



Stockholms universitet

OBS! Läs noga igenom anvisningarna i tentamen, t.ex. hur du ska skriva svaren. Det är ditt ansvar som student att följa de anvisningar som ges.

NOTE! Read the examination instructions carefully, e.g. how to write the answers. It is your responsibility as a student to follow the given instructions.

Skriv din anonymiseringskod och dagens datum på allt material du lämnar in.
(Enter your anonymization code and today's date on all submitted materials)

Anonymiseringskod (Anonymization code)	3	1	1	-	0	1	0	9	-	J	G	N
Datum (Date YYYY-MM-DD)	2023-09-22						Plats nr. (Seat No.)	103				

Kurs/Kurskod (Course/Course code)	ST 1101
Kursmoment (Course component)	Statistik och dataanalys 1

Fylls i av tentamensvärd (To be filled in by invigilator)

Direkt i skrivning: (kryss)		Svarsblankett: (kryss)		Lösa svarsblad: (antal)	5
--------------------------------	--	---------------------------	--	----------------------------	---

Lämnat in blankt: (kryss)		Dator: (kryss)	
------------------------------	--	-------------------	--

Inlämningstid: 10:43

Signatur tentamensvärd: _____

Fylls i av lärare/examinator (To be filled in by teacher/examinator)

Betyg:	A	Poäng:	100
--------	---	--------	-----

Signatur rättande lärare/examinator: _____



Datum: (Date YYYY-MM-DD)	2023-09-22	Kurs/Kurskod: (Course/Course code)	ST1101	Sidnr.: (Page no.)	1								
Anonymiseringskod (Anonymization code)	3	1	1	-	0	1	0	9	-	J	G	N	1

a) Kategoriska variabler består av olika kategorier, med eller utan rangordning, t.ex om man är vegetarian, vegan eller äter kött. En numerisk variabel har rangordnade värden, t.ex hur ofta man äter kött under en vecka.	Uppg.nr.: (Task no.)	1
b) En outlier är en observation som tydligt avviker från resten av observationerna	Lärarens kommentar: (Teacher's note)	
c) Bias är systematiska fel i ett insamlat data-material som oftast beror på insamlingsmetoden. Exempel: En undersökning som vill undersöka välmåendet hos personer mellan 18 till 85 år skickar ut enkäten endast digitalt. Risken med att enbart använda digital metod är att äldre inte har samma tillgång och kunskap att kunna svara på enkäten och det blir ett systematiskt bortfall av äldre personer och därmed inte representativt för populationen, man vill undersöka		
d) En dold variabel är en variabel vi inte har med i vår modell men som <u>förtfarande påverkar sambandet i vår modell.</u> Exempel: Socioekonomisk bakgrund kan ha ett starkt samband med lön. Men en	Poäng: (Points)	

dold variabel kan vara utbildningsnivå som är en mellanliggande variabel. SocEk \rightarrow utbildning \rightarrow lön

Det är alltså socioekonomisk bakgrund som påverkar ens utbildningsnivå, som i sin tur påverkar lön.

Sambandet mellan soc-ek och lön försvinner om vi inkluderar utbildningsnivå.

Uppg.nr.:
(Task no.)

Lärarens
kommentar:
(Teacher's
note)

Poäng:
(Points)



Datum: (Date YYYY-MM-DD)	2023-09-22	Kurs/Kurskod: (Course/Course code)	ST1101	Sidnr.: (Page no.)	2
Anonymiseringskod (Anonymization code)	3 1 1 - 0 1 0 9 - J G N				

a, Förklaringsvariabel: variabeln som man förväntar sig ska påverka en annan variabel

Uppg.nr.:
(Task no.)

2

Responsvariabel: variabeln som man förväntar sig ska förändras när vi ändrar värde på vår förklaringsvariabel

Lärarens kommentar:
(Teacher's note)

4

b, Residualen är skillnaden mellan det observerade värdet i ett datamaterial och det predicerade värdet från en regressionsmodell

$$e = y - \hat{y}$$

4

- c, (1) Residualerna ska vara normalfördelade
(2) Residualerna ska ha en konstant varians

6

d, Att jämföra R-kvadrat mellan två modeller kan vara riskabelt då R-kvadrat ökar ju fler variabler modellen har, oavsett om de bidrar till modellen eller inte, därmed bör det undvikas. Att en modell är mer flexibel betyder inte heller att den är den bästa modellen, en för flexibel modell kan vara överanpassad till datan och inte fånga det generella mönstret på ett bra sätt. Den andra ingenjären har därmed inte rätt i sitt resonemang.

Adjusted R-kvadrat eller jämföra vilken

Poäng:
(Points)

modell som har lägst RMSE är att föredra
för att utvärdera vilken modell som är bäst

Uppg.nr.:
(Task no.)

Lärarens
kommentar:
(Teacher's
note)

6

Bäst!

Poäng:
(Points)



a) $n = 138 + 101 + 60 + 60 + 47 + 39 = 445$

svar: 445 kunder svarade

Uppg.nr.: (Task no.)

