Stockholms universitet
Statistiska institutionen

Skriftlig tentamen avseende kursen **Introduktion till statistik för statsvetare** (ST131G) 2013-04-25

****

Skrivtid: 5 timmar.
Hjälpmedel: Miniräknare, samt vidhäftat formelblad.
Genomgång: 2013-05-07 klockan 12.00 i sal B705.

****

Tentamen består av fem uppgifter vilka kan ge maximalt fyra poäng vardera, totalt tjugo poäng. Då en uppgift i sin tur består av två eller flera deluppgifter värderas dessa lika. För full poäng på en uppgift/deluppgift krävs att tydliga, fullständiga och välmotiverade lösningar samt svar inlämnas. Lycka till !!! / Peter Claësson

****

1.

Statistik över nederbörden under 50 år, för ett geografiskt område, sammanställdes i nedanstående tabell.

Nederbörd (mm)	100-599	600-699	700-799
Antal år	25	15	10

a) Ange vad som här är variabel, vilken variabeltyp och vilken datanivå som föreligger, samt vad som är frekvens. Motivera.
b) Illustrera fördelningen ovan i ett lämpligt diagram, samt motivera valet av diagram.
c) Såväl medelvärde som median beräknades, och man erhöll 520 respektive 600. Hur ska dessa värden tolkas?
d) Ange tre olika exempel på spridningsmått, samt diskutera vilket av de tre spridningsmåtten som här kan anses vara att föredra.

2.

Utgifter vid en myndighet har granskats för ett antal år. Här ges även konsumentprisindex (KPI) respektive år.

År	1990	2000	2005	2010
Utgifter (Mkr)	32	40	72	96
KPI	208	260	280	303

a) Beräkna en indexserie i löpande priser (basår 1990) avseende de aktuella utgifterna.
b) Beräkna därefter en motsvarande indexserie i fasta priser (1990 års prisnivå).
c) Illustrera även de två ovan beräknade indexserierna i ett lämpligt diagram. Motivera.
d) Hur stor är den procentuella utgiftsförändringen mellan år 1990 och år 2010?

v g v

3.

För ett speciellt bilmärke, av en viss årsmodell, har bilarnas värdeminskning beroende på körsträcka noterats.

Körsträcka (tusentals mil)	3,8	4,6	4,9	5,4	5,5	7,1
Försäljningspris (tusentals kronor)	210	203	196	193	189	169

En regressionslinje anpassades, varpå man erhöll följande ekvation: $\hat{y} = 259{,}1 - 12{,}6x$ (avrundade värden).

a) Rita in observerade värden från ovanstående tabell i ett lämpligt diagram och rita även in regressionslinjen.
b) Efter nödvändiga beräkningar fås att determinationskoefficienten blir ca 0,99. Hur ska detta värde tolkas?
c) Ge en prediktion med hjälp av regressionslinjen för en bil som har gått 5000 mil och kommentera kritiskt.
d) Vid analysen erhålles så kallade residualer. Redogör för vad som avses därmed och hur dessa ska tolkas.

4.

Antag att ett slumpmässigt urvalsförfarande ska genomföras bland företag i ett visst län för att se om intresse föreligger bland företagsledarna rörande nyetablering inom ett framtida planerat köpcentrum.

Precisera vad som här kan anses vara mål- respektive rampopulation och redogör för samt ge även exempel på följande åtta begrepp/metoder, utifrån det givna antagandet.

a) Övertäckningsfel,
b) Undertäckningsfel,
c) Bortfallsfel,
d) Urvalsfel,
e) Obundet slumpmässigt urval,
f) Systematiskt urval,
g) Stratifierat urval,
h) Urvalsfraktion.

5.

Redogör för nedanstående åtta begrepp/metoder samt ge realistiska exempel på respektive begrepp/metod.

a) Ordinaldata,
b) Gruppenkät,
c) Probes,
d) Dikotom variabel,
e) Exponentiell trend,
f) Hansen-Hurwitz metod,
g) Prestigebias,
h) Väntevärdesriktig skattning.

1. Beskrivande statistik

1.1 Medelvärde, varians, standardavvikelse

Ett statistiskt material består av n observationer

$$x_1, x_2, \ldots, x_n$$

Medelvärdet är

$$\bar{x} = \frac{x_1 + x_2 + \ldots + x_n}{n} = \frac{\sum x}{n} \tag{1.1.1}$$

Variansen är

$$s^2 = \frac{\sum (x - \bar{x})^2}{n-1} = \frac{\sum x^2 - (\sum x)^2 / n}{n-1} \tag{1.1.2}$$

Standardavvikelsen är

$$s = \sqrt{s^2} \tag{1.1.3}$$

När materialet redovisas i en **frekvenstabell**, där värdet x_i förekommer med frekvensen f_i, är medelvärdet och variansen

$$\bar{x} = \frac{\sum f_i x_i}{n} \tag{1.1.4}$$

$$s^2 = \frac{\sum f_i (x_i - \bar{x})^2}{n-1} = \frac{\sum f_i x_i^2 - (\sum f_i x_i)^2 / n}{n-1} \tag{1.1.5}$$

Räkneregler

Om $y = a + bx$, där a och b är konstanter, är

$$\bar{y} = a + b\bar{x} \tag{1.1.6}$$

$$s_y^2 = b^2 s_x^2 \tag{1.1.7}$$

1.2 Regression, korrelation

Regressionslinjen är $y=a+bx$.

