

STOCKHOLMS UNIVERSITET

Statistiska Institutionen

Mikael Havasi

## Omentamen för Introduktion till statistik för statsvetare

Onsdag, 2014-05-28, Klockan 16:00-21:00, Brunnsvikssalen

Tillåtna hjälpmedel: Formelblad (som vidhäftats tentamen) och miniräknare, utan lagrade formler eller text

Utlämning av tentan: 13 juni, klockan 17-18, B705

Efter utsatt datum och tid för utlämning, kan tentan endast hämtas ut från studentexpeditionen

För poäng krävs tydliga svar! Tentan kan ge 100 poäng totalt. För varje fråga anges det vad respektive fråga motsvarar i poäng.

*Lycka till!*

1. För sex hushåll mättes variabeln "antal familjemedlemmar". Följande värden erhöles:

-----  
3            1            2            2            7            2  
-----

- Räkna ut medelvärdet och medianen. Förklara hur dessa lägesmått skall tolkas var för sig och i relation till varandra.(4 p)
- Räkna ut variansen och standardavvikelsen. Ge en tolkning av det mått av dessa två som bäst beskriver spridningen.(4 p)
- Vad är det för variabeltyp och datanivå som föreligger här? Motivera ditt svar.(4 p)
- Illustrera fördelningen i ett lämpligt diagram tillsammans med beskrivande kommentarer, Motivera ditt val av diagram.(4 p)

2. Variabeln "omsättningen i miljoner kronor" för företag inom en viss bransch kan tänkas förklaras av variabeln "spenderade kronor i tusental på reklam". För ett urval av 15 företag beräknas en enkel linjär regressionsmodell:

$$y = 250 + 5 * X$$

a) Vilka av de två variablerna är den beroende och den oberoende variabeln enligt modellen?(2 p)

b) Vilken av de två koefficienterna är interceptet och vilken är lutningskoefficienten? Tolka dessa två koefficienter.(4 p)

c) För den givna modellen beräknas korrelationskoefficienten till 0.61.

– Vad innebär korrelationskoefficienten generellt?(4 p)

– Ge en tolkning av det givna värdet på korrelationskoefficienten i detta fall.(2 p)

d) Beräkna determinationskoefficienten och tolka detta värde.(4 p)

3. Man är inom en kommun intresserad av befolkningens ( $\geq 18$  år) åsikter om kollektivtrafiken i kommunen. Då en totalundersökning blir för dyr blir man tvungen att genomföra en urvalsundersökning.

Redogör för nedanstående statistiska begrepp/metoder och ge exempel utifrån den givna situationen(3 poäng för varje delfråga):

- |                               |                 |
|-------------------------------|-----------------|
| a) Mål- och rampopulation     | f) Hoppfråga    |
| b) Obundet slumpmässigt urval | g) Täckningsfel |
| c) Stratifierat urval         | h) Mätfel       |
| d) Kvoturval                  | i) Urvalsfel    |
| e) Urvalsfraktion             | j) Bortfallfel  |

4. a) Genomsnittspriset för ekologisk lättmjölk under ett år är nedan angivet för tre år:

År	2011	2012	2013
Pris	7,85	7,89	8,01

Gör en enkel indexserie där 2011 är basår.(2 p)

b) Nedan ges genomsnittspriset och sålda kvantiteter i hundratusental av två ekologiska mjölksorter under tre år:

År	2011	2012	2013
Lättmjölk, Pris	7,85	7,89	8,01
Lättmjölk, Sålda förpackningar	31	37	69
Standardmjölk, Pris	8,56	8,78	8,95
Standardmjölk, Sålda förpackningar	34	40	70

Räkna ut två sammansatta index genom att använda Laspeyres och Paasches index och sätt 2011 som basår. Tolka eventuella skillnader mellan de två indexen(6 p)

5. Redogör *kortfattat* för nedanstående åtta begrepp/metoder samt ge realistiska exempel på respektive begrepp/metod (3p för varje delfråga):

