

Tentamentsskrivning i **Matematisk Statistik med Metoder MVE490**

Tid: den 29 oktober, 2016

Examinatorer: Kerstin Wiklander och Erik Broman.

Jour: Claes Andersson (tel 0734031540).

Hjälpmedel: miniräknare, egenhändigt skriven formelsamling om två A4 fram och bak (dvs 4 sidor) samt tabeller (delas ut på plats).

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. Hjälp Emil med att besvara frågorna nedan. (13p)

- (a) Man har slumpmässigt valt ut en insatslägenhet bland de som är till salu i ett område. Till denna har man registrerat ett antal variabler med följande värden:

Utgångspris: 2 495 000,-

Avgift/månad: 5329,-

Antal rum: 4

Boarea: $91.5 m^2$

Fasadens material: Tegel

Fasadens färg: Röd

Avstånd till närmaste spårvagnshållplats: 318 m

Antal P-platser utanför huset: 42

Ange för varje variabel om den är kvalitativ eller kvantitativ. I fallet kvantitativ, ange även om det är en diskret eller en kontinuerlig variabel.

Ange vilken typ av lägesmått och spridningsmått man kan använda till kvalitativ variabel respektive till kvantitativ variabel.

Antag att man har ett stickprov på sju lägenheter. Värdena för antal rum blev: 2 3 1 2 5 5 3.

Vilken typ av illustration är lämplig för att sammanfatta dessa data? Konstruera en och glöm inte att namnge axlarna.

- (b) Om man i en datamängd med stickprovsstorlek $n = 12$ har beräknat standardavvikelsen till $s = 4.3$. Vad är då skattningen av den teoretiska standardavvikelsen för medelvärdet?

- (c) Fyll i tre av de fyra rutorna nedan med signifikansnivån α (Typ I-fel), med β (Typ II-fel) samt med styrkan $1 - \beta$.

		Verklighet: H ₀ är sann	Verklighet: H ₀ är falsk
Beslut:	Förkasta inte H ₀		
Beslut:	Förkasta H ₀		

- (d) I en artikel beskrivs en studie där ett t-test använts för att undersöka om två olika tillverkningsmetoder skiljer sig åt när det gäller hållfasthet. Signifikansnivån α hade valts till 5% och p-värdet rapporterades ha fått värdet 0.09.

Vilken princip skall användas för att dra en slutsats med hjälp av ett p-värde? Ge också en skiss på hur ett p-värde kan illustreras i ovanstående frågeställning.

Vilket av påståendena nedan är en rätt tolkning av resultatet? Du måste motivera ditt val.

1. Skillnad mellan \bar{x}_1 och \bar{x}_2 är signifikant.
 2. Skillnad mellan μ_1 och μ_2 är signifikant.
 3. Skillnaden mellan medelvärdena är relevant.
 4. Skillnaden mellan väntevärdena är relevant.
 5. Skillnad mellan \bar{x}_1 och \bar{x}_2 är inte signifikant.
 6. Skillnad mellan μ_1 och μ_2 är inte signifikant.
 7. De två medelvärdena är lika.
 8. De två väntevärdena är lika.
 9. Det är 95% säkert att medelvärdena är lika.
 10. Det är 95% säkert att väntevärdena är lika.
2. Betrakta två händelser A, B och antag att $\mathbb{P}(A) = 0.5$ och $\mathbb{P}(B) = 0.3$ (6p)
- (a) Om $\mathbb{P}(A \cap B) = 0.1$, vad blir då $\mathbb{P}(A \cup B)$ och $\mathbb{P}(A^c \cup B)$?
 - (b) Om istället A, B är oberoende, vad blir då $\mathbb{P}(A \cup B)$ och $\mathbb{P}(A^c \cup B)$?
 - (c) Om C, D är två händelser som uppfyller att $\mathbb{P}(C) = 0.2$, $\mathbb{P}(D) = 0.6$ och $\mathbb{P}(C \cup D) = 0.7$, kan då C, D vara oberoende?

