

Proportions test

Kom ihåg: F14 Vi vill skatta proportionen p i en population med en viss egenskap.

Stickprovet $X = \#$ utvalda med egenskapen och $X \sim \text{Bin}(n, p)$ där $n = \text{stickprovets storlek}$.

Vi använder (punkt-)skattningen $\hat{p} = \frac{X}{n}$.

F18 Vi vill testa (t.ex.)

$$H_0: p = p_0$$

$$H_1: p \neq p_0$$

Vi använde att $\hat{p} \approx N(np_0, np_0(1-p_0))$

om $p = p_0$ och n stort (så att normalapprox är ok). Här var $\hat{p}$ test-statistikan.

Man kunde lika gärna ha använt

$$\frac{\hat{p} - np_0}{\sqrt{np_0(1-p_0)}} \approx N(0, 1).$$

Hur gör vi om n är litet?

Tal: Siershan Saida gissar färgen på 20 kort som dras slumpmässigt med återläggning. Om

$X = \#$ kort som gissas rätt $\sim \text{Bin}(20, p)$.

Hitta RR (förlastningsregion) för testet $H_0: p = \frac{1}{4}$

$$H_1: p > \frac{1}{4}$$

på signifikansnivån $\alpha = 5\%$, och skissa teststyrkan som funktion av p .

L: Vi använder test-statistikan $\bar{X}$.
RR kommer ha formen $[c, 20]$. c bestäms av sambandet

$$\begin{aligned} \alpha = 0.05 &= P(\text{förlasta } H_0 \mid H_0 \text{ sann}) \\ &= P(\bar{X} \geq c) \text{ om } \bar{X} \sim \text{Bin}(20, \frac{1}{4}). \end{aligned}$$

Räkningar ger att

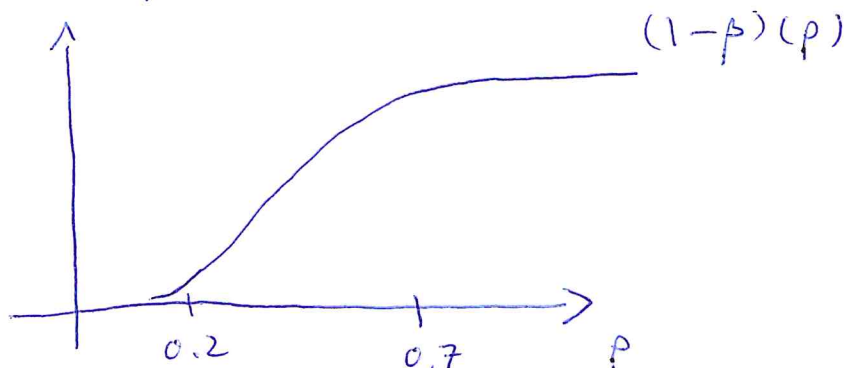
$$P(\bar{X} \geq 8) = 0.101 \text{ och } P(\bar{X} \geq 9) = 0.041.$$

Vi väljer därför $RR = [9, 20]$.

Teststyrkan $1 - \beta = P(\text{förlasta } H_0 \mid H_1 \text{ sann})$
 $= P(\bar{X} \geq 9) \text{ om } \bar{X} \sim \text{Bin}(20, p).$

Tabell:

p	0.27	0.3	0.35	0.4	0.45
$1 - \beta$	0.064	0.113	0.238	0.404	0.586



Anm: 1/ Om Saida är magisk med $p=0.4$ upptäcks detta med sann. 0.404.

2/ Om 1000 personer gör testet på $\alpha=0.05$ nivån kommer antalet som av ren slump får över 9 rätt vara

$\text{Bin}(1000, 0.041)$ - fördelat.

"fishing"

Konfidsensintervall (för p):

Vi antar att n är stort så att

$$\bar{X} \approx N(\mathbb{E}[\bar{X}], \text{Var}(\bar{X})) = N(np, np(1-p))$$

$$\Rightarrow \frac{\bar{X} - np}{\sqrt{np(1-p)}} \approx N(0, 1)$$

Steg 1: Använd referensvariabeln

$$\frac{\bar{X} - np}{\sqrt{np(1-p)}} \approx N(0, 1)$$

Steg 2: Givet signifikansnivån α , hitta kvantilerna $z_{\alpha/2}$ och $-z_{\alpha/2}$.

