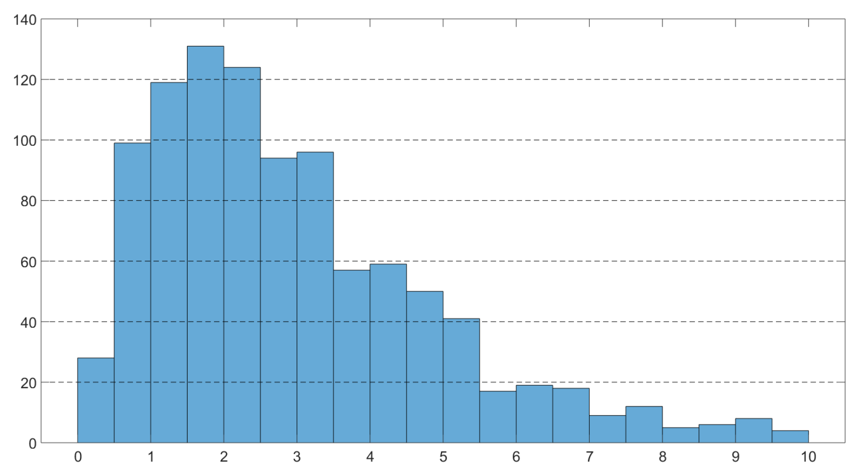
Tentamentsskrivning i **Matematisk Statistik med Metoder MVE490**

Tid: 1 november 2018
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": 20 till 29
betyg "4": 30 till 39 poäng
betyg "5": 40 eller fler poäng.

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

1. I figuren nedan kan du se ett histogram över ett stickprov av storlek $n = 1000$.
   (a) Ungefär hur många observationer ligger i intervallet $[4; 4.5]$? (2p)
   (b) Ungefär hur många procent av stickprovet ligger i intervallet $[0.5; 1.5]$? (2p)
   (c) Skulle du säga att stickprovet är skevt? (1p)
   (d) Vad tror du om stickprovets medelvärde och median? Är medelvärdet eller medianen störst, eller är de ungefär lika stora? (1p)



Lösning:

(a) Den stapeln ser ut att vara just lite lägre än 60-strecket, så lite mindre än 60 obervationer ligger i det intervallet.

(b) Detta intervall täcker två staplar, så vi vill inkludera både stapeln för intervallet $[0.5; 1]$ och stapeln för intervallet $[1; 1.5]$. Det ser ut att vara knappt 100 observationer i det första intervallet och knappt 120 i det andra, så totalt knappt $100 + 120 = 220$. Eftersom vi har totalt 1000 observationer så blir detta ca 22%.

(c) Det ser ut som att stickprovet är skevt. Vi ser inga 'extrema' observationer åt vänster men väldigt många observationer åt höger ligger långt ifrån den stora delen av stickprovet.

(d) När stickprovet är skevt åt detta håll så kommer medelvärdet att vara betydligt större än medianen.

2. X är en diskret slumpvariabel vars sannolikhetsfunktion ges av följande tabell.

k	0	1	2	3
$P(X = k)$	4c	0.25	0.5	c

(a) Bestäm värdet på konstanten c. (2p)

(b) Beräkna väntevärde och varians av X. (2p)

(c) Låt $Y = X/2 + 1$. Ta fram sannolikhetsfunktionen för Y och presentera den i en tabell likt den ovan. (2p)

Lösning:

(a) Ett villkor för att det ska vara en sannolikhetsfunktion är att $\sum P(X = k) = 1$. I vårt fall har vi att

$$\sum_{k=0}^{3} P(X = k) = 4c + 0.25 + 0.5 + c = 5c + 0.75.$$

Om detta ska vara lika med 1 så får vi att

$$1 = 5c + 0.75 \Rightarrow 5c = 0.25 \Rightarrow c = 0.05.$$

(b) Vi använder definitionen av väntevärde samt tabellen ovan med $c = 0.05$:

$$\begin{aligned}E(X) =& \sum_{k=0}^{3} kP(X = k)\\ =&0 \times P(X = 0) + 1 \times P(X = 1) + 2 \times P(X = 2) + 3 \times P(X = 3)\\ =&0 \times 0.2 + 1 \times 0.25 + 2 \times 0.5 + 3 \times 0.05\\ =&0 + 0.25 + 1 + 0.15 = 1.4.\end{aligned}$$

