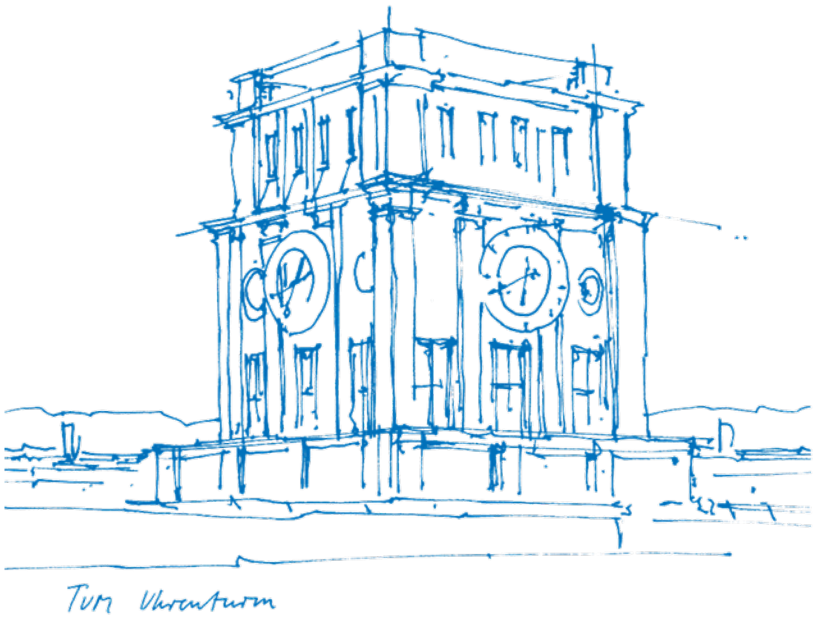
Empirical Research Methods - Lecture 6

Prof. Dr. Helmut Farbmacher
Technical University of Munich
TUM School of Management
Munich, May 27, 2024



Linear Regression (Part IV)

Outline of today's lecture

Interpretation of Parameters

- Logarithms in Regressions
- Polynomials
- Interactions
- Categorical Regressors (two categories)
- Categorical Regressors (multiple categories)

Hypothesis Testing

- Different ways to convey similar information
- Sidenote 1: Significance vs Relevance
- Sidenote 2: Multiple testing

Interpretation of Parameters

Logarithms in Regressions

Continuous Regressors (Recap)

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

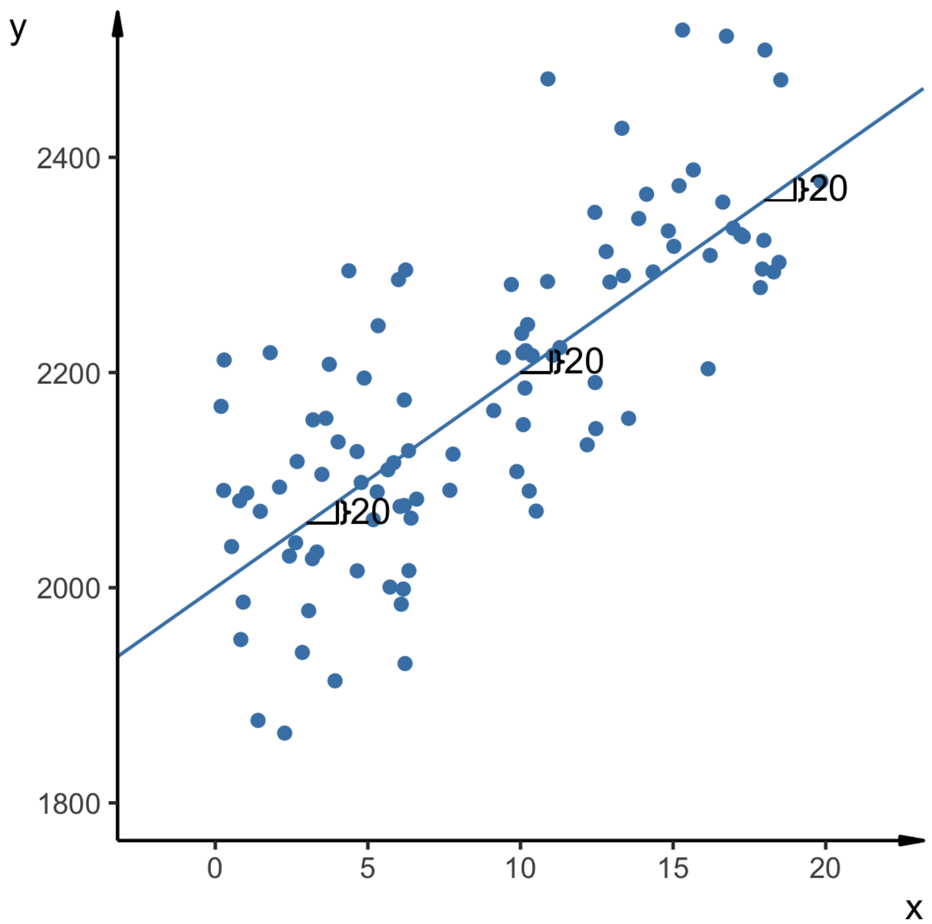
When a regressor x is continuously distributed, we define the **marginal effect** of a change in x as derivative

$$\frac{\partial \mathrm{E}(y|x)}{\partial x} = \beta$$

In the linear regression model the derivative is simply the slope parameter (β).

That is the marginal effect of x is **constant** in a linear model. Particularly, it does <u>not</u> depend on the level of y or x so far.

Continuous Regressors (Recap)



Logarithms in Regressions

We can also model nonlinear relations in a linear regression model. Note that the "linearity" in a linear regression is just referring to the parameters in the model.

That is, the following regressions are linear regression models,

$$\begin{aligned} y &= \alpha + \beta ln(x) + u \\ ln(y) &= \alpha + \beta x + u \\ ln(y) &= \alpha + \beta ln(x) + u \end{aligned}$$

although the natural logarithm $ln(\cdot)$ clearly is a nonlinear function.

The variables being log-transformed must take on positive values only.

Logarithms in Regressions

Linear-Log:	$y = \alpha + \beta ln(x) + u$	A 1% change in x is associated with a change in y of 0.01β.
Log-Linear:	$ln(y) = \alpha + \beta x + u$	A change in x by one unit (Δx = 1) is associated with a 100*β% change in y.
Log-Log:	$ln(y) = \alpha + \beta ln(x) + u$	A 1% change in x is associated with a β% change in y. β is thus the elasticity of y with respect to x.

Logarithms in Regressions

Data info:

The *ceosal2* dataset contains information on U.S. CEOs, including age, degree, salary, tenure, etc.from 1990.

