



Interpretation

Lecture 12

Louis SIRUGUE

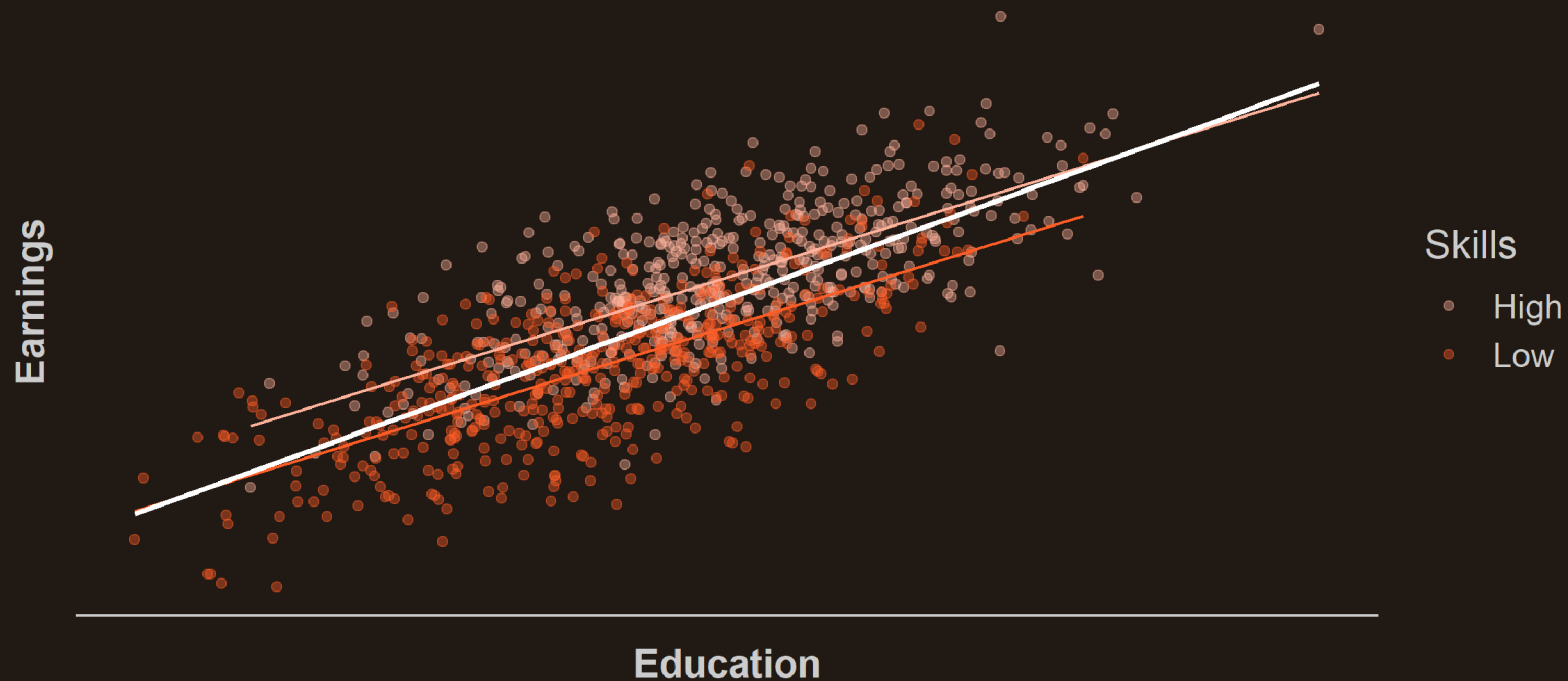
CPES 2 - Fall 2023



Quick reminder

Omitted variable bias

- If a third **variable** is correlated with both x and y , it would **bias the relationship**
 - We must then **control** for such variables
 - And if we can't we must acknowledge that our estimate is not causal with '*ceteris paribus*'

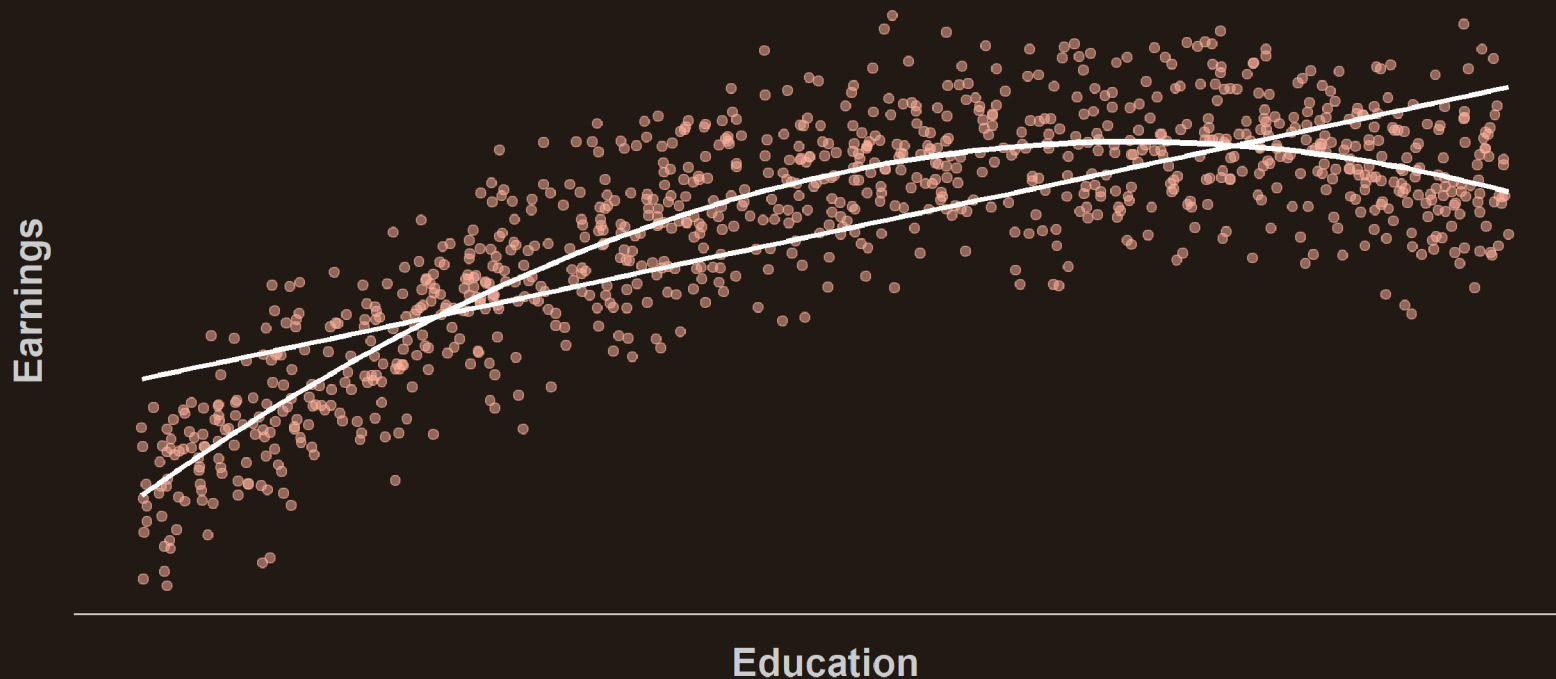




Quick reminder

Functional form

- Not capturing the **right functional** form correctly might also lead to biased estimations:
 - Polynomial order, interactions, logs, discretization matter
 - **Visualizing the relationship** is key





Quick reminder

Selection bias

- **Self-selection** is also a common threat to causality
- What is the impact of going to a better neighborhood on your children outcomes?
 - We cannot just regress children outcomes on a mobility dummy
 - Individuals who move may be different from those who stay: **self-selection issue**
 - Here it is not that the sample is not representative of the population, but that **the outcomes of those who stayed are different from the outcomes those who moved would have had, if they had stayed**

Simultaneity

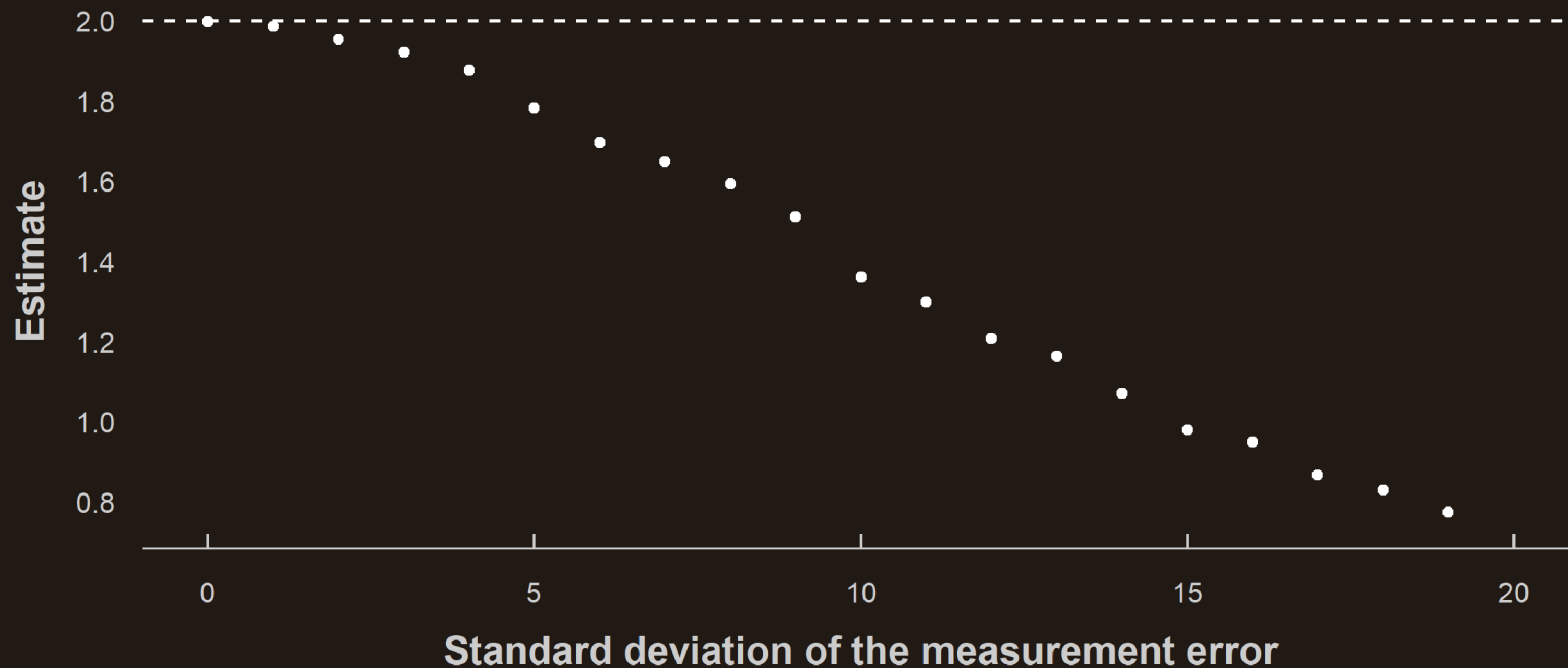
- Consider the relationship between **crime** rate and **police coverage** intensity
- What is the **direction of the relationship?**
 - We cannot just regress crime rate on police intensity
 - It's likely that more crime would cause a positive response in police activity
 - And also that police activity would deter crime



