Statistikk

Jo Eidsvik

Matematiske fag, NTNU

To utvalg

Utvalg 1:

- Data $X_1, X_2, \ldots, X_{n_1}$ uavhengige og identisk fordelt $X_i \sim N(\mu_1, \sigma_1^2)$.
- Estimator for forventning $\bar{X} = \frac{1}{n_1} \sum_{i=1}^{n_1} X_i$.
- Estimator for varians $S_1^2 = \frac{1}{n_1 - 1} \sum_{i=1}^{n_1} (X_i - \bar{X})^2$.

Utvalg 2:

- Data $Y_1, Y_2, \ldots, Y_{n_2}$ uavhengige og identisk fordelt $Y_j \sim N(\mu_2, \sigma_2^2)$.
- Estimator for forventning $\bar{Y} = \frac{1}{n_2} \sum_{j=1}^{n_2} Y_j$.
- Estimator for varians $S_2^2 = \frac{1}{n_2 - 1} \sum_{j=1}^{n_2} (Y_j - \bar{Y})^2$.

Varianser kjent - Konfidens intervall for $\mu_1 - \mu_2$

1. $Z = \frac{\bar{X}-\bar{Y}-(\mu_1-\mu_2)}{\sqrt{\sigma_1^2/n_1+\sigma_2^2/n_2}}$, $Z \sim N(0,1)$.
2. Konfidensintervall: $P(\bar{X} - \bar{Y} - z_{1-\alpha/2}\sqrt{\sigma_1^2/n_1 + \sigma_2^2/n_2} < \mu_1 - \mu_2 < \bar{X} - \bar{Y} - z_{\alpha/2}\sqrt{\sigma_1^2/n_1 + \sigma_2^2/n_2}) = 1 - \alpha$

Variansene er lik ($\sigma_1^2 = \sigma_2^2 = \sigma^2$) men ukjent

Estimator for varians σ^2 er $S^2_{\text{pool}} = \frac{n_1-1}{n_1+n_2-2} S_1^2 + \frac{n_2-1}{n_1+n_2-2} S_2^2$.

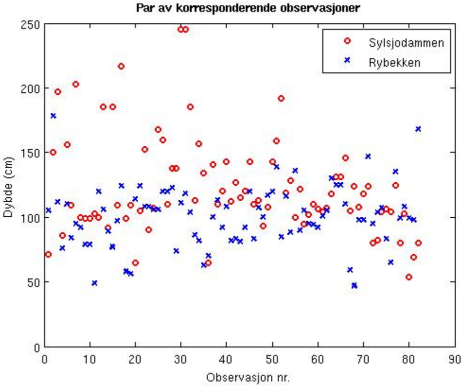
$S^2_{\text{pool}} = \frac{1}{n_1+n_2-2} \left(\sum_{i=1}^{n_1} (X_i - \bar{X})^2 + \sum_{j=1}^{n_2} (Y_j - \bar{Y})^2 \right)$

1. $T = \frac{\bar{X}-\bar{Y}-(\mu_1-\mu_2)}{s_{\text{pool}}\sqrt{1/n_1+1/n_2}}$, $T \sim t_{n_1+n_2-2}$.
2. Konfidensintervall: $P(\bar{X} - \bar{Y} - t_{n_1+n_2-2,1-\alpha/2} s_{\text{pool}} \sqrt{1/n_1 + 1/n_2} < \mu_1 - \mu_2 < \bar{X} - \bar{Y} - t_{n_1+n_2-2,\alpha/2} s_{\text{pool}} \sqrt{1/n_1 + 1/n_2}) = 1 - \alpha$

Andre tilfeller

- Varianser σ_1^2 og σ_2^2 er ukjente og ulik. Blir approksimativ T-fordeling.
- Parring: $n_1 = n_2 = n$ og $D_i = X_i - Y_i \sim N(\mu_d, \sigma_d^2)$. Blir som en 1-utvalg T situasjon.

Eksempel to-utvalg med parring



Eksempel to-utvalg med parring

- Snødybde Sylsjødammen $X_1, X_2, \ldots, X_n$ og Rybekken $Y_1, Y_2, \ldots, Y_n$.
- Antakelser $D_i = X_i - Y_i \sim N(\mu_D, \sigma_D^2)$, $i = 1, \ldots, n$, uavhengige.
- Vi søker konfidensintervall for μ_D.
- Estimator $\bar{D} = \frac{1}{n} \sum_{i=1}^{n} D_i$
- $s_D^2 = \frac{1}{n-1} \sum_{i=1}^{n} (D_i - \bar{D})^2$
- $T = \frac{\bar{D} - \mu_D}{s_D / \sqrt{n}} \sim t_{n-1}$

Eksempel - to utvalg med parring

$$P(\bar{D} - t_{n-1,0.975} s_D/\sqrt{n} < \mu_D < \bar{D} - t_{n-1,0.025} s_D/\sqrt{n}) = 0.95$$

$n = 82$, $\bar{d} = 22$. $s_D = 41$. $t_{81,0.025} = -1.99$
$\mu_{D,Lower} = 13$, $\mu_{D,Upper} = 31$ Konfidensintervallet på nivå $1 - \alpha = 0.95$ viser at Sylsjødammen har forventet mer snø enn Rybekken. μ_D er trolig ikke 0.