STOCKHOLMS UNIVERSITET
Statistiska institutionen
Jessica Franzén

TENTAMEN I STATISTIKENS GRUNDER 1

2017-02-15

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

Uppgift 1. (20 poäng)
Bland elitidrottare utförs kontinuerligt dopningstester. Proceduren vid dopningstester är att man först gör ett enkelt dopningstest s.k. screening (om screeningtestet visar ett positivt resultat går man vidare och utför mer avancerade tester). Vid detta första enkla screeningtest finns alltid en risk att man drar fel slutsatser, dvs antingen att screeningen visar att idrottaren är dopad fast han/hon inte är det eller att screeningen visar att idrottaren inte är dopad fast han/hon är det. Man vet att 2 procent av alla elitidrottare är dopade och 4 procent av alla screeningtest visar positivt. Sannolikheten att en slumpmässigt vald elitidrottare är dopad och samtidigt visar ett positivt resultat på sceeningtestet är 0.019.

a) Vad är sannolikheten att screeningen av en elitidrottare visar att idrottaren är dopad fast han/hon inte är det?

b) Vad är sannolikheten att screeningen visar att idrottaren inte är dopad fast han/hon är det?

Uppgift 2. (20 poäng)
Längden på graviditeter, X dagar, antas vara normalfördelad med väntevärde 266 dagar och standardavvikelse 16 dagar. En graviditet som inte avviker mer än 14 dagar från förväntad längd sägs vara av normal längd.

a) Vad är sannolikheten att en graviditet får normal längd?

Längden på fyra gravida väninnors graviditeter betecknas X_1, X_2, X_3, X_4. De antas vara obereonde och normalfördelade med samma väntevärde och standardavvikelse som ovan.

b) Vad är sannolikheten att alla fyra graviditeterna blir av normal längd?

c) Vad är sannolikheten att minst en av de fyra väninnorna föder för tidigt?

d) Räkna ut väntevärde och varians för $\overline{Y} = \frac{\sum_{i=1}^{n} X_i}{n}$ d.v.s. för medelvärdet av längden på de fyra vänninornas graviditeter.

Uppgift 3 (20 poäng)
Den beräknade inflationen X under nästa månad följer fördelningsfunktionen

$$F(x) = \frac{x}{5}$$

X är en kontinuerlig variabel i intervallet 0 till 5.

a) Vad är sannolikheten att inflationen blir mer än 3 procent?

b) Vad är sannolikheten att inflationen blir mellan 1.5 och 2 procent?

c) Vad är sannolikheten att inflationen blir mellan 1.5 och 2 procent om man redan vet att den blir högst 2 procent?

Uppgift 4 (20 poäng)
Betygen på en tenta i en klass om 90 studenter fördelar sig enligt följande:
A: 21%, B: 18% C: 10% D: 12% E: 14% F: 25%

Man väljer slumpmässigt 20 tentor med återläggning bland de 90 tentorna. Vad är sannolikheten att man bland de 20 dragna tentorna får

a) exakt 2 tentor med betyg A

b) färre än 8 tentor med betyg F

c) Man väljer nu slumpmässigt ut 50 tentor med återläggning. Vad är sannolikheten att fler än 35 av dessa har godkänt resultat d.v.s. betyg A, B, C, D eller E?

d) Man väljer slumpmässigt ut tentor utan återläggning tills man har fått 4 st med betyg A. Ange två skäl till varför Binomialfördelningen inte kan användas som modell för antal dragna tentor.

Uppgift 5. (20 poäng)
Vi har två stokastiska variabler med följande sannolikhetsfördelningar

x	1	2	3
$f(x)$	0.6	0.3	0.1

y	2	3	4
$f(y)$	0.05	0.5	0.45

Vidare vet man att
$f_{X|Y}(1|2) = 0.2$
$f_{X|Y}(2|3) = 0.4$
$f_{X|Y}(1|3) = 0.5$
$f_{X|Y}(3|2) = 0.8$

a) Ange den simultana frekvensfunktionen för X och Y.

b) Beräkna korrelationen mellan X och Y. Tolka resultatet.

c) Beräkna variansen av $T = 2X - 3Y$.