Regressionskoefficienten är

$$b = \frac{\sum (x - \bar{x})(y - \bar{y})}{\sum (x - \bar{x})^2} = \frac{\sum xy - \sum x \sum y / n}{\sum x^2 - (\sum x)^2 / n} \tag{1.2.1}$$

$$a = \bar{y} - b\bar{x} \tag{1.2.2}$$

Korrelationskoefficienten är

$$r = \frac{\sum (x - \bar{x})(y - \bar{y})}{\sqrt{\sum (x - \bar{x})^2 \sum (y - \bar{y})^2}} = \tag{1.2.3}$$

$$= \frac{\sum xy - \sum x \sum y / n}{\sqrt{\left[\sum x^2 - (\sum x)^2 / n\right]\left[\sum y^2 - (\sum y)^2 / n\right]}} = b \frac{s_x}{s_y}$$

Residualvariansen är

$$s_e^2 = \frac{n-1}{n-2} s_y^2 (1 - r^2) = \frac{1}{n-2}\left(\sum y^2 - a \sum y - b \sum xy\right) \tag{1.2.4}$$

1.3 Prisindex

Laspeyres index är

$$\frac{\sum p_t q_0}{\sum p_0 q_0} \cdot 100 \tag{1.3.1}$$

Paasches index är

$$\frac{\sum p_t q_t}{\sum p_0 q_t} \cdot 100 \tag{1.3.2}$$

Statistiska institutionen

Stockholms universitet

Rättningsblad

Datum: 25/4-2013

Sal: Laduvikssalen

Tenta: Statistik för statsvetare

Kurs: Introduktion till statistik för statsvetare

ANONYMKOD: SFS-0025

☒ 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					4
Lär.ant.									
M	3.5	M	2.5	2.5					

POÄNG	BETYG	Lärarens sign.
16.5	B	

SU, STATISTIK

Skrivsal: Laduvikssalen Anonymkod: SFS-0025 Blad nr: 1

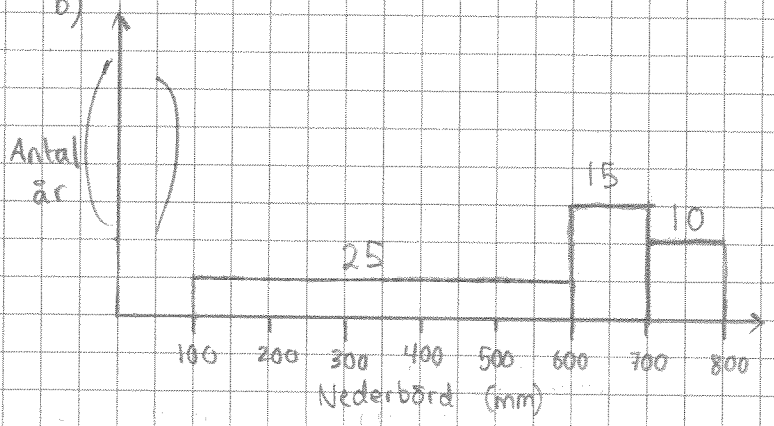
1. a) Variabel = Nederbörd (mm)

Variabeltyp = kvantitativ variabel (kontinuerlig - går att vara väldigt exakt vid mätningen med decimaler etc. klassindelad)

Datanivå = kvotdata, eftersom det existerar en absolut nollpunkt och meningsfulla kvoter går att bilda (exempel: 200 mm nederbörd är dubbelt så mycket som 100 mm etc)

Frekvens = Antal år 1

b) Histogram



1

Jag valde att presentera datan i formen av ett histogram eftersom kvantitativ data främst presenteras med hjälp av stolpdiagram och histogram, men histogram kan vara att föredra ifall materialet är klassindelat, vilket det är i detta fall.

c) Medelvärdet 520 innebär det värde man får ifall man dividerar summan av alla värden vi har för vår variabel (nederbörd) med antalet observationer, i detta fallet 50. Detta värde vittnar om var tyngdpunkten ligger i vår fördelning, vilket tycks stämma bra om man tittar på fördelningen i diagrammet. Det genomsnittliga värdet för nederbörden under dessa 50 år var alltså 520 mm per år.