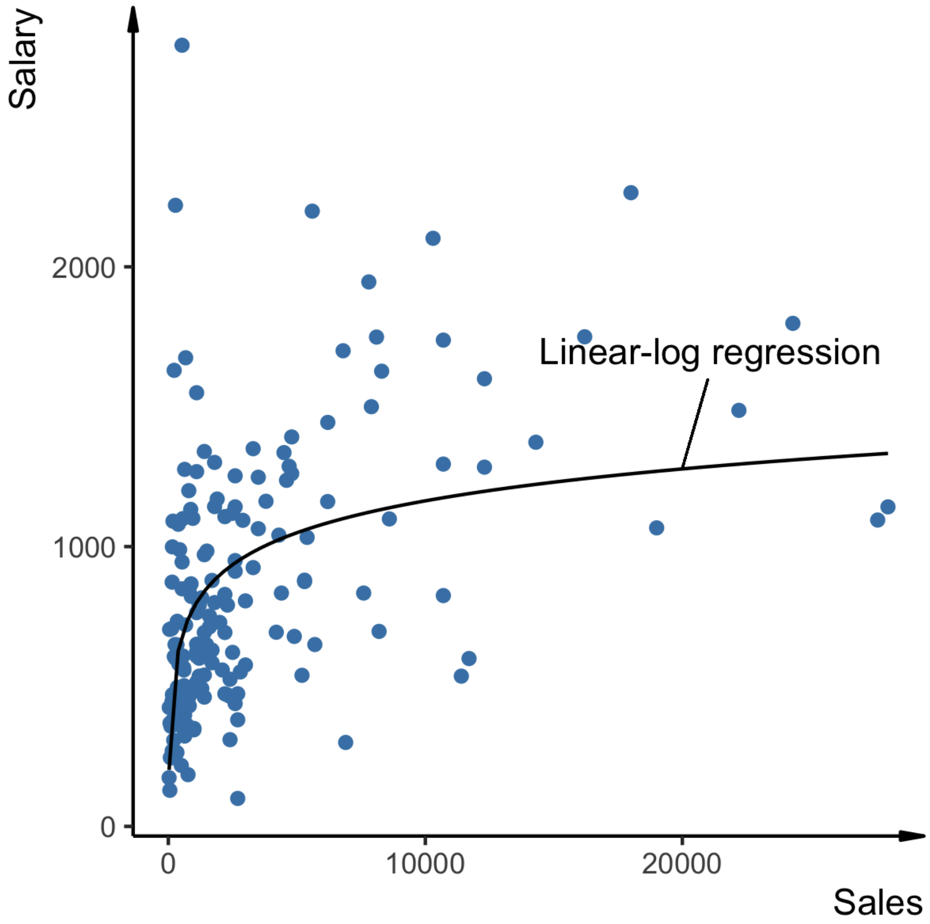
Variables:

salary	compensation of the CEO (in $1000)
sales	firm sales, millions

Data can be downloaded here:
`https://econpapers.repec.org/paper/bocbocins/ceosal2.htm`.

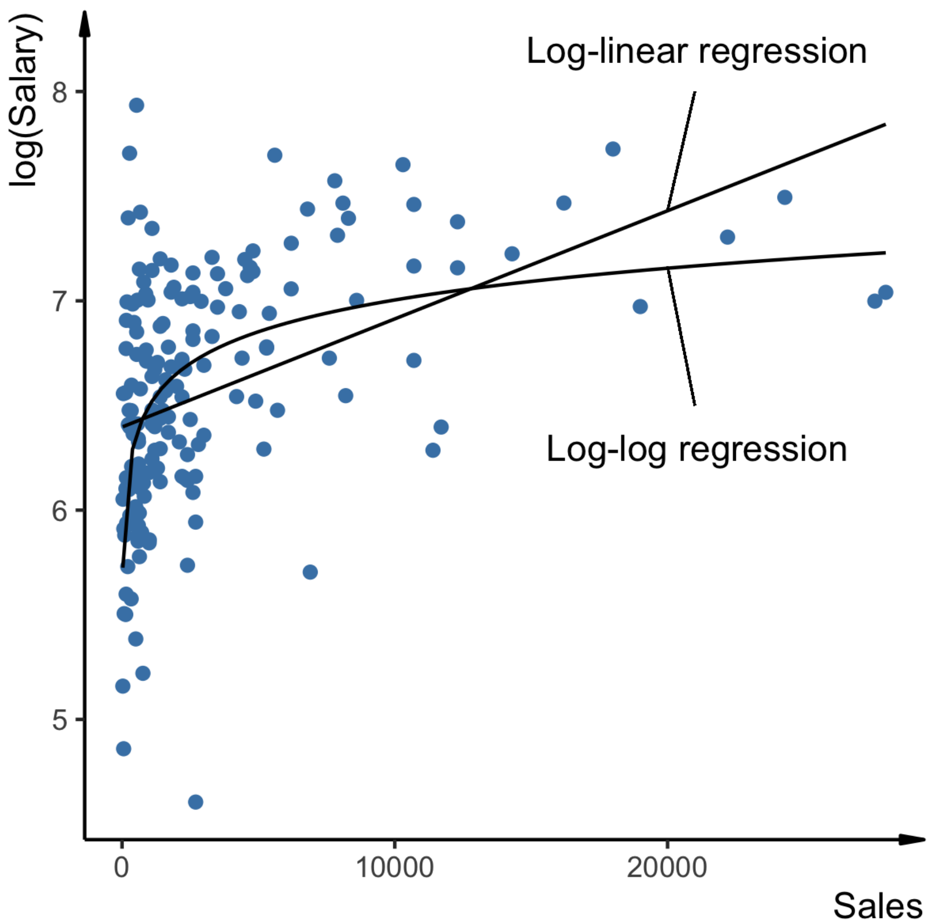
You can also load the data via the *wooldridge* package in *R*, use the function *data("ceosal2")* after installing the package.

Logarithms in Regressions



Applied Econometrics

Logarithms in Regressions



Interpretation of Parameters

Polynomials

Polynomials

Example: quadratic effect of job experience

$$wage = \beta_0 + \beta_1 educ + \beta_2 exp + \beta_3 exp^2 + u$$

Wage is explained by education and a quadratic function of experience.

The model has three explanatory variables:
education, experience and experience squared.

The marginal effect of experience is:

$$\frac{\partial \mathrm{E}(wage|\boldsymbol{x})}{\partial exp} = \beta_2 + 2\beta_3 exp$$