- |                   |                           |
|-------------------|---------------------------|
| a) Prestigebias   | f) Chi <sup>2</sup> -test |
| b) Pilotstudie    | g) Allmänna födelsetalet  |
| c) Snöbollsurval  | h) Konfidensintervall     |
| d) Ordinalskala   | i) Nonsenssamband         |
| e) Binär variabel | j) Residual               |

# 1. Beskrivande statistik

## 1.1 Medelvärde, varians, standardavvikelse

Ett statistiskt material består av  $n$  observationer

$$x_1, x_2, \dots, x_n$$

Medelvärdet är

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

Variansen är

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

Standardavvikelsen är

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

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

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

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

**Räkneregler**

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

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

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

## 1.2 Regression, korrelation

Regressionslinjen är  $y = a + bx$ .

Regressionskoefficienten är

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

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

Korrelationskoefficienten är

$$r = \frac{\sum (x - \bar{x})(y - \bar{y})}{\sqrt{\sum (x - \bar{x})^2 \sum (y - \bar{y})^2}} = \frac{\sum xy - \sum x \sum y / n}{\sqrt{[\sum x^2 - (\sum x)^2 / n][\sum y^2 - (\sum y)^2 / n]}} = \frac{b s_x}{s_y} \quad (1.2.3)$$

Residualvariansen är

$$s_e^2 = \frac{n-1}{n-2} s^2 (1-r^2) = \frac{1}{n-2} (\sum y^2 - a \sum y - b \sum xy) \quad (1.2.4)$$

## 1.3 Prisindex

Laspeyres index är

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

Paasches index är

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

# Rättningsblad

**Datum:** 28/5-2014

**Sal:** Brunnsvikssalen

**Tenta:** Statistik för statsvetare

**Kurs:** Introduktion till statistik för statsvetare

**ANONYMKOD:**

SFS-0004

4

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 34
Lär.ant. 14	6	24	8	22					

POÄNG 74	BETYG C	Lärarens sign. Munster
-------------	------------	---------------------------

1a) medelvärde:  $\bar{x} = \frac{\sum x}{n}$

$$\frac{3 + 1 + 2 + 2 + 7 + 2}{6} = \underline{\underline{2,8}}$$

median:

$$\left. \begin{array}{l} 1 \\ 2 \\ 2 \\ 2 \\ 3 \\ 7 \end{array} \right\} \frac{2+2}{2} = \underline{\underline{2}}$$

SVAR: medelvärdet är 2,8 & betraktas som ett genomsnittsmått för alla observationer.

Medianen är 2 & är ett bättre lägesmått än medelvärdet om det förekommer extremvärden

1b) varians:  $s^2 = \frac{\sum (x - \bar{x})^2}{n-1} = \frac{(17 - 2,8)^2}{6-1} = \frac{201,64}{5} = \underline{\underline{40,328}}$

standardavvikelsen:  $s = \sqrt{s^2} = \sqrt{40,328} = \underline{\underline{6,35}}$  Lite fel

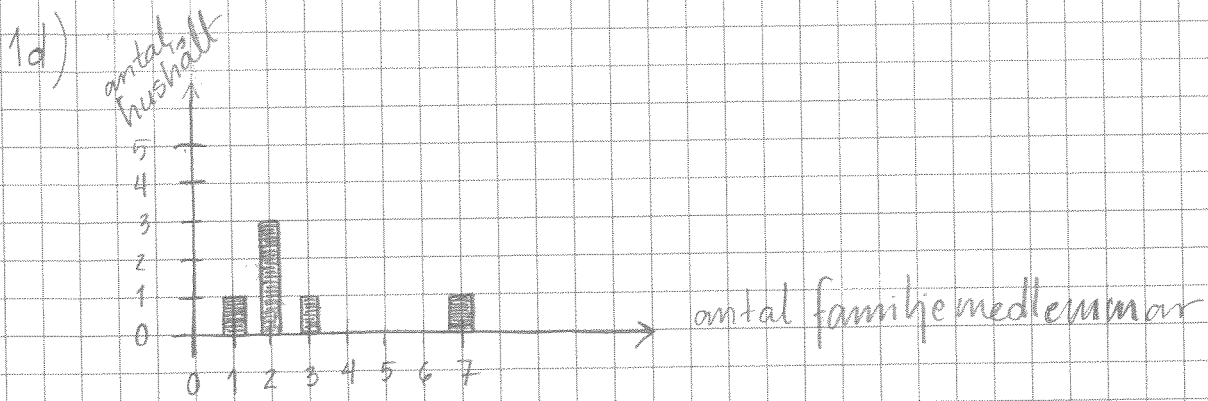
SVAR: Variansen är 40,328 & är ett spridningsmått som beskriver

