STOCKHOLMS UNIVERSITET
Statistiska institutionen
Michael Carlson (examinator)
Ulf Högnäs

HT2018

TENTAMEN I GRUNDLÄGGANDE STATISTIK FÖR EKONOMER

2018-11-29

Skrivtid:	kl. 15.00 - 20.00
Godkända hjälpmedel:	Miniräknare utan lagrade formler och text
Bifogade hjälpmedel:	Häftet *Formelsamling och Tabeller över statistiska fördelningar* (återlämnas efter skrivningen)

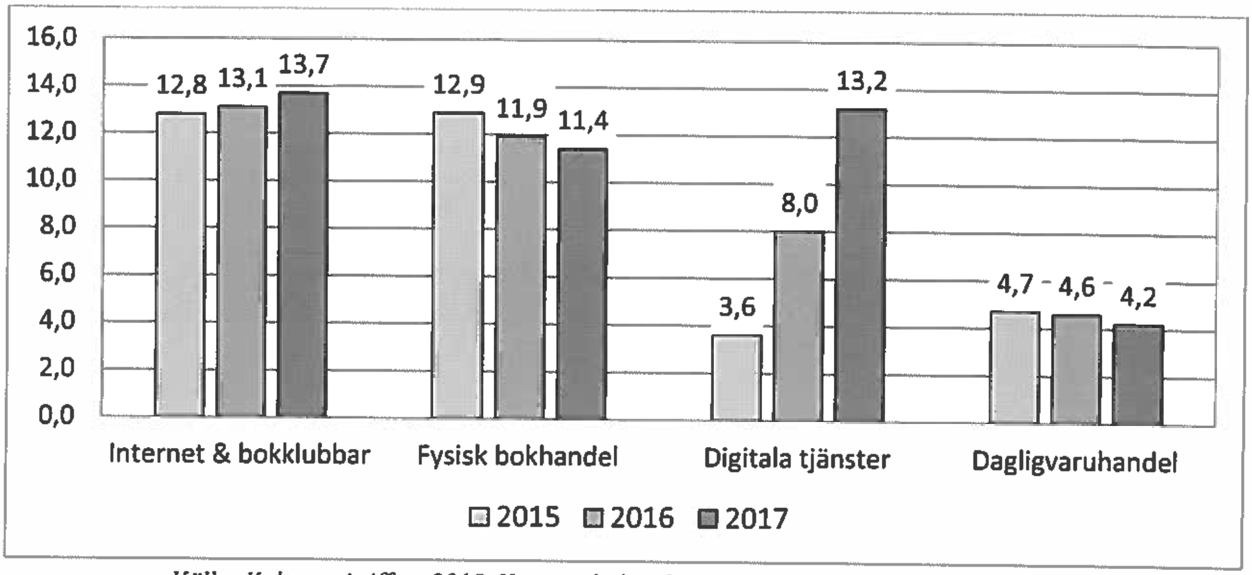
- Tentamen består av 7 uppgifter, i förekommande fall uppdelade i deluppgifter. Maximalt antal poäng anges per deluppgift.
- **Uppgift 1 – 5**: Svar lämnas på särskild **SVARSBILAGA**,
  - Flervalsfrågor där ett av fem alternativ är korrekt svar.
  - Har fler än ett svarsalternativ markerats för en deluppgift 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.
- Lösningsförslag läggs ut på Mondo kort efter tentamen.

LYCKA TILL!

Uppgift 1



Källa: *Kulturen i siffror 2018*; Kort om kultur 2018:1. Myndigheten för kulturanalys.

a) Diagrammet ovan visar antalet sålda böcker (fysiska och digitala) i miljoner via olika försäljningskanaler åren 2015–2017. Siffrorna omfattar ca 75 % av all litteratur som säljs i Sverige. Vilket av följande påståenden är <u>inte</u> rätt? (5p)

