1. Consider a sample of size N of observations on Y_i. Assume that $E[Y_i] = \mu$, and $E[(Y_i - \mu)^2] = \sigma^2$.

a) Consider the sample mean $\overline{Y} = \frac{1}{N}\sum_i Y_i$. Show that

$$E[(\overline{Y} - \mu)^2] = \sigma^2/N.$$

if we can show that $E[\bar{Y}] = \mu$ we may use the variance property:

$$E[\bar{Y}] = E\left[\frac{1}{n}\sum Y_i\right] = \frac{1}{N}\sum_i E[y_i] = \mu$$

So:

$$\begin{aligned} E[(\bar{Y}-\mu)^2] &= \mathrm{Var}(\bar{Y}) \\ &= \mathrm{Var}\left(\frac{1}{n}\sum y_i\right) \\ &= \frac{1}{n^2}\sum \mathrm{Var}(y_i) \\ &= \frac{1}{n^2}\cdot n\sigma^2 \\ &= \frac{\sigma^2}{n} \end{aligned}$$

b) Consider the sum:

$$D = \frac{1}{N}\sum_i (Y_i - \overline{Y})^2$$

Using the fact that $(Y_i - \overline{Y}) = ((Y_i - \mu) + (\mu - \overline{Y}))$ show that:

$$D = \frac{1}{N}\sum_i (Y_i - \mu)^2 - \frac{1}{N}\sum_i (\overline{Y} - \mu)^2$$

* Rearrange $(Y_i - \mu) + (\mu - \bar{Y}) = (Y_i - \mu) - (\bar{Y} - \mu)$

$$\begin{aligned} D &= \frac{1}{N}\sum\left((Y_i - \mu) - (\bar{Y}-\mu)\right)^2 \\ &= \frac{1}{N}\sum\left[(y_i-\mu)^2 - 2(y_i-\mu)(\bar{Y}-\mu) + (\bar{Y}-\mu)^2\right] \\ &= \frac{1}{N}\sum(y_i-\mu)^2 - 2\frac{1}{N}\sum(y_i-\mu)(\bar{Y}-\mu) + \frac{1}{N}\sum(\bar{Y}-\mu)^2 \\ &= \frac{1}{N}\sum(y_i-\mu)^2 - \frac{2}{n}\cdot n(\bar{Y}-\mu)^2 + (\bar{Y}-\mu)^2 \end{aligned}$$

$\Rightarrow$

$$\frac{1}{N}\sum(y_i - \mu) = \bar{Y} - \mu$$

$$= \frac{1}{n}\sum(y_i - \mu)^2 - 2(\bar{Y} - \mu)^2 + (\bar{Y} - \mu)^2$$

$$= \frac{1}{n}\sum(y_i - \mu)^2 - (\bar{Y} - \mu)^2 \qquad \text{since } \frac{1}{n}\sum_{i=1}^{N} = 1$$

$$= \frac{1}{n}\sum(y_i - \mu)^2 - \frac{1}{n}\sum(\bar{Y} - \mu)^2 \qquad \square$$

c) Using your answer in (a) show that $E[D] = \frac{N-1}{N}\sigma^2$.

$$E[D] = E\left[\frac{1}{n}\sum(y_i - \mu)^2 - \frac{1}{n}\sum(\bar{Y} - \mu)^2\right]$$

$$= \frac{1}{N}\sum E[(y_i - \mu)^2] - \frac{1}{n}\sum E[(\bar{Y} - \mu)^2]$$

$$= \sigma^2 - \frac{1}{n}\sigma^2$$

$$= \frac{\sigma^2 n - \sigma^2}{n}$$

$$= \frac{N-1}{n}\sigma^2$$

d) Using (c), show that

$$E[\frac{1}{N-1}\sum_i (Y_i - \overline{Y})^2] = \sigma^2$$

$$E[D] = \frac{(N-1)}{N}\sigma^2$$

$$E\left[\frac{1}{N}\sum(Y_i - \bar{Y})^2\right] = \frac{(N-1)}{N}\sigma^2 \qquad \text{* Rearrange}$$

$$E\left[\frac{1}{N-1}\sum(Y_i - \bar{Y})^2\right] = \sigma^2$$

2. Suppose we have a sample of size N made up of two groups, denoted 0 and 1. Let D_i be a dummy variable with $D_i = 1$ indicating membership in group 1. Finally, let N_0 and N_1 represent the size of the two subgroups (so $N = N_0 + N_1$).

