

TENTAMEN I STATISTIKENS GRUNDER 1

2012-10-03

Skrivtid: kl 9.00 - 14.00
Godkända hjälpmedel: Miniräknare, språklexikon
Bifogade hjälpmedel: Formelblad, statistiska tabeller

- Tentamen består av 5 uppgifter, i förekommande fall uppdelade i deluppgifter. Maximalt antal poäng anges per uppgift.
 - För full poäng på en uppgift krävs tydliga, utförliga och väl motiverade lösningar.
 - Kontrollera alltid dina beräkningar och lösningar! Slarvfel kan också ge poängavdrag!
-

Uppgift 1.

Anta att händelserna A , B och C inträffar med sannolikheterna $P(A) = 0,4$; $P(B) = 0,5$ respektive $P(C) = 0,8$. Anta vidare att snitthändelserna inträffar med sannolikheterna $P(A \cap B) = 0,2$; $P(A \cap C) = 0,32$ respektive $P(B \cap C) = 0,4$.

- a) Visa att de tre händelserna är parvis oberoende, dvs. att paren (A, B) , (A, C) respektive (B, C) är statistiskt oberoende. (4p)
- b) Givet sannolikheterna ovan, ange sannolikheterna $P(\bar{A} \cup B)$ och $P(A|C)$. (4p)
- c) Anta nu dessutom att $P(A \cap B \cap C) = 0,2$. Är de tre händelserna tillsammans statistiskt oberoende? (6p)
- d) Anta fortfarande att $P(A \cap B \cap C) = 0,2$. Ange $P(\bar{A} \cap B \cap C)$ och $P(A \cap B \cap \bar{C})$. (6p)

Uppgift 2.

Ett postorderföretag noterar tiden mellan utskick av beställningar och mottagande av betalning för varorna mätt i veckor. Fördelningen över tiden ges i tabellen nedan:

X = antal veckor till betalning	1	2	3	4	5
Andel	0,16	0,48	0,24	0,08	0,04

- a) Ange fördelningsfunktionen för tiden mellan utskick och betalning i tabellform. Illustrera både frekvens- och fördelningsfunktion i lämpliga grafer. (6p)
- b) Ange den betingade frekvensfunktionen för X givet att $1 < X < 5$. (6p)
- c) Beräkna väntevärde och standardavvikelse för X givet att $1 < X < 5$. (8p)

Uppgift 3.

Vid en bostadsundersökning valde man slumpmässigt ut ett antal familjer boende i hyreslägenheter. För varje familj registrerade man X = antalet barn och Y = antalet rum exkl. kök. I tabellen nedan anges den simultana frekvensfunktionen:

	$Y = 1$	2	3	4
$X = 0$	0,11	0,09	0,06	0,01
1	0,07	0,12	0,12	0,02
2	0,02	0,06	0,16	0,04
3	0	0,03	0,04	0,02
4	0	0	0,02	0,01

- Ange de marginella frekvensfunktionerna (sannolikhetsfunktionerna) för X resp. Y och beräkna deras respektive väntevärde och varians. (6p)
- Beräkna kovariansen och korrelationen mellan X och Y . Är X och Y oberoende? (6p)
- Förklara kortfattat (max $\frac{1}{2}$ - 1 A4) begreppet *kausalitet* och ange tre kriterier som man brukar ange som krav för att det (eventuellt) ska föreligga kausalitet. Relatera diskussion till exemplet ovan. Är det rimligt att tro att det råder ett kausalt förhållande mellan X och Y ? (8p)

Uppgift 4.

Ett godisföretag tillverkar halstabletter. Vikten för en förpackning antas vara normalfördelad med väntevärde 21,4 och varians 0,16.

- Beräkna sannolikheten att en slumpmässigt vald förpackning väger mindre än 20,89 gram. (6p)
- 20 förpackningar väljs ut slumpmässigt och vägs. Avrunda sannolikheten från uppgift a) och beräkna sannolikheten att fler än fem förpackningar väger mindre än 20,89 gram. (6p)
- Senare väljs 100 förpackningar ut och vägs. Beräkna approximativt sannolikheten att fler än fem förpackningar väger mindre än 20,89 gram. (8p)

Uppgift 5.

Anta att man har n stycken normalfördelade stokastiska variabler $X_1, X_2, \dots, X_n$. Alla har samma väntevärde och varians; dvs. $E(X_i) = a$ och $V(X_i) = b$ för alla $i = 1, \dots, n$. De är dessutom korsvis oberoende.