Medianens värde på 600 vittnar istället om villet som är det mittersta värdet när värdena är upprädade efter storleksordning, villet i detta fall blir mellan värde 25 och 26 ($\frac{50+1}{2} = 25{,}5$).
I det här fallet tycks medianen ge den mest rättvisande bilden av nederbörden under dessa 50 år, eftersom "extremvärdena" till höger (tex 700 mm och över) förskjuter medelvärdet. 2

d) Variationsbredd = högsta värdet - lägsta värdet
Kvartilavstånd = tredje kvartilen - första kvartilen (kvartilerna delar in materialet i fyra lika stora delar)
Standardavvikelsen = Visar på "den genomsnittliga spridningen" kring medelvärdet
Variationsbredden är lätt att räkna ut och förstå men ger oss i detta fall inte mycket information om spridningen. Eftersom jag förordade ~~medelvärdet~~ medianen (stavfel V2) i tidigare uppgift är det naturligt att föredra kvartilavståndet framför standardavvikelsen som spridningsmått i detta fall. Eftersom standardavvikelsen visar på spridningen kring medelvärdet (oftast att förorda vid högre datanivåer som intervall- och kvotdata) och medelvärdet i vårt fall riskerar att ge en missvisande bild pga förskjutningen i materialet, torde kvartilavståndet som hör samman med med medianen vara ett bättre spridningsmått i detta fall. Kvartilerna ger ökad information om materialets fördelning. 1

-/ Bra!

SU, STATISTIK

Skrivsal: Laduvikssalen Anonymkod: SFS-0025 Blad nr: 2

Löpande priser

2. | År | 1990 | 2000 | 2005 | 2010 |
|---|---|---|---|---|
| Utgifter | 32 | 40 | 72 | 96 |
| KPI | 208 | 260 | 280 | 303 |

Hur kronans värde ändrat sig

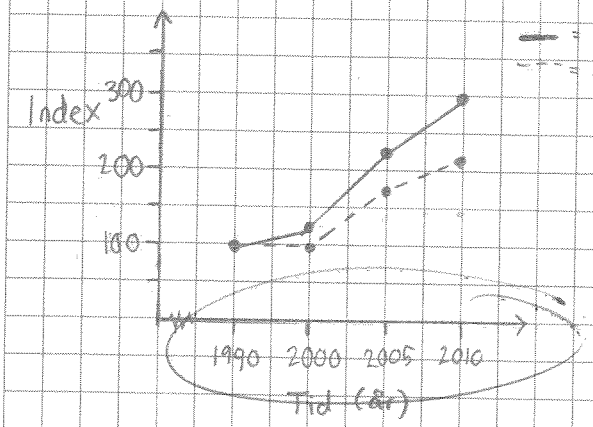
a)

År	Index (löpande)	KPI	Deflator	Priser (fasta)	Index (fast)
1990	100 *	208	1 *	32 *	100 *
2000	125 *	260	0,8	32 *	100
2005	225	280	0,74	53,28	167
2010	300	303	0,69	66,24	207
	* 32/32		* 208/208	* 32 · (208/208)	* 32/32
	* 40/32			* 32 · (208/260)	

1

1

c)



0,5

Ingen deflatorstrans?

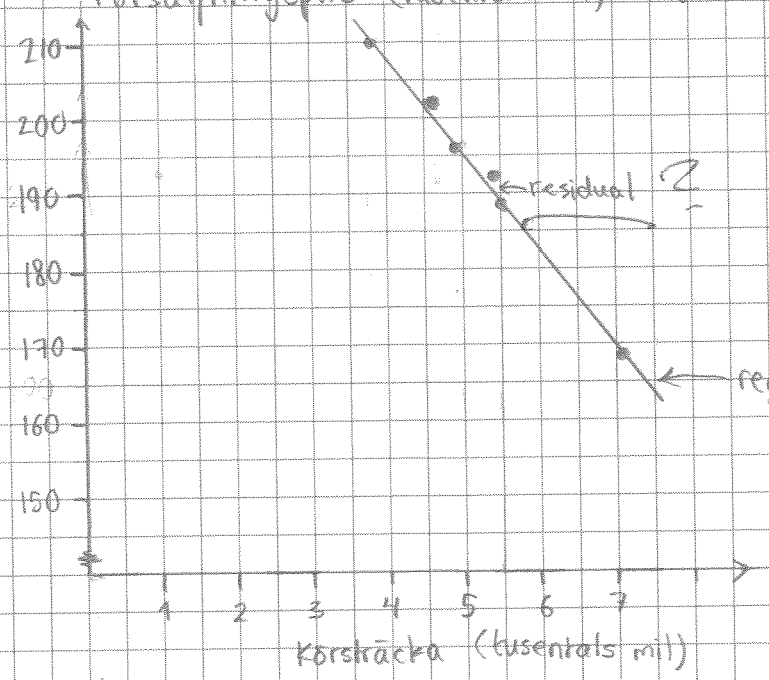
Jag valde att illustrera indexutvecklingarna med hjälp av ett linjediagram eftersom dessa är lämpade för att beskriva just utvecklingar över tid, och för att de exempelvis underlättar jämförelser som i detta fall.

d) I löpande priser har den procentuella utgiftsförändringen innebärit en ökning med 200%. I fasta priser har den istället innebärit en ökning med 107%. Dessa procentuella förändringar är framräknade med år 1990 som utgångspunkt, i de fasta priserna har vi underlättat jämförelsen genom att ange priserna i kronans nivå för år 1990.

