STOCKHOLMS UNIVERSITET
Statistiska institutionen
Ulf Högnäs (examinator)

VT 2020

TENTAMEN I GRUNDLÄGGANDE STATISTIK FÖR EKONOMER

2020-02-13

Skrivtid:	kl. 16.00 - 21.00
Godkända hjälpmedel:	Miniräknare utan lagrade formler och text
Bifogade hjälpmedel:	Häftet *Formelsamling och Tabeller över statistiska fördelningar* (återlämnas efter skrivningen)

- Tentamen består av 7 uppgifter, i förekommande fall uppdelade i deluppgifter. Maximalt antal poäng anges per deluppgift.

- **Uppgift 1 – 5**: Svar lämnas på särskild **SVARSBILAGA**,
  - Totalt 12 flervalsfrågor där ett av fem alternativ är korrekt svar.
  - Har fler än ett svarsalternativ markerats för en deluppgift ges noll poäng.
  - Uträkningar lämnas ej in för dessa, om uträkningar ändå lämnas in kommer de inte att beaktas vid bedömningen.

- **Uppgift 6 – 7**: Svar med **FULLSTÄNDIGA REDOVISNINGAR** ska lämnas in.
  - Använd endast skrivpapper som tillhandahålls i skrivsalen.
  - För full poäng på en uppgift krävs tydliga, utförliga och väl motiverade lösningar.
  - Kontrollera alltid dina beräkningar och lösningar! Slarvfel kan också ge poängavdrag!

- Tentamen kan maximalt ge 60 + 40 = 100 poäng och för godkänt resultat krävs minst 50.

- Betygsgränser:

A:	90 – 100 p
B:	80 – 89 p
C:	70 – 79 p
D:	60 – 69 p
E:	50 – 59 p
Fx:	40 – 49 p
F:	0 – 40 p

OBS! Fx och F är underkända betyg som kräver omexamination. Studenter som får betyget Fx kan alltså inte komplettera för högre betyg.

LYCKA TILL!

Uppgift 1

a) Tabellen nedan visar tio femåringars spargrisbesparingar i kronor. **Beräkna första och tredje kvartilen, enligt denna kurs metod.** (5p)

44	61	34	3	17	26	19	87	23	86

(A) $Q_1 = 16.75$; $Q_3 = 67.25$

(B) $Q_1 = 16.75$; $Q_3 = 67.75$

(C) $Q_1 = 18.25$; $Q_3 = 67.25$

(D) $Q_1 = 18.50$; $Q_3 = 67.25$

(E) $Q_1 = 18.50$; $Q_3 = 67.75$

b) Tabellen nedan visar simultanfördelningen av två variabler x and y, som kommer från ett stickprov av storlek $n = 10$:

Frekvens	$y = 0$	$y = 1$
$x = 0$	4	1
$x = 1$	4	1

Beräkna kovariansen mellan x och y. (5p)

(A) -0.222

(B) -0.200

(C) 0

(D) 0.200

(E) 0.222

Uppgift 2

a) Den Europeiska unionen beställde en studie över hur ofta dess invånare, i åldrarna 15 år och äldre, är fysiskt aktiva (B_1 = *Ofta;* B_2 = *Ibland;* B_3 = *Sällan eller Aldrig*). Resultaten delades också in i två åldersgrupper (A = *15-24 år gammal,* $\bar{A}$ = *25 och äldre*). En person väljs slumpmässigt; de simultana sannolikheterna för var och en av dem sex kombinationerna presenteras i tabellen nedan. Notera att B_1, B_2 och B_3 är parvis disjunkta och att $\bar{A}$ är komplementet av A.

	B_1	B_2	B_3
A	0.03	0.05	0.08
$\bar{A}$	0.12	0.21	0.51

Vilket av följande påståenden är falskt? (5p)

(A) $P(\bar{A}) = 0.84$

(B) $P(B_1 \mid A) = 0.20$

(C) $P(B_2 \mid \bar{A}) = 0.25$

(D) $P(B_2 \cup \bar{A}) = 0.89$

(E) $P(A \cap \overline{B_2}) = 0.11$

b) Tabellen nedan beskriver en slumpvariabel X. **Beräkna väntevärdet μ och standardavvikelsen σ för X.** (5p)