Statistiska institutionen

Stockholms universitet

Rättningsblad

Datum: 15/2-2017

Sal: Värtasalen

Tenta: Statistikens grunder 1

Kurs: Statistikens grunder 15 hp (dag)

ANONYMKOD: SG-0047

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

OBS! SKRIV ÄVEN PÅ BAKSIDAN AV SKRIVBLADEN

Markera besvarade uppgifter med kryss

1	2	3	4	5	6	7	8	9	Antal inl. blad
X	X	X	X	X					6
Lär.ant. 20	19	20	15	19,5					

POÄNG	BETYG	Lärarens sign.
93,5	A	GF

SU, STATISTIK

Skrivsal: Vårtasalen Anonymkod: SG-0047 Blad nr: 1

Uppg 1

A = Elitidrottare är dopad $\bar{A}$ = Elitidrottare ej dopad

B = Screening visar positivt $\bar{B}$ = Screening visar negativt

$P(A) = 0{,}02$ $P(B) = 0{,}04$ $P(A \cap B) = 0{,}019$

Vi ritar en fyrfältstabell och sätter in värdena

	A	$\bar{A}$	
B	0,019	0,021	0,04
$\bar{B}$	0,001	0,959	0,96
	0,02	0,98	1

a) Vi vill ta reda på $P(B|\bar{A})$, alltså att screeningen visar positivt givet att idrottaren ej är dopad.

Vi använder multiplikationssatsen:

$$P(\bar{A} \cap B) = P(B|\bar{A}) \cdot P(\bar{A})$$

$$P(B|\bar{A}) = \frac{P(\bar{A} \cap B)}{P(\bar{A})} \Rightarrow \frac{0{,}021}{0{,}98} \approx 0{,}0214 \approx 0{,}021$$

Svar: Slh för att screeningen visar positivt givet att idrottaren inte är dopad är 0,021

10

b) Ta reda på $P(\bar{B}|A)$.

$$P(\bar{B}|A) = \frac{P(A \cap \bar{B})}{P(A)} \Rightarrow \frac{0{,}001}{0{,}02} = 0{,}05$$

Svar: Slh att screeningen visar negativt givet att idrottaren är dopad är 0,05

10

SU, STATISTIK

Skrivsal: Vårtasalen Anonymkod: SG-0047 Blad nr: 2

Uppg 2 $\mu = 266$ $\sigma = 16$ $X \sim N(266, 16^2)$

a) $P(252 < X < 280)$ är slh för normal graviditetslängd

$P(252 < X < 280) = P(X < 280) - P(X < 252) =$

$P\left(\frac{X-\mu}{\sigma} < \frac{280-266}{16}\right) - P\left(\frac{X-\mu}{\sigma} < \frac{252-266}{16}\right) = P(Z < 0{,}875) - P(Z < -0{,}875)$

$= \Phi(0{,}875) - \Phi(-0{,}875) = \Phi(0{,}875 - (1 - \Phi(0{,}875)))$

Vi får avrunda $0{,}875 \approx 0{,}88$ för att hitta värdet i tabellen

Enligt tabell får vi då:

$0{,}81057 - (1 - 0{,}81057) = 0{,}62114 \approx 0{,}6211$

5

Svar: Slh för att en graviditet får normal längd är 0,6211

b) Fyra gravida kvinnor vad är slh för att alla får normal längd?

$n = 4$ $p = 0{,}6211$ $X \sim bin(4, 0{,}6211)$

$P(X=4) = f(4) = \binom{4}{4} 0{,}6211^4 \, 0{,}3789^0 = \underbrace{\frac{4!}{4!\,0!}}_{1} \cdot 0{,}6211^4 \cdot 0{,}3789^0 = 0{,}14881$

5

$\approx 0{,}1488$

Svar: Slh för att alla 4 graviditeter får normal längd är 0,1488

c) Slh för att minst 1 av de 4 föder för tidigt?