1

3 a)

Körsträcka (tusentals mil)	3,8	4,6	4,9	5,4	5,5	7,1
Försäljningspris (tusentals kr)	210	203	196	193	189	169



Jag valde att illustrera informationen i ett spridningsdiagram eftersom dessa illustrerar just spridning på ett bra sätt och kan visa på samband mellan olika variabler.

b) En determinationskoefficient på 0,99 innebär att 99 % av variationerna (observationerna) i försäljningspris kan förklaras av variabeln körsträcka. Denna koefficient får man ut av att upphöja korrelationskoefficienten (r) i två, r blir då ca 0,99... vilket bekräftar det starka samband vi ser i diagrammet.

c) En bil som har gått 5000 mil bör kosta kring $\hat{y} = 259{,}1 - 12{,}6x = 259{,}1 - 12{,}6 \cdot 5 = 196{,}1$ i tusentals kronor, enligt vår modell. (går även att uppskatta i diagrammet). Med vår modell och övriga värden som underlag tycks denna prediktion vara rimlig

d) Residualer är den skillnad som uppstår mellan vår modells förväntade värden, och våra uppmätta värden. I vårt fall har vi ingen stor residualspridning och alltså talar detta för ett starkt samband mellan variablerna.

Eftersom siffran 5 befinner sig mitt i vår modell blir prediktionen relativt säker. Värden som går långt utanför med osäkra prediktioner

Kritisk kommentar till c

3.5 (4)

4. Vår målpopulation i detta fall bör rimligen vara alla företag i länet som har intresse av att etablera sig i ett köpcenter, exempelvis sådana vars verksamhet kretsar kring detaljvaruhandel. Rampopulationen blir då de av dessa företag som finns med i någon lista eller förteckning över verksamma företag. När vi fått tag i denna förteckning kan ett eventuellt övertäckningsfel vara att vi får med (företag i rampopulationen) som inte sysslar med denna typ av verksamhet, eller som har gått i konkurs (listan kanske inte är uppdaterad). Ett undertäckningsfel kan istället vara ett företag med (eventuellt) denna typ av intresse, men som ännu inte har blivit registrerat. Från denna lista hade vi kunnat utnyttja exempelvis ett systematiskt urval för att välja ut undersökningsobjekt på ett (hur dåligt?) slumpartat sätt (har hög precision så länge det inte föreligger någon typ av periodicitet som sammanfaller med vårt systematiska urval). Ett bortfallsfel (vad för ett "fel"?) skulle kunna vara att företagsledare vars företag blivit utvalda inte svarar på de enkäter vi har skickat ut. Ett urvalsfel kan uppstå om det till exempel existerar periodicitet i förteckningen som vi drar vårt systematiska urval från, så att detta skapar en förvridning i vårt urval. (Systematiskt urval: siffra mellan 1 och 10 slumpas fram (vad?), därefter väljs var n:te enhet i förteckningen ut). Ifall vi

föredrar att göra vårt urval med ett så kallat "obundet slumpmässigt urval" innebär det att vi lottar ut (med hjälp av dator vid fall av många undersökningsobjekt) ett visst antal enheter ur vår förteckning (utan återläggning). Alla enheter måste ha samma chans att bli utvalda. I ett stratifierat urval kan vi motverka eventuella snedvridningar som slumpen kan medföra genom att välja ut vissa faktorer som vi tror påverkar vår variabel (inställning till nyetablering i ett planerat köpcentrum), exempelvis kön. Om vi vill vara säkra på att våra resultat ska spegla vår rampopulation så mycket som möjligt kan vi då välja ut strata z ur vår rampopulation, exempelvis att vi kollar på ett visst antal kvinnor/män som är proportionellt med fördelningen i rampopulationen. Se förra sida? I det aktuella fallet kanske ett företags storlek eller fysiska läge i länet har relevans och då kan vi göra ett urval som försäkrar oss om att vi inte går miste om dessa viktiga aspekter. En urvalsram kan är en del av vårt urval, i vårt fall exempelvis några få utvalda företagsledare som besvarat vår enkät.

När vi har valt ut våra stratum gör vi vårt slumpmässiga urval utifrån dessa!

