

Stockholms universitet  
Statistiska institutionen

**Skriftlig tentamen** avseende kursen **Introduktion till statistik för statsvetare**

\*\*\*\*\*

Fredag 27 maj 2016 kl. 15:00-20:00

Skrivtid: 5 timmar.

Hjälpmedel: Miniräknare, samt vidhäftat formelblad.

**Genomgång: 2016-06-01 klockan 15:00 till 16:00 i sal B705. (Observera att tiden är flyttad)**

-----

Tentamen består av fem uppgifter vilka totalt kan ge maximalt 10 poäng vardera, totalt 50 poäng. **För full poäng på en uppgift krävs att tydliga, fullständiga och välmotiverade lösningar samt svar inlämnas.**

Lycka till! / Per Marcus

\*\*\*\*\*

1.

Polisen genomför hastighetskontroller under en veckas tid vid en 110-sträcka straxt norr om Gävle. Följande resultat erhöles.

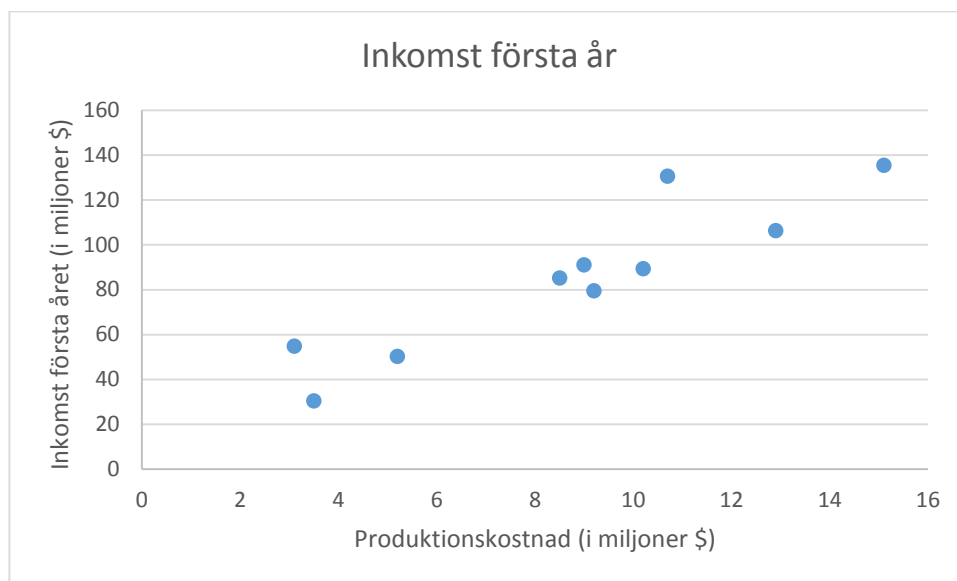
Hastighet	Antal bilar (i tusental)
<91	1
91-100	5
101-110	15
111-120	17
121-130	11
>130	7

- Ange vad som här är variabel och vad som är frekvens. För variabeln, ange variabeltyp och datanivå.
- Illustrera fördelningen i ett lämpligt diagram samt motivera val av diagram.
- Ange tre olika exempel på lägesmått, samt diskutera vilka för- och nackdelar som föreligger med dessa lägesmått.
- Ange tre olika exempel på spridningsmått, samt diskutera vilket av spridningsmått som här kan vara att föredra.

2.

Under ett år går det mycket bra för att filmbolag. En undersökning gällande inkomster och kostnader (i miljoner \$) för filmproduktion gjordes varvid följande regressionslinje anpassades,  $\hat{y} = 15,51 + 7,98x$ .

- Hur ska de erhållna värdena i regressionslinjen tolkas?
- Efter nödvändiga beräkningar fås determinationskoefficienten till ca 0,92. Hur ska detta värde tolkas?
- Vi analysen erhålls så kallade residualer. Redogör vad som avses med dessa och hur dessa ska tolkas.
- Följande diagram visar på de parvisa observationer som använts för studien. Rita ut regressionslinjen samt residualerna i diagrammet. (Förslag, Skapa ett eget diagram i din lösning)



3.

Kiruna landsting anställer Dig för att genomföra en undersökning gällande blöjanvändning i Kiruna kommun bland barn i åldrarna 0-4 år.

- Precisera vad som här kan vara målpopulation och rampopulation.
- Totalfelet består av fem feltyper. Precisera de fem feltyperna, vad de innebär samt ge exempel på dessa i denna undersökning.

4.

En omfattande undersökning skall genomföras bland slumpmässigt utvalda pensionärer i åldern 65-90 år i en viss svensk kommun. Huvudsyftet med undersökning är att förstå vad pensionärernas inställning till äldrevården i kommunen är.

Ange vad som är målpopulation samt rampopulation samt ge exempel på följande urvalsmetoder (där även deras för- och nackdelar ska kommenteras, givet undersökningssituationen som antas föreligga). Ange även vilka urvalsmetoder som är sannolikhetsurval respektive icke-sannolikhetsurval.

- a. Stratifierat urval
- b. Systematiskt urval
- c. Kvoturval
- d. Bekvämlighetsurval

5.

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

- a. Prestigebias
- b. Dikotom variabel
- c. Konfidensintervall
- d. Väntevärdesriktig skattning
- e. Ordinalskala
- f. Probes
- g. Nonsenskorrelation
- h. Intervallskala
- i. Kontinuerlig variabel
- j. Kumulativ frekvens

## 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

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

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

Residualvariansen är

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

### 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

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