Polynomials

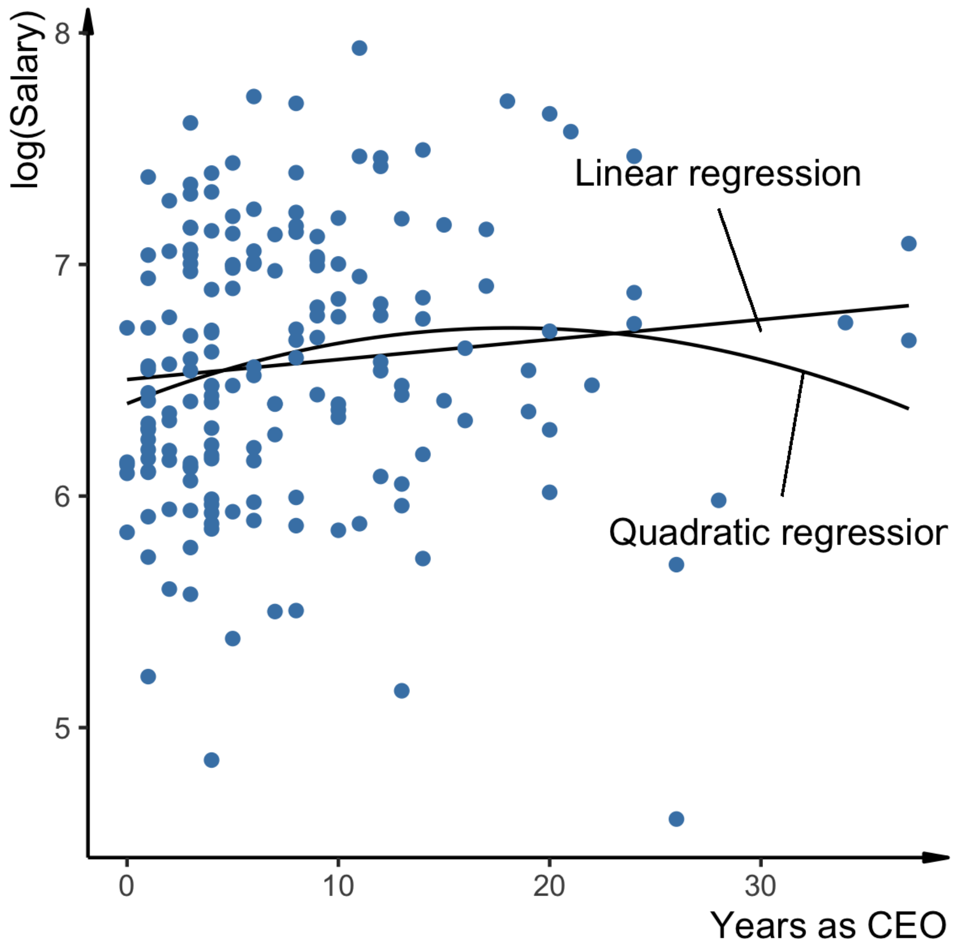
Example: CEO salary

$$log(salary) = \beta_0 + \beta_1 log(sales) + \beta_2 ceoten + \beta_3 ceoten^2 + u$$

Model assumes a (ceteris paribus) constant elasticity relationship between CEO salary and the sale of the firm.

Model assumes a (ceteris paribus) quadratic relationship between log CEO salary and his or her tenure with the firm.

Polynomials



Interpretation of Parameters

Interactions

Binary & Continuous Regressors (Recap)

$$wage = \beta_0 + \beta_1 female + \beta_2 educ + u$$

When a regressor x is **binary**, we get as marginal effect a group difference

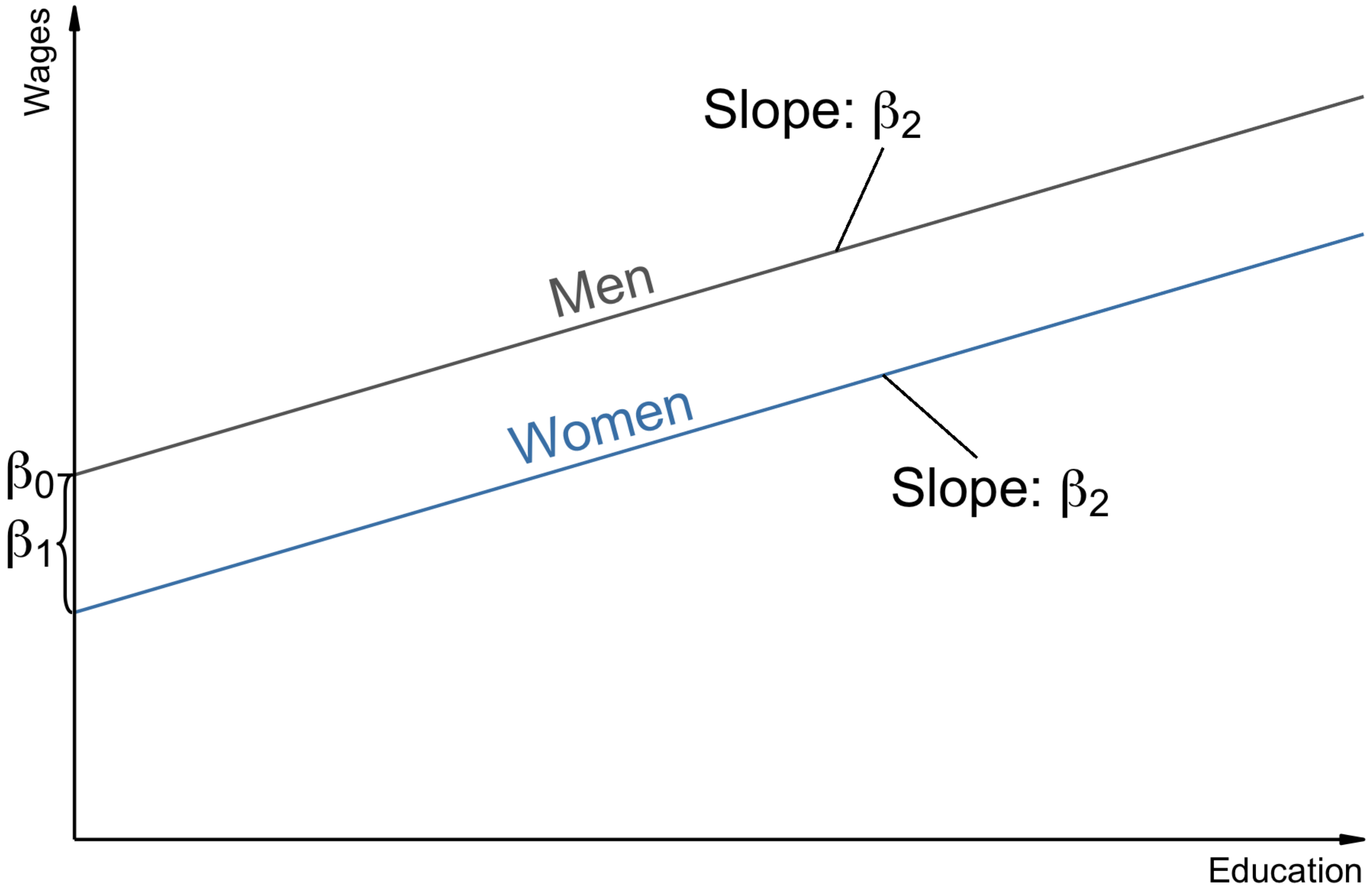
$$\beta_1 = \mathrm{E}[wage|educ, female = 1] - \mathrm{E}[wage|educ, female = 0]$$

$$\beta_2 = \frac{\partial \mathrm{E}[wage|educ, female]}{\partial educ}$$

hence, β_1 is an **intercept shifter** and β_2 is the slope parameter.

Interpretation: β_1 gives the difference in wages between women and men, holding education fixed.

Binary & Continuous Regressors (Recap)



Interactions

An important practical technique is the use of **interaction terms**.

