Stockholms universitet
Statistiska institutionen

Skriftlig tentamen avseende kursen **Introduktion till statistik för statsvetare** (ST131G) 2015-04-28

****

Skrivtid: 5 timmar.
Hjälpmedel: Miniräknare, samt vidhäftat formelblad.
Genomgång: 2015-05-12 klockan 16.00 i sal D289.

****

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. Betygskriterier enligt Kursbeskrivning för innevarande termin. En sjugradig betygsskala (enligt A-F) tillämpas. Lycka till !!! / Peter Claësson

****

1.

För samtliga aktiva medlemmar i ett politiskt parti finns statistisk avseende hur många år man varit medlem.

a) Ange noggrant vad som här är variabel, vilken variabeltyp och vilken datanivå som föreligger, samt vad som är frekvens.
b) Två typvärden erhölls i den aktuella fördelningen! Redogör för hur det kunde komma sig att inte endast ett typvärde erhölls.
c) Även medelvärde och median bestämdes, där det visade sig att medelvärdet översteg medianens värde! Redogör för hur det kunde komma sig att så var fallet.
d) Ange tre olika exempel på spridningsmått, samt diskutera vilket av de tre spridningsmåtten som här kan anses vara att föredra.

2.

En myndighets utgifter har granskats för ett antal år. Även konsumentprisindex (KPI) respektive år föreligger.

År	1980	1990	2000	2005	2010
Utgifter (Mkr)	40	80	88	92	96
KPI	100	208	260	280	303

a) Beräkna en indexserie i löpande priser (basår 1980) för de aktuella utgifterna.
b) Beräkna därefter en motsvarande indexserie i fasta priser (1980 års prisnivå).
c) Illustrera de två därmed beräknade indexserierna i ett lämpligt diagram.
d) Hur stor är den procentuella utgiftsförändringen mellan år 1980 och år 2010?

v g v

3.

För en mindre stat har man under flera år studerat hur arbetslösheten (uttryckt i %) beror på inflationstakten (likaså uttryckt i %), varvid följande regressionslinje anpassades: $\hat{y} = 8{,}25 - 0{,}10x$ (avrundade värden).

a) Hur ska de erhållna värdena i regressionslinjens ekvation ovan tolkas?
b) Efter nödvändiga beräkningar fås att determinationskoefficienten blir ca 0,75. Hur ska detta värde tolkas?
c) Ge en prediktion med hjälp av regressionslinjen ett år då inflationstakten är 5 %, och kommentera kritiskt.
d) Vid analysen erhölls så kallade residualer. Redogör för vad som avses därmed och hur dessa ska tolkas.

4.

En omfattande enkätundersökning ska genomföras bland ett slumpmässigt urval av ungdomar i åldrarna 16-24 år boende i en viss svensk kommun, där huvudsyftet är att få information om ungdomars vilja att söka sig ifrån kommunen eller inte.

Ange noggrant vad som kan sägas utgöra mål- respektive rampopulation och redogör för, och ge exempel på, följande urvalsmetoder (där även deras för- och nackdelar ska kommenteras), givet undersökningssituationen som antas föreligga.

a) Obundet slumpmässigt urval,
b) Systematiskt slumpmässigt urval,
c) Stratifierat urval,
d) Klusterurval.

5.

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

a) Histogram,
b) Lådagram,
c) Cyklisk variation,
d) Dikotom variabel,
e) Exponentiell trend,
f) Hansen-Hurwitz metod,
g) Prestigebias,
h) Konfidensintervall.

1. Beskrivande statistik

1.1 Medelvärde, varians, standardavvikelse

Ett statistiskt material består av n observationer

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

Medelvärdet är

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

Variansen är

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

Standardavvikelsen är

$$s = \sqrt{s^2} \qquad (1.1.3)$$

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

$$\bar{x} = \frac{\sum f_i x_i}{n} \qquad (1.1.4)$$

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

Räkneregler

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

$$\bar{y} = a + b\bar{x} \qquad (1.1.6)$$

$$s_y^2 = b^2 s_x^2 \qquad (1.1.7)$$

1.2 Regression, korrelation

Regressionslinjen är $y=a+bx$.

