STOCKHOLMS UNIVERSITET
Statistiska institutionen
Jessica Franzén

TENTAMEN I STATISTIKENS GRUNDER 1

2017-09-26

Skrivtid: 10.00-15.00
Godkända hjälpmedel: Miniräknare
Tentamen består av fem uppgifter. För full poäng på en uppgift krävs tydliga, utförliga och väl motiverade lösningar.

Uppgift 1. (20 poäng)
Ett företag har en fabrik i vart och ett av tre länder och de tre fabrikerna tillverkar oberoende av varandra samma produkt. Produkterna samlas på ett gemensamt lager för vidare leverans till återförsäljare. Fabriken i Kina är den modernaste och står för 40 procent av den totala produktionen, medan produktionen i Indien och Vietnam står för vardera 30 procent. Tyvärr uppstår ibland fel i produktionen. I Kina är i genomsnitt 4 procent av de tillverkade produkterna felaktiga. Motsvarande siffra i Indien är 10 procent och i Vietnam 6 procent.

a) Vad är sannolikheten att en kund köper en felaktig produkt?

b) Vad är sannolikheteten att en felaktig produkt är tillverkad i Kina?

Uppgift 2 (20 poäng)
Den stokastiska variabeln X kan anta värdena $x = 0, 1, 2$ och 3 och har frekvensfunktionen

$$f_X(x) = \frac{12}{25(x+1)}$$

a) Beräkna $F_X(2)$ och förklara vad det är du räknat ut.

Den stokastiska variabeln Y kan anta värdena $y = 0$ och 1 och har frekvensfunktionen

$$f_Y(y) = 0.25^y(1-0.25)^{1-y}$$

X och Y är oberoende.

b) Ange den simultana frekvensfunktionen för X och Y i tabellform.

c) Beräkna väntevärde och varians för $T = 6X + 2Y$

Uppgift 3. (20 poäng)
Du vill komma in i en port som är låst med en fyrsiffrig portkod. Du vet vilka fyra siffror som ingår i portkoden men inte i vilken ordning de ska vara. Siffrorna är 0,1,7 och 8.

a) Du prövar en kod med slumpmässigt vald ordning på de 4 siffrorna.Vad är sannolikheten att du slår rätt kod?

b) Ytterligare 19 personer kommer efter dig och vill in i porten och de har samma information som du dvs att siffrorna 0,1,7 och 8 ingår men inte i vilken ordning. Var och en försöker, oberoende av varandra, med en slumpmässigt vald ordning. Vad är sannolikheten att minst en av er slår rätt kod?

Uppgift 4 (20 poäng)
Låt T vara medeltemperaturen i havet under juli månad på Mallorca. Vi antar att T är normalfördelad med väntevärde 24.8 grader och standardavvikelse 2.2 grader.

a) Vad är sannolikheten att medeltemperaturen i juli ett slumpmässigt valt år ligger under 22 grader?

b) Vad är sannolikheten att medeltemperaturen i juli ligger under 22 grader fler än 10 år av ett sekel (100 år)? Ange eventuella antaganden som måste vara uppfyllda för dina beräkningar.

c) Beräkna gränsvärdet där det är 95 procent säkert att medeltemperaturen i juli ligger över detta värde.

Uppgift 5. (20 poäng)
Vägförvaltningen i ett län har fått i uppdrag att planera en ny väg som bl.a. kräver att det byggs två tunnlar och en liten bro. I den första planläggningsfasen fokuseras det på variablerna

$$\begin{aligned} U &= \text{byggprojektets varaktighet (i veckor efter startdatum)} \\ K &= \text{kostnad per vecka under bygget (i miljoner kronor)} \end{aligned}$$

Båda variablerna är osäkra och vägplanläggarna anger sannolikhetsfördelningen för det två variablerna. Variabeln U kan anta värdena 100, 180 och 400 med sannolikheten 0.3, 0.4 respektive 0.3. Variabeln K kan anta värdena 1 och 2 med sannolikheten 0.6 respektive 0.4. Vi får också veta att simultansannolikheten $f_{U,K}(400, 1) = 0$ och att den betingade sannolikheten $f_{K|U}(1|180) = 0.75$.

a) Ange den simultana frekvensfunktionen för U och K.

b) Beräkna korrelationen mellan U och K. Tolka resultatet.

c) Är U och K oberoende? Förklara kopplingen mellan korrelation och beroende. Är två okorrelerade variabler alltid oberoende? Varför eller varför inte?

