# Linear Regression with a Single Continuous Predictor

PSY 612

# Review: General Linear Models & Model Comparisons

# The basic data analysis equation

**Data = Model + Error**

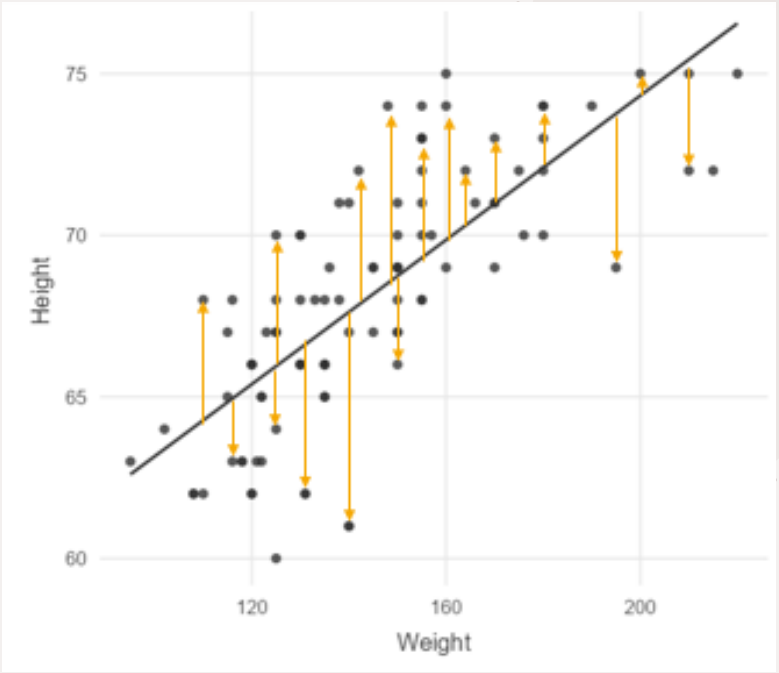
- **Data** = the raw, empirical observations that we want our model to explain
- **Model** = an algebraic expression that describes a geometric shape (e.g., a line, a plane, etc.) that captures the general pattern observed in the data
- **Error** = the part of the data left unexplained by the model

The **best-fitting model** is the one that best does best at predicting the data by **minimizing the sum of squared errors (SSE)**.

# The best-fitting model

Sum of squared errors:

- $SSE = \Sigma(Y_i - Y'_i)^2$
  - $Y_i$ = each participant's raw score in the data set
  - $Y'_i$ = the score predicted by the model for each participant
  - $(Y_i - Y'_i)$ = the error for each participant; the difference between their actual and predicted scores (aka, **residual**)
  - $(Y_i - Y'_i)^2$ = each error (or residual) squared



# The general equation for a linear model

$$Y_i = \beta_0 + \beta_1 X_{i1} + \beta_2 X_{i2} + \beta_3 X_{i3} + \ldots + \beta_j X_{ij} + \varepsilon_i$$

Data | Model | Error

This equation describes the relationship between the predictor variables (X's) and outcome variable (Y) *at the population level*:

- $Y_i$ = Every participant's raw score on the outcome variable
- βs are called **model parameters**
  - $\beta_0$ = the model's intercept
  - $\beta_1$ to $\beta_j$ = The slope contribution corresponding to each of the predictors in the model
- $X_{i1}$ to $X_{ij}$ = Every participant's raw scores on each of the predictor variables
- $\varepsilon_i$ = Every participant's error

# The estimate of a linear model

$$Y_i = b_0 + b_1X_{i1} + b_2X_{i2} + b_3X_{i3} + \ldots + b_jX_{ij} + e_i$$

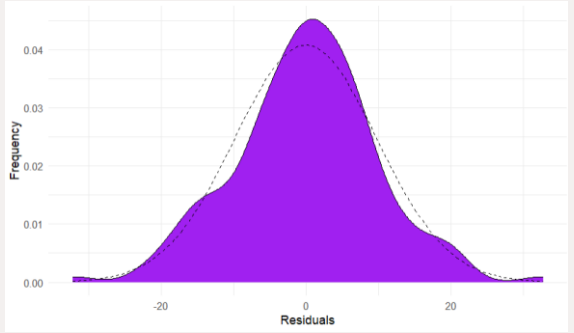
Data — Model — Error

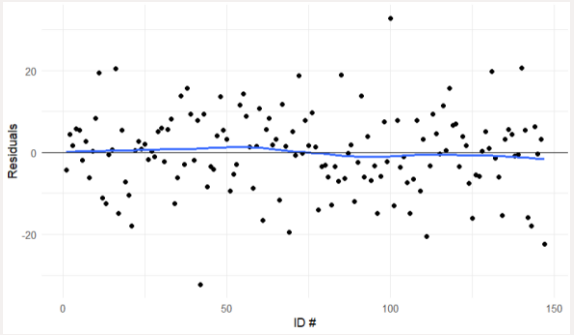
We typically do not have population-level data and instead have to ***estimate the equation using sample data*** that represents the linear relationship between the predictor(s) (X's) and outcome variable (Y):

