

TENTAMEN I GRUNDLÄGGANDE STATISTIK FÖR EKONOMER

2019-03-22

Skrivtid: kl. 9.00 - 14.00

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.

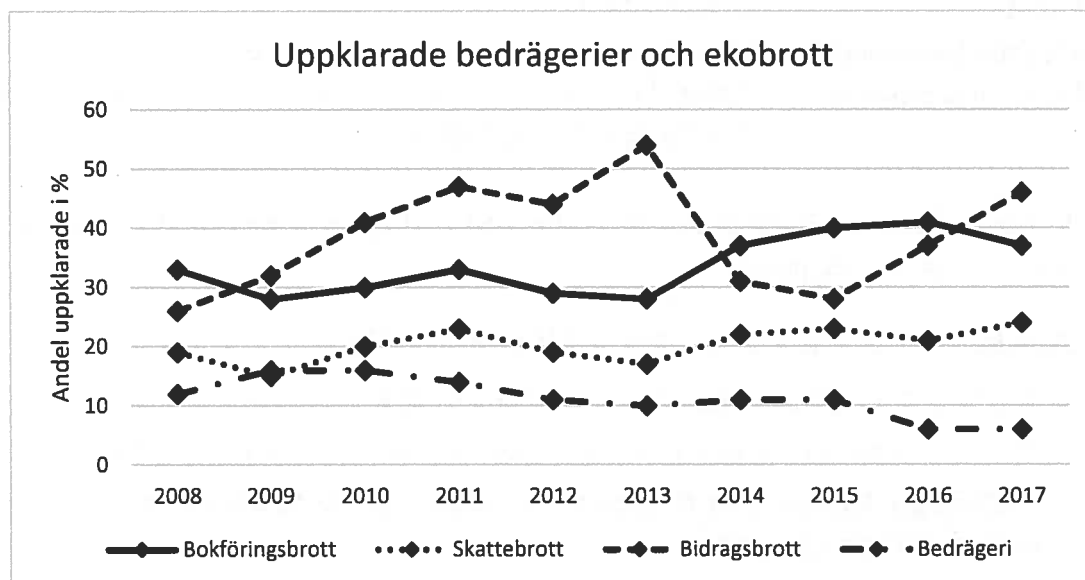
- Lösningförslag läggs ut på Mondo kort efter tentamen.

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 eller anser att något är fel med uppgiften, skriv ditt svar på Svarsbilagan vid den aktuella frågan och kommentera på baksidan av bladet.

Uppgift 1

Brottsförebyggande rådet publicerar regelbundet brottsstatistik i Sverige. I diagrammet nedan visas t.ex. andelen upklarade fall av anmälda bedrägerier och ekobrott.



- a) Utifrån ovanstående diagram, vilket av följande påståenden skulle du säga är **inte** är rätt? (5p)
- Årtal är en diskret numerisk variabel på intervallskala.
 - Brottstyp är en kategorisk variabel på nominalskala.
 - Skattebrott uppvisar en mycket svag uppåtgående trend och endast en vag cyklicitet.
 - Antalet ekobrott totalt sett har varit relativt stabilt sedan 2008.
 - Andel upklarade brott är en kontinuerlig numerisk variabel med utfallsrum (0,1).

Ett helt annat datamaterial innehåller uppgifter om de anställdas månadslöner i ett visst företag. Följande fem deskriptiva mått beskriver lönerna i tusental kronor (tkr):

$$\min = 20 \quad Q_1 = 30 \quad \text{Median} = 34 \quad Q_3 = 45 \quad \max = 115$$

- b) Vilket av följande påståenden är **falskt**? (5p)
- Minst hälften av de anställda tjänar 34 tkr eller mindre
 - Fler än hälften tjänar mer än medianlönen
 - Minst hälften av de anställda har en lön i intervallet 30 – 45 tkr
 - Kvartilavståndet (IQR) är 15 tkr
 - Ingen tjänar 15 tkr

Uppgift 2

Tittarvanorna för streaming TV granskas för två olika konsumentgrupper, yngre tittare (A) och äldre tittare (\bar{A}). Man klassificerade tittarvanorna för bägge grupperna i tre kategorier enligt $B_1 =$ aldrig, $B_2 =$ ibland och $B_3 =$ dagligen. En person dras slumpmässigt och sannolikhetsfördelningen över de sex olika kombinationerna av åldersgrupp och tittarvanor ges i tabellen nedan.

