STOCKHOLMS UNIVERSITET
Statistiska institutionen
Jessica Franzén

OMTENTAMEN I STATISTIKENS GRUNDER 2

2017-04-18

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

Förklara följande begrepp:
a) Samplingfördelning.
b) Typ I-fel och Typ II-fel i ett hypotestest och hur de är relaterade till varandra.

c) Lista vilka komponenter en tidsserie kan bestå av och förklara respektive komponents innebörd för tidsserien.
d) Lista de olika beslutskriterierna inom beslutsteorin som faller under "beslut under osäkerhet" och förklara vad respektive kriterium går ut på.

Uppgift 2

Man vill testa batteritiden för två mobiltelefoner av olika modell. Man mäter tiden i timmar för båda telefonerna (med liknande användningsmönster) tills batterierna måste laddas.

Telefon 1:	5.9	5.5	6.8	6.4	7.0	6.6	7.7	6.9	6.2
Telefon 2:	5.3	5.6	5.1	6.3	5.8	5.7	7.2	7.0	6.1

a) Är batteritiden för Telefon 1 längre än batteritiden för Telefon 2? Utför ett lämpligt test på signinfikansnivån 0.05.

b) Man vill göra en mer nogrann undersökning av batteritiden för de två telefonmodellerna. Man mäter därför batteritiden enligt ovan men den här gången för 50 slumpmässigt valda telefoner av modell 1 och 35 stycken slumpmässigt vakda telefoner av modell 2. Medelvärdet och variansen för de två telefonmodellerna visas i tabellen. Utför ett test för att testa om batteritiden för Telefonmodell 1 är längre än batteritiden för Telefonmodell 2 genom att räkna ut exakt p-värde och kommentera resultatet. Ange också vilka antaganden du gör.

	Medelvärde	Varians	Antal
Telefon 1	6.504	0.452	50
Telefon 2	6.201	0.497	35

Uppgift 3
En kedja med matbutiker undersöker inställningen till den nya sortens självbetjäningskassor genom att tillfråga 802 slumpmässigt utvalda kunder. Av dessa har 424 angett att de föredrar traditionella kassor.

a) Testa om proportionen som föredar traditionella kassor är större än 50% på signifikansnivån 10%.

b) Beräkna testets p-värde.

c) Avgör om följande allmänna påståenden om signifikansnivå och p-värde är korrekta eller ej. Motivera ditt svar:

i) Om en nollhypotes förkastas på signifikansnivån 0.05 måste den även förkastas på signifikansnivån 0.01.

ii) Om en nollhypotest kan förkastas på signifikansnivån 0.025 men inte på signifikansnivån 0.01 innebär det att p-värdet är mellan 0.01 och 0.025.

Uppgift 4
Till 650 slumpmässigt utvalda kunder i olika inkomstkategorier som alla har bostadslån i en viss bank ställde man i en enkät frågan:

Klarar du av en räntehöjning i storleksordningen en dubblering av räntan?

1) Ja, utan förändringar i levnadsstandard
2) Ja, men med sänkningar av levnadsstandard
3) Nej, blir tvungen att flytta till billigare boende

Man valde att slumpmässigt välja ut 200 höginkomstattagare, 400 medelinkomsttagare och 50 låginkomsttagare. Svaren sammanställdes i följande tabell:

	Höginkomsttagare	Medelinkomsttagare	Låginkomstagare
Ja, utan förändringar	29%	18%	38%
Ja, med sänkning av lev. standard.	70%	79%	60%
Nej	1%	3%	2%
Antal	200	400	50

Banken har en hypotes om att inkomstnivån inte påverkar förmågan att hantera en kraftig räntehöjning, då man antar att personer med hög inkomst anpassat sina utgifter till inkomsterna och därmed är minst lika sårbara som personer med lägre inkomst. Undersök med ett lämpligt test om det finns några skillnader i hur man klarar av en räntehöjning beroende på inkomstnivå. Använd signifikansnivån 5%.