standardavvikelsen är 6,35 & är ett mått som i genomsnitt beskriver hur långt från 2 medelvärdet ( $\bar{x}$ )

1c) Variabeltyp = kvantitativ - behöver ej operationaliseras för att beräknas numerisk - antalet familjemedlemmar = beräkningsbart utan att begreppet operationaliseras

Data nivå:

Kvotskala - det finns en absolut nollpunkt där enheter betyder ingen familjemedlem, (men då finns det heller inget hushåll.)



Stapelldiagram ger en snabb & enkel översikt över hur fördelningen över antalet familjemedlemmar i de tillfrågade (respondenternas) hushåll. / 14

2a) Den beroende variabeln (y) är "spenderade kronor..."  
 Den oberoende variabeln (x) är "omsättning i..."  
 D.v.s. att ju högre omsättning desto mer pengar spenderas på reklam. 0 Tvärtom

2b) Interceptet:

lutningskoefficienten:

0

2c) Korrelationskoefficienten (0,61) berättar hur starkt det linjära sambandet är. Kan anta värden fr  $-1 \rightarrow +1$  där  $-1$  = starkt negativt samband &  $+1$  = starkt positivt samband. Ju närmre 0 desto svagare samband. 0,61 = positivt samband. / 7

2d) Determinationskoefficienten visar igenomsnitt hur mycket variationerna för y-variabeln påverkas av variationerna i x-variabeln.

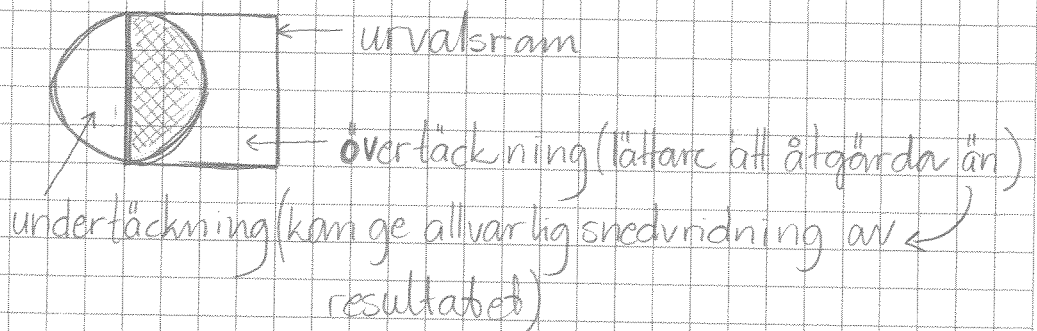
2

/ 6

3a) Målpopulation = de element man idealiskt sett vill undersöka (måste finnas i en ram).

ex. → Slumpmässigt utvalda personer ( $\geq 18$  år) som finns bland kommunens befolkning.

Rampopulation = de enheter som finns inom en ram.



ex. → Kommunens befolkning som är  $\geq 18$  år

3b) OSU - förhållandvis enkelt att genomföra.

ex. → Man slumpar fram personer som får ingå i undersökningen.

