

FACIT: Tentamen L9MA30, LGMA30

Matematik 3 för lärare, åk 7-9, Sannolikhetslära och statistik,
Matematik 3 för gymnasielärare, Sannolikhetslära och statistik

2016-10-21 kl. 08.30-12.30

Examinator och Jour: Henrike Häbel, tel. 073 0671252 alt. 031 772 8296

Hjälpmedel: Valfri miniräknare (tomt minne och ingen möjlighet att koppla upp sig till nätet; kontrolleras av examinatorn), (dubbelsidigt) A4 ark med egna anteckningar samt med skrivningen utdelade tabellsidor och formelsamling.

Maxpoäng: 30. För godkänd krävs minst 15 p, för Väl Godkänd 24 p. Eventuell bonuspoäng från 2016 adderas. Lösningar skall redovisas till varje uppgift. Uppgifterna är ej ordnade efter svårighetsgrad. Rättningsprotokoll anslås senast tre veckor efter tentamen.

1. Visa att för tre händelser A , B och C i något utfallsrum Ω med $0 < P(C) < 1$ samt

$$P(A|C) \geq P(B|C) \text{ och } P(A|C^*) \geq P(B|C^*)$$

gäller

$$P(A) \geq P(B).$$

Ledning: Du kan använda lagen om total sannolikhet.

2p

Lösningsförslag:

$$P(A) \stackrel{LTS}{=} P(C)P(A|C) + P(C^*)P(A|C^*) \geq P(C)P(B|C) + P(C^*)P(B|C^*) = P(B) \quad (2p)$$

2. Varje elev måste välja 6 matematiska uppgifter under en tentamen, 3 från linjär algebra och 3 från matematisk statistik. Det finns 5 möjliga uppgifter för varje ämne och 10 elever som tar tentamen. Vad är sannolikheten att två slumpmässigt valda elever har valt samma kombination?

2p

Lösningförslag:

Det finns

$$\binom{5}{3} \binom{5}{3} = 100 \quad (1p)$$

möjliga kombinationer. Om vi tar två elever, spelar det ingen roll vilken kombination den ena har, men det finns bara en gynnsam kombination (utfall) för den andra. Därmed blir svaret

$$P(\text{samma kombination}) = \frac{1}{100}. \quad (1p)$$

3. En elev söker till lärarprogrammet på universitet A och B. Han tror att möjligheten att komma in på de olika programmen är 0.7 för A och 0.8 för B. Låt X vara totala antalet program eleven kommer in på av dessa två. Händelserna att komma in på olika program antas oberoende. Bestäm sannolikhetsfunktionen och fördelningsfunktionen för X .

3p

Lösningförslag:

$$p_X(k) = \begin{cases} P(A^*)P(B^*) = 0.3 \cdot 0.2 = 0.06 & , k = 0 \\ P(A)P(B^*) + P(B)P(A^*) = 0.7 \cdot 0.2 + 0.8 \cdot 0.3 = 1 - (0.06 + 0.56) = 0.38 & , k = 1 \\ P(A)P(B) = 0.7 \cdot 0.8 = 0.56 & , k = 2 \end{cases} \quad (2p)$$

och

$$F_X(x) = \begin{cases} 0 & , x < 0 \\ 0.06 & , 0 \leq x < 1 \\ 0.44 & , 1 \leq x < 2 \\ 1 & , x \geq 2 \end{cases} \quad (1p)$$

4. Anna och Petter spelar Trivial Pursuit med fyra kategorier och Anna har precis kommit i mål. Petter ska, slumpmässigt och utan att titta på nästa frågekort, bestämma från vilken kategori Anna ska svara på den sista, avgörande frågan för att ta hem spelet.

- a) Anna är bra på 'underhållning' och 'natur och vetenskap'. På frågor från de två kategorierna svarar hon rätt med 90% sannolikhet och på frågor från de två övriga kategorierna bara med 30% sannolikhet. Rita ett trädidiagram och beräkna sannolikheten för att Anna svarar rätt.

2p

- b) Givet att hon svarar fel på den sista frågan, vad är sannolikheten att Petter har valt kategori 'underhållning'?