Uppgift 5
En bild säger mer än 1000 ord säger ordspråket. För att utreda sanningshalten i detta lät man 40 slumpmässigt utvalda personer betrakta en bild under 60 minuter och under tiden skriva ner de ord som bilden fick dem att relatera till. Medelvärdet blev 1238 ord och standardavvikelsen 304 ord.

a) Undersök om ordspråket är sant genom att bilda ett lämpligt 99%-igt konfidensintervall.

b) Hur stort urval ska man dra för att bredden på det 99%-iga konfidensintervall ska vara högst 200 ord?

c) Om man för ett givet konfidensintervall ökar både konfidensgraden och urvalstorleken, kan man då säga vad som händer med bredden på konfidensintervallet? Motivera.

Statistiska institutionen

Stockholms universitet

Rättningsblad

Datum: 18/4-2017

Sal: Brunnsvikssalen

Tenta: Statistikens grunder 2

Kurs: Statistikens grunder

ANONYMKOD: SGO-0014

☒ 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					5
Lär.ant. 19	11	20	19	17					

POÄNG	BETYG	Lärarens sign.
86	B	JF

SU, STATISTIK

Skrivsal: Brunnsvikssalen Anonymkod: S40.com Blad nr: 1

1. a) Vid varje stickprov kommer olika värde vi den få dvs urvalskarakteristika. Urvalskarakteristikan kan sålades beskrivas som en stokastisk variabel. Den sannolikhetsfördelning som denna stokastiska variabel antar är vad vi kallar samplingfördelning.

5

b) Typ 1-fel är då man förkastar nollhypotesen då den i själva verket är sann. α betecknar sth att begå typ I fel.
Typ 2-fel är då man inte förkastar nollhypotesen fast den är falsk. β betecknar sannolikheten att begå typ II fel.

4

$\alpha = P(\bar{X} \leq x \mid H_0 \text{ är sann})$

$\beta = P(\bar{X} \geq x \mid H_0 \text{ är falsk}) = 1 - P(\bar{X} \leq x \mid H_0 \text{ är falsk})$

c) Trend → Anger den långsiktiga utvecklingen för tidserien.

Konjunktur → Återkommande variationer i tidserien som inte kan förutspås

Säsong → Återkommande variationer som kan förutspås på förhand.

Slump → Helt och hållet slumpmässiga variationer som ej är återkommande och som inte kan förutspås.

5

d) <u>Maximin</u>: Välj det handlingsalternativ som maximerar nyttan i det sämsta utfallet.

<u>Maximax</u>: Välj det handlingsalternativ som maximerar nyttan i det bästa utfallet.

<u>Minimax regret</u>: Välj det handlingsalternativ som minimerar skillnaden mellan det bästa och det sämsta utfallet givet handlingen

<u>Laplace</u>: Då vi inte vet slh för respektive utfall så tillskriver vi dem lika stora slh. $\left(\frac{1}{n}\right)$ och maximerar sedan nyttan givet dessa sannolikheter.

5

SU, STATISTIK

Skrivsal: Brunnsviksalen Anonymkod: SGO-0014 Blad nr: 2

2.

a)

Telefon 1	5,9	5,5	6,8	6,4	7,0	6,6	7,7	6,9	6,2
Telefon 2	5,3	5,6	5,1	6,3	5,8	5,7	7,2	7,0	6,1
D	0,6	-0,1	1,7	0,1	1,2	0,9	0,5	-0,1	0,1

* Parvisa beroende observationer (upprepade försök på samma två telefoner)
* $n < 30 \rightarrow$ t-test Nej!

Fel test!

$H_0: \mu_1 = \mu_2$

$H_1: \mu_1 > \mu_2$ (ensidigt test)

Signifikansnivå: 5%

Beslutsregel: Förkasta H_0 om $T_{obs} > T_{krit}$

$$T_{obs} = \frac{\bar{D} - D_0^{(=0)}}{S_D/\sqrt{n}} = \frac{0,544}{0,7379/\sqrt{9}} = 2,21$$

$$S_D^2 = \frac{1}{1-n}\sum(D_i - \bar{D})^2 = \frac{4,9 - 0,544}{8} = 0,545$$

$$S_D = \sqrt{0,545} = 0,7379$$

$T_{krit}(8) = 1,860$

$T_{obs} > T_{krit}$

3

Svar: Vi kan förkasta H_0 på 5% signifikansnivå. Telefon 1 har längre batteritid av dessa två.

b) ✱ Oberoende parvisa observationer! N_{C1}

- 50 slumpmässigt valda telefoner av modell 1
- 35 slumpmässigt valda telefoner av modell 2.

