STOCKHOLMS UNIVERSITET
Statistiska institutionen
Jessica Franzén

OMTENTAMEN I STATISTIKENS GRUNDER 1

2017-10-30

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

a) Förklara vad som menas med den klassiska sannolikhetsdefinitionen samt den frekventistiska sannolikhetsdefinitionen. Förklara skillnaden mellan dem.
b) Förklara de tre begreppen "snitthändelse", "komplementhändelse" och "ömsesidigt uteslutande händelser". Rita Venndiagram för att förtydliga förklaringarna.
c) Förklara skillnaden mellan en diskret och kontinuerlig fördelning. Rita ett exempel på vardera fördelning.
d) Ange de fyra skalnivåerna för data och förklara vad varje skalnivå innebär.
e) Förklara skillnaden mellan observationsstudier och experimentella studier.

Uppgift 2

Då Axel var på besök i Rom fick han problem med språket. Axel talar engelska men inte italienska och bara 1 av 5 italienare behärskar engelska. Under en promenad i centrum är det något lättare att träffa någon som talar engelska, då 1 av 5 personer som befinner sig i Roms centrum är utländska turister. Bland dessa turister är del hela 1 av 2 som behärskar engelska.

a) Vad är sannolikheten att en slumpvis person som Axel träffar i centrum talar engelska?

b) Vad är sannolikheten att denna slumpvisa person i a) (som talar engelska) är italienare?

c) På en kort promenad i centrum träffar Axel 16 personer. Vad är sannolikheten att det är lika många utländska turister som italienare bland dessa 16 personer?

d) Vad är sannolikheten att det är minst 8 utländska turister bland de 16 personerna i c)?

Uppgift 3
Kvadratmeterpriset på en slumpmässigt vald lägenhet i Småstad är normalfördelad med väntevärde 12 000 kronor och standardavvikelse 2000 kronor.

a) Vad är sannolikheten att en slumpmässigt vald lägenhet kostar mer än 15 000 kronor per kvadratmeter.

b) Vad är sannolikheten att en slumpmässigt vald lägenhet kostar mellan 10 000 och 12 000 kronor per kvadratmeter.

c) Vad är sannolikheten att en lägenhet kostar mer än 10 000 kronor per kvadratmeter, givet att vi vill göra en bra affär och bara är intresserade av lägenheter som kostar under genomsnittet d.v.s. 12 000 kornor per kvadratmeter?

d) Beräkna talet P där 10 procent av lägenheterna har ett högre kvadratmeterpris än P.

Uppgift 4
Ett tillverkningsföretag köper in komponenter från en leverantör. Komponenterna levereras i stora partier och innan ett parti accepteras görs en kvalitetskontroll. Ett slumpmässigt uval av 50 enheter kontrolleras. Om urvalet innehåller högst två felaktiga enheter accepteras partiet men om fler än två felaktiga enheter påträffas skickar man tillbaka hela partiet till leverantören.

a) Beräkna sannolikheten att ett parti accepteras om det innehåller 3 procent felaktiga enheter.

b) Beräkna förväntat antal felaktiga varor i ett urval av 50 enheter om partiet innehåller 3 procent felaktiga enheter.

c) Beräkna sannolikheten att ett urval om 300 enheter innehåller fler än 15 felaktiga enheter om partiet innehåller 3 procent felaktiga enheter.

Uppgift 5
Samma tillverkningsföretag som i uppgift 4 köper nu in en helt annan typ av komponent från en annan leverantör. Fördelningsfunktionen för antal defekta komponenter i en förpackning om 10 stycken komponenter ges i tabellen. Det förekommer aldrig fler än 3 defekta komponenter i en förpackning.

x	0	1	2	3
$F(x)$	0.83	0.92	0.99	1

a) Ange frekevensfunktionen för antal defekta komponenter i en förpackning om 10 stycken. Vad är sannolikheten att det finns 3 defekta komponenter i en förpackning?

b) Beräkna förväntat antal defekta komponenter i en förpackning.

c) Vi utgår ifrån att det i en förpackning finns 3 defekta komponenter av 10. Två komponenter dras slumpmässigt från dessa 10 utan återlägggning. Ange frekvensfunktionen för den stokastiska variabeln Y som anger antal defekta komponenter bland de två dragna komponenterna.