Vi har även att

$$\begin{aligned}E(X^2) =& \sum_{k=0}^{3} k^2 P(X=k) \\ =& 0^2 \times P(X=0) + 1^2 \times P(X=1) + 2^2 \times P(X=2) + 3^2 \times P(X=3) \\ =& 0 \times 0.2 + 1 \times 0.25 + 4 \times 0.5 + 9 \times 0.05 \\ =& 0 + 0.25 + 2 + 0.45 = 2.7\end{aligned}$$

vilket vi använder oss av för att beräkna variansen:

$$Var(X) = E(X^2) - E(X)^2 = 2.7 - 1.4^2 = 0.74.$$

(c) Eftersom X kan anta värdena 0,1,2 och 3 så kan Y anta värdena 1, 1.5, 2 och 2.5. Sannolikheterna för vart och ett för dessa värden är densamma som motsvarande sannolihet för X. Vi får då följande tabell.

k	1	1.5	2	2.5
$P(Y=k)$	0.2	0.25	0.5	0.05

3. Du slår två tärningar, en röd och en grön. A är händelsen att den röda tärningen blir 5, B är händelsen att den gröna tärningen blir udda, C är händelsen att summan av tärningarna blir udda, D är händelsen att summan av tärningarna blir 7 och E är händelsen $A \cap B$.

(a) Är B och C oberoende? (2p)

(b) Är A och D oberoende? (2p)

(c) Beräkna $P(C \cup E)$. (2p)

Lösning:

(a) Här är det nog lättast att testa om $P(C|B) = P(C)$. $P(C|B)$ är sannolikheten att summan av tärningarna blir udda givet att den gröna tärningen blir udda, vilket är lika med sannolikheten att den röda tärningen blir jämn, vilket är lika med 0.5.

När vi slår två tärningar så har vi 36 olika och lika sannolika utfall, 18 av dessa leder till ett udda resultat. Alltså har vi att

$$P(C) = \frac{18}{36} = \frac{1}{2} = 0.5$$

alltså är $P(C|B) = P(C)$ och således är B och C oberoende.

(b) Här är det nog lättast att testa om $P(D|A) = P(D)$. $P(D|A)$ är sannolikheten att summan av tärningarna blir sju givet att den röda tärningen blir 5. Om den röda tärningen blir 5 så måste den gröna tärningen bli 2 för att summan ska bli 7, detta händer med sannolikhet 1/6. Vi har alltså att $P(D|A) = 1/6$.

För att beräkna $P(D)$ kan vi börja med att lista alla utfall i D. Vi har att

$$D = \{(1,6),(2,5),(3,4),(4,3),(5,2),(6,1)\}$$

så antalet utfall i D är 6. Vi får alltså att

$$P(D) = \frac{\text{antal utfall i } D}{\text{antal utfall totalt}} = \frac{6}{36} = \frac{1}{6}.$$

Vi har alltså att $P(D|A) = P(D)$ och således är A och D oberoende.

(c) Så händelsen E är händelsen att både A och B händer, dvs händelsen att den röda tärningen blir 5 och att den gröna tärningen blir udda. Vi listar alla utfall i E:

$$E = \{(5,1),(5,3),(5,5)\}.$$

Eftersom C är händelsen att summan av tärningarna är udda och inga utfall i E ger en udda summa så vet vi att C och E är disjunkta och vi vill alltså använda oss av formeln

$$P(C \cup E) = P(C) + P(E).$$

Vi räknade ut $P(C)$ i (a)-uppgiften: $P(C) = 0.5$. Det återstår att beräkna $P(E)$. Vi ser att det finns 3 utfall i E, så vi får att

$$P(E) = \frac{\text{antal utfall i } D}{\text{antal utfall totalt}} = \frac{3}{36} = \frac{1}{12}$$

och således

$$P(C \cup E) = P(C) + P(E) = \frac{1}{2} + \frac{1}{12} = \frac{6+1}{12} = \frac{7}{12} \approx 0.5833.$$

4. 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) En diskret slumpvariabel kan bara anta heltalsvärden.

   (b) Om X är en slumpvariabel så kan $E(X)$ inte vara negativ.

   (c) Om X är en slumpvariabel så kan $Var(X)$ inte vara negativ.

   (d) Om A och B är disjunkta händelser så gäller att $P(A|B) = 0$.

