

TENTAMEN I GRUNDLÄGGANDE STATISTIK FÖR EKONOMER 2018-10-30

Skrivtid:	kl. 15.00 - 20.00
Godkända hjälpmedel:	Miniräknare utan lagrade formler och text
Bifogade hjälpmedel:	Häftet <i>Formelsamling och Tabeller över statistiska fördelningar</i> (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.

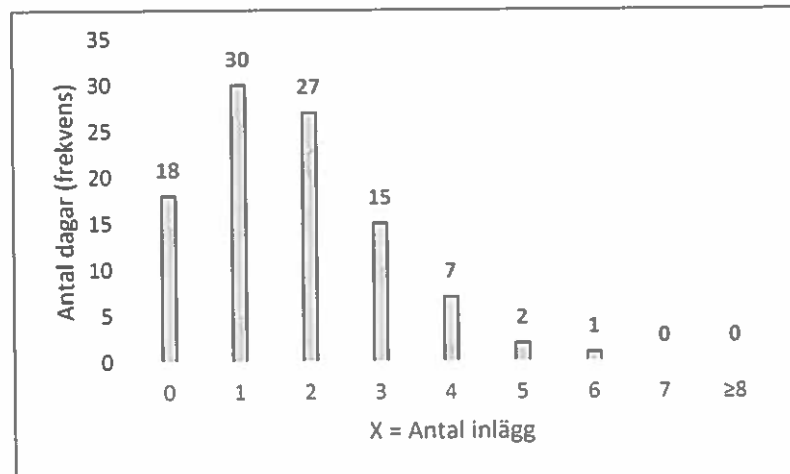
- Lösningförslag läggs ut på Mondo kort efter tentamen.

LYCKA TILL!

Endast svar ska anges för uppgifterna 1-5 med deluppgifter. Markera rätt svar på Svarsbilagan. OBS! Om du inte hittar rätt svar bland de givna svarsalternativen, skriv ditt svar och eventuell kommentar i marginalen på Svarsbilagan vid den aktuella frågan.

Uppgift 1

En Facebook-användare har fört bok över antalet inlägg som hon gjort per dag under $n = 100$ dagar. Vissa dagar har noll inlägg gjorts, andra dagar är det fler. I diagrammet nedan visas frekvensfördelningen för $x =$ antal inlägg per dag. De observerade frekvenserna anges ovanför varje stapel.



- a) Beräkna och ange kvartilavståndet (IQR) för antal inlägg per dag enligt metoden som beskrivs i kurslitteraturen. TIPS! De kumulativa frekvenserna är användbara. (5p)
- A. $IQR = 1,50$
 - B. $IQR = 2,75$
 - C. $IQR = 1,00$
 - D. $IQR = 1,75$
 - E. Inget av ovanstående är rätt, IQR går inte att beräkna med givna data.

En tidsserie brukar ibland beskrivas som en summa, eller produkt, av olika komponenter som var för sig avspeglar olika egenskaper i tidsserien man studerar.

- b) Vilket av följande påståenden är inte rätt i samband med komponentmetoden? (5p)
- A. Glidande medelvärden används för att skatta trenden (el. trendcykel).
 - B. I en multiplikativ modell är medelvärdet av säsongskomponenterna lika med 1.
 - C. Additiva modeller förknippas typiskt med konstant säsongvariation.
 - ? D. Säsongskomponenterna skattas efter att trenden har rensats bort från tidsserien.
 - ? E. Trenden antas alltid vara linjärt ökande alternativt linjärt avtagande med tiden.

Uppgift 2

Ett företag som arbetar med företagshälsa genomförde en studie där fann man att åldersfördelningen bland personer som motionerade regelbundet var följande: 50 % tillhörde åldersgrupp A_1 (18-29 år), 30 % tillhörde grupp A_2 (30-54 år) och 20 % tillhörde grupp A_3 (55-67 år). Vidare fann man i varje åldersgrupp en varierande andel rökare. Andelen rökare i A_1 var 20 %, i A_2 var andelen 30 % och i A_3 var det 5 % som rökte. Gruppen rökare betecknas med R .

a) En person väljs slumpmässigt ur populationen av motionärer. Vad är sannolikheten att den personen är en rökare? (5p)

A. 0,1833

B. 0,20

C. 0,55

D. 0,10

E. 0,1667

b) Den slumpmässigt dragna personen visade sig vara rökare. Vad är sannolikheten att den personen tillhör åldersgruppen 30-54 år? (5p)