Regressionskoefficienten är

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

$$a = \bar{y} - b\bar{x} \qquad (1.2.2)$$

Korrelationskoefficienten är

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

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

Residualvariansen är

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

1.3 Prisindex

Laspeyres index är

$$\frac{\sum p_t q_0}{\sum p_0 q_0} \cdot 100 \qquad (1.3.1)$$

Paasches index är

$$\frac{\sum p_t q_t}{\sum p_0 q_t} \cdot 100 \qquad (1.3.2)$$

Statistiska institutionen

Stockholms universitet

Rättningsblad

Datum: 28/4-2015

Sal: Brunnsvikssalen

Tenta: Statistik för statsvetare

Kurs: Statistik för statsvetare

ANONYMKOD: SFS ~~0025~~ 0026

☒ 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. 3.5	4	3.5	3	~~1.5~~ 2					

POÄNG	BETYG	Lärarens sign.
16	B	

SU, STATISTIK

Skrivsal: BR Anonymkod: SFS0026 Blad nr: 1

1 a) Variabeln är antal år, det är en kvantitativ variabel som ligger på en kvotskalanivå och den är även ✓ diskret, då den är heltalig och endast kan anta ett ändligt/fast antal värden, klassindelas, rangordnas, bilda meningsfulla kvoter samt har en absolut nollpunkt. Frekvensen är de antal gånger ett visst värde har uppkommit d.v.s. hur många som varit medlemmar i x antal år (här). * samt har ekvidistans

b) Typvärdet är det mest frekvent förekommande värdet, att det i detta fall blev två st typvärden visar på att det har varit ett lika stort antal personer som varit medlemmar i två olika tidsperioder tex att 20pers är det största värdet som uppkom och detta uppkom både för 5 år och 12 år.

c) Att medelvärdet överstiger medianens värde här och göra med att det är en sned fördelning av medlemsår där några medlemmar har varit med väldigt länge (mycket längre än övriga förmodligen) som gör att genomsnittet ökar, medan medianen - - ?

d) Exempel på tre spridningsmått är standardavvikelse, kvartilavstånd, variationsvidd/bredd.
Standardavvikelsen visar hur variationerna i genomsnitt avviker från det aritmetiska medelvärdet, kvartilavståndet visar inom vilket intervall de 50% mest centrala observationerna befinner sig och variationsbredden är skillnaden mellan det största och minsta värdet. Vanligtvis är standardavvikelsen att föredra, men eftersom medelvärdet här överstiger medianen och tyder på en sned fördelning är kvartilavståndet (i likhet med medianen) att föredra här. Då det visar vilka antal medlemsår som är vanligast vilket kan vara av intresse vid den givna frågan. Även variationsbredden skulle kunna beräknas om man är intresserad av hur mycket det skiljer sig mellan den nyaste och "äldste" medlemmen, dock kan detta tyckas vara mindre viktigt

35

2

Index löp = $\frac{Pris_x}{Pris_{bas}}$ API APKIDI

Index — årtal

KPI bas / KPI NV = deflator

deflator × index

a) År Utg. (mkr) Index löp. (%)