Statistiska institutionen

Stockholms universitet

Rättningsblad

Datum: 30/10/17

Sal: Ugglevik

Tenta: Statistikens grunder

Kurs: Statistikens grunder 1

ANONYMKOD: SG-0029

☑ 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					4
Lär.ant. 14	20	16	18	20					

POÄNG	BETYG	Lärarens sign.
88	B	JF

SU, STATISTIK

Skrivsal: Ugglevikssalen Anonymkod: SG-0029 Blad nr: 1

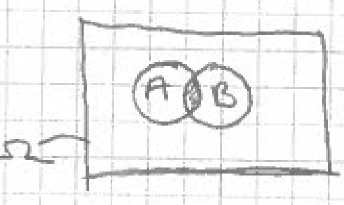
1a.) Klassisk sannolikhetsdefinition: I den klassiska sannolikhetsteorin tittar man på sannolikheten att en stokastisk händelse inträffar. Exempelvis ett kast med tärning. När man kastar en tärning så är sannolikheten (slh) 1 att just någon händelse i utfallsrummet Ω inträffar. Dock så är slh att få en 1:a $1/6$. När man väl kastar en tärning så kanske man inte får en 1:a efter 6 försök, eller ens i det 6000:e försöket. Men när tärningskasten $\to \infty$ så är slh $1/6$ att man får en 1:a. Samma princip gäller också exempelvis om man kastar ett mynt.

I den frekventistiska sannolikhetsdefinitionen så har man räknat ut sannolikheten efter att ha tittat på de olika utfallen eller händelserna i utfallsrummet.

2

b.) Snitthändelse – är slh att exempelvis en händelse A och B inträffar samtidigt exempelvis:

Venndiagram



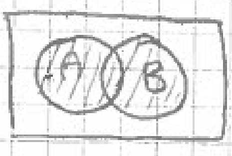
Med andra ord, en snitthändelse är $P(A \cap B)$ och motsvaras alltså av det skuggade området i figuren. Kan också skrivas som:

$$P(A \cap B) = P(B|A) \times P(A) = P(A|B) \times P(B)$$

k

Komplementhändelse: Är slh att antingen händelse A inträffar, eller händelse B inträffar, eller att båda inträffar samtidigt

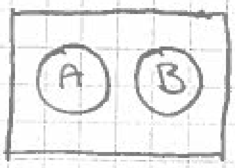
Venndiagram



alltså $P(A \cup B) = P(A) + P(B) - P(A \cup B)$

Det motsvaras av det streckade området i ← diagrammet.

Ömsesidigt uteslutande

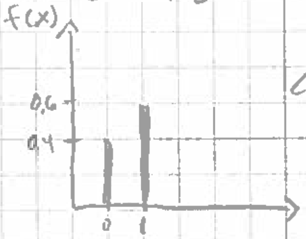


Här kan inte båda händelserna inträffa samtidigt. Den ena händelsen utesluter den andra. $P(A) + P(B)$. Ex antingen äter jag en kaka nu $P(A)$ eller så gör jag det inte $P(B)$.

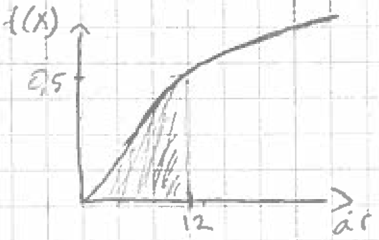
3

c.) Vid en diskret fördelning så har man ex en fördelning över antalet flickor i en 6:e klass. Antingen är man flicka eller så är man pojke. Vi säger att den procentuella andelen flickor är 60 % och de tilldelas nummer 1.