Steg 3: Lös ut p ur $-z_{\alpha/2} \leq \frac{\bar{X} - np}{\sqrt{np(1-p)}} \leq z_{\alpha/2}$

⌈ Kom ihåg: $1 - \alpha \approx P\left(-z_{\alpha/2} \leq \frac{\bar{X} - np}{\sqrt{np(1-p)}} \leq z_{\alpha/2}\right)$ ⌋

Detta är lite bråligt så

Steg 3': Hitta ett K.I. för p ur villkoret

$$\mathbb{P}\left(-z_{\alpha/2} \leq \frac{\bar{X} - np}{\sqrt{n\hat{p}(1-\hat{p})}} \leq z_{\alpha/2}\right)$$

$\approx$
 $\uparrow$
 hoppas vi

$$\mathbb{P}\left(-z_{\alpha/2} \leq \frac{\bar{X} - np}{\sqrt{np(1-p)}} \leq z_{\alpha/2}\right).$$

Vi får

$$\begin{aligned} & \mathbb{P}\left(-z_{\alpha/2} \leq \frac{\bar{X} - np}{\sqrt{n\hat{p}(1-\hat{p})}} \leq z_{\alpha/2}\right) \\ &= \mathbb{P}\left(-z_{\alpha/2} \sqrt{n\hat{p}(1-\hat{p})} \leq \bar{X} - np \leq z_{\alpha/2} \sqrt{n\hat{p}(1-\hat{p})}\right) \\ &= \mathbb{P}\left(\frac{\bar{X}}{n} - z_{\alpha/2} \sqrt{\frac{\hat{p}(1-\hat{p})}{n}} \leq p \leq \frac{\bar{X}}{n} + z_{\alpha/2} \sqrt{\frac{\hat{p}(1-\hat{p})}{n}}\right). \end{aligned}$$

Så ett $(1-\alpha)100\%$ K.I. för p blir

$$I_p = \hat{p} \pm z_{\alpha/2} \sqrt{\frac{\hat{p}(1-\hat{p})}{n}}$$

- Anm: 1/ Man kan kolla $n\hat{p}$ och $n(1-\hat{p})$ för att motivera CLT. (normalapprox).
- 2/ För små n kan man använda kvantiler för binomialfördelningen.

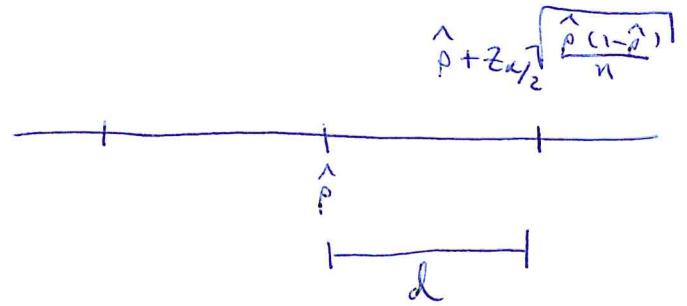
Stickprovets storlek

Bredden $2d$ på K.I. ovan ges av

$$2d = 2 z_{\alpha/2} \sqrt{\frac{\hat{p}(1-\hat{p})}{n}}$$

Givet en föreskriven bredd d ,
hur väljer vi n ?

$$n = \frac{z_{\alpha/2}^2 \hat{p}(1-\hat{p})}{d^2}$$



Detta kräver dock kännedom om $\hat{p}$.
Om en sådan inte finns kan vi använda
att $\hat{p}(1-\hat{p}) \leq \frac{1}{4}$ (övn 9.9). Det fungerar
därför alltid att ta

$$n = \frac{z_{\alpha/2}^2}{4d^2}.$$

Jämförelse av två proportioner

Tal: Två väljarundersökningar genomfördes
med 4 månaders mellanrum. Resultatet för
parti Q blev som följer:

- 1/ 143 av $n_1 = 1217$ tillfrågade stöttar Q
- 2/ 139 " $n_2 = 1078$ " " "

a) Ange en punktskattning för $p_1 - p_2$.

