



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	0	0	4	-	E	P	E
Datum (Date YYYY-MM-DD)	2023/03/27							Plats nr. (Seat No.)		7		

Kurs/Kurskod (Course/Course code)	ST1101
Kursmoment (Course component)	Tentamen 1

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

Direkt i skrivning: (kryss)		Svarsblankett: (kryss)		Lösa svarsblad: (antal)	3
--------------------------------	--	---------------------------	--	----------------------------	---

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

Inlämningstid: 11:09

Signatur tentamensvärd: Be

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

Betyg:	B	Poäng:	83
--------	---	--------	----

Signatur rättande lärare/examinator: Chng





1) $n = 504$ $\bar{x} = 140,2971$
 $\bar{y} = 139,3844$

Rökare = $\frac{171}{504}$

icke rökare = $\frac{333}{504}$

3p

Uppg.nr.:
(Task no.)

1

Lärarens kommentar:
(Teacher's note)

a)

	antal personer	
Rökare	171	171
icke Rökare	333	333
Marginal	504	504

b) Vi ser för rökare är medelvärdet för systoliskt blodtryck högre än för icke rökare. Rökarens medelvärde $\approx 140,3$ medan icke rökarens medelvärde är $\approx 139,4$. Vi ser att det finns några få outliers i båda fördelningarna. Icke rökarens Q_4 är ungefär på samma nivå som Rökarens Q_1 , så vi kan generellt säga att Rökare har högre blodtryck än icke Rökare. Den nedre kvartilen för Rökare är på \approx samma nivå som den tredje Q_3 för icke rökare. Vilket innebär att $\approx 75\%$ av icke rökare har lägre blodtryck än de lägsta datan för Rökare. Vidare ser vi att spridningen verkar vara jämnt fördelad i de båda grupperna och det verkar som att variansen är liknande mellan de två olika grupperna

	min	Q1	medel	Q3	Max
icke rökare:	$\approx 138,6$	138,7	139,3844	$\approx 139,7$	$\approx 140,2$
Rökare:	$\approx 139,5$	139,7	140,2971	$\approx 140,4$	$\approx 141,3$

5p
Poäng:
(Points)

c) $n=6$ $\bar{x} = \frac{\sum x_i}{n}$ $s_x^2 = \frac{\sum (x_i - \bar{x})^2}{n-1}$ $s_x = \sqrt{s_x^2}$

Sidnr:
(Page no.)

2

140,44
141,16
139,67
139,36
139,68
139,29
839,6

$\bar{x} = \frac{839,6}{6} \approx 139,93333$

Uppg.nr.:
(Task no.)

1

Lärens
kommentar:
(Teacher's
note)

139,29 139,36 139,67 139,68 140,44 141,16

median = $139,67 + 139,68 = 139,525$

Obs	$x - \bar{x}$	$(x - \bar{x})^2$
1	$140,44 - 139,93333 = 0,50667$	0,256714488
2	$141,16 - 139,93333 = 1,22667$	1,504719289
3	$139,67 - 139,93333 = -0,26333$	0,069342688
4	$= -0,57333$	0,328707288
5	$= -0,25333$	0,0064176088
6	$139,29 - 139,93333 = -0,64333$	0,413873488
		2,579774851

$s_x^2 = \frac{2,579774851}{6-1} = 0,51595497$

$s_x = \sqrt{0,51595497} = 0,71830055$

c) Stickprovets standardavvikelse är $\approx 0,7183$ vilket

innebär att de observerade personernas systoliska

blodtryck skiljer sig i genomsnitt $0,71833 \text{ mmHg}$

från det observerade medelvärdet $139,9333 \text{ mmHg}$.

Alltså de flesta observationer (68%) hittar vi

en standardavvikelse från medelvärdet. $139,933 \pm 0,7183$

/9 p

d) Vi ser att histogrammet är Unimodalt och normal-fördelat. Det verkar inte finnas några särskilda outliers och fördelningen är väldigt symmetrisk. Histogrammet är inte skevt, men det finns en tendens till en lite längre svans på höger sida.

Histogrammet är bell-shaped och vi ser att antagandet om att det inte finns ett samband mellan rökning och systoliskt blodtryck stämmer då skillnaden mellan rökare och icke rökare är centrerad vid 0

Op.