Statistiska institutionen

Stockholms universitet

Rättningsblad

Datum: 26/09/17

Sal: Uggleviksalen

Tenta: Statistikens grunder

Kurs: Statistikens grunder

ANONYMKOD: SG-0013

☒ Jag godkänner att min tenta får läggas ut anonymt på hemsidan som studentsvar.

OBS! SKRIV ÄVEN PÅ BAKSIDAN AV SKRIVBLADEN

SORRY LÄSTE INTE ORDENTLIGT

Markera besvarade uppgifter med kryss

1	2	3	4	5	6	7	8	9	Antal inl. blad
X	X	X	X	X					7st
Lär.ant. 20	20	18	17	19					

JK

POÄNG	BETYG	Lärarens sign.
94	A	[signature]

SU, STATISTIK

Skrivsal: Uggleviksalen **Anonymkod:** SG-0013 **Blad nr:** 1.

Uppgift 1.

A = vara kommer från Kina $P(A) = 0,4$

B = —||— Indien $P(B) = 0,3$

C = —||— Vietnam $P(C) = 0,3$

X = felaktig $P(X) = ?$

$\bar{X}$ = hel $P(\bar{X}) = 1 - P(X)$

Vi vet också att:

$P(X|A) = 0,04$ $P(X|B) = 0,1$ $P(X|C) = 0,06$

där $P(X|Y) = \dfrac{P(X \cap Y)}{P(Y)} \rightarrow P(X \cap Y) = P(X|Y) \cdot P(Y)$

Det ger följande värden för vår simultanfördelning:

$P(X \cap A) = 0,04 \cdot 0,4 = 0,016$

$P(X \cap B) = 0,1 \cdot 0,3 = 0,03$

$P(X \cap C) = 0,06 \cdot 0,3 = 0,018$

	A	B	C	
X	0,016	0,03	0,018	0,064
$\bar{X}$	0,384	0,27	0,282	0,936
	0,4	0,3	0,3	1

a) $P(X \cap A) + P(X \cap B) + P(X \cap C)$

$= 0,016 + 0,03 + 0,018 = 0,064$

$= P(X)$

1,0

b) $P(A|X) = \dfrac{P(A \cap X)}{P(X)}$

$P(A|X) = \dfrac{0,016}{0,064} = 0,25$

$P(A|X) = 0,25$

1,0

SU, STATISTIK

Skrivsal: Ugglevikssalen **Anonymkod:** SG-0013 **Blad nr:** 2

Uppgift 2.

X är en stokastisk variabel som kan anta värderna 0, 1, 2, 3

$$f(x) = \frac{12}{25(x+1)}$$

Beräkningar f(x)

$f(0) = \frac{12}{25 \cdot 1} = 0,48$

$f(1) = \frac{12}{50} = 0,24$

$f(2) = \frac{25}{75} = 0,16$

$f(3) = \frac{25}{75} = 0,12$

1. Tabell för X:

X	f(x)	xf(x)	$x^2f(x)$	F(x)
0	0,48	0	0	0,48
1	0,24	0,24	0,24	0,72
2	0,16	0,32	0,64	0,88
3	0,12	0,36	1,08	1
	Σ1	Σ0,92	Σ1,96	

a) F(2) = 0,88 och beskriver sannolikheten för att X är likamed eller mindre än 2 alltså $X \leq 2$

6

2. Tabell för y

y	f(y)	yf(y)	$y^2f(y)$	F(y)
0	0,75	0	0	0,75
1	0,25	0,25	0,25	1
		Σ0,25	Σ0,25	

Beräkningar för f(y)

$f(0) = 0,25^0 \cdot (1-0,25)^{1-0}$

$= 1 \cdot 0,75 = 0,75$

$f(1) = 0,25^1 \cdot (1-0,25)^{1-1}$

$= 0,25 \cdot 1 = 0,25$

b) Den simultana frekvensfunktionen

	y = 0	1	
X = 0	0,36	0,12	0,48
1	0,18	0,06	0,24
2	0,12	0,04	0,16
3	0,09	0,03	0,12
	0,75	0,25	1

Vid stokastiskt oberoende gäller:

$P(X \cap Y) = P(X) \cdot P(Y)$

c) $\underbrace{E(T) = E(6X+2Y) = 6E(X)+2E(Y)}$

Ges av räkneregler i formelblad

$E(X) = \sum_x x f(x)$ av tabell för X ges $0{,}92$

$E(Y) = \sum_y y f(y)$ av tabell för y ges $0{,}25$

$E(T) = 6 \cdot 0{,}92 + 2 \cdot 0{,}25 = 6{,}02$ ✓C

$E(T) = 6{,}02$

Ges av formel

$V(T) = V(6x+2y) = 6^2 V(X) + 2^2 V(Y) + 2 \cdot 6 \cdot 2 \operatorname{cov}(X,Y)$