L: $\hat{p}_1 = \frac{\bar{X}_1}{n_1}$ och $\hat{p}_2 = \frac{\bar{X}_2}{n_2}$ där

$\bar{X}_k = \#$ anhängare av Q vid undersökning k .

T OBS: $\bar{X}_k$ är en s.v. "innan-perspektiv"

Vi har att $\bar{X}_k \sim \text{Bin}(n_k, p_k)$. Vi låter

$$\widehat{p_1 - p_2} = \hat{p}_1 - \hat{p}_2 \text{ och observerar att}$$

$$\mathbb{E}[\hat{p}_1 - \hat{p}_2] = \mathbb{E}[\hat{p}_1] - \mathbb{E}[\hat{p}_2] = p_1 - p_2 \text{ s. a.}$$

$\widehat{p_1 - p_2}$ är VVR. Data ger

$$\hat{p}_1(x) - \hat{p}_2(x) = \frac{143}{1217} - \frac{139}{1078} \approx -0.0114$$

b) Bestäm ett $(1-\alpha)100\%$ K.I. för $p_1 - p_2$.

1: Vi har att (normalapprox)

$$\hat{p}_k \approx N(\mathbb{E}[\hat{p}_k], \text{Var}(\hat{p}_k)) = N(p_k, \frac{p_k(1-p_k)}{n_k})$$

för $k=1, 2$. $\Rightarrow$

$$\hat{p}_1 - \hat{p}_2 \approx N\left(p_1 - p_2, \frac{p_1(1-p_1)}{n_1} + \frac{p_2(1-p_2)}{n_2}\right) \quad \text{ty}$$

$$\mathbb{E}[\hat{p}_1 - \hat{p}_2] = p_1 - p_2, \quad \text{Var}(\hat{p}_1 - \hat{p}_2) \underset{\substack{\uparrow \\ \text{ober.}}}{=} \text{Var}(\hat{p}_1) + \text{Var}(\hat{p}_2).$$

För att komma vidare ersätter vi

$\text{Var}(\hat{p}_1 - \hat{p}_2)$ med dess skattning

$$s_{\hat{p}_1 - \hat{p}_2}^2 = \frac{\hat{p}_1(1-\hat{p}_1)}{n_1} + \frac{\hat{p}_2(1-\hat{p}_2)}{n_2}.$$

"Som vanligt" där $Z \sim N(0, 1)$

$$1-\alpha = \mathbb{P}(-z_{\alpha/2} \leq Z \leq z_{\alpha/2})$$

$$\approx \mathbb{P}\left(-z_{\alpha/2} \leq \frac{\hat{p}_1 - \hat{p}_2 - (p_1 - p_2)}{s_{\hat{p}_1 - \hat{p}_2}} \leq z_{\alpha/2}\right)$$

$$= P(\hat{p}_1 - \hat{p}_2 - z_{\alpha/2} S_{\hat{p}_1 - \hat{p}_2} \leq p_1 - p_2 \leq \hat{p}_1 - \hat{p}_2 + z_{\alpha/2} S_{\hat{p}_1 - \hat{p}_2}).$$

Enligt $(1-\alpha)100\%$ K.I. för $p_1 - p_2$ blir

alltså

$$I_{p_1 - p_2} = \hat{p}_1 - \hat{p}_2 \pm z_{\alpha/2} \sqrt{\frac{\hat{p}_1(1-\hat{p}_1)}{n_1} + \frac{\hat{p}_2(1-\hat{p}_2)}{n_2}} //$$

c) Vad blir det numeriska K.I. om $\alpha=0.01$?

2: Vi har att $z_{0.005} \approx 2.575$ och att $\hat{p}_1 \approx 0.1175$, $\hat{p}_2 \approx 0.1289$ s.a. det numeriska

K.I. blir

$$I_{p_1 - p_2} \approx -0.0114 \pm 2.575 \cdot 0.0138 \\ \approx [-0.047, 0.024]. //$$

Ans: Har vi fog att hävda att sympatierna har ändrats (dvs $p_1 \neq p_2$)?

