

Föreläsning 8

Kapitel 9 och 10 – sid 230-284

Samband mellan kvalitativa och kvantitativa variabler

Agenda

- Samband mellan kvalitativa variabler
- Chitvåtest för analys av frekvenstabell och korstabell
- Samband mellan kvantitativa variabler

Samband mellan kvalitativa variabler

Exempel

- En bilförsäljare har sålt bilar av ett visst märke som tillverkas i tre färger: röd, svart och silver. Försäljningen av bilar i respektive färg en viss månad presenteras i följande tabell.

Färg	Antal sålda bilar (f)
Röd	13
Svart	21
Silver	17
Totalt	51

- Finns det någon skillnad i popularitet mellan färgerna på 5% signifikansnivå?

Chitvåtest (χ^2) för analys av frekvenstabell

- Krav:
 - det råder oberoende mellan grupperna i tabellen. Innebörden i detta är att samma element (person) endast får ingå i en grupp.
 - max 20% av de förväntade frekvenserna är mindre än 5
 - alla förväntade frekvenser är större än 1.
- Steg 1: Välj signifikansnivå och formulera hypoteser

H_0 : Det finns inga skillnader i frekvens mellan grupperna

H_a : Det finns skillnader i frekvens mellan grupperna

χ^2 -test för analys av frekvenstabell

- Steg 2: Testvariabeln

$$\chi_{test}^2 = \sum_{i=1}^V \frac{(O_i - E_i)^2}{E_i}$$

- V är antalet grupper
- O_i är de **observerade** frekvenserna
- E_i är de **förväntade** frekvenserna (kraven gäller för dessa), och beräknas genom:

$$E_i = \frac{\sum O_i}{V}$$

χ^2 -test för analys av frekvenstabell

- Steg 3: Beräkna kritiska området
 - Ett χ^2 -test har alltid kritiskt område till höger av fördelningen

$$\chi_{krit}^2 = \chi_{V-1;\alpha}^2$$

- Steg 4: Dra slutsats och tolka
 - Hamnar χ_{test}^2 i det kritiska området \Rightarrow förkasta H_0 .