Quick reminder

Measurement error

- **Measurement error** in the independent variable also induces a bias
 - The resulting estimation would mechanically be **downward biased**
 - The **noisier** the measure, the **larger the bias**



Quick reminder

Randomized Controlled Trials

- A Randomized Controlled Trial (RCT) is a type of experiment in which the thing we want to know the impact of (called the treatment) is **randomly allocated** in the population
 - The two **groups** would then have the same characteristics on expectation, and would be **comparable**
 - It is a way to obtain **causality** from randomness
- RCTs are very **powerful tools** to sort out issues of:
 - Omitted variables
 - Selection bias
 - Simultaneity
- But RCTs are **not immune** to every problem:
 - The sample must be representative and large enough
 - Participants should comply with their treatment status
 - Independent variables must not be noisy measures of the variable of interest
 - ...



Today: Interpretation

1. Point estimates

- 1.1. Continuous variables
- 1.2. Discrete variables
- 1.3. Log vs. level

2. Practice interpretation

3. Regression tables

- 3.1. Layout
- 3.2. Reported significance
- 3.3. R squared

4. Wrap up!



Today: Interpretation

1. Point estimates

- 1.1. Continuous variables
- 1.2. Discrete variables
- 1.3. Log vs. level

1. Point estimates

1.1. Continuous variables

- In this first part, we're going to consider the **relationship** between:
 - The **income level** of young parents
 - The **health** of their **newborn**
- Consider first the following specification of the two variables:
 - A continuous measure of **annual household income in euros**
 - A continuous measure of **birth weight in grams**

$$\text{Birth weight}_i = \alpha + \beta \times \text{Household income}_i + \varepsilon_i$$

```
lm(birth_weight ~ household_income, data)$coefficients
```

```
##      (Intercept) household_income
## 3.134528e+03      2.213871e-03
```

→ How would you interpret $\hat{\beta}$ here? (Note that e+03 and e-03 mean $\times 10^3$ and $\times 10^{-3}$)

1. Point estimates

1.1. Continuous variables

- When both x and y are continuous, the **general** template for the **interpretation** of $\hat{\beta}$ is:

"Everything else equal, a 1 [unit] increase in [x] is associated with an [in/de]crease of [beta] [units] in [y] on average."

- So in our case the **adequate interpretation** would be:

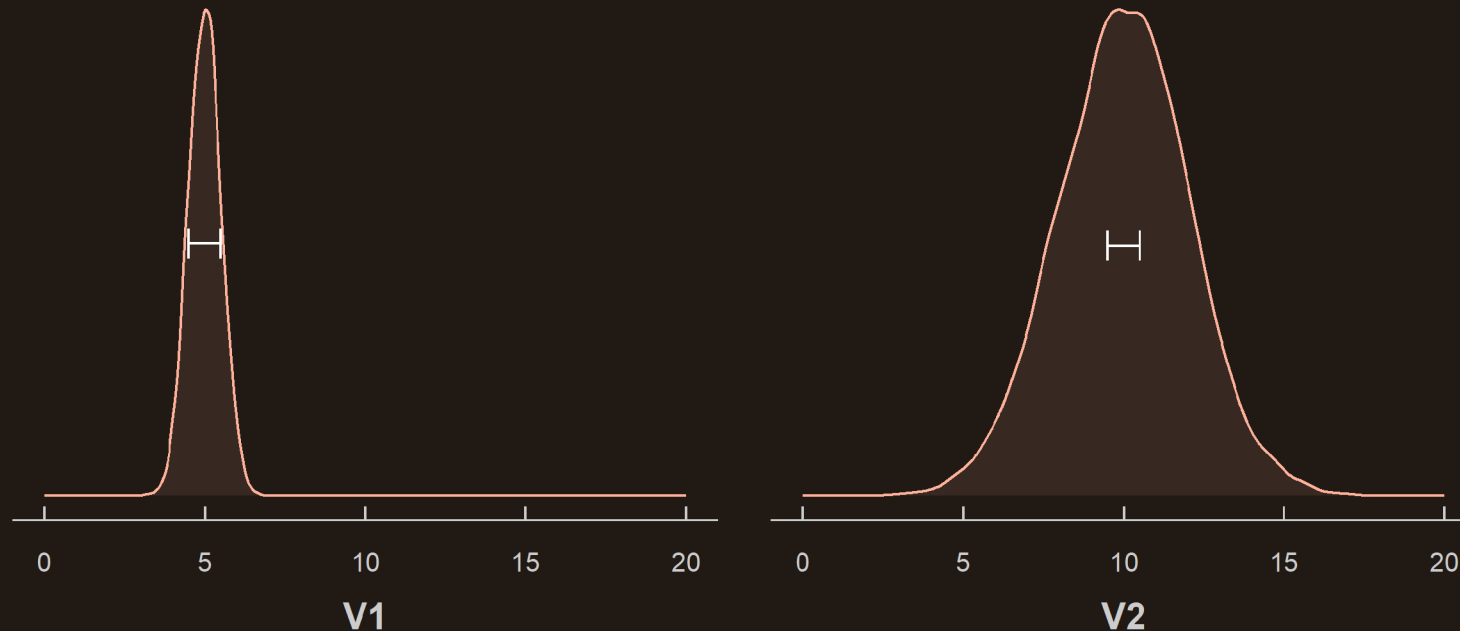
*"Everything else equal, a **1 euro increase in annual household income** is associated with an **increase of 0.002 gram in newborn birth weight** on average."*

- But it would be even better to **interpret** the results for a **meaningful variation** of x
 - For an annual household income, a **1 euro variation** is not really meaningful
 - 1 euro increase \rightarrow 0.002 gram increase \Leftrightarrow **1,000 euro increase** \rightarrow 2 gram increase

1. Point estimates

1.1. Continuous variables

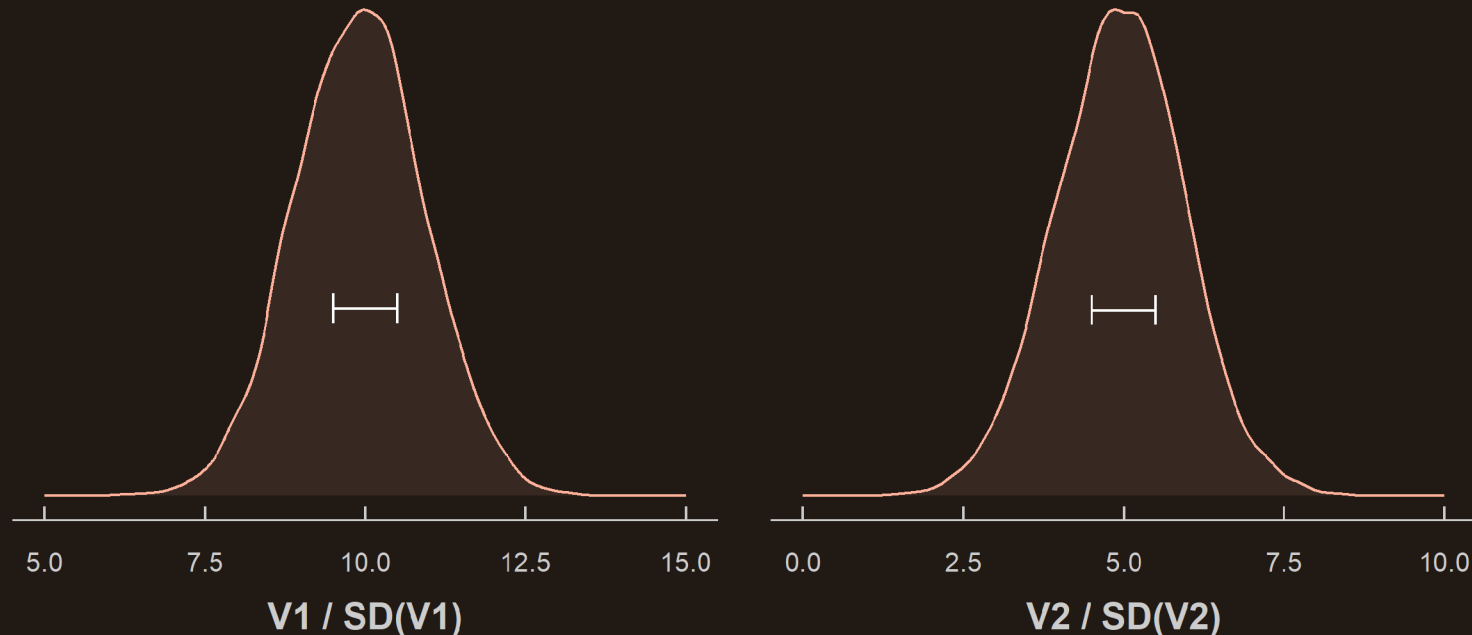
- A common way to obtain a coefficient for a **meaningful variation** of x is to **standardize** x
 - If we divide x by $SD(x)$, the 1 **unit increase** in $\frac{x}{SD(x)}$ is equivalent to an $SD(x)$ **increase** in x
 - An $SD(x)$ change in x is meaningful: it's low if x is very concentrated and high if x is highly spread out



