STOCKHOLMS UNIVERSITET
Statistiska institutionen
Jessica Franzén

TENTAMEN I STATISTIKENS GRUNDER 2

2018-03-16

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. Samtliga formler som används för beräkningar ska anges tydligt.

Uppgift 1

a) Du tar ett slumpmässigt urval och räknar ut ett 95%-igt konfidensintervall för populationsmedelvärdet μ. Vilket eller vilka av följande värden kommer garanterat ligga inom det uträknade intervallet? Motivera.
i) 0 ii) μ iii) $\overline{x}$ iv) 1.96

b) Vid ett hypotestest får du ett p-värde på 0.003. Motivera varför eller varför inte respektive påstående är sant eller falskt.
i) Vi förkastar H_0 på alla rimliga signifikansnivåer.
ii) Ett p-värde anger sannolikheten att nollhypotesen är sann.
iii) Ett p-värde måste anta ett värde mellan 0 och 1.

c) Vad betyder väntevärdesriktig skattning? Ge ett exempel på en väntevärdesriktig skattning och visa att den är väntevärdesriktig.

d) Förklara begreppet ändlighetskorrektion samt när och varför den används.

Uppgift 2

En däcktillverkare har intresse av att uppskatta den genomsnittliga sträckan som den senaste modellens vinterdäck håller innan de slits ut. 6 slumpmässigt utvalda däck har testkörts tills de blivit utslitna med följande resultat i tusentals kilometer:

51.2 52.8 44.8 59.2 46.4 48.0

a) Antag att $\sigma = 5.5$ och beräkna ett 95%-igt konfidensintervall för den genomsnittliga sträckan som däcken håller. Redogör för eventuella nödvändiga antaganden.

b) Hur stort urval av däck skulle krävas för att längden av ett 95%-igt konfidensintervall skulle bli högst 6 (tusen kilometer).

Uppgift 3

Forskare har studerat sambandet mellan graviditetslängd och utbildningsnivå på de födda barnen senare i livet. Data i tabellen kommer från den aktuella studien och består av 4953 slumpmässigt valda graviditeter uppdelade på de födda barnens utbildningsnivå som vuxna.

	Graviditetslängd (veckor)					
	22-27	28-32	33-36	37-42	>43	Totalt
Ej gymnasieutbildning	14	34	140	1010	81	1279
Gymnasieutbildning eller högre	26	65	343	3032	208	3674

a) Testa med lämpligt test om det råder ett beroende mellan graviditetslängd och utbildningsnivå på de födda barnen som vuxna. Använd signifikansnivån 5 procent.

b) De flesta graviditeter varar mellan 37-42 veckor. Testa med lämpligt test om andelen för tidigt födda (graviditetslängd 36 veckor eller kortare) är större än 0.1. Sök p-värdet och tolka resultatet.

Uppgift 4

En läkare mäter blodtryck under vila på 10 slumpmässigt utvalda patienter med diabetes, en sjukdom som ofta ger förhöjt blodtryck. Patienterna behandlas sedan med en kognitiv terapi under en månad varpå blodtrycket mäts på nytt under vila. Följande resultat erhålles:

Patient	1	2	3	4	5	6	7	8	9	10
Blodtryck före	153	148	139	126	149	135	138	132	121	135
Blodtryck efter	148	138	140	121	141	120	131	126	121	140

a) Testa med lämpligt test hypotesen att terapin har någon blodtryckssänkande effekt på signifikansnivån 5 procent. Ange eventuella nödvändiga antaganden.

b) Man vill testa samma hypotes som i a) d.v.s. om terapin har någon blodtryckssänkande effekt. Denna gång jämför man dock en grupp om 10 slumpmässigt valda personer (med diabetes) som inte genomgått terapin med 10 helt andra slumpmässigt valda personer (med diabetes) som har genomgått terapin. Ska man använda samma testvariabel som i a)? Motivera. Om inte, ange den nya testvariabeln och vad som skiljer de två testvariablerna åt och varför.

Uppgift 5

Världsmarknadspriset i kronor/ton för stål under de senaste 6 månaderna ges i tabellen.

	Okt 2017	Nov 2017	Dec 2017	Jan 2018	Feb 2018	Mars 2018
Pris (kr/ton)	680	700	675	715	730	740

a) Räkna ut index för priserna de senaste 6 månaderna med mars 2018 som basmånad.

