3. Statistical Tests

- So far, we only discussed how to obtain consistent estimates of a parameter and how to interpret regression estimates
- An important part of empirical work is to test whether the estimated parameters differ from a hypothesized quantity
- We obtain the estimate of $\hat{\beta}$ from a sample
- To do a statistical test we need to know something about the distribution of $\hat{\beta}$
- Useful thought experiment: think of the variance in the obtained estimates $\hat{\beta}$ when you would draw different samples from a population
  - When the variance of $\hat{\beta}$ is small then you can be very certain that $\hat{\beta}$ is close to the true β
  - When the variance of $\hat{\beta}$ is