Notera att B_1 , B_2 och B_3 är sinsemellan disjunkta och att de tillsammans fyller upp hela utfallsrummet (samtliga tittare har en och endast en vana); detta gäller förstås även för A och \bar{A} eftersom de är varandras komplement (man är antingen yngre eller äldre).

	B_1	B_2	B_3
A	0,10	0,15	0,15
\bar{A}	0,25	0,20	0,15

a) Vilket av följande påståenden är **falskt**? (5p)

- A. $P(A \cap B_2) = 0,14$ ung och tittar ibland
- B. $P(\bar{A} \cup B_1) = 0,95$ äldre eller tittar aldrig eller både och
- C. $P(B_3|A) = 0,15$ tittar ofta givet ung
- D. $P(B_2 \cap \bar{B}_3) = 0,70$ tittar ibland och man tittar inte dagligen
- E. $P(B_2|B_1) = 0$ tittar ibland givet tittar aldrig

Du väntar på en leverans av ny TV. Budfirman har meddelat att de kommer någon gång mellan kl. 12 och kl. 16 så du har tagit ledigt på eftermiddagen för att kunna ta emot dem. Du inser att din väntetid från kl.12 kan betraktas som en slumpvariabel. Du har kommit fram till att en slumpmodell för $X =$ väntetid i timmar kan beskrivas med den kumulativa sannolikhetsfunktionen

$$P(X \leq x) = F(x) = \frac{x}{4} \quad \text{där } 0 \leq x \leq 4$$

b) Vad är sannolikheten att du får vänta mellan 1,5 och 3 timmar innan de kommer enligt denna modell? (5p)

- A. 0,125
- B. 0,250
- C. 0,375
- D. 0,500
- E. 0,750

Uppgift 3

Två slumpvariabler är normalfördelade enligt $X \sim N(10, 25)$ respektive $Y \sim N(8, 9)$.

a) Anta att du gör $n = 9$ observationer på variabeln X . Vad är sannolikheten att få ett stickprovsmedelvärde som är större än 11, dvs. vad är $P(\bar{X} > 11)$?

- A. 0,159
- B. 0,274
- C. 0,359
- D. 0,579
- E. 0,726

b) Du gör nu en observation på variabeln X och en på variabeln Y . Anta också att korrelationen mellan X och Y är $\rho_{XY} = 0,42$. Vad är sannolikheten att $2X + 3Y$ är mindre än 40, dvs. vad är $P(2X + 3Y < 40)$? Tips: Beräkna kovariansen mellan X och Y först.

- A. 0,261
- B. 0,374
- C. 0,401
- D. 0,599
- E. 0,626

Sannolikheten att en enskild observation på X är större än 11 är lika med 0,58, nästan exakt. Beteckna denna sannolikhet med $P = 0,58$. Du observerar $n = 5$ oberoende observationer på X . Låt $W =$ antalet X av $n = 5$ som är större än 11.

c) Vad är sannolikheten för händelsen $W = 2$? Räkna så exakt som möjligt. OBS! Du kan inte använda Tabell 7 för att lösa uppgiften eftersom tabellen inte omfattar vare sig $P = 0,58$ eller $P = 0,42$.

- A. 0,249
- B. 0,353
- C. 0,013
- D. 0,336
- E. 0,074

OBS! Svartalternativen i a) – c) har avrundats till 3 decimaler. Välj det som är närmast.