17

Poäng:
(Points)



a)

	Under	Över	Hälsosam	
Man	295	72	73	440
Kvinna	560	163	37	760
Marginal	855	235	110	Total: 1200

Ingen förd. ans. vem. / 0 p

Uppg.nr.:
(Task no.)
2

Lärarens kommentar:
(Teacher's note)

b) $\frac{72}{1200} = 0,06$ / 3 p

c)

	Under	Över	Hälsosam	
Man	295/1200	72/1200	73/1200	440/1200
Kvinna	560/1200	163/1200	37/1200	760/1200
	855/1200	235/1200	110/1200	1

/ 5 p.

d)

	Kvinna	Man
Under	0,7368	0,16705
Över	0,2145	0,1636
Hälsosam	0,0487	0,1659

Det verkar som att fler män tenderar att ha en hälsosam kroppsutfattning. En större andel av kvinnorna anser sig vara överviktiga i jämförelse med männen. Bland både män och kvinnor anser de flesta att de är underviktiga.

/ 6 p.

e) Kvinnor $N(68, 10)$

$Z_{Elsa} = -0,2$

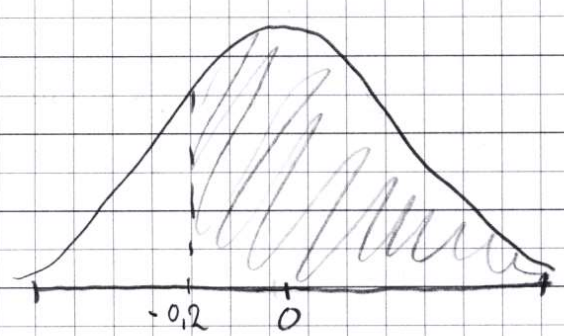
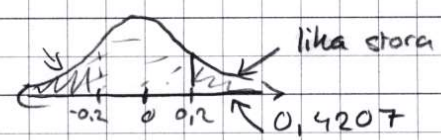
$X > P(z = -0,2) = 1 - Z(-0,2)$

$= 1 - 0,4207 = 0,5793$

$Z(0,2) = 0,5793$

Vi kan använda oss av symmetrifördelning

$1 - 0,5793 = 0,4207$



$Z = \frac{X - \bar{x}}{S_x}$

$-0,2 = \frac{X - 68}{10}$

$X = 10(-0,2) + 68$

$X = 66$

22

e) 57,93% av kvinnorna väger mer än Elsa som väger 66kg

Poäng:
(Points)
8 p

3.) $n=102$

Sidnr.:
(Page no.)

4

Uppg.nr.:
(Task no.)

3

Lärarens
kommentar:
(Teacher's
note)

$$\hat{y} = b_0 + b_1 x$$

$$r_{xy} = \frac{S_{xy}}{S_x \cdot S_y}$$

$$S_{xy} = \frac{\sum (x_i - \bar{x})(y_i - \bar{y})}{n-1}$$

$$b_1 = \frac{S_{xy}}{S_x^2} = \boxed{r_{xy}} \cdot \boxed{\frac{S_y}{S_x}}$$

$$b_0 = \bar{y} - b_1 \bar{x}$$

$$S_x^2 = \frac{\sum (x_i - \bar{x})^2}{n-1} = \frac{751,8852}{101} = \approx 7,4444$$

$$S_x = \sqrt{S_x^2} = \sqrt{7,4444} \approx 2,7284$$

$$S_y^2 = \frac{\sum (y_i - \bar{y})^2}{n-1} = \frac{1820813411}{101} = \approx 18027855,55$$

$$S_y = \sqrt{S_y^2} = \sqrt{18027855,55} = \approx 4245,9222$$

$$S_{xy} = \frac{\sum (x_i - \bar{x})(y_i - \bar{y})}{n-1} = \frac{675804,1}{101} \approx 6691,129703$$

$$r = \frac{S_{xy}}{S_x \cdot S_y} = \frac{6691,129703}{2,7284 \cdot 4245,9222} \approx 0,5775896$$

$$b_1 = r \cdot \frac{S_y}{S_x} = 0,5775896 \cdot \frac{4245,9222}{2,7284} = 898,8420193$$

$$\bar{x} = \frac{\sum x_i}{n} = \frac{1095,28}{102} = 10,73803922$$

$$\bar{y} = \frac{\sum y_i}{n} = \frac{693586}{102} = 6797,901961$$

$$b_0 = \bar{y} - b_1 \bar{x} = 6797,901961 - 898,8420193(10,73803922) = -2853,89895$$

a) $\hat{y} = -2853,8989 + 898,8420 \text{ medelstudietid}$

Då studietiden är 0 så är medelinkomsten $-2853,8989$ vilket inte är möjligt. Därför har interceptet i denna modell ingen betydelse.