(25)

SU, STATISTIK

Skrivsal: Ladugårdssalen Anonymkod: SFS-0025 Blad nr: 4

5a) Ordinaldata = den näst lägsta datanivån. Innebär att datan kan anta olika värden som går att rangordna. Kvalitativ. (Har ingen absolut nollpunkt och är ej meningsfull att räkna kvoter från!) Ett vanligt exempel på ordinaldata är när man vid mätning av människors attityd använder en skala, tex 1-5.

b) När man delar ut en enkät till många individer på en gång, exempelvis en klass. Billig metod eftersom man når ut till många individer utan att behöva resa över en stor geografisk yta. Litet bortfall.

c) Probes —

d) Dikotom variabel = variabel som kan anta två värden, exempelvis 1 eller 0. Används vid kodning av kvalitativa variabler och kan sedan utnyttjas för att räkna ut ett medelvärde. Exempel: Man = 1 kvinna = 0, $\frac{5}{10}$ = 50% män.

e) Exponentiell trend är en trend som ej ökar på ett konstant sätt i absoluta tal (vilket i sådana fall medför en linjär trend), utan skillnaden på y-axeln för varje steg i x-axeln (lutningen) ökar hela tiden, i absoluta tal.

Ex. 2

f) Hansen-Hurwitz metod

g) Prestigebias: När en undersökning innehåller frågeställningar som kan framkalla en viss snedvridning i resultatet, eftersom undersökningsobjekten visar en tendens att besvara frågeställningarna på ett prestigeartat sätt. Exempelvis om någon i en undersökning av hur mycket pengar personer spenderar på tekniskprylar i månaden tycker att det finns en prestige i att lägga ner mycket pengar på detta, och därför uppger ... ??

h) Väntevärdesriktig skattning: En skattning är väntevärdesriktig om den antar samma värde som parametern för hela populationen. En skattning från ett enda urval är ofta otillräcklig, men om man exempelvis räknat ut medelvärdet från ett flertal olika slumpmässigt utvalda urval bör detta medelvärde stämma överens med medelvärdet för hela populationen, skattningen är då väntevärdesriktig.

25

Mkt bra!

Mats Peter

Statistiska institutionen

Stockholms universitet

Rättningsblad

Datum: 25/4-2013

Sal: Laduvikssalen

Tenta: Statistik för statsvetare

Kurs: Introduktion till statistik för statsvetare

ANONYMKOD: SFS-0013

☑ 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	X	X	5
Lär.ant.									
3.5	4	3.5	2.75	2.5					

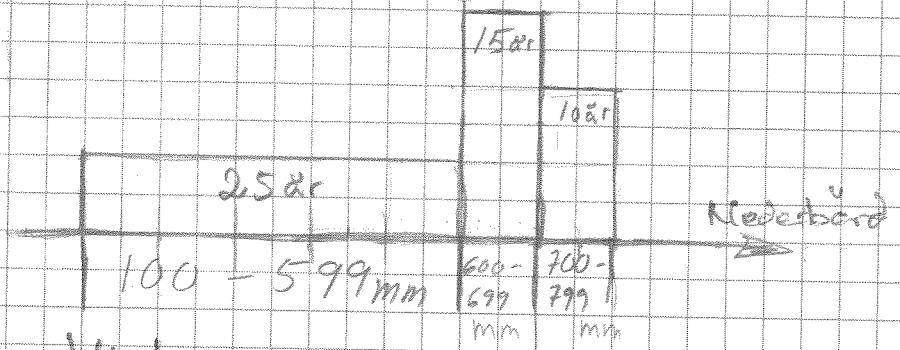
POÄNG	BETYG	Lärarens sign.
16.25	B	

SU, STATISTIK

Skrivsal: Laduriksalen Anonymkod: SFS-0013 Blad nr: 1

1. a) Variabeln är nederbörden
Variabeln är kvantitativ och kontinuerlig
Variabeln är på kvotdatanivå.
Detta för att nederbörden kan anta alla värden i intervallen (beroende på hur noga man mäter)
Dessutom finns det en (meningsfull) absolut nollpunkt och 2 mm är dubbelt så mycket som 1 mm, alltså det är meningsfullt att bilda kvoter.
Frekvensen är antalet år, eftersom dessa är angivna som hur många år med den nederbörden. 1

b) Jag har valt att rita ett histogram.



Histogram
(Detta diagram) passar eftersom variabeln är kvantitativ och kontinuerlig. (1)

c) • medianen ska är det mittersta värdet, hälften av observationerna är mindre än den och hälften är mer än den. (det är ett lägesmått)

• medelvärdet är också ett lägesmått som innebär att man räknat ut hur mycket nederbörd som skulle kommit, om det hade kommit lika mycket varje år.