Uppgift 4

För en liten grupp slumpmässigt utvalda nyfödda barn mättes vikten i gram vid förlossning (X) och efter tre dagar (Y) varvid följande resultat erhöles:

Barn i	1	2	3	4	5	6	7	8
Vikt vid förlossning x_i	3500	3200	2900	3700	3000	3800	3300	3800
Vikt efter tre dagar y_i	3100	3000	2800	3400	2900	3600	3100	3700

a) Vilket är ett 95 % konfidensintervall för den genomsnittliga viktnedgången i gram? (5p)

- A. (111; 289)
- B. (128; 272)
- C. (138; 262)
- D. (-167; 567)
- E. (-135; 535)

OBS! Svarsalternativen har avrundats till heltal. Välj det som är närmast.

b) Vilket av följande antaganden eller påståenden är inte relevant eller t.o.m. direkt felaktigt för beräkningen och tolkningen av konfidensintervallet i a) ovan? (5p)

- A. Barnen är oberoende av varandra.
- B. Centrala gränsvärdesatsen är inte tillämplig i detta fall.
- C. Stickproven för X respektive Y är oberoende av varandra.
- D. Viktnedgången för ett slumpvist valt barn antas vara normalfördelad.
- E. Standardfelet för skattningen av den genomsnittliga viktnedgången är ca 37,80.

Uppgift 5

Ett stort företag analyserade år 2013 samtliga medarbetares språkkunskaper, både i moderbolaget och i alla dess dotterbolag. Man ville veta hur många språk medarbetarna behärskade och beräknade andelen som behärskade ett språk (40 %), två språk (40 %) respektive tre eller fler språk (20 %). Under år 2018 gjordes en urvalsundersökning där $n = 500$ slumpmässigt valda medarbetare tillfrågades. Av de utvalda angav 180 ett språk, 188 två språk och 132 tre eller fler språk.

Du får i uppdrag att avgöra om fördelningen över antal språk som bolagets samtliga medarbetare behärskar har förändrats sedan 2013. Du bestämmer dig för ett χ^2 -test.

a) Ange det observerade värdet för testvariabeln. (5p)

- A. 0,071
- B. 9,195
- C. 10,57
- D. 12,96
- E. 14,87

b) Med en signifikansnivå $\alpha = 1 \%$, vilken slutsats är korrekt? (5p)

- A. H_0 förkastas, fördelningen har förändrats
- B. H_0 förkastas inte, fördelningen har förändrats
- C. H_0 förkastas, samma fördelning nu som då
- D. H_0 förkastas inte, samma fördelning nu som då
- E. H_0 förkastas inte, fördelningen har förändrats

c) Inom vilket av följande intervall ligger p -värdet för detta test? (5p)

- A. 5 – 100 %
- B. 2,5 – 5 %
- C. 1 – 2,5 %
- D. 0,5 – 1 %
- E. 0,1 – 0,5 %

Fullständig redovisning krävs för följande uppgifter. Använd separata pappersark för uppgift 6 resp. uppgift 7.

Uppgift 6

Man samlar in data för att studera och jämföra andelen företag i två olika branscher med avseende på kvinnlig representation i företagsstyrelserna. Man får följande resultat:

$$\text{Bransch 1: } n_1 = 247 \quad \hat{p}_1 = 0,43$$

$$\text{Bransch 2: } n_2 = 352 \quad \hat{p}_2 = 0,46$$

- Testa på 5 % signifikansnivå om motsvarande populationsandelar (P_1 och P_2) kan antas vara lika eller inte. En noggrann och detaljerad lösning fordras. Men den här gången behöver du inte ange villkor och antaganden för testet. (10p)
- Beräkna ett 95 % konfidensintervall för P_2 vid användande av enbart stickprovet för bransch 2. Formel, formel med insatta numeriska värden och resultat fordras. (5p)
- Säg att man kan anta att P_1 och P_2 är lika, dvs. $P_1 = P_2 = P$. Beräkna ett 95 % konfidensintervall för P . Hur mycket kortare blir konfidensintervallet jämfört med det i b) ovan? (5p)

Uppgift 7.