- Låt $Y = X_1 + X_2 + \dots + X_n$. Ange väntevärde och varians för Y . (5p)
- Låt $W = Y/n$. Ange väntevärde och varians för W . Vad brukar W kallas? (7p)
- Anta att $n = 100$, $a = 10$ och $b = 25$. Vilket värde på w ger $P(W > w) = 0,025$? Obs! Då alla X_i är normalfördelade blir W också normalfördelad. (8p)

FORMLER

Additionssatsen: $P(A \cup B) = P(A) + P(B) - P(A \cap B)$

Multiplikationssatsen: $P(A \cap B) = P(A|B) \cdot P(B) = P(B|A) \cdot P(A)$

Väntevärde för diskret s.v. X :

$$E(X) = \sum_{x \in \Omega_x} x f(x)$$

Varians för diskret s.v. X :

$$V(X) = \sum_{x \in \Omega_x} [x - E(X)]^2 f(x) = \sum_{x \in \Omega_x} x^2 f(x) - [E(X)]^2 = E(X^2) - E(X)^2$$

Kovarians och korrelation mellan två diskreta s.v. X och Y , ($\mu_x = E(X)$ och $\mu_y = E(Y)$):

$$\text{Cov}(X, Y) = \sum_{x \in \Omega_x} \sum_{y \in \Omega_y} (x - \mu_x)(y - \mu_y) f(x, y) = \sum_{x \in \Omega_x} \sum_{y \in \Omega_y} xy f(x, y) - \mu_x \mu_y$$

$$\text{Corr}(X, Y) = \frac{\text{Cov}(X, Y)}{SD(X) \cdot SD(Y)} = \frac{\text{Cov}(X, Y)}{\sqrt{V(X)V(Y)}}$$

Räkneregler för väntevärden och varianser (a, b och c är konstanter och X och Y är stokastiska variabler)

$$E(c) = c$$

$$V(c) = 0$$

$$E(X + c) = E(X) + c$$

$$V(X + c) = V(X)$$

$$E(aX) = aE(X)$$

$$V(aX) = a^2 V(X)$$

$$E(aX + bY + c) = aE(X) + bE(Y) + c \quad V(aX + bY + c) = a^2 V(X) + b^2 V(Y) + 2ab \text{Cov}(X, Y)$$

Binomialfördelningen:

$$f(x) = \binom{n}{x} p^x (1-p)^{n-x} = \frac{n!}{x! (n-x)!} p^x (1-p)^{n-x}$$

Poissonfördelningen:

$$f(x) = \frac{\lambda^x e^{-\lambda}}{x!}$$

LÖSNINGSFÖRSLAG (kortfattade)

Uppgift 1.

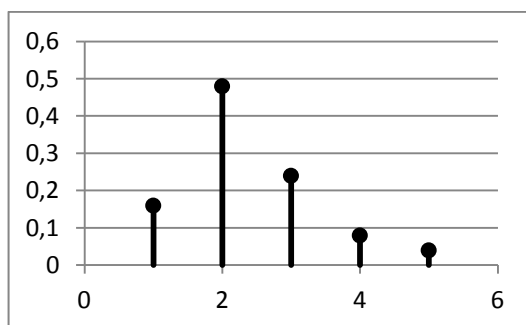
- a) $P(A \cap B) = 0,2 = 0,4 \cdot 0,5 = P(A) \cdot P(B) \Rightarrow$ oberoende
 $P(A \cap C) = 0,32 = 0,4 \cdot 0,8 = P(A) \cdot P(C) \Rightarrow$ oberoende
 $P(B \cap C) = 0,4 = 0,5 \cdot 0,8 = P(A) \cdot P(B) \Rightarrow$ oberoende
- b) $P(\bar{A} \cup B) = [\text{additionssatsen} + A \text{ och } B \text{ oberoende}] = P(\bar{A}) + P(B) - P(\bar{A}) \cdot P(B)$
 $= 1 - P(A) + P(B) - [1 - P(A)] \cdot P(B) = 0,6 + 0,5 - 0,6 \cdot 0,5 = 0,8$
 $P(A|C) = [\text{ty } A \text{ och } C \text{ oberoende}] = P(A) = 0,4$
- c) $P(A) \cdot P(B) \cdot P(C) = 0,4 \cdot 0,5 \cdot 0,8 = 0,16 \neq 0,2 = P(A \cap B \cap C) \Rightarrow$ sammantagna finns ett beroende mellan A , B och C .
- d) De två delmängderna $(\bar{A} \cap B \cap C)$ och $(A \cap B \cap C)$ är disjunkta och slår man ihop dem, dvs. skapar unionen av dem, får man $(B \cap C)$. Rita ett Venndiagram! Alltså, $P(\bar{A} \cap B \cap C) + P(A \cap B \cap C) = P(B \cap C) \Rightarrow P(\bar{A} \cap B \cap C) = 0,4 - 0,2 = 0,2$
På motsvarande sätt kan man resonera kring $(A \cap B \cap \bar{C})$ och $(A \cap B \cap C)$; alltså $P(A \cap B \cap \bar{C}) + P(A \cap B \cap C) = P(A \cap B) \Rightarrow P(A \cap B \cap \bar{C}) = 0,2 - 0,2 = 0$

Uppgift 2.