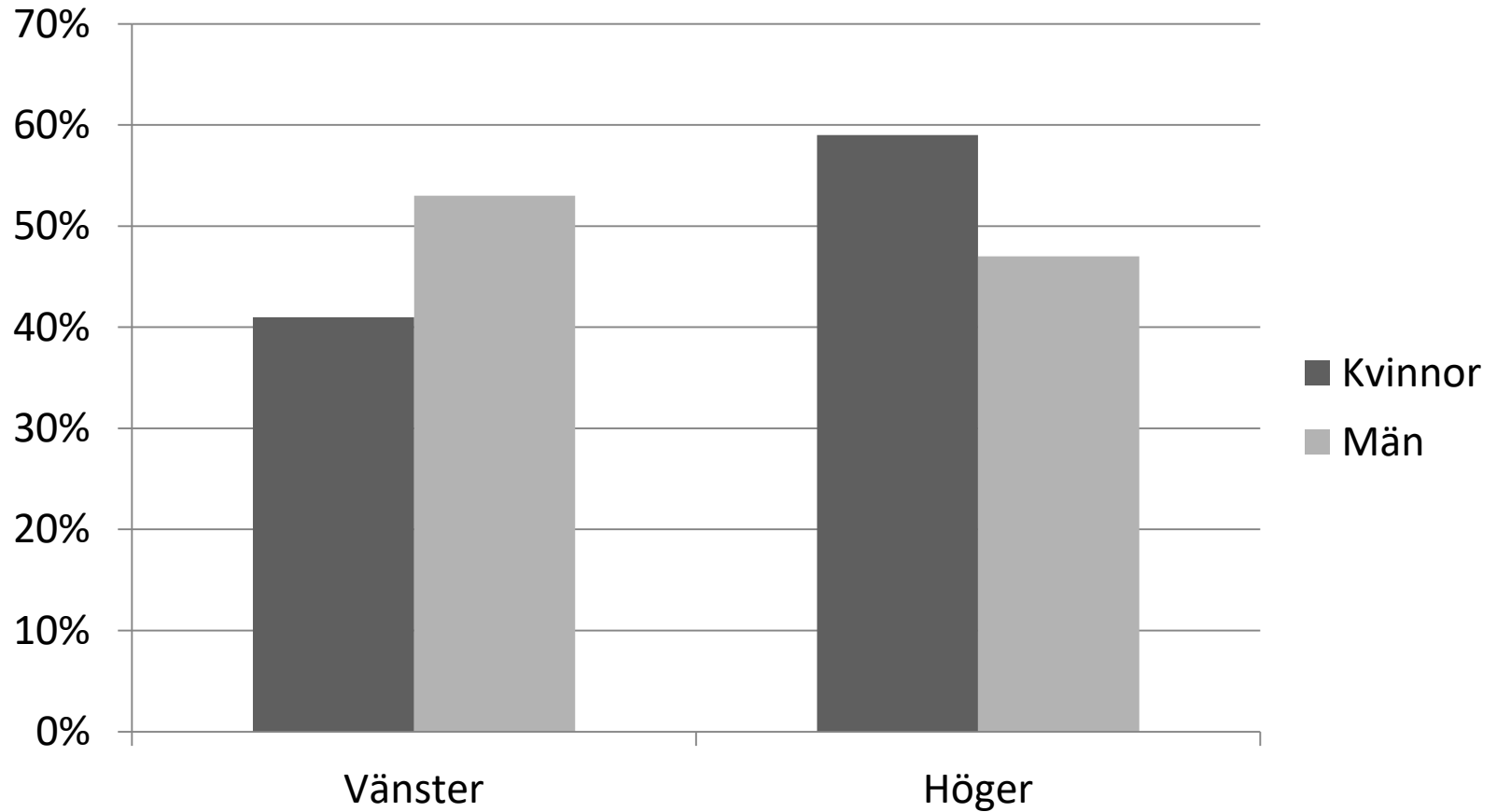
Exempel

- Man drar ett OSU om medlemmar ur en stor politiskt oberoende organisation, och frågar dels om kön, dels om politisk tillhörighet (vänster eller höger). Följande resultat erhålles.

	Vänster	Höger
Kvinna	98	141
Man	67	59

Går det att på 5 procents signifikansnivå påvisa några skillnader mellan kvinnor och män som är medlemmar i organisationen i fråga om politisk tillhörighet?

Exempel (forts.)



χ^2 -test för analys av korstabell

- Krav:
 - det råder oberoende mellan cellerna. Innebörden är att samma element (person) inte får förekomma i flera celler i tabellen
 - max 20% av de förväntade frekvenserna är mindre än 5
 - alla förväntade frekvenser är större än 1
 - Steg 1: Hypoteser
 - H_0 : *Det finns inga skillnader i fördelning mellan grupperna (alternativt: Det finns inget samband mellan grupperna)*
 - H_a : *Det finns skillnader i fördelning mellan grupperna (alternativt: Det finns samband mellan grupperna)*
-

χ^2 -test för analys av korstabell

- Steg 2: Testvariabeln

$$\chi_{test}^2 = \sum_{i=1}^W \frac{(O_i - E_i)^2}{E_i}$$

- W är antalet celler i korstabellen
- O_i är de **observerade** frekvenserna
- E_i är de **förväntade** frekvenserna (kraven gäller för dessa), och beräknas genom:

$$E_i = \frac{\text{radtot} \cdot \text{koltot}}{\text{tottot}}$$

χ^2 -test för analys av korstabell

- Steg 3: Kritiskt område

$$\chi_{krit}^2 = \chi_{(r-1)(c-1); \alpha}^2$$

– r och c är antalet rader respektive kolumner

- Steg 4: Dra slutsats och tolka
 - Hamnar χ_{test}^2 i det kritiska området \Rightarrow förkasta H_0 .

Sammanslagningar av variabler vid χ^2 -test

- Om inte kraven uppfylls måste ibland sammanslagningar av alternativ genomföras
- Dessa måste ske på logiskt vis, t.ex. måste ordningen bibehållas
- Vid nominalskala är det enklast att skapa ett nytt alternativ som kallas "Övrigt"