basår	1980	40	100
	1990	80	200*
	2000	88	220
	2005	92	230
	2010	96	240

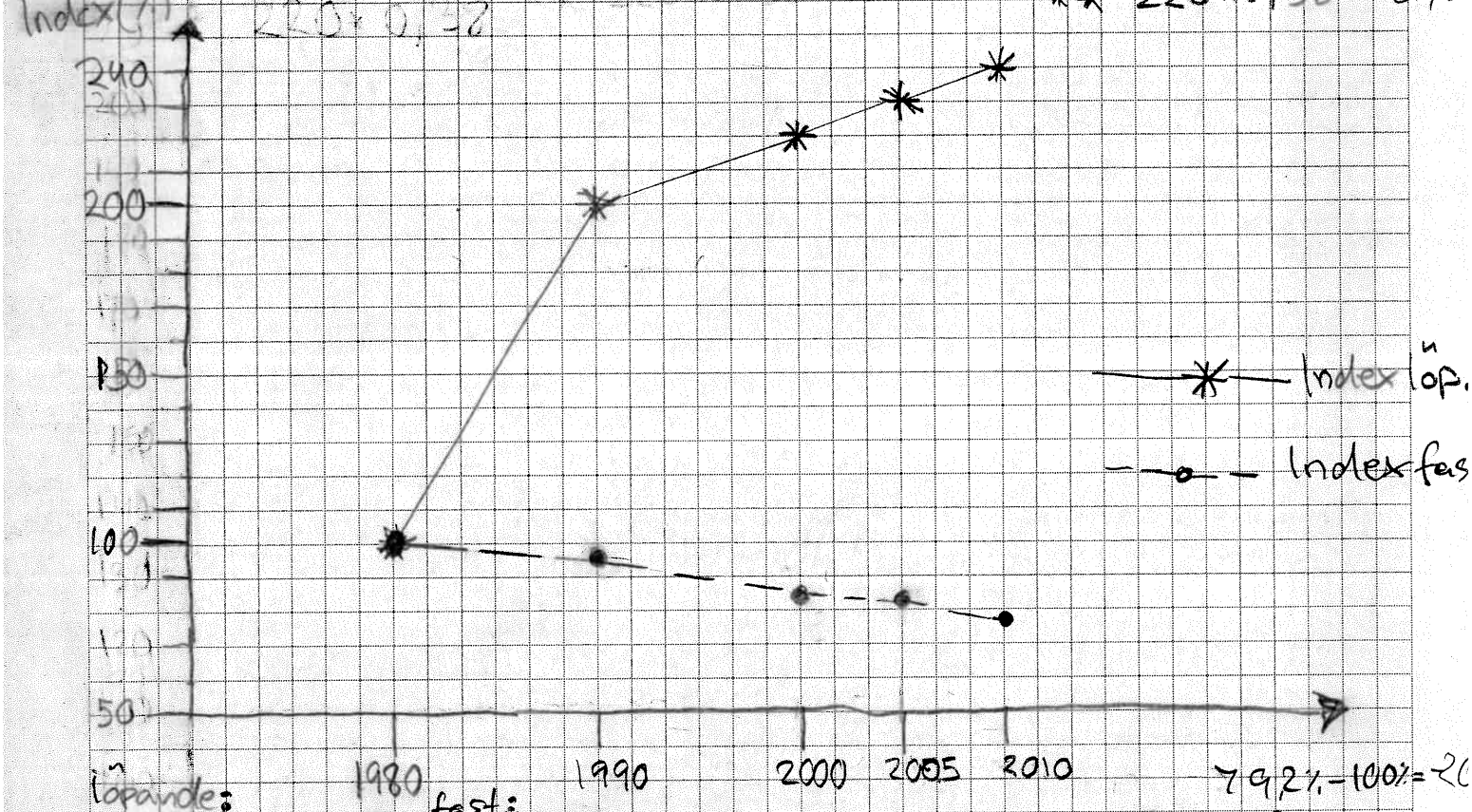
* $\frac{80}{40} = 2 = 200\%$

b)

	År	Utg. (mkr)	KPI	Index löp. (%)	Deflator	Index fast (%)
basår	1980	40	100	100	100/100 = 1	100
	1990	80	208	200	100/208 ≈ 0,48	96
	2000	88	260	220	100/260 ≈ 0,38	83,6**
	2005	92	280	230	100/280 ≈ 0,36	82,8
	2010	96	303	240	100/303 ≈ 0,33	79,2

** 220 × 0,38 = 83,6

c)



löpande: fast:

d) $\frac{240}{100} \approx 2,4 = 240\% \approx 240\% - 100\% = 140\%$ $\frac{79,2}{100} = 0,792 = 79,2\%$ 79,2% − 100% = −20

Den procentuella utgiftsförändringen mellan år 1980 och 2010 var i löpande priser 140% och −20,8% i fasta

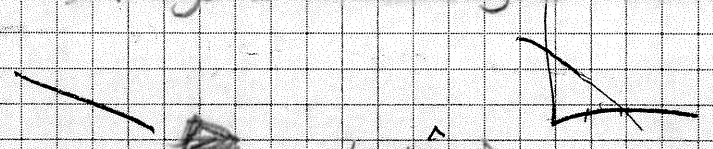
Bra! /4

SU, STATISTIK

Skrivsal: BR Anonymkod: SFS0026 Blad nr: 2

3. $\hat{Y} = 8{,}25 - 0{,}10x$

a) Regressionslinjen visar att det linjära sambandet är negativt det vill säga att när den ena variabeln ökar minskar den andra. När X ökar minskar Y. Regressionslinjen visar det genomsnittliga värdets



← Tolka koefficienterna !!!