3c) Stratifierat urval = man delar in populationen i olika strata / delgrupper som gärna får vara så olika / heterogena varandra som möjligt för att man på så vis kan dra slutsatser fr en grupp jämf m om man tittat på hela populationen (som vid en OSU)

→ ex. delar in populationen geografiskt el. i ålder. De som bor i en viss del av kommunen tycker så jämf m andra som tycker så.



3d) kvoturval: risk för okänd undertäckning & snedvridning  
man fyller sin kvot, ex. man vill spegla befolkningen  
i stort & fyller sina kvoter i % så att x antal  
kvinnor %-uett sett är detsamma som %-en  
kvinnor i hela befolkningen.  
→ man letar efter befolkningens  $\leq 18$  år & "höjer sig" när man  
fyllt en viss kvot / kommit upp i ett bestämt antal svar

3e) Urvalsfraktion

3f) Hoppfrågor: används vid enkätundersökningar.

ex. Röker du?

svarsalt: ja

nej (om nej, hoppa till fråga x)

kan användas för att urskilja individer

ex. → i telefonundersökning

Åker du kollektivt?

ja

nej, (om nej gå till fråga 6)

3g) täckningsfel: ingår i det "totala felet" / Non sampling error  
undertäckning-individer som ej ingår i urvalsramen  
övertäckning-enheter som ingår i urvalsramen,  
men ej i målpopulationen, lättare fel  
att upptäcka & åtgärda än undertäckning

→ ex. vid en enkätundersökning kan man fråga  
respondenten om den är  $\geq 18$  år för att  
sälla ut dem som ej är det.

3h) mätfel: kan orsakas av ex. mätinstrument, respondenters svar  
 Non-sampling-error, ingår vid det "totala felet"

→ ex. respondenter var bor ej i kommunen. 1

3i) urvalsfel - den urvalsram man arbetar efter är  
 ej optimal för resultatet/det undersökningen  
 är menat att ge svar på.

→ ex. man tillfrågar personer som ej ingår i kommunens befolkning el. ej ingår i rätt ålderskategori.

3j) bortfallsfel: man får ej svar fr de respondenter som ingår i målpopulationen. Kan omfattningen vara individbortfall (helt uteblivet svar) el. partiellt bortfall (ex. vissa frågor lämnas obesvarade).

Non-sampling-error  
 ingår vid det totala felet

→ ex. bortfallsfelet blir stort så undersökningen kan ej anses hålla hög kvalitet.

TQM = innehåll, aktualitet, jämförbarhet, tillförlitlighet, validitet

år	2011	2012	2013
pris	7,85	7,89	8,01
index	100	$\frac{7,89}{7,85} \times 100$ <u>= 100,51</u>	$\frac{8,01}{7,85} \times 100$ <u>= 102,04</u>

2

1/24

4b) Laspeyres  $\frac{\sum p_t q_0}{\sum p_0 q_0} \times 100$  Paasches  $\frac{\sum p_t q_t}{\sum p_0 q_t} \times 100$

2012

varav	$p_t \times q_0$	$p_0 \times q_0$	$p_t \times q_t$	$p_0 \times q_t$
lättmj.	244,59	243,35	2911,93	2910,45
stand.mj.	298,52	291,04	351,2	342,4
totalt:	543,11	534,39	643,13	632,85

Laspeyres 2012 =  $\frac{543,11}{534,39} \times 100 = \underline{\underline{101,63}}$

Paasches 2012 =  $\frac{643,13}{632,85} \times 100 = \underline{\underline{101,62}}$

2013

varav	$p_t \times q_0$	$p_0 \times q_0$	$p_t \times q_t$	$p_0 \times q_t$
lättmj.	248,31	243,35	552,69	541,65
stand.mj.	304,3	291,04	626,5	599,2
totalt	552,61	534,39	1179,19	1140,85

Laspeyres 2013 =  $\frac{552,61}{534,39} \times 100 = \underline{\underline{103,41}}$

Paasches 2013 =  $\frac{1179,19}{1140,85} \times 100 = \underline{\underline{103,36}}$

6

8

5a) prestigebias = ex. vid enkätundersökningar ger respondenterna ett mer el. mindre medvetet förskönande svar som ej är helt sanningens enligt.  
 ex. hur mycket man bränner

5b) pilotstudie = testgenomgång av ex. en enkät för att upptäcka ev. brister som man ej upptäckt innan.

ex. testa så att frågorna i en enkät är förståeliga och svarsalternativen är utförliga.