Exempel

- På ett företag angav chefer och övriga hur många dagar i veckan de motionerade

Antal dagar	Chefer	Övriga
0	6	38
1	8	19
2	5	23
3	5	10
4	0	4
5	1	3
6	1	2
7	0	1

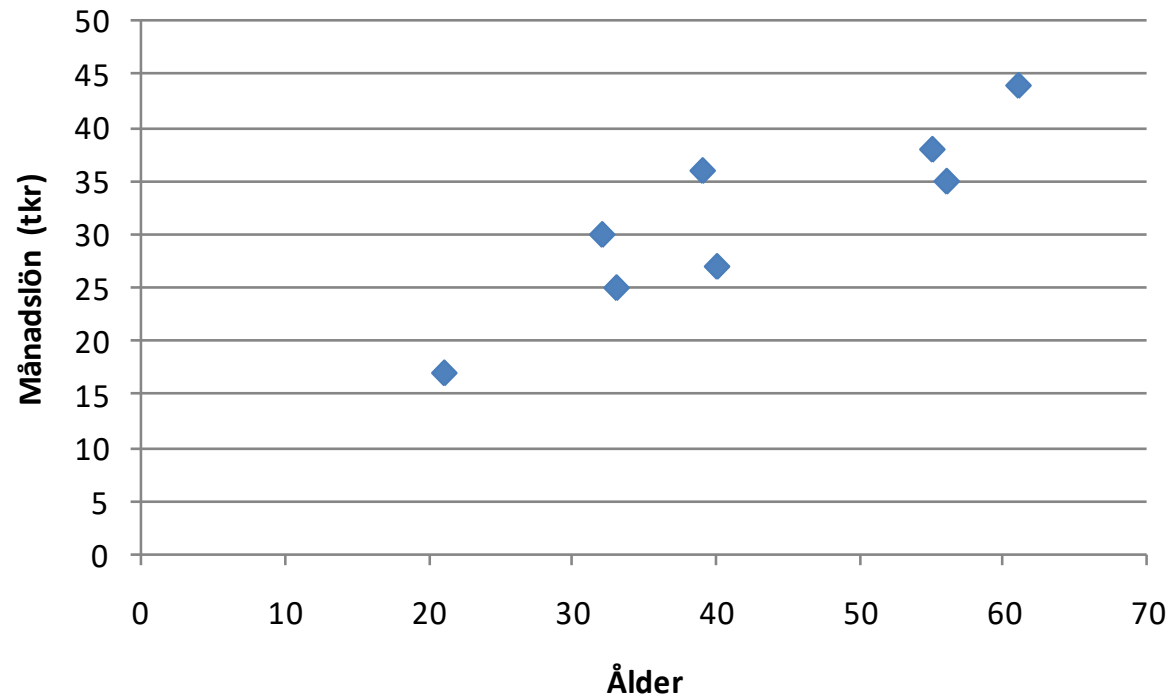
Samband mellan kvantitativa variabler

Exempel

- Ett företag har ett större antal säljare anställda. Vi har dragit ett OSU om 8 av dessa och för varje utvald person undersökt ålder och månadslön (i tusentals kronor). Föreligger det något samband mellan ålder och månadslön för säljare vid företaget?

Säljare	Månadslön	Ålder
1	17	21
2	30	32
3	27	40
4	35	56
5	44	61
6	38	55
7	36	39
8	25	33

Exempel



Att studera i ett spridningsdiagram

- Är sambandet linjärt?
 - Undersök om punkterna faller längs en tänkt rät linje

- Lutar punktsvärmen?
 - Lutar punkterna uppåt är det ett positivt samband, nedåt ett negativt samband

Att studera i ett spridningsdiagram

- Hur starkt är sambandet?
 - Undersök hur tätt observationerna ligger längs den tänkta räta linjen. Om de är utspridda är sambandet svagt, om de ligger nära linjen är sambandet starkt
- Finns det några observationer som kraftigt avviker från de övriga?
 - Dessa observationer kallas för uteliggare och kan (men inte behöver) bero på felmätning eller felinmatning

Korrelationskoefficienten

- Matematiskt mått för styrkan av ett linjärt samband mellan två variabler

$$r = \frac{SS_{XY}}{\sqrt{SS_{XX} \cdot SS_{YY}}} = \frac{\sum_{i=1}^n (x_i - \bar{x})(y_i - \bar{y})}{\sqrt{\sum_{i=1}^n (x_i - \bar{x})^2 \cdot \sum_{i=1}^n (y_i - \bar{y})^2}}$$

- Korrelationskoefficienten antar värden mellan -1 och $+1$.
 - Ju närmare -1 desto starkare negativt linjärt samband
 - Ju närmare $+1$ desto starkare positivt linjärt samband
 - Om korrelationskoefficienten är nära 0 finns inget linjärt samband

Tolkning av korrelationskoefficienten

- Vi tolkar *absolutvärdet* av korrelationskoefficienten (betecknas $|r|$):

$ r $	Samband
> 0.85	Mycket starkt
$0.65 - 0.85$	Starkt
$0.35 - 0.65$	Måttligt
$0.20 - 0.35$	Svagt
< 0.20	Mycket svagt

Korrelation \neq Kausalitet