(Residualvariansen) →

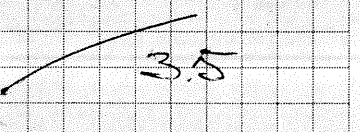
X	$\hat{Y}$
1	8,15
2	8,05

b) Determinationskoefficienten visar på det linjära sambandets styrka uttryckt i procent och ju närmare 1 det är desto starkare blir det (1=100%) detta värde (0,75) visar att sambandet är relativt starkt. $0 \leq r^2 \leq 1$ När den är 1 är residualerna = 0 och alla erhållna obs. stämmer med linjen.

c) $\hat{y} = 8{,}25 - 0{,}10 \times 5 = 7{,}75$

enligt regressionslinjens ekvation blir arbetslösheten 7,75% när inflationen är 5%, dock bör prediktionerna (prognoser) med hjälp av denna göras med extrem försiktighet. Denna "trend" (riktning) (prognos) behöver inte alls vara förhållande / korrekt.

d) Residualerna visar på avståndet mellan regressionslinjen (det genomsnittliga värdet) och de erhållna värdena. Om det är positivt är de erhållna värdena större än genomsnitt om det är lägre är de erhållna värdena mindre än genomsnitt.



SU, STATISTIK

Skrivsal: BR Anonymkod: SFS0026 Blad nr: 3

4 Målpopulationen är ungdomar 16-24 år boende i den specifika kommuner, det är dessa man vill undersöka.

Rampopulationen är de man kan undersöka (ramen) för urvalet, i denna situation blir det givet att undersöka ungdomar 16-24 som är skrivna i kommunen vid undersökningstillfället.

a) ett obundet slumpmässigt urval (OSU) innebär att alla i populationen har samma sannolikhet att bli individer valda (ett s.k sannolikhetsurval) och är inte styrt/riktat på något vis (obundet). (ex. lottning) Detta kan vara att föredra i och med att det med hjälp av detta blir ett helt slumpmässigt urval, vilket var något som eftersöktes. Dock så kan detta vara ofördelaktigt i det hänseende att det kan finnas eller annan egenskap många ungdomar som är en viss ålder och därför göra att dessa kan ha en större chans att bli representerade medan andra missas helt. om det t.ex endast finns ett fåtal 22åringar och många 17åring

b) Ett systematiskt slumpmässigt urval innebär att ett OSU först utförs på de första 1-10 individerna i registret (detta urval utförs när individerna i en population är nummererade) därefter tas man exempelvis var 10:e om stickprovet ska vara 10% av populationen, alltså om nr. 4 blir slumpad kommer också nr. 14, 24, 34, 44 osv. med i stickprovet. Eftersom vår population är ordnade i ett register,

vg.v. →

4. b) fortsättning:

över var de är skrivna, så numreras de och plockas ut enkelt mha. denna metod. Vidare kan en bred åldersfördelning även fås här ifall de ordnas efter födelsedatum, vilket är fördelaktigt för vår undersökning. Dock så tas inte andra bakgrundsvariabler hänsyn till här så det skulle kunna slumpa sig som så att stickprovet endast innefattar/har en överdriven representation av ex. killar, eller personer boende i ett visst område. Varför?

c) Stratifierat urval innebär att man delar in populationen i olika delpopulationer (stratan) mha en viss stratifieringsvariabel för att det inte ska bli missvisande resultat om det finns en överrepresentativ med individer med vissa egenskaper, samt för att kunna särredovisa de olika stratanas resultat. När uppdelningen av populationen har gjorts så utförs ett OSU inom varje strata.

Detta kan vara bra för vår undersökning då det gör det möjligt för oss att även dela upp populationen efter exempelvis kön eller geografiskt område för att då kunna se hur individer med dessa egenskaper vill sölja sig från kommunen eller ej. Negativt här är att det dock antas att individer med en viss egenskap tänker lika, vilket givetvis inte behöver vara fallet, utan den/de individ/er som undersöks kan ju skilja sig från gruppen som helhet

d) Ett klusterurval undersöker grupper av element (individer) för att möjliggöra jämförelser sinsemellan. Dessa grupper slumpas fram mha OSU. Här är man ute efter att se hur grupperna i sin helhet skiljer sig och inte de enskilda individerna inom dessa grupper.