Först måste vi räkna ut slh att föda för tidigt:

$\mu = 266$ $\sigma = 16$ $X \sim N(266, 16^2)$

$P(X < 252) = P\left(\frac{X-\mu}{\sigma} < \frac{252-266}{16}\right) = \Phi(-0{,}875) = 1 - \Phi(0{,}88) = 1 - 0{,}81057$

$= 0{,}18943 \approx 0{,}1894$

R

Forts uppg 2

Slh att minst 1 av de 4 föder för tidigt:

$P(Z \geq 1)$ när $Z \sim Bin(4, 0{,}1894)$

$$P(Z \geq 1) = 1 - P(Z \leq 0) = 1 - f(0) = 1 - \binom{4}{0} 0{,}1894^0 0{,}8106^4 =$$

$$1 - \underbrace{\frac{4!}{0!4!}}_{1} \cdot 0{,}1894^0 \cdot 0{,}8166^4 \approx 0{,}5682$$

Svar: Slh för att minst 1 av de fyra föder för tidigt är $\underline{0{,}5682}$

5

d) Väntevärde & varians för $\bar{Y} = \frac{\sum_{i=1} x_i}{n}$, dvs medelvärdet av längden på de 4 kvinnornas graviditeter

$$E(\bar{Y}) = E\left(\frac{X_1 + X_2 + X_3 + X_4}{n}\right) = E\left(\frac{1}{n} \cdot (X_1 + X_2 + X_3 + X_4)\right)$$

$$= \{\text{mha räkneregler}\} = \frac{1}{n} E(X_1 + X_2 + X_3 + X_4)$$

$$= \frac{1}{n}\left(E(X_1) + E(X_2) + E(X_3) + E(X_4)\right) = \frac{1}{n} \cdot 4E(X) \Rightarrow$$

$$\frac{1}{4} \cdot 4 \cdot 266 = \underline{266}$$

R

$$V(\bar{Y}) = V\left(\frac{X_1 + X_2 + X_3 + X_4}{n}\right) = V\left(\frac{1}{n} \cdot (X_1 + X_2 + X_3 + X_4)\right)$$

$$= \{\text{mha räkneregler}\} = \left(\frac{1}{n}\right)^2 \cdot V(X_1 + X_2 + X_3 + X_4) =$$

$$\left(\frac{1}{n}\right)^2 \cdot \left(V(X_1) + V(X_2) + V(X_3) + V(X_4)\right) = \left(\frac{1}{n}\right)^2 \cdot 4 \cdot V(X) \Rightarrow$$

$$= \left(\frac{1}{4}\right)^2 \cdot 4 \cdot \underbrace{256}_{16^2} = 16384$$

Svar: Väntevärdet $E(\bar{Y}) = \underline{266}$ & $V(\bar{Y}) = \underline{16384}$

4

SU, STATISTIK

Skrivsal: Vartasalen Anonymkod: SG-0047 Blad nr: 3

Upg 3 X = bestående inflationen under nästa månad

$F(x) = \frac{x}{5}$, X är kont variabel intervall 0-5

a) Slh att inflationen blir mer än 3%

$P(X>3) = 1 - P(X<3) = 1 - F(3) = 1 - \frac{3}{5} = 0,4$

Svar: Slh att inflationen blir mer än 3% är $\underline{\underline{0,4}}$ 6

b) Slh att inflationen blir mellan 1,5% & 2%

$P(1,5 < X < 2) = P(X<2) - P(X<1,5) = F(2) - F(1,5) = \frac{2}{5} - \frac{1,5}{5} = 0,1$

Svar: Slh att inflationen blir mellan 1,5% & 2% är $\underline{\underline{0,1}}$ 6

c) Slh att inflationen blir mellan 1,5 & 2 om man redan vet att den högst kan bli 2.