Lösning:

- (a) Vi har att

$$\mathbb{P}(A \cup B) = \mathbb{P}(A) + \mathbb{P}(B) - \mathbb{P}(A \cap B) = 0.5 + 0.3 - 0.1 = 0.7$$

och

$$\begin{aligned} \mathbb{P}(A^c \cup B) &= \mathbb{P}(A^c) + \mathbb{P}(B) - \mathbb{P}(A^c \cap B) \\ &= 1 - \mathbb{P}(A) + \mathbb{P}(B) - (\mathbb{P}(B) - \mathbb{P}(A \cap B)) \\ &= 1 - \mathbb{P}(A) + \mathbb{P}(A \cap B) = 1 - 0.5 + 0.1 = 0.6. \end{aligned}$$

- (b) Om A, B är oberoende så har vi att $\mathbb{P}(A \cap B) = \mathbb{P}(A)\mathbb{P}(B) = 0.5 * 0.3 = 0.15$. Räkningarna ovan ger då

$$\mathbb{P}(A \cup B) = \mathbb{P}(A) + \mathbb{P}(B) - \mathbb{P}(A \cap B) = 0.5 + 0.3 - 0.15 = 0.65$$

och

$$\mathbb{P}(A^c \cup B) = 1 - \mathbb{P}(A) + \mathbb{P}(A \cap B) = 0.5 + 0.15 = 0.65.$$

- (c) Vi ser att

$$\mathbb{P}(C \cap D) = \mathbb{P}(C) + \mathbb{P}(D) - \mathbb{P}(C \cup D) = 0.2 + 0.6 - 0.7 = 0.1.$$

Då $\mathbb{P}(C)\mathbb{P}(D) = 0.2 * 0.6 = 0.12$ är C, D inte oberoende.

3. Glenn projekterar ett större villaområde för VillaByggare AB. Glenn har att välja mellan två byggmetoder för fasaderna. Antingen väljer han en traditionell metod, eller också väljer han den nya enstegstätade metoden.

Kostnaden för en fasad som byggs med den traditionella metoden är (kontinuerligt) likformigt fördelat på intervallet $[500000, 650000]$ SEK. Denna metod har sannolikhet 0.93 att inte drabbas av fukt och mögelskador under den 10års period som VillaByggarna ger garanti på villorna.

Om fasaden byggs med den nya enstegstätade metoden är kostnaden istället likformigt fördelat på intervallet $[200000, 400000]$ SEK. Denna metod har sannolikhet 0.4 att drabbas av fukt och mögelskador under 10års perioden.

Om garantin löses ut räknar VillaByggarna att det medför en extrakostnad på 1000000 i båda fallen. (8p)

- (a) Vilken metod bör Glenn välja?
 (b) Glenn valde metod nr 2 för alla villorna i hela projektet. Byggnationen består av 341 likadana villor. Om vi antar att alla mögel och fuktskador uppkommer oberoende av varandra, vad är då sannolikheten att fler än 140 villor drabbas?

Lösning:

- (a) Låt X_1 vara totala kostnaden för metod 1, och X_2 för metod 2. X_1 består av två delar: $U_1 \sim U[500000, 650000]$ som är byggkostnaden och G_1 som är garantikostnaden. Vi har att

$$\begin{aligned} \mathbb{E}[X_1] &= \mathbb{E}[U_1 + G_1] = \mathbb{E}[U_1] + \mathbb{E}[G_1] \\ &= \frac{500000 + 650000}{2} + 0 * \mathbb{P}(G_1 = 0) + 1000000 * \mathbb{P}(G_1 = 1000000) \\ &= 575000 + 70000 = 645000. \end{aligned}$$

På liknande sätt blir

$$\begin{aligned}\mathbb{E}[X_2] &= \mathbb{E}[U_2 + G_2] = \mathbb{E}[U_2] + \mathbb{E}[G_2] \\ &= \frac{200000 + 400000}{2} + 0 * \mathbb{P}(G_2 = 0) + 1000000 * \mathbb{P}(G_2 = 1000000) \\ &= 300000 + 400000 = 700000.\end{aligned}$$

Förutom att det är mer etiskt är det även kostnadseffektivt att välja metod nr 1.

- (b) Låt $X = \#$ villor som drabbas. Vi har då att $X \sim \text{Bin}(341, 0.4)$ och vi söker

$$\mathbb{P}(X \geq 140).$$

Här måste vi använda oss av normalapproximation, och har då att $X \approx N(np, np(1-p)) = N(136.4, 81.84)$. Därmed har vi att

$$\begin{aligned}\mathbb{P}(X \geq 140) &= \mathbb{P}\left(\frac{X - 136.4}{\sqrt{81.84}} \geq \frac{140 - 136.4}{\sqrt{81.84}}\right) \\ &\approx \mathbb{P}(Z \geq 0.3979) = 1 - \mathbb{P}(Z \leq 0.3979) \approx 1 - 0.654 = 0.346.\end{aligned}$$

4. Sara spelar på Lotto. Det finns 40 nummer att välja mellan, och varje rad består av 7 nummer. (4p)

- (a) Hur många rader finns det totalt?
(b) Hur många rader innehåller nummer 1?