A. 0,09

B. 0,5454

C. 0,45

D. 0,30

E. 0,2222

Företaget analyserade även antal träningspass per dag bland icke-rökarna. Man fann att det hela kunde beskrivas med följande sannolikhetsfördelning där X = antalet pass för en slumpmässigt vald person en slumpvist vald dag:

x	0	1	2	3
$P(X = x)$	0,4	0,3	0,2	0,1

c) Beräkna och ange väntevärdet och standardavvikelsen för antalet träningspass per dag. (5p)

A. $\mu_X = 1,5$ $\sigma_X = 1,0$

B. $\mu_X = 1,5$ $\sigma_X = 1,118$

C. $\mu_X = 1,0$ $\sigma_X = 1,0$

D. $\mu_X = 1,0$ $\sigma_X = 1,118$

E. $\mu_X = 1,0$ $\sigma_X = 2,0$

Uppgift 3

Alla som köper avokado vet att det kan vara svårt att bedöma om den är bra eller inte. Erfarenhetsmässigt vet du att ca 75 % är bra och resterande 25 % är antingen omogna eller bruna och övermogna, dvs. inte bra. Eftersom du gillar avokado och vill vara säker på att få några bra till middagsbjudningen köper du $n = 8$ stycken frukter. Låt X beteckna antalet bra frukter som du får.

a) Beräkna och ange sannolikheten att du får högst 4 bra frukter, dvs. $P(X \leq 4)$. (5p)

- A. 0,114
- B. 0,208
- C. 0,886
- D. 0,027
- E. 0,087

Låt X beteckna tiden i minuter som en slumpvist vald anställd ägnar åt kafferaster och Y tiden som används för privat surfande på webben under en normal arbetsdag. Anta att X och Y båda är normalfördelade slumpvariabler med väntevärden $\mu_X = 30$ respektive $\mu_Y = 40$ samt varianser $\sigma_X^2 = 100$ respektive $\sigma_Y^2 = 400$. Definiera sedan en ny slumpvariabel $W = X + Y$. Variansen för W är $\sigma_W^2 = 180$.

b) Beräkna och ange sannolikheten $P(25 \leq Y \leq 65)$. (5p)

- A. 0,121
- B. 1,121
- ? C. 0,332
- D. 0,879
- E. 0,668

c) Beräkna och ange sannolikheten ρ_{XY} = korrelationskoefficienten mellan X och Y . TIPS! Du måste först beräkna kovariansen mellan X och Y ; använd σ_W^2 och formelhäftet! (5p)

- A. -0,80
- B. 0,90
- C. 0,80
- D. -1,60
- E. -0,90

OBS! Svarsalternativen i a) och b) har avrundats till 3 decimaler.

Uppgift 4

Inför en lansering av en ny produkt ville ett marknadsföringsföretag göra en demografisk analys av olika kommuner. Man ville bl.a. jämföra medelinkomsten i Alvesta och Boden. Två slumpmässiga stickprov omfattande $n_A = 100$ personer i Alvesta och $n_B = 100$ personer i Boden samlades in.

Stickprovsmedelvärdet i Alvesta var $\bar{x}_A = 266$ tkr och standardavvikelsen $s_A = 38$ tkr. I Boden var medelvärdet $\bar{x}_B = 281$ tkr och standardavvikelsen $s_B = 40$ tkr.

Anta att inkomsterna i respektive kommun är normalfördelade och att du ska beräkna ett 95 % konfidensintervall för **differensen** (skillnaden) mellan de två kommunernas medelinkomst, dvs. för $\mu_A - \mu_B$.

- a) Beräkna och ange **felmarginalen** för differensen mellan de två kommunernas medelinkomst. TIPS! Du behöver inte använda alla sifferuppgifter som du fick. (5p)

- A. 7,647
- B. 21,63
- C. 9,075
- D. 10,81
- E. 1,731

- b) Om man hade satt som krav att felmarginalen för differensen högst skulle vara lika med 8, hur stort stickprov hade man behövt från respektive kommun? (5p)

OBS! Du känner ju inte till populationsvarianserna σ_A^2 respektive σ_B^2 . Använd därför stickprovsvarianserna i dina beräkningar. Dessutom ska du utgå ifrån att stickproven ska vara lika stora i de två kommunerna, dvs. $n_A = n_B$.