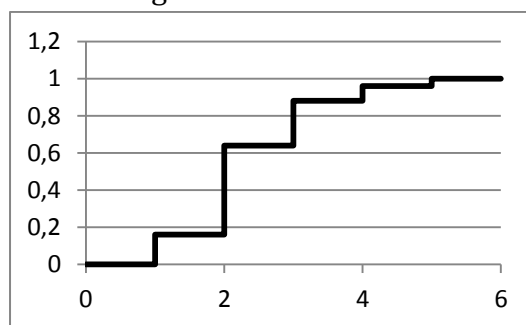
- a) Fördelningsfunktionen i tabellform:

$X = \text{antal veckor till betalning}$	1	2	3	4	5
Andel = $f(x)$	0,16	0,48	0,24	0,08	0,04
Fördelningsfunktion = $F(x)$	0,16	0,64	0,88	0,96	1,00

Frekvens- el. sannolikhetsfunktion



Fördelningsfunktion



- b) $P(1 < X < 5) = P(2 \leq X \leq 4) = P(2) + P(3) + P(4) = 0,80$. Den betingade fördelningen $f_{X|2 \leq X \leq 4}(x) = f(x)/P(2 \leq X \leq 4)$ för $x = 2, 3$ eller 4 och $f_{X|2 \leq X \leq 4}(x) = 0$ annars.

I tabellform:

x	1	2	3	4	5	Summa
$f_{X 2 \leq X \leq 4}(x)$	–	0,6	0,3	0,1	–	1

c) $E(X | 2 \leq X \leq 4) = 2 \cdot 0,6 + 3 \cdot 0,3 + 4 \cdot 0,1 = 2,5$

$$E(X^2 | 2 \leq X \leq 4) = 2^2 \cdot 0,6 + 3^2 \cdot 0,3 + 4^2 \cdot 0,1 = 6,7$$

$$V(X | 2 \leq X \leq 4) = E(X^2 | 2 \leq X \leq 4) - E(X | 2 \leq X \leq 4)^2 = 6,7 - 2,5^2 = 0,45;$$

$$SD(X | 2 \leq X \leq 4) = \sqrt{0,45} \approx 0,671$$

Uppgift 3.

a) Marginella frekvensfunktionerna (sannolikhetsfunktionerna) för X resp. Y

x	0	1	2	3	4	Summa
$\sum_y f(x, y) = f_X(x)$	0,27	0,33	0,28	0,09	0,03	1

y	1	2	3	4	Summa
$\sum_x f(x, y) = f_Y(y)$	0,2	0,3	0,4	0,1	1

$$E(X) = 0 \cdot 0,27 + 1 \cdot 0,33 + 2 \cdot 0,28 + 3 \cdot 0,09 + 4 \cdot 0,03 = 1,28$$

$$E(X^2) = 0^2 \cdot 0,27 + 1^2 \cdot 0,33 + 2^2 \cdot 0,28 + 3^2 \cdot 0,09 + 4^2 \cdot 0,03 = 2,74$$

$$V(X) = E(X^2) - E(X)^2 = 1,1016 (\approx 1,10)$$

$$E(Y) = 1 \cdot 0,2 + 2 \cdot 0,3 + 3 \cdot 0,4 + 4 \cdot 0,1 = 2,4$$

$$E(Y^2) = 1^2 \cdot 0,2 + 2^2 \cdot 0,3 + 3^2 \cdot 0,4 + 4^2 \cdot 0,1 = 6,6$$

$$V(Y) = E(Y^2) - E(Y)^2 = 0,84$$

b) $E(XY) = \sum_x \sum_y xyf(x, y) = 0 \cdot 1 \cdot 0,11 + 0 \cdot 2 \cdot 0,09 + \dots$ [varje kombination av x och y] ...
 $\dots + 4 \cdot 3 \cdot 0,02 + 4 \cdot 4 \cdot 0,01 = 3,49$

$$\text{Cov}(X, Y) = E(XY) - E(X) \cdot E(Y) = 3,49 - 1,28 \cdot 2,4 = 0,418$$

$$\text{Corr}(X, Y) = \text{Cov}(X, Y) / \sqrt{V(X)V(Y)} = 0,418 / \sqrt{1,1016 \cdot 0,84} = 0,434535\dots (\approx 0,435)$$

