Förra lektionen: Stickprov i par

$$\xi_k \in N(\mu_k, \sigma_1), k = 1, 2, \ldots, n \qquad \eta_k \in N(\mu_k + \Delta, \sigma_2), k = 1, 2, \ldots, n$$

Oberoende par (ξ_k, η_k). (Dock tillåtet med beroende mellan ξ_k och η_k.)

- Bilda $\zeta_k = \eta_k - \xi_k \in N(\Delta, \frac{\sigma}{\sqrt{n}})$, med observerade värden z_k.
- Räkna ut medelvärde $\bar{z}$ och skattad standardavvikelse s. (formelblad)
- Ett konfidensintervall för Δ med konfidensgrad $1-\alpha$ är $[\bar{z} - a, \bar{z} + a]$ med $a = t_{\alpha/2}(n-1)\frac{s}{\sqrt{n}}$. (För okänt σ.)

Nu: Två oberoende stickprov (jämförelse **sid 216**)

$$\xi_k \in N(\mu_1, \sigma), k = 1, 2, \ldots, n_1 \qquad \eta_k \in N(\mu_2, \sigma), k = 1, 2, \ldots, n_2$$

Alla variabler oberoende.
Skillnad: Samma fördelning för alla k, samma σ, tillåtet att $n_1 \neq n_2$.

Två oberoende stickprov, (fortsättning)

$$\xi_k \in N(\mu_1, \sigma), k = 1, 2, \ldots, n_1 \qquad \eta_k \in N(\mu_2, \sigma), k = 1, 2, \ldots, n_2$$

Alla variabler oberoende.

- Ger att $\bar{\xi} \in N(\mu_1, \frac{\sigma}{\sqrt{n_1}})$, $\bar{\eta} \in N(\mu_2, \frac{\sigma}{\sqrt{n_2}})$ och $\bar{\xi} - \bar{\eta} \in N\left(\mu_1 - \mu_2, \sigma\sqrt{\frac{1}{n_2} + \frac{1}{n_2}}\right)$.
- Från observerade stickprov ($x_k \leftrightarrow \xi_k$, $y_k \leftrightarrow \eta_k$), räkna ut $\bar{x}$, $\bar{y}$, s_x, s_y.
- Formel (8.6), **sid 214** ger bättre skattning av σ:

$$s^2 = {\sigma^2_{\text{obs}}}^* = \frac{(n_1 - 1)s_x^2 + (n_2 - 1)s_y^2}{n_1 - 1 + n_2 - 1} \qquad \left(= \frac{s_x^2 + s_y^2}{2} \text{om } n_1 = n_2 \right)$$

- **Sid 216:** Konfidensintervall för $\mu_1 - \mu_2$ med konfidensgrad $1 - \alpha$: $[\bar{x} - \bar{y} - a, \bar{x} - \bar{y} + a]$ med $a = t_{\alpha/2}(n_1 - 1 + n_2 - 1)s\sqrt{\frac{1}{n_1} + \frac{1}{n_2}}$.

Lab 1 och 2

Om vi antar normalfördelade kroppsmått och en systematisk skillnad mellan dominerande och icke dominerande sida i Lab 1 och 2, så kan vi beteckna observationerna på följande sätt för $k = 1, 2, 3, \ldots, 18$

	Icke dominerande sida	**Dominerande sida**
Kvinnor	$\xi_k^{\mathrm{I}} \in N({\mu_k}^{\mathrm{K}}, \sigma_1)$	$\xi_k^{\mathrm{D}} \in N({\mu_k}^{\mathrm{K}} + \Delta^{\mathrm{K}}, \sigma_2)$
Män	$\eta_k^{\mathrm{I}} \in N({\mu_k}^{\mathrm{M}}, \sigma_1)$	$\eta_k^{\mathrm{D}} \in N({\mu_k}^{\mathrm{M}} + \Delta^{\mathrm{M}}, \sigma_2)$

$(\sigma_1 = \sigma_2)$

Problemtyp 1: Undersök skillnaden mellan de två sidorna för kvinnor

Stickprov i par, ty *tydlig* parbildning mellan variablerna i oberoende par $(\xi_k^{\mathrm{D}}, \xi_k^{\mathrm{I}})$ med variation mellan paren så att vi *ej* får samma systematiska skillnad Δ^{K} (d v s modellfel) om vi bildar andra par $(\xi_j^{\mathrm{D}}, \xi_m^{\mathrm{I}})$.

- Bilda $\zeta_k = \xi_k^{\mathrm{D}} - \xi_k^{\mathrm{I}} \in N(\Delta^{\mathrm{K}}, \sigma)$ med observerade värden z_k.
  ($\sigma = \sqrt{\sigma_1^2 + \sigma_2^2}$ **bara** om σ_1, σ_2 oberoende, men det kvittar när σ_1, σ_2 är okända.)
- Konfidensintervall för Δ^{K} med konfidensgrad $1 - \alpha$: $[\bar{z} - a, \bar{z} + a]$ med $a = t_{\alpha/2}(17)\frac{s}{\sqrt{18}}$ (som på sid 211 i boken).

Alltså: Stickprov i par när observationer hör tydligt ihop parvis.

	Icke dominerande sida	**Dominerande sida**
Kvinnor	$\xi_k^{\mathrm{I}} \in N(\mu^{\mathrm{K}}, \sigma_1)$	$\xi_k^{\mathrm{D}} \in N(\mu^{\mathrm{K}} + \Delta^{\mathrm{K}}, \sigma_2)$
Män	$\eta_k^{\mathrm{I}} \in N(\mu^{\mathrm{M}}, \sigma_1)$	$\eta_k^{\mathrm{D}} \in N(\mu^{\mathrm{M}} + \Delta^{\mathrm{M}}, \sigma_2)$

$(\sigma_1 = \sigma_2)$

Problemtyp 2: Undersök skillnad kvinna–man för icke dominerande sida

Nu samma μ^{K} och μ^{M} för alla k (d v s väntevärden för *hela* populationen). Modellantagande (sid 216) stämmer då *både* för stickprov i par och för två stickprov, men nu ingen tydlig parindelning (kvittar vilket ξ_j^{I} som paras ihop med vilket η_m^{I}). **Tumregel:** Vi väljer då att väga ihop två oberoende stickprov (ett för män, ett för kvinnor) som på sid 215:

- Räkna ut $\bar{x}, \bar{y}, s_{\mathrm{K}}, s_{\mathrm{M}}$ och $\sigma^*_{\text{obs}} = \sqrt{\frac{(18-1)s_{\mathrm{K}}^2 + (18-1)s_{\mathrm{M}}^2}{18-1+18-1}} = \sqrt{\frac{s_{\mathrm{K}}^2 + s_{\mathrm{M}}^2}{2}}$.
- Konfidensintervall för $\mu^{\mathrm{M}} - \mu^{\mathrm{K}}$ med konfidensgrad $1 - \alpha$: $[\bar{y} - \bar{x} - a, \bar{y} - \bar{x} + a]$ med $a = t_{\alpha/2}(18 + 18 - 2)\sigma^*_{\text{obs}}\sqrt{\frac{1}{18} + \frac{1}{18}}$.

Vi räknar här ut punktskattningen σ^*_{obs} från från dubbelt så många observerade värden som metoden för stickprov i par, vilket (lite förenklat) i allmänhet ger mindre konfidensintervall än med den metoden när det ej finns någon tydlig parbildning mellan observationerna att utnyttja.