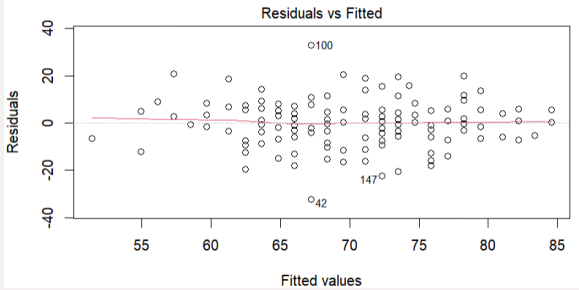
- $Y_i$ = Participants' raw scores on the outcome variable in the sample
- $b$'s are called **parameter estimates**
  - $b_0$ = estimate of the model's intercept
  - $b_1$ - $b_j$ = estimates of the slope contribution corresponding to each of the predictors in the model
- $X_{i1}$ - $X_{ij}$ = Participants' raw scores on each of the predictor variables in the sample
- $e_i$ = Each participant's error in the sample

# Assumptions underlying a linear model

- Errors are **normally distributed**
- Errors are **independent**
- Errors are **equally distributed** across the range of fitted values (i.e., homogeneity of variance)



# Models we've covered

**Zero parameter model:**

- $Y_i = B_0 + \varepsilon_i$

**One parameter model:**

- $Y_i = \beta_0 + \varepsilon_i$

**Model with a single categorical predictor with 2 levels (# of codes = *m* - 1):**

- $Y_i = \beta_0 + \beta_1 Code_i + \varepsilon_i$

**Model with a single categorical predictor with 2+ levels (# of codes = *m* – 1):**

- $Y_i = \beta_0 + \beta_1 Code1_i + \beta_2 Code2_i + .... + \beta_j CodeJ_i + \varepsilon_i$

**Model with multiple categorical predictors (# of codes = *m* – 1, inclusion of interaction effects):**

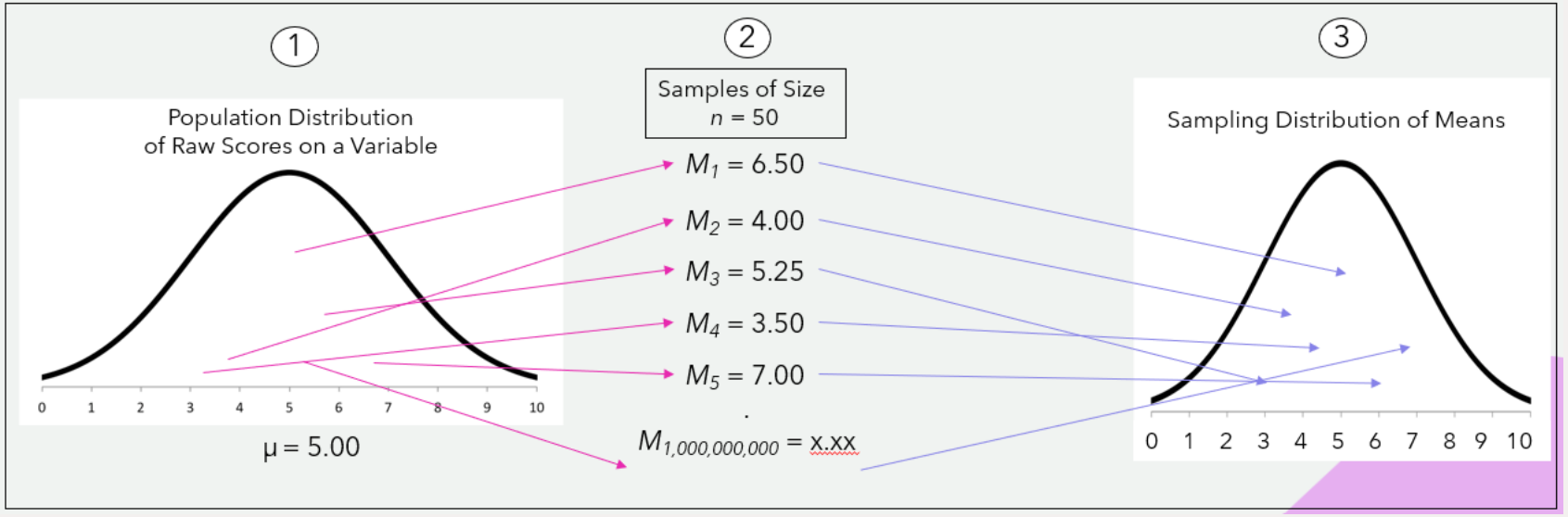
- $Y_i = \beta_0 + \beta_1 Code1_i + \beta_2 Code2_i + .... + \beta_j CodeJ_i + \varepsilon_i$