Man skattar att index för stålpriset i april 2018 uttryckt med basmånad mars 2018, blir 95, 105 eller 110 med sannolikheterna 0.3, 0.5 och 0.2

b) Beräkna förväntat pris på stål i april 2018 baserat på sannolikhetsskattningarna av april månads prisindex.

Ett bolag som bedriver handel på stålmarkanden överväger tre olika investeringsalternativ; 1. Investera 100 miljoner i köp av stål, 2. Inte göra någon investering alls, 3. Investera 50 miljoner i köp av stål. Företagets vinst (eller förlust) i miljoner kronor för de tre olika indexutfallen och de tre möjliga investeringsalternativen, framgår av tabellen.

	Utfall Index		
	95	105	110
Alt 1	-5	5	10
Alt 2	0	0	0
Alt 3	-2.5	2,5	5

c) Antag att man inte vet sannolikhetsfördelningen för indexutfallet. Hjälp företaget att välja investeringsalternativ med minimax-regret kriteriet.

Statistiska institutionen

Stockholms universitet

Rättningsblad

Datum: 16/3 2018

Sal: Ugglevikssalen

Tenta: Statistikens grunder 2

Kurs: Statistikens grunder

ANONYMKOD: 0015-UNS

☒ 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					9
Lär.ant. 18	18	19	15	20					

POÄNG	BETYG	Lärarens sign.
90	A	GJF

SU, STATISTIK

Skrivsal: Uggleviks salen Anonymkod: 0015-UNS Blad nr: 1

Uppgift 1

a.) Det ända vär det av de 4 alternativet som garanterat kommer att finnas inom intervallet är $\bar{x}$. μ Kommer att göra det i ca 95% av fallen om man upprepar försöket under samma förutsättningar många gånger. $\bar{x}$ Kommer ligga i mitten av intervallet. $\bar{x} \pm \alpha_{/2} \frac{\sigma}{\sqrt{n}}$

2

b.) i.) Sant, förutsatt att vi bedömer alla rimliga signifikansnivåer som allt över 0,3% och vanligtvis är 1% det lägsta.

2

ii.) Sant, P-värde anger sannolikheten att få det observerade värdet eller ett ännu osannolikare/extremare värde, om H_0 är sant. Dvs ett p-värde på ~~ter~~ 0,003 motsvarar 0,3% s/h att få det observerade värdet om H_0 är sant.

V 0

iii.) Sant, eftersom att p-värdet är sannolikheten för ett utfall så måste det lägsta värdet vara 0 och det högsta 1.

2

SU, STATISTIK

Skrivsal: Uggleviksssalen Anonymkod: 0015-UNS Blad nr: 2

c.) Väntevärdesriktig skattning innebär att det förväntade värde för $\bar{x}$ är det sanna μ.

$E(X)=\mu$

$E(\bar{X})=E\left(\frac{\sum X_i}{n}\right)=\frac{1}{\cancel{n}}\cdot\cancel{n}E(X)=E(X)=\mu$

6

d.) Ändlighetskorrektion behöver användas när man ska beräkna varians för en ändlig population utan återläggning. Formeln är då $\frac{N-n}{N-1}$

Man behöver ej använda ändlighetskorrektion om $\frac{N-n}{N-1}>0{,}95$ eftersom att påverkan av att det är utan återläggning då är liten.

6

SU, STATISTIK

Skrivsal: Uggleviksalen Anonymkod: 0015-UNS Blad nr: 3

Uppgift 2 Information: $n=6$

Urval: 51,2 52,8 44,8 59,2 46,4 48,0

a.) $\sigma=5,5$

Antaganden: Oberoende observationer n.f. X

Eftersom σ är känd så använder vi oss av Z. Dvs vi använde Normal approximering $N(0,1)$ enligt CGS.

$$\bar{x}=\frac{\sum x_i}{n}=\frac{302,4}{6}=50,4$$

95% K.I $\bar{x}\pm z_{\alpha/2}\frac{\sigma}{\sqrt{n}} = 50,4\pm 1,96\frac{5,5}{\sqrt{6}}$