5c) snöbollsurval = en individ får leda till en ytterligare som får leda till en ytterligare osv.

Okänt under täckningsfel!

ex. använd vid undersökningar om brottslighet, missbruk, prostitution (där individer kan vara svåra att finna)

5d) ordinalskalor = kvantitativ variabel som kan rangordnas  
 ex. används vid åsikts- & attitydskalor

5e) binär variabel = dikotom/0/1-variabel

kan endast ta två värden

ex. "är du gift?" → ja/nej

5f) Chi-två-test = beräknar skillnaden i  $\chi^2$ -led för konstanterna a & b.

5g) Allmänna födelseåret - vid demografiska undersökningar undersöker ex. en kohort grupp för att kunna göra jämförelser över tid.

5h) konfidensintervall = berättar ex i genomsnitt utifrån medelvärdet var en viss % av stickproven befinner sig utifrån standardavvikelsen.

ex. på ett konfidensintervall på 95% ligger stickproven m 95% säkerhet inom intervallet & 5% ligger utanför. Vill man öka säkerheten så ökar man antalet stickprov.

5i) nonsenssamband = sambandet mellan x & y-faktor finns till synes, men påverkas i själva verket av en annan faktor.

ex. de som har hög lön är längre än de som

har en låg lön → egentligen är det faktorn "kön" som påverkar lönen & ej längden på personen.

5j) Residual = punkt som finns i ett regressionsdiagram nära el. längre ifrån regressionslinjen  
Z (spridningsdiagram)

22



Stockholms  
universitet

Statistiska institutionen

## Rättningsblad

**Datum:** 28/5-2014

**Sal:** Brunnsvikssalen

**Tenta:** Statistik för statsvetare

**Kurs:** Introduktion till statistik för statsvetare

**ANONYMKOD:**

SFS-0007

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					5 2k
Lär.ant.	13	4	24	6	20					

POÄNG	BETYG	Lärarens sign.
67	D	Mina Hem

1. a) • Medelvärde:  $\frac{3+1+2+2+7+2}{6} \approx 2,83 = \bar{x}$

- 2,83 är medelvärdet och visar därmed det mittersta värdet genom en generell beräkning, även mer noggrann.

• Median: 1 2 2 2 3 7

↓

$\frac{2+2}{2} = 2$

- Medianen är det absolut mittersta värdet bland de värden som presenterats. Man räddar upp värden i storleksordning för att få fram det mittersta värdet, 2.

- Det är i snitt 2,83 familjemedlemmar i varje hushåll. Det blir en aning problematisk med detta resultat då 0,83 inte kan räknas som en person. Medianen som är 2 ger ingen detaljerad beskrivning av antalet familjemedlemmar. 3

b) • Variansen:  $s^2 = \frac{\sum (x - \bar{x})^2}{n-1} \Rightarrow s_n^2 = (x - \bar{x})^2$

①  $s_1^2 = (1 - 2,83)^2 = 3,3489$

②  $s_2^2 = (2 - 2,83)^2 = 0,6889$

③  $s_3^2 = \text{---} = 0,6889$

④  $s_4^2 = \text{---} = 0,6889$

⑤  $s_5^2 = (3 - 2,83)^2 = 0,0289$

⑥  $s_6^2 = (7 - 2,83)^2 = 17,3889$

$\Rightarrow s_1^2 + s_2^2 + s_3^2 + s_4^2 + s_5^2 + s_6^2 = 22,8334 = \sum s_n^2$

$s^2 = \frac{22,8334}{6-1} = 4,56668 \approx 4,57$

- Standardavvikelsen är det mått som bäst beskriver spridningen (då det tillhör spridningsmått) som ligger på ca 2,14.