Vi kan se att medelinkomsten ökar i genomsnitt med \$898,8420 per år av studietid.

/ 8 p.

b) $R^2 = r^2 = 0,5775896^2 = 0,33360974$

$\approx 33,36\%$ av variationen i medelinkomst kan förklaras av variabeln medelstudietid

/ 3 p

Poäng:
(Points)



Datum: (Date YYYY-MM-DD)	2023-03-27	Kurs/Kurskod: (Course/Course code)	ST101	Sidnr.: (Page no.)	
Anonymiseringskod (Anonymization code)	311-0004-EP E				5

c)	$se = \sqrt{Se^2}$ $Se^2 = \frac{\sum (y_i - \hat{y}_i)^2}{n-2} = \frac{1213392025}{100} = 1213392,25$ $se = \sqrt{1213392,25} = 1101,540853$	Uppg.nr.: (Task no.)	3
c)	$Se = \$1101,540853$ De faktiska värdena för medelinkomst skiljer sig med en standardavvikelse på \$1101,54 från de predikterade värdena. / 3 p.	Lärarens kommentar: (Teacher's note)	
d)	Medel inkomst 8403 Medel studietid 14,62 $\hat{y} = -2853,8989 + 898,8420(14,62) = 10287,17114$ $8403 - 10287,17114 = -1884,17114$		
d)	Residual = -1884,17114 USD / 3 p.		
e)	Om medelstudietiden hålls konstant kommer medelinkomsten att öka i genomsnitt med värdet på b_2 om "minst 80% har en universitetsutbildning". Om dummyvariabeln inte är sann, det vill säga "mindre än 80% har en universitetsutbildning" kommer inte variabeln påverka medelinkomsten. Jag förväntar mig att b_2 kommer vara positiv då längre studietid ger högre lön så förväntar jag mig ett liknande positivt samband mellan lön och universitetsutbildning.		
	8 p URA! 25		
		Poäng: (Points)	

4a) a) $\widehat{\text{age}} = -11,41659 + 0,16225 \text{ belly} + 0,10685 \text{ hdlength}$
 $b_0 = -11,41659$ $b_1 = 0,16225$ $b_2 = 0,10685$

b_1 = om variabeln huvudlängd i cm hålls konstant kommer ålder i år i genomsnitt att öka med 0,16225/midjeomkrets i cm

b_2 = Om variabeln midjeomkrets i cm hålls konstant kommer ålder i år i genomsnitt att öka med 0,10685/huvudlängd i cm

Residualernas medelvärde går ej att avläsa från modellen. Däremot kan vi se en residual analys med min, Q1, median, Q3 och Max.

$\bar{e} = 0$ för minsta kvadratmetoden (1-2p)

b) $R^2 = 0,2142$ 21,42% av variationen i $\log \text{Age}$ kan förklaras av variablerna belly och hdlength

c) $\widehat{\log \text{Age}} = -4,42863 + 0,05562 \text{ belly} + 0,04104 \text{ hdlength}$
 $-4,42863 + 0,05562(34,5) + 0,04104(90,7) = 1,212588$

Svar: 1,212588 $\log \text{År}$ eller 16,315 år $10^{1,212588} = 16,315$
 expl. : ? ~~no~~ - ok! (5p)

d) $\frac{100}{4} = 25$

$n=100$ $K=4$ $(25\% \cdot 4 = 100\%$

Svar: Inom en given fold innehåller 25% testdata alltså $0,25 \times 100 = 25\%$ testdatan och 75% träningsdata. Detta beror på att vi har 4 foldar och vi vill testa modellerna med 100 observationer som är uppdelat i 4 foldar. Totalt sett vill vi alltså testa 100 observationer på "4 försök" och delar upp varje fold i 25% testdata och 75% träningsdata

3p

19



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!