x	-2	-1	0	1	2
$P(X = x)$	0.2	0.3	0.2	0.1	0.2

(A) $\mu = -0.2\,;\, \sigma = 1.2$

(B) $\mu = -0.2\,;\, \sigma = 1.4$

(C) $\mu = -0.4;\, \sigma = 1.4$

(D) $\mu = -0.2\,;\, \sigma = 1.96$

(E) $\mu = -0.4\,;\, \sigma = 1.96$

Uppgift 3

a) Som en del av en undersökning får personer i ett oberoende och lika fördelat slumpmässigt urval av 20 Svenskar frågan "Såg du TV-programmet *Eurovision Song Contest* förra året?" Antag att den verkliga proportionen av svenskar som såg programmet förra året är 25% och att varje person svarar sanningsenligt. **Vad är sannolikheten att <u>8 eller fler</u> (av dem 20 som tillfrågats) svarar "ja"?** Välj det alternativ som ligger närmast ditt svar. (5p)

(A) 1%

(B) 3%

(C) 4%

(D) 10%

(E) 96%

b) Två studenter, Anna and Bea, skriver en tenta. Antag att deras respektive tentamenspoäng kan beskrivas med två normalfördelade slumpvariabler, X and Y:

Anna: $X \sim N(70, 6^2)$
Bea: $Y \sim N(60, 8^2)$

Antag att X och Y är oberoende. **Beräkna sannolikheten att Anna får högre poäng än Bea, enligt denna modell.** Välj det alternativ som ligger närmast ditt svar. (5p)

(A) 16%

(B) 24%

(C) 76%

(D) 79%

(E) 84%

c) Under en månad får E-handlaren *Pebbles'R'Us* 600 unika besök. Varje unikt besök resulterar i ett köp med sannolikhet 2%. Antag att besöken är statistiskt oberoende och att vart och ett besök leder till som mest ett köp under månaden. **Beräkna den approximativa sannolikheten att de 600 unika besöken leder till <u>20 eller fler</u> köp** Välj det alternativ som ligger närmast ditt svar. Ledtråd: ett köp kan ses som "lyckat försök". (5p)

(A) 0%

(B) 1%

(C) 5%

(D) 10%

(E) 25%

Uppgift 4

a) I ett oberoende och lika fördelat stickprov av 900 tyska skolbarn fann man att 90 av barnen var vänsterhänta. **Baserat på detta stickprov, vilket av följande interval är ett <u>90%</u> konfidensintervall för andelen vänsterhänta skolbarn i Tyskland?** Välj det alternativ som ligger närmast ditt svar. (5p)

(A) (0.0804, 0.1196)

(B) (0.0836, 0.1164)

(C) (0.0941, 0.1059)

(D) (0.0951, 0.1049)

(E) (0.0998, 0.1002)

b) En nutritionist undersöker mängden kilokalorier (kcal) i två populära veganska "energy bars", *Raw* och *Omega* (hon misstänker att innehållsförteckningarna är felaktiga). Nutritionisten samlar in ett slumpmässigt urval av 10 energy bars av vardera sort och mäter deras kilokaloriinnehåll. Antag att de verkliga populationsvarianserna är desamma och att deras kilokaloriinnehåll är normalfördelade. **Baserat på nutritionistens tabell nedan, bilda ett 95% konfidensintervall för skillnaden i kcal-innehåll,** $\mu_{\mathbf{Raw}} - \mu_{\mathbf{Omega}}$. Välj det alternativ som ligger närmast ditt svar. (5p)

Brand	n	$\bar{x}$	s^2
Raw	10	200	62.0
Omega	10	190	66.0

(A) (1.907, 18.093)

(B) (2.483, 17.517)

(C) (2.988, 17.012)

(D) (4.397, 15.603)

(E) (4.685, 15.315)

c) **Vilket av följande alternativ är korrekt?** (5p)

(A) Ett statistiskt signifikant resultat är synonymt med ett p-värde som är större än α.

(B) Konfidensintervallets längd kallas också *felmarginal*.

(C) För att minska standardfelet hos en estimator med 50% behöver man typiskt dubbla stickprovsstorleken.

(D) Om stickprovet är stort nog och populationens varians är ändlig kan man använda centrala gränsvärdessatsen för att säga att stickprovsmedelvärdet kommer att vara approximativt normalfördelat.

(E) Väntevärdet för en icke-väntevärdesriktig estimator är lika med den parameter som den skattar.