3

b) svar: 39 kunder i Malmö var missnöjda

Lärarens kommentar: (Teacher's note)

c) $\frac{1}{1} \dots$

Stockholm: $138 + 60 = 198$

$$\frac{198}{445} \approx 0,45 = 45\%$$

Göteborg: $101 + 47 = 148$

$$\frac{148}{445} \approx 0,33 = 33\%$$

Malmö: $60 + 39 = 99$

$$\frac{99}{445} \approx 0,22 = 22\%$$

100%

Ja: $138 + 101 + 60 = 299$

$$\frac{299}{445} \approx 0,67 = 67\%$$

Nej: $60 + 47 + 39 = 146$

$$\frac{146}{445} \approx 0,33 = 33\%$$

100%

Tolkning: Totalt sett var de flesta kunder som svarade från Stockholm, 45% av alla svarande. 67% av alla svarande kunder var nöjda med sitt senaste köp

4

d) Stockholm

Ja $138/198 \approx 0,697 = 69,7\%$

Nej $60/198 \approx 0,303 = 30,3\%$

Göteborg

Ja $101/148 \approx 0,682 = 68,2\%$

Nej $47/148 \approx 0,318 = 31,8\%$

Poäng: (Points)

Malmö

Ja $60/99 \approx 0,606 \approx 60,6\%$

Nej $39/99 \approx 0,394 \approx 39,4\%$

e, När man jämför kundnöjdhet mellan städerna ser man att Stockholm har högst andel nöjda kunder med 69,7% nöjda kunder, följt av Göteborg på 68,2%. Staden med högst andel missnöjda kunder är Malmö med 39,4%. Skillnaderna mellan Stockholm och Göteborg är väldigt små, medan Malmö skiljer sig mer från de andra städerna.

Uppg.nr.:
(Task no.)

Lärarens
kommentar:
(Teacher's
note)

8

4

Poäng:
(Points)



Datum: (Date YYYY-MM-DD)	2023-09-22	Kurs/Kurskod: (Course/Course code)	ST1101	Sidnr.: (Page no.)	
Anonymiseringskod (Anonymization code)	3 1 1 - 0		1 0 9 - J G N		4

a) $\bar{x} = 55$ poäng $s_x = 8$ poäng

$$z_x = \frac{x - \bar{x}}{s_x} = \frac{59 - 55}{8} = 0,5 \rightarrow 0,6915$$

Andelen 0,6915 har lägre poäng än Anna.

$$1 - 0,6915 = 0,3085 \approx 31\%$$

svar: ungefär 31% av alla som gjort proveret har fått högre poäng än Anna

b) 90:e percentilen $\rightarrow z = 1,28$

$$z_x = \frac{x - \bar{x}}{s_x}$$

$$1,28 = \frac{x - 55}{8}$$

$$x = 55 + 1,28 \cdot 8$$

$$x = 65,24 \approx 65 \quad \text{66 avrundat uppåt.}$$

svar: 65 poäng behövs för att tillhöra de högsta 10 procenten.

c) IQR = Q3 - Q1 | Q3 = 75% Q1 = 25%

$$75\% \rightarrow z = 0,67 \rightarrow 0,67 = \frac{x - 55}{8}$$

$$x = 55 + 0,67 \cdot 8$$

$$x = 60,36$$

Uppg.nr.:
(Task no.)

4

Lärarens
kommentar:
(Teacher's
note)

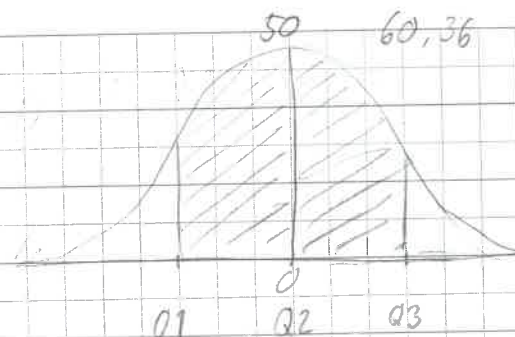
4

4

Poäng:
(Points)

$$Q_3 = 60,36 \text{ poäng}$$

$$\bar{x} = Q_2 = 55 \text{ poäng}$$



Uppg.nr.:
(Task no.)

Lärarens
kommentar:
(Teacher's
note)

$$\text{Skillnad mellan } Q_3 \text{ och } Q_2 = 60,36 - 55 = 5,36 \text{ poäng}$$

$$\text{Skillnad mellan } Q_1 \text{ och } Q_2 \text{ är också } 5,36 \text{ poäng}$$

$$Q_1 = 55 - 5,36 = 49,64$$

$$\left(z = \frac{49,64 - 55}{8} = -0,67 \rightarrow 0,7486 \right)$$

$$1 - 0,7486 = 0,2514 \approx 25\%$$

$$IQR = Q_3 - Q_1 = 60,36 - 49,64 = 10,72$$

svar: IQR är 10,72 poäng

$$d, \quad IQR = Q_3 - Q_1$$

$$Q_3 = \log(4,1) = e^{\log(4,1)} = 60,39$$

$$Q_1 = \log(3,9) = e^{\log(3,9)} = 49,40$$

$$IQR = 60,39 - 49,40 = 10,99$$

svar: IQR är 10,99, ungefär detsamma som
svaret i deluppgift c.