En livsmedelskedja i USA genomförde ett försök där man ville studera hur $Y =$ veckoförsäljningen i antal paket av en viss kaffesort påverkades av $X =$ exponeringsytan i kvadratfot. Ytan varierades slumpmässigt mellan 3, 6 och 9 kvadratfot under tolv veckor. Andra kaffesorter upptog under hela tiden 3 kvadratfots exponeringsyta. Låt oss också göra antagandet att ingen prisvariation förekom under försöksperioden. Anta att sambandet kan beskrivas med en enkel linjär regressionsmodell:

$$Y_i = \beta_0 + \beta_1 X_i + \varepsilon_i$$

Observationerna från försöket ges på nästa sida tillsammans med en del beräkningar.

- Skatta parametrarna β_0 och β_1 i modellen med minsta-kvadrat-metoden och tolka parameterskattningen av β_1 . Är en motsvarande tolkning av skattningen av β_0 meningsfull i detta fall? Förklara kortfattat. (5p)
- I tabellen över observationerna finns residualerna e_i för observationerna $i = 1 - 9$ men de tre sista, $e_{10} - e_{12}$, saknas. Använd dina resultat i a) ovan för att beräkna dessa residualer. Beräkna därefter residualvariansen. OBS! Om du inte har ett resultat i a) kan du ange *hur* du beräkningarna ska utföras. (5p)
- Anta att du ska beräkna 95 % konfidensintervall för den genomsnittliga försäljningen när $X = 3, 6$ respektive 9 kvadratfot. För vilket av dessa värden på X blir intervallet kortast och varför? OBS! Du behöver inte göra omfattande beräkningar för att besvara frågan. (5p)
- Plotta residualerna e_i mot den förklarande variabeln x_i och kommentera resultatet. Är det något som ser konstigt ut som kan påverka vår tro på modellen eller ser allt bra ut? Förklara kortfattat. OBS! Om du inte har beräknat de tre residualerna som saknades i b) ovan, använd de värden som finns. (5p)

Bilaga till Uppgift 7

Data och residualer

Nr i	Försäljning y_i	Exponeringsyta x_i	Residual e_i
1	526	6	8,000
2	421	3	1,875
3	581	6	63,000
4	640	9	23,125
5	412	3	-7,125
6	500	9	-116,875
7	444	6	-74,000
8	443	3	23,875
9	580	9	-36,875
10	570	6	
11	376	3	
12	723	9	

Deskriptiv statistik

	Försäljning i antal y	Exponeringsyta x
Medelvärde	518	6
Standardfel	30,11644	0,73855
Standardavvikelse	104,32641	2,55841
Varians	10884	6,54545
Minimum	376	3
Q1	426,5	3
Median	513	6
Q3	580,75	9
Maximum	723	9
Summa	6216	72
Antal	12	12
Kovarians x, y	215,727	



Stockholms
universitet

Statistiska institutionen

Rättningsblad

Datum: 22/03/19

Sal: Brunnsvikssalen

Tenta: Statistik för ekonomer

Kurs: Grundläggande statistik för ekonomer

ANONYMKOD:

0008-KDR

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.	10	15	10	15	16	19			

POÄNG	BETYG	Lärarens sign.
95	A	MR

**SVARSBILAGA till Tentamen i Grundläggande statistik för ekonomer
2019-03-22**

Skrivsal: Brunnsvikssalen

Anonymkod: 0008-KDR (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"/>	5
	b)	<input type="checkbox"/>	<input checked="" type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	5
Uppgift 2	a)	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input checked="" type="checkbox"/>	5
	c)	<input type="checkbox"/>	<input type="checkbox"/>	<input checked="" type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	5
Uppgift 3	a)	<input type="checkbox"/>	<input checked="" type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	5
	b)	<input type="checkbox"/>	<input type="checkbox"/>	<input checked="" type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	5
	c)	<input checked="" type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	5
Uppgift 4	a)	<input checked="" type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	5
	b)	<input type="checkbox"/>	<input type="checkbox"/>	<input checked="" type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	5
Uppgift 5	a)	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input checked="" type="checkbox"/>	<input type="checkbox"/>	5
	b)	<input checked="" type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	5
		<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input checked="" type="checkbox"/>	5

60

11

b) a) $\alpha = 0,05$

$$\hat{p}_0 = \frac{(247 \cdot 0,43) + (352 \cdot 0,46)}{247 + 352} = 0,4476293823$$

$H_0: \hat{p}_2 = \hat{p}_1$

$H_1: \hat{p}_2 \neq \hat{p}_1$

dubbelsidigt test $Z_{\alpha/2} = 1,96 = Z_{KRIT}$

Om $|Z_{OBS}| > Z_{KRIT}$ förkastar H_0

test variabel $Z = \frac{\hat{p}_2 - \hat{p}_1 - 0}{\sqrt{\hat{p}_0(1-\hat{p}_0)(\frac{1}{n_2} + \frac{1}{n_1})}} \sim N(0,1)$

$$Z_{OBS} = \frac{0,46 - 0,43 - 0}{\sqrt{\hat{p}_0(1-\hat{p}_0)(\frac{1}{352} + \frac{1}{247})}} = 0,7268638573$$

SWAR a)

Slutsats: 0,7268638573 är inte större än 1,96 vilket innebär att vi på 5% nivån inte förkastar H_0 . På 5% nivån kan populationsandelarna antas vara lika.

10

b)

$$Z_{\alpha/2} = Z_{0,05/2} = 1,96$$

formel: $\hat{p} \pm z_{\alpha/2} \cdot \sqrt{\frac{\hat{p} \cdot (1-\hat{p})}{n}}$

$$0,46 \pm 1,96 \cdot \sqrt{\frac{0,46(1-0,46)}{352}}$$

KI: 95% $[0,41; 0,51]$

SWAR b)

felmarginal: 0,0520667578

avrundat till två decimaler; 0,05

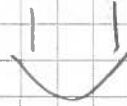
15

fortsättning \rightarrow

$$c) P_1 = P_2 = P \Rightarrow \hat{P}_1 - \hat{P}_2 = 0$$

$$Z_{\alpha/2} = Z_{0,05/2} = 1,96$$

$$\text{formel: } \hat{P}_1 - \hat{P}_2 \pm Z_{\alpha/2} \cdot \sqrt{\frac{\hat{P}_1(1-\hat{P}_1)}{n_1} + \frac{\hat{P}_2(1-\hat{P}_2)}{n_2}}$$
$$-0,03 \pm 1,96 \cdot \sqrt{\frac{0,43(1-0,43)}{247} + \frac{0,46(1-0,46)}{352}}$$



felmarginal: 0,0807650698

avrundat till

två decimaler: 0,08

SVAR c)

$$\text{KI}_{95\%}: [-0,11; 0,05]$$

avstånd intervall: 0,16

avstånd i b): 0,1

Konfidensintervallet i b) blir 0,06 enheter kortare än konfidensintervallet i c)

Detta är KI för differensen $P_1 - P_2$

Sök KI för $P = P_1 = P_2$

hur ska du skatta P?