4

• Standardavvikelse:  $s = \sqrt{s^2}$

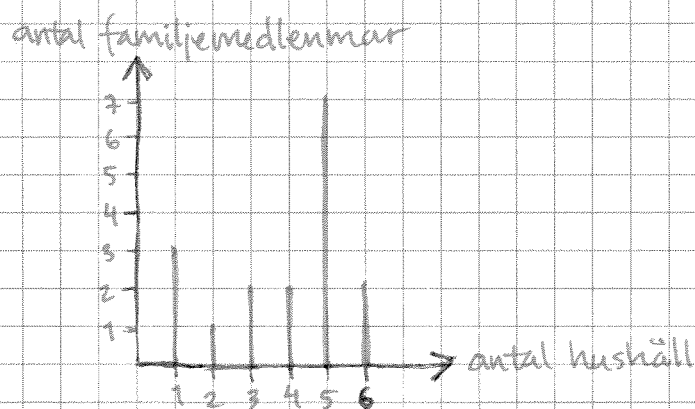
$s = \sqrt{4,57} \approx 2,14$

VGV  
⇒

c) • Variabeltypen som föreligger i detta exempel är diskret kvantitativ variabel eftersom man söker efter antal familjemedlemmar i varje hushåll i sex hushåll. Det handlar om heltal då antalet inte kan anta vilket värde som helst, därav diskret.

• Datanivån som föreligger i detta exempel är intervall då det endast går att se en yttlig skillnad mellan antalet familjemedlemmar i varje hushåll (ex. i hushåll nr. 1 bor det 2st fler än i hushåll nr. 2.). Den har ingen absolut nollpunkt då det inte kan bo 0 personer i ett hushåll. Grundläggande räknesätt är tillåtna/kan fungera som addition och subtraktion.

d) Det diagram som skulle vara lämpligt i detta fall med tanke på variabeltypen och datanivån är ett stolpsdiagram.



3

13



2. a)  $y = 250 + 5 \cdot x$

- Enligt modellen (formeln) är den beroende variabeln (y) "omsättningen i miljoner kronor" & den oberoende variabeln är "spenderade kronor i tusental på reklam" (x). 2

b) interceptet är 250 & lutningskoefficienten är 5.

2

c) - Korrelationskoefficienten (r) visar orsakssambandet mellan den beroende och den oberoende variabeln.

d)

4

3. a) Målpopulation är den populationen man egentligen vill undersöka som i detta fall är kommunens befolkning som är 18 år och äldre.

Rampopulationen är som ett redskap för att undersöka målpopulationen genom en ram (register/förteckning).  
I detta fall är rampop. alla registrerade inom kommunen som är 18 år och äldre. 3

b) OSU är ett slumpmässigt urval som tillåter alla individer eller element inom en undersökning samma sannolikhet att komma med i urvalet. Därav obundet och icke-styrt urval.  
I detta fall blir det att man slumpmässigt väljer ut individer som är 18 år och äldre med hjälp av en ram (register), alltså den registrerade delen (rampop.). 3

c) Stratifierat urval är ett urval som ger samma sannolikhet för alla individer <sup>element</sup> att komma med i ett urval genom att dela in dessa i stratum som bör vara så lika som möjligt för ett "rättvist" resultat. När man delat in individer/element i stratum (som grupper) väljer man sedan slumpmässigt ut några från varje stratum och låter dessa bilda urvalsgruppen.  
Detta för att minoriteter ska få samma chans att representeras i ett urval som majoriteter för att sedan ge ett någorlunda generaliserbart resultat.