A. Antal sålda böcker är en diskret numerisk variabel på kvotskalenivå.

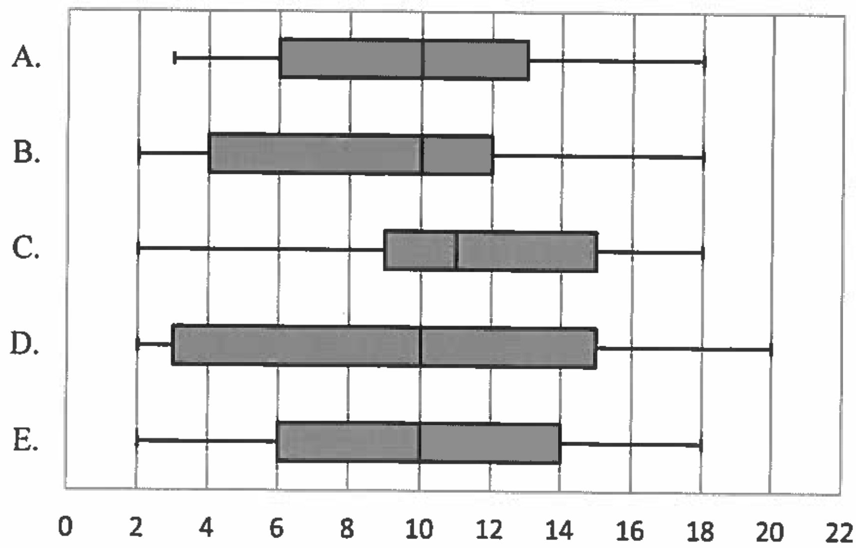
B. Årtal är en diskret numerisk variabel på kvotskalenivå.

C. Försäljningen av digitala tjänster 2017 ökade med ca 65 % jämfört med året innan.

D. Försäljningskanal är en kategorisk variabel på nominalskalenivå.

E. Totalt sett ökar bokförsäljningen i Sverige.

b) Du får ett datamaterial bestående av $n = 9$ numeriska värden: $5, 12, 6, 10, 3, 11, 2, 18, 12$. Vilken av boxplottarna A-E i diagrammet nedan hör ihop detta material? (5p)



Uppgift 2

Låt X = antalet kontrakt som ett företag som säljer systemlösningar tar hem per månad. Baserat på försäljningsstatistik har man funnit att sannolikhetsfunktionen för X med god approximation beskrivas med funktionen

$$P(X = x) = \frac{4 - x}{10} \quad \text{för } x = 0,1,2,3$$

a) Ange väntevärde och varians för X. (5p)

A. $\mu_X = 1{,}0$ $\quad\sigma_X^2 = 1{,}414$

B. $\mu_X = 1{,}5$ $\quad\sigma_X^2 = 1{,}667$

C. $\mu_X = 1{,}5$ $\quad\sigma_X^2 = 1{,}25$

D. $\mu_X = 1{,}0$ $\quad\sigma_X^2 = 2{,}0$

E. $\mu_X = 1{,}0$ $\quad\sigma_X^2 = 1{,}0$

Företagets kunder kan välja olika applikationer som tillägg till sina system. Företaget grupperar dessa applikationer i tre olika kategorier som vi kan benämna A, B och C. Baserat på tidigare erfarenheter och pga. av hur systemen är konstruerade vet man att A väljs oberoende av både B och C samt att B och C är disjunkta, dessa två typer kan inte ingå i samma system. Slutligen har man att sannolikheterna för var och en av de tre kategorierna är $P(A) = 0{,}40$; $P(B) = 0{,}30$ och $P(C) = 0{,}20$.

b) Vad är sannolikheten att inget tillägg väljs av en kund? TIPS: Rita ett Venndiagram! (5p)

A. 0,1

B. 0,36

C. 0,336

D. 0,3

E. Kan inte beräknas med ledning av ovanstående (information saknas)

Uppgift 3

Aktier i ett visst företag handlas i stor mängd på Nasdaq Stockholm och enligt en värderingsmodell är $X =$ värdet på företagets aktie om tre månader en normalfördelad slumpvariabel med väntevärde $\mu_X = 200$ kr och variansen $\sigma_X^2 = 25$. Du köper 10 000 aktier för 196 kr styck.

a) Vad är sannolikheten att din totala vinst överstiger 100 000 kr? (5p)

A. 0,885

B. 0,595

C. 0,405

D. 0,115

E. 0,190

Låt $X =$ beloppet i kronor som en slumpvist utvald kund handlar för i en viss butik under veckorna före jul. Beloppet är alltså en slumpvariabel och den kan dessutom antas vara normalfördelad med väntevärde $\mu = 800$ och standardavvikelse $\sigma = 225$. Man tar ett stickprov på $n = 9$ kunder och beräknar stickprovsmedelvärdet.