(1)

d) typvärde, kvartilavstånd, standardavvikelse

Eftersom det är kvotdata vi har här är standardavvikelsen att föredra 0.5

(2) a)

År	utgifter	Löpande index	Deflation	Fast index
1990	32	100	1	100
2000	40	125	$\frac{208}{260} = 0,8$	100
2005	72	225	0,74286	167,14
2010	96	300	0,68647	205,94

$$\text{Löpande index} = \frac{\text{Jämförelseåret}}{\text{basåret}} \cdot 100$$

ex $\frac{40}{32} = 1,17647 \cdot 100 = 117,647$ 1

Deflatorn (0,8 94,08 1,25) 1

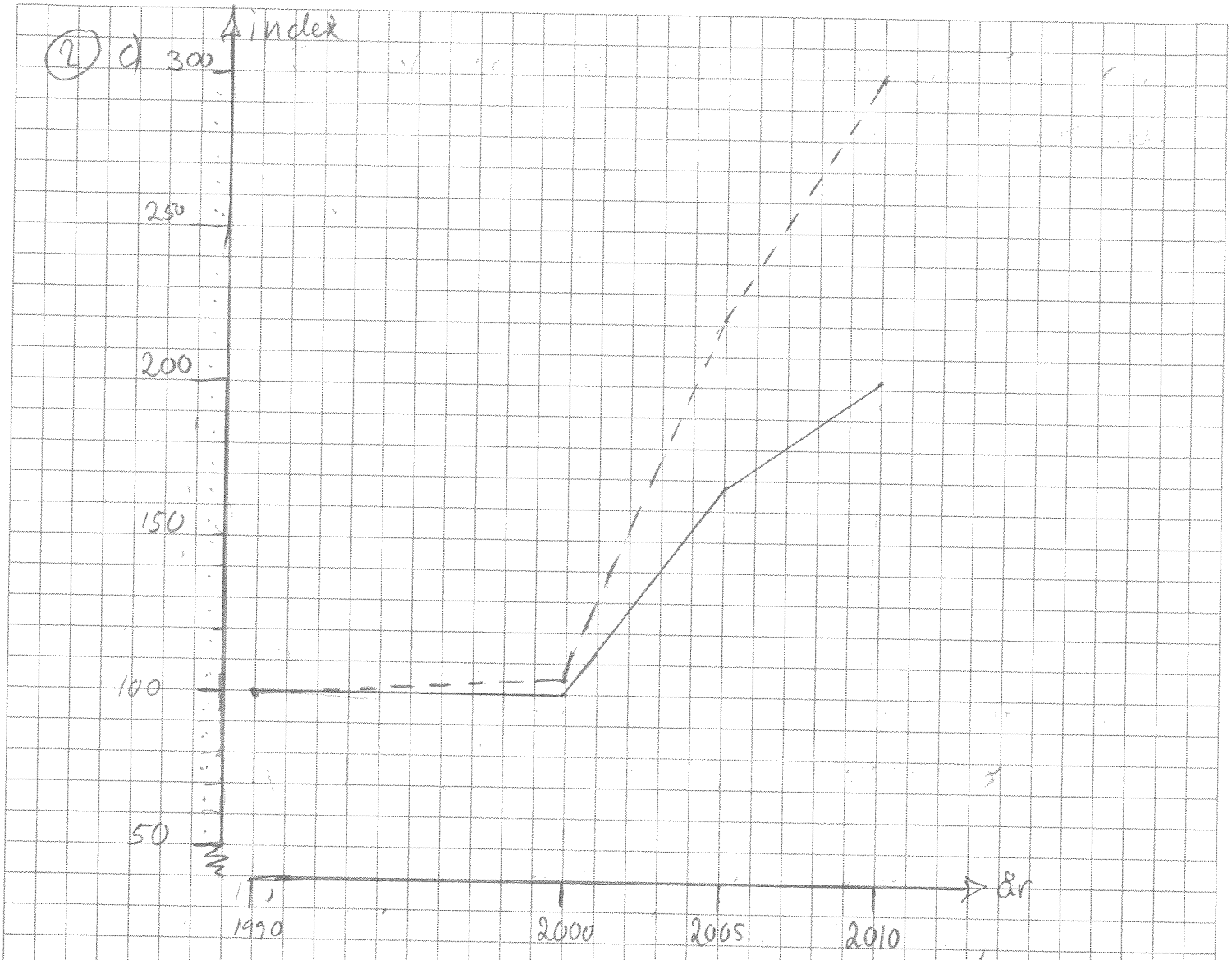
Fast index = index · deflatorn

$$\text{Deflation} = \frac{\text{KPI för basåret}}{\text{KPI för jämförelseåret}}$$

SU, STATISTIK

Skrivsal: Ladu rikssalen Anonymkod: SFS-0013 Blad nr: 2

(2) a)