X, Y oberoende ty $\text{Corr}(X, Y) \neq 0$ (och även $\text{Cov}(X, Y) \neq 0$)

c) Se kurslitteratur och föreläsningssanteckningar. Kortfattat: Kausalitet = orsaks-samband mellan händelser, egenskaper, variabler, något påverkar något annat. Krav = assymetri, kontrollerbarhet och tidsfördröjning (med korta förklaringar). Ex. diskussion: Det finns ett positivt linjärt samband mellan lägenhetens storlek och antalet barn, ju fler barn desto större lägenhet. Statistiska samband är dock inte ett bevis på kausalitet. Lägenhetens storlek kan kanske till viss del påverkas av familjens storlek. Å andra sidan kanske vissa väntar med att skaffa sig barn tills de har en större lägenhet. Kan finnas andra variabler som påverkar bägge (ekonomi?).

Uppgift 4.

a) X = vikten i en förpackning, antas att $X \sim N(21,4 ; 0,16)$

$$Z = (X - 21,4)/0,4 \text{ ger } Z \sim N(0;1)$$

$$\begin{aligned}
 P(X < 20,89) &= P(Z < (20,89 - 21,4)/0,4) = P(Z < -1,275) \approx P(Z < -1,28) = \\
 &= P(Z > 1,28) = 1 - P(Z \leq 1,28) = 1 - \Phi(1,28) = [\text{avläst från tabell}] = 1 - 0,89973 = \\
 &= 0,10027 \approx 0,10
 \end{aligned}$$

- b) Y = antal förpackningar av $n = 20$ som väger mindre än 20,89 gram.

$Y \sim \text{Bin}(20; 0,10)$ där $p = 0,10$ är från uppgift a.

$$P(Y > 5) = 1 - P(Y \leq 5) = [\text{avläst från tabell}] = 1 - 0,98875 = 0,01125 \approx 0,011$$

- c) Y = antal förpackningar av $n = 100$ som väger mindre än 20,89 gram.

$Y \sim \text{Bin}(100; 0,10)$

$$E(Y) = np = 100 \cdot 0,10 = 10 \text{ och } V(Y) = np(1-p) = 9$$

Alt. 1 Normalapproximering, motiveras av tumregeln $np(1-p) > 5$.

Låt $Z = (Y-10)/3$ och approximera med $Z \sim N(0;1)$

$$\begin{aligned}
 P(Y > 5) &= 1 - P(Y \leq 5) \approx 1 - P(Z \leq (5-10+1/2)/3) = 1 - P(Z \leq -1,5) = \\
 &= 1 - P(Z > 1,5) = 1 - (1 - P(Z \leq 1,5)) = P(Z \leq 1,5) = \Phi(1,5) = [\text{avläst från tabell}] \\
 &= 0,93319 \approx 0,933
 \end{aligned}$$

Alt. 2 Poissonapproximering, motiveras nästan av tumregeln $p < 0,10$.

Approximera fördelningen för Y med $Y \sim \text{Po}(np) = \text{Po}(10)$.

$$P(Y > 5) = 1 - P(Y \leq 5) \approx [\text{avläst från tabell}] = 1 - 0,06709 = 0,93291 \approx 0,933$$

$$\textbf{Alt. 3} \text{ Exakt svar. } P(Y > 5) = 1 - P(Y \leq 5) = 1 - \sum_{k=0}^5 \binom{100}{k} 0,1^k 0,9^{100-k} = 0,94242$$

Uppgift 5.

a) $E(Y) = E(X_1 + X_2 + \dots + X_n) = E(X_1) + E(X_2) + \dots + E(X_n) = n \cdot a$

$$V(Y) = V(X_1 + X_2 + \dots + X_n) = V(X_1) + V(X_2) + \dots + V(X_n) = n \cdot b$$

b) $E(W) = E(Y/n) = \frac{1}{n}E(Y) = \frac{1}{n} \cdot n \cdot a = a$

$$V(W) = V(Y/n) = \frac{1}{n^2}E(Y) = \frac{1}{n^2} \cdot n \cdot b = \frac{b}{n}$$

W brukar kallas medelvärde.

c) $W \sim N(a, \frac{b}{n})$; insättning ger $W \sim N(10, \frac{25}{100})$.

$$Z = (W-10)/(5/10) = 2W - 20 \Leftrightarrow W = 10 + Z/2 \text{ och där } Z \sim N(0,1).$$

$$P(W > w) = P(Z > z) = 0,025. \text{ Från Tabell 2 läses att } P(Z > 1,96) = 0,025$$

$$z = 1,96 \Rightarrow w = 10 + 1,96/2 = 10,98 \text{ dvs. } P(W > 10,98) = 0,025$$