b) Vad är sannolikheten att stickprovsmedelvärdet är <u>mindre</u> än 700 kr? (5p)

A. 0,908

B. 0,092

C. 0,123

D. 0,903

E. 0,056

I butiken lockar man kunderna med ett specialerbjudande; om de registrerar sig som VIP-kunder för att få reklam och erbjudanden via e-post så får de omedelbart en bonus på 50 kr som de kan handla för. Anta att sannolikheten att en slumpvist utvald kund ska nappa på erbjudandet är 0,40.

c) Vad är sannolikheten att fler än 5 kunder av 12 tillfrågade ska nappa på erbjudandet? (5p)

A. 0,335

B. 0,842

C. 0,158

D. 0,665

E. 0,438

OBS! Svarsalternativen i a) – c) har avrundats till 3 decimaler.

Uppgift 4

Ett flygbolag ville undersöka sina kunders vilja att betala en extra avgift på 10 € för att få tillgång till WiFi och Internet under bolagets långflygningar. Man frågade $n = 100$ flygresenärer och av dessa svarade 55 att de var villiga att betala den extra avgiften.

a) Ange vilket av följande alternativ som är ett 95 % konfidensintervall för $P =$ andelen resenärer som är villiga att betala avgiften. (5p)

A. (0,452 ; 0,648)

B. (0,468 ; 0,632)

C. (0,545 ; 0,555)

D. (0,540 ; 0,560)

E. (0,352 ; 0,548)

Flygbolaget var mycket intresserade av resultaten men tyckte att felmarginalen var för stor. Man ville utöka undersökningen genom att fråga en ännu större grupp och därmed dra ner felmarginalen till högst 0,01 men med bibehållen konfidens, dvs. 95 %.

OBS! När du beräknar den totala stickprovsstorleken n som behövs för att få ner felmarginalen till högst 0,01 ska du använda det värde på P som ger störst varians och därmed störst felmarginal. Vilket är det värdet?

b) Ange den totala stickprovsstorleken som krävs för att felmarginalen garanterat ska vara högst 0,01? (5p)

A. $n = 97$

B. $n = 4900$

C. $n = 6765$

D. $n = 9508$

E. $n = 9604$

Uppgift 5.

Olika sätt att exponera en produkt i butiker, dvs. skyltning, placering mm., kan ge olika resultat i försäljning. Ett företag som tillverkar batterier testar två olika sätt att exponera sina produkter i butiker där produkterna säljs; låt A respektive B beteckna de två olika sätten.

Du ska nu testa om plan A är bättre än plan B dvs. om den genomsnittliga försäljningen μ_A för plan A är större än den genomsnittliga försäljningen μ_B för plan B.

a) Vilket av följande alternativ är korrekta noll- och mothypoteser? (4p)

A. $H_0: \mu_A = \mu_B$ mot $H_1: \mu_A \neq \mu_B$

B. $H_0: \mu_A \neq \mu_B$ mot $H_1: \mu_A = \mu_B$

C. $H_0: \mu_A = \mu_B$ mot $H_1: \mu_A > \mu_B$

D. $H_0: \bar{x}_A = \bar{x}_B$ mot $H_1: \bar{x}_A > \bar{x}_B$

E. $H_0: \mu_A > \mu_B$ mot $H_1: \mu_A = \mu_B$

Ett urval av $n_A = 30$ butiker exponerar produkterna enligt plan A och ett annat urval av $n_B = 35$ butiker exponerar produkterna enligt plan B. Efter undersökningsperioden fick man stickprovsmedelvärdena $\bar{x}_A = 810$ och $\bar{x}_B = 700$ och standardavvikelserna $s_A = 200$ och $s_B = 300$ för respektive plan.

b) Ange det observerade värdet på din testvariabel. (6p)

A. 1,760

B. 2,460

C. 1,261

D. −1,760

E. 1,708

c) Givet att du använder signifikansnivån $\alpha = 0{,}05$, vilket av följande alternativ är en korrekt slutsats? (5p)

A. $|z_{\text{obs}}| < z_{0,025} = 1{,}96 \Rightarrow H_0$ förkastas inte

B. $z_{\text{obs}} > z_{0,05} = 1{,}6449 \Rightarrow H_0$ förkastas

C. $t_{\text{obs}} > t_{63;0,05} \approx 1{,}670 \Rightarrow H_0$ förkastas inte

D. $|t_{\text{obs}}| < t_{63;0,025} \approx 1{,}998 \Rightarrow H_0$ förkastas inte

E. $z_{\text{obs}} > -z_{0,05} = -1{,}96 \Rightarrow H_0$ förkastas

Fullständig redovisning krävs för följande uppgifter.