1. Point estimates

1.1. Continuous variables

- A common way to obtain a coefficient for a **meaningful variation** of x is to **standardize** x
 - If we divide x by $SD(x)$, the 1 **unit increase** in $\frac{x}{SD(x)}$ is equivalent to an $SD(x)$ **increase** in x
 - An $SD(x)$ change in x is meaningful: it's low if x is very concentrated and high if x is highly spread out



1. Point estimates

1.1. Continuous variables

- Note that if you **standardize both x and y** , the resulting $\hat{\beta}$ equals the **correlation** between x and y
 - To show that, let's first rewrite the formula of the beta coefficient:

$$\hat{\beta} = \frac{\text{Cov}(x, y)}{\text{Var}(x)} = \frac{\text{Cov}(x, y)}{\text{SD}(x) \times \text{SD}(x)}$$

$$\hat{\beta} = \frac{\text{Cov}(x, y)}{\text{SD}(x) \times \text{SD}(x)} \times \frac{\text{SD}(y)}{\text{SD}(y)}$$

$$\hat{\beta} = \frac{\text{Cov}(x, y)}{\text{SD}(x) \times \text{SD}(y)} \times \frac{\text{SD}(y)}{\text{SD}(x)}$$

$$\hat{\beta} = \text{Cor}(x, y) \times \frac{\text{SD}(y)}{\text{SD}(x)}$$

1. Point estimates

1.1. Continuous variables

- Starting with the previous expression, the $\hat{\beta}$ **coefficient with the standardized variables** writes:

$$\hat{\beta} = \frac{\text{Cov}\left(\frac{x}{\text{SD}(x)}, \frac{y}{\text{SD}(y)}\right)}{\text{SD}\left(\frac{x}{\text{SD}(x)}\right) \times \text{SD}\left(\frac{y}{\text{SD}(y)}\right)} \times \frac{\text{SD}\left(\frac{y}{\text{SD}(y)}\right)}{\text{SD}\left(\frac{x}{\text{SD}(x)}\right)}$$

- But by construction, the standard deviation of a standardized variable is 1:

$$\hat{\beta} = \frac{\text{Cov}\left(\frac{x}{\text{SD}(x)}, \frac{y}{\text{SD}(y)}\right)}{1 \times 1} \times \frac{1}{1}$$

$$\hat{\beta} = \text{Cov}\left(\frac{x}{\text{SD}(x)}, \frac{y}{\text{SD}(y)}\right)$$

$$\hat{\beta} = \frac{\text{Cov}(x, y)}{\text{SD}(x) \times \text{SD}(y)} = \text{Cor}(x, y)$$

- Learn the cheatsheet on moments properties:

$$\begin{aligned} \text{Var}(aX) \\ = \\ a^2 \text{Var}(X) \end{aligned}$$

1. Point estimates

1.2. Discrete variables

- Consider the following specification of the two variables:
 - A categorical variable for **annual household income divided in terciles**
 - Still a continuous measure of **birth weight in grams**

$$\text{Birth weight}_i = \alpha + \beta_1 T2_i + \beta_2 T3_i + \varepsilon_i$$

- Recall that when including a **categorical variable** in a regression, a **reference category** must be **omitted**

```
lm(birth_weight ~ income_tercile, data)$coefficients
```

```
##      (Intercept) income_tercileT2 income_tercileT3
##      3112.83162          88.24778          222.65414
```

→ How would you interpret $\hat{\beta}_1$ and $\hat{\beta}_2$ here?

1. Point estimates

1.2. Discrete variables

- With a discrete x , the interpretation of the coefficient must be **relative to the reference category**:

"Everything else equal, belonging to the [x category] is associated with a [beta] [unit] [higher/lower] average [y] relative to the [reference category]."

- So in our case, the **adequate interpretations** would be:

*"Everything else equal, belonging to the **second income tercile** is associated with a **88 grams higher average birth weight** relative to the **first income tercile**."*

*"Everything else equal, belonging to the **third income tercile** is associated with a **223 grams higher average birth weight** relative to the **first income tercile**."*

- And the **intercept** is the **average birth weight** for newborns to parents in the **first income tercile**

1. Point estimates

1.2. Discrete variables

- Consider now the following specification of the two variables:
 - A continuous measure of **annual household income in euros**
 - A **binary variable** taking the value 1 if **the newborn is underweight** and 0 otherwise

$$\text{Underweight}_i = \alpha + \beta \times \text{Household income}_i + \varepsilon_i$$

```
lm(underweight ~ household_income, data)$coefficients
```

```
##      (Intercept) household_income
## 5.214013e-02    -4.084787e-07
```

→ How would you interpret $\hat{\beta}$ here?

→ And would you consider its magnitude high?

1. Point estimates

1.2. Discrete variables

- With a **binary y variable**, the coefficient must be interpreted in **percentage points**:

"Everything else equal, a 1 [unit] increase in [x] is associated with a $[\text{beta} \times 100]$ percentage point [in/de]crease in the probability that [y equals 1] on average."

- So in our case, the **adequate interpretation** would be:

*"Everything else equal, a **1 euro increase in annual household income** is associated with a **0.00004 percentage point decrease in the probability that the newborn is underweight** on average."*

- Here the **interpretation** would be more **meaningful**:
 - For a **1,000 euro** increase → 0.04 percentage point decrease
 - Compared to the **typical probability** to have an underweight newborn



1. Point estimates

1.2. Discrete variables

- The mean of a dummy variable corresponds to the share of 1s:

```
mean(data$underweight)
```

```
## [1] 0.037
```

- We can also compute the probability that $y = 1$ for the average x with our estimated coefficients:

```
5.214013e-02 + mean(data$household_income) * -4.084787e-07
```

```
## [1] 0.03700001
```

For the **average household**, a **1,000 euro increase** in annual income would be associated with a **0.0004 / 0.037 \approx 1% decrease in the probability** that the newborn is **underweight**

1. Point estimates

1.2. Discrete variables

- Finally if **both** the y and the x variables are **discrete**, the coefficient must be interpreted:
 - In **percentage points**
 - **Relative** to the reference category

$$\text{Underweight}_i = \alpha + \beta_1 T2_i + \beta_2 T3_i + \varepsilon_i$$

```
lm(underweight ~ income_tercile, data)$coefficients
```

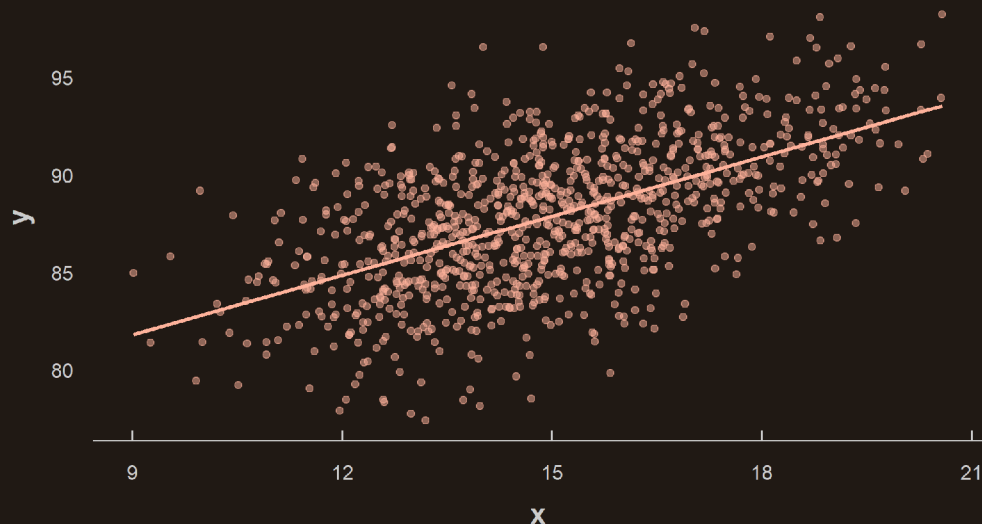
```
##      (Intercept) income_tercileT2 income_tercileT3
##      0.07207207      -0.03903904      -0.06608405
```

*"Everything else equal, belonging to the **second income tercile** is associated with a **3.9 percentage point lower probability that the newborn is underweight** relative to the **first income tercile**."*

1. Point estimates

1.3. Log vs. level

- Consider now the following hypothetical relationship:



- The slope tells us by how many **units** the y variable would increase for a **1 unit** increase in x
- But often times in Economics we're interested in the elasticity between the two variables:
 - What is the expected **percentage change** in y for a **one percent increase** in x ?