a) Consider an OLS regression model:

$$y_i = \alpha + \beta D_i + u_i$$

Using the first order conditons for the OLS estimates $(\widehat{\alpha}, \widehat{\beta})$, show that

$$\begin{aligned} \widehat{\alpha} &= \frac{1}{N_0} \sum_{i \in 0} y_i, \\ \widehat{\alpha} + \widehat{\beta} &= \frac{1}{N_1} \sum_{i \in 1} y_i \end{aligned}$$

$$\min_{\alpha, \beta} \frac{1}{n} \sum (y_i - \alpha - \beta D_i)^2$$

FOC

$$\alpha: \frac{1}{n} \sum \left[-2(y_i - \hat{\alpha} - \hat{\beta} D_i) \right] = 0 \qquad (1)$$

$$\beta: \frac{1}{n} \sum \left[-2 D_i (y_i - \hat{\alpha} - \hat{\beta} D_i) \right] = 0 \qquad (2)$$

(2) $$\frac{1}{N} \sum D_i y_i - \frac{1}{N} \sum D_i \hat{\alpha} - \frac{1}{N} \sum D_i \hat{\beta} = 0$$

when $D_i = 1$

$$\frac{1}{N_1} \sum_{i \in 1} y_i - \frac{1}{N_1} \sum_{i \in 1} \hat{\alpha} - \frac{1}{N_1} \sum_{i \in 1} \hat{\beta} = 0$$

Rearranging

$$\boxed{\hat{\alpha} + \hat{\beta} = \frac{1}{N_1} \sum_{i \in 1} y_i}$$

(1) $$\frac{1}{N} \sum (y_i - \hat{\alpha} - \hat{\beta} D_i) = 0$$

when $D_i = 0$:

$$\frac{1}{N_0} \sum_{i \in 0} y_i - \frac{1}{N_0} \sum_{i \in 0} \hat{\alpha} = 0$$

Rearrangin

$$\boxed{\hat{\alpha} = \frac{1}{N} \sum_{i \in 0} y_i}$$

b) Now consider a case where there are 3 groups: $0, 1, 2$ and there are 2 dummies: $D_{1i} = 1$ if i is in group 1, and 0 otherwise, and $D_{2i} = 1$ if i is in group 2, and 0 otherwise, with the regression model:

$$y_i = \alpha + \beta_1 D_{1i} + \beta_2 D_{2i} + u$$

i) find the first order conditions for minimizing the sum of squared residuals, which define the OLS estimator.

ii) using the FOC show that the OLS estimators for the 3-group model will have

$$\begin{aligned}
\widehat{\alpha} &= \frac{1}{N_0}\sum_{i\in 0} y_i \\
\widehat{\alpha} + \widehat{\beta}_1 &= \frac{1}{N_1}\sum_{i\in 1} y_i \\
\widehat{\alpha} + \widehat{\beta}_2 &= \frac{1}{N_2}\sum_{i\in 2} y_i
\end{aligned}$$

Let $D_i = \begin{pmatrix} 1 \\ D_1 \\ D_2 \end{pmatrix}$, $\beta = \begin{pmatrix} \alpha \\ \beta_1 \\ \beta_2 \end{pmatrix}$

$$\min_{\beta} \frac{1}{N} \Sigma \left(y_i - D_i' \beta \right)^2$$

FOC: $\beta : \frac{1}{N} \Sigma (-2) D_i \left(y_i - D_i' \hat{\beta} \right) = 0$

$$\frac{1}{N} \Sigma D_i \left(y_i - D_i' \hat{\beta} \right) = 0$$

$$\frac{1}{N} \Sigma \begin{pmatrix} 1 \\ D_1 \\ D_2 \end{pmatrix} \left(y_i - \hat{\alpha} - \hat{\beta}_1 D_1 - \hat{\beta}_2 D_2 \right) = 0$$