   (e) Låt $X_1, X_2, \ldots, X_{n_1}$ och $Y_1, Y_2, \ldots, Y_{n_2}$ vara två stickprov från samma fördelning. Om n_1 är större än n_2 så är variansen av $\bar{X}$ större än variansen av $\bar{Y}$.

   (f) Ju högre signifikansnivå desto längre blir konfidensintervallet.

   (g) Ju större stickprov desto kortare blir konfidensintervallet.

(h) Om p-värdet är mindre än signifikansnivån så förkastar vi nollhypotesen på denna signifikansnivå.

Lösning:

(a) Falskt. En diskret slumpvariabel kan anta vilka värden som helst, men bara ett uppräkneligt antal.

(b) Falskt. Väntevärdet av en slumvariabel kan mycket väl vara negativt.

(c) Sant. Variansen av en slumpvariabel kan aldrig vara negativ.

(d) Sant. Om A och B är disjunkta så gäller att $P(A \cap B) = 0$ och således $P(A|B) = P(A \cap B)/P(B) = 0/P(B) = 0$.

(e) Falskt. Vi har att $Var(\bar{X}) = \sigma^2/n_1$ och att $Var(\bar{Y}) = \sigma^2/n_2$ där σ^2 är variansen i fördelningen. Om n_1 är större än n_2 så har vi alltså att variansen av $\bar{X}$ är mindre än variansen av $\bar{Y}$.

(f) Falskt. Ju högre signifikansnivå desto mindre blir värdet $z_{\frac{\alpha}{2}}$ och desto kortare blir konfidensintervallet.

(g) Sant. Ju större stickprov, dvs ju större värde på n, desto mindre blir variansen för punktskattaren och desto kortare blir konfidensintervallet.

(h) Sant. Det är precis detta som är regeln.

5. För att utföra en kvalitetskontroll av ett parti lampor tar man slumpmässigt ut 9 lampor och studerar hur många av dessa som är defekta. För att kontrollen ska bli godkänd får högst två av de uttagna lamporna ha någon typ av defekt. Man vet att under produktionen av lamporna får ungefär 10% av alla lampor någon typ av defekt. Antag att defekterna är oberoende av varandra.

(a) Låt X vara en slumpvariabel för antalet defekta lampor i kontrollen. Vilken fördelning har X? (2p)

(b) Vad är sannolikheten att ett parti kommer bli godkänd? (2p)

(c) En dag när Vilma ska göra en kontroll väljer hon ut 9 lampor som hon ska undersöka och konstaterar att partiet blir godkänt. Dock efter undersökningen råkar hon tappa en icke defekt lampa så att den går sönder.

Om vi förutsätter att hon fortfarande studerar samma 9 lampor, vad är då sannolikheten att partiet nu inte blir godkänt? (2p)

Tips: Hur måste de övriga lamporna ha varit för att olyckan ska ändra kontrollresultatet?

Lösning:

(a) Eftersom vi studerar enheter som fungerar med en viss sannolikhet och är oberoende av varandra blir $X \sim Bin(9, 0.1)$

(b) Vi söker sannolikheten att ett parti blir godkänt. Detta motsvarar att max 2 lampor har en defekt. Sannolikheten för detta kan skrivas som

$$P(X \leq 2) = \binom{9}{0} 0.1^0 (1-0.1)^9 + \binom{9}{1} 0.1^1 (1-0.1)^8 + \binom{9}{2} 0.1^2 (1-0.1)^7 = 0.947$$

(c) Vi vet att partiet först var godkänt vilket innebär att det max kan ha funnits 2 lampor av de 9 studerande som hade en defekt. För att partiet inte länge ska vara godkänt efter att Vilma har råkat förstöra en lampa så måste antalet lampor nu vara större än tre. Detta betyder att av de övriga 8 lamporna i partier så måste 2 av dem redan ha en defekt. Sannolikheten för detta kan skrivas som $P(Y = 2)$ där $Y \sim Bin(8, 0.1)$.