12om ihåg: $p_1 - p_2$ skattas av

$$\hat{p}_1 - \hat{p}_2 \approx N\left(p_1 - p_2, \frac{p_1(1-p_1)}{n_1} + \frac{p_2(1-p_2)}{n_2}\right)$$

där $\hat{p}_1 = \frac{\bar{X}_1}{n_1}$ resp. $\hat{p}_2 = \frac{\bar{X}_2}{n_2}$ där $\bar{X}_k \sim \text{Bin}(n_k, p_k)$

Test av två proportioner

Vi är intresserade av följande test

I $H_0: p_1 = p_2$ II $H_0: p_1 = p_2$ III $H_0: p_1 = p_2$

$H_1: p_1 > p_2$ $H_1: p_1 < p_2$ $H_1: p_1 \neq p_2$

Anm: Boken tar upp $H_0: p_1 - p_2 = (p_1 - p_2)_0$.

Vi skippar detta,

Under H_0 så är $p_1 = p_2 = p$. Därför blir

$$\frac{p_1(1-p_1)}{n_1} + \frac{p_2(1-p_2)}{n_2} = p(1-p)\left(\frac{1}{n_1} + \frac{1}{n_2}\right).$$

Vi skattar p med

$$\hat{p} = \frac{\bar{X}_1 + \bar{X}_2}{n_1 + n_2} = \frac{n_1 \hat{p}_1 + n_2 \hat{p}_2}{n_1 + n_2} \quad \text{så att}$$

$$\hat{p}_1 - \hat{p}_2 \approx N\left(0, p(1-p)\left(\frac{1}{n_1} + \frac{1}{n_2}\right)\right)$$

$$\approx N\left(0, \hat{p}(1-\hat{p})\left(\frac{1}{n_1} + \frac{1}{n_2}\right)\right)$$

Vår test-statistika blir därför

$$\frac{\hat{p}_1 - \hat{p}_2}{\sqrt{\hat{p}(1-\hat{p})\left(\frac{1}{n_1} + \frac{1}{n_2}\right)}} \approx N(0, 1).$$

Tal (forts): 12 om ihåg

Tillfälle 1: 143 av $n_1 = 1217$

" 2: 139 av $n_2 = 1078$

Testa $H_0: p_1 = p_2$

$H_1: p_1 < p_2$

på signifikansnivån $\alpha = 0.05$,

L: Vi beräknar p-värdet för att se om detta understiger 0.05,

Vi använder test-statistikan

$$T = \frac{\hat{p}_1 - \hat{p}_2}{\sqrt{\hat{p}(1-\hat{p})\left(\frac{1}{n_1} + \frac{1}{n_2}\right)}} \approx N(0,1) \quad (\text{om } H_0 \text{ sann})$$

Om H_1 sann $\Rightarrow T$ bör vara negativ,

$$\begin{aligned} p\text{-värde} &= P(T(X) \leq T(x) \mid H_0 \text{ sann}) \\ &= P(Z \leq T(x)) \quad \text{där } Z \sim N(0,1), \end{aligned}$$

Data ger $\hat{p}_1 \approx 0.1175$, $\hat{p}_2 \approx 0.1289$ och $\hat{p} \approx 0.1229$. Vidare blir

$$\sqrt{\hat{p}(1-\hat{p})\left(\frac{1}{n_1} + \frac{1}{n_2}\right)} \approx 0.0137$$

$$\text{s. a. } T(x) \approx -0.8332,$$

$$p\text{-värde} = P(Z \leq -0.8332) \approx 0.203 > 0.05 = \alpha,$$

Vi kan alltså ej förkasta H_0 på signifikansnivån $\alpha = 0.05$

Två stickprov (normalförd):

Vi har två oberoende stickprov

$\bar{X} = (X_1, \dots, X_n)$ och $\bar{Y} = (Y_1, \dots, Y_m)$ där

$X_k \sim N(\mu_1, \sigma_1^2)$ och $Y_k \sim N(\mu_2, \sigma_2^2)$.