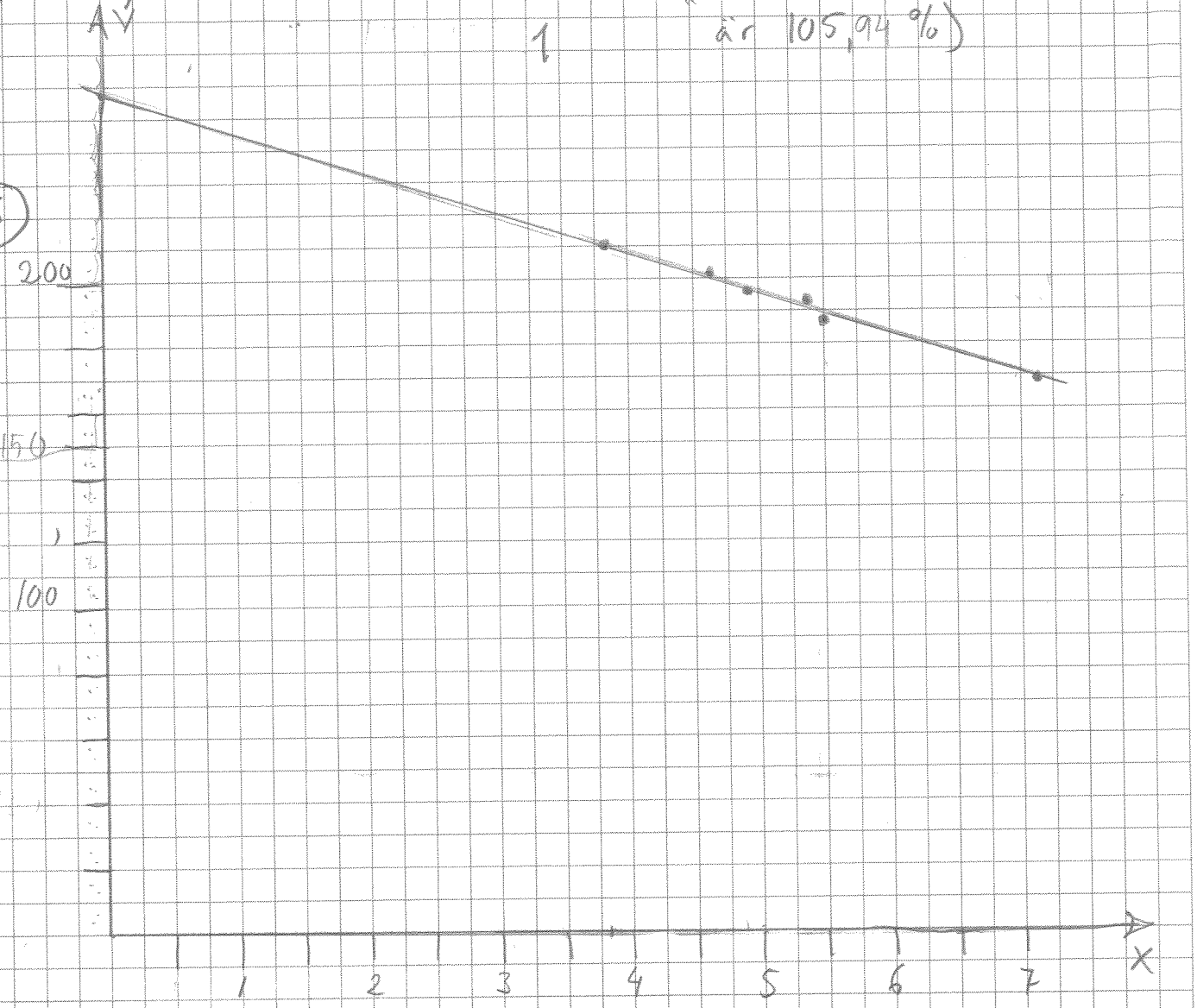


——— = fast index för myndighetens utgifter

– – – – = löpande " "

- Jag valde att rita ett linjediagram eftersom det passar bra att visa förändring över tid med.

d) utgifterna har ökat med 200% (men den reella ökningen är 105,94 %) 1



(3)

x = körsträckan

Y = Försäljningspris

1

b) Determinationskoefficienten betyder i det här fallet hur mycket av (förändringen) variationerna av försäljningspriset som beror på körsträckan. (i generella termer hur mycket av förändringen av y som beror på den oberoende variabeln (x)) - - - och här står "0.99" därmed fetstil? (0.75)

c) ca 195 tusen kr. Det är en relativt säker bedömning eftersom det är en interpolation

Be-räkning?

