

# Industriell matematik och statistik, LMA136 2013/14

Repetition

Statistikor

Hypotestest

Konfidensintervall

Styrka

26 Februari 2014

# Disposition

Föreläsning 10

Repetition

Statistikor

Hypotestest

Konfidensintervall

Styrka

Repetition

Statistikor

Hypotestest

Konfidensintervall

Styrka

## Repetition

## Statistikor

## Hypotestest

## Konfidensintervall

## Styrka

- ▶ Punktskattare;  $\bar{X} = \frac{1}{n} \sum_{i=1}^n X_i$ ,  
 $\hat{\sigma}^2 = \frac{1}{n-1} \sum_{i=1}^n (X_i - \bar{X})^2$ ,  $\hat{p} = \frac{X}{n}$
- ▶ Väntevärdesriktighet,  $\mathbb{E}[\hat{\theta}] = \theta$ .
- ▶ Effektivitet,  $\text{Var}(\hat{\theta})$ , önskar skattare med liten varians.
- ▶ Intervallskatting (Använda centrala gränsvärdessatsen på skattarna.) Anger ett intervall som  $\theta$  ligger inom med viss sannolikhet.
  - ▶ Statistika  $Z$  (eller  $T$ ) har fördelning  $F$  ( $Z$  en funktion av  $\hat{\theta}$  och  $\theta$ )
  - ▶  $q_{1-\alpha/2}$  är  $(1 - \alpha/2)$ -kvantil för fördelningen  $F$ .  
 $P(-q_{1-\alpha/2} < Z < q_{1-\alpha/2}) = 1 - \alpha$
  - ▶ Lös ut den sökta parametern ur olikheterna
  - ▶  $(1-\alpha)$ -konfidensintervallet för  $\theta$  definieras av det resulterande intervallet.

# Konfidsensintervall

- Fördelningar; Normalfördelningen, Student-t fördelning,  $\chi^2$ -fördelning
- Om  $F$  = standardiserad normalfördelning

| $q_{1-0.10/2}$ | $q_{1-0.05/2}$ | $q_{1-0.01/2}$ |
|----------------|----------------|----------------|
| 1.64           | 1.96           | 2.58           |

- Om parametern är normalfördelad men variansen för parametern är okänd så gäller att  $F$  = Student-t fördelning med  $n - 1$  frihetsgrader.

| Antal observationer | $q_{1-0.10/2}$ | $q_{1-0.05/2}$ | $q_{1-0.01/2}$ |
|---------------------|----------------|----------------|----------------|
| $n = 5$             | 2.13           | 2.78           | 4.60           |
| $n = 10$            | 1.83           | 2.26           | 3.25           |
| $n = 100$           | 1.66           | 1.98           | 2.63           |

- $(1-\alpha)$ -konfidsensintervall för väntevärden ges av

$$(\bar{x} - q_{1-\alpha/2}S, \bar{x} + q_{1-\alpha/2}S)$$

då  $S = \sigma/\sqrt{n}$  om standardavvikelsen känd och  $S = s/\sqrt{n}$  då  $\sigma$  okänd.

Repetition

Statistikor

Hypotestest

Konfidsensintervall

Styrka

- Väntevärdesstatistikor; ersätt  $\sigma^2$  med skattaren  $\hat{\sigma}^2$  i uttrycket för

statistikan om  $\sigma$  okänd.

| Statistika                                | Fördelning | Tumregler                  |
|-------------------------------------------|------------|----------------------------|
| $\frac{\bar{X} - \mu}{\sigma / \sqrt{n}}$ | $N(0, 1)$  | $\sigma$ känd              |
| $\vdots$                                  | $t(n - 1)$ | $\sigma$ okänd, skatta $s$ |

- Väntevärdesdifferensstatistikor;

$\Delta(X, Y) = \bar{X} - \bar{Y}$ ,  $S(\sigma_X, \sigma_Y) = \sqrt{\text{Var}(\Delta(X, Y))}$ , ersätt  $\sigma_X, \sigma_Y$  i uttrycket för

statistikan med skattarna av respektive standardavvikelse om de är okända (ihopviktade

skattaren om  $\sigma_X = \sigma_Y$ ).