# Testing significance of predictor(s) using model comparisons

# Sampling Distributions

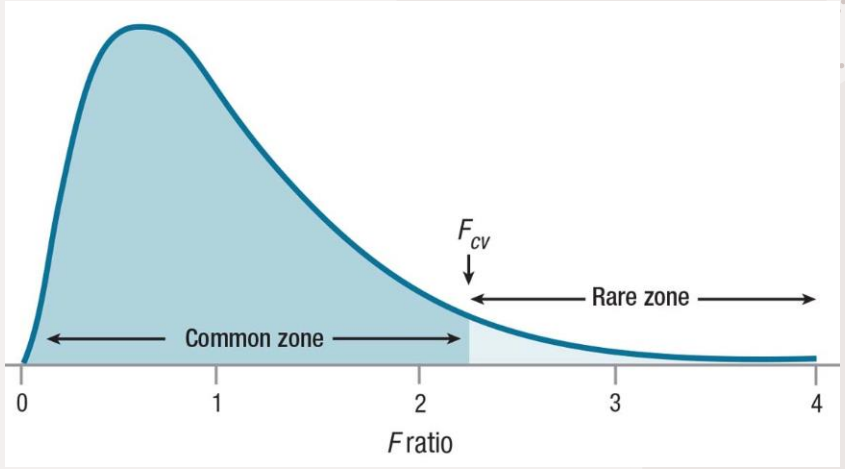
- The **sampling distribution** of a statistic is a distribution that demonstrates how a statistic could vary across all possible samples of a given size.

- **Example:** Sample Distribution of Means



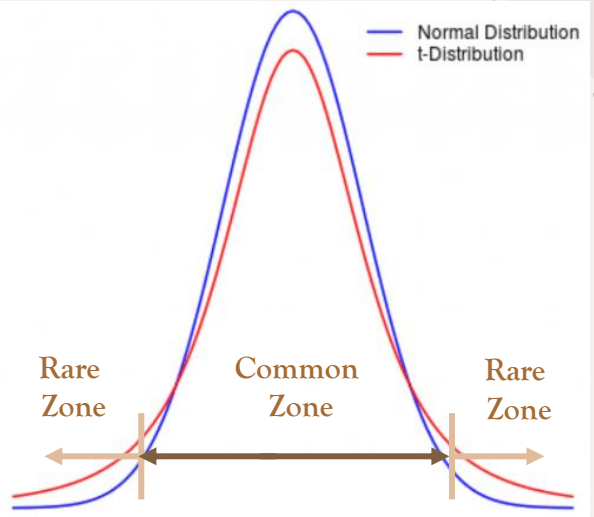
# Sampling Distributions

- A **sampling distribution** can be constructed for any type of statistic.
  - Sampling distribution of means
  - Sampling distribution of *F*-statistics
  - Sampling distribution of *t*-statistics, etc.




The **sampling distribution is used to represent the statistics that would be possible to obtain via sampling *if the null hypothesis is true*.**

- Key to how we test the **significance** of a statistic
- A *p*-value is **the probability of obtaining our particular test statistic** (*F*-statistic or *t*-statistic) **if the null hypothesis is true**



# Testing significance using model comparisons

**Step 1: Construct the Model C / Model A comparison**

- Model A includes all of the predictors
- Model C excludes *only* the predictor(s) we would like to test the significance of

**Step 2: State the null (and alternative) hypotheses**

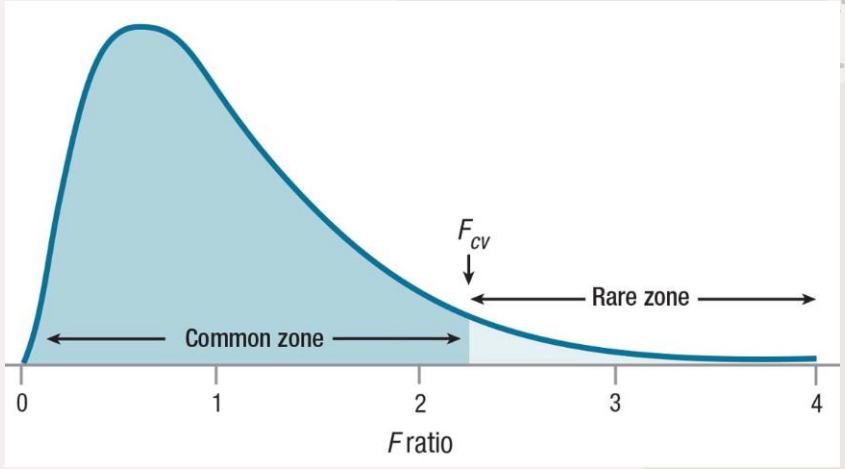
- The null hypothesis is particularly important – this is the hypothesis we actually end up testing!