Uppgift 5

En godistillverkare producerar chockladdoppade jordnötter. Varje jordnöt har ett färgglatt yttre skal och det finns $K = 4$ möjliga färger. Tillverkaren hävdar att varje färg förekommer med samma sannolikhet. Om en påse innehåller ett visst antal jordnötter så är alltså de förväntade andelarna desamma för varje färg. Du väljer slumpmässigt en påse med $n = 200$ jordnötter och räknar jordnötternas färger: 45 är röda, 40 är gröna, 60 är gula, och 55 är bruna. Du väljer att använda ett χ^2-test för att avgöra om varje färg är lika vanligt förekommande eller om färgernas sannolikheter skiljer sig, vid 5% signifikansnivå.

a) **Vad är beslutsregeln och vad är nollhypotesen?** (5p)

(A) Förkasta H_0 om $\chi^2_{obs} > 3.841$; H_0: varje färg har sannolikhet 25%

(B) Förkasta H_0 om $\chi^2_{obs} > 5.991$; H_0: varje färg har sannolikhet 25%

(C) Förkasta H_0 om $\chi^2_{obs} > 5.991$; H_0: någon av sannolikheterna är <u>inte</u> 25%

(D) Förkasta H_0 om $\chi^2_{obs} > 7.815$; H_0: varje färg har sannolikhet 25%

(E) Förkasta H_0 om $\chi^2_{obs} > 7.815$; H_0: någon av sannolikheterna är <u>inte</u> 25%

b) **Vad är testvariabelns värde?** Välj det alternativ som ligger närmast ditt svar. (5p)

(A) 0.0115

(B) 3.31

(C) 5.00

(D) 7.93

(E) 8.44

Uppgift 6

I en sömnstudie mätte en grupp forskare hur många timmars sömn en grupp slumpmässigt utvalda (och frivilliga) patienter sov under en natt. En vecka senare fick samma grupp av åtta patienter ett sömnmedel innan deras sömn mättes ytterligare en natt. Tabellen visar hur länge varje patient sov, med och utan sömnmedlet. Testa på 5% signifikansnivå om sömnmedlet gör att patienter sover längre.

Patient	1	2	3	4	6	7	8
Utan sömnmedel	8.6	6.9	7.8	7.2	7.6	7.6	7.7
Med sömnmedel	9.5	7.6	7.7	8.6	9.2	9.5	8.5

a) Ange dina antaganden, hypoteser, teststatistika, kritiskt värde och beslutsregel. (8p)
b) Slutför dina beräkningar, ange dina slutsatser och tolka resultatet. (6p)
c) Förklara kortfattat skillnaden mellan typ I-fel och typ II-fel. Ge ett kortfattat exempel på ett typ I fel. Du kan använda studien i denna uppgift för att hitta på ett exempel. (6p)

Uppgift 7

Ett bokförlag har precis givit ut en ny biografi. Tabellen nedan innehåller data från ett slumpmässigt urval av åtta bokhandlare där y är antalet sålda böcker och x är pris i euro. För att underlätta era beräkningar innehåller tabellen också några ytterligare siffror.

i	x_i	y_i	x_i^2	y_i^2	$x_i y_i$	$\hat{y}_i$
1	15	29	225	841	435	38,38596
2	20	38	400	1444	760	29
3	19	34	361	1156	646	30,87719
4	21	39	441	1521	819	27,12281
5	15	43	225	1849	645	38,38596
6	26	10	676	100	260	17,73684
7	19	19	361	361	361	30,87719
8	25	20	625	400	500	19,61404
Summa	160	232	3314	7672	4426	232

a) Skriv ner en regressionsmodell med pris (X) som oberoende variabel och antalet sålda böcker (Y) som beroende variabel. Skatta modellens parametrar med minsta kvadratmetoden. (6p)

b) Beräkna residualvarianserna och förklaringsgraden. Tolka förklaringsgraden. (6p)

c) Testa formellt om modellens lutningskoefficient är skild från noll vid 5% signifikansnivå. Ange dina antaganden, hypoteser, teststatistika, kritiskt värde och beslutsregel, beräkningar och slutsats. (8p)

Statistiska institutionen

Stockholms universitet

Rättningsblad

Datum: 13/2-2020

Sal: Brunnsvikssalen

Tenta: Statistik för ekonomer (svenska versionen)

Kurs: Grundläggande statistik för ekonomer

ANONYMKOD: 0028-SBL

☒ 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	X	X			3
Lär.ant.									

BeKs

POÄNG	BETYG	Lärarens sign.
63	D	

STOCKHOLMS UNIVERSITET
Statistiska Institutionen

VT 2020

SVARSBILAGA – GRUNDLÄGGANDE STATISTIK FÖR EKONOMER
2020-02-13

Rum: BR

Anonymkod: 0028-SBL *(skriv tydligt!)*

Markera tydligt ditt svar med ett kryss (X) i vald ruta.