| Statistika                                               | Fördelning         | Tumregler                                                                                                                                    |
|----------------------------------------------------------|--------------------|----------------------------------------------------------------------------------------------------------------------------------------------|
| $\frac{\Delta - (\mu_X - \mu_Y)}{S(\sigma_X, \sigma_Y)}$ | $N(0, 1)$          | $\sigma_X, \sigma_Y$ kända                                                                                                                   |
| $\vdots$                                                 | $N(0, 1)$          | $n_X, n_Y$ stora                                                                                                                             |
| $\vdots$                                                 | $t(n_X + n_Y - 2)$ | $\sigma_X = \sigma_Y = \sigma$                                                                                                               |
| $\vdots$                                                 | $t(df)$            | $df \approx \frac{(\frac{s_X^2}{n_X} + \frac{s_Y^2}{n_Y})^2}{\frac{(\frac{s_X^2}{n_X})^2}{n_X - 1} + \frac{(\frac{s_Y^2}{n_Y})^2}{n_Y - 1}}$ |

Repetition

Statistikor

Hypotestest

Konfidensintervall

Styrka

Repetition

**Statistikor**

Hypotestest

Konfidensintervall

Styrka

$$\blacktriangleright \hat{p} = \frac{X}{n}, \hat{\sigma}^2 = \frac{\sum_{i=1}^n (X_i - \bar{X})^2}{n-1}$$

| Storhet           | Statistika                                                | Fördelning    |
|-------------------|-----------------------------------------------------------|---------------|
| Andelsparameter   | $\frac{\hat{p} - p}{\sqrt{\frac{\hat{p}(1-\hat{p})}{n}}}$ | $N(0, 1)$     |
| Standardavvikelse | $\frac{(n-1)\hat{\sigma}^2}{\sigma^2}$                    | $\chi^2(n-1)$ |

Objektivt testa om antagande passar

- ▶ Hypoteser;  $H_0, H_1$
- ▶ Signifikansnivå;  $\alpha = P(\text{Förkasta } H_0 \text{ sann} | H_0 \text{ sann})$
- ▶ Testvariabel med fördelning under nollhypotesen samt testregel

## Exempel

Antag att längden för en slumpmässigt vald manlig professionell handbollsspelare är normalfördelad med standardavvikelse  $\sigma = 10$  och okänt väntevärde.

Uppgift: Undersök om den förväntade längden för en handbollsspelare är längre än medlet för svenska män på signifikansnivån 5% och ett urval av  $n = 100$

handbollsspelare där  $\bar{x} = 193.2\text{cm}$ . (medellängd för män i Sverige 181.5cm år 2013, källa: centrum för pediatrik växtforskning)..

- ▶  $H_0 : \mu = 181.5$ , Mothypotes  $H_1 : \mu > 181.5$
- ▶ Signifikansnivå,  
 $\alpha = P(\text{Förkasta } H_0 \text{ sann} | H_0 \text{ sann}) = 0.05$ ,
- ▶ Testvariabel:  $Z = \frac{\bar{X} - \mu}{\sigma/\sqrt{n}} \sim N(0, 1)$  givet  $H_0$   
 $P(Z > 1.64) = \alpha = 0.05$ ,  
 Testregel: Förkasta  $H_0$  om  $Z_{Obs} > 1.64$ .
- ▶  $\bar{x} = 193.2\text{cm}$ ,  $Z_{Obs} = \frac{\bar{x} - \mu}{\sigma/\sqrt{n}} = \frac{193.2 - 181.5}{10/\sqrt{100}} = 11.7$
- ▶  $Z_{Obs} = 11.7 > 1.64$  därav  $Z_{Obs}$  i förkastelseområdet och  $H_0$  förkastas.

Repetition

Statistikor

Hypotestest

Konfidsintervall

Styrka

# Hypotestest

5 st delsteg;

- ▶ Definiera nollhypotes,  $H_0$ , samt mothypotes,  $H_1$ .
- ▶ Definiera signifikansnivån  
 $\alpha = P(\text{Förkasta } H_0 \text{ sann} | H_0 \text{ sann})$
- ▶ Definiera testvariabeln (statistikan)  $Z, T$ , med tillhörande kvantiler  $q$  för förkastelseområdet.
- ▶ Beräkna observerade testvariabeln  $Z_{Obs}, T_{Obs}$
- ▶ Förkasta  $H_0$  (och acceptera  $H_1$ ) om testvariabeln inom förkastelseområdet, annars acceptera  $H_0$ .