Vi vill jämföra μ_1 och μ_2 och inför

därför $\theta = \mu_1 - \mu_2$ som skattas mha

$$\hat{\theta} = \bar{X} - \bar{Y}.$$

Enligt tidigare föreläsningar har vi att

$$\bar{X} \sim N(\mu_1, \frac{\sigma_1^2}{n}), \quad \bar{Y} \sim N(\mu_2, \frac{\sigma_2^2}{m})$$

Detta ger att

$$\bar{X} - \bar{Y} \sim N(\mu_1 - \mu_2, \frac{\sigma_1^2}{n} + \frac{\sigma_2^2}{m}).$$

Fall 1 (σ_1, σ_2 kända): Låt $D = \frac{\sigma_1^2}{n} + \frac{\sigma_2^2}{m}$

$$\Rightarrow R = \frac{\bar{X} - \bar{Y} - \theta}{D} \sim N(0, 1)$$

$$\Rightarrow 1 - \alpha = P(-z_{\alpha/2} \leq R \leq z_{\alpha/2})$$

$$= P(\bar{X} - \bar{Y} - z_{\alpha/2} D \leq \theta \leq \bar{X} - \bar{Y} + z_{\alpha/2} D)$$

$$\Rightarrow I_{\mu_1 - \mu_2} = \bar{X} - \bar{Y} \pm z_{\alpha/2} D$$

- Tal: I en undersökning uppskattades halten dioxin i två sjöar. Den första kontrollerades på $n=20$ platser och den andra på $m=29$. Resultaten anses vara oberoende och $N(\mu_1, \sigma_1^2)$ resp. $N(\mu_2, \sigma_2^2)$ -fördelade. Om $\sigma_1^2=1.4$ och $\sigma_2^2=1.9$, hitta ett 95% K.I. för $\mu_1 - \mu_2$ om data gav $\bar{x}=5.3$ och $\bar{y}=6.1$

L: Vi får att $D = \sqrt{\frac{\sigma_1^2}{n} + \frac{\sigma_2^2}{m}} \approx 0.368$.

Tabell ger $z_{0.025} \approx 1.96$ så ett 95% numeriskt K.I. blir

$$\begin{aligned} I_{\mu_1 - \mu_2} &= \bar{x} - \bar{y} \pm z_{0.025} D \approx -0.8 \pm 1.96 \cdot 0.368 \\ &\approx [-1.52, -0.079], // \end{aligned}$$

- Vi vill även kunna hantera situationen när σ_1^2 och σ_2^2 är okända. Man skiljer då på fallen $\sigma_1^2 = \sigma_2^2$ och $\sigma_1^2 \neq \sigma_2^2$.

För att avgöra vilket som är aktuellt så vill vi testa om $\sigma_1^2 = \sigma_2^2$.

Hypotestest av varianser

Här behöver vi bara betrakta två möjliga test

$$\text{I} \quad H_0: \sigma_1^2 = \sigma_2^2$$

$$\text{II} \quad H_0: \sigma_1^2 = \sigma_2^2$$

$$H_1: \sigma_1^2 > \sigma_2^2$$

$$H_1: \sigma_1^2 \neq \sigma_2^2$$

"Om $\sigma_2^2 > \sigma_1^2$ så byt namn på dom"

Vi använder $S_1^2 = \frac{1}{n-1} \sum_{k=1}^n (\bar{x}_k - \bar{x})^2$ för att

skatta σ_1^2 och liknande använder vi S_2^2

för att skatta σ_2^2 .

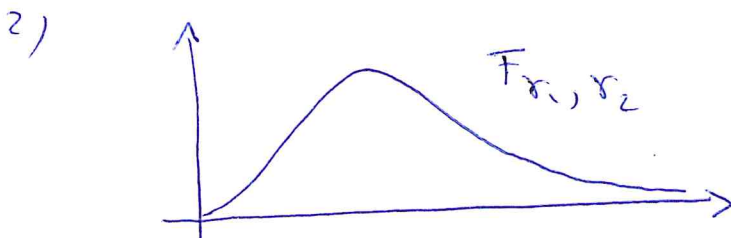
Vi har följande sats:

Sats: Om $\sigma_1^2 = \sigma_2^2$ så gäller att

$$\frac{S_1^2}{S_2^2} \sim F(n-1, m-1)$$

dvs F-fördelad med $n-1$ resp. $m-1$ frihetsgrader,

Anm: Vi använder tabell för att hitta sannolikheter för F-fördelningarna



Tal (forts): Låt data vara som i förra talet men antag nu att σ_1^2 och σ_2^2 är okända. Data gav $S_1^2 \hat{=} 1.7$, $S_2^2 \hat{=} 0.8$

Testa $H_0: \sigma_1^2 = \sigma_2^2$

$$H_1: \sigma_1^2 > \sigma_2^2$$

på signifikansnivån $\alpha = 0.05$.

L: Vi använder test-statistikan

$$T = \frac{S_1^2}{S_2^2} \sim F(n-1, m-1) = F(19, 28)$$

Tabell (sid 705) säger att

$$P(T > 1.694) \approx 0.1$$

Tabell (sid 711) säger att

$$P(T > 1.972) \approx 0.05.$$

I vårt fall har vi att $\frac{S_1^2(x)}{S_2^2(y)} \approx 2.125$

så att

$$p\text{-värdet} = P(T > 2.125) < 0.05.$$

Vi förkastar därför H_0 .

Tal: Testa $H_0: \sigma_1^2 = \sigma_2^2$

$$H_1: \sigma_1^2 \neq \sigma_2^2$$

på signifikansnivån $\alpha = 0.05$,

L: Man kan visa att

$$P\left(\frac{\max(S_1^2, S_2^2)}{\min(S_1^2, S_2^2)} > a\right) = 2 P(T > a) \quad \forall a > 0$$

med T som ovan.

Här fick vi att

$$\tilde{T}(x, y) = \frac{\max(s_1^2, s_2^2)}{\min(s_1^2, s_2^2)} \approx 2.125$$

så att

$$p\text{-värdet} = P(\tilde{T}(X, Y) > \tilde{T}(x, y))$$

$$= 2 \cdot P(T(X, Y) > \tilde{T}(x, y))$$

$$= 2 \cdot \underbrace{P(T(X, Y) > 2.125)} = ?$$

Vi kan g hitta $\rightarrow$ i tabell.

Antag att vi har två stickprov

$\underline{X} = (X_1, \dots, X_n)$ och $\underline{Y} = (Y_1, \dots, Y_m)$ där

$$X_k \sim N(\mu_1, \sigma^2) \text{ och } Y_k \sim N(\mu_2, \sigma^2)$$

(olika μ , samma σ^2). Enligt F20:

$$\bar{X} - \bar{Y} \sim N(\mu_1 - \mu_2, \sigma^2(\frac{1}{n} + \frac{1}{m})).$$

Hur skall σ^2 skattas?

$$s_{\underline{X}}^2 = \frac{1}{n-1} \sum_{k=1}^n (X_k - \bar{X})^2 \text{ och } s_{\underline{Y}}^2 = \frac{1}{m-1} \sum_{k=1}^m (Y_k - \bar{Y})^2$$

skattar båda σ^2 . Man kan visa att

$$S_p^2 = \frac{(n-1)s_{\underline{X}}^2 + (m-1)s_{\underline{Y}}^2}{n+m-2} \quad \text{"p för pooled"}$$

är VVR och har minst varians av

alla skattare på formen $\gamma s_{\underline{X}}^2 + (1-\gamma) s_{\underline{Y}}^2$.