I detta fall skulle man kunna dela in de som åker varje dag i kollektivtrafiken i ett stratum, de som åker några gånger veckan i ett stratum och de som åker någon gång i månaden i ett stratum. Sedan gör man ett urval ur varje stratum beroende på storleken av varje stratum. 3

vgr.  
⇒

## d) Kvoturval

○

e) Urvalsfraktion är det bestämda antalet individer/element  
Z man vill ha i sitt urval. I detta fall är det inte angivet, men  
bör bestämmas då en avgränsning underlättar undersökningen.

f) Hoppfråga är en fråga som kan uteläsa en rad följdfrågor  
om man ger ett visst svar. I detta fall om respondenten  
får frågan "Sitter du när du åker i kollektivtrafiken?"  
Om nej, hoppa till fråga 7. En följdfråga till denna skulle  
kunna vara "Är sittplatserna bekväma?", om vi utgår ifrån  
att respondenten svarar "nej" på huvudfrågan blir det svårt  
Z för denne att besvara följdfrågan då denne inte vet hur det  
känns att sitta på bussen/tåget. Följdfrågan blir irrelevant.

g) Täckningsfel är ett icke-urvalsfel som i stort innebär att  
man undersöker en annan population än den man planerat  
undersöka. Det finns två typer av täckningsfel, ena är  
övertäckning och innebär att man undersöker individer/element  
som inte tillhör målpop. Den andra kallas undertäckning  
och menar på att man missat, man får inte med individer/element  
Z som egentligen bör finnas med. I detta fall om man  
skulle få med yngre än 18 år eftersom denne är registrerad  
eller om man förbörar en del 18-åringar och äldre unga.  
att dessa ej är registrerade.

h) Måttfel är ett systematiskt fel och står för skillnaden mellan  
ernållet och sant värde. I detta fall om mätinstrumenten  
Z är felaktiga till följd av undertäckning, man förbörar  
en del av målpopulationen och får felaktiga resultat, eller om  
respondenterna skulle ljuga i sina svar. (Kan kopplas till bortfallsfel)

3. i) Urvalsfel sker i koppling till det totala urvalet, man gör ett felaktigt urval. 1

j) Bortfallsfel är ett icke-urvalsfel som tyder på förlorad data som man planerat erhålla. Det kan bero på att respondenten missuppfattat frågan, har språksvårigheter eller att undersökaren missuppfattar respondentens svar. Att det sker tolkningsfel eller att man inte får tag på några av de man vill undersöka. I detta fall om delar av målpop. har språk svårigheter eller att de väljer att inte svara över huvudtaget. Ett bortfall av viktig info sker. 3

4. a)

År	Pris	Index (baserat på 2011)
2011	7,85	$7,85 / 7,85 = 1 \cdot 100 = 100$
2012	7,89	$7,89 / 7,85 \approx 1,005 \cdot 100 = 100,5$
2013	8,01	$8,01 / 7,85 \approx 1,02 \cdot 100 = 102$

/24

2

b) Laspeyres index 
$$\frac{\sum p_t \cdot q_0}{\sum p_0 \cdot q_0} \cdot 100$$

2011: 
$$\frac{(7,85 \cdot 31) + (8,56 \cdot 34)}{(7,85 \cdot 31) + (8,56 \cdot 34)} = \frac{534,39}{534,39} = 1 \cdot 100 = \underline{\underline{100}}$$

2012: 
$$\frac{(7,89 \cdot 31) + (8,78 \cdot 34)}{534,39} = \frac{543,11}{543,39} \approx 0,999 \cdot 100 = \underline{\underline{99,9}}$$

2013: 
$$\frac{(8,01 \cdot 31) + (8,95 \cdot 34)}{543,39} = \frac{552,61}{543,39} \approx 1,017 \cdot 100 = \underline{\underline{101,7}}$$

vgr. →

b) Paasches index  $\frac{\sum p_k \cdot q_k}{\sum p_0 \cdot q_k} \cdot 100$

2011:  $\frac{(7,85 \cdot 31) + (8,56 \cdot 34)}{(7,85 \cdot 31) + (8,56 \cdot 34)} = 1 \cdot 100 = \underline{\underline{100}}$