c) Forts. (vilket betyder att den ligger innanför det intervall vi studerat).
Dock är inte mina streck knivskarpa och även om determinationskoefficienten är relativt hög kan det finnas andra faktorer som påverkar priset, exempel kan den vara billigare om förra ägarens hund tuggat sönder sätena. (1)

d) En residual är avståndet från en punkt till regressionslinjen. Den visar hur mycket observationen avviker från modellen. Pos/neg sidorna? (0.75)

(4) målpopulationen = företagsledarna i det aktuella länet
rampopulation = det register över målpopulationen.
Jag tänker mig att det aktuella länet passande nog precis upprättat ett komplett register över företags ledarna i länet.

a) Överäckningsfel i det här fallet skulle kunna vara att även företagsledare från andra län skulle vara med i registret.
Alltså att rampopulationen innehåller fler individer än målpopulationen. 0.5

b) Underträckning är motsatsen till överträckning. I vårt fall skulle det inte alla företagsledare (i det aktuella länet) finnas med i registret.

Alltså: Rampopulationen täcker in färre än målpopulationen 0.5

c) Bortfallsfel innebär att bortfallet är speciellt på något sätt. I vårt fall skulle det innebära att en grupp av företagsledare med något gemensamt som skulle påverkat resultatet av undersökningen. Vägrar/inte kan svara. 0.5
ex kanske en grupp som är inte intresserad av köpcentrumet men som hatar postentäter

d) Urvalsfel är nästan omöjligt, som blir när vi genom vårt urvalsmetod får fram ett stickprov. Detta stickprov av företagsledare kanske inte är proportionellt mot målpopulationens av företagsledare. Därför kommer kanske våra resultat heller inte vara i enlighet med målpopulationens. Om man gjort ett korrekt urvalsförfarande kan man beräkna urvalsfelet. (0.25)

(slumpmässigt)

e) Obundet slumpmässigt urval = alla individer (företagsledare) i rampopulationen (registret) har en känd sannolikhet att komma med i urvalet. Sannolikheten är större än noll, vilket på urvalet genomförs på så vis att ... (0.25)

f) Systematiskt urval: alla i rampopulationen (i vårt fall registret över företagsledare) får en siffra. Sedan slumpar man en siffra och väljer ut den individen (företagsledaren). Där efter bestämmer man ett intervall och hoppar och tar ex var tionde företagsledare från den första (ex 2, 12, 22, 32 osv) (0.25)

g) Stratifierat urval innebär att man delar in registret efter någon bakgrunds variabel exempelvis företagsledare i storlek respektive små företag och slumpar ut företagsledare ur respektive grupp. (0.5)

OBS! Kan vara förödande, man måste veta hur registret är organiserat!

h) Differansen mellan urvalet och målpopulationen (0)

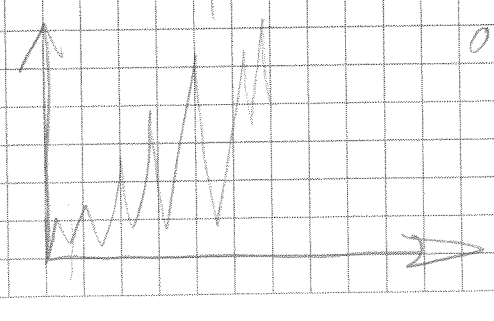
(5) a) Ordinaldata är data som kan kategoriseras och rangordnas, men som inte har ekvidistans eller en meningsfull nollpunkt. Man kan inte bilda kvoter mellan enheterna.
Ett exempel är betyg, A, B, C, D, E, F / MVG, VG, G
0.5

b) En enkät som delas ut till en bestämd grupp som besvaras vid ett gemensamt tillfälle, men enskilt
Ex en enkät om drogvanor som delas ut i en skolklass att besvaras innan lektionen börjar. (enskilt)
0.5

c) Probes är i förväg förberedda förtydliganden som intervjuaren kan ge till respondenten om hen inte förstår frågan.
ex att intervjuaren förklarar vad som menas med "inkomst" i fråga ...
0.5

d) En dikotom variabel innebär att den bara kan anta två värden
Exempelvis kön (Kvinna/Man) brukar räknas som en dikotom variabel.
0.5

e) En exponentiell trend kan se ut såhär



och T·S·C·E
är multiplikativ ✓

e) Kort. det är en ökning eller minskning som aldrig kan bli noll och den sker över en längre tidsperiod.
ex på trend är ett företag som växt under en längre period och förväntas fortsätta.

f) Hansen-Hurwitz metod är en metod för att beräkna regression.

g) Prestigebias är ett systematiskt fel som beror på att respondenten vill framställa sig själv i bättre dagar, ex att man anger att man minnsann tränar 5 gånger i veckan fast man bara tränar 2

h) en väntevärdesriktig skattning är en skattning som stämmer överens med mål populationens. Alltså en parameter i urvalet är väntevärdesriktig om den stämmer överens/ligger nära det motsvarande i populationen. På vilket sätt?

Mvh Peter