MLR Issues

William Brasic

The University of Arizona

Functional Form Misspecification

Definition 1: Functional Form Misspecification

A functional form misspecification occurs when we don't correctly specify the relationship between the independent and explanatory variables.

Functional Form Misspecification

Example 1: Omitted Variable Bias

When the true model is $wage_i = \beta_0 + \beta_1 educ_i + \beta_2 exper_i + u_i$, but we exclude experience from the regression.

- Omitted Variable Bias (OVB): The size of the bias is determined by the size of β_2 and the correlation between education and experience.
  - Violation of MLR Assumption 4.

Functional Form Misspecification

Example 2: Model Overspecification

When the true model is $y_i = \beta_0 + u_i$, but we estimate $y_i = \beta_0 + \beta_1 x_i + u_i$

- This is an example of *model overspecification*
  - Likely want to include covariates that lead to meaningful increases in the adjusted R-squared.
- $\widehat{\beta}_0$ and $\widehat{\beta}_1$ unbiased and consistent, but have larger variance (less efficient)
  - $\mathbb{E}\left[\widehat{\beta}_0\right] = \beta_0$.
  - $\mathbb{E}\left[\widehat{\beta}_1\right] = 0$.

Functional Form Misspecification

Example 2: Model Overspecification

Under MLR Assumptions 1-4,

$$\widehat{\beta}_1 = \frac{\sum_{i=1}^{n} \left(x_i - \overline{X}\right)\left(y_i - \overline{Y}\right)}{\sum_{i=1}^{n} \left(x_i - \overline{X}\right)^2}$$

Functional Form Misspecification

Example 2: Model Overspecification

Under MLR Assumptions 1-4,

$$\begin{aligned}\widehat{\beta}_1 &= \frac{\sum_{i=1}^{n}\left(x_i-\overline{X}\right)\left(y_i-\overline{Y}\right)}{\sum_{i=1}^{n}\left(x_i-\overline{X}\right)^2} \\ &= \frac{\sum_{i=1}^{n}\left(x_i-\overline{X}\right)\left(\beta_0+u_i-\overline{Y}\right)}{\sum_{i=1}^{n}\left(x_i-\overline{X}\right)^2}\end{aligned}$$

Functional Form Misspecification

Example 2: Model Overspecification

Under MLR Assumptions 1-4,

$$\begin{aligned}\widehat{\beta}_1 &= \frac{\sum_{i=1}^{n}\left(x_i-\overline{X}\right)\left(y_i-\overline{Y}\right)}{\sum_{i=1}^{n}\left(x_i-\overline{X}\right)^2} \\ &= \frac{\sum_{i=1}^{n}\left(x_i-\overline{X}\right)\left(\beta_0+u_i-\overline{Y}\right)}{\sum_{i=1}^{n}\left(x_i-\overline{X}\right)^2} \\ &= \frac{\sum_{i=1}^{n}\left(x_i-\overline{X}\right)u_i}{\sum_{i=1}^{n}\left(x_i-\overline{X}\right)^2}\end{aligned}$$

because $\sum_{i=1}^{n}\left(x_i-\overline{X}\right)=0$. Taking expectation of each side gives $\mathbb{E}\left[\widehat{\beta}_1 \mid x_i\right]=0$ because MLR Assumptions two and four imply $\mathbb{E}[u_i \mid x_i]=0$ for all observations.

Functional Form Misspecification

Example 2: Model Overspecification

Under MLR Assumptions 1-4,

$$\begin{aligned}\mathbb{E}\left[\widehat{\beta}_0 \mid x_i\right] &= \mathbb{E}\left[\overline{Y} - \widehat{\beta}_1 \overline{X} \mid x_i\right] \\ &= \mathbb{E}\left[\overline{Y} \mid x_i\right] - \mathbb{E}\left[\widehat{\beta}_1 \mid x_i\right] \overline{X} \\ &= \mathbb{E}\left[y_i \mid x_i\right] \\ &= \beta_0.\end{aligned}$$

Functional Form Misspecification

Example 3: Model Misspecification

When the true model is $y_i = \beta_0 + x_i^{\beta_1} + u_i$, but we estimate $y_i = \beta_0 + \beta_1 x_i + u_i$.

- This is an example of *model misspecification* where MLR Assumptions 1 is violated.
- $\widehat{\beta}_0$ and $\widehat{\beta}_1$ are biased.