K.I. för $\mu_1 - \mu_2$:

För att beräkna ett K.I. behöver vi en referensvariabel:

$$Z = \frac{(\bar{X} - \bar{Y}) - (\mu_1 - \mu_2)}{\sqrt{\sigma^2(\frac{1}{n} + \frac{1}{m})}} \sim N(0, 1)$$

men är bara användbar om σ^2 är känt.
Om σ^2 är okänt låter vi

$$T = \frac{(\bar{X} - \bar{Y}) - (\mu_1 - \mu_2)}{\sqrt{S_p^2(\frac{1}{n} + \frac{1}{m})}} \sim t(n+m-2)$$

Vi har sedan att

$$1 - \alpha = P\left(-t_{\alpha/2}(n+m-2) \leq \frac{(\bar{X} - \bar{Y}) - (\mu_1 - \mu_2)}{\sqrt{S_p^2 \left(\frac{1}{n} + \frac{1}{m}\right)}} \leq t_{\alpha/2}(n+m-2)\right)$$

så att ett $(1 - \alpha)100\%$ K.I. för $\mu_1 - \mu_2$ blir

$$I_{\mu_1 - \mu_2} = \bar{X} - \bar{Y} \pm t_{\alpha/2}(n+m-2) S_p(\bar{X}, \bar{Y}) \sqrt{\frac{1}{n} + \frac{1}{m}}$$

Tal (forts): Två sjöar, vi mäter dioxinhalten

$n=20$, $m=29$, $\hat{\mu}_1 = \bar{X} = 5.3$, $\hat{\mu}_2 = \bar{Y} = 6.1$. Vi antar nu att $\sigma_1^2 = \sigma_2^2 = \sigma^2$ är okänd. Data ger att

$$s_x^2 = 1.7 \quad \text{och} \quad s_y^2 = 0.8$$

Hitta ett 95% K.I. för $\mu_1 - \mu_2$.

1: Vi slår ihop skattningarna av σ^2 :

$$s_p^2 = \frac{(n-1)s_x^2 + (m-1)s_y^2}{n+m-2} = \frac{19 \cdot 1.7 + 28 \cdot 0.8}{20+29-2} \approx 1.164$$

Tabell sid 700 ger att $t_{0.025}^{(47)} \approx 2.012$.

Ett 95% numeriskt K.I. för $\mu_1 - \mu_2$ blir därtför

$$I_{\mu_1 - \mu_2} = \bar{X} - \bar{Y} \pm 2.012 \cdot \sqrt{1.164} \sqrt{\frac{1}{20} + \frac{1}{29}} \approx [-1.43, -0.17].$$

Hypotestest Vi har följande möjliga test:

I: $H_0: \mu_1 = \mu_2$ II: $H_0: \mu_1 = \mu_2$ III: $H_0: \mu_1 = \mu_2$

$H_1: \mu_1 > \mu_2$

$H_1: \mu_1 < \mu_2$

$H_1: \mu_1 \neq \mu_2$

Tal (forts): Vi vill testa huruvida

$\mu_1 \neq \mu_2$, Testa

om detta stämmer på $\alpha = 0.05$ nivå.

2: Vi antar inte längre att $\sigma_1^2 = \sigma_2^2$ så

vi får börja med att testa detta. Vi

väljer här $\alpha = 0.05$. Testet blir

$$H_0: \sigma_1^2 = \sigma_2^2$$

$$H_1: \sigma_1^2 \neq \sigma_2^2$$

Enligt förra föreläsningen använder vi
test-statistikan

$$\tilde{T}(\underline{X}, \underline{Y}) = \frac{\max(s_{\underline{X}}^2, s_{\underline{Y}}^2)}{\min(s_{\underline{X}}^2, s_{\underline{Y}}^2)}$$

och förkastar H_0 om

$$p\text{-värdet} = \mathbb{P}(\tilde{T}(\underline{X}, \underline{Y}) \geq \tilde{T}(\underline{x}, \underline{y}))$$

$$= 2 \mathbb{P}(T(\underline{X}, \underline{Y}) \geq \tilde{T}(\underline{x}, \underline{y})) \leq \alpha.$$

Här är $T(\underline{X}, \underline{Y}) = \frac{s_{\underline{X}}^2}{s_{\underline{Y}}^2} \sim F(n-1, m-1)$

och tabell (sid 711) gav att

$$p\text{-värde} = 2 \mathbb{P}(T(\underline{X}, \underline{Y}) \geq 2.125) \approx 2 \cdot 0.034 > 0.05$$

så vi köper att $\sigma_1^2 = \sigma_2^2$, dvs vi förkastar
ej H_0 .