- A. 746
- B. 365
- C. 183
- D. 94
- E. 1462

Uppgift 5

Sverige tar emot ensamkommande barn som söker asyl. Tre av de vanligaste ursprungsländerna som barnen kom från under ett visst år var Afghanistan, Somalia och Syrien. Anta att det under en månad totalt kom $n = 160$ ensamkommande barn fördelat på de tre länderna och kön enligt följande tabell (siffrorna har justerats men är baserade på statistik från Migrationsverket):

Antal	Afghanistan	Somalia	Syrien
Pojkar	50	30	40
Flickor	10	10	20

Du får i uppdrag att genomföra ett χ^2 -test på signifikansnivån 5 % för att pröva om könsfördelningen för de tre länder är lika eller olika.

a) Ange korrekt beslutregel för ditt test. (4p)

- A. om $\chi_{\text{obs}}^2 > 2,920 \Rightarrow$ förkasta H_0
- B. om $\chi_{\text{obs}}^2 > 5,991 \Rightarrow$ förkasta H_0
- C. om $\chi_{\text{obs}}^2 > 3,841 \Rightarrow$ förkasta inte H_0
- D. om $\chi_{\text{obs}}^2 > 7,378 \Rightarrow$ förkasta H_0
- E. om $\chi_{\text{obs}}^2 < 4,303 \Rightarrow$ förkasta inte H_0

b) Med givna data, ange testvariabelns observerade värde och din slutsats. (6p)

- A. $\chi_{\text{obs}}^2 = 4,875 \Rightarrow H_0$ förkastas, könsfördelningen för de tre länder är lika
- B. $\chi_{\text{obs}}^2 = 0,556 \Rightarrow H_0$ förkastas inte, könsfördelningen för de tre länder är lika
- C. $\chi_{\text{obs}}^2 = 4,444 \Rightarrow H_0$ förkastas inte, könsfördelningen för de tre länder är lika
- D. $\chi_{\text{obs}}^2 = 4,444 \Rightarrow H_0$ förkastas inte, könsfördelningen för de tre länder är olika
- E. $\chi_{\text{obs}}^2 = 4,875 \Rightarrow H_0$ förkastas inte, könsfördelningen för de tre länder är olika

Fullständig redovisning krävs för följande uppgifter.

Använd separata pappersark för uppgift 6 resp. uppgift 7.

Uppgift 6

Vid en analys av produktionskostnader inom en stor koncern studerade man sambandet mellan ställtider (dvs. tid då produktionen ligger nere) och lönsamhet. Man hade uppgifter för åtta olika fabriker och skattade en enkel linjär regressionsmodell enligt:

$$\text{Modell 1: } Y_i = \beta_0 + \beta_1 X_i + \varepsilon_i$$

där Y = lönsamhet i % av investeringen och X = andel i % av total tid som var ställtid.

Senare kom man ihåg att de fyra första fabrikerna 1-4 tillverkade mobiltelefoner medan de andra, 5-8, tillverkade andra produkter. Under perioden hade marknaden för mobiltelefoner dessutom haft ett kraftigt uppsving med stora vinster. Man utökade därför modellen med ytterligare en förklaringsvariabel enligt:

$$\text{Modell 2: } Y_i = \beta_0 + \beta_1 X_i + \beta_2 D_i + \varepsilon_i$$

där D är en dummyvariabel sådan att $D = 1$ för fabriker som tillverkade mobiltelefoner och 0 annars. Data och resultaten för dessa modellskattningar hittar du på nästa sida. Vissa delar av utskriften saknas dock.

- Beräkna 90 % konfidensintervall för lutningskoefficienten för X = andel ställtid i **Modell 1** och i **Modell 2**. Tolka och jämför resultaten. Kommentera. Är resultaten rimliga? (6p)
- Åskådliggör datamaterialet och sambandet mellan X och Y i ett lämpligt diagram. Markera på lämpligt sätt de fabriker som tillverkar mobiltelefoner och de som inte gör det. Rita även in den skattade regressionslinjen för **Modell 1** i ditt diagram. (4p)
- Kan du med hjälp av diagrammet i b) förklara resultaten i a) uppgiften? Förklara! (4p)
- Förklara kortfattat begreppet förklaringsgrad. Beräkna sedan den justerade förklaringsgraden för respektive modell. Jämför och kommentera resultaten. (6p)

Uppgift 7.