Intervallet ändras alltså så att F(2) motsvarar slh i F(2) har i intervallet 0-5 slh $\left(F(2) = \frac{2}{5} = 0,4\right)$ 0,4 att inträffa, detta blir den nya "hela".

Alltså blir slh att inflationen är mellan 1,5 & 2 $= \frac{0,1}{0,4} = 0,25$

Svar: Slh att inflationen är mellan 1,5% & 2% när man redan vet att den är högst 2% är $\underline{\underline{0,25}}$

8

SU, STATISTIK

Skrivsal: Vartasalen **Anonymkod:** SG-0047 **Blad nr:** 4

Uppg 4 90 tentor, man drar slumpmässigt 20

a) Vad är Slh att man får 2 tentor med A av de 20 dragna.

Slh för att få A är 0,21.

$n = 20 \quad p = 0{,}21 \quad X \sim Bin(20, 0{,}21)$

$P(X=2) = \cancel{f(0) + f(1) + f(2)} = \cancel{\underbrace{\binom{20}{0}}_{1} \cdot 0{,}21^0 \cdot 0{,}79^{20} + \underbrace{\binom{20}{1}}_{20} \cdot 0{,}21^1 \cdot 0{,}79^{19}}$

$\boxed{+ \underbrace{\binom{20}{2}}_{190} \cdot 0{,}21^2 \cdot 0{,}79^{18} \approx 0{,}1770}$

1

Svar: Slh för att få exakt 2 tentor med A av de 20 dragna är $\underline{\underline{0{,}1770}}$

b) Vad är Slh att man får färre än 8 med F av de 20 dragna

Slh att få F = 0,25

$n = 20 \quad p = 0{,}25 \quad Y \sim Bin(20, 0{,}25)$

$P(X<8) = P(X \leq 7) = \{\text{enl tabell}\} = 0{,}89819 \approx 0{,}8982$

Svar: Slh för att få mindre än 8, dvs 7 eller färre med F av de 20 är $\underline{\underline{0{,}8982}}$

4

$\Rightarrow$

c) forts. uppg 4
$n=50$ $p=\{\text{Sannolikheten att få godkänt}\}=0{,}75$
$X \sim Bin(50, 0{,}75)$
Slh att fler än 35 är godkända
$P(X>35)=P(X\geq 36)=1-P(X\leq 35)=$ finns ej i tabell, för stort n. Är kontinuitetskorrektion möjligt?
Villkor:
$$\boxed{\begin{array}{ll} n\cdot p>5 & \text{då } p\leq 0{,}5 \\ n\cdot p\cdot(1-p)>5 & \text{då } p>0{,}5 \end{array}}$$
K
Vi prövar: Vårt $p>0{,}5$ därför prövas om $n\cdot p\cdot(1-p)>5$
$50\cdot 0{,}75\cdot 0{,}25 \approx 9{,}375.$
Ja!, kontinuitetskorrektion möjlig!
$$P(X\geq 36)=P\left(\frac{X-\frac{1}{2}-np}{\sqrt{np(1-p)}}>\frac{36-\frac{1}{2}-37{,}5}{\sqrt{50\cdot 0{,}75\cdot 0{,}25}}\right)=P(\underbrace{Z>-0{,}6531}_{\text{avrundar mer}})$$
avrundat
$$=1-P(Z\leq -0{,}6531)=1-\Phi(-0{,}6531)=1-(1-\Phi(0{,}65))=\{\text{enl tabell}\}=$$
$$=1-(1-0{,}74215)=0{,}74215\approx 0{,}7422$$
Svar: Slh att fler än 35 av de 50 tentorna har godkänt resultat är $\underline{\underline{0{,}7422}}$

8

d) 1. Binomialfördelningen används bara då det är med återläggning
2. Man kan inte få fram ett n värde då det kan bli olika fem dragning, allt mellan $n=4$ & $n=90$