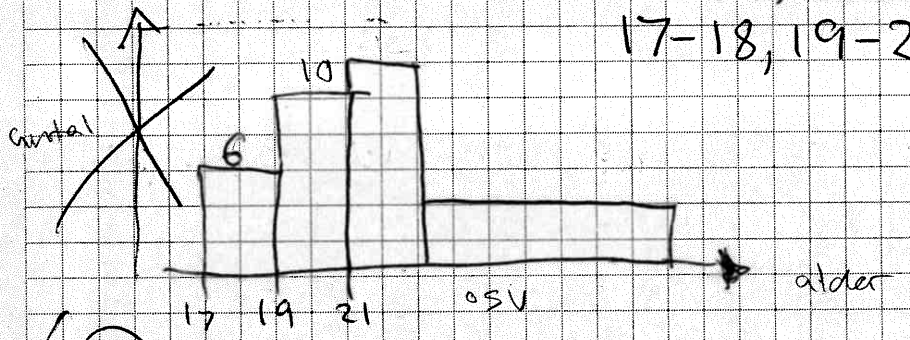
Detta blir negativt för vår undersökning i och med att den är ute efter att ta reda på hur dessa slumpmässigt valda ungdomar tycker (enskilt) för att i och med detta troligen se tendenser hos vissa individer (och kanske lyckas lösa problem med utflyttning)

Dock så kan denna metod vara resursbesparande. Samt att jämförelser mellan t.ex. geografiska områden kan göras.

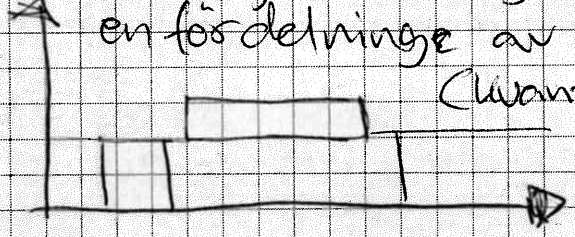
SU, STATISTIK

Skrivsal: BR **Anonymkod:** SFS 0026 **Blad nr:** 4

5. a) Ett diagram som redovisar kvantitativa variabler som är klassindelade tex. ~~[illegible]~~ åldersfördelning i en klass på universitetet ~~[illegible]~~ 17-18, 19-20, 21-22, 23-

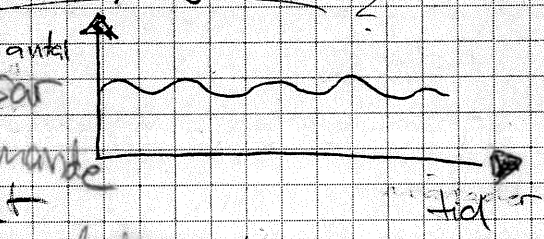


b) Ett diagram redögor på ett tydligt sätt hur en fördelninge av ett antal observationer (kvantitativa variabler) ser ut.



???

c) Cyklisk variation (ett samband, linjärt) ?



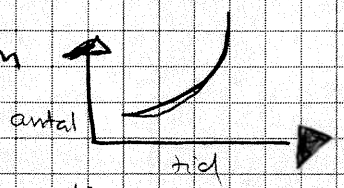
när ett antal observationer uppvisar en viss tendens som är återkommande som exempelvis att det historiskt sett har fötts många fler barn det andra kvartalet jäntemot resten av året. (En återkommande trend) → fr säsongvariation!

en variabel som endast kan kategoriseras ?

d) En dikotom variabel är en nominell variabel som endast kan anta två värden ?

Ja/nej, man/kvinna

e) exponentiell trend linjärt
är ett samband som ökar
(eller minskar) exponentiellt.
Exempelvis när man planterar en kaninstam
så börjar det lägt och ökar exponentiellt.