Utgå ifrån uppgift 6. Du vill nu jämföra medelvärdena för de två grupperna av fabriker med avseende på lönsamhet, dvs. Y variabeln. Låt μ_M beteckna väntevärdet för fabriker som tillverkar mobiler och μ_A väntevärdet för fabriker som inte tillverkar mobiler. Betrakta de data som ges på nästa sida som två oberoende stickprov från respektive grupp.

- Beräkna medelvärdena och varianserna för de två stickproven. (2p)
- Testa sedan hypotesen att μ_M och μ_A är lika mot hypotesen att μ_M är större än μ_A på signifikansnivån 5 %. Ange antaganden, hypoteser, testvariabel, beslutsregel, beräkningar och slutsats. (10p)
- Förklara kortfattat vad som menas med Typ I och Typ II fel samt, signifikansnivå och styrka och hur dessa begrepp hänger ihop. Vad är ett p -värde och hur kan man använda det? Max en A4 bör räcka. (8p)

Typ 1 : Oskyldigt dömd

Typ 2 : Skyldig men frigiven

Bilaga till Uppgift 6 och 7

Data	Fabrik	Y = Lönsamhet (%)	X = Andel ställtid (%)	D = Mobil	xy	\bar{x}	\bar{y}
	1	10	60	1	600	47,75	7
	2	9	55	1	495		
	3	8	65	1	520		
	4	11	53	1	587		
	5	4	45	0	180		
	6	6	29	0	174		
	7	5	35	0	175		
	8	3	40	0	120		

Modell 1

ANOVA					
	Fg	KvS (SS)	Mkv (MS)	F	p-värde F
Regression (R)	1	26,975	26,975	4,901	0,06878
Residual (E)	6	33,025	5,504		
Totalt (T)	7	60,000			

	Koefficienter	Standardfel	t-kvot	p-värde
Konstant	-0,445	3,464	-0,129	0,90188
Ställtid (X)	0,156	0,070	2,214	0,06878

Modell 2

ANOVA					
	Fg	KvS (SS)	Mkv (MS)	F	p-värde F
Regression (R)	2	56,018	28,009	35,166	0,00113
Residual (E)	5	3,982	0,796		
Totalt (T)	7	60,000			

	Koefficienter	Standardfel	t-kvot	p-värde
Konstant	10,558	2,249	4,695	0,00536
Ställtid (X)	-0,163	0,059	-2,749	0,04037
Mobil (D)	8,415	1,394	6,038	0,00179

**SVARSBILAGA till Tentamen i Grundläggande statistik för ekonomer
2018-10-30**

Skrivsal: _____

Anonymkod: _____ (skriv tydligt!)

Markera ditt svar med ett tydligt kryss (X) i rutorna nedan.

OBS! Endast ett kryss per uppgift. Har fler än ett svarsalternativ markerats ges noll poäng.

OBS! Om du efter att ha kontrollerat dina beräkningar ordentligt kommer fram till att svaret inte finns bland de angivna svarsalternativen, skriv ditt svar i marginalen till höger och kommentera (använd baksidan om det behövs).

		A	B	C	D	E
Uppgift 1	a)	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
	b)	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
Uppgift 2	a)	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
	b)	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
	c)	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
Uppgift 3	a)	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
	b)	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
	c)	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
Uppgift 4	a)	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
	b)	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
Uppgift 5	a)	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
	b)	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>



Rättningsblad

Datum: 30/10/18

Sal: Värtasalen

Tenta: Statistik för ekonomer

Kurs: Grundläggande statistik för ekonomer

ANONYMKOD:

0142-PBM

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			4
Lär.ant. 10	15	15	10	10	20	20			

POÄNG	BETYG	Lärarens sign.
100	A	ME

Sic itur ad astra /ME

**SVARSBILAGA till Tentamen i Grundläggande statistik för ekonomer
2018-10-30**

Skrivsal: Värtasalen

Anonymkod: 0102-ABM (skriv tydligt!)

Markera ditt svar med ett tydligt kryss (X) i rutorna nedan.

OBS! Endast ett kryss per uppgift. Har fler än ett svarsalternativ markerats ges noll poäng.

OBS! Om du efter att ha kontrollerat dina beräkningar ordentligt kommer fram till att svaret inte finns bland de angivna svarsalternativen, skriv ditt svar i marginalen till höger och kommentera (använd baksidan om det behövs).