Functional Form Misspecification

Example 3: Model Misspecification

Under MLR Assumptions 2-4, (MLR assumption 1 is broken because of the nonlinearity)

$$\widehat{\beta}_1 = \frac{\sum_{i=1}^{n} \left(x_i - \overline{X}\right)\left(y_i - \overline{Y}\right)}{\sum_{i=1}^{n} \left(x_i - \overline{X}\right)^2}$$

Functional Form Misspecification

Example 3: Model Misspecification

Under MLR Assumptions 2-4, (MLR assumption 1 is broken because of the nonlinearity)

$$\begin{aligned}\widehat{\beta}_1 &= \frac{\sum_{i=1}^{n}\left(x_i-\overline{X}\right)\left(y_i-\overline{Y}\right)}{\sum_{i=1}^{n}\left(x_i-\overline{X}\right)^2} \\ &= \frac{\sum_{i=1}^{n}\left(x_i-\overline{X}\right)\left(\beta_0+x_i^{\beta_1}+u_i-\overline{Y}\right)}{\sum_{i=1}^{n}\left(x_i-\overline{X}\right)^2}\end{aligned}$$

Functional Form Misspecification

Example 3: Model Misspecification

Under MLR Assumptions 2-4, (MLR assumption 1 is broken because of the nonlinearity)

$$\begin{aligned}
\widehat{\beta}_1 &= \frac{\sum_{i=1}^{n}\left(x_i-\overline{X}\right)\left(y_i-\overline{Y}\right)}{\sum_{i=1}^{n}\left(x_i-\overline{X}\right)^2} \\
&= \frac{\sum_{i=1}^{n}\left(x_i-\overline{X}\right)\left(\beta_0+x_i^{\beta_1}+u_i-\overline{Y}\right)}{\sum_{i=1}^{n}\left(x_i-\overline{X}\right)^2} \\
&= \frac{\sum_{i=1}^{n}\left(x_i-\overline{X}\right)x_i^{\beta_1}}{\sum_{i=1}^{n}\left(x_i-\overline{X}\right)^2} + \frac{\sum_{i=1}^{n}\left(x_i-\overline{X}\right)u_i}{\sum_{i=1}^{n}\left(x_i-\overline{X}\right)^2}
\end{aligned}$$

because $\sum_{i=1}^{n}\left(x_i-\overline{X}\right)=0$.

Functional Form Misspecification

Example 3: Model Misspecification

Taking expectation of both sides yields

$$\mathbb{E}\left[\widehat{\beta}_1 \mid x_i\right] = \mathbb{E}\left[\frac{\sum_{i=1}^{n}\left(x_i - \overline{X}\right)x_i^{\beta_1}}{\sum_{i=1}^{n}\left(x_i - \overline{X}\right)^2} + \frac{\sum_{i=1}^{n}\left(x_i - \overline{X}\right)u_i}{\sum_{i=1}^{n}\left(x_i - \overline{X}\right)^2}\middle| x_i\right]$$

Functional Form Misspecification

Example 3: Model Misspecification

Taking expectation of both sides yields

$$\begin{aligned}
\mathbb{E}\left[\widehat{\beta}_1 \mid x_i\right] &= \mathbb{E}\left[\frac{\sum_{i=1}^{n}\left(x_i-\overline{X}\right)x_i^{\beta_1}}{\sum_{i=1}^{n}\left(x_i-\overline{X}\right)^2}+\frac{\sum_{i=1}^{n}\left(x_i-\overline{X}\right)u_i}{\sum_{i=1}^{n}\left(x_i-\overline{X}\right)^2}\middle| x_i\right] \\
&= \mathbb{E}\left[\frac{\sum_{i=1}^{n}\left(x_i-\overline{X}\right)x_i^{\beta_1}}{\sum_{i=1}^{n}\left(x_i-\overline{X}\right)^2}\middle| x_i\right] \\
&\neq \beta_1
\end{aligned}$$

because $\mathbb{E}[u_i \mid x_i] = 0$ for each observation by MLR Assumptions 2 and 4.

