

## F13

---

Konfidensintervall  
och  
hypotesprövning  
Chi-tvåtest

## Förra gången (F12)

---

- Stratifierat urval
  - Dela in populationen i homogena strata med avseende på stratifieringsvariabeln
  - Välj stratifieringsvariabel som har ett samband med undersökningsvariabeln (som är korrelerad med undersökningsvariabeln)
  - Dra slumpmässigt urval ur varje stratum

## Förra gången (F12)

---

- Systematiskt urval
  - Beräkna  $r=N/n$  och avrunda nedåt
  - Välj slumpmässigt ett tal mellan 1 och  $r$
  - Välj sedan vart  $r$ :te tal
- Gruppurval
  - Dra ett slumpmässigt urval av grupper av element
  - Gör sedan en totalundersökning i varje grupp/kluster, alt. Dra ett urval ur varje grupp

## OSU-formler (OSU utan återläggning)

---

- Vill skatta medelvärde,  $\mu$

$$\bar{x} = \frac{\sum_{i=1}^n x_i}{n} \quad \hat{V}(\bar{X}) = \left(1 - \frac{n}{N}\right) \frac{s^2}{n}$$

- Vill skatta andel,  $P$

$$p = \frac{\sum_{i=1}^n x_i}{n} \quad \text{där } x\text{-värdena antar värdena } 0 \text{ eller } 1 \quad \hat{V}(p) = \left(1 - \frac{n}{N}\right) \frac{p(1-p)}{n-1}$$

- Vill skatta total,  $\tau$

$$\hat{\tau} = N \times \bar{x} \quad \hat{V}(\hat{\tau}) = N^2 \times \hat{V}(\bar{X})$$

## Stratifierat urval-formler (med OSU utan återläggning i varje stratum)

- Vill skatta medelvärde,  $\mu$

$$\bar{x}_{str} = W_1 \times \bar{x}_1 + W_2 \times \bar{x}_2 + \dots + W_L \times \bar{x}_L \quad \text{där } W_i \text{ är vikter och } L \text{ är antal strata}$$

$$\hat{V}(\bar{X}_{str}) = W_1^2 \left(1 - \frac{n_1}{N_1}\right) \frac{s_1^2}{n_1} + W_2^2 \left(1 - \frac{n_2}{N_2}\right) \frac{s_2^2}{n_2} + \dots + W_L^2 \left(1 - \frac{n_L}{N_L}\right) \frac{s_L^2}{n_L}$$

- Vill skatta andel, P

$$p_{str} = W_1 \times p_1 + W_2 \times p_2 + \dots + W_L \times p_L$$

$$\hat{V}(p) = \sum_{i=1}^L W_i^2 \left(1 - \frac{n_i}{N_i}\right) \frac{p_i(1-p_i)}{n_i - 1} \quad \text{där x-värdena antar värdena 0 eller 1}$$

- Vill skatta total,  $\tau$

$$\hat{\tau}_{str} = N \times \bar{x}_{str} \quad \hat{V}(\hat{\tau}_{str}) = N^2 \times \hat{V}(\bar{X}_{str})$$

## Konfidensintervall

- Bilda k.i. som skattningen +/- 1,96 gånger standardavvikelsen för skattningen.
  - 1,96 fås från tabell över normalfördelningen
  - Om vi önskar andra konfidensgrader kan de också fås från tabell över normalfördelning
    - 1,64 motsvarar ett 90%igt k.i.
    - 2,58 motsvarar ett 99%igt k.i.

## Samband mellan variabler

---

- X och Y på intervall/kvotskalenivå
  - $r_{XY}$
  - Mäter styrkan av det linjära sambandet
    - $-1 \leq r_{XY} \leq 1$
  - $R^2$ =andelen förklarad variation
    - Determinationskoefficienten
- X och Y på ordinalskalenivå
  - $r_s$
  - Rangkorrelationskoefficient
    - $-1 \leq r_s \leq 1$

## Samband mellan variabler

---

- X och Y på nominalskalenivå
  - Chi-tvåtest
  - Ex. finns det något samband mellan kön och åsikt om privat barnomsorg?
  - Skiljer sig åldersfördelningen mellan könen?

## Hypoteser

---

- $H_0$ : Inget samband mellan kön och åsikt
  - Oberoende mellan kön och åsikt
  - Åsiktsfördelningen är samma för män och kvinnor
- $H_1$ : Det finns ett samband mellan kön och åsikt
  - Det råder beroende mellan kön och åsikt
  - Åsiktsfördelningen är ej samma för män och kvinnor

## Hypotesprövning

---

- Vid hypotesprövning ställer vi upp en noll-hypotes ( $H_0$ ) och en alternativ hypotes ( $H_1$ ).
- Vi samlar sedan in data och testar nollhypotesen.
- Låt oss testa nollhypotesen att det *inte* finns något samband mellan kön och åsikt.