Poäng:
(Points)

5

7



a) Intercept: När medeltemperaturen är 0 säljs i genomsnitt 132 glassar

Uppg.nr.: (Task no.)

5

Lutningskoefficienten: När temperaturen ökar med 1°C förväntas antalet sålda glassar öka med 17 st den dagen, i genomsnitt

Lärarens kommentar: (Teacher's note)

4

b) $\hat{y}_{24} = 132 + 17 \cdot 24 = 540$

4

svår: När det är 24 grader C förväntas glasskiöskan sälja 540 glassar

c) r_{xy} = korrelationskoefficienten mellan medeltemp och antal sålda glassar
 s_x = standardavvikelsen för medeltemperaturen
 s_y = standardavvikelsen för antalet sålda glassar
 \bar{x} = medelvärdet av uppmätta medeltemperaturen
 \bar{y} = medelvärdet av antalet sålda glassar

Verifiering:

$$b_1 = r_{xy} \cdot \frac{s_y}{s_x} = 0,7 \cdot \frac{190}{7,8} \approx 17,05 \approx 17$$

$$b_0 = \bar{y} - b_1 \bar{x} = 439 - 17,05 \cdot 18 = 132,1 \approx 132$$

4

d) b_0 : När medeltemperaturen är 0 grader C och det inte regnar säljs i genomsnitt 290 glassar

Poäng: (Points)

b_1 : När medeltemperaturen ökar med 1 grad C förväntas glassförsäljningen öka i genomsnitt med 8 glassar, givet att regn hålls konstant

b_2 : Dagar då det regnar minskar försäljningen av glassar i genomsnitt med 206 glassar jämfört med dagar då det inte regnar, givet att medeltemperaturen är densamma.

$$e) \hat{y} = 290 + 8x - 206x \quad \begin{array}{l} x=1 \text{ regn} \\ x=0 \text{ inte regn} \end{array}$$

$$\hat{y}_{x=1} = 290 + 8x - 206 \cdot 1$$

$$\hat{y}_{x=1} = 84 + 8x \rightarrow \text{regn}$$

$$\hat{y}_{x=0} = 290 + 8x \rightarrow \text{inte regn}$$

$$\hat{y} = 290 + 8x + 206x \quad \begin{array}{l} x=0 \text{ regn} \\ x=1 \text{ inte regn} \end{array}$$

$$\boxed{\hat{y} = 84 + 8x + 206x}$$

$$\hat{y}_{x=0} = 84 + 8x - 206 \cdot 0$$

$$\hat{y}_{x=0} = 84 + 8x \rightarrow \text{regn}$$

$$\hat{y}_{x=1} = 84 + 8x + 206 \cdot 1$$

$$\hat{y}_{x=1} = 290 + 8x \rightarrow \text{inte regn}$$

Ekvationen blir densamma oavsett om vi ledar regn som 1 eller 0

Regler i skrivsalen

- Följ tentamensvärds anvisningar.
- Väskor och ytterkläder ska placeras på anvisad plats.
- Placera ID-handling väl synlig på bordet framför dig.
- Ingen student får lämna skrivsalen under de första 30 minuterna.
- Endast en student i taget får besöka toaletten. Vid toalettbesök skriv ditt namn och klockslag på avsedd lista. Efter toalettbesöket ska du åter ange klockslag på listan.
- Elektronisk utrustning som mobiltelefon eller Smartwatch ska vara avstängd och placerad på anvisad plats.
- Under tentamen gäller tystnad – det är förbjudet att prata, eller på annat sätt kommunicera, med andra studenter under pågående tentamen.
- Innan tentamenshandlingarna lämnas in; skriv sidnummer, anonymiseringskod och datum på alla inlämnade papper.

Om något är oklart – fråga gärna tentamensvärden. Lycka till!

Rules in the examination hall

- Follow the invigilator's instructions.
- Bags and outerwear must be placed at the designated place.
- Place your ID document clearly visible on the table in front of you.
- No student may leave the examination hall for the first 30 minutes.
- Only one student at a time may visit the toilet. Before visiting the toilet, write your name and time on the intended list. After the toilet visit, enter the time on the list again.
- Electronic equipment such as a mobile phone or Smartwatch must be switched off and placed at the designated place.
- During the exam, silence applies – you are not allowed to talk, or otherwise communicate, with other students during the exam.
- Before submitting the examination documents; remember to write the page number, anonymization code, and date on all papers.

Please do not hesitate to ask the invigilator if anything is unclear. Good luck!