Använd separata pappersark för uppgift 6 resp. uppgift 7.

Uppgift 6

En rikstäckande snabbmatskedja genomförde en undersökning för att bl.a. ta reda på vilket betalningssätt som deras kunder föredrar och faktiskt använder; Swish, bank- eller kreditkort alternativt kontanter. Under en vecka valdes slumpvist $n = 152$ kunder, man noterade vilket betalningssätt som kunden valde samt frågade dem bland annat om deras ålder. Kunderna delades in i tre åldersgrupper. Man fick följande resultat:

Antal	Swish	Kort	Kontanter	Totalt
Unga	12	22	4	38
Medel	16	52	8	76
Äldre	4	26	8	38
Totalt	32	100	20	152

a) Testa på 5 % signifikansnivå om betalningssätt som kunderna väljer beror på ålder. Redovisa antaganden, hypoteser, testvariabel samt beslutsregel och kritisk gräns. (8p)

b) Redovisa sedan dina beräkningar, dra en slutsats och tolka slutsatsen i ord. (8p)

c) Förklara kortfattat vad som menas med begreppet *iid* och varför det ofta är önskvärt att ett stickprov är *iid*. (4p)

Uppgift 7

Man ville veta om studenters resultat på en delkurs B i ett visst ämne kan förklaras av deras resultat på delkurs A (som kommer före B) och deras gymnasiebetyg. Med ett stickprov av storlek $n = 30$ studenter skattade man två linjära regressionsmodeller:

Modell 1: $Y_i = \beta_0 + \beta_1 X_{1i} + \varepsilon_i$

Modell 2: $Y_i = \beta_0 + \beta_1 X_{1i} + \beta_2 X_{2i} + \varepsilon_i$

där Y = poäng på B-kursens sluttentamen, X_1 = poäng på A-kursens sluttentamen och X_2 är betyg från gymnasiet (på en skala 0-20). På följande sida finns Excel-utskrifter och sammanfattande statistik för de tre variablerna. Använd dessa för att besvara följande frågor:

a) Testa hypotesen att lutningskoefficienten β_2 är skild från noll givet att β_1 är skild från noll i Modell 2. Du behöver inte redovisa vilka antaganden som krävs. (8p)

b) Beräkna med hjälp av Modell 1 ett 95 % prediktionsintervall för en slumpmässigt vald students poäng på sluttentamen i delkurs B, givet att studenten fick X_1 =80 poäng för delkurs A. Förklara hur intervallet ska tolkas. (8p)

a) Förklara <u>kortfattat</u> begreppet multikollinearitet. Varför är det problematiskt? (4p)

Fullständig redovisning krävs för följande uppgifter.

Använd separata pappersark för uppgift 6 resp. uppgift 7.

Uppgift 6

En rikstäckande snabbmatskedja genomförde en undersökning för att bl.a. ta reda på vilket betalningssätt som deras kunder föredrar och faktiskt använder; Swish, bank- eller kreditkort alternativt kontanter. Under en vecka valdes slumpvist $n = 152$ kunder, man noterade vilket betalningssätt som kunden valde samt frågade dem bland annat om deras ålder. Kunderna delades in i tre åldersgrupper. Man fick följande resultat:

Antal	Swish	Kort	Kontanter	Totalt
Unga	12	22	4	38
Medel	16	52	8	76
Äldre	4	26	8	38
Totalt	32	100	20	152

a) Testa på 5 % signifikansnivå om betalningssätt som kunderna väljer beror på ålder. Redovisa antaganden, hypoteser, testvariabel samt beslutsregel och kritisk gräns. (8p)

b) Redovisa sedan dina beräkningar, dra en slutsats och tolka slutsatsen i ord. (8p)

c) Förklara kortfattat vad som menas med begreppet *iid* och varför det ofta är önskvärt att ett stickprov är *iid*. (4p)

Uppgift 7

Man ville veta om studenters resultat på en delkurs B i ett visst ämne kan förklaras av deras resultat på delkurs A (som kommer före B) och deras gymnasiebetyg. Med ett stickprov av storlek $n = 30$ studenter skattade man två linjära regressionsmodeller:

Modell 1: $Y_i = \beta_0 + \beta_1 X_{1i} + \varepsilon_i$

Modell 2: $Y_i = \beta_0 + \beta_1 X_{1i} + \beta_2 X_{2i} + \varepsilon_i$

där Y = poäng på B-kursens sluttentamen, X_1 = poäng på A-kursens sluttentamen och X_2 är betyg från gymnasiet (på en skala 0-20). På följande sida finns Excel-utskrifter och sammanfattande statistik för de tre variablerna. Använd dessa för att besvara följande frågor:

a) Testa hypotesen att lutningskoefficienten β_2 är skild från noll givet att β_1 är skild från noll i Modell 2. Du behöver inte redovisa vilka antaganden som krävs. (8p)

b) Beräkna med hjälp av Modell 1 ett 95 % prediktionsintervall för en slumpmässigt vald students poäng på sluttentamen i delkurs B, givet att studenten fick X_1 =80 poäng för delkurs A. Förklara hur intervallet ska tolkas. (8p)

a) Förklara <u>kortfattat</u> begreppet multikollinearitet. Varför är det problematiskt? (4p)

Statistiska institutionen

Stockholms universitet

Rättningsblad

Datum: 29/11-2018

Sal: Värtasalen

Tenta: Statistik för ekonomer

Kurs: Grundläggande statistik för ekonomer

ANONYMKOD: 0089-LMK

☒ 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			
Lär.ant. 10	10	10	13	15	16	6			

10+3

POÄNG	BETYG	Lärarens sign.
80	B	ME

STOCKHOLMS UNIVERSITET
Statistiska institutionen

HT2018 A-B

SVARSBILAGA till Tentamen i Grundläggande statistik för ekonomer 2018-11-29

Skrivsal: Vårtasalen

Anonymkod: 0089 - LMK *(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 på baksidan.

		A	B	C	D	E	
Uppgift 1	a)		X				R
	b)		X				R
Uppgift 2	a)					X	R
	b)				X		R
Uppgift 3	a)			X			
	b)		X				R
	c)	X					R
Uppgift 4	a)	X					R
	b)					X	R
Uppgift 5	a)			X			R
	b)	X					R
	c)		X				R

SU, STATISTIK

Skrivsal: Vårtasalen **Anonymkod:** 0089 - LMK **Blad nr:** 1

Uppgift 6:

A): Det test som ska göras är ett så kallat Homogenitetstest (oberoende test), ett χ^2 - test!

- Antaganden: Man antar att de olika åldrarnas svar är helt oberoende av varandra, och att fördelningen emellan dessa är lika (iid). 0+1

Eftersom antalet tillfrågade, med andra ord, stickprovsstorleken är större än 30 $(n > 30)$ $(n = 152)$, kan man approximera denna fördelning mot en normalfördelning enligt CGS. -1

- Hypoteser: Ho: betalningssätten beror på åldern

mot ... Tvärtom

H1: betalningssätten beror ej på åldern, minst en variabel visar annat. -2

- Testvariabel: $$\chi^2 = \sum_{i=1}^{r} \sum_{j=1}^{c} \frac{(O_{ij} - E_{ij})^2}{E_{ij}} \quad \text{där } E_{ij} = \frac{R_i C_j}{n}$$ R

- Kritisk gräns: $(r-1)(c-1) = (3-1)(3-1) = 2 \cdot 2 = 4$ frihetsgrader

$\chi^{4}_{0,05} = \left[\text{tabell } 4\right] = 9,488$ då $\alpha = 0,05$

- Beslutsregel: Förkasta Ho om $\chi^2_{OBS} > \chi^2_{KRIT} = 9,488$ R

χ^2 observerat

χ^2 kritisk

5

Uppgift 6: (fortsättning)

B): O_{ij} = observerat antal

O_{ij}	Swish	Kort	Kontanter	Totalt
UNGA	12	22	4	38
MEDEL	16	52	8	76
ÄLDRE	4	26	8	38
Totalt	32	100	20	152

E_{ij}	Swish	Kort	Kontanter	Totalt
UNGA	8 ①	25	5	38
MEDEL	16 ②	50	10	76
ÄLDRE	8	25	5 ③	38
Totalt	32	100	20	152

$$E_{ij} = \frac{R_i C_j}{n}$$

Ex:

① $\frac{38 \cdot 32}{152} = 8$

② $\frac{76 \cdot 32}{152} = 16$

③ $\frac{38 \cdot 20}{152} = 5$

$\downarrow \chi^2 = (O_{ij} - E_{ij})^2 / E_{ij} \downarrow$

				Totalt
UNGA	2	0,36	0,20	2,56
MEDEL	0	0,08	0,40	0,48
ÄLDRE	2	0,04	1,80	3,84
Totalt	4	0,48	2,4	6,88 R

$= \chi^2 = 6{,}88$

Slutsats: $\chi^2_{OBS} = 6{,}88 < \chi^2_{KRIT} = 9{,}488$

H_0 förkastas INTE, vilket betyder att betalningssätten beror på åldern! Konsekvensfel

8

SU, STATISTIK

Skrivsal: Vårtasalen **Anonymkod:** 0089-LMK **Blad nr:** 2

Uppgift 6: (fortsättning)

(c): iid betyder kort att fördelningen ~~mellan~~ för olika ~~variabler eller komponenter~~ är densamma, i ett stickprov. observationer + oberoende

Detta är ofta önskvärt för då kan man lättare göra olika sorters tester och även dra slutsatser. ok

Om Exempelvis fördelningen inte är densamma i ett stickprov, det kan vara att en andel av svaren går mot en normalfördelning och andra andelar inte, kan slutsatser inte dras, iallafall inte lätt eller på ett bra vis. ok

3

SU, STATISTIK

Skrivsal: Värtasalen Anonymkod: 0089 - LMK Blad nr: 3

Uppgift 7:

A):

Uppgift 7: (fortsättning)

B): Modell 1: $Y_i = \beta_0 + \beta_1 X_{1i} + \varepsilon_i$ 　95% prediktionsintervall

$\alpha = 0,05$ 　 $n = 30$

Den skattade linjära modellen $\longrightarrow$ $\hat{y}_i = b_0 + b_1 x_i$

$\underbrace{b_0 = 26,907 \quad b_1 = 0,8492}_{\text{BILAGA!}}$ 　 $\underbrace{x_1 = 80}_{\text{Frågan...}}$

$\hat{Y} = 26,907 + 0,8492 \cdot 80 \longrightarrow \hat{y} = 94,843$ R

Prediktionsintervall för prediktionen $\hat{y}$ givet $X = x$:

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

$$* \; s_e^2 = [\text{residualvariansen}] = \frac{\sum_{i=1}^{n} e_i^2}{n-k-1}$$

$e = [\text{enligt bilagan}] = 28$? Detta är frihetsgradstalet! −2

$e^2 = 28^2 = 784$ 　 $n = 30$ 　 $k = 1$ (1 förklaringsvariabel)

$$s_e^2 = \frac{784}{30-1-1} = \frac{784}{28} = 28$$

$$t_{n-2;\,\alpha/2} = t_{28;\,0,025} \; [\text{tabell 3}] = 2,048$$

följdfel

$$\longrightarrow 94,843 \pm 2,048 \cdot \sqrt{28\left(1 + \frac{1}{30} + \frac{(80-80,733)^2}{29 \cdot 28}\right)}$$

−1 　 ska vara $\neq S_{x_1}^2$

$$\longrightarrow 94,843 \pm 2,048 \cdot \sqrt{28(1,033995019)}$$

$$\longrightarrow 94,843 \pm 2,048 \cdot 5,38069314$$

$$\longrightarrow 94,843 \pm 11,0196591$$ följdfel

$\rightsquigarrow$ (83,82334 ; 105,8626599) $\longleftarrow$ INTERVALLET

Slutsats: Med 95% prediktion (ej troligt) var intervallen (83,82334 ; 105,8626599).

Förklara vad intervallet säger oss! −2

4

SU, STATISTIK

Skrivsal: Vårtasalen **Anonymkod:** 0089-LMK **Blad nr:** 4

Uppgift 7: (fortsättning)

(: Multikollinearitet

förklaringsvariabler

Multikollinearitet sker när olika ~~linjära modeller~~ har för hög korrelation emellan varandra, med andra ord är deras linjära funktion är för lik; Linjerna är nästan densamma som varandra.

Problematiken där är simpelt att man inte kan dra bra slutsatser, de olika modellerna/funktionerna är för lika varandra. Därför är det viktigt att när man gör en så kallad regressionsanalys att man inte väljer variabler med för lika värde!

2