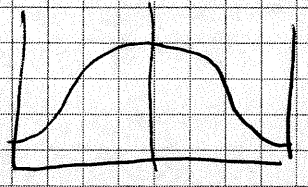


f) Hansen-Hurwitz metoden (är när man
ställer "svårare" frågor i slutet av ett frågoformulär
och börjar med de "lättare" om bakgrundsvariablerna
för att få respondenten mer villig att svara på
dessa. ✓)

g) Prestigebias är ett systematiskt fel (svar) som
uppkommer i och med att respondenten känner ett
socialt tryck till att anpassa sina svar till
"acceptabla" värderingar, t.ex. så kan en
respondent påsta att hon/han har gått på
teater fler gånger än hon/han har pga. prestige
och då göra att resultatet av undersökn blir missvisande.

h) konfidensintervallet är oftast 95%-igt,
redovisar normalfördelning.

Det är 95% konfidens att
stickproven (av en viss storlek)



stämmer överens med populationen
och därmed är väntevärdesriktiga i genomsnitt över alla stickprov
(när ett stickprov är framtaget mha osu stämmer
parametrar såsom medelvärde för stickprovet
överens med medelvärdet för populationen ???)

Statistiska institutionen

Stockholms universitet

Rättningsblad

Datum: 28/4-2015

Sal: Brunnsvikssalen

Tenta: Statistik för statsvetare

Kurs: Statistik för statsvetare

ANONYMKOD: SFS-0014

☒ 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. 3.5	4	3	~~2.5~~ 3	2.5					

POÄNG	BETYG	Lärarens sign.
16	B	

SU, STATISTIK

Skrivsal: Brunnsviksalen **Anonymkod:** SFS-0014 **Blad nr:** 1

1

a) Variabel = Hur många år man varit medlem

Variabeltyp = Variabeln är kvantitativ och kontinuerlig då den anges i numeriska värden som kan innehålla decimaler (ex 2,5 år osv)

Datanivå = Kvotdata föreligger då en absolut nollpunkt finns samt att svaren kan delas in i kvoter (ex 4 år är dubbelt så många som 2)

Frekvens = Antalet undersökta, alltså de aktiva medlemmarna i det politiska partiet

b) Anledningen till att två typvärden erhålls var antagligen på grund av att det fanns två lika stora grupper som var för sig stod för den högsta frekvensen. Exempelvis 20 individer som svarade att de varit medlemmar i 20 år samt lika många som svarade att de varit medlemmar i 15 år.

c) Om partiet består av ett fåtal äldre personer som varit aktiva medlemmar betydligt längre än resterande individer i partiet så utgör de äldre personerna ett extremvärde som gör att medelvärdet (genomsnittet) ökar. Medianen är dock inte lika känslig för dessa extremvärden då den går ut på att identifiera det mest centrala värdet då samtliga observationer (svar) är uppställda i stegrande ordning.

d) Tre olika spridningsmått är standardavvikelsen, variationsvidden och kvartilavståndet. Då standardavvikelsen är kopplat till medelvärdet är detta ej att föredra då medelvärdet i detta fall är påverkat av extremvärden (som nämnts i föregående fråga). I det aktuella fallet är det istället kvartilavståndet som kan tänkas ge bäst vidare information, detta då kvartilavståndet fungerar bäst då det mest tillförlitliga typvärdet är medianen. P2

3.5

(2) a) b)

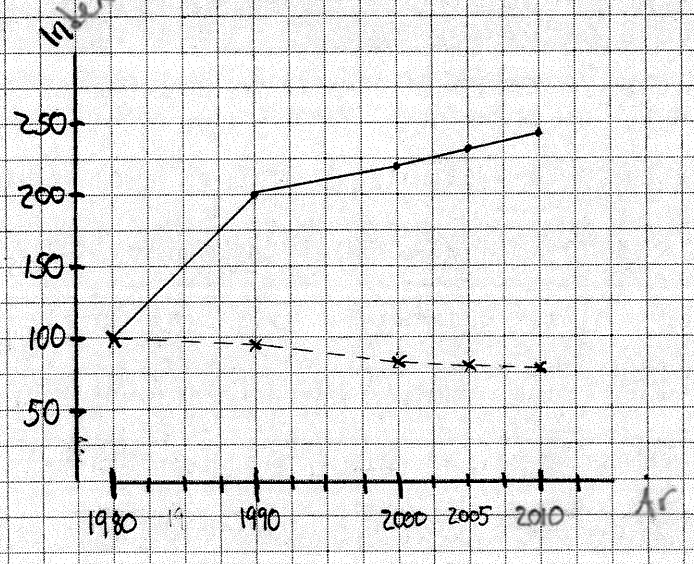
År	Utgifter	Index löp	KPI	Deflator	Index fast
(Basår) 1980	40	100	100	1	100
1990	80	200*)	208	0,48*)	96*)
2000	88	220	260	0,39	85,8
2005	92	230	280	0,36	82,8
2010	96	240	303	0,33	79,2