Disket fördelning



Kontinuerlig fördelning: Istället för att titta på andelen flickor och pojkar i 6 e klassen kanske man bestämmer sig för att titta på åldern istället. Här kan man se att slh är 0,5 att eleverna är under 12 år. Man tittar dock inte på sannolikheten att någon just är 12 år. Definitionen 12 år är lite luddig. Dvs en elev kanske är 12 år på dagen men hon är också ett visst antal timmar, eller minuter eller sekunder, eller millisekunder så som men att de sträcker sig mot oändligheten - så det går inte att mäta så som i en diskret fördelning.



4

d.) nominalskala, ordinalskala, intervall, kvotskala, ~~[illegible]~~.
De anger bl.a hur bearbetande mätvärden är. Där de två första skalorna inte säger särskilt mycket om datan.

1

e.) Experimentella studier är sådana som sker i en kontrollerad miljö med minst en grupp som har blivit utsatt för en s.k. behandling och en kontrollgrupp. Vilken som hamnar i vilken grupp är slumpmässigt. Exempelvis kan man göra ett experiment för att kontrollera om ett visst läkemedel har en effekt eller inte. Då får den ena gruppen behandlingen under x lång tid och den andra gruppen tror att de får behandlingen. Även vetskapen om vilken grupp som får behandlingen kan vara oviss för de som utför experimentet. Sedan jämför man resultatet från de båda grupperna med varandra när materialet är insamlat. Om det sedan skiljer sig mellan de båda grupperna kan man säga att läkemedlet har haft effekt. Man försöker ofta eliminera olika effekter som kan inverka på resultatet. Så exempelvis kan man dela in individerna i experimentet i olika par och sedan jämföra

SU, STATISTIK

Skrivsal: Ugglevikssalen Anonymkod: SG-0029 Blad nr: 2

forts. 1 e.) paren. Paren kan vara konstruerade på så sätt att dei innehåller liknande aspekter så som exempelvis ålder, vikt, utbildning, etnicitet osv beroende på vad man vill studera.

Men experimentella studier är ofta svåra att utföra och kostsamma. Därför utförs ofta observationsstudier - som kan vara kvalitativa eller kvantitativa.

Här tar man oftast ut ett urval ur en population (istället för att göra en totalundersökning på hela populationen). Dessa individer ska också väljas slumpmässigt med återläggning för att observationen ska vara oberoende (eller så har man en stor population så att ändlighetskorrektionen $(N-n)/(N-1) \approx 1$.

Vid en observationsstudie kan man exempelvis titta på inkomsten och ~~...~~ på hur mycket folk konsumerar (K) för att se om det råder ett kausalt samband.

Man kan exempelvis samla data och göra en regression för att skatta denna möjliga effekt av inkomsten där:

$\widehat{K}_i = \beta_0 + \beta_1 \text{Inkomst}_i + \varepsilon_i$ (β_0 är interceptet och β_1 är den skattade effekten av inkomsten och ε_i är en "error" term).

Också här vill man kontrollera för eventuella bakomliggande variabler/effekter som kan förvränga det skattade resultatet och därför lägga till

$$\widehat{K}_i = \beta_0 + \beta_1 \text{Inkomst}_i + \beta_n X_i + \varepsilon_i$$

där $\beta_n X_i$ står för eventuella bakomliggande variabler som kan orsaka bias i estimatorn ifall de inte tas med.

4

SU, STATISTIK

Skrivsal: Utdelareskrivsalen   Anonymkod: SG-0029   Blad nr: 3

2 a.) E = Pratar engelska $\bar{E}$ = pratar ej engelska
T = Turist $\bar{T}$ = ej turist

$T = \frac{1}{5} = 0{,}2$

$\bar{T} = 1 - 0{,}2 = 0{,}8$

	E	$\bar{E}$	
T	0,1	0,1	0,2
$\bar{T}$	0,16	0,64	0,8
	0,26	0,74	1

Söker efter P(E)

Givet att man är Turist så är slh 1/2 att man talar engelska. så:

$P(E|T) = 0{,}5$

$P(E|T) = \frac{P(E \cap T)}{P(T)} = \quad 0{,}5 = \frac{P(E \cap T)}{0{,}2}$

$P(E \cap T) = 0{,}5 \times 0{,}2 = 0{,}1$

Givet att man inte är Turist så är slh 1/5 att man talar engelska $P(E|\bar{T}) = 0{,}2$

$P(E|\bar{T}) = \frac{P(E \cap \bar{T})}{P(\bar{T})} = \quad P(E \cap \bar{T}) = 0{,}2 \times 0{,}8 = 0{,}16$

Svar: slh att en slumpvis person talar engelska är
$P(E) = P(E \cap \bar{T}) + P(E \cap T) = 0{,}1 + 0{,}16 = \underline{\underline{0{,}26}}$ 5

b.) Nya utfallsrummet = 0,26   andelen Italienare
som också talar engelska $P(\bar{T} \cap E) = 0{,}16$

$\frac{0{,}16}{0{,}26} \approx 0{,}61538$ Svar: Slh att en slumpvis vald person som talar engelska är ca 0,6154.   5

c.) $n = 16$ T = turister $T \sim Bin(n=16, p=0{,}2)$

$P(T = 8) = P(T \leq 8) - P(T \leq 7) = 0{,}99852 - 0{,}99300 =$
från tabell = 0,00552

Svar: slh är ca 0,0055.   5

d.) forts ⇒ $P(T \geq 8) = 1 - P(T \leq 7) = 1 - 0{,}9930 = 0{,}007$
från tabell

Svar: slh är 0,007 att det är minst 8 utländska turister.   5

↘

3 a.) X = kvadratmeterpris på lägenhet $X \sim N(12000, 2000^2)$.

Söker efter $P(X > 15000) \Rightarrow 1 - P\left(\frac{X-\mu}{\sigma} \leq \frac{15000-12000}{2000}\right) =$

$1 - (Z \leq 1{,}5) = 1 - \Phi(1{,}5) = 1 - 0{,}93319 = 0{,}06681$

Svar: slh är ca 0,0668

5

b.) söker efter $P(10\,000 \leq X \leq 12\,000)$

$P\left(\frac{10\,000 - 12\,000}{2000} \leq \frac{X-\mu}{\sigma} \leq \frac{12\,000 - 12\,000}{2000}\right)$

$P(\overset{A}{-1} \leq Z \leq \overset{b}{0}) = B - A \Rightarrow \Phi(0{,}00) - \Phi(-1) = \Phi(0{,}00) - (1 - \Phi(1)) =$

$\Phi(0{,}00) + \Phi(1) - 1 = 0{,}5 + 0{,}84134 - 1 = \underline{0{,}34134}$ Svar: slh är 0,34134

5

c.) $P(X > 10\,000 \mid X < 12\,000)$

$P(X \leq 10\,000)$ för att räkna ut och ta bort slh att X är mindre än eller lika med 10 000 kr/m².

$P\left(Z \leq \frac{10\,000 - 12\,000}{2000}\right) = P(Z \leq -1) = 1 - \Phi(1) =$

$1 - 0{,}84134 = 0{,}15866$

$P(X > 10000 \mid X < 12\,000) = 0{,}34134 - 0{,}15866 = 0{,}18268$

Svar: slh är $\underline{\underline{0{,}18268}}$

1

d.) $P\left(\frac{P-\mu}{\sigma} > \frac{P - 12\,000}{2000}\right) = 0{,}1$

$1 - P\left(Z < \frac{P - 12000}{2000}\right) = 0{,}1$ $\left(Z < \frac{P-12000}{2000}\right) = 0{,}9$

$0{,}9 \approx \Phi(1{,}\cancel{29})$ 8 $\Rightarrow P - 12\,000 = 2000 \times 1{,}29$