Eftersom X och Y är oberoende är $\operatorname{cov}(X,Y) = 0$

$V(X) = \sum_x x^2 f(x) - E(X)$ av tabell för x ges

$V(X) = 1{,}96 - 0{,}92^2 = 1{,}1136$

$V(y) = \sum_y y^2 f(y) - E(y)$ av tabell för y ges

$V(y) = 0{,}25 - 0{,}25^2 = 0{,}1875$

SU, STATISTIK

Skrivsal: Uggleviksalen **Anonymkod:** SG-0013 **Blad nr:** 3

forts. 2

$V(T) = 6^2 \cdot 1{,}136 + 2^2 \cdot 0{,}1875 + 2 \cdot 6 \cdot 2 \cdot 0$

$V(T) = 40{,}8342$

K

8

SU, STATISTIK

Skrivsal: Ugglenkssalen **Anonymkod:** SLG-0013 **Blad nr:** 4

Uppgift 3.

Möjliga kombinationer:

0, 1, 7, 8	1, 7, 8, 0	7, 8, 0, 1	8, 7, 1, 0
0, 7, 1, 8	1, 8, 7, 0	7, 0, 8, 1	8, 1, 7, 0
0, 8, 7, 1	1, 0, 8, 7	7, 1, 8, 0	8, 0, 7, 1
0, 7, 8, 1	1, 8, 0, 7	7, 8, 1, 0	8, 7, 0, 1
0, 8, 1, 7	1, 7, 0, 8	7, 1, 0, 8	8, 0, 1, 7
0, 1, 8, 7	1, 0, 7, 8	7, 0, 1, 8	8, 1, 0, 7

a) Det finns sammanlagd 24 uttal = 4!
Om vi antar att sannolikheten för alla uttag är samma är sannolikheten att så givet uttall

$\frac{1}{24} = 0{,}0417$ (avrundat på 4 decimaler) 10

b) Ny stokastisk variabel $Y \sim bin(19, 0{,}0417)$

$P(Y \geq 1) = 1 - P(Y \leq 0) = 1 - F(0)$

$F(x) = \binom{n}{x} p^x \cdot (1-p)^{n-x}$

$F(0) = \binom{19}{0}^{20} 0{,}0417^0 \cdot (1-0{,}417)^{19} = 0{,}4452$

$P(Y \geq 1) = 1 - 0{,}4452 = 0{,}5548$ 8

Sannolikheten för att minst 1 av 19 slår rätt är ungefär $\approx 0{,}55$

SU, STATISTIK

Skrivsal: Ugglevikssalen Anonymkod: SG-0013 Blad nr: 5

Uppgift 4

T = medeltemperatur i havet under Juli

$T \sim N(\mu = 24{,}8, \sigma = 2{,}2)$

a) Beräkna $P(T \leq 22)$

Normalfördelningsfunktionen ger:

$$P(T \leq t) = P\left(Z \leq \frac{t-\mu}{\sigma}\right)$$

$$P(T \leq 22) = P\left(Z \leq \frac{22-24{,}8}{2{,}2}\right) = \Phi(-1{,}27)$$

tabell ger att $Z = 0{,}89796$ då $\Phi(1{,}27)$. För $\Phi(-1{,}27)$ gäller $1 - \Phi(1{,}27)$. Sannolikheten för $T \leq 22$ är $0{,}10204 \approx 0{,}1$

4

b) Vi antar att sannolikheten $P(T \leq 22)$ är konstant över tid.

Vi definierar en ny stokastisk variabel Y = antal gånger $T \leq 22$ förekommer

$Y \sim bin(100, 0{,}1)$ dvs $n = 100$ $p = 0{,}1$

Beräkna $P(Y \geq 11)$. Vi antar att "större än" innebär att $Y \neq 10$.

$P(Y \geq 11) = 1 - P(Y \leq 10)$ VC

Vänd

Formel ger: Kolla att normalapprox är OK

$$P(Y \leq y) = P\left(Z \leq \frac{y + \frac{1}{2} - np}{\sqrt{np(1-p)}}\right)$$ där μ av $X \sim bin = np$

och V av $X \sim bin = np(1-p)$

Det ger att:

$$P(Y \leq 10) = P\left(Z \leq \frac{10,5 - 100 \cdot 0,1}{\sqrt{100 \cdot 0,9 \cdot 0,1}}\right) = \Phi(0,17) \rightarrow$$

$= 0,56749$

$$P(Y \geq 11) = 1 - 0,56749 = 0,43251$$

Sannolikheten att medeltemperaturen under ett sekel ligger under 22° i sammanlagt fler än 10 år är $\approx 0,43$ Antaganden?