→ *The log transformation can be used to easily get an approximation of that*



1. Point estimates

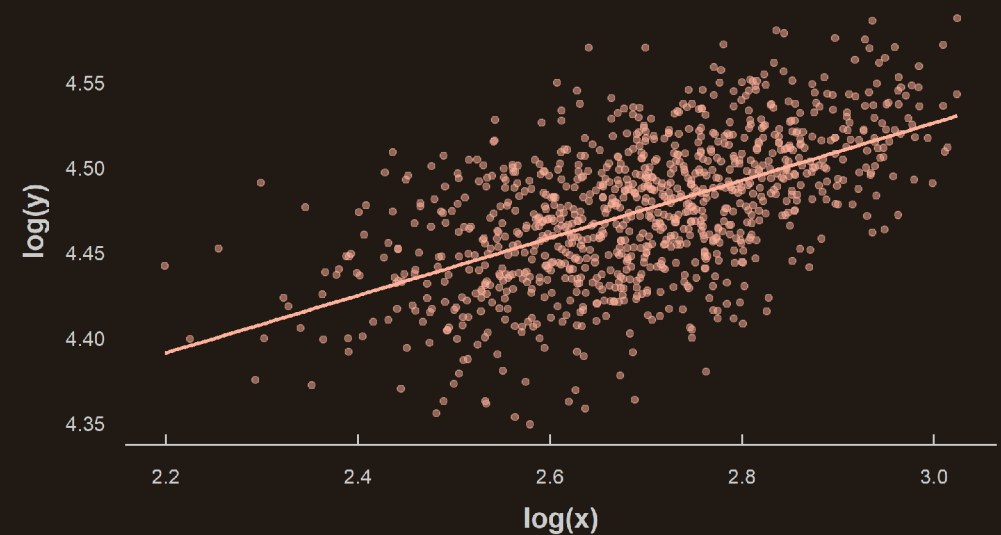
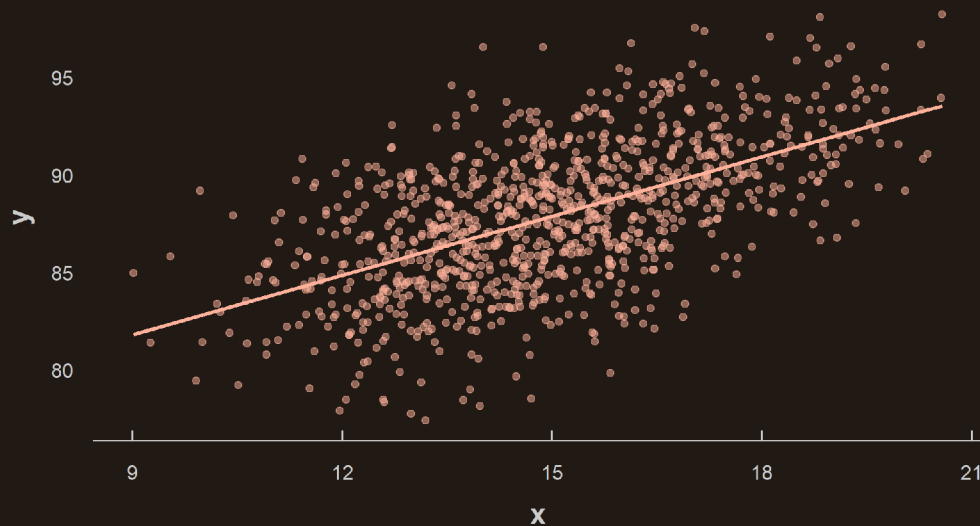
1.3. Log vs. level

- Instead of considering

$$y_i = \alpha_{lvl} + \beta_{lvl}x_i + \varepsilon_i$$

- We consider

$$\log(y_i) = \alpha_{log} + \beta_{log} \log(x_i) + \varepsilon_i$$

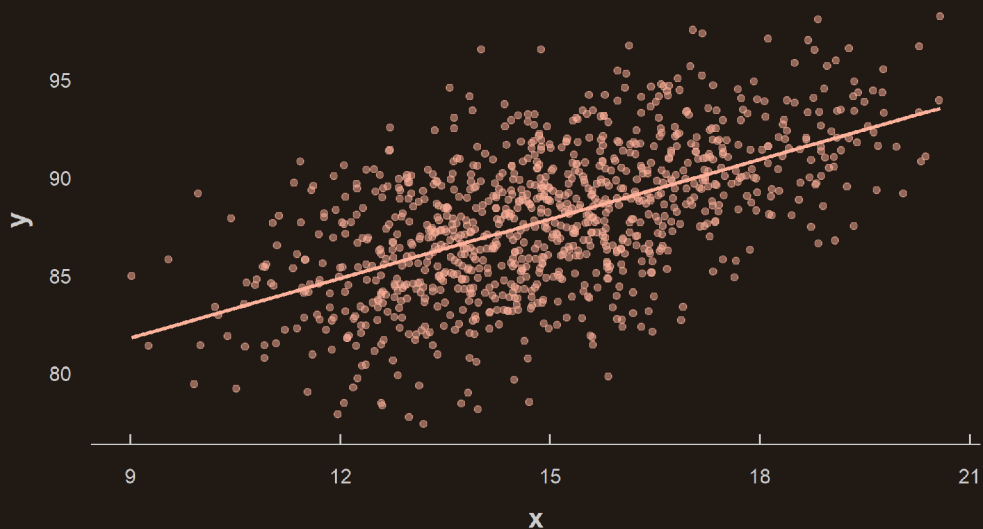




1. Point estimates

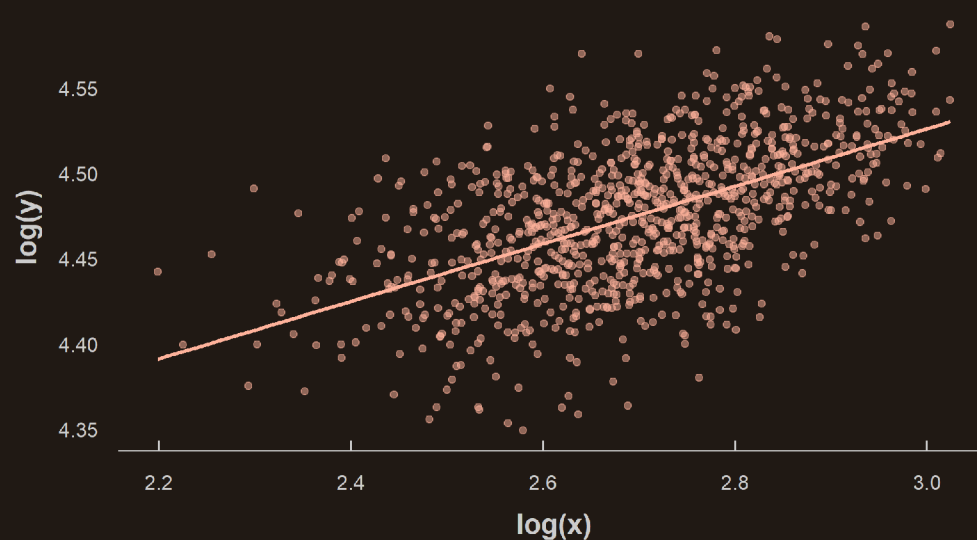
1.3. Log vs. level

$$\widehat{\beta}_{lvl} = 1.0121933$$



$$(15 \div 100) \times \widehat{\beta}_{lvl} \approx (15 \div 100) \times 1.0121933 \\ \approx 0.151829$$

$$\widehat{\beta}_{log} = 0.16875$$



$$0.151829 \div 90 = 0.0016870 \\ \approx \beta_{log} \%$$

1. Point estimates

1.3. Log vs. level

- Thus the interpretation differs depending on whether variables are in log or in level:
 - When variables are in **level** we should interpret the coefficients in terms of **unit** increase
 - When variables are in **log** we should interpret the coefficients in terms of **percentage** increase

Interpretation of the regression coefficient

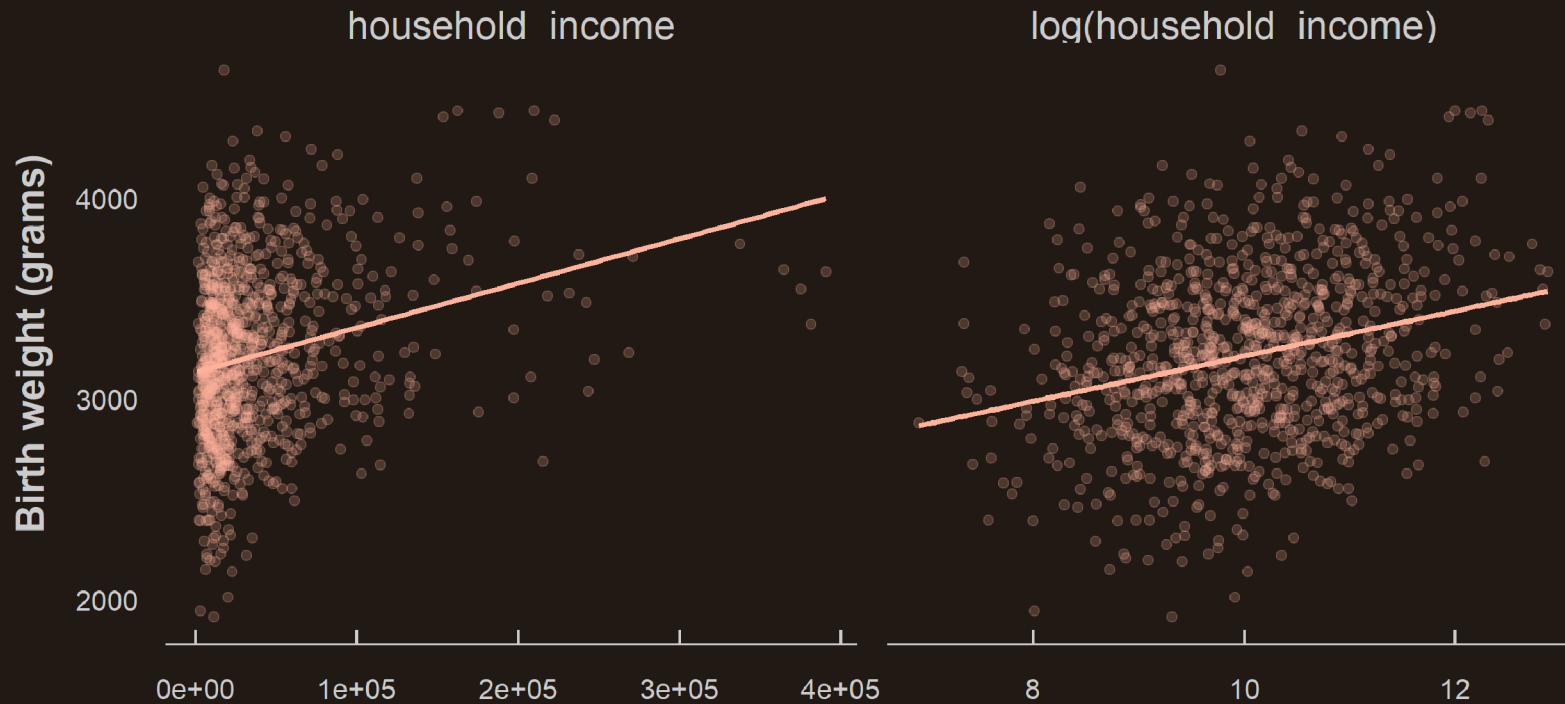
	y	log(y)
x	$\hat{\beta}$ is the unit increase in y due to a 1 unit increase in x	$\hat{\beta} \times 100$ is the % increase in y due to a 1 unit increase in x
log(x)	$\hat{\beta} \div 100$ is the unit increase in y due to a 1% increase in x	$\hat{\beta}$ is the % increase in y due to a 1% increase in x