$$\Rightarrow \frac{1}{N} \Sigma (y_i - \hat{\alpha} - \hat{\beta}_1 D_1 - \hat{\beta}_2 D_2 = 0 \qquad (1)$$

$$\frac{1}{N} \Sigma \left(y_i - \hat{\alpha} - \hat{\beta}_1 D_1 - \hat{\beta}_2 D_2 \right) D_{1,i} = 0 \qquad (2)$$

$$\frac{1}{N} \Sigma \left(y_i - \hat{\alpha} - \hat{\beta}_1 D_1 - \hat{\beta}_2 D_2 \right) D_{2i} = 0 \qquad (3)$$

Using Eq (1) yields

$$\frac{1}{N}\sum\left(y_i - \hat{\alpha} - \hat{\beta}_1 D_1 - \hat{\beta}_2 D_2\right) = 0$$

when $D_1, D_2 = 0$

$$\Rightarrow \frac{1}{N_0}\sum_{i\in 0}(y_i - \hat{\alpha}) \Rightarrow \frac{1}{N_0}\sum_{i\in 0}\hat{\alpha} = \frac{1}{N_0}\sum_{i\in 0} y_i$$

$$\Rightarrow \boxed{\hat{\alpha} = \frac{1}{N_0}\sum_{i\in 0} y_i}$$

Using Eq (2) yields

$$\frac{1}{N}\sum\left(y_i - \hat{\alpha} - \hat{\beta}_1 D_1 - \hat{\beta}_2 D_2\right) D_{1,i} = 0$$

when $D_1 = 1, D_2 = 0$

$$\frac{1}{N_1}\sum_{i\in 1}\left(y_i - \hat{\alpha} - \hat{\beta}_1\right) = 0$$

$$\Rightarrow \boxed{\hat{\alpha} + \hat{\beta} = \frac{1}{N_1}\sum_{i\in 1} y_i}$$

Using Eq (3) yields

$$\frac{1}{N}\sum\left(y_i - \hat{\alpha} - \hat{\beta}_1 D_1 - \hat{\beta}_2 D_2\right) D_{2,i} = 0$$

when $D_1 = 0, D_2 = 1$

$$\frac{1}{N_2}\sum_{i\in 2}\left(y_i - \hat{\alpha} - \hat{\beta}_2\right) = 0$$

$$\Rightarrow \boxed{\hat{\alpha} + \hat{\beta}_2 = \frac{1}{N_2}\sum_{i\in 2} y_i}$$

3. In Lecture 3 we showed that the j^{th} row of the population regression coefficient β^* from the model

$$y_i = x_i\beta^* + u_i \qquad (1)$$

can be obtained by first getting the residual from an auxilliary regression of x_{ji} on all the other $x's$:

$$x_{ji} = x'_{(\sim j)i}\pi + \xi_i \qquad (2)$$

then forming:

$$\beta^*_j = E[\xi_i^2]^{-1}E[\xi_i y_i] \qquad (3)$$

Suppose that we also "residualized" the dependent variable using an auxilliary regression of y_i on all the other $x's$:

$$y_i = x'_{(\sim j)i}\lambda + \phi_i \qquad (4)$$

Show that its also true that:

$$\beta^*_j = E[\xi_i^2]^{-1}E[\xi_i\phi_i] \qquad (5)$$

(In other words, we could residualize BOTH x_{ji} AND y_i and still get the same right anwer).

$$E[\varepsilon_i y_i] = E[\varepsilon_i [x'_{(\sim j)i}\lambda + \phi_i]]$$
$$= E[\varepsilon_i (x'_{(\sim j)i}\lambda)] + E[\varepsilon_i \phi_i]$$

by equation (2) ε_i is orthogonal to every $x'_{(\sim j)i}$, $E[\varepsilon_i (x'_{(\sim j)i}\lambda)] = 0$.

$$\Rightarrow E[\varepsilon_i y_i] = E[\varepsilon_i \phi_i]$$

Therefore, we may write $\beta^*_j = E[\varepsilon_i^2]^{-1} E[\xi_i \phi_i]$