$P = 12\,000 + (1{,}2\cancel{9} \times 2000) = 14580$ 8 Svar $P = 14580$

5

SU, STATISTIK

Skrivsal: Ugglevikssalen Anonymkod: SG-0029 Blad nr: 4

4 a.) $n = 50$ $p = 0{,}03$ felaktiga enheter

X = felaktiga enheter $X \sim Bin(n=50, p=0{,}03)$

$P(X \leq 2) = P(X=0) + P(X=1) + P(X=2)$ 0,03 finns ej i tabell

$\binom{n}{x} p^x \cdot (1-p)^{n-x} = \frac{50!}{0!(50-0)!} \cdot 0{,}03^0 \cdot 0{,}97^{50} = 1 \cdot 1 \cdot 0{,}97^{50} \approx 0{,}21807$

$\frac{50!}{1!(50-1)!} \cdot 0{,}03^1 \cdot 0{,}97^{49} = 50 \times 0{,}03 \times 0{,}97^{49} \approx 0{,}33721$

$\frac{50!}{2!\,48!} \cdot 0{,}03^2 \cdot 0{,}97^{48} = 1225 \cdot 0{,}03^2 \cdot 0{,}97^{48} = 0{,}25552$

$\Rightarrow 0{,}21807 + 0{,}33721 + 0{,}25552 = 0{,}8108$

Svar: Slh är 0,8108 att ett parti accepteras.

8

b.) $E(X) = n \cdot p$ $50 \times 0{,}03 = 1{,}5$ st

Svar: ungefär 2 st

4

c.) $n = 300$ $p = 0{,}03$ $X \sim Bin(300, 0{,}03)$

$E(X) = n \cdot p = 300 \cdot 0{,}03 = 9$

Tittar ifall villkoren för normalapproximation är uppfyllda då n är väldigt stort:

$n \cdot p > 5$ då $p \leq 0{,}5$ $n \cdot (1-p) > 5$ då $p > 0{,}5$

- villkor uppfyllda då $n \cdot p = 9 > 5$

$V(X) = n \cdot p(1-p) = 9 \cdot 0{,}97 = 8{,}73$ $\sqrt{V(X)} \approx 2{,}9547$

Söker efter $P(X > 15)$ = $P(X \geq 16)$ kontinuitetskorrektion: $P(X > 14{,}5) = 1 - P(X \leq 14{,}5)$

$1 - P\left(Z \leq \frac{\cancel{14{,}5}\ 15{,}5 - 9}{\sqrt{8{,}73}}\right) = 1 - P(Z \leq 1{,}86) = 1 - \Phi(1{,}86) =$

$1 - 0{,}96856 = 0{,}03144$ Svar: Slh är 0,03144 att urvalet innehåller fler än 15 st felaktiga enheter.

6

5a.)

x	$F(x)$		$f(x)$
0	0,83	0,83 =	0,83
1	0,92	0,92 − 0,83 =	0,09
2	0,99	0,99 − 0,92 =	0,07
3	1	1 − 0,99 =	0,01

7

Slh att det finns 3 defekta komponenter är 0,01

b.)

x	$f(x)$	$x \cdot f(x)$
0	0,83	$0 \cdot 0,83 = 0$
1	0,09	$1 \cdot 0,09 = 0,09$
2	0,07	$2 \cdot 0,07 = 0,14$
3	0,01	$3 \cdot 0,01 = 0,03$

$$E(X) = \sum_{i=1}^{n} x_i \cdot f(x_i) =$$

$$0,09 + 0,14 + 0,03 = 0,26$$

3

Svar: förväntat antal defekta komponenter är 0,26 (- så knappt någon alls.)

c.) Antal defekta $= \frac{3}{10}$

y = defekt komponent utan återläggning:

$$y = 0 = \frac{7}{10} \cdot \frac{6}{9} = \frac{7}{15}$$

$$y = 1 = \left(\frac{3}{10} \cdot \frac{7}{9}\right) + \left(\frac{7}{10} \cdot \frac{3}{9}\right) = \frac{7}{15}$$

$$y = 2 = \frac{3}{10} \cdot \frac{2}{9} = \frac{1}{15}$$

y	Svar: $f(y)$:
0	7/15 ≈ 0,4667
1	7/15 ≈ 0,4667
2	1/15 ≈ 0,666

10