1. Point estimates

1.3. Log vs. level

- Let's give it a try with our example on household income and birth weight
 - We've already seen that because income is log-normally distributed, it should be included in log



1. Point estimates

1.3. Log vs. level

- So what would be your interpretation of the slope estimated from the following regression?

$$\text{Birth weight}_i = \alpha + \beta \log(\text{Household income}_i) + \varepsilon$$

```
lm(birth_weight ~ log(household_income), data)$coefficients
```

```
##           (Intercept) log(household_income)
##           2091.2323           112.3234
```

- With a continuous y in **level** and a **logged** x variable, the template would be:

"Everything else equal, a 1 percent increase in [x] is associated with a [beta/100] [unit] [in/de]crease in [y] on average."

- So in our case, the **adequate interpretation** would be:

*"Everything else equal, a **1 percent increase in annual household income** is associated with a **1.12 grams increase in the birth weight of the newborn** on average."*

Overview



1. Point estimates ✓

- 1.1. Continuous variables
- 1.2. Discrete variables
- 1.3. Log vs. level

2. Practice interpretation

3. Regression tables

- 3.1. Layout
- 3.2. Reported significance
- 3.3. R squared

4. Wrap up!

Overview



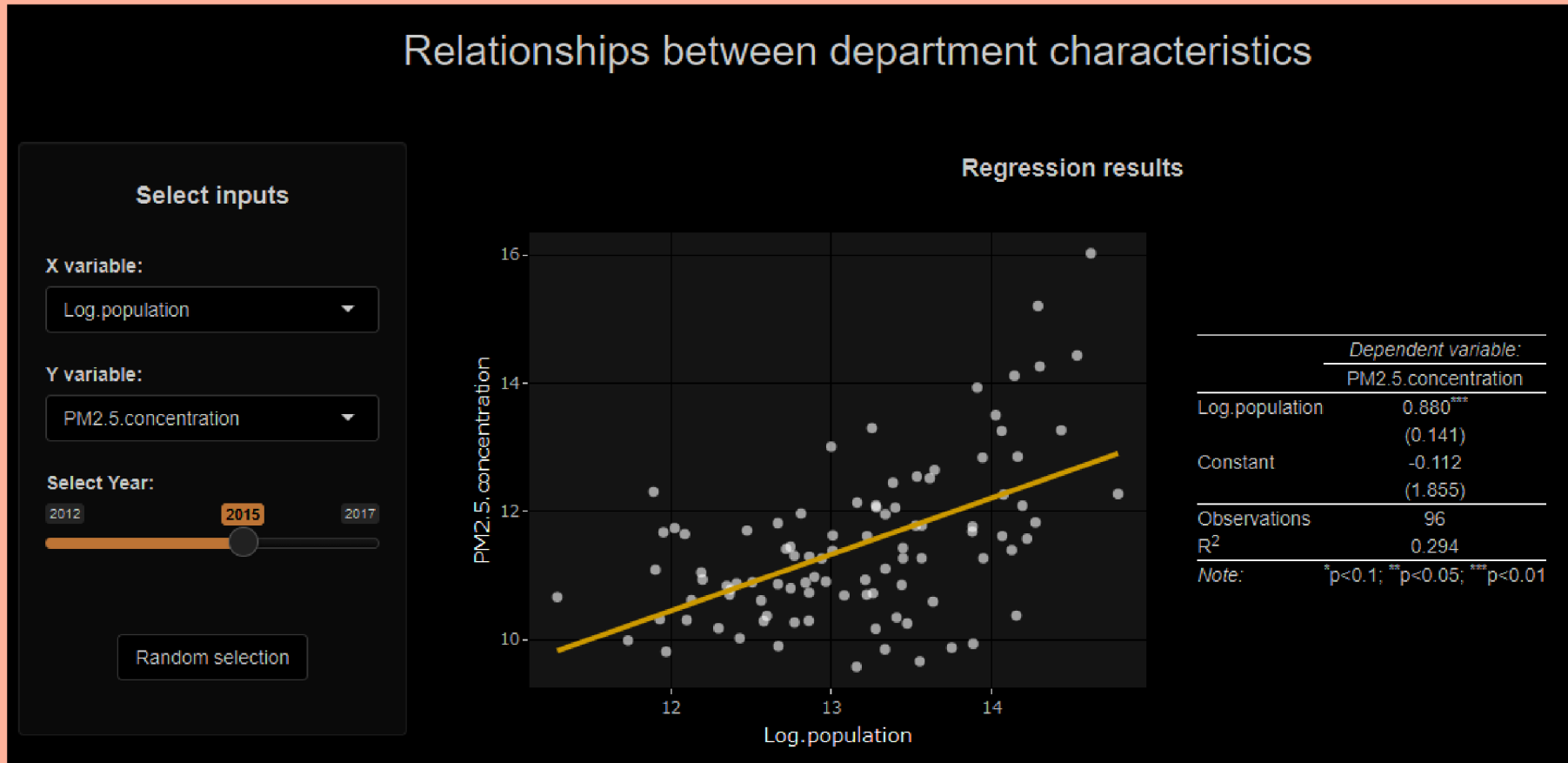
1. Point estimates ✓

- 1.1. Continuous variables
- 1.2. Discrete variables
- 1.3. Log vs. level

2. Practice interpretation

2. Practice interpretation

→ Let's practice coefficient interpretation with randomly generated relationships:



Overview



1. Point estimates ✓

- 1.1. Continuous variables
- 1.2. Discrete variables
- 1.3. Log vs. level

2. Practice interpretation ✓

3. Regression tables

- 3.1. Layout
- 3.2. Reported significance
- 3.3. R squared

4. Wrap up!

Overview



1. Point estimates ✓

- 1.1. Continuous variables
- 1.2. Discrete variables
- 1.3. Log vs. level

2. Practice interpretation ✓

3. Regression tables

- 3.1. Layout
- 3.2. Reported significance
- 3.3. R squared

3. Regression tables

3.1. Layout

- So far we've been **used to** regression results **displayed this way:**

```
lm(birth_weight ~ household_income, data)$coefficients
```

```
##      (Intercept) household_income
## 3.134528e+03    2.213871e-03
```

- Or with the more exhaustive **summary()** coefficients output:

```
summary(lm(birth_weight ~ household_income, data))$coefficients
```

```
##              Estimate  Std. Error  t value  Pr(>|t|)
## (Intercept) 3.134528e+03 1.656840e+01 189.187165 0.000000e+00
## household_income 2.213871e-03 2.808507e-04  7.882732 8.355367e-15
```

→ But in **formal** reports and academic papers, the **layout** of regression tables is **a bit different**

3. Regression tables

3.1. Layout

	<i>Dependent variable:</i>	
	Birth weight	
	(1)	(2)
Household income	0.002 ^{***} (0.0003)	0.002 ^{***} (0.0003)
Girl (ref: Boy)		-135.218 ^{***} (34.838)
Constant	3,134.528 ^{***} (16.568)	3,246.365 ^{***} (34.257)
Observations	1,000	963
<i>Note:</i>	* p<0.1; ** p<0.05; *** p<0.01	

Regression tables often contain multiple regressions:

- With **one regression in each column**
 - Regression models are numbered
 - Dependent variable mentioned above
- And one variable in **each row**
 - With the **point estimate**
 - And a **precision measure** below
- **General info** on each model **at the bottom**
- A **symbology** for the **p-value** testing whether the coefficient is significantly different from 0 or not

3. Regression tables

3.1. Layout

	<i>Dependent variable:</i>	
	Birth weight	
	(1)	(2)
Household income	0.002 ^{***} (0.0003)	0.002 ^{***} (0.0003)
Girl (ref: Boy)		-135.218 ^{***} (34.838)
Constant	3,134.528 ^{***} (16.568)	3,246.365 ^{***} (34.257)
Observations	1,000	963

Note: * p<0.1; ** p<0.05; *** p<0.01

It makes it easy to compare the different models:

- We can **add controls progressively**
 - Check the **stability** of the main **coefficient**

→ *If it gets significantly closer to 0 it might indicate that the raw relationship was fallaciously driven by a confounding factor*

- And **compare general statistics**
 - N is lower in the second regression
 - It means that there are missing values
 - Could this induce a selection bias?