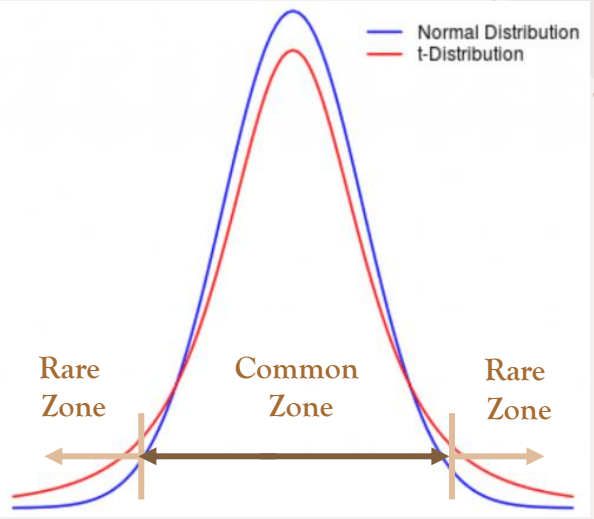
**Step 3: Estimate the parameters in Model C & Model A and evaluate each model's fit to the data:**

- SSE(C) & SSE(A)
- SSR = SSE(A) – SSE(C)

**Step 5a: Construct a sampling distribution representing the results one would theoretically expect to obtain *if the null hypothesis is true***

- Often using a *t*- or *F*- sampling distribution



# Testing significance using model comparisons

**Step 5b: State the researcher's willingness to make a Type I error, which determines *where* on the sampling distribution a test statistic must fall to be significant**

- Typically $\alpha = .05$

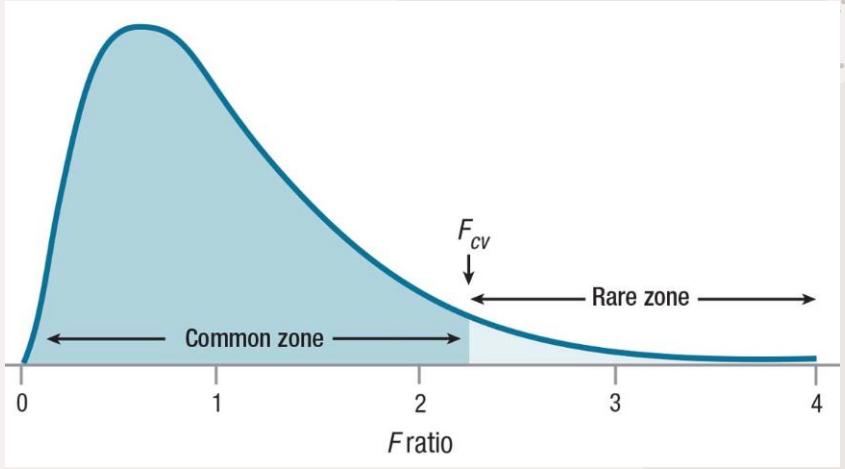
**Step 6: Calculate the test statistic**

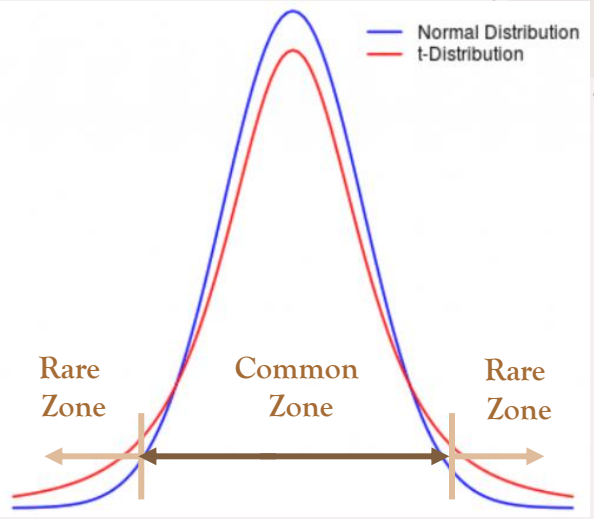
- $F = \dfrac{\text{additional variance explained by Model A}}{\text{variance left unexplained by Model A}}$
- $t = \dfrac{\text{parameter estimate}}{\text{standard error}}$

**Step 7: To assess significance, obtain**

- A *p*-value corresponding to the test statistic, and/or
- A confidence interval around the parameter estimate of interest

**Step 8: Calculate a measure of effect size**



# Linear Regression with a Single Continuous Predictor

# Linear regression with a single continuous predictor

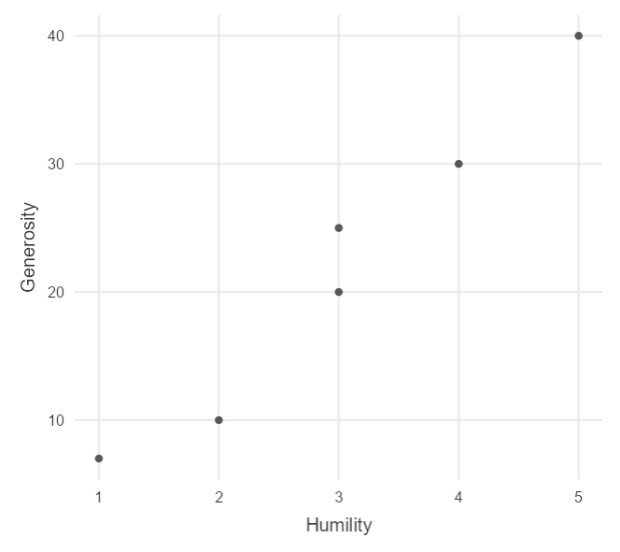
**Example:** Humility has been defined as a combination of having a willingness to see oneself accurately, an appreciation of others' strengths, and an openness to others' feedback. A researcher is interested in whether people's humility is a predictor of their generosity, or their willingness to help others.