Functional Form Misspecification

Example 3: Model Misspecification

Under MLR Assumptions 2-4, (MLR assumption 1 is broken because of the nonlinearity)

$$\mathbb{E}\left[\widehat{\beta}_0 \mid x_i\right] = \mathbb{E}\left[\overline{Y} - \widehat{\beta}_1 \overline{X} \mid x_i\right]$$

Functional Form Misspecification

Example 3: Model Misspecification

Under MLR Assumptions 2-4, (MLR assumption 1 is broken because of the nonlinearity)

$$\begin{aligned}\mathbb{E}\left[\widehat{\beta}_0 \mid x_i\right] &= \mathbb{E}\left[\overline{Y} - \widehat{\beta}_1 \overline{X} \mid x_i\right] \\ &= \mathbb{E}\left[\widehat{\beta}_0 + \overline{X^{\widehat{\beta}_1}} - \widehat{\beta}_1 \overline{X} \mid x_i\right] \\ &\neq \beta_0.\end{aligned}$$

Functional Form Misspecification

Definition 2: RESET (Regression Specification Error Test) Test

A RESET (Regression Specification Error Test) Test is a way of determining if we have a misspecification issue:

1. Estimate $y_i = \boldsymbol{x}_i'\boldsymbol{\beta} + u_i$.
2. Obtain $\widehat{y}_i$.
3. Estimate $y_i = \boldsymbol{x}_i'\boldsymbol{\beta} + \delta_1\widehat{y}_i^2 + \delta_2\widehat{y}_i^3 + u_i$.
4. Carry out an F-test of $H_0 : \delta_1 = \delta_2 = 0$ versus $H_A : \delta_1 \neq 0$ or $\delta_2 \neq 0$.

- Rejection of the null implies some misspecification issue.
- The F-statistic has two numerator degrees of freedom and $n - k - 2$ denominator degrees of freedom.

Level or Natural Log for Outcome?

Question 1: Natural Log or Level for Outcome?

When should we use y_i versus $\ln(y_i)$ as our outcome variable?

Level or Natural Log for Outcome?

Question 1: Natural Log or Level for Outcome?

When should we use y_i versus $\ln(y_i)$ as our outcome variable?

Answer to Question 3

Really it depends on your preference. Here are some tips:

1. When $y_i \leq 0$, we can't use logs.
2. When y_i can take on on large values, likely want to use logs.
3. When percentage changes seem more informative, likely want to use logs.

High Multicollinearity

Question 2: High Multicollinearity

We know $X'X$ is singular because of perfect multicollinearity we know the OLS solution does not exist, but what happens with non-perfect yet high multicollinearity?

High Multicollinearity

Question 2: High Multicollinearity

We know $X'X$ is singular because of perfect multicollinearity we know the OLS solution does not exist, but what happens with non-perfect yet high multicollinearity?

Answer to Question 2

Higher correlation among explanatory variables means $(X'X)^{-1}$ will be large (yet still invertible) meaning $\mathbb{V}\left[\widehat{\boldsymbol{\beta}} \mid x_i\right] = \sigma^2 (X'X)^{-1}$ will be large.

- Variance is inflated.
- Standard errors are inflated.
- Inference becomes unreliable.

Omitted Variable Bias (OVB)

Definition 3: Omitted Variable Bias (OVB)

Omitted Variable Bias (OVB) occurs when we exclude a relevant explanatory variable from model.

- This could, but not necessarily, create an endogeneity problem so $\text{Cov}(v_i, x_i) \neq 0$ which makes our estimators inconsistent and would bias our estimators because this also implies $\mathbb{E}[v_i \mid x_i] \neq 0$

Omitted Variable Bias (OVB)

Example 4: Omitted Variable Bias (OVB)

Suppose the true model is $\boldsymbol{y} = X\boldsymbol{\beta} + Z\boldsymbol{\gamma} + \boldsymbol{u}$, but we estimate $\boldsymbol{y} = X\boldsymbol{\beta} + \boldsymbol{v}$ where $\boldsymbol{v} = Z\boldsymbol{\gamma} + \boldsymbol{u}$ so Z is the set of excluded explanatory variables.

Omitted Variable Bias (OVB)

Example 4: Omitted Variable Bias (OVB)

$$\begin{aligned}\widehat{\boldsymbol{\beta}} &= (X'X)^{-1} X'\boldsymbol{y} \\ &= (X'X)^{-1} X' (X\boldsymbol{\beta} + Z\boldsymbol{\gamma} + \boldsymbol{u})\end{aligned}$$

Omitted Variable Bias (OVB)

Example 4: Omitted Variable Bias (OVB)

$$\begin{aligned}\widehat{\boldsymbol{\beta}} &= (X'X)^{-1} X'\boldsymbol{y} \\ &= (X'X)^{-1} X' (X\boldsymbol{\beta} + Z\boldsymbol{\gamma} + \boldsymbol{u}) \\ &= \boldsymbol{\beta} + (X'X)^{-1} X'Z\boldsymbol{\gamma} + (X'X)^{-1} X'\boldsymbol{u}.\end{aligned}$$