Suppose we want to allow that the effect of schooling differs b/w women and men

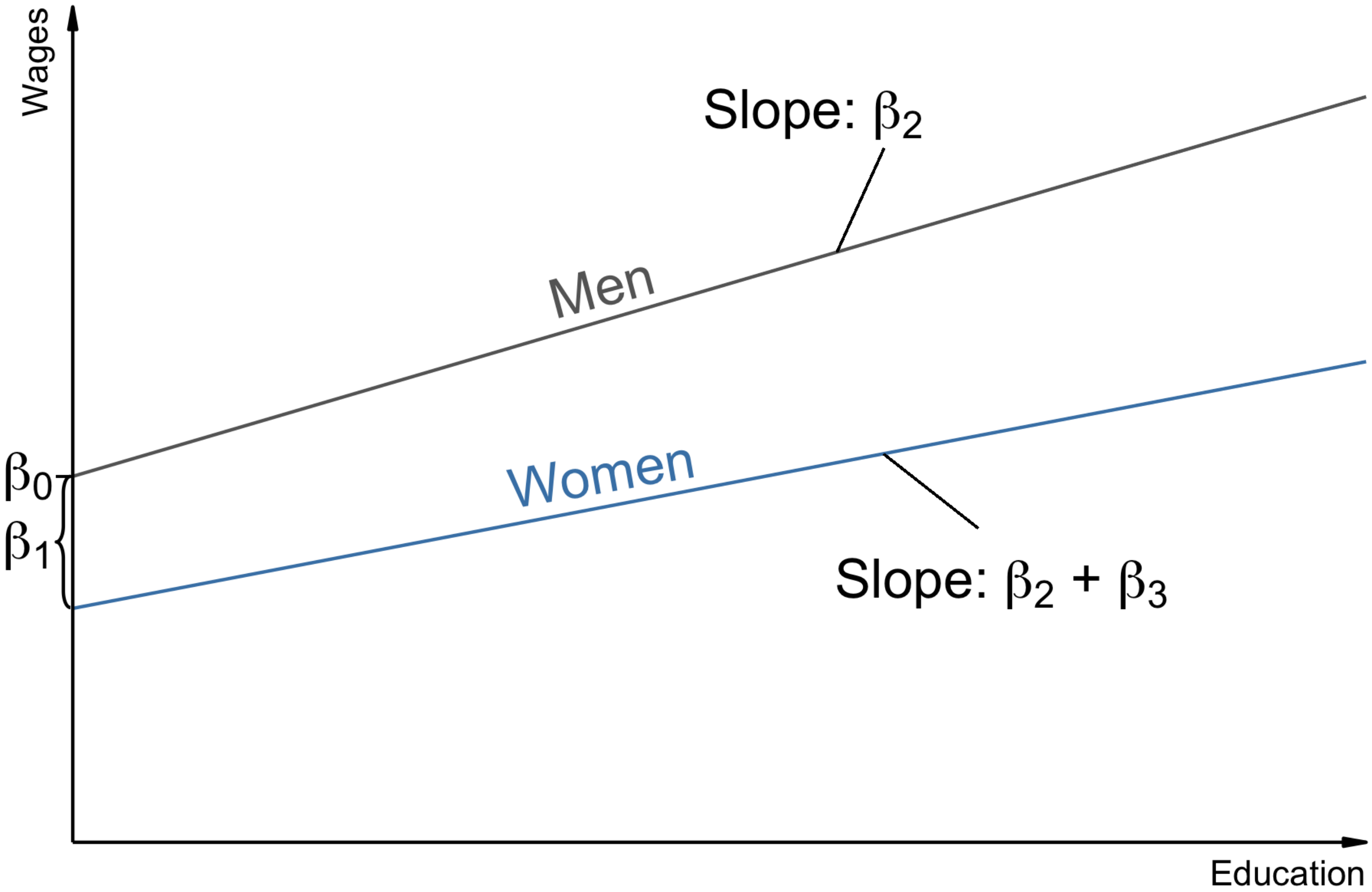
$$wage = \beta_0 + \beta_1 female + \beta_2 educ + \beta_3 female \times educ + u$$

The marginal effect of *educ* on *wage* now is

$$\frac{\partial \mathrm{E}(wage|\boldsymbol{x})}{\partial educ} = \beta_2 + \beta_3 female$$

It is straightforward to test for an interaction effect $H_0 : \beta_3 = 0$ vs $H_1 : \beta_3 \neq 0$.

Interactions: Graphical Illustration



Interpretation of Parameters

Categorical Regressors (two categories)

Categorical Regressors (two categories)

Consider a categorical variable with mutually exclusive categories.

When using dummy variables, one category has to be omitted.

For example, $male + female = 1$, there is perfect collinearity between *male*, *female* and the *constant* (**dummy variable trap**).

We call the omitted category the **reference or baseline category**.

1. $wage = \beta_0 + \beta_1 educ + \beta_2 female + u \rightarrow$ Reference category are men ($female = 0$)

   or

2. $wage = \alpha_0 + \alpha_1 educ + \alpha_2 male + u \rightarrow$ Reference category are women ($male = 0$)

Categorical Regressors (two categories)

How does our dataset look like with a dummy variable? Say, we have a constant, educ (in years) and a dummy for female (3rd variable):

$$\begin{pmatrix} 1 & 9 & 1 \\ 1 & 12 & 1 \\ 1 & 8 & 0 \\ 1 & 15 & 1 \\ 1 & 15 & 0 \end{pmatrix}$$

Now, let's add a dummy variable for male (4th variable) in this dataset:

$$\begin{pmatrix} 1 & 9 & 1 & 0 \\ 1 & 12 & 1 & 0 \\ 1 & 8 & 0 & 1 \\ 1 & 15 & 1 & 0 \\ 1 & 15 & 0 & 1 \end{pmatrix}$$

$male + female = 1$, **perfect collinearity** between *male*, *female* and the *constant*.

Categorical Regressors (two categories)

Statistical software drops one of the dummies if you include *male* and *female*.

General rule: We always have to leave one of the categories out of the regression (**reference/baseline category**)

Does it matter which category we use as reference? Not really!

We can change the reference category

$$\begin{aligned} wage &= \beta_0 + \beta_1 educ + \beta_2 female + u \\ &= \beta_0 + \beta_1 educ + \beta_2 (1 - male) + u \\ &= (\beta_0 + \beta_2) + \beta_1 educ - \beta_2 male + u \\ &= \alpha_0 + \alpha_1 educ + \alpha_2 male + u \end{aligned}$$

where $\alpha_0 = \beta_0 + \beta_2$, $\alpha_1 = \beta_1$, and $\alpha_2 = -\beta_2$

The constant differs, the sign (of the dummy) flips, but the conclusion is the same.

Categorical Regressors (two categories)

Data info:

The Current Population Survey (*cps09mar*) contains information on employment, earnings, educational attainment, income etc. for 57.000 U.S. households (March 2009).

Variables:

earnings	total annual earnings
education	years of education (based on highest degree)
female	1 if female, 0 otherwise
male	1 if male, 0 otherwise

Data can be downloaded here: `https://www.ssc.wisc.edu/~bhansen/econometrics/`.