Humility was measured on a 1 (strongly disagree) to 5 (strongly agree) scale using items like, "Even when I disagree with others, I can recognize that they have sound points." Generosity was measured based on how much money participants offered to donate to a local charity that was raising money to house people during winter. Participants' scores are shown in the table to the right.

| Humility | Generosity |
|---|---|
| 1 | 7 |
| 4 | 30 |
| 3 | 25 |
| 5 | 40 |
| 2 | 10 |
| 3 | 20 |
| $M_{Humility} = 3$ | $M_{Generosity} = 22$ |

# Data Visualization

- When fitting a linear model, it's important for the researcher to first *visualize* the relationships being tested to ensure they follow the pattern of a straight line




| Humility | Generosity |
|---|---|
| 1 | 7 |
| 4 | 30 |
| 3 | 25 |
| 5 | 40 |
| 2 | 10 |
| 3 | 20 |

$M_{Humility} = 3$ $M_{Generosity} = 22$

# Centering the continuous predictor

- Continuous predictors are often mean-centered prior to using them in a model

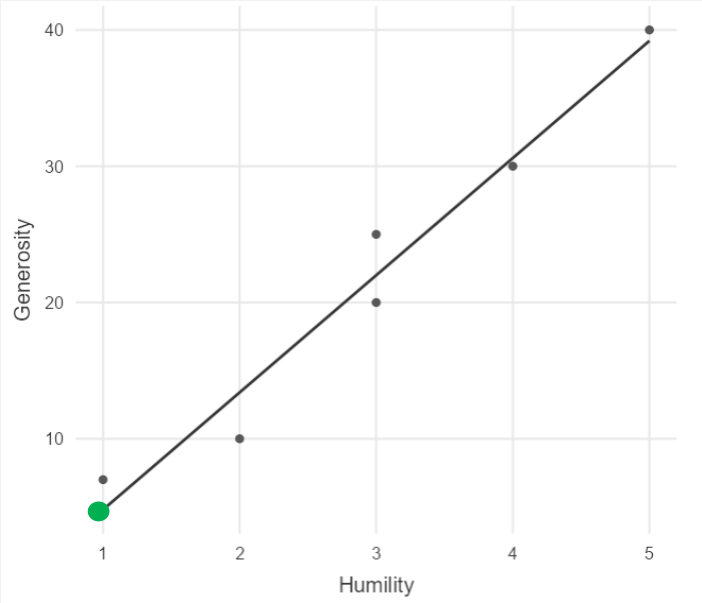
- To mean-center a predictor, simply subtract the mean of the predictor variable from each of the raw scores

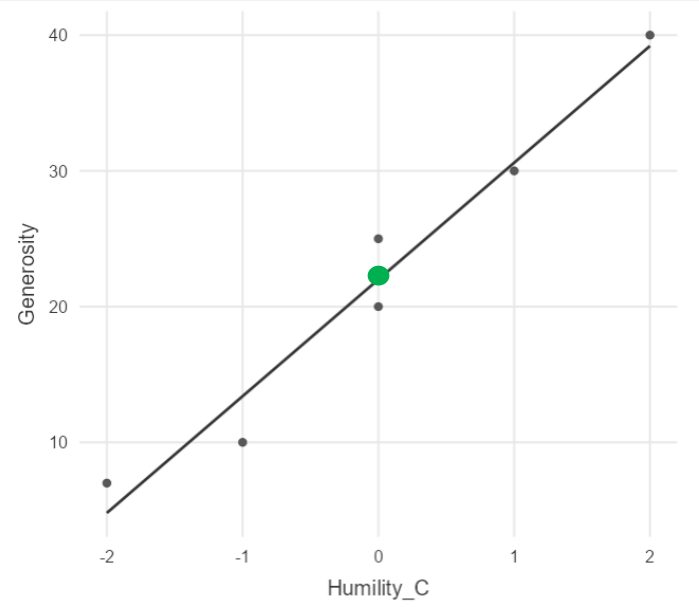
  - Expresses each person's score in terms of how far away it is from the mean in raw units

| Humility | Humility_C |
|---|---|
| 1 | -2 |
| 4 | 1 |
| 3 | 0 |
| 5 | 2 |
| 2 | -1 |
| 3 | 0 |
| $M_{Humility} = 3$ | $M_{Humility\_C} = 0$ |

# Centering the continuous predictor

- Centering shifts the predictor along the x-axis:
  - $b_0$ becomes the predicted score on Y for someone who scores equal to the *mean* of X
  - The meaning of $b_1$ does not change and the fit of the model does not change