		A	B	C	D	E
Uppgift 1	a)	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input checked="" type="checkbox"/>	<input type="checkbox"/>
	b)	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input checked="" type="checkbox"/>
Uppgift 2	a)	<input type="checkbox"/>	<input checked="" type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
	b)	<input type="checkbox"/>	<input type="checkbox"/>	<input checked="" type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
	c)	<input type="checkbox"/>	<input type="checkbox"/>	<input checked="" type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
Uppgift 3	a)	<input checked="" type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
	b)	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input checked="" type="checkbox"/>
	c)	<input checked="" type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
Uppgift 4	a)	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input checked="" type="checkbox"/>	<input type="checkbox"/>
	b)	<input type="checkbox"/>	<input type="checkbox"/>	<input checked="" type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
Uppgift 5	a)	<input type="checkbox"/>	<input checked="" type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
	b)	<input type="checkbox"/>	<input type="checkbox"/>	<input checked="" type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>



Uppgift 6a) Modell 1för att skatta lutningskoefficienten (β_1) för X =areal ställtid:

$$KI: b_1 \pm t_{n-k-1, \alpha/2} \cdot s_{b_1}$$

Enligt excel utskrift för modell 1:

$$b_1 = 0,156$$

$$s_{b_1} = 0,070$$

Enligt t-tabel:

$$n=8, \quad k=1 \quad \alpha=0,1 \text{ (10\%)}$$

$$t_{8-1-1; 0,05} = 1,943$$

90% konfidensintervall för β_1 i modell 1 = $0,156 \pm 1,943 \times 0,070$
 $= 0,156 \pm 0,13601$

$$KI = (0,01999, 0,29201)$$

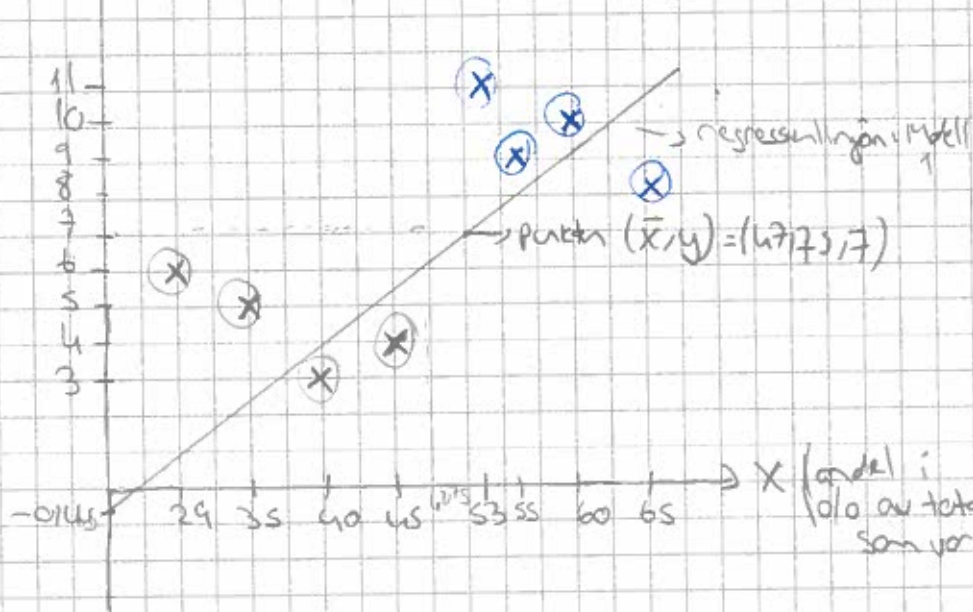
Modell 2Enligt excel utskrift $\rightarrow b_1 = -0,163 \quad s_{b_1} = 0,059$ Enligt t-tabel $\rightarrow t_{n-k-1, \alpha/2} = t_{5; 0,05} = 2,015 \quad \left(\begin{array}{l} k=2, n=8 \\ \alpha=0,1 \end{array} \right)$ 90% konfidensintervall för β_1 för modell 2:

$$b_1 \pm t_{5; 0,05} \cdot s_{b_1} = -0,163 \pm 2,015 \times 0,059 = -0,163 \pm 0,118885$$

$$= (-0,281885, -0,044115)$$

Om vi kollar på modell 1, ^{ned} 90% konfidens ligger lutningskoefficient β_1 intervallet $(0,01999, 0,29201)$ medan i modell 2 ligger detta i $(-0,281885, -0,044115)$.
 Modell 1 resultat är lite rimligt för att areal i ställtid borde ha ledas till generelltvis ökning i lönsamhet. Detta resultat händer på grund av modellen saknar en viktig förklaringsvariabel (mobiltelefon tillverknings-
 dummy variabel).
 Men modell 2 KI verkar rimligt med negativ samband mellan ställtid och lönsamhet.