7

c) Beräkna värdet på t när $P(T > t) = 0,95$

I tabell för normalfördelningens kvantiler kan utläsas att då $P(Z > z_a) = 0,95$ är $z_a = (-1,6449)$

Eftersom $z = \frac{x - \mu}{\sigma}$ kan vi räkna ut t.

$$-1,6449 = \frac{t - 24,8}{2,2}$$

$$(-1,6449 \cdot 2,2) + 24,8 = t$$

$$t = 21,18122$$

6

Avrundat till hela grader ger det att det är 0,95 sannolikt att medeltemperaturen i havet är över 21° i Juli månad.

SU, STATISTIK

Skrivsal: Ugglevikssalen **Anonymkod:** SG-0013 **Blad nr:** 6

Uppgift 5.

U = byggprojekters varaktighet i veckor
K = Kostnad per vecka i miljarder kronor

$\Omega(U)$: 100, 180, 400
$\Omega(K)$: 1, 2

$P(U=100) = 0{,}3$ $P(U=180) = 0{,}4$ $P(U=400) = 0{,}3$
$P(K=1) = 0{,}6$ $P(K=2) = 0{,}4$

Vi vet också att:
$P(U=400, K=1) = 0$ och att
$P(K=1 \mid U=180) = 0{,}75$

a)

		K 1	2	
	100	0,3	0	0,3
U	180	0,3	0,1	0,4
	400	0	0,3	0,3
		0,6	0,4	1

Beräkningar

$P(K \mid U) = \dfrac{P(K \cap U)}{P(U)}$ det ger →

$0{,}75 = \dfrac{P(1 \cap 180)}{0{,}4}$

$P(1 \cap 180) = 0{,}3$

Med hjälp av faktumet att
$P(U \cap K) + P(U \cap \bar{K}) = P(U)$
kan de andra sannolikheterna beräknas.

6

b) Formel ger att $Corr(U,K) = \frac{Cov(U,K)}{\sqrt{V(U)V(K)}}$

För att räkna ut korrelationen måste vi allra först räkna ut kovariansen och de enskilda varianserna för K och U.

U	f(U)	Uf(U)	$U^2f(U)$
100	0,3	30	3000
180	0,4	72	12960
400	0,3	120	48000
Σ	1	222	63960

K	f(K)	Kf(K)	$K^2f(K)$
1	0,6	0,6	0,6
2	0,4	0,8	1,6
Σ	1	1,4	2,2

$$V(U) = \sum_U U^2 f(U) - \left(\sum_U U f(U)\right)^2$$

$V(U) = 63960 - 222^2 = 14676$ ✓

$$V(K) = \sum_K K^2 f(K) - \left(\sum_K K f(K)\right)^2$$

$V(K) = 2,2 - 1,4^2 = 0,24$ ✓

För att räkna ut Cov(U,K) krävs E(U,K):

	1	2
100	30	0
180	54	36
400	0	240

$E(U,K) = 30 + 54 + 36 + 240 = 360$

Formel ger:

$$Cov(U,K) = \sum_U \sum_K UK f_{U,K}(U,K) - E(U) \cdot E(K)$$

där $E(U) = \sum_U U f(U)$ och $E(K) = \sum_K K f(K)$

Det ger:

$Cov(U,K) = 360 - 222 \cdot 1,4 = 49,2$ ✓

SU, STATISTIK

Skrivsal: Uggleviksalen Anonymkod: SG-0013 Blad nr: 7

Fort upg 5

Korrelationen mellan U och K blir $\frac{49,2}{\sqrt{14676 \cdot 0,24}}$

$= 0,8290020994 \approx 0,83$ ✓

Att korrelationen är 0,83 innebär att värdena på K och U beror mycket av varandra (samspelar i hög grad) och att sambandet är positivt - alltså att när U ökar, ökar även K

10

c) U och K är alltså två beroende variabler vilket betyder att sannolikheten för K = k varierar beroende på värdet på U och vice versa.

Om 2 variabler är oberoende kommer $E(X,Y) = E(X)E(Y)$ och därför ge $cov(X,Y) = 0$

om kovariansen är noll kommer även korrelationen bli 0.

Det finns variabler som har kovariansen 0 men som är beroende. Man kan därför endast säga att en oberoende variabel ska ha cov och corr = 0 men inte att alla variabler med cov, och corr = 0 är oberoende.

Varför?

3