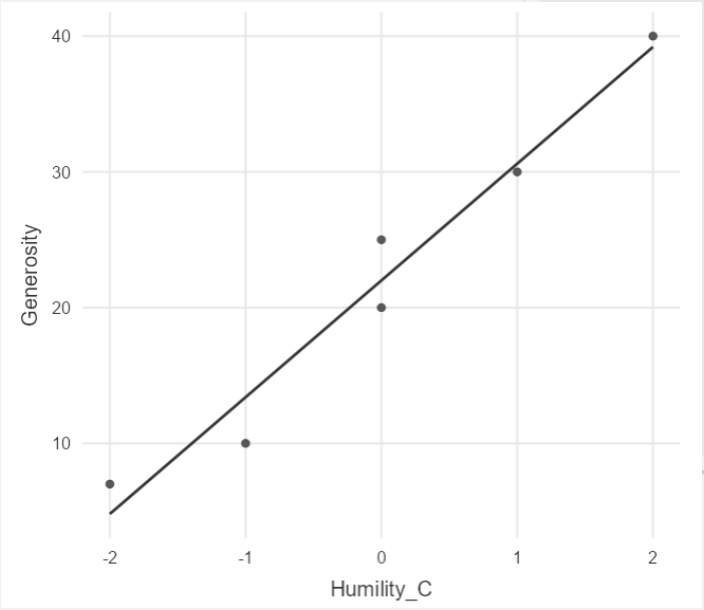


# Centering the continuous predictor

## When do you center a continuous predictor?

- When the researcher wants the y-intercept to occur at a meaningful value
- When there are multiple continuous predictors, and interaction effects between them, in the model.
  - Mean-centering the continuous predictors will reduce their redundancy with the interaction term – more to come on this!




If there is only a single continuous predictor in the model, it is up to the researcher.

- For this example, we'll use the centered predictor (Humility_C)

# Model Comparison

## Model Comparison:

Model C: $Y_i = \beta_0 + \varepsilon_i$ PC = 1

Model A: $Y_i = \beta_0 + \beta_1 Humility\_C_i + \varepsilon_i$ PA = 2

## Null & Alternative Hypotheses:

$H_0$: $\beta_1 = 0$

$H_1$: $\beta_1 \neq 0$

# Estimate the parameters in Model C

Model C: $Y_i = \beta_0 + \varepsilon_i$

- When only a single numerical value is used to predict scores on Y, the mean of Y does the best at minimizing SSE

Estimate of Model C:

- $Y_i = 22 + e_i$

| Humility_C | Generosity |
|---|---|
| -2 | 7 |
| 1 | 30 |
| 0 | 25 |
| 2 | 40 |
| -1 | 10 |
| 0 | 20 |

$M_{Humility\_C} = 0$ $M_{Generosity} = 22$

# Evaluate Fit of Model C

| $Y_i$ Generosity | $Y'$ Value predicted by model | $(Y_i - Y'_i)$ | $(Y_i - Y'_i)^2$ |
|---|---|---|---|
| 7 | 22 | -15 | 225 |
| 30 | 22 | 8 | 64 |
| 25 | 22 | 3 | 9 |
| 40 | 22 | 18 | 324 |
| 10 | 22 | -12 | 144 |
| 20 | 22 | -2 | 4 |

Estimate of Model C:

$Y_i = 22 + e_i$

$SSE(C) = \Sigma(Y_i - Y'_i)^2$

$SSE(C) = 770$

# Estimate the parameters in Model A

Model A: $Y_i = \beta_0 + \beta_1 \text{Humility\_C}_i + \varepsilon_i$

The values of $b_1$ and $b_0$ that best minimize SSE for a model with a single continuous predictor:

- $b_1 : \frac{\Sigma(x_i - M_x)(y_i - M_y)}{\Sigma(X - M_x)^2} = \frac{SP}{SS_x}$
- $b_0 : M_Y - b_1 M_X$

- *See derivation on page 78 of your textbook*

| Humility_C | Generosity |
|---|---|
| -2 | 7 |
| 1 | 30 |
| 0 | 25 |
| 2 | 40 |
| -1 | 10 |
| 0 | 20 |

$M_{\text{Humility\_C}} = 0$ $M_{\text{Generosity}} = 22$

# Estimate the parameters in Model A

| $X_i$ Humility_C | $Y_i$ Generosity | $(X_i - M_X)$ | $(Y_i - M_Y)$ | $(X_i - M_X)*(Y_i - M_Y)$ | $(X_i - M_X)^2$ |
|---|---|---|---|---|---|
| -2 | 7 | -2 | -15 | 30 | 4 |
| 1 | 30 | 1 | 8 | 8 | 1 |
| 0 | 25 | 0 | 3 | 0 | 0 |
| 2 | 40 | 2 | 18 | 36 | 4 |
| -1 | 10 | -1 | -12 | 12 | 1 |
| 0 | 20 | 0 | -2 | 0 | 0 |
| $M_X = 0$ | $M_Y = 22$ | | | SP = 86 | $SS_X = 10$ |

$$b_1 = \frac{\Sigma(x_i - M_x)(y_i - M_y)}{\Sigma(X - M_x)^2} = \frac{SP}{SS_x}$$