✱ Antaganden:
- Slumpmässiga observationer
- Normalfördelning

$n_1 > 30$, $n_2 > 30$ } n.f. enligt CGS.

	Medelvärdet	Varians	n
Telefon 1:	6,504	0,452	50
Telefon 2:	6,201	0,497	35

H_0: $\mu_1 = \mu_2$ (ensidigt test)

H_A: $\mu_1 > \mu_2$

$$Z = \frac{\bar{X} - \bar{Y} - D_0^{(=0)}}{\sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}} = \frac{6,504 - 6,201}{\sqrt{\frac{0,452}{50} + \frac{0,497}{35}}} \approx \cancel{1,30}\ 1,99$$

$$P(Z > 1,30 \mid H_0 \text{ är sann}) = 1 - P(Z < 1,30) = 1 - \phi(1,30)$$
$$= 1 - 0,90320 = 0,0968$$

0,02

Svar: P-värdet blev 0,0968 vilket innebär att skillnaden i batteritid är statistiskt signifikant på 9,68% nivå. Det innebär att slh att vi får ett extremare värde än det vi observerade givet att H_0 är sann är 9,68%.
På 10% signifikansnivå kan vi således förkasta H_0.
Telefoner av modell 1 har längre batteritid än telefoner av modell 2 enligt dessa data.

8

SU, STATISTIK

Skrivsal: Brunnsvikssalen Anonymkod: SG0014 Blad nr: 3

3.

$n = 802$

a) $P = \frac{424}{802} \approx 0{,}528 \approx 52{,}8\%$

$H_0: \quad \pi_0 = 0{,}5$ (Ensidigt test)

$H_A: \quad \pi_A > 0{,}5$

Signifikansnivå: 10%

Beslutsregel: Förkasta H_0 om $Z_{obs} > Z_{krit}$

Uträkning av Z_{obs}: $Z = \frac{P - \pi_0}{\sqrt{\frac{\pi_0(1-\pi_0)}{n}}} = \frac{0{,}528 - 0{,}5}{\sqrt{\frac{0{,}528 \cdot 0{,}471}{802}}} = 1{,}689$

$Z_{krit} = Z_\alpha = Z_{0,1} = 1{,}2816$

$Z_{obs} > Z_{krit}$

Svar: Vi kan förkasta H_0 på 10% signifikansnivå. En majoritet föredrar traditionella klassar.

b) P-värde: $P(Z > Z_{obs} \mid H_0 \text{ är sann})$

$\rightarrow P(Z > 1{,}689 \mid \pi_0 = 0{,}5)$

$= P(Z > 1{,}682) = 1 - P(Z < 1{,}682) = 1 - \Phi(1{,}682)$

$\approx 1 - \Phi(1{,}68) = 1 - 0{,}95352 = 0{,}04648$

$\approx 4{,}6\%$

Svar: P-värdet är 4,6%

r r r u r

7

3

c) i) Nej. Något som är statistiskt signifikant på 10% behöver ingalunda vara det på 5%.
Exempelvis måste testvariabeln (Z) för normalfördelningar vara över 1,2816 för att vara signifikant på 10% och över 1,6449 för att vara signifikant på 5%. Vi inser därför att Z_{obs} kan vara mellan 1,2816 och 1,6449 och därmed vara signifikant på 10% nivå men inte på 5% nivå.

(5)

ii) Ja. Det illustreras i deluppgift a.
P-värdet ger den lägsta signifikansnivån då vi kan förkasta H_0. Så om vi vet att något är signifikant på 5% nivå kan vi räkna ut sannolikheten att få ett extremare värde än det vi observerade givet att H_0 är sann och det bör ligga mellan vad vi vet är signifikant och vad vi vet är icke-signifikant.

(5)

SU, STATISTIK

Skrivsal: Brunnsvikssalen   Anonymkod: SCO-0014   Blad nr: 4

4. Först räknas procentsatserna om till antal. Sedan slås Låg/medelinkomstklassen ihop så att $E(n) > 5$. Därefter kan vi genomföra ett homogenitetstest genom att räkna ut χ^2_{obs}.

Du slår ihop "Ja med förändring" och "Nej". Du har gjort detta rätt nedan men vänt på beteckningarna.

H_0: Fördelning av svar skiljer sig inte mellan inkomstgrupper

H_1: Fördelningen skiljer sig åt.

R

Sig. nivå: 5%

Beslutsregel: Förkasta H_0 om $\chi^2_{obs} > \chi^2_{krit}(k_1-1)(k_2-1)$

Obs.	Ja, utan förändringar	Ja, med förändringar	Nej	tot
Höginkomst	58	72	19	149
Låg/medelinkomst	142	328	31	501
	$\frac{200}{650}$	$\frac{400}{650}$	$\frac{50}{650}$	

E(n)	Ja, utan...	Ja, med...	Nej
Höginkomst	$\frac{200}{650} \cdot 149 = 45$	$\frac{400}{650} \cdot 149 = 91$	$\frac{50}{650} \cdot 149 = 11$
Låg/medelinkomst	$\frac{200}{650} \cdot 501 = 154$	$\frac{400}{650} \cdot 501 = 308$	$\frac{50}{650} \cdot 501 = 38$

$$\chi^2_{obs} = \sum \frac{(n - E(n))^2}{E(n)}$$

Vänd!

$$\chi^2_{obs} = \frac{(58-45)^2}{45} + \frac{(72-91)^2}{91} + \frac{(19-11)^2}{11} + \frac{(142-154)^2}{154} + \frac{(328-308)^2}{308}$$

$$+ \frac{(31-38)^2}{38} \approx 16{,}991$$

$$\chi_{krit}(2-1)(3-1) = \chi^2_{krit}(2) = 5{,}991$$

Svar: Vi kan förkasta nollhypotesen på 5% nivå. Fördelningen av svar skiljer sig åt mellan ålderskategrupper. Därmed är det troligt att bankernas hypotes inte stämmer.

19

SU, STATISTIK

Skrivsal: Brunnsviksalen Anonymkod: SG0-0014 Blad nr: 5

5. X = "Antal nedskrivna ord"

$n > 30$

→ X approximativt n.f enligt CGS.

$X \sim N(0,1)$

$n = 40$

$\bar{x} = 1238$

$S = 304$

a) 99% KI: $\bar{x} \pm z_{0,01/2} \cdot \frac{S}{\sqrt{n}}$

$= 1238 \pm 2,5758 \cdot \frac{304}{\sqrt{40}}$

$= 1238 \pm 123,8$

$[1114,2 \ ; \ 1361,8]$

Är ordspråket sant?

Bredd $= \underbrace{2 \cdot z_{\alpha/2} \cdot \frac{S}{\sqrt{n}}}_{B} = 2B$ 6

b) $2B = 200$

$B = 100$

→ $100 = 2,5758 \cdot \frac{304}{\sqrt{n}}$

$100^2 = 2,5758^2 \cdot \frac{304^2}{n}$

$n = \frac{2,5758^2 \cdot 304^2}{100^2} = 15,8$ 61,32

?

Svar: n måste vara 16 för att konfidensintervallets bredd ska vara högst 200.

c) Nej. Ökar man konfidensgraden så ökar bredden på intervallet. ~~Att~~ öka urvalet har emellertid motsatt effekt, med större urval minskar bredden på intervallet.

1. $\bar{x} \pm \left(z_{\alpha/2}\uparrow \cdot \frac{s}{\sqrt{n}}\right)\uparrow$

2. $\bar{x} \pm \left(z_{\alpha/2} \cdot \frac{s}{\sqrt{n}\uparrow}\right)\downarrow$

Huruvida intervallet blir bredare eller ej beror på vilken ökning som är störst.

4