2p

Lösningförslag:

- a) Trädidiagram (1p)

$$P(\text{Anna svara rätt}) = P(A) = 2 \frac{0.9}{4} + 2 \frac{0.3}{4} = \frac{3}{5} = 0.6 \quad (1p)$$

b)

$$P(A^*) = \sum_{i=1}^4 P(B_i)P(A^*|B_i) = 1 - 0.6 = 0.4 \quad (0.5\text{p})$$

Låt B_1 vara händelse att *kategori* är 'underhållning'.

$$\begin{aligned} P(B_1|A^*) &\stackrel{\text{Bayes}}{=} \frac{P(B_1)P(A^*|B_1)}{P(A^*)} \quad (1\text{p}) \\ &= \frac{\frac{1}{4} \cdot \frac{1}{10}}{\frac{4}{10}} = \frac{\frac{1}{40}}{\frac{4}{10}} = \frac{1}{16} = 0.0625 \quad (0.5\text{p}) \end{aligned}$$

5. Låt X och Y vara två kontinuerliga stokastiska variabler. Betrakta den simultana täthetsfunktionen

$$f_{XY}(x, y) = \begin{cases} 4xy & , 0 \leq x, y \leq 1 \\ 0 & , \text{annars} \end{cases}$$

a) Beräkna de marginella täthetsfunktionerna $f_X(x)$ och $f_Y(y)$.

2p

b) Är X och Y oberoende?

1p

c) Beräkna $E(X + Y)$ och $V(X + Y)$.

3p

Lösningsförslag:

a)

$$f_X(x) = \int_0^1 4xy dy = \left[4x \frac{y^2}{2} \right]_0^1 = 2x, \quad \forall 0 \leq x \leq 1 \quad (1\text{p})$$

och

$$f_Y(y) = \int_0^1 4xy dx = \left[4y \frac{x^2}{2} \right]_0^1 = 2y, \quad \forall 0 \leq y \leq 1. \quad (1\text{p})$$

b) X och Y är oberoende eftersom

$$f_{X,Y}(x, y) = 4xy = 2x \cdot 2y = f_X(x)f_Y(y). \quad (1\text{p})$$

c) Eftersom X och Y är oberoende gäller $E(X+Y) = E(X)+E(Y)$ och $V(X+Y) = V(X)+V(Y)$.

$$E(X) = E(Y) = \int_0^1 2x^2 dx = \left[2 \frac{x^3}{3} \right]_0^1 = \frac{2}{3} \quad (1\text{p})$$

$$V(X) = V(Y) = \int_0^1 2x^3 dx - \left(\frac{2}{3} \right)^2 = \left[2 \frac{x^4}{4} \right]_0^1 - \frac{4}{9} = \frac{2}{4} - \frac{4}{9} = \frac{1}{18} = 0.056 \quad (1\text{p})$$

och därmed blir $E(X + Y) = \frac{2}{3} + \frac{2}{3} = \frac{4}{3} = 1.333$ och $V(X + Y) = \frac{1}{18} + \frac{1}{18} = 0.111 \quad (1\text{p})$.

6. Paprikor som slumpmässigt väljs från ett stort parti väger i genomsnitt 95 gram, och vikten varierar enligt en fördelning med en standardavvikelse som är 20 gram. Antag att successivt valda paprikor har en vikt som är oberoende av varandra och följer en fördelning med dessa parametrar.

a) Du väljer nu 100 paprikor. Vad är sannolikheten att dessa tillsammans väger minst 10 kg?

3p

b) Borde standardavvikelsen vara större eller mindre för att sannolikheten för att paprikorna tillsammans skall väga minst 10 kg skall vara större än 0.01?

2p

Lösningförslag:

a) Betrakta $X \approx N(100 \cdot 95, \sqrt{100} \cdot 20) = N(9500, 200)$ (1p) och

$$P(X \geq 10000) = 1 - P(X \leq 10000) = 1 - \Phi\left(\frac{10000 - 9500}{200}\right) = 1 - \Phi(2.5) = 1 - 0.9938 = 0.0062 \quad (2p)$$

b)

$$P(X \geq 10000) = 1 - \Phi(\lambda_\alpha) \stackrel{!}{\geq} 0.01 \Leftrightarrow \frac{10000 - 9500}{10\sigma} \leq 2.32 \Leftrightarrow \sigma \geq \frac{500}{10 \cdot 2.32} = 21.55.$$

Det vill säga att standardavvikelsen borde vara större (21.55 istället för 20). Här kan man också resonera (utan att beräkna σ) att standardavvikelsen borde vara större för att då är det mer sannolikt att 100 paprikor väger mer än i a). (2p)

7. Antag att vi har gjort 15 mätningar av längden på ett speciellt stag till båtar. Dessa stag kan antas vara approximativt normalfördelade. Från mätningarna beräknades delsummorna

$$\sum_{i=1}^{15} x_i = 199 \text{ och } \sum_{i=1}^{15} x_i^2 = 2787.$$

Beräkna ett 95 % konfidensintervall för väntevärdet μ av den stokastiska variabeln X = stag längden. Ger detta intervall grund att förkasta nullhypotesen att $\mu = 10$?