Omitted Variable Bias (OVB)

Example 4: Omitted Variable Bias (OVB)

Taking expectations of both sides yields

$$\mathbb{E}\left[\widehat{\boldsymbol{\beta}} \mid X\right] = \mathbb{E}\left[\left(X'X\right)^{-1} X'\left(X\boldsymbol{\beta} + Z\boldsymbol{\gamma} + \boldsymbol{u}\right) \mid X\right]$$

Omitted Variable Bias (OVB)

Example 4: Omitted Variable Bias (OVB)

Taking expectations of both sides yields

$$\begin{aligned}\mathbb{E}\left[\widehat{\boldsymbol{\beta}} \mid X\right] &= \mathbb{E}\left[\left(X'X\right)^{-1}X'\left(X\boldsymbol{\beta}+Z\boldsymbol{\gamma}+\boldsymbol{u}\right) \mid X\right] \\ &= \boldsymbol{\beta}+\mathbb{E}\left[\left(X'X\right)^{-1}X'Z\boldsymbol{\gamma} \mid X\right]\end{aligned}$$

because $\mathbb{E}\left[\boldsymbol{u} \mid X\right] = 0$ by MLR Assumption 4.

So for $\widehat{\boldsymbol{\beta}}$ to be unbiased, we need $\mathbb{E}\left[\left(X'X\right)^{-1}X'Z \mid X\right] = \mathbf{0}$.

When would this happen?

Omitted Variable Bias (OVB)

Example 4: Omitted Variable Bias (OVB)

$\widehat{\boldsymbol{\beta}}$ is unbiased when X and Z are orthogonal so $X'Z = 0$ implying $\mathbb{E}[X'Z] = \mathbf{0}$:

- Dot product between columns of X and columns of Z is $\mathbf{0}$ meaning they are linearly independent, i.e., no column of X is a linear function of any other column(s) of Z.

Omitted Variable Bias (OVB)

Example 4: Omitted Variable Bias (OVB)

$\widehat{\boldsymbol{\beta}}$ is unbiased when X and Z are orthogonal so $X'Z = 0$ implying $\mathbb{E}[X'Z] = \mathbf{0}$:

- Dot product between columns of X and columns of Z is $\mathbf{0}$ meaning they are linearly independent, i.e., no column of X is a linear function of any other column(s) of Z.

In practice, omitted variable bias occurs when the relevant variable(s) Z is correlated with both the independent variable(s) X and the dependent variable $\boldsymbol{y}$, but the excluded variable(s) Z is left out of the model.

Omitted Variable Bias (OVB)

Example 4: Omitted Variable Bias (OVB)

Recall

$$\mathbb{E}\left[\widehat{\boldsymbol{\beta}} \mid X\right] = \boldsymbol{\beta} + \mathbb{E}\left[\left(X'X\right)^{-1} X'Z\boldsymbol{\gamma} \mid X\right].$$

$\widehat{\boldsymbol{\beta}}$ is unbiased if

- $\gamma = 0$ so variables in Z don't impact $\boldsymbol{y}$.
- $\mathbb{E}\left[X'Z\right] = 0$ (orthogonality/exogeneity condition).

Omitted Variable Bias (OVB)

Example 4: Omitted Variable Bias (OVB)

Recall

$$\mathbb{E}\left[\widehat{\boldsymbol{\beta}} \mid X\right] = \boldsymbol{\beta} + \mathbb{E}\left[\left(X'X\right)^{-1} X'Z\boldsymbol{\gamma} \mid X\right].$$

$\widehat{\boldsymbol{\beta}}$ is unbiased if

- $\gamma = 0$ so variables in Z don't impact $\boldsymbol{y}$.
- $\mathbb{E}\left[X'Z\right] = 0$ (orthogonality/exogeneity condition).

Otherwise, we're screwed and $\widehat{\boldsymbol{\beta}}$ will be biased!

Direction of OVB

TABLE 3.2 Summary of Bias in $\tilde{\beta}_1$ When x_2 Is Omitted in Estimating Equation (3.40)

	$\text{Corr}(x_1, x_2) > 0$	$\text{Corr}(x_1, x_2) < 0$
$\beta_2 > 0$	Positive bias	Negative bias
$\beta_2 < 0$	Negative bias	Positive bias

- For instance, if x_2 has a positive effect on y ($\beta_2 > 0$) and x_1 and x_2 are positively correlated, then $\widehat{\beta}_1$ will be inflated.
- Bias is multiplicative.