*) $\frac{80}{40} = 2 \cdot 100$ (Basår)

*) $\frac{100}{208} = 0,48$ (Basår)

*) $0,48 \cdot 200 = 96$

c)



SU, STATISTIK

Skrivsal: Brunnsviksalen **Anonymkod:** SFS-0014 **Blad nr:** 2

d) I löpande index har den procentuella utgiftsförändringen ökat med 140 % medan den i fast index har minskat med 20,8 % ($\frac{79,2}{100} = 0,792 \rightarrow 1 - 0,792 = 0,208$ (20,8 %))

3) a) Ekvation: $\hat{y} = 8,25 - 0,10x$ $(\hat{y} = a - bx)$

Variabeln a är i detta fall 8,25 och står för arbetslösheten i % då inflationen är 0. a-värdet är alltså det värde där linjen korsar y-axeln då $x = 0$.

Variabeln b är i detta fall den förändring som sker varje gång som x (inflationen i %) förändras med ... ?

b) Determinationskoefficienten uttryckt i % blir i detta fall 75 % vilket innebär att i 75 % av de iakttagelser som gjorts så råder ett samband mellan de två variablerna (arbetslöshet och inflation).

c) $\hat{y} = 8,25 - 0,10 \cdot 5$

$\hat{y} = 8,25 - 0,5 = 7,75$

Ett år då inflationstakten är 5 % skall alltså arbetslösheten motsvara 7,75 %. Arbetslöshet behöver dock inte följa samma kurva hela tiden och det kan finnas betydligt fler faktorer än inflation som påverkar förändringen av arbetslöshet. Då samtliga faktorer ej här är medräknade så finns en stor osäkerhet i denna prediktion.

SFS-0014

d) Residualer är de punkter som avviker från den räta linjen. 2
En punkt som hamnar under linjen kallas för negativ residual vilket innebär att en underskattning av punktens värde föreligger. En punkt som ligger över linjen kallas positiv residual och innebär en överskattning av punktens värde.
Finns många residualer som avviker markant från den räta linjen så kan man även anta att det samband som den räta linjen står för är tämligen svagt. Ju närmre residualerna ligger den räta linjen desto starkare är sambandet mellan variablerna.

3

(4) Rampopulationen för undersökningen kan tänkas utgöras av
→ för "klusterurvalet"!
ett register över folkboktörda individer i kommunen medan målpopulationen utgörs av de individer man helt idealt vill undersöka, alltså ungdomar i åldrarna 16-24 år.
Ett obundet slumpmässigt urval är likt att dra lappar ur en hatt. Positivt med detta urval är att alla individer har lika stor sannolikhet att komma med i urvalet. Givet till situationen kan dock denna urvalsmetod vara komplicerad då målpopulationen är så "smal" vilket gör att risker för urvalsfel kan uppstå genom att individer som man faktiskt ämnar att undersöka ej kommer med i urvalet. 2
Vid ett systematiskt slumpmässigt urval kan man använda det register som finns (ramen). Man kan då t.ex. slumpmässigt dra ett startnummer i ett visst spann av siffror beroende på ramens storlet. Exempelvis dras ett nummer av de 10 första i registret och sedan går man systematiskt till väga genom 2

hela registret. Dras nummer 8 så är detta startnummret, efter det kommer nummer 18, 28, 38 osv om man går i steg om 10. Denna metod är bra då den likt föregående bygger på slumpen. Det finns dock en del nackdelar med systematiskt slumpmässigt urval. För det första kan det vara så att det register som används är systematiskt skrivet vilket i värsta fall kan innebära att det urval du får fram innehåller mycket dålig variation. Säg att jämna nummer i registret endast är kvinnor och du drar nummer 2. Detta kommer således att innebära att inga män kommer med i urvalet. För det andra kan detta urval till den givna situationen innebära en risk att den målpopulation som finns ej kommer med i urvalet.