$$P(Y = 2) = \binom{8}{2} 0.1^2 (1 - 0.1)^{8-2} = 0.149$$

6. I en stor simklubb pågår diskussioner om eventuellt inköp av så kallade hajdräkter. För att spendera pengar på de nya hajdräkterna vill klubben försäkra sig om att simtiderna hos medlemmarna förbättras nog mycket. De lånar därför in hajdräkter till 8 slumpmässigt utvalda medlemmar och utför tester på distansen 50 frisim. Om den genomsnittliga tidsförbättringen är större än 0.5 sekunder så bestämmer de sig för att köpa in hajdräkter till alla medlemmar. De 8 utvalda medlemmarna simmar varsitt lopp utan hajdräkt och varsitt lopp med hajdräkt. Resultaten, dvs tiderna i sekunder, blev följande

Medlem	1	2	3	4	5	6	7	8
Utan hajdräkt	28.21	29.13	28.23	26.19	27.06	27.75	30.96	26.28
Med hajdräkt	27.85	28.62	27.56	25.27	26.37	27.12	29.92	25.74
Skillnad	0.36	0.51	0.67	0.92	0.69	0.63	1.04	0.54

Vi kan anta att tiderna är normalfördelade.

Formulera och testa rätt hypoteser för att säkerställa att den genomsnittliga tidsförbättringen är stor nog för att inköpet ska göras. Använd signifikansnivå $\alpha = 0.05$. Är resultaten tillräckligt bra för att vi ska vilja köpa in hajdräkter till alla medlemmar? (6p)

Lösning: Låt μ_d beteckna medelvärdet för hur mycket snabbare medlemmarna simmar med hajdräkt. Vi vill testa hypteserna

$$H_0 : \mu_d \leq 0.5 \qquad \text{mot} \qquad H_1 : \mu_d > 0.5.$$

Eftersom tiderna är normalfördelade och stickprovet är litet så använder vi teststatistikan

$$T = \frac{\bar{D} - D_0}{s_d/\sqrt{n}}$$

som är t-fördelad med $n-1$ frihetsgrader. Vi vet att $n = 8$ och att $D_0 = 0.5$ så vi måste ta reda på $\bar{D}$ och s_d. Vi beräknar $\bar{D}$ genom att ta medelvärdet av tidsskillnaderna, dvs medelvärdet av den nedersta raden från tabellen ovan: $\bar{D} = \frac{1}{8}\sum D_k = 0.67$. Vi tar sen fram s_d för dessa värden:

$$s_d = \sqrt{s_d^2} = \sqrt{\frac{1}{n-1}\sum(\bar{D} - D_k)^2} = \sqrt{\frac{1}{7}\sum(0.67 - D_k)^2} = 0.2204.$$

Vi kan nu beräkna vår observerade teststatistika:

$$t = \frac{0.67 - 0.5}{0.2204/\sqrt{8}} = 2.1817.$$

Vi tittar i T-tabellen och ser att det kritiska värdet för $\alpha = 0.05$ och antalet frihetsgrader 7 är 1.895. Eftersom vår observerade teststatistika 2.1817 är större än (ligger längre ifrån 0 än) det kritiska värdet 1.895 så förkastar vi H_0 på signifikansnivå $\alpha = 0.05$.

Vi drar slutsatsen att den genomsnittliga tidsförbättringen är större än 0.5 sekunder och klubben borde alltså köpa in hajdräkter till alla medlemmar.

7. Låt $X_1, X_2, \ldots, X_n$ vara ett stickprov från en normalfördelning med okänt väntevärde μ och okänd varians σ^2.

   (a) Låt $Y = 0.4 \times X_1 + 0.6 \times X_2$. Är Y en väntevärdesriktig skattare för μ? (2p)

   (b) Du har gjort en undersökning och samlat in ett stickprov av storlek $n = 10$ från fördelningen. Ditt observerade stickprov blev:

4.45	3.56	5.20	3.51	16.94	-2.83	-6.27	-2.70	-2.47	0.83

   Gör ett 90%-igt konfidensintervall för μ. (4p)

Lösning:

(a) För att Y ska vara en väntevärdesriktig skattare för μ så måste vi har att $E(Y) = \mu$. Vi använder egenskaper för väntevärdet och får att

$$\begin{aligned} E(Y) =& E(0.4 \times X_1 + 0.6 \times X_2) \\ =& 0.4E(X_1) + 0.6E(X_2) \\ =& 0.4\mu + 0.6\mu = \mu. \end{aligned}$$

Ja, Y är en väntevärdesriktig skattare för μ.