$50,4\pm 4,4$

$\boxed{\text{Svar: 95\% K.I = 46 till 54,8}}$

8

b.) $2B$ = Hela K.I B = Halva K.I

$B=z_{\alpha/2}\frac{\sigma}{\sqrt{n}}$ $\sqrt{n}=z_{\alpha/2}\frac{\sigma}{B}$

$3=1,96\frac{5,5}{\sqrt{n}}$ $\sqrt{n}=1,96\frac{5,5}{3}=3,5933$ $\sqrt{n}^2=n=12,9118$

$\boxed{\text{Svar: Vi behöver 13 observationer för att längden av ett 95\%-igt konfidensintervall ska bli högst 6 (tusen km).}}$

10

SU, STATISTIK

Skrivsal: Ugglevikssalen Anonymkod: 0015-UNS Blad nr: 4

Uppgift 3 Information: $n = 4953$

	Graviditetslängd (veckor) 22-27	28-32	33-36	37-42	>43	Tot
Ej gymnasieutb	14	34	140	1010	81	1274
Gymnasieutb eller högre	26	65	343	3032	208	3674
	40	99	483	4042	289	4953

a.) X^2-oberoende test

H_0: Oberoende råder

H_A: Beroende råder

$5-1 \cdot 2-1 = 4 \cdot 1 = 4$

H_0 förkastas om $X^2_{obs} > X^{2\,(4)}_{0,05} = 9,488$

Vi ser att alla celler har ett värde >5.

$E(n_i)$

	22-27	28-32	33-36	37-42	>43	Tot
Ej gymn. utb	10,33	25,56	124,72	1043,75	74,63	1274
Gymn. utb eller högre	29,67	73,44	358,28	2998,25	214,37	3674
	40	99	483	4042	289	4953

För att räkna ut förväntad värde för varje cell så räknar vi ((marginalvärde$_1$/tot population) · (marginalvärde$_2$/tot pop) · total population). Ex för första cellen $4953 \cdot (1274/4953) \cdot (40/4953) = 10,33$. Vi gör detta för alla celler.

För att få vårt X^2_{obs} så summerar vi (differensen i kvadrat och delar på förväntat värde) $\sum \frac{(n_i - E(n_i))^2}{E(n_i)}$

$X^2_{obs} = 1,3 + 2,77 + 1,87 + 1,09 + 0,54 + 0,45 + 0,97 + 0,65 + 0,38 + 0,19 = 10,21$

Svar: $X^2_{obs} = 10,21 > 9,488$. Dvs H_0 förkastas och vi kan påvisa att det faktiskt råder ett beroende.

10

SU, STATISTIK

Skrivsal: Ugglevikssalen Anonymkod: 0015-UNS Blad nr: 5

b)

H_0: Andel för tidigt födda $= 0{,}1$

H_A: Andel för tidigt födda $> 0{,}1$

Testvariabel $Z = \dfrac{P - H_0}{\sqrt{P(1-P)/n}}$ approx $N(0,1)$ CGS

$\pi(1-\pi)$ om H_0 är sann

H_0 förkastas om Z_{obs} har ett väldigt lågt p-värde eftersom att vi ej har en angiven signifikansnivå

Urvalsproportioner:

För tidigt födda $= \dfrac{40+99+483}{4953} = 0{,}12558$

Övriga $= \dfrac{4042+289}{4953} = 0{,}87442$

$P = 0{,}12558$ $n \cdot P = 619 > 5$ $n \cdot (1-P) > n \cdot P$ ✓

$P(1-P) = 0{,}1098$

P-värde: $1 - \Phi(5{,}43) \approx 0$ ✓

$$\frac{0{,}12558 - 0{,}1}{\sqrt{0{,}1098/4953}} = 5{,}43$$

Svar: H_0 förkastas eftersom att P-värdet är ungefär 0. Dvs sannolikheten att få vårat observerade värde eller något ännu mer osannolikt är ca 0%.
Vi kan alltså påvisa att andelen för tidigt födda är större än 0,1.

9

SU, STATISTIK

Skrivsal: Ugglevikssalen Anonymkod: 0015-UNS Blad nr: 6

Uppgift 4 Information: $n=10$

Patient	1	2	3	4	5	6	7	8	9	10
Blodtryck före	153	148	139	126	149	135	138	132	121	135
Blodtryck efter	148	138	140	121	141	120	131	126	121	140

a) Hypotestest ~~för~~ parvisa observationer

$H_0: \cancel{\mu_D = 0}$ $\mu_x = \mu_y$ Sig. Nivå: 5%

$H_A: \cancel{\mu_D < 0}$ $\mu_x > \mu_y$ Antagandena?