2012:  $\frac{(7,89 \cdot 37) + (8,78 \cdot 40)}{(7,85 \cdot 37) + (8,56 \cdot 40)} = \frac{643,13}{632,85} \approx 1,016 \cdot 100 = \underline{\underline{101,6}}$

2013:  $\frac{(8,01 \cdot 69) + (8,95 \cdot 70)}{(7,85 \cdot 69) + (8,56 \cdot 70)} = \frac{1179,19}{1140,85} \approx 1,034 \cdot 100 = \underline{\underline{103,4}}$

ÅR	2011	2012	2013
Laspeyres index	100	99,9	101,7
Paasches index	100	101,6	103,4

Rätt

- Dessa index söker efter olika saker som tydligt kan ses vid jämförandet av formelerna. Laspeyres söker resultat i jämförelse med basåret (2011) medan Paasches jämför resultat för nuvarande tidpunkt / framåt.

4

6

5. a) Prestigebias tillhör icke-urvalsfel (bortfallsfel/mätfel) och uppkommer vid prestige fyllda frågor. Ex. om intervjuaren ställer en fråga till en intervjuperson som har och göra med dennes plats i företaget/samhället (status) kan denne överdriva eller ljuga i sitt svar för att leva upp till eller skapa en viss bild av sig själv (hålla högt prestige nivå). }

b) Pilotstudie är som en testundersökning av en planerad undersökning som görs för att upptäcka eventuella fel innan man utför själva undersökningen. Man tar ett så kallat stickprov. Ex. kan vara att jag valt att undersöka arbetslösheten bland ungdomar i Stockholm.

3 Innan jag skickar ut mina enkäter får jag ett stickprov ur mitt urval och testar hur bra min enkät är (pilotstudie).

c) Snöbollsurval är ett icke-slumpmässigt urval som blir aktuellt när (känsliga) undersökningar där individer/element inom målpop. är svåra att få tag på. Om man hittar en kan den i sin tur leda till en annan, osv. Ex. om jag undersöker våldtäcket och vill få tag på våldtäcketsoffer.

Det finns inget register över dessa och många är anonyma.

Jag kan få tag på ett offer som i sin tur leder mig

3 till ett eller flera offer och som i sin tur leder mig vidare till andra offer, osv. Detta är snöbollsurval.

d) Ordinalskala är den andra av fyra datanivåer. Den visar ytliga (tydliga) skillnader mellan variabler, och det går därmed inte att räkna på detta (inga räknesätt är tillåtna)

3 Det är en datanivå som kategoriserar variabler som ex. betygsskalan A-F, där A är högst & F lägst.

e) Binär variabel är en variabel som kan ha flera utfall till skillnad från dikotom som har endast två. Ex. en fråga med tre eller fyra svarsalternativ.

h) Konfidensintervall innebär att medelvärdet med en viss procent (relativ) säkerhet finns i det erhållna intervallet och räknas därmed som väntevärdesriktigt. Detta beräknas med en formel  $a \pm 1,96 \cdot \frac{\sum (x_i - \bar{x})(y_i - \bar{y})}{\sum (x_i - \bar{x})^2}$  och beror på hur många procent som anges.

i) Nonsenssamband är som en falsk kausalitet (orsakssamband) där man tror att två variabler skapat ett visst resultat, men det i själva verket är en "yttre"/utomstående variabel som påverkat. Ex. tv-abonnemang ökar trots att allt fler får sämre syn (köper glasögon). Man slutar inte titta på tv för att man har fått sämre syn utan man köper glasögon till följd av ekonomiskt tillväxt.

j) Residual är avståndet mellan den skattade linjen och punkten (variabelvärdet). Ex.