(b) Från våra observationer tar vi fram vårt observerade stickprovsmedelvärde $\bar{x} = \sum_{k=1}^{10} x_k = 2.02$ som kommer vara mitten av vårt konfidensintervall. Eftersom stickprovet kommer från en normalfördelning och n är relativt litet så vill vi använda formlerna

$$\ell = \bar{x} - t_{\alpha/2,n-1}\frac{s}{\sqrt{n}}$$

$$u = \bar{x} + t_{\alpha/2,n-1}\frac{s}{\sqrt{n}}.$$

Vi tar fram s genom formeln

$$s = \sqrt{s^2} = \sqrt{\frac{1}{n-1}\sum_{k=1}^{n}(\bar{x}-x_k)^2} = \sqrt{\frac{1}{9}\sum_{k=1}^{10}(2.02-x_k)^2} = \cdots = 6.49.$$

Från tabell VI får vi att $t_{0.05,9} = 1.833$. Vi sätter in allt detta i formlerna och får att

$$\ell = 2.02 - 1.833\frac{6.49}{\sqrt{10}} = -1.74$$

$$u = 2.02 + 1.833\frac{6.49}{\sqrt{10}} = 5.79.$$

Vårt observerade 90%-iga konfidensintervall blev alltså $[-1.74; 5.79]$.

8. Kalle gillar att spela poker och han är även duktig på matematik. Han har just fått jobb på ett nätcasino där han ska räkna på sannolikheter inom pokerspel. Hans nya chef hävdar att om man blir tilldelad en pokerhand slumpmässigt så är chansen exakt 50% att man får par eller bättre, dvs proportionen av alla pokerhänder med par eller bättre är 0.5. Kalle tycker att det verkar märkligt att denna proportion skulle vara exakt 0.5 så han vill göra ett test för att se om det verkligen kan stämma. Han delar 200 pokerhänder slumpmässigt till sig själv och får par eller bättre i 114 av händerna.

   (a) Hjälp Kalle att utföra hypotestestet. Använd signifikansnivå $\alpha = 0.05$. Ska Kalle förkasta nollhypotesen på denna signifikansnivå? (4p)

   (b) Är p-värdet för testet större eller mindre än 0.05? Tips: du behöver inte ta fram p-värdet för att svara på frågan. (1p)

   (c) Är det möjligt att vi har begått ett typ-I-fel? Är det möjligt att vi har begått ett typ-II-fel? (1p)

   Lösning: Låt p vara den riktiga sannolikheten att få par eller bättre om man blir tilldelad en pokerhand slumpmässigt.

   (a) Eftersom det bara är en avvikelse från 0.5 han är intresserad av, dvs inte större än eller mindre än, så är det ett tvåsidigt test vi bör göra. Hypoteserna blir alltså

   $$H_0: p = 0.5 \qquad H_1: p \neq 0.5.$$

   Vi har att $np_0 \geq 15$ och att $n(1-p_0) \geq 15$ så vi kan använda teststatistikan

   $$Z = \frac{\hat{p}-p_0}{\sqrt{p_0(1-p_0)/n}}$$

som är $N(0,1)$-fördelad. I vårt fall är $p_0 = 0.5$, $\hat{p} = 114/200 = 0.57$ och $n = 200$ så vår observerade teststatistika är

$$z = \frac{0.57 - 0.5}{\sqrt{0.5(1-0.5)/200}} = 1.98.$$

Vi tittar i $N(0,1)$ efter de kritiska värdena och ser att $\pm z_{\frac{\alpha}{2}} = \pm z_{0.025} = \pm 1.96$. Eftersom den observerade teststatistikan ligger utanför de kritiska värdena så förkastar vi nollhypotesen. På signifikansnivå $\alpha = 0.05$ så kan vi dra slutsatsen att chefen har fel!

(b) Eftersom vi förkastade nollhypotesen på signifikansnivå $\alpha = 0.05$ så vet vi att p-värdet är mindre än 0.05. Generellt sett förkastar nollhypotesen på signifikansnivå α om och endast om p-värdet är mindre än α.

(c) Typ-I-fel begås när man förkastar en sann nollhypotes. Eftersom vi förkastade nollhypotesen så är det möjligt att vi har begått ett typ-I-fel. Typ-II-fel begås när man inte förkastar en falsk nollhypotes. Eftersom vi förkastade nollhypotesen så kan vi inte ha begått ett typ-II-fel.