Categorical Regressors (two categories)

$$lm(formula = earnings \sim female + education, data = cps09mar)$$

MODEL INFO:
Observations: 50742
Dependent Variable: earnings
Type: OLS linear regression

MODEL FIT:
$F(2,50739) = 5810.40$, $p = 0.00$
$R^2 = 0.19$
Adj. $R^2 = 0.19$

Standard errors: OLS

	Est.	S.E.	t val.	p
(Intercept)	-40230.70	1089.59	-36.92	0.00
female	-20770.84	423.35	-49.06	0.00
education	7480.62	76.27	98.08	0.00

$\widehat{\beta}_2 = \ldots$

Categorical Regressors (two categories)

$$lm(formula = earnings \sim male + education, data = cps09mar)$$

```
MODEL INFO:
Observations: 50742
Dependent Variable: earnings
Type: OLS linear regression

MODEL FIT:
F(2,50739) = 5810.40, p = 0.00
R² = 0.19
Adj. R² = 0.19

Standard errors: OLS
-------------------------------------------------------
                        Est.       S.E.    t val.      p
----------------- ----------- --------- --------- ------
(Intercept)        -61001.55    1119.75    -54.48   0.00
male                20770.84     423.35     49.06   0.00
education            7480.62      76.27     98.08   0.00
-------------------------------------------------------
```

$\widehat{\alpha}_2 = \ldots$

Interpretation of Parameters

Categorical Regressors (multiple categories)

Categorical Regressors (multiple categories)

When a regressor x can take on **multiple discrete** values (e.g. occupation, industry, region, educational degrees), we get as marginal effects several group differences.

Let D represent a categorical variable assuming J **distinct** values.

We do <u>not</u> want to estimate

$$y = \beta_0 + \beta_1 D + u$$

because β_1 has <u>no</u> precise interpretation. All group differences would be set to β_1 . Simply recoding categories would give different estimates.

Solution: create $J-1$ dummy variables (omit one, J_{ref}, which becomes the reference category)

$$d_{ij} = \begin{cases} 1 & \text{if } D_i = j \\ 0 & \text{otherwise} \end{cases}$$

Categorical Regressors (multiple categories)

Example: highest education obtained; $degree \in \{$ BSc, MSc, PhD$\}$

$$degree_1 = \begin{cases} 1 & \text{if } degree = \text{BSc} \\ 0 & \text{otherwise} \end{cases}$$

$$degree_2 = \begin{cases} 1 & \text{if } degree = \text{MSc} \\ 0 & \text{otherwise} \end{cases}$$

$$degree_3 = \begin{cases} 1 & \text{if } degree = PhD \\ 0 & \text{otherwise} \end{cases}$$

Note that degrees are defined as "highest degree obtained", and so they are **mutually exclusive**.

Categorical Regressors (multiple categories)

i	D	d_1	d_2	d_3
1	category A	1	0	0
2	category B	0	1	0
3	category A	1	0	0
4	category C	0	0	1
5	category C	0	0	1
6	category B	0	1	0
7	category B	0	1	0
8	category A	1	0	0
9	category A	1	0	0
10	category C	0	0	1

Categorical Regressors (multiple categories)

We can then estimate the following model:

$$wage = \beta_0 + \beta_1 degree_2 + \beta_2 degree_3 + u$$

where "BSc" is the **baseline/reference category**.

The intercept β_0 is the mean outcome of those in the reference group (i.e. BSc)

β_1 is the difference in wages between Master ($degree = 2$) and Bachelor degree (baseline):

$$\beta_1 = \mathrm{E}(wage|MSc) - \mathrm{E}(wage|BSc)$$

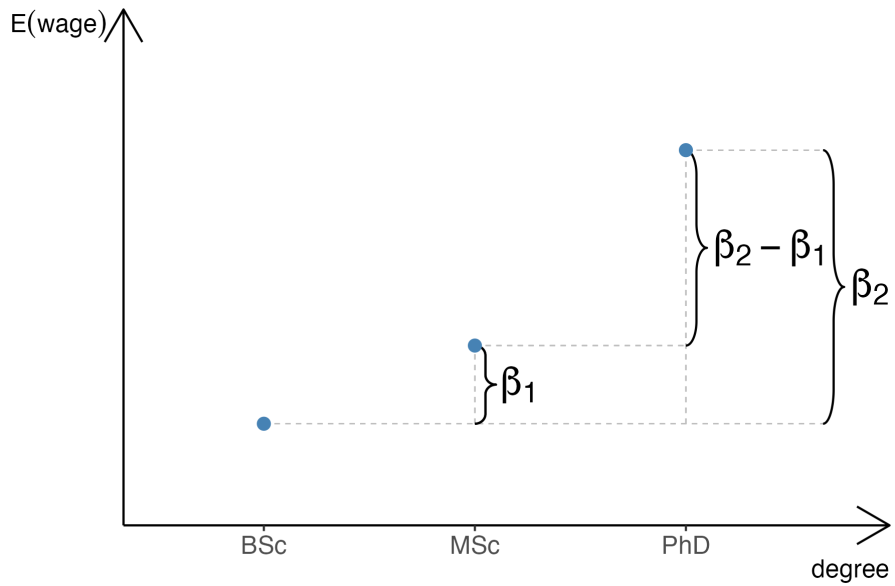
β_2 is the difference in wages between PhD ($degree = 3$) and Bachelor degree (baseline):

$$\beta_2 = \mathrm{E}(wage|PhD) - \mathrm{E}(wage|BSc)$$

Categorical Regressors (multiple categories)

If we want to learn the difference between two groups of a categorical regressor, we only need to contrast their coefficients. We get the difference b/w PhD ($degree = 3$) and Master ($degree = 2$) by comparing β_1 and β_2, that is

$$\begin{aligned}\beta_2 - \beta_1 &= \mathrm{E}(wage|PhD) - \mathrm{E}(wage|BSc) - [\mathrm{E}(wage|MSc) - \mathrm{E}(wage|BSc)] \\ &= \mathrm{E}(wage|PhD) - \mathrm{E}(wage|MSc)\end{aligned}$$



Categorical Regressors (multiple categories)

We can estimate

$$y = \beta_0 + \sum_{j \neq J_{ref}} \beta_j d_j + u$$

The interpretation of the regression coefficients is

$$\beta_j = \mathrm{E}(y|d = j) - \mathrm{E}(y|d = J_{ref})$$

and therefore **relative to the reference category** (J_{ref}).

The intercept β_0 is here the average outcome of those in the reference group.

If we want to compare group k with group m, we need to compare their coefficients:

$$\beta_k - \beta_m = \mathrm{E}(y|d = k) - \mathrm{E}(y|d = m)$$

Hypothesis Testing (in large samples)

Hypothesis Testing

We consider the following (bivariate) regression:

$$y_i = \alpha + \beta x_i + u_i\,.$$

Suppose we want to test $H_0 : \beta = \beta_{H_0}$ vs $H_1 : \beta \neq \beta_{H_0}$ for an hypothesized value (β_{H_0}).

For example, we want to test whether the private returns to education are equal to 1000 € ($\beta_{H_0} = 1000$) or whether the private returns to education are equal to 0 € ($\beta_{H_0} = 0$), that is education, x, has no effect at all on wages, y.

Keep in mind:

- β is the true (unknown) value in the population
- $\widehat{\beta}$ is the OLS estimate in our sample
- β_{H_0} is a (known) hypothesized value (often this value is simply 0)

Hypothesis Testing

We are looking for sufficient evidence against the null hypothesis.

Idea: Construct a test with a known distribution under the null hypothesis (H_0).
Reject the null hypothesis if the value of the test statistic is too large.