3. Regression tables

3.2. Reported significance

	<i>Dependent variable:</i>	
	Birth weight	
	(1)	(2)
Household income	0.002 ^{***} (0.0003)	0.002 ^{***} (0.0003)
Girl (ref: Boy)		-135.218 ^{***} (34.838)
Constant	3,134.528 ^{***} (16.568)	3,246.365 ^{***} (34.257)
Observations	1,000	963
<i>Note:</i>	* p<0.1; ** p<0.05; *** p<0.01	

It makes it easy to compare the different models:

- The **evolution** of the **significance** matters as well
 - The main coefficient should stay significant
- But don't rely too much on the **symbology**
 - Thresholds are **not always the same**
 - **Sometimes** there are **none**
- Instead, keep in mind this **rule of thumb:**

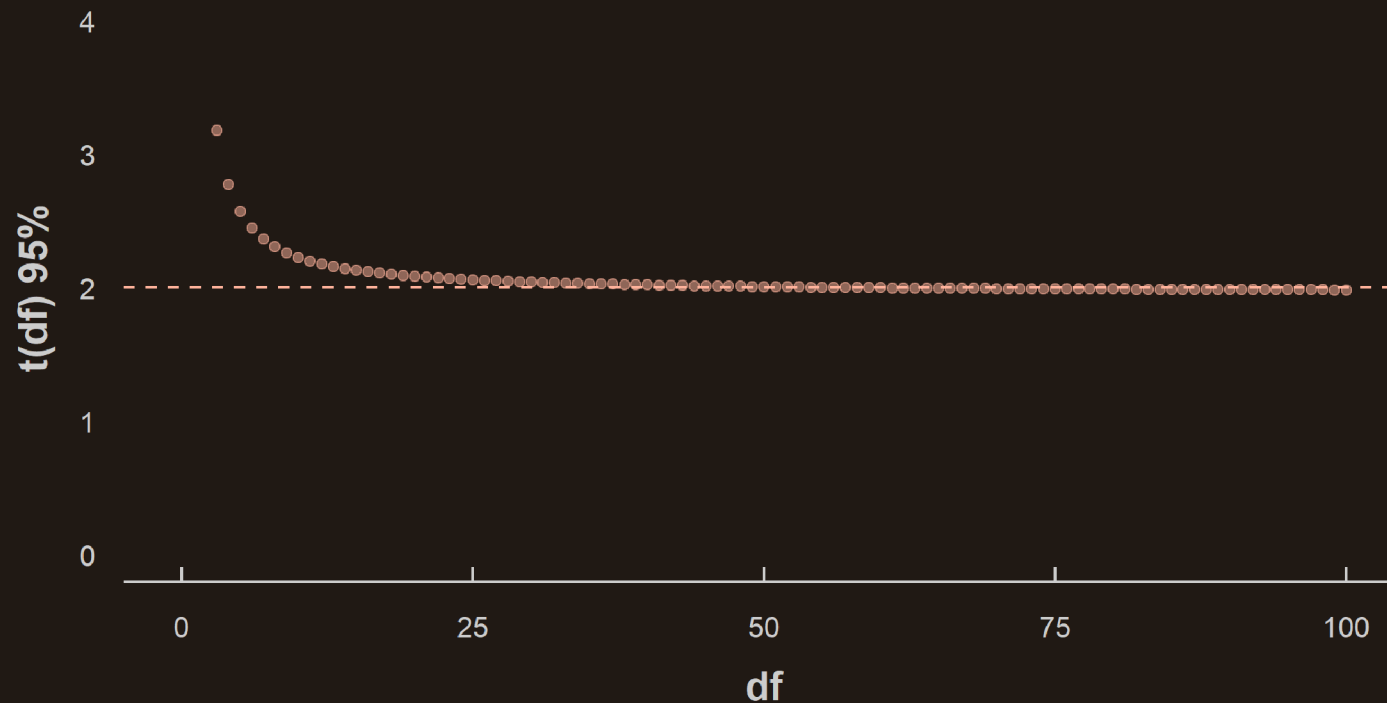
→ A coefficient \approx twice larger than its standard error has a p-value of \approx 5%

3. Regression tables

3.2. Reported significance

- Remember the formula for the **confidence interval**:
 - We can **fix** the **confidence level** $1 - \alpha$ to 95% and check **how t varies with df**

$$\hat{\beta} \pm t(df)_{1-\frac{\alpha}{2}} \times se(\hat{\beta})$$



3. Regression tables

3.2. Reported significance

- **As soon as** you have **about 20 observations more than** you have **parameters** to estimate:
 - The ***t* value** gets very **close to 2**
 - And as **df** increases it quickly converges to ≈ 2
- The coefficient is statistically significant if the lower bound of its (absolute) confidence interval is larger than 0
 - Which is an easy calculation if we **approximate the *t* value by 2**
 - A reasonable approximation for a back of the envelope calculation unless there are very few observations
- The (*absolute*) lower bound of the CI writes:

$$|\hat{\beta}| - t(df)_{1-\frac{\alpha}{2}} \times \text{se}(\hat{\beta})$$

$$|\hat{\beta}| - 2 \times \text{se}(\hat{\beta}) > 0$$

$$|\hat{\beta}| > 2 \times \text{se}(\hat{\beta})$$

So if the **coefficient** is clearly more than **twice larger** than its **standard error**, it must be **statistically significant** at the **5% significance level**

→ But sometimes the p-value or the confidence interval is reported instead of the standard error



3. Regression tables

3.2. Reported significance

	<i>Dependent variable:</i>	
	Birth weight	
	(1)	(2)
Household income	0.002 ^{***} p = 0.000	0.002 ^{***} p = 0.000
Girl (ref: Boy)		-135.218 ^{***} p = 0.0002
Constant	3,134.528 ^{***} p = 0.000	3,246.365 ^{***} p = 0.000
Observations	1,000	963

Note: * p<0.1; ** p<0.05; *** p<0.01

	<i>Dependent variable:</i>	
	Birth weight	
	(1)	(2)
Household income	0.002 ^{***} (0.002, 0.003)	0.002 ^{***} (0.002, 0.003)
Girl (ref: Boy)		-135.218 ^{***} (-203.500, -66.936)
Constant	3,134.528 ^{***} (3,102.055, 3,167.002)	3,246.365 ^{***} (3,179.223, 3,313.507)
Observations	1,000	963

Note: * p<0.1; ** p<0.05; *** p<0.01



3. Regression tables

3.3. R squared

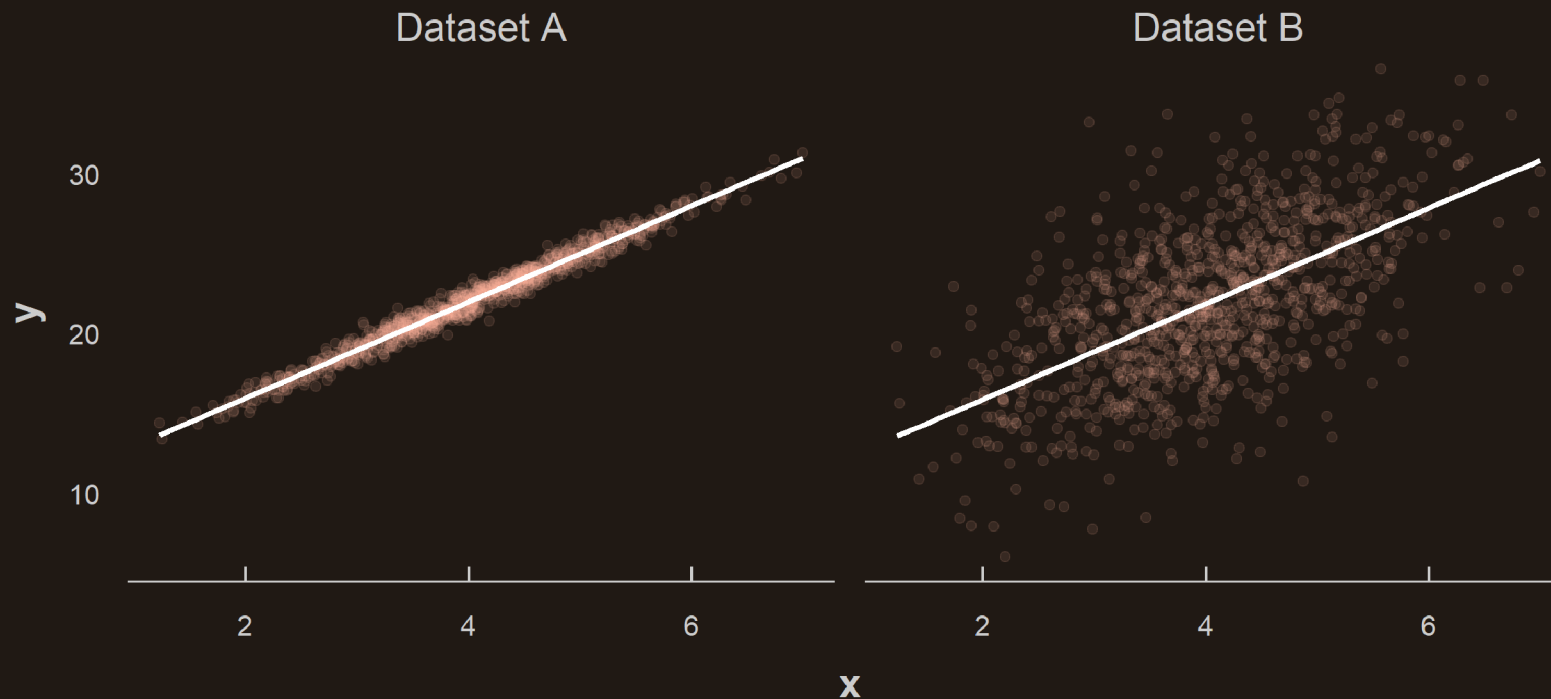
- In **regression tables**, the R^2 of the model is **always reported** below the number of observations
 - The R^2 captures how well the **model fits the data**
 -