2

SU, STATISTIK

Skrivsal: Värtasalen **Anonymkod:** SG-0047 **Blad nr:** 5

Uppg 5

X	f(x)
1	0,6
2	0,3
3	0,1

Y	f(Y)
2	0,05
3	0,5
4	0,45

$f_{X|Y}(1|2)=0,2$ $f_{X|Y}(2|3)=0,4$ $f_{X|Y}(1|3)=0,5$ $f_{X|Y}(3|2)=0,8$

a) Ange simultana frekvensfunktionen

Y \ X	1	2	3	
2	0,01	0	0,04	0,05
3	0,25	0,2	0,05	0,5
4	0,34	0,1	0,01	0,45
	0,6	0,3	0,1	

K

a) då de är beroende:

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

$P(1\cap 2)=0,2\cdot 0,05=0,01$

$P(2\cap 3)=0,4\cdot 0,5=0,2$

$P(1\cap 3)=0,5\cdot 0,5=0,25$

$P(3\cap 2)=0,8\cdot 0,05=0,04$

Svar: Se tabell

8

b) För att beräkna korrelationen beräknas först kovariansen

$Cov(X,Y)=\sum_x\sum_y xy\, f_{X,Y}(x,y)-E(X)E(Y)$

Vi börjar med att beräkna väntevärden & varianser för X & Y

X	f(x)	x·f(x)	$x^2\cdot f(x)$
1	0,6	0,6	0,6
2	0,3	0,6	1,2
3	0,1	0,3	0,9
	1	1,5	2,7

Y	f(Y)	Y·f(Y)	$Y^2\cdot f(Y)$
2	0,05	0,1	0,2
3	0,5	1,5	4,5
4	0,45	1,8	7,2
	1	3,4	11,9

$E(X)=1,5$ K $E(Y)=3,4$ K

$V(X)=2,7-1,5^2=0,45$ K $V(Y)=11,9-3,4^2=0,34$ K forts upp 5

Sen beräknar jag $x \cdot y$ i tabell

		X = 1	X = 2	X = 3
Y	2	2	4	6
	3	3	6	9
	4	4	8	12

Sen beräknar jag $x \cdot y \cdot f_{X,Y}(x,y)$ i tabell

		X = 1	X = 2	X = 3
Y	2	0,02	0	0,24
	3	0,75	1,2	0,45
	4	1,36	0,8	0,12

Sen beräknar jag $\sum xy f_{X,Y}(x,y)$:

$0{,}02+0{,}75+1{,}36+0+1{,}2+0{,}8+0{,}24+0{,}45+0{,}12 = 4{,}94$

Sen tar jag 4,94 minus väntevärdena för X & Y multiplicerade

$4{,}94 - 1{,}5 \cdot 3{,}4 = -0{,}16$

$Cov(X,Y) = -0{,}16$ K

$$Corr(X,Y) = \frac{Cov(X,Y)}{\sqrt{V(X)V(Y)}} = \frac{-0{,}16}{\sqrt{0{,}45 \cdot 0{,}34}} \approx -0{,}409$$ K

Svar: korrelationen är $-0{,}409$, detta innebär att det är ett ganska svagt negativt linjärt samband, hade varit starkt vid -1.

8

SU, STATISTIK

Skrivsal: Vårtasalen **Anonymkod:** SG-0047 **Blad nr:** 6

~~Forts uppg 5.~~

Variansen av $T = 2X - 3Y$

$V(T) = V(2x - 3y) = 2^2 V(X) + 3^2 V(Y) - 2 \cdot 2 \cdot 3 \, Cov(X, Y)$

$= 4 \cdot 0{,}45 + 9 \cdot 0{,}34 - 2 \cdot 2 \cdot 3 \cdot -0{,}16 = \cancel{2{,}94}$ 6,78

Svar: Variansen av $T = 2X - 3Y$ är $\cancel{2{,}94}$

3,5