Ingredients: $\widehat{\beta} \xrightarrow{d} N\left(\beta, \mathrm{Var}(\widehat{\beta})\right)$

The t-statistic is simply a transformation of $\widehat{\beta}$ such that it would be **standard** normally distributed under the null hypothesis (i.e. if $\beta = \beta_{H_0}$ is true)

$$t = \frac{\widehat{\beta} - \beta_{H_0}}{\sqrt{\mathrm{Var}_{\widehat{\beta}}}} \left[= \frac{\widehat{\beta} - \beta_{H_0}}{\mathrm{SE}} \right]$$

A large (in absolute values) t-statistic is evidence against the null hypothesis.

Hypothesis Testing

The t-statistic behaves asymptotically like a **standard** normal random variable if the null hypothesis is true (i.e. under H_0).

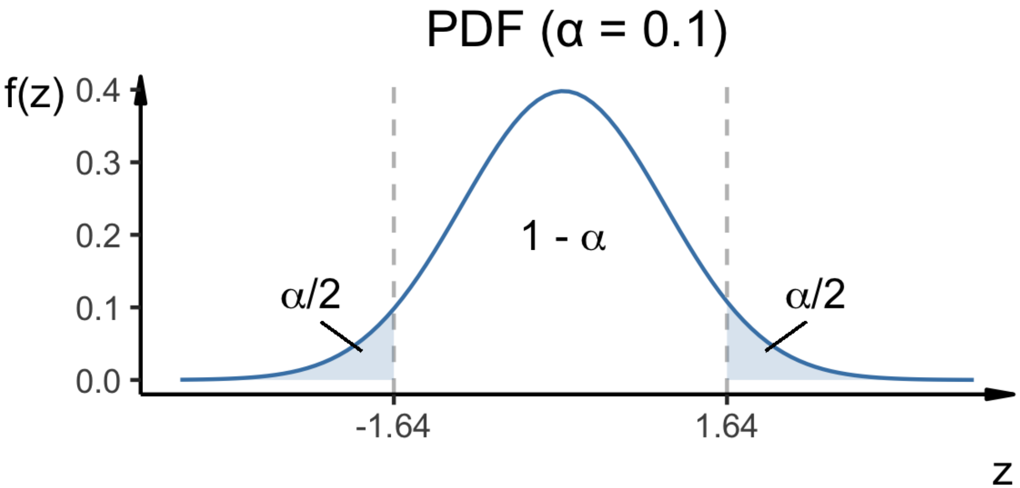
This is a convenient result because we then need to remember only the critical values of a **standard** normal distribution:

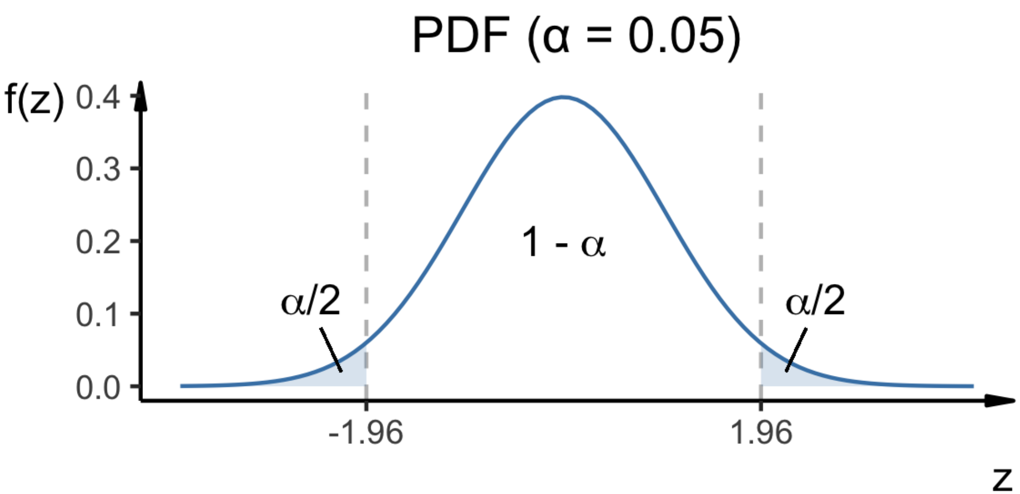
If $|t| > 2.58$ (or $|t| > 1.96$ or $|t| > 1.64$), we can reject the null hypothesis H_0 at the 1% (or 5% or 10%) level.

The term **significance level** refers to a pre-chosen probability (e.g., 5% significance level or $\alpha = 0.05$).

Many authors refer to *statistically significant* if $|t| > 1.96$ and *statistically highly significant* if $|t| > 2.58$.

Hypothesis Testing





Hypothesis Testing

A ***p*-value**, or probability value, is the probability of finding the observed (or a more extreme) t-statistic when the null hypothesis would indeed be true.

You can interpret the p-value as an **empirical significance level**: it denotes the significance level at which we can just reject the null hypothesis.

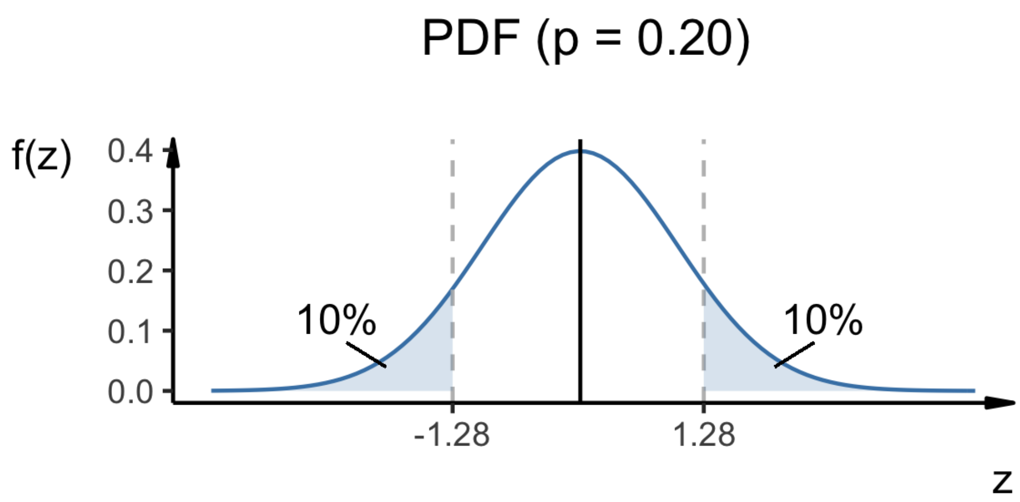
To calculate a p-value on your own, you need:

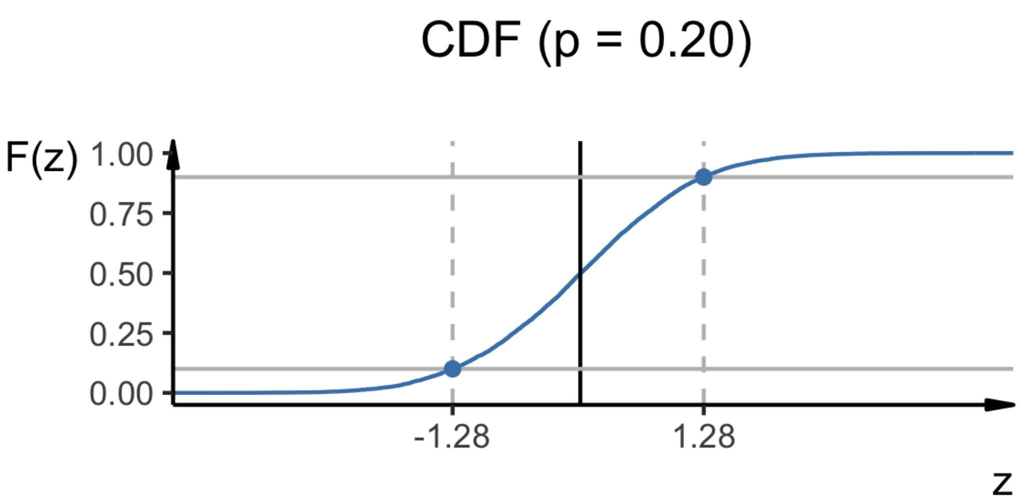
- the absolute value of the t-statistic: $|t|$
- the standard normal cdf: $\Phi(\cdot)$

$$p = 2\left(1 - \Phi\left(|t|\right)\right)$$

A small p-value is evidence against the null.

Hypothesis Testing





Hypothesis Testing

A **confidence interval** (*CI*) is the point estimate plus and minus a measure for the (likely) variation in that estimate.

To calculate a *CI* on your own, you need:

- the point estimate: $\widehat{\beta}$
- the standard error of this point estimate: SE
- the critical value of the intended significance level: $c = \Phi^{-1}(1 - \alpha/2)$

$$CI = \widehat{\beta} \pm c \cdot \mathrm{SE}$$

Remember the critical values: 2.58 (1% level); 1.96 (5% level); 1.64 (10% level)

If the hypothesized value β_{H_0} is outside the confidence interval, we can interpret this as evidence against the null.

Hypothesis Testing

Different ways to convey similar information

Different ways to convey similar information

1. Asterisks to indicate the level of significance, e.g., *** significant at 1% level; ** significant at 5% level; * significant at 10% level
2. Standard errors in parentheses
3. Confidence intervals
4. *p*-values in parentheses
5. *t*-statistics in parentheses

Different ways to convey similar information

Which way should I pick? Depends on your audience (and changes over time).

More and more journals rule out fixed thresholds of significance (e.g., see SMJ Author Guidelines in the subsection “Reporting Results of Statistical Analyses”)

Guidelines Regarding Empirical Research in *SMJ*

Reporting Results of Statistical Analyses

SMJ no longer accepts papers for publication that report or refer to cutoff levels of statistical significance (p-values). In statistical studies, authors should report either standard errors or exact p-values (without asterisks) or both, and should interpret these values appropriately in the text. Rather than referring to specific cutoff points, the discussion could report confidence intervals, explain the standard errors and/or the probability of observing the results in the particular sample, and assess the implications for the research questions or hypotheses tested.

Different ways to convey similar information

More and more journals rule out fixed thresholds of significance (e.g., see AER Guidelines in the subsection “Style Guide”)

aeaweb.org

Tables

-- Columns must be in vertical (or portrait) orientation.
-- Tables must be no more than 9 columns wide including row headings.
-- Number your tables consecutively with Arabic numerals.
-- Use only horizontal lines and additional blank space to show space distinction.
-- Do not use shading.
-- Do not abbreviate in column headings.
-- To denote sections of a table, use panel A, panel B, etc.
-- Place a zero in front of the decimal point in all decimal fractions (e.g., 0.357, not .357).
-- For footnotes pertaining to specific table entries, footnote keys should be lowercase letters (a, b, c, etc.). Do not use asterisks to denote significance of estimation results. Report the standard errors in parentheses.
-- Place source lines after the footnotes. If the citation is a complete sentence, place a period at the end of the source line.
-- Include full citations of sources in the references.

Different ways to convey similar information

Potential pitfalls: *p*-hacking and publication bias.

Addressing these problems may be more important than the way how we present our results.

Lengthy discussions:
Statistical Significance, p-Values, and the Reporting of Uncertainty (Imbens, JEP 2021)
or here:
In Praise of Confidence Intervals (Romer, AER P&P 2020)

Additional discussions (not that important from an applied point of view):
in Statistics
in Operations Research

Sidenote 1: Significance vs Relevance

Statistical significance

- Measured by t-tests, p-values, confidence intervals, standard errors
- Addresses concerns about precision of estimates

Economic relevance

- Measured by coefficient magnitudes (i.e. the absolute value of $\widehat{\beta}$)
- Addresses concerns about policy relevance

Potential pitfalls

Overemphasize **economically irrelevant** variables that are statistically significant (if your sample size is large, you can easily find tiny effects).

Discard **economically relevant** variables that are statistically insignificant (if your sample size is small, you can easily miss important variables).

Sidenote 2: Multiple testing

Testing a **single** parameter of a *multivariate* regression works in the same way as described on the previous slides.

If we want to test an hypothesis that involves **multiple** parameters of a multivariate regression, we need more sophisticated methods.

If you want to know more, join my lecture “Microeconometric Methods for Big Data”.

Recommended reading

For next week please read chapter:

8.2 Panel Data: Estimation

in `https://mixtape.scunning.com/index.html`

Illustration asterisks

Go back

Dickstein et al. (AER P&P, 2015): Market Size and Health Insurance Premiums

AEA PAPERS AND PROCEEDINGS

TABLE 2—REGION-LEVEL ANALYSES

	Number of insurers		Premium	
	(1)	(2)	(3)	(4)
log population	0.652***	0.645***	−108.9***	−137.2**
	(0.187)	(0.221)	(24.09)	(61.65)
log land area (100s of sq. miles)		−0.212*		203.0***
		(0.129)		(68.83)
Fraction population urban		−3.015***		1095.8**
		(1.030)		(510.1)
Fraction pop urban squared		3.094***		−1,047.1*
		(1.095)		(597.4)
Observations	398	398	398	398
R^2	0.619	0.659	0.621	0.656

Notes: Specifications 2 and 4 include as controls: median income, share of households with income 25K–100K, Medicare Geographic Adjustment Factor, share of adult population in 40–64 age bin, percent of employed population working in establishments with fewer than 10 employees, and number of short-term general hospitals. Price regressions include as an additional control the deductible of the second lowest priced silver plan. All regressions include state fixed effects. Standard errors (clustered at the state level) in parentheses.