OBS! Bara ett kryss per fråga. Om mer än ett alternativ har angivits så kommer inga poäng att ges för den frågan.

OBS! Om du, efter noggrann kontroll, är övertygad att rätt alternativ inte finns med bland svarsalternativen, så kan du använda marginalen för att förklara hur du resonerar. Du bör i detta fall ändå markera ett av svarsalternativen.

	A	B	C	D	E	
1a	☐	☐	☐	☒	☐	
1b	☐	☐	☒	☐	☐	
2a	☐	☒	☐	☐	☐	
2b	☐	☒	☐	☐	☐	
3a	☐	☐	☒	☐	☐	–
3b	☒	☐	☐	☐	☐	–
3c	☐	☐	☐	☐	☒	–
4a	☐	☒	☐	☐	☐	
4b	☐	☒	☐	☐	☐	
4c	☐	☐	☐	☐	☒	–
5a	☐	☐	☐	☒	☐	
5b	☐	☐	☒	☐	☐	

40/60

SU, STATISTIK

Skrivsal: BR **Anonymkod:** 0028-SBL **Blad nr:** 1

Uppgift 6

a)

$n = 7$

Antaganden: μ_x = utan sömnmedel μ_y = med sömnmedel

X och Y är normalfördelade $\alpha = 0{,}05$

X och Y har parvisa beroenden

σ_D^2 är okänd; $\sigma_x^2 = \sigma_y^2$

Hypoteser: H_0: $\mu_x - \mu_y = 0$ H_1: $\mu_x - \mu_y < 0$

Teststatistika: $t_{n-1} = \frac{\bar{d} - \mu_0}{s_d/\sqrt{n}}$

$t_{krit} = 1{,}943$ om vi testar med 5% signifikansnivå

Beslutsregel: förkasta H_0 om $t_{obs} > 1{,}943$

/7

b) $t_{obs} = \frac{\bar{d} - \mu_0}{s_d/\sqrt{n}}$

	1	2	3	4	5	6	7
x = utan sömn.	8.6	6,9	7.8	7.2	7.6	7.6	7.7
y = med sömn.	9,5	7.6	7.7	8.6	9.2	9.5	8.5
d = y − x	0,9	0,7	−0,1	1,4	1,6	1,9	0,8

$\bar{d} = \frac{0{,}9 + 0{,}7 - 0{,}1 + 1{,}4 + 1{,}6 + 1{,}9 + 0{,}8}{7} =$

$= 1{,}029$

$s_d^2 = \frac{\sum |d_i - \bar{d}|^2}{n-1} = 0{,}4457$

$s_d = \sqrt{s_d^2} = \sqrt{0{,}4457} = 0{,}6676$

→

$$t_{obs} = \frac{1{,}029 - 0}{0{,}6676/\sqrt{7}} = \frac{1{,}029}{0{,}2523} = 4{,}078$$

$4{,}078 > 1{,}943$ vilket innebär att $t_{obs} > t_{krit}$ och att H_0 förkastas. Detta betyder att sömnmedlet gör att patienter sover längre.

/6

c) Typ I-fel är att förkasta en sann nollhypotes vilket även kallas för signifikansnivå och betecknas med α

Typ II-fel är att inte förkasta en falsk nollhypotes och betecknas ofta med β

/4

SU, STATISTIK

Skrivsal: BR Anonymkod: 0028-SBL Blad nr: 2

Uppgift 7

a) $Y = \beta_0 + \beta_1 X_i$

$\bar{x} = \frac{160}{8} = 20$

$\bar{y} = \frac{232}{8} = 29$

$\beta_0 = \bar{y} - \beta_1 \cdot \bar{x} =$
$= 29 - (-1{,}877 \cdot 20) =$
$= 66{,}54$

$Y = 66{,}54 + (-1{,}877 \cdot X_i)$

$\beta_1 = \frac{Cov(x,y)}{S_x^2}$

$\beta_1 = \frac{\frac{\sum(x_i - \bar{x})(y_i - \bar{y})}{n-1}}{\frac{\sum(x_i - \bar{x})^2}{n-1}}$

$Cov(x,y) = -30{,}57$

$S^2_x = 16{,}285$

$\beta_1 = \frac{-30{,}57}{16{,}285} = -1{,}877$

6

b) $S_e^2 = \frac{\sum(y_i - \hat{y}_i)^2}{n-k-1} = \frac{\sum(y_i - \hat{y}_i)^2}{5} =$

$= \frac{542{,}28}{5} = 108{,}456$

$k = 2$ förklaringsvariabler

0