OVB in Practice

Question 3: OVB in Practice

How often do we have an OVB problem in practice?

OVB in Practice

Question 3: OVB in Practice

How often do we have an OVB problem in practice?

Answer to Question 3

You will likely always have an OVB issue. In practice, you want to use the bias direction as a rough approximation of the sign and report it for transparency. OVB is a huge issue in modern econometrics because with it we can't identify the causal impact of our covariates on our outcome.

Practical OVB Scenarios

Example 5: Practical OVB Scenarios

1. Regression of income on whether one is married or not, but exclude their physical attractiveness.

Practical OVB Scenarios

Example 5: Practical OVB Scenarios

1. Regression of income on whether one is married or not, but exclude their physical attractiveness.
2. Regression of income on whether one has a college degree, but exclude ability.

Practical OVB Scenarios

Example 5: Practical OVB Scenarios

1. Regression of income on whether one is married or not, but exclude their physical attractiveness.
2. Regression of income on whether one has a college degree, but exclude ability.
3. Regression of income on whether one drinks, but exclude job stress level.

OVB Solutions

Question 4: OVB Solutions

What solutions exist to the OVB problem?

OVB Solutions

Question 4: OVB Solutions

What solutions exist to the OVB problem?

Answer to Question 4

Unfortunately, there is no statistical test for OVB. However, we can obtain an instrumental variable (IV) to instrument for the endogeneous regressor of interest. To be a valid IV, we must have that

1. The IV itself is exogenous.
2. The IV is sufficiently correlated with the endogenous regressor of interest.
3. The IV can only impact the outcome through the endogenous regressor of interest (exclusion restriction).

Measurement Error

Definition 4: Measurement Error

Measurement error occurs when we mismeasure either the outcome or explanatory variables.

- Typically, when the outcome is mismeasured, our estimators are unbiased, yet have higher variance.
- However, when an explanatory variable is mismeasured, we are likely screwed with biased estimators.

Measurement Error in Outcome

Example 6: Measurement Error in the Outcome

Suppose the true model is $\boldsymbol{y} = X\boldsymbol{\beta} + \boldsymbol{u}$, but we observe $\boldsymbol{y}^* = \boldsymbol{y} + \boldsymbol{e}$ so $\boldsymbol{y} = \boldsymbol{y}^* - \boldsymbol{e}$. Then, we estimate

$$\boldsymbol{y}^* = X\boldsymbol{\beta} + \boldsymbol{u} + \boldsymbol{e}.$$

- $\boldsymbol{y}$ is measured with error $\boldsymbol{e}$ giving the observed $\boldsymbol{y}^*$.

Measurement Error in Outcome

Example 6: Measurement Error in the Outcome

Under MLR Assumptions 1-4,

$$\begin{aligned}\widehat{\boldsymbol{\beta}} &= (X'X)^{-1}X'\boldsymbol{y}^* \\ &= \boldsymbol{\beta} + (X'X)^{-1}X'\boldsymbol{u} + (X'X)^{-1}X'\boldsymbol{e}.\end{aligned}$$

Measurement Error in Outcome

Example 6: Measurement Error in the Outcome

Under MLR Assumptions 1-4,

$$\begin{aligned}\mathbb{E}\left[\widehat{\boldsymbol{\beta}} \mid X\right] &= \mathbb{E}\left[\boldsymbol{\beta} + (X'X)^{-1}X'\boldsymbol{u} + (X'X)^{-1}X'\boldsymbol{e} \mid X\right] \\ &= \boldsymbol{\beta} + (X'X)^{-1}X'\mathbb{E}\left[\boldsymbol{e} \mid X\right].\end{aligned}$$

Thus, for $\widehat{\boldsymbol{\beta}}$ to be unbiased, we not only need $\mathbb{E}[\boldsymbol{u} \mid X] = 0$, but also $\mathbb{E}[\boldsymbol{e} \mid X] = 0$.

- As long as the measurement error in the outcome is zero on average, we are good.