7) a) $Y_i = \beta_0 + \beta_1 X_i + \epsilon_i \Rightarrow \hat{y} = b_0 + b_1 \cdot X$

$b_1 = \frac{Cov(x,y)}{S_x^2} = \frac{215,727}{6,54545} = 32,95831455$
 avrundas till heltal ≈ 33

$b_0 = \bar{y} - b_1 \bar{x} = 518 - 33 \cdot 6 = 320$

SVAR a)

$\hat{y} = 320 + 33x$

5

tolkning b_1 : för varje kvadratmeter exponeringsyta ökar vden genomsnittliga velo-försäljningen i antal paket med 33.

tolkning b_0 : om exponeringsytan är 0 så är den genomsnittliga velo-försäljningen 320. Men detta blir lite konstigt för om det inte finns någon produkt att köpa bilt försäljningen 0. Däremot behövs b_0 när $x \geq 1$ för att få den genomsnittliga velo-försäljningen.

b) $e_i = y_i - \hat{y}_i$

	y_i	\hat{y}_i	e_i
10	570	518	52
11	376	419	-43
12	723	617	106
$S_e^2 = 3136,350088$			

$S_e^2 = \frac{\sum_{i=1}^n e_i^2}{n - k - 1} = \frac{31363,50088}{10}$

$k=1$
 $n=12$

$\sum_{i=1}^n e_i^2 = 31363,50088$

e_i^2
64
3,515625
3969
534,673129
50,765625
13659,76563
5976
570,015625

W

c)

$$\hat{y} \pm t_{n-2, \alpha/2} \cdot \sqrt{S_e^2 \left(\frac{1}{n} + \frac{(x - \bar{x})^2}{(n-1)S_x^2} \right)}$$

SVAR c)

den delen som påverkas av att x byter värde är $(x - \bar{x})^2$ $\bar{x} = 6$

$$(n-1)S_x^2$$

vilket innebär att felmarginaler dvs.

hela uttrycket efter $\hat{y} \pm$ blir som minst

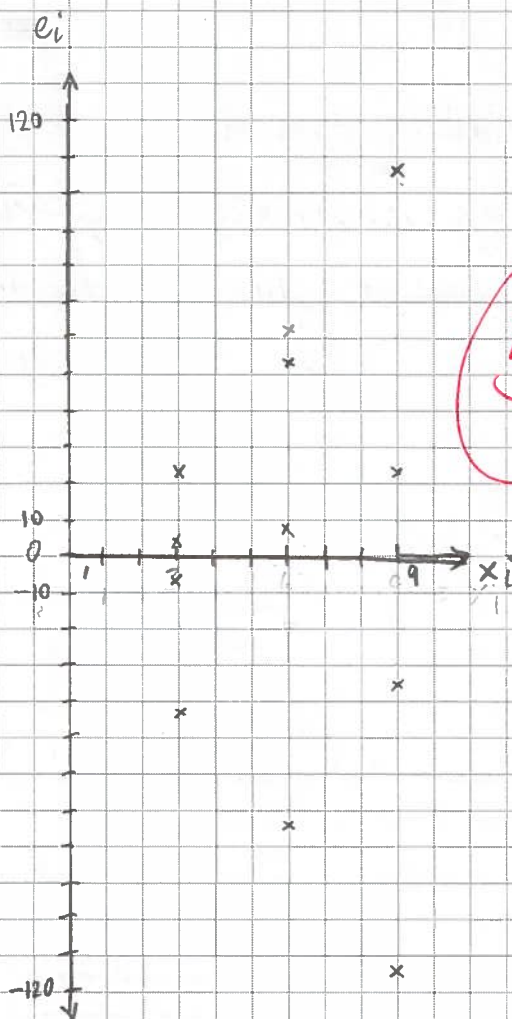
ju närmare x är \bar{x} . Eftersom $b = \bar{x}$ så

blir intervallet kortast då. 309 är lika långt

från \bar{x} så dessa kommer få samma felmarginal.

d)

SVAR d)



Residualerna ökar i variation desto större exponeringsyta, vilket tyder på att exponeringsyta sig självt förklarar inte variationen i y tillräckligt bra. Jag skulle inte rekommendera denna modell. Aningen komplettera med ex. prisdifferens eller byt ut modellen.

17