3. Regression tables

3.3. R squared

- In **regression tables**, the R^2 of the model is **always reported** below the number of observations
 - The R^2 captures how well the **model fits the data**
 - The model has a **good fit (high R^2)** on dataset A but a **poor fit (low R^2)** on dataset B





3. Regression tables

3.3. R squared

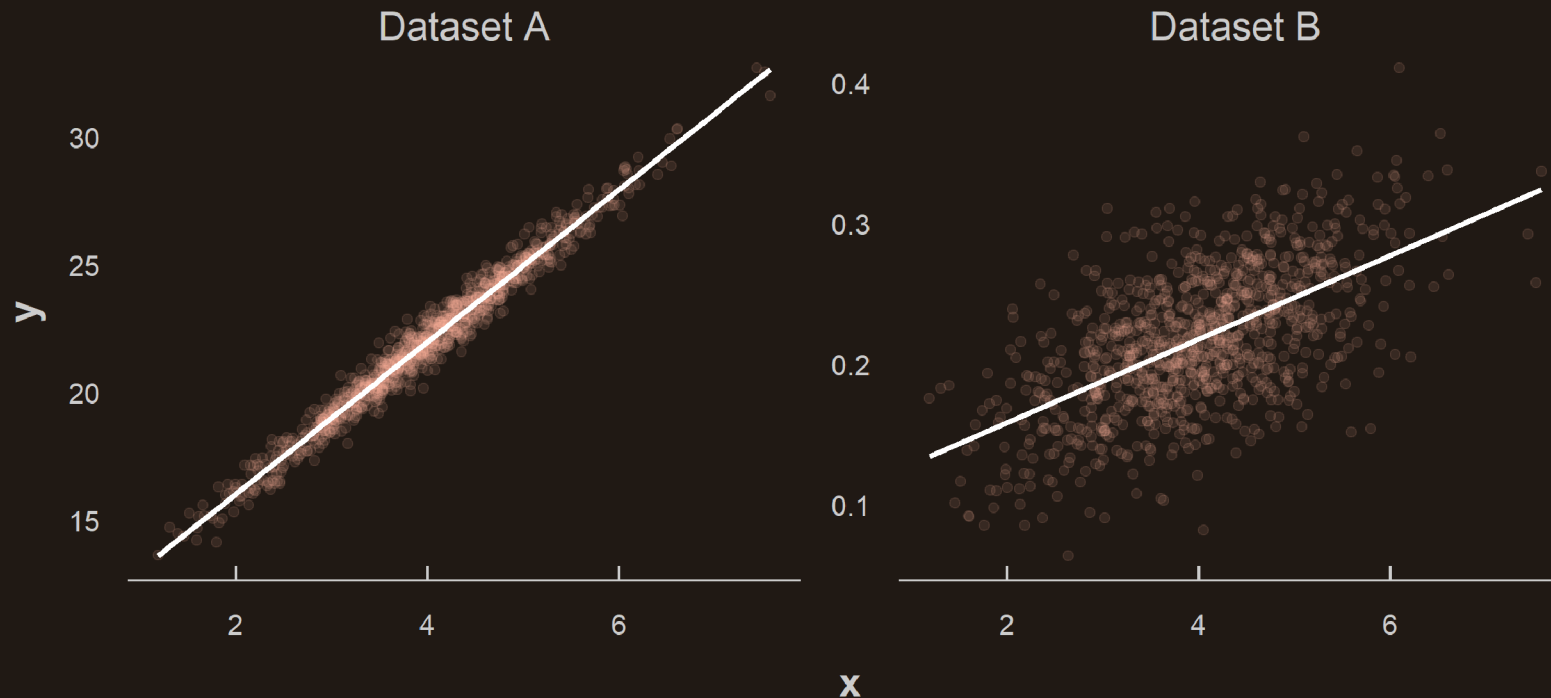
- The **standard error** already gives an idea on the goodness of the fit, but it is expressed in the **same unit as y**
 - So we **cannot compare** two different models based on that statistic
 -



3. Regression tables

3.3. R squared

- The **standard error** already gives an idea on the goodness of the fit, but it is expressed in the **same unit as y**
 - So we **cannot compare** two different models based on that statistic
 - The standard error of the slope would be larger on dataset A than on dataset B

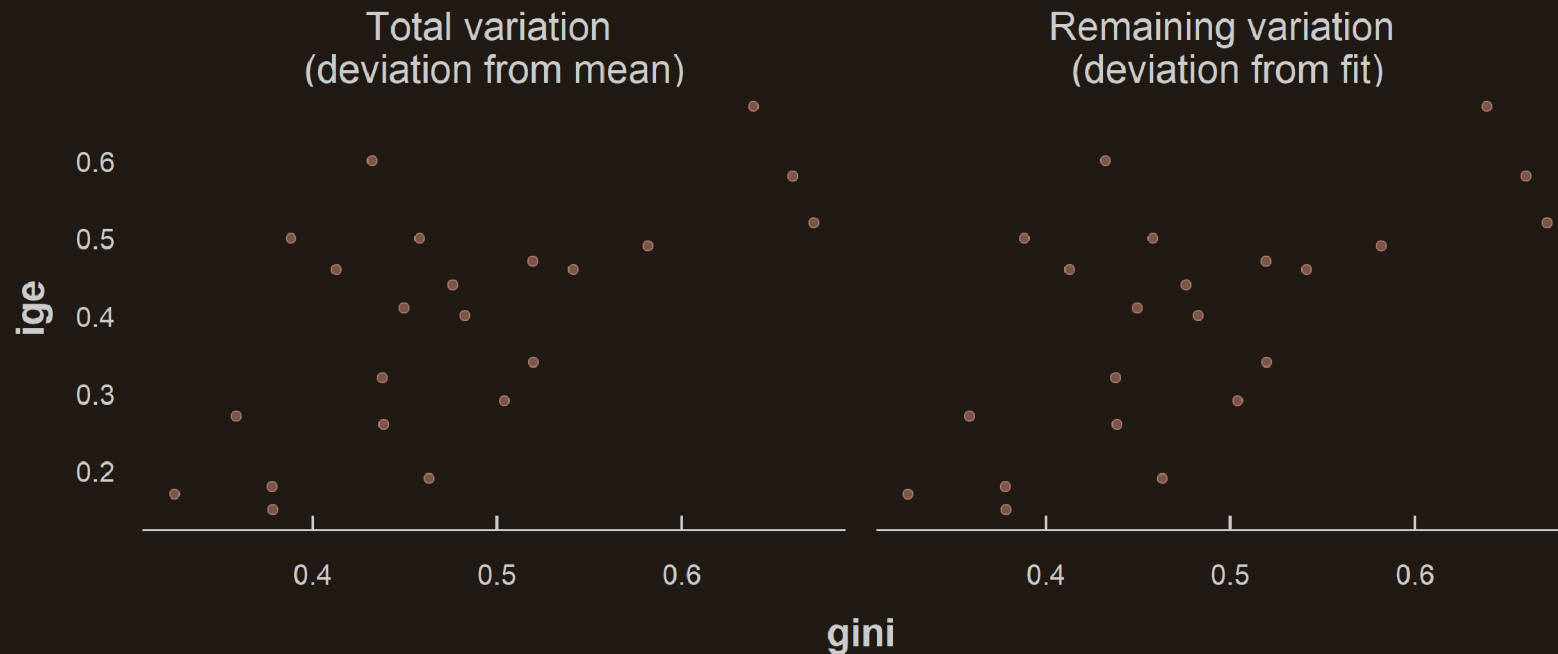




3. Regression tables

3.3. R squared

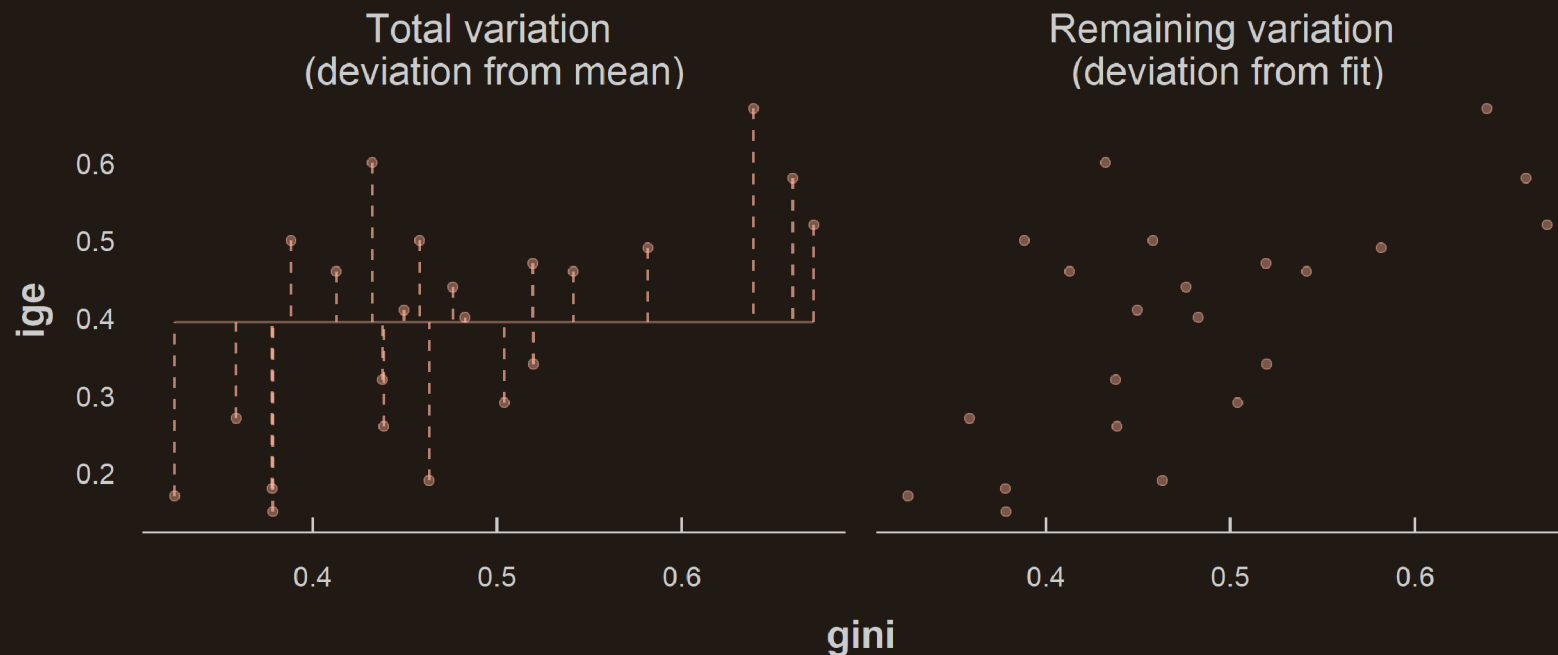
- The R^2 captures the **goodness of fit** as the **percentage** of the y variation captured by the model, from:
 -
 -