b) y (lönansat i % av investering)



$$\bar{x} = \frac{\sum x_i}{8} = 47,75$$

$$\bar{y} = \frac{\sum y_i}{8} = 7$$

\otimes fabriker som tillverkar mobiltelefoner

\otimes inte tillverkar mobiltelefoner

Medelvärde punkterna \bar{x} (47,75) och \bar{y} (7) ligger på ^{stättade} regressionslinjen i modell 1. Den stättade regressionslinjen lutning är 0,156 och interceptet är -0,1115.

b) Modell 1 tar inte hänsyn till om en fabrik tillverkar mobiltelefoner eller inte, medan modell 2 har 2 förklaringsvariabel (en dummyvariabel + tillägg). I frågan a) modell 1 hade en positiv lutningskoefficient för variabel x som är inte rimligt. Det är inte logiskt att fabriker som har högre ställt tid har högre lönsamhet. När vi kollar på diagrammet ovan så kan vi se att den verkliga positiva sambandet är om en fabrik producerar mobiltelefoner eller inte, så modell 1 saknar en viktig förklaringsvariabel i modellen (specifikation bias) vilket leder till inkorrekt modell och lutningskoeff. Dvs, genom att kolla på diagrammet 1 kan vi förklara varför lutningskoeff för x = andel ställt tid i modell 1 är positiv medan i modell 2 negativ

Uppgift 6 forts.

R^2 kan ta värden mellan 0% - 100%

(bd) Förklaringsgrad betecknas som R^2 och detta visar hur bra en modell för att förklara variationen i beroende variabel. I modell 1 finns det bara ett förklaringsvariabel (X = andel i % total tid som var ställdid), så R^2 i modell 1 visar hur bra variationen i X att förklara variationen i Y (lönsett i % av löntuttagen). I modell 2 har vi 2 förklaringsvariabler (X och dummyvariabel modelltyp) och där R^2 visar hur mycket av variationen i Y kan förklaras med variationen i de 2 förklaringsvariabler.

$$\bar{R}^2 = 1 - \frac{SSE / (n - k - 1)}{SST / (n - 1)} \quad \left(\text{Justerad } R^2 \right)$$

Modell 1

Enligt utskrift \rightarrow $SSE = 33,025$ $SST = 60,000$ $k = 1$ $n = 8$

$$\bar{R}^2 = 1 - \frac{(33,025/6)}{(60,000/7)} = 1 - \frac{5,504166667}{8,571428571} = 0,3578 = \underline{\underline{35,78\%}}$$

Modell 2

Enligt utskrift \rightarrow $SSE = 3,982$ $SST = 60,000$ $k = 2$ $n = 8$

$$\bar{R}^2 = 1 - \frac{(3,982/5)}{(60,000/7)} = 1 - \frac{0,7964}{8,571428571} = 0,9071 = \underline{\underline{90,71\%}}$$

Man kan se att modell 2 är mycket bättre att förklara variation i lönsamhet jämfört med modell 1. \bar{R}^2 tar hänsyn till hur många förklaringsvariabler finns i modellen så det är tillförlitlig för att jämföra 2 modeller även om de har annorlunda antal förklaringsvariabler. Resultatet är förväntat för att vi visste redan att Modell 1 sakar en signifikant förklaringsvariabel (Om vi testade bara modell 1, då hade vi p -värde = 0,068 > 0,05, och kunde vi från till att modellen är inte användbar - 5% signifikans nivå). Man kan också se att Modell 1's SSE är mycket högre än Modell 2.

Uppgift 7

- a) Stickprov M - fabriker som tillverkar mobiltelefoner
Stickprov A - fabriker som tillverkar mobiltelefoner

Medelvärde för stickprov M $\Rightarrow \frac{\sum y_M}{n} = \frac{10+9+8+11}{4} = \underline{9,5}$
av lensenhet

Varians $S_{y_M}^2 = \frac{\sum_{i=1}^n y_i^2 - n \cdot \bar{y}_M^2}{n-1} = \frac{(10^2+9^2+8^2+11^2) - 4 \cdot (9,5)^2}{4-1} = \underline{1,67}$

Medelvärde för lensenhet i stickprov $A = \frac{\sum y_A}{n} = \frac{10+6+5+3}{4} = \underline{4,5}$

Varians $S_{y_A}^2 = \frac{(10^2+6^2+5^2+3^2) - 4 \cdot (4,5)^2}{4-1} = \underline{1,67}$

b) Antagande

Stickproven är oberoende från varandra, och y betecknar en slumpvariabel som visar lensenhet. Observationer i stickprov M är oberoende från varandra och har samma fördelning, dvs. lensenhet väntevärde μ_M och varians σ_M^2 (iid). De är normalfördelade.