När detta är avklarat så återvänder vi till

$$H_0: \mu_1 = \mu_2$$

$$H_1: \mu_1 \neq \mu_2$$

Vi använder test-statistikan

$$T = \frac{(\bar{X} - \bar{Y}) - (\mu_1 - \mu_2)}{\sqrt{s_p^2 \left(\frac{1}{n} + \frac{1}{m} \right)}} \sim t(n+m-2),$$

Vi beräknar p-värdet: (F18)

$$p\text{-värde} = 2P(T(X, Y) \geq |T(x, y)|)$$

där vi under H_0 har att

$$T(X, Y) = \frac{\bar{X} - \bar{Y}}{\sqrt{s_p^2 \left(\frac{1}{n} + \frac{1}{m} \right)}} \sim t(n+m-2)$$

så att

$$T(x, y) = \frac{\bar{x} - \bar{y}}{\sqrt{s_p^2(x, y) \left(\frac{1}{n} + \frac{1}{m} \right)}} \approx \frac{5.3 - 6.1}{\sqrt{1.164 \left(\frac{1}{20} + \frac{1}{29} \right)}} \approx -2.55,$$

Tabell (sid 700) ger att

$$\begin{aligned} p\text{-värde} &= 2P(T(X, Y) \geq | -2.55 |) = 2P(T(X, Y) \geq 2.55) \\ &= 2(1 - P(T(X, Y) \leq 2.55)) \approx 2(1 - 0.993) = 2 \cdot 0.007 \\ &= 0.014 \end{aligned}$$

så vi förkastar H_0 på signifikansnivån

$$\alpha = 0.05,$$

Stickprov i par

Dyker upp naturligt t.ex. när samma experiment upprepas vid olika tidpunkter eller genomförs med olika metoder,

Tal: Halten bly uppmäts i 20 vattendrag som ligger i anslutning till en nyetablerad fabrik. Mätning genomfördes innan och ett år efter att fabriken startades. Resultatet av mätningarna blev som följer:

Drag nr:	1	2	3	4	...	19	20
Innan:	1.7	1.6	2.7	4.3		0.7	2.9
Efter:	1.9	1.5	3.1	3.9		0.8	3.4

Under antagandet att mätningarna kommer från normalfördelningar, bestäm ett 99% K.I. för skillnaden i halten bly.

Anm: Låt μ_k = halten bly i vattendrag k innan fabriken startas.

Det är rimligt att förvänta sig skillnader mellan vattendragen, dvs $\mu_1 \neq \mu_2$
etc

Vi antar dock att skillnaden Δ innan/efter är densamma för alla vattendragen, dvs

Halt innan	Halt efter ett år
μ_1	$\mu_1 + \Delta$
μ_2	$\mu_2 + \Delta$
$\vdots$	$\vdots$
μ_{20}	$\mu_{20} + \Delta$

Vi vill skatta Δ !

Vad innebär normalantagandet?

Jo! $\bar{X}_k =$ uppmätt halt innan

$\bar{Y}_k =$ " " efter

$$\Rightarrow \bar{X}_k \sim N(\mu_k, \sigma_1^2) \text{ och } \bar{Y}_k \sim N(\mu_k + \Delta, \sigma_2^2)$$

och därmed blir $Z_k = \bar{Y}_k - \bar{X}_k \sim N(\Delta, \sigma_1^2 + \sigma_2^2)$

2: Vi skattar Δ med $\hat{\Delta} = \bar{Z}$ där

$$Z_k \sim N(\Delta, \sigma^2) \quad (\sigma^2 = \sigma_1^2 + \sigma_2^2), \text{ Vi får att}$$

$$R = \frac{\hat{\Delta} - \Delta}{S(Z)/\sqrt{n}} \sim t(n-1),$$

Data ger oss $\bar{Z} = 0.42$ och $S_Z = 0.32$ s.a.

$$I_{\Delta} = \bar{Z} \pm t_{0.005}^{(19)} \frac{S_Z}{\sqrt{201}} = 0.42 \pm 2.86 \frac{0.32}{\sqrt{201}} \approx \{0.215, 0.625\} //$$

Anm: Parat stickprov reduceras till ett stickprov!