Problem 1

$$\bar{Y} = \sum_{i}^{n} y_i \qquad \mathbb{E}(y_i) = \mu, \quad Var(y_i) = \sigma^2$$

$$\rho_{ij} = corr(y_i, y_j) := \begin{cases} 1 & i = j \\ \rho & |i-j| = 1 \\ 0 & |i-j| > 1 \end{cases}$$

1) • Show that $Var(\bar{Y}) = \frac{\sigma^2}{n}\left(1 + 2\rho\left(\frac{n-1}{n}\right)\right)$

By definition: $corr(y_i, y_j) := \frac{cov(y_i, y_j)}{\sigma_i \cdot \sigma_j}$

$$Var(\bar{Y}) = \frac{1}{n^2}\left(\sum_{i=1}^{n} var(y_i) + 2\sum_{i<j} cov(y_i, y_j)\right)$$

$$= \frac{1}{n^2}\left(\sum_{i} \sigma^2 + 2\sum \sigma_i \sigma_j \, corr(y_i, y_j)\right)$$

$$= \frac{1}{n^2}\left(n\sigma^2 + 2\sigma^2 \sum_{i<j} \rho_{ij}\right)$$

$[|i-j| > 0 \rightarrow \rho_{ij} = 0]$ $= \frac{1}{n^2}\left(n\sigma^2 + 2\sigma^2 \sum_{i=0}^{n-1} \rho_{i,i+1}\right)$

$[|i-j| = 1 \rightarrow \rho_{ij} = \rho]$ $= \frac{1}{n^2}\left(n\sigma^2 + 2\sigma^2 (n-1)\rho\right)$

$$= \frac{\sigma^2}{n}\left(1 + 2\rho\frac{(n-1)}{n}\right) \blacksquare$$

- Show that $\mathbb{E}(s^2) = \sigma^2 (1 - 2\frac{\rho}{n})$

$$s^2 := \frac{1}{n-1} \sum (y_i - \bar{y})^2$$

$$= \frac{1}{n-1} \sum_i^n y_i^2 - 2 y_i \bar{y} + \bar{y}^2$$

$$= \frac{1}{n-1} \sum y_i^2 - \frac{2}{n-1} \sum_{i=1}^n \bar{y} y_i + \frac{n}{n-1} \bar{y}^2$$

$$\mathbb{E}(s^2) = \frac{1}{n-1} \sum \mathbb{E}(y_i^2) - \frac{n}{n-1} \mathbb{E}(\bar{y}^2)$$

$$\left[\begin{aligned} &\mathbb{E}(y_i^2) = \sigma^2 + \mu^2 \\ &E(\bar{y}^2) = Var(\bar{y}) + \mathbb{E}(\bar{y})^2 \end{aligned} \right]$$

$$= \frac{n}{n-1} (\sigma^2 + \mu^2) - \frac{n}{n-1} \left(\frac{\sigma^2}{n} \left(1 + 2\rho \frac{(n-1)}{n}\right) + \mu^2 \right)$$

$$\frac{n}{n-1} \left(\sigma^2 + \cancel{\mu^2} - \frac{\sigma^2}{n} - \frac{2\rho\sigma^2 \cancel{n}}{\cancel{n^2}} + \frac{2\rho\sigma^2}{n^2} - \cancel{\mu^2} \right)$$

$$\frac{\sigma^2}{n-1} \left(n - 1 - \frac{2\rho n - 2\rho}{n} \right)$$

$$\frac{\sigma^2}{n-1}\left(n - 1 - \frac{2\rho}{n}(n-1)\right)$$

$$\frac{\sigma^2}{\cancel{n-1}}\cancel{(n-1)}\left(1-\frac{2\rho}{n}\right) = \sigma^2\left(1-\frac{2\rho}{n}\right)$$

3) $CI := \left[\bar{y} \pm \frac{s}{\sqrt{n}} t_{n-1}^{1-\frac{\alpha}{2}}\right]$

$s := \sqrt{\frac{1}{n-1}\sum(y_i-\bar{y})^2}$

Problem 2

$y_{ij} := \mu_j + \varepsilon_{ij} \qquad j=1,\dots,n_i \qquad i=0,1$

1) $\bar{y}_i := \frac{1}{n_i}\sum_{j=1}^{n_i} y_{ij}$ $\qquad \hat{S}^2 := \hat{\sigma}^2 = \frac{\sum_{i=1}^{n_0}(y_{0i}-\bar{y}_0)^2 + \sum_{i=1}^{n_1}(y_{1i}-\bar{y}_1)^2}{n_0+n_1-2}$

$\rightarrow t = \frac{\bar{y}_1-\bar{y}_0}{\hat{S}\sqrt{1/n_0+1/n_1}}$

2)

source	DoF	SS	MS	F
Groups	1	SSbetween	SS between	$(n-2)\frac{SS\ between}{SS\ within}$
Error	$n-2$	SS within	SS within $/(n-2)$	
Total	$n-1$	SS total		

$SS_{between} := \sum_{i=1}^{k} n_i(\bar{y}_{i\cdot} - \bar{y}_{\cdot\cdot})^2 = n_0(\bar{y}_{0\cdot}-\bar{y}_{\cdot\cdot})^2 + n_1(\bar{y}_{1\cdot}-\bar{y}_{\cdot\cdot})^2$

$SS_{within} := \sum_{j=1}^{n_0}(y_{0j}-\bar{y}_{0\cdot})^2 + \sum_{j=1}^{n_1}(y_{1j}-\bar{y}_{1\cdot})^2$

$SS_{total} := \sum_{j=1}^{n_0}(y_{0j}-\bar{y}_{\cdot\cdot})^2 + \sum_{j=1}^{n_1}(y_{1j}-\bar{y}_{\cdot\cdot})^2$

where

$$y_{..} := \frac{1}{N}\sum_{i=1}^{k} n_i\hat{\beta}_i = \frac{n_0\hat{\beta}_0 + n_1\hat{\beta}_1}{n_0+n_1} = \frac{n_0\bar{y}_{0\cdot} + n_1\bar{y}_{1\cdot}}{n_0+n_1}$$

$$\hat{\beta}_i = \bar{y}_{i\cdot}$$

$$\bar{y}_{i\cdot} := \frac{1}{n_i}\sum_{j=1}^{n_i} y_{ij}$$

$$t^2 = \left(\frac{\bar{y}_1 - \bar{y}_0}{\hat{s}\sqrt{1/n_0 + 1/n_1}}\right)^2 = \frac{(\bar{y}_{1\cdot} - \bar{y}_{0\cdot})^2}{\frac{\sum_{i=1}^{n_0}(y_{0i} - \bar{y}_{0\cdot})^2 + \sum_{i=1}^{n_1}(y_{1i} - \bar{y}_{1\cdot})^2}{n_0+n_1-2}\cdot\left(\frac{1}{n_0}+\frac{1}{n_1}\right)}$$

$$= \frac{\bar{y}_{1\cdot}^2 - 2\bar{y}_{1\cdot}\bar{y}_{0\cdot} + \bar{y}_{0\cdot}^2}{\frac{SS\ within}{(n-2)}\cdot\left(\frac{1}{n_0}+\frac{1}{n_1}\right)} = (n-2)\frac{\frac{1}{\left(1/n_0 + 1/n_1\right)}\left(\bar{y}_1 - \bar{y}_0\right)^2}{SS\ within}$$

$$= (n-2)\frac{SS\ between}{SS\ within} = F$$

3) The power depends if the two groups have similar variance and equal sizes or not.

If yes ⇒ then we expect same power.

If no, t-test could be wrong, therefor the power for t statistic is smaller