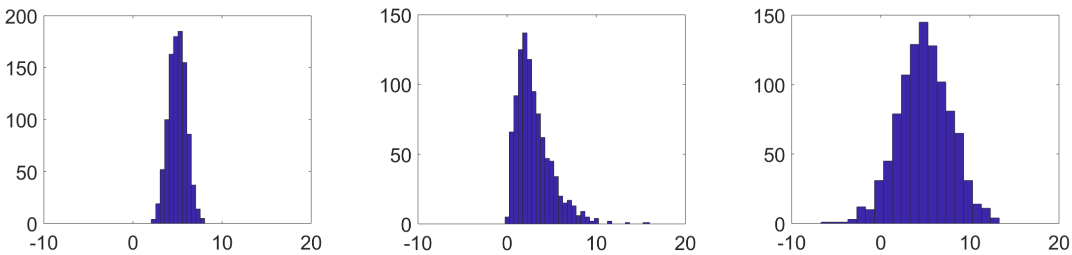
   (d) Stickprov 1 ser ut att komma från en normalfördelning.

(e) Stickprov 2 ser ut att komma från en normalfördelning.

(f) Stickprov 3 ser ut att komma från en normalfördelning.

(g) Alla tre stickproven ser ut att vara skeva.

(h) I allmänhet så är histogram ett lämpligt sätt att presentera både kvalitativ och kvantitativ data.



(a) Stickprov 1 **(b)** Stickprov 2 **(c)** Stickprov 3

Figure 1: Histogram över tre olika stickprov

Lösning:

(a) Falskt. Både stickprov 1 och 3 ser ut att ha ett medelvärde på ungefär 5.

(b) Sant. Det är mycket större spridning i stickprov 3, dvs större varians.

(c) Sant. Kvartilavstånd är, precis som variansen, ett mått på spridningen i ett stickprov.

(d) Sant. Histogrammet för stickprov 1 ser typiskt ut för en normalfördelning.

(e) Falskt. Histogrammet för stickprov 2 är skevt, vilket normalfördelningen inte är.

(f) Sant. Samma här som för stickprov 1.

(g) Falskt. Det är bara stickprov 2 som är skevt, de två andra är väldigt symmetriska.

(h) Falskt. Histogram är inte lämpligt för kvalitativ data.