Samma antagande är också tillämpligt på stickprov A där $y_i \sim N(\mu_A, \sigma_A^2)$. Vi vet inte σ_M^2 och σ_A^2 men vi kan skatta dem med stickproven $S_{y_M}^2$ och $S_{y_A}^2$ (vi antar att σ_M^2 och σ_A^2 är lika).

På grund av oberoende antagningar kan vi också anta att

$$\bar{M} \sim N\left(\bar{\mu}_M, \frac{\sigma_M^2}{n}\right) \text{ och } \bar{A} \sim N\left(\bar{\mu}_A, \frac{\sigma_A^2}{n}\right)$$

(\bar{M} och \bar{A} är slumpriktor som betecknar medelvärdet av stickproven
 $\bar{\mu}_M = \frac{\sum y_M}{n}$ $\bar{\mu}_A = \frac{\sum y_A}{n}$)

Hypotes

$H_0: \mu_M - \mu_A = 0$ mot $H_1: \mu_M - \mu_A > 0$ $\alpha = 0,05$

Uppgift 7b fortsätter

Testvariabel

\bar{J}_M^2 och \bar{J}_A^2 observerade och $n_M, n_A < 30$, båda normalfördelade

$$T_{n_M+n_A-2} = \frac{M - A - D_0}{s.p. \sqrt{\frac{1}{n_M} + \frac{1}{n_A}}} \quad \text{där } s.p.^2 = \frac{(n_M-1)S_M^2 + (n_A-1)S_A^2}{n_M+n_A-2}$$

\rightarrow +fördelad med $u+u-2=6$ frihetsgrad

Kritisk gräns och beslutregel

$\alpha = 5\%$

för att det är en ensidig test, $t_{kritisk} = t_{u+u-2, 0,05}$

$$= t_{6, 0,05} = \underline{\underline{1,943}}$$

forkasta H_0 om $t_{obs} > t_{kritisk} = 1,943$

Beräkningar

Vi vet att $\bar{M} = 9,15$, $\bar{A} = 4,15$, $S_M^2 = 1,167$ och $S_A^2 = 1,167$

$$s.p.^2 = \frac{(4-1) \cdot 1,167 + (4-1) \cdot 1,167}{4+4-2} = 1,167 \rightarrow s.p. = \sqrt{1,167} = 1,292286798$$

$$t_{obs} = \frac{9,15 - 4,15 - 0}{1,292286798 \sqrt{\frac{1}{4} + \frac{1}{4}}} = \underline{\underline{5,1471356553}}$$

Slutsats

$t_{obs} = 5,1472$ och $t_{kritisk} = 1,943$, så $t_{obs} > t_{kritisk}$. Detta betyder att vi kan forkasta H_0 ($\mu_M = \mu_A = 0$) på 5% signifikansnivå. Det är statistisk säkerställning på 5% att μ_M är större än μ_A .

10

\rightarrow

7c

Typ 1 fel: att förkasta en sann nollhypotes

Typ 2 fel: att inte förkasta en falsk nollhypotes.

Signifikansen betecknas med α och beskriver sannolikheten att förkasta en sann nollhypotes, dvs. sannolikheten att göra en typ 1 fel. Detta bestäms gotttyckligt men brukar vara 1% eller 5%. Testets styrka betecknas som $(1-\beta)$ och betyder att sannolikheten att förkasta en falsk nollhypotes. Detta hänger ihop med typ 2 fel för att sannolikheten att göra ett typ 2 fel är lika med β medan testets styrka är $1-\beta$.

P-värde är den lägsta nivå som man kan förkasta nollhypotes.

Om p-värde är mindre än signifikansen α , då kan man förkasta nollhypotes för det testet. Om p-värde är högre än signifikansen α , då kan man inte förkasta nollhypotesen för det testet.

~~8~~

Sic itur ad astra!