3. Regression tables

3.3. R squared

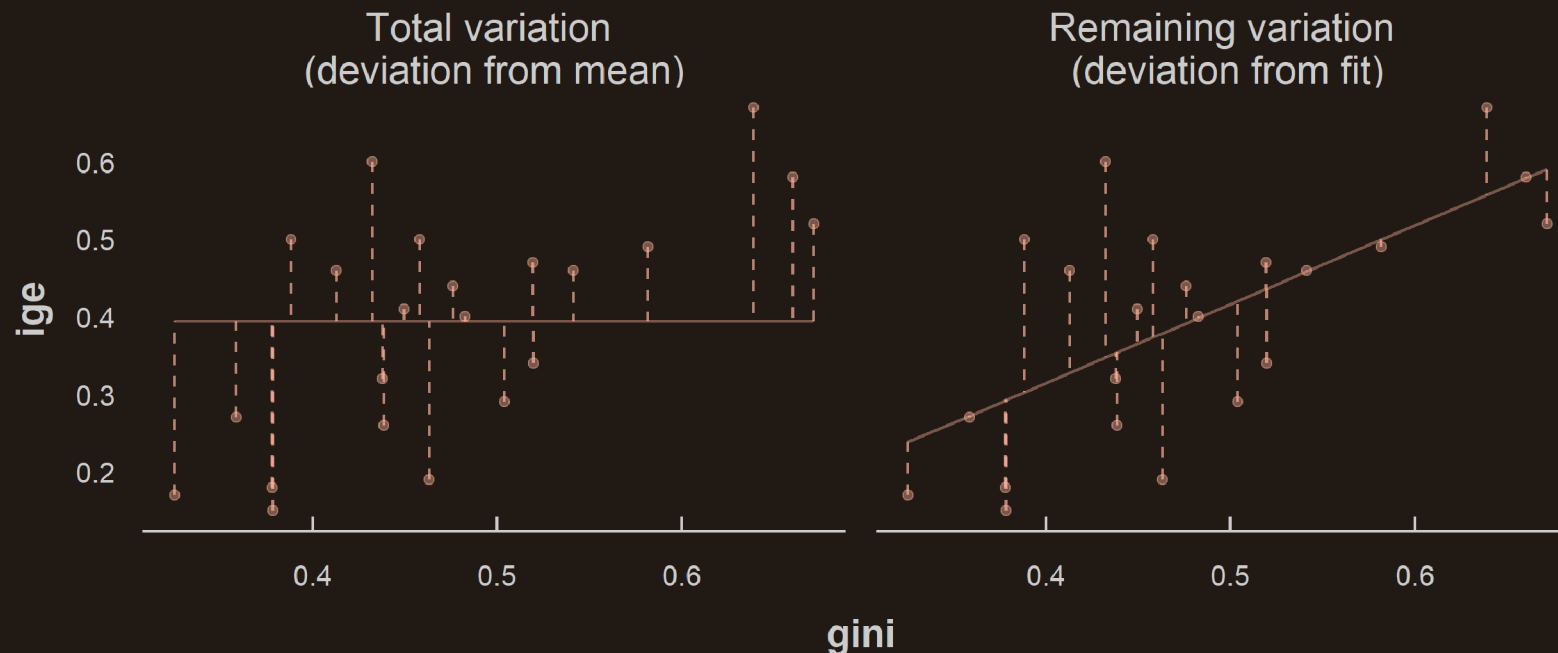
- The R^2 captures the **goodness of fit** as the **percentage** of the y variation captured by the model, from:
 - The **total variation** of the y variable (its variance $\sum_{i=1}^n (y_i - \bar{y})^2$)
 -



3. Regression tables

3.3. R squared

- The R^2 captures the **goodness of fit** as the **percentage** of the y variation captured by the model, from:
 - The **total variation** of the y variable (its variance $\sum_{i=1}^n (y_i - \bar{y})^2$)
 - The **remaining variation** of the y variable once its modeled (the sum of squared residuals $\sum_{i=1}^n \hat{\epsilon}_i^2$)



3. Regression tables

3.3. R squared

- We can then obtain a proper formula from the following reasoning

$$\text{Total variation} = \text{Explained variation} + \text{Remaining variation}$$

$$\frac{\text{Explained variation}}{\text{Total variation}} = 1 - \frac{\text{Remaining variation}}{\text{Total variation}}$$

$$\frac{\text{Explained variation}}{\text{Total variation}} = 1 - \frac{\sum_{i=1}^n \hat{\varepsilon}_i^2}{\sum_{i=1}^n (y_i - \bar{y})^2} \equiv R^2$$

- Because all the terms are sums of squares, we usually talk about:
 - **Total Sum of Squares** (TSS)
 - **Explained Sum of Squares** (ESS)
 - **Residual Sum of Squares** (RSS)

3. Regression tables

3.3. R squared

- Note that the **TSS** is actually the **variance of y** :
 - So the **R^2** is interpreted as the **share of the variance of y** which is **explained** by the model
 - And as such, the **R^2** is always comprised **between 0 and 1**

$$R^2 = 1 - \frac{\sum_{i=1}^n \hat{\varepsilon}_i^2}{\sum_{i=1}^n (y_i - \bar{y})^2} = \frac{\text{Explained variation}}{\text{Total variation}}$$

- An undesirable property of the **R^2** is that it **mechanically increases** with the number of **dependent variables**
 - Such that with many variables the **R^2** tends to overestimate the goodness of the fit
 - This is why you will sometimes see some **Adjusted R^2**

$$\text{Adjusted } R^2 = 1 - \frac{(1 - R^2)(n - 1)}{n - \#parameters}$$

Overview



1. Point estimates ✓

- 1.1. Continuous variables
- 1.2. Discrete variables
- 1.3. Log vs. level

2. Practice interpretation ✓

3. Regression tables ✓

- 3.1. Layout
- 3.2. Reported significance
- 3.3. R squared

4. Wrap up!

4. Wrap up!

Standard interpretations

- When both x and y are continuous, the **general** template for the **interpretation** of $\hat{\beta}$ is:

"Everything else equal, a 1 [unit] increase in [x] is associated with an [in/de]crease of [beta] [units] in [y] on average."

- With a discrete x , the interpretation of the coefficient must be **relative to the reference category**:

"Everything else equal, belonging to the [x category] is associated with a [beta] [unit] [higher/lower] average [y] relative to the [reference category]."

- With a **binary y variable**, the coefficient must be interpreted in **percentage points**:

"Everything else equal, a 1 [unit] increase in [x] is associated with a [beta \times 100] percentage point [in/de]crease in the probability that [y equals 1] on average."

4. Wrap up!

Interpretations with variable transformation

Standardization

- To standardize a variable is to **divide it by its SD**
 - The variation of a standardized variable should not be **interpreted** in units but **in SD**
 - For instance if x and y are continuous and x is standardized, the interpretation becomes:

*"Everything else equal, a 1 **standard deviation** increase in $[x]$ is associated with an $[in/de]$ crease of $[beta]$ $[units]$ in $[y]$ on average."*

- If both x and y are standardized, the slope is the correlation coefficient between x and y

Log-transformation

- The log transformation allows to interpret the coefficient in percentage:

Interpretation of the regression coefficient

	y	log(y)
x	$\hat{\beta}$ is the unit increase in y due to a 1 unit increase in x	$\hat{\beta} \times 100$ is the % increase in y due to a 1 unit increase in x
log(x)	$\hat{\beta} \div 100$ is the unit increase in y due to a 1% increase in x	$\hat{\beta}$ is the % increase in y due to a 1% increase in x

4. Wrap up!

Regression table layout

	Birth weight	
	(1)	(2)
Household income	0.002 ^{***} (0.0003)	0.002 ^{***} (0.0003)
Girl (ref: Boy)		-135.218 ^{***} (34.838)
Constant	3,134.528 ^{***} (16.568)	3,246.365 ^{***} (34.257)
Observations	1,000	963
R ²	0.059	0.074

Note: * p<0.1; ** p<0.05; *** p<0.01

Regression tables often contain multiple regressions:

- With **one regression in each column**
- And one variable in **each row**
 - With the **point estimate**
 - And a **precision measure** below
- **General info** on each model **at the bottom**
 - Number of observations
 - $R^2 = 1 - \frac{\sum_{i=1}^n \hat{\epsilon}_i^2}{\sum_{i=1}^n (y_i - \bar{y})^2}$
- A **symbology** for the **p-value** testing whether the coefficient is significantly different from 0 or not