Measurement Error in Outcome

Example 6: Measurement Error in the Outcome

Even when $\widehat{\boldsymbol{\beta}}$ is unbiased,

$$\begin{aligned}\mathbb{V}\left[\widehat{\boldsymbol{\beta}} \mid X\right] &= \left(\sigma_{\boldsymbol{u}}^2 + \sigma_{\boldsymbol{e}}^2\right)(X'X)^{-1} \\ &> \sigma_{\boldsymbol{u}}^2 (X'X)^{-1}.\end{aligned}$$

when $\boldsymbol{u}$ and $\boldsymbol{e}$ are uncorrelated. Thus, variance of our estimators is inflated is the presence of mismeasured outcome variable.

Measurement Error in Outcome

Question 5: Measurement Error in the Outcome

Say we wish to regress income on education. More educated individuals may be more likely to over report their income. This would create measurement error in the outcome that is driven through an individual's education level. What would the direction of the bias be here?

Measurement Error in Outcome

Question 5: Measurement Error in the Outcome

Say we wish to regress income on education. More educated individuals may be more likely to over report their income. This would create measurement error in the outcome that is driven through an individual's education level. What would the direction of the bias be here?

Answer to Question 5

Higher education levels likely lead to high levels of income. Moreover, education levels and the measurement error are positively correlated. Thus, the estimate of the education parameter would be biased upward. Other possible arguments?

Measurement Error in Regressors

Example 7: Measurement Error in Regressors

Suppose the true model is $\boldsymbol{y} = X\boldsymbol{\beta} + \boldsymbol{u}$, but we observe $X^* = X + E$ so $X = X^* - E$. Then, we estimate

$$\boldsymbol{y} = (X^* - E)\boldsymbol{\beta} + \boldsymbol{e}.$$

- Covariates in X are measured with error E giving the observed X^*.

Measurement Error in Regressors

Example 7: Measurement Error in Regressors

Under MLR Assumtions 1-4,

$$\mathbb{E}\left[\widehat{\boldsymbol{\beta}} \mid X^*\right] = \beta \frac{\mathbb{V}\left[X\right]}{\mathbb{V}\left[X\right] + \mathbb{V}\left[E\right]}.$$

- Attenuation Bias!

Measurement Error in Regressors

Definition 5: Attenuation Bias

Recall that $X^* = X + E$ where X^* are the observed regressors, X are the true regressors, and E is the measurement error.

While higher variance in the regressors can reduce our estimator's variance, higher variance measurement error typically leads to attenuation bias. This means that the estimated coefficients $\widehat{\beta}$ are biased towards zero. The larger the variance of the measurement error relative to the variance of the true regressors, the more severe the attenuation bias.

Simultaneity

Definition 6: Simultaneity

The simultaneity problem occurs when a regressor causes and is caused by the outcome.

- This is often called reverse causality.

Simultaneity

Example 8: Simultaneity Scenarios

Some regression models with a simultaneity problem include:

1. Regressing an individual's income on whether or not they drink alcohol.
2. Regressing an individual's pay on whether or not they are in a union.
3. Regressing price on quantity.

Simultaneity

Example 9: Simultaneity

Consider the classical supply and demand framework:

$$q_s = \alpha_1 p + u$$
$$q_d = \alpha_2 p + e$$

where q_s and q_d are quantity supplied and demanded, respectively, and p is price.

Simultaneity

Example 9: Simultaneity

In equilibrium, quantity supplied equals quantity demanded:

$$q_s = q_d = q.$$

Simultaneity

Example 9: Simultaneity

Therefore, we can write:

$$q = \alpha_1 p + u \quad \text{(Supply equation)}$$
$$q = \alpha_2 p + e \quad \text{(Demand equation).}$$

Simultaneity

Example 9: Simultaneity

Setting the equations equal gives:

$$\begin{aligned} \alpha_1 p + u &= \alpha_2 p + e \\ (\alpha_1 - \alpha_2)p &= e - u \\ p &= \frac{e-u}{\alpha_1 - \alpha_2}. \end{aligned}$$

Thus, p is endogenous (correlated with both e and u) so α_1 and α_2 will be biased.

Random Sampling

Question 6: Random Sampling

How do we determine if MLR Assumption 2 of random sampling is met? If we were interested in estimating the income returns to education and collected a sample of individuals all with income above 100K, would this sample be random?

Random Sampling

Question 6: Random Sampling

How do we determine if MLR Assumption 2 of random sampling is met? If we were interested in estimating the income returns to education and collected a sample of individuals all with income above 100K, would this sample be random?

Answer to Question 6

Largely our own intuition and no, this sample is not random since the average income isn't even 100K.

- There is no statistical test.

Thank You!