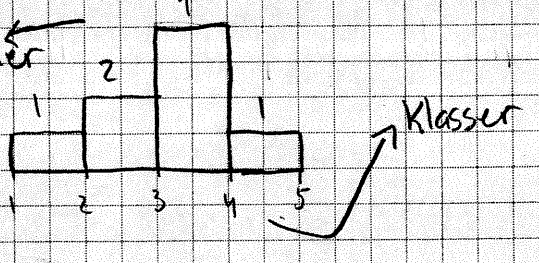
Vid ett stratifierat urval delas populationen in i olika grupper (strata). Ur varje strata görs sedan ett urval, detta urval kan vara beroende av storleken på de strata som vill undersökas. Ett stratum med många individer får därför ett större urval än ett stratum med få individer så att proportionerna blir rätt. Viktigt är också att de olika strata som finns ska skilja sig åt i stor grad. Ett alternativ till stratifierat urval i denna situation kan därför vara att dela in målpopulationen i två strata (män och kvinnor). Denna metod är mycket fördelaktig då man vill jämföra olika grupper (typer av individer) med varandra. Bristen är dock att urvalet i den givna situationen redan är mycket begränsat, dels till geografiskt område men

också till ålder. Detta kan medföra att en jämförelse av betydelse mellan män och kvinnor kan bli svår. (???)
Klusterurval, även kallat gruppurval innebär att grupper av individer slumpas eller väljs ut och där samtliga individer i gruppen får delta i undersökningen. Denna urvalsmetod är bra då det går att nå många individer och då den ger en hög svarstrehvens. Det finns dock risker för att de olika individerna i gruppen påverkar varandra vilket gör att svaren ej behöver överensstämma med verkligheten. Ett klusterurval hade i detta fall kunnat göras genom de skolor som finns i kommunen. Ett stort antal skolklasser hade kunnat ta del av undersökningen. (Bra metod då individerna i skolan även utgör målpopulationen).

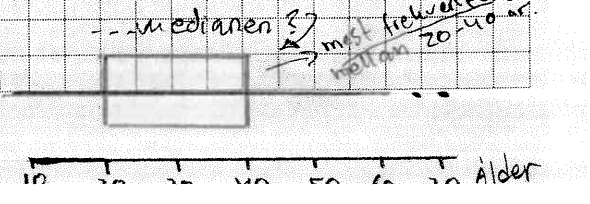
Tänker Du här på en s.k. gruppenkät?

(3)

5) a) Histogram = En diagramtyp som förespråkas vid kvantitativt och klassindelat material. Diagrammet utgörs av ihopsatta staplar där stapelns höjd visar antalet observationer och stapelns bredd visar klassens storlek.



b) Lådagram = En diagramtyp som framställs ur blad och grendiagram. Lådan visar var flest antal individer "befinner sig" i undersökningen. Lådagram kan t.ex. användas för att illustrera åldersfördelningen i en organisation. Lådan kommer att placeras där den mest frekventa åldern förekommer.



c) Cyklisk variation ≠ Resultatet av samma undersökning varierar i olika cykler. Kan t.ex. vara en undersökning som påverkas av årstider. ✓

d) Dikotom variabel = En variabel som endast kan anta två värden. Exempelvis en variabel som utgörs av värdena Ja eller Nej.

e) Exponensiell trend = En långsiktig och beständig minskning eller ökning av en linjär kurva. Exempelvis ett samband vars utvecklingskurva hela tiden har samma lutning. ✓

f) Hansen-Hurwitz metod = En metod för att minska det bortfall som uppstått i en undersökning genom att byta datainsamlingsmetod för det bortfall som uppstått. Exempelvis ringa upp och hålla en telefonintervju med de personer som ej svarat på en postenkät-undersökning.

g) Prestigebias = Respondenten svarar oärligt på frågor för att få sig själv att framstå i bättre dager. Kan t.ex. förekomma vid känsliga frågor så som alkoholkonsumtion där respondenten inte vill erkänna att den dricker mer än vad man bör.

h) Konfidensintervall ≠ Beskriver var en viss procent av ett stickprov befinner sig utifrån medelvärdet med hjälp av standardavvikelsen. Mäter olika stickprovs förhållande till medelvärde och standardavvikelse. ✓

2.5