3p

Lösningförslag:

Vi får ett $100(1 - \alpha)\%$ konfidensintervall med $\alpha = 0.05$ för μ med

$$I_\mu = \left[\bar{x} \pm t(n-1)_{\frac{0.05}{2}} \sqrt{\frac{s_x^2}{n}} \right] \quad (1p)$$

och

$$\bar{x} = \frac{199}{15} = 13.267, \quad s_x^2 = \frac{1}{15-1}(2787 - 15 \cdot 13.267^2) = 10.486, \quad t(15-1)_{0.025} = 2.145 \quad (0.5p)$$

Så vi får

$$I_\mu = \left[13.267 \mp 2.145 \sqrt{\frac{10.486}{15}} \right] = [11.474, 15.060] \quad (0.5p)$$

$10 \notin I_\mu$, så därmed finns grund att förkasta nullhypotesen att $\mu = 10$. (1p)

8. Betrakta följande del av en artikel publicerad den 6 juli 2016 på aftonbladet.se.

‘I en annan undersökning, utförd av Quinnipiacuniversitetet, som också presenterades nyligen hade Clinton ett försprång på bara två procentenheter.

Hon fick då stöd av 42 procent av de tillfrågade jämfört med 40 procent för Trump.

Felmarginalen i den nya undersökningen, där 1 610 registrerade väljare intervjuades, var plus/minus 2.4 procentenheter, enligt TT.’

Med vilken konfidensnivå beräknades felmarginalen? (Antag en oändlig population av registrerade väljare.)

3p

Antal väljare som röstar för Clinton antas $\text{Bin}(1610, p)$ -fördelade, där p är den sanna proportionen väljare som röstar för Clinton. Skatta p med $\hat{p} = 0.42$ (0.5p). Eftersom $n\hat{p}(1 - \hat{p}) \approx 392 > 10$ får man använda normalapproximationen (0.5p) och får

$$I_p = [\hat{p} \pm 0.024] = \left[0.42 \pm \lambda_? \sqrt{\frac{0.42(1 - 0.42)}{1610}} \right] = [0.42 \pm \lambda_? 0.0123] \quad (1p)$$

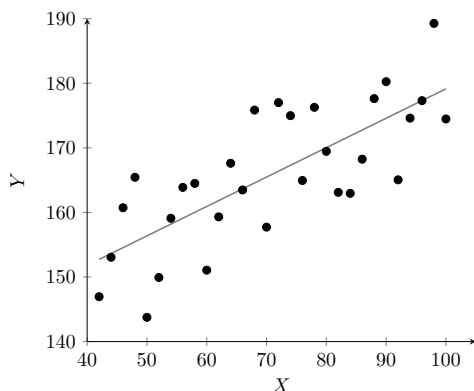
Det vill säga att $\lambda_? = 0.024/0.0123 \approx 1.96 = \lambda_{0.05/2}$ (0.5p). Så konfidensnivån har antagligen varit 5% (0.5p).

9. a) Titta på regressionslinjerna anpassade till datamaterial i A) och B) nedan. Vad tror du korrelationskoefficienten

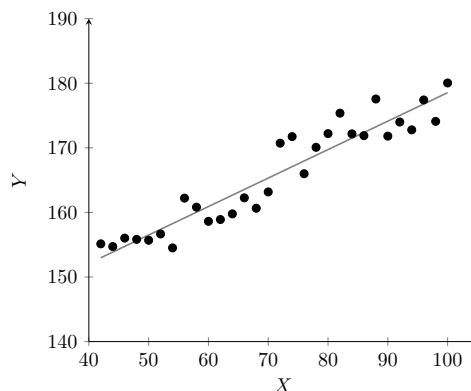
$$\rho = \frac{C(X, Y)}{D(X)D(Y)}$$

kan vara för X och Y i A) respektiv B)? Motivera ditt svar.

A)



B)



1p

- b) Varför skattas kovariansen $C(X, Y)$ med

$$s_{xy}^2 = \frac{1}{n-1} \sum_{i=1}^n (x_i - \bar{x})(y_i - \bar{y}_i),$$

där $\bar{x}$ och $\bar{y}$ är medelvärden?

Lösningsförslag:

- a) Regressionslinjen i A och B har positivt löpning, därför måste korrelationskoefficienten också vara positiv. Spridningen av observationer är större i A än i B och därför måste $0 < \rho_A < \rho_B$. (1p)
- b) Här kan man resonera fritt. T.ex. att kovariansen är definierat som $E((X - E(X))(Y - E(Y)))$ och då sätter man bara i observationer och medelvärdena. Man delar med $(n - 1)$, så att det passar ihop med stickprovvarianser. (1p)

Lycka till!