$$b_1 = \frac{86}{10} = 8.60$$

$$b_0 = M_Y - b_1 M_X$$

$$b_0 = 22 - (8.6*0) = 22$$

Estimate of Model A:

$$Y_i = 22 + 8.60*\text{Humility\_C}$$

# Evaluate Fit of Model A

| $Y_i$ Generosity | $X_i$ Humility_C | $Y_i' = 22 + 8.6*\text{Humility\_C}$ | $(Y_i - Y'_i)$ | $(Y_i - Y'_i)^2$ |
|---|---|---|---|---|
| 7 | -2 | 22 + (8.6*-2) = 4.8 | 2.2 | 4.84 |
| 30 | 1 | 22 + (8.6*1) = 30.6 | -0.6 | 0.36 |
| 25 | 0 | 22 + (8.6*0) = 22 | 3 | 9 |
| 40 | 2 | 22 + (8.6*2) = 39.2 | 0.8 | 0.64 |
| 10 | -1 | 22 + (8.6*-1) = 13.4 | -3.4 | 11.56 |
| 20 | 0 | 22 + (8.6*0) = 22 | -2 | 4 |

Estimate of Model A:

$Y_i = 22 + 8.60*\text{Humility\_C}$

$SSE(A) = \Sigma(Y_i - Y'_i)^2$

$SSE(A) = 30.4$

# Comparing Fit of Model A vs Model C

SSR is how much additional error is explained by Model A compared to Model C:

- SSR = SSE(C) – SSE(A)
- SSR = 770 – 30.4 = 739.6

PRE describes the reduction in error as a proportion:

- PRE = $\frac{SSR}{SSE(C)} = \frac{739.6}{770} = 0.96$

Model A explains 96% more of the variability in generosity scores compared to Model C.

# Summary Table

| Source | *SS* | *df* | *MS* | *F* | *p* |
|---|---|---|---|---|---|
| Reduced | SSR = SSE(C) – SSE(A) | $df_{\text{Reduced}}$ = PA - PC | MSR = $\frac{SSR}{df_{Reduced}}$ | $F = \frac{MSR}{MSE}$ | Use R to obtain |
| Model A | SSE(A) = $\Sigma(Y_i - Y'_i)^2$ | $df_{\text{ModelA}}$ = n - PA | MSE = $\frac{SSE(A)}{df_{ModelA}}$ | | |

# Summary Table

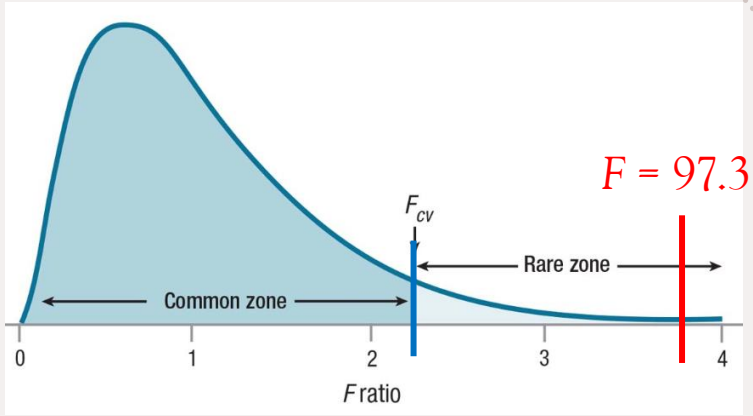
```
Analysis of Variance Table

Response: generosity
            Df Sum Sq Mean Sq F value    Pr(>F)
humility_c   1  739.6   739.6  97.316 0.0005924 ***
Residuals    4   30.4     7.6
---
```

| | *SS* | *df* | *MS* | *F* | *p* |
|---|---|---|---|---|---|
| Reduced | 739.6 | 2-1 = 1 | 739.6 | $F = \frac{739.6}{7.6} = 97.32$ | $p < .001$ |
| Model A | 30.4 | 6-2 = 4 | 7.6 | | |

The *F*-statistic indicates where we land on a sampling distribution of *F*-statistics that would be expected *if the null hypothesis is true*.

- The *p*-value indicates the probability of obtaining our results *if the null hypothesis is true*




$F = 97.32, p < .001$

$F_{CV}(1, 4) = 7.709$

# Effect Size

The significance (or non-significance) of a predictor does **not** communicate the **size**, or practical importance, of the predictor

- **Report the effect size** along with the significance of predictors

etaSquared() function provides:

- $\eta^2 = \frac{SSR}{SS_{Total}}$
  - The proportion of total variability in the DV that is related to the predictor variable(s)
- $\eta^2_{\text{Partial}} = \frac{SSR}{SSE(C)}$
  - The proportion of the variability in the DV that is related to the predictor variable(s) that was left unexplained by the other predictors in the model

For our example:

$\eta^2 = \frac{739.6}{770} = 0.96$

$\eta^2_{\text{Partial}} = \frac{739.6}{770} = 0.96$

- $\eta^2$ is equal to PRE when Model C is the one parameter model
- $\eta^2$ is equal to $\eta^2_{\text{Partial}}$ when there is only a single predictor in the model