Vid varje dragning väljs 7 nummer ut. För högsta vinst måste man ha spelat en rad där alla 7 nummer matchas.

- (c) Ange en lämplig fördelning för X = första numret som väljs.
(d) Ange en lämplig fördelning för Y = raden som väljs.

Lösning:

- (a) $\binom{40}{7}$
(b) Om ettan räknas bort har vi 39 nummer kvar varav 6 skall väljas ut. Vi får $\binom{39}{6}$.

Vid varje dragning väljs 7 nummer ut.

- (c) X är likformigt fördelad på mängden $\{1, 2, \dots, 40\}$
(d) Y är likformigt fördelad på mängden av alla rader.

5. Lisa gillar hjortar. Hon ställer sig därför i en glänta i skogen under en timme och spanar efter hjortar. Lisa antar att antalet hjortar hon får se under en given timme är Poisson-fördelat med parameter $\lambda = 1.2$ (6p)

- (a) Vad är sannolikheten att Lisa får se minst 3 hjortar?
 (b) Vad är sannolikheten att Lisa får se minst 3 hjortar om vi vet att hon fått syn på minst en?
 (c) Är Lisas antagande om Poisson-fördelningen rimlig? Motivera varför/varför inte.

Lösning: Låt $X = \#$ hjortar som Lisa får se.

- (a) Vi har att

$$\begin{aligned}\mathbb{P}(X \geq 3) &= 1 - \mathbb{P}(X \leq 2) \\ &= 1 - \mathbb{P}(X = 0) - \mathbb{P}(X = 1) - \mathbb{P}(X = 2) \\ &= 1 - e^{-\lambda} \left(1 + \frac{\lambda^1}{1!} + \frac{\lambda^2}{2!} \right) \approx 0.1205\end{aligned}$$

- (b) Vi söker $\mathbb{P}(X \geq 3|X \geq 1)$ och har att

$$\mathbb{P}(X \geq 3|X \geq 1) = \frac{\mathbb{P}(X \geq 3)}{\mathbb{P}(X \geq 1)} = \frac{\mathbb{P}(X \geq 3)}{1 - e^{-\lambda}} \approx 0.1725$$

- (c) Om vi antar att hjortarna vid varje given tidpunkt befinner sig likformigt fördelade i skogen och att de rör sig oberoende av varandra så är det ett rimlig fördelning. Dessa antaganden kan dock med fog kritiseraras.

6. Kalle planerar att utföra ett tvåstickprovs t-test. (4p)

- (a) Ange förutsättningarna för att kunna använda ett sådant test som statistisk metod vid hypotestest.
 (b) Kalle studerade absorptionstiden (i minuter) av vätska i två olika material. Värdena blev:
 Material 1: 1.03 0.65 0.35 0.43 0.53 2.59
 Material 2: 0.87 1.13 1.18 0.32 3.87 0.79
 Bör man använda t-test i denna situation? Varför/varför inte?
 (c) Använd en av datamängderna i (b). Räkna som om punktskattningarna av de två parametrarna i modellen enligt (a)-uppgiften är de sanna värdena. Vad är i så fall risken att få ett orimligt värde enligt modellen, dvs sannolikheten att få ett negativt värde?

7. Mäklare från två olika firmor har oventande om varandra fått göra värderingar av ett antal bostäder. Upplägget av jämförelsen gjordes så att de fick värdera samma lägenheter. Resultatet blev: (5p)

Objekt nr:	1	2	3	4	5	6
Mäklare A:	2 395000	475000	1 385000	1 775000	885000	795000
Mäklare B:	2 310000	510000	1 290000	1 680000	830000	770000

- (a) Vilken statistisk testmetod skall användas och hur ser modellen ut? Testa på signifikansnivå 5% om Mäklare A generellt gör en högre värdering än Mäklare B. Beskriv hela arbetsgången och förklara dina beteckningar.
 - (b) Vad blir svaret om frågeställningen istället gäller att upptäcka en skillnad mellan mäklarna? Visa hur du drar den slutsatsen.
8. En genomgång gjordes av 200 slumpmässigt valda fönsterkarmar i ett stort bostadsområde. Det upptäcktes att 33 hade sådana defekter att de måste bytas ut. (4p)
- (a) Beräkna ett konfidensintervall för proportionen defekta i hela området. Använd konfidensgrad 99%. Undersök också om förutsättningarna för beräkningarna är uppfyllda.
 - (b) Hur skulle den teoretiska standardavvikelsen skattas om det istället hade handlat om ett test?