***Significant at the 1 percent level.
**Significant at the 5 percent level.
*Significant at the 10 percent level.

Illustration standard errors

Go back

Cubas et al. (AER P&P, 2021): Work-Care Balance and the Gender Wage Gap

150 AEA PAPERS AND PROCEEDINGS MAY 2021

Table 1—Gender Gap in Work and Household Care

	Day dummies (1)	+ Dem. controls (2)	+ Usual hours (3)	+<50 hours (4)
Panel A. Working hours				
Female gap in work hours	−0.746	−0.742	−0.317	−0.279
	(0.0424)	(0.0429)	(0.0415)	(0.0427)
Observations	16,073	16,073	16,073	16,073
Average hours, men	8.691			
Average hours, women	7.943			
Average hours, total	8.441			
Panel B. Household care				
Female gap in household care hours	0.388	0.387	0.310	0.272
	(0.0196)	(0.0196)	(0.0190)	(0.0194)
Observations	16073	16073	16073	15483
Average hours, men	0.726			
Average hours, women	1.114			
Average hours, total	0.856			
Panel C. Household care in prime time				
Incidence of household care 8 to 5	0.150	0.150	0.135	0.117
	(0.00755)	(0.00755)	(0.00757)	(0.00787)
Observations	14,386	14,386	14,386	13896
Average, men	0.197			
Average, women	0.347			
Average, total	0.246			

Notes: The table is based on ATUS respondents who are 18–65 years old, who report usual weekly hours >=35 in the CPS, who are married with at least one child in the household, and whose diary day is a weekday. "Work" corresponds to hours spent on "work and work-related activities," which does not include travel or commuting time. "Household care" corresponds to hours spent on "caring for and helping household members," which does not include housework. "Incidence of household care 8–5" is equal to one if the respondent reported nonzero household care between 8 AM and 5 PM. For work and household care hours, we restrict the sample to those who report nonzero time spent on work-related activities. For the incidence measure, we restrict the sample to those who report nonzero time spent on work-related activities at the work site. Each column reports the coefficient on the "female" dummy with various controls. Column 1 includes day and year fixed effects. Column 2 includes age, education category, and race fixed effects. Column 3 adds usual weekly hours reported in the CPS. Column 4 only includes workers who reported usual weekly hours of less than 50.

Source: Data are from the 2003–2018 ATUS

Illustration confidence intervals

Go back

VOL. 109 NO. 1 SAMPAT AND WILLIAMS: HOW DO PATENTS AFFECT FOLLOW-ON INNOVATION 219

TABLE 2—PATENTS AND FOLLOW-ON INNOVATION ON HUMAN GENES CLAIMED IN ACCEPTED/REJECTED PATENT APPLICATIONS: REGRESSION ESTIMATES

	log of follow-on innovation in 2011–2012	Any follow-on innovation in 2011–2012
	(1)	(2)
Panel A. Scientific publications		
Patent granted	0.0019	−0.0014
	(0.0060)	(0.0054)
Mean of dependent variable	0.1104	0.1094
Observations	15,524	15,524
Panel B. Clinical trials		
Patent granted	0.0006	−0.0015
	(0.0080)	(0.0043)
Mean of dependent variable	0.1038	0.0659
Observations	15,524	15,524
Panel C. Diagnostic test		
Patent granted	–	−0.0092
	–	(0.0056)
Mean of dependent variable	–	0.1199
Observations	–	15,524

Notes: This table estimates differences in follow-on innovation on genes claimed in at least one granted patent relative to genes claimed in at least one patent application but never in a granted patent. The sample for these regressions is constructed from gene-level data, and includes genes claimed in at least one patent application in our USPTO human gene patent application sample ($N = 15{,}524$). Each coefficient is from a separate regression. Estimates are from ordinary-least-squares models. Heteroskedasticity robust standard errors.

more econometrically appropriate than modeling the outcome in levels. We focus on the log of follow-on innovation and (separately) an indicator for any follow-on innovation.

Given the absence of strong visual evidence for a difference in follow-on innovation across patented and non-patented genes, our focus here is on what magnitudes of effects can be ruled out by our confidence intervals. Across these specifications, our 95 percent confidence intervals tend to reject declines or increases in follow-on innovation on the order of more than 5–15 percent. For brevity, we focus on interpreting the log coefficients. For our measures of follow-on scientific research (publications; panel A of Table 2) and commercialization (clinical trials; panel B of Table 2), the 95 percent confidence intervals can reject declines or increases of more than 2 percent. For our measure of diagnostic test availability (only measured as a

Illustration p-value

Go back

Müller and Schwieren (J. of Business Economics, 2020)

Table 5 Tobit regression on y, the amount returned by Player 2 for those players who received a positive amount from Player 1

Variable	Coeff.	SE	p-value
x	1.645***	0.269	0.000
Neuroticism	– 0.004	0.043	0.933
Extraversion	0.005	0.046	0.915
Opennness	0.078	0.065	0.236
Agreeableness	– 0.027	0.049	0.580
Conscientiousness	0.080	0.049	0.220
Female	– 0.061	2.133	0.904
Age	0.039	0.369	0.915
n	47		
pseudo R^2	0.103		

This table shows the coefficients (2nd column), standard errors (3rd column) and p-values (4th column) from a tobit regression. The dependent variable is the amount returned by Player 2

*, **, *** Indicate significance at the 10%, 5% and 1% level respectively

Illustration t-statistics

Go back

Cawley (JHR, 2004): The Impact of Obesity on Wages

Table 2
Coefficients and t-Statistics from Log Wage Regressions for Males

	White Males			Black Males			Hispanic Males		
	OLS	OLS with Lag Weight	Fixed Effects	OLS	OLS with Lag Weight	Fixed Effects	OLS	OLS with Lag Weight	Fixed Effects
Column Number	1	2	3	4	5	6	7	8	9
BMI	−0.001	−0.003	−0.0001	0.004	0.005	0.003	−0.007	−0.009	−0.002
	(−0.83)	(−1.54)	(−0.21)	(2.19)	(2.04)	(1.96)	(−3.12)	(−3.15)	(−1.03)
Weight in pounds	−0.0002	0.0005	0.0001	0.0006	0.0007	0.0005	−0.0011	−0.001	−0.0003
	(−1.01)	(−1.68)	(0.07)	(2.21)	(2.06)	(1.96)	(−3.47)	(−3.18)	(−0.90)
Underweight	−0.14	0.005	−0.035	−0.099	−0.046	0.013	0.029	0.107	−0.005
	(−3.05)	(0.07)	(−1.12)	(−2.75)	(−0.70)	(0.28)	(0.44)	(1.13)	(−0.12)
Overweight	0.039	0.016	0.022	0.031	0.019	0.014	−0.025	−0.021	0.018
	(3.04)	(1.05)	(2.63)	(1.87)	(0.94)	(1.14)	(−1.13)	(−0.82)	(1.15)
Obese	−0.033	−0.075	0.013	0.043	0.042	0.031	−0.066	−0.100	0.023
	(−1.73)	(−3.05)	(0.89)	(1.80)	(1.29)	(1.64)	(−2.21)	(−2.41)	(0.91)
Number of observations	29,410	12,410	29,410	13,414	6,128	13,414	9,070	4,079	9,070

Notes:
1) Data: NLSY males.
2) One of three measures of weight is used: BMI, weight in pounds (controlling for height in inches) or the three indicator variables for clinical weight classification: underweight, overweight, and obese (where healthy weight is the excluded category).
3) For BMI and weight in pounds, coefficients and *t* statistics are listed. For indicators of clinical weight calssification, the percent change in log wages associated with a change in the indicator variable from 0 to 1 and *t* statistics are listed.
4) Other regressors include: number of children ever born, age of youngest child, general intelligence, highest grade completed, mother's highest grade completed, father's highest grade completed, years of actual work experience, job tenure, age, year, and indicator variables for marital status, county unemployment rate, current school enrollment, part-time job, white collar job, and region of residence.

Cawley

Applied Econometrics

Contact

Helmut Farbmacher
office.econometrics@mgt.tum.de