Testets p-värde (beror av mothypotesen)

| Mothypotes             | p-värde                               |
|------------------------|---------------------------------------|
| $H_1 : \mu \neq \mu_0$ | $P( Z  > Z_{Obs}   H_0 \text{ sann})$ |
| $H_1 : \mu > \mu_0$    | $P(Z > Z_{Obs}   H_0 \text{ sann})$   |
| $H_1 : \mu < \mu_0$    | $P(Z < Z_{Obs}   H_0 \text{ sann})$   |

Repetition

Statistikor

Hypotestest

Konfidensintervall

Styrka

## Exempel

Antal kunder inne på akademibokhandeln i femmanhuset kl1200 varje dag kan modelleras som en Poissonprocess med intensitet  $\lambda$ .

Undersök om intensiteten  $\lambda = 10$  kan accepteras på signifikansnivån 5% från en månads samling data,  $x = \{x_1, x_2, \dots, x_{22}\}$  med  $\bar{x} = 7.5$ .

►  $H_0 : \lambda = 10, H_1 : \lambda \neq 10$

►  $\alpha = 0.05$

►  $Z = \frac{\bar{X} - \lambda}{\sqrt{\lambda/n}} \sim N(0, 1)$  givet  $H_0$

$P(|Z| > 1.96) = 0.05$ , Testregel: Förkasta  $H_0$  om  $|Z_{Obs}| > 1.96$ .

►  $\bar{x} = 7.5$  samt  $n = 22$  ger

$$Z_{Obs} = \frac{\bar{x} - \lambda}{\sqrt{\lambda/n}} = \frac{7.5 - 10}{\sqrt{10/22}} = -3.7$$

►  $|Z_{Obs}| > 1.96$  dvs i förkastelseområdet och  $H_0$  förkastas på signifikansnivå 5%.

Repetition

Statistikor

Hypotestest

Konfidsintervall

Styrka

Beräkna ett  $(1-\alpha)$ -konfidensintervall för  $\mu$  baserat på data  $x$ .

Hypotestest med  $H_0 : \mu = \mu_0$ ,  $H_1 : \mu \neq \mu_0$

Förkastar  $H_0$  på signifikansnivå  $\alpha$  om konfidensintervallet inte innehåller  $\mu_0$ .

## Exempel

Antag att längden för en slumpmässigt vald manlig trädgårdsarbetare är normalfördelad med okända parametrar  $\mu, \sigma$ .

Uppgift: Undersök om den förväntade längden för en trädgårdsarbetare är lägre än medlet för svenska män på signifikansnivån 5% och ett urval av  $n = 10$  trädgårdsarbetare där  $\bar{x} = 179.1\text{cm}$  och  $s = 9.3$ . (medellängd för män i Sverige 181.5cm.)

►  $H_0 : \mu = 181.5$ , Mothypotes  $H_1 : \mu < 181.5$

► Signifikansnivå,

$$\alpha = P(\text{Förkasta } H_0 \mid H_0 \text{ sann}) = 0.05,$$

► Testvariabel:  $T = \frac{\bar{X} - \mu}{\sqrt{\hat{\sigma}^2/n}} \sim t(n-1)$  givet  $H_0$

$$P(T < -1.83) = \alpha,$$

Testregel: Förkasta  $H_0$  om  $T_{Obs} < -1.83$ .

►  $\bar{x} = 179.1\text{cm}$ ,  $T_{Obs} = \frac{\bar{x} - \mu}{s/\sqrt{n}} = \frac{179.1 - 181.5}{9.3/\sqrt{10}} = -0.78$

►  $T_{Obs} = -0.78 > -1.83$  därav  $T_{Obs}$  ej i förkastelseområdet och  $H_0$  accepteras.

Repetition

Statistikor

Hypotestest

Konfidensintervall

Styrka

Styrka  $\beta$ .

|                 | $H_0$ sann   | $H_1$ sann    |
|-----------------|--------------|---------------|
| Tror $H_0$ sann | Ok           | fel av typ II |
| Tror $H_1$ sann | fel av typ I | Ok            |

$P(\text{Förkasta } H_0 \text{ sann} | H_0 \text{ sann}) = \alpha$      $(1-\alpha) = \text{testets konfidensgrad}$

$P(\text{Tror } H_0 \text{ sann} | H_1 \text{ sann}) = \beta$      $(1-\beta) = \text{testets styrka}$

| Domstol | Skyldig      | Icke skyldig  |
|---------|--------------|---------------|
| Döms    | Ok           | fel av typ II |
| Frias   | fel av typ I | Ok            |