# Performing the Analysis in R

Import (or set up) the data:

```r
generosity <- c(7,30,25,40,10,20)
humility <- c(1,4,3,5,2,3)
data <- cbind.data.frame(generosity, humility)
```

Center the continuous predictor (if desired):

```r
data$humility_c <- scale(humility, center = TRUE, scale = FALSE)
```

Fit the model:

```r
model_center <- lm(generosity ~ humility_c, data = data)
```

```r
model_uncenter <- lm(generosity ~ humility, data = data)
```

# Performing the Analysis in R

Output (Centered predictor):

```
> anova(model_center)
Analysis of Variance Table

Response: generosity
            Df Sum Sq Mean Sq F value    Pr(>F)
humility_c   1  739.6   739.6  97.316 0.0005924 ***
Residuals    4   30.4     7.6
```

```
> summary(model_center)

Call:
lm(formula = generosity ~ humility_c, data = data)

Residuals:
   1    2    3    4    5    6
 2.2 -0.6  3.0  0.8 -3.4 -2.0

Coefficients:
            Estimate Std. Error t value Pr(>|t|)
(Intercept)  22.0000     1.1255  19.548 4.04e-05 ***
humility_c    8.6000     0.8718   9.865 0.000592 ***
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 2.757 on 4 degrees of freedom
Multiple R-squared:  0.9605,    Adjusted R-squared:  0.9506
F-statistic: 97.32 on 1 and 4 DF,  p-value: 0.0005924
```

Output (Uncentered predictor):

```
> anova(model_uncenter)
Analysis of Variance Table

Response: generosity
           Df Sum Sq Mean Sq F value    Pr(>F)
humility    1  739.6   739.6  97.316 0.0005924 ***
Residuals   4   30.4     7.6
```

```
> summary(model_uncenter)

Call:
lm(formula = generosity ~ humility, data = data)

Residuals:
   1    2    3    4    5    6
 2.2 -0.6  3.0  0.8 -3.4 -2.0

Coefficients:
            Estimate Std. Error t value Pr(>|t|)
(Intercept)  -3.8000     2.8472  -1.335 0.252899
humility      8.6000     0.8718   9.865 0.000592 ***
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 2.757 on 4 degrees of freedom
Multiple R-squared:  0.9605,    Adjusted R-squared:  0.9506
F-statistic: 97.32 on 1 and 4 DF,  p-value: 0.0005924
```

# 95% Confidence Interval

- $b_1$ represents our best estimate of the relationship between the predictor and outcome variable
  - However, $b_1$ is only a **point estimate** of $\beta_1$.
    - $b_1$ is unlikely to be exactly equal to $\beta_1$ due to sampling error.
  - Instead, we can construct a 95%CI around $b_1$ to represent the **range of values** within which we are more confident the true value of $\beta_1$ may exist:

$$95\%CI = b_1 \pm \sqrt{\frac{F_{CV} * MSE}{SS_x}}$$

```
> confint(model_center)
                 2.5 %   97.5 %
(Intercept) 18.875214 25.12479
humility_c   6.179551 11.02045
```

- $b_1$ is our point estimate
- $F_{CV}$ is the *F*-critical value corresponding to (1, *n*-PA) degrees of freedom and $\alpha = .05$
- MSE = $\frac{SSE(A)}{n - PA}$
- $SS_X = \Sigma(X - M_x)^2$

**For our example:**

- Lower bound: $8.6 - \sqrt{\frac{7.709*7.6}{10}} = 6.18$
- Upper bound: $8.6 + \sqrt{\frac{7.709*7.6}{10}} = 11.02$

# Difference between correlation and linear regression?

What would the correlation between humility and generosity be if we calculated it?

- Why isn't $r$ equal to $b_1$ for a model with a single continuous predictor?

```
> cor(data$humility, data$generosity)
[1] 0.980061
```

The correlation describes the ***standardized*** linear relationship between two continuous variables.

- If you standardize (i.e., transform into z-scores) the raw data before performing a linear regression analysis with a single continuous predictor, $b_1$ will equal the zero-order correlation

```r
data$generosity_std <- scale(generosity, center = TRUE, scale = TRUE)
data$humility_std <- scale(humility, center = TRUE, scale = TRUE)

model_std <- lm(generosity_std ~ humility_std, data = data)
```

- Recall: $z = \frac{X - M_X}{SD}$

**Q:** How would you interpret the meaning of $b_1$ now?

```
Coefficients:
                          Estimate            Std. Error t value Pr(>|t|)
(Intercept)  0.00000000000000008884 0.09069238239845851812   0.000 1.000000
humility_std 0.98006095755288613613 0.09934852726704040959   9.865 0.000592

(Intercept)
humility_std ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 0.2222 on 4 degrees of freedom
Multiple R-squared:  0.9605,    Adjusted R-squared:  0.9506
F-statistic: 97.32 on 1 and 4 DF,  p-value: 0.0005924
```