## Korstabell

	Kvinnor	Män	
Pos	8	13	21
Neg	12	7	19
Vet ej	4	6	10
	24	26	50

Finns det något samband mellan kön och åsikt?

## Korstabell

	Kvinnor	Män	
Pos	8 33%	13 50%	21 42%
Neg	12 50%	7 27%	19 38%
Vet ej	4 17%	6 23%	10 20%
	24 100%	26 100%	50 100%

Finns det något samband mellan kön och åsikt?

## Korstabell

	Kvinnor		Män		
Pos	8 33%	10,1 42%	13 50%	10,9 42%	21 42%
Neg	12 50%	9,1 38%	7 27%	9,9 38%	19 38%
Vet ej	4 17%	4,8 20%	6 23%	5,2 20%	10 20%
	24 100%		26 100%		50 100%

Jämför observerade frekvenser med förväntade frekvenser. Om det *inte* finns något samband mellan kön och åsikt förväntar vi oss att 42% av kvinnorna, dvs 10,1 samt 42% av männen, dvs 10,9 svarat positivt. Men våra *observerade* värden är 8 resp. 13. Skiljer sig dessa så mycket åt att vi tror att det finns ett samband?

## Chi-tvåtest

- $O_i$ =observerade frekvenser
- $E_i$ =förväntade frekvenser
- Testvariabel:

$$\chi^2 = \sum_{i=1}^k \frac{(O_i - E_i)^2}{E_i}$$

- k=antal celler i korstabellen

$$\begin{aligned} \chi_{obs}^2 &= \frac{(8-10,1)^2}{10,1} + \frac{(13-10,9)^2}{10,9} + \frac{(12-9,1)^2}{9,1} + \\ &+ \frac{(7-9,9)^2}{9,9} + \frac{(4-4,8)^2}{4,8} + \frac{(6-5,2)^2}{5,2} = 2,86 \end{aligned}$$

## Chi-tvåtest

- Är våra *observerade* värden tillräckligt annorlunda än de *förväntade* värdena – så som vi skulle förvänta oss om  $H_0$  är sann?
- Om  $H_0$  är sann så följer chitvå-måttet en chi-två fördelning med  $(r-1) \times (c-1)$  frihetsgrader (där  $r$  är antal rader och  $c$  är antal kolumner)
  - Test:
  - Ställ upp nollhypotes och alternativ hypotes
  - Välj signifikansnivå (vanligtvis 5%)
  - Om  $\chi^2_{\text{obs}} > \chi^2_{\text{krit}} \longrightarrow$  förkasta  $H_0$  (på 5% sign.nivå)
    - Vi tror inte att våra avvikelser kan ha uppkommit av slumpen (pga att vi dragit ett stickprov)
  - Om  $\chi^2_{\text{obs}} < \chi^2_{\text{krit}} \longrightarrow$  förkasta ej  $H_0$  (på 5% sign.nivå)
    - Våra avvikelser kan ha uppkommit av slumpen (pga att vi dragit ett stickprov)

## Chi-tvåtest: vårt exempel

- $\chi^2_{\text{obs}} = 2,86$
- $\chi^2_{\text{krit}} = 5,99$ 
  - $(3-1) \times (2-1) = 2$  frihetsgrader
  - Sign.nivå 5%
  - Kritiska värdet kan fås från tabell över chi-tvåfördelning
- $\chi^2_{\text{obs}} < \chi^2_{\text{krit}} \longrightarrow$  Förkasta ej  $H_0$  på 5% sign.nivå
  - Vi har ej stöd för att det finns ett samband mellan kön och åsikt
- P-värde (p-value=prob-value)
  - Sannolikheten att få ett värde på testvariabeln som är minst lika extremt som det observerade givet att  $H_0$  är sann
  - Om p-värdet är större än signifikansnivån kan vi ej förkasta  $H_0$

# Chi-tvåtest i Minitab

- Stat → Tables → Cross-tabulation and Chi-square

## Tabulated statistics: Åsikt; Kön

Rows: Åsikt Columns: Kön

	Kvinnor	Män	All
Neg	12 50,00	7 26,92	19 38,00
Pos	8 33,33	13 50,00	21 42,00
Vet ej	4 16,67	6 23,08	10 20,00
All	24 100,00	26 100,00	50 100,00

Cell Contents: Count  
% of Column

Pearson Chi-Square = 2,831; DF = 2; P-Value = 0,243  
Likelihood Ratio Chi-Square = 2,856; DF = 2; P-Value = 0,240

\* NOTE \* 1 cells with expected counts less than 5