H_0 förkastas om $t_{obs} < t_{0,05}^{(9)} = -1,833$ (10−1)

Testvariabel:

$$T = \frac{\bar{D}}{\sqrt{s_D^2/n}} \quad t(n-1)$$

$$\bar{D} = \frac{\sum d_i}{n}$$

$$s_D^2 = \frac{\sum d_i^2 - n\bar{D}^2}{n-1} = \frac{550 - 10 \cdot -5,5^2}{10-1} = 27,5$$

	1	2	3	4	5	6	7	8	9	10	Tot
d	-5	-10	1	-5	-8	-15	-7	-6	0	5	-55
d^2	25	100	1	25	64	225	49	36	0	25	550

$\bar{D}$: $-55/10 = -5,5$

$$\frac{-5,5}{\sqrt{27,5/10}} = -3,3166$$

Svar:
H_0 förkastas eftersom $-3,3166 < -1,833$.
Dvs, vi kan påvisa att terapin har en blodtryckssänkande effekt.

9

SU, STATISTIK

Skrivsal: Uggleviktssalen Anonymkod: 0015-UNS Blad nr: 7

b.) Eftersom att observationerna nu istället är 2 st oberoende variabler så behöver vi justera vårt test och testvariabel. Vi behöver även anta att båda variabler har samma varians. dvs $\sigma_1^2 = \sigma_2^2$ 4

Testvariabel i uppgift a:

$$t = \frac{\bar{D}}{\sqrt{S_D^2/n}} \qquad t^{(n-1)}$$

Den nya testvariabeln blir istället:

$$t = \frac{\bar{X}_1 - \bar{X}_2}{\sqrt{S_p^2\left(\frac{1}{n_1} + \frac{1}{n_2}\right)}} \qquad t^{(n_1+n_2-2)}$$

p betyder här poolad.

Dvs, S_p^2 är den poolade variansen

$$S_p^2 = \frac{(n_1-1)\cdot S_1^2 + (n_2-1)S_2^2}{n+m-2}$$

Skillnaden är att vi inte bara bryr oss om differensen utan här jämför hela grupperna.

6

SU, STATISTIK

Skrivsal: Ugglevikssalen Anonymkod: 0015-UNS Blad nr: 8

Uppgift 5 Information:

	10/17	11/17	12/17	01/18	02/18	03/18
Pris (kr/ton)	680	700	675	715	730	740

a.) Basår = 03/18

Svar:

Index med 03/18 som basår (2 decimaler):

	Index	
10/17	91,89	$\frac{680}{740} \cdot 100$
11/17	94,59	$\frac{700}{740} \cdot 100$
12/17	91,22	$\frac{675}{740} \cdot 100$
01/18	96,62	$\frac{715}{740} \cdot 100$
02/18	98,65	$\frac{730}{740} \cdot 100$
03/18	100	Basår $\frac{740}{740} \cdot 100$

$\frac{\text{Måt år}}{\text{Bas år}} \cdot 100$

4

b.) April 2018 stålpris

03/18 = Bas = 740

Sannolikheter: 95 = 0,3 105 = 0,5 110 = 0,2

$(740 \cdot 0{,}95 \cdot 03) + (740 \cdot 1{,}05 \cdot 0{,}5) + (740 \cdot 0{,}10 \cdot 0{,}2) =$ Förväntat stålpris för april 2018

$= 762{,}2$

$\boxed{\text{Svar: Förväntat stålpris för april 2018 är 762,2 kr}}$

8

SU, STATISTIK

Skrivsal: Uggleviksssalen **Anonymkod:** 0015-UNS **Blad nr:** 9

c.)

	95	105	110	regret			Är
Alt 1	-5	5	10	5	0	0	5
Alt 2	0	0	0	0	5	10	10
Alt 3	-2,5	2,5	5	2,5	2,5	5	5
Max	0	5	10				

Svar: En ligt minimax-regret kriteriet borde företaget välja alt 1 eller alt 2 där max regret är 5. Eftersom att regret är 0 för de andra naturtillstånden vid alt 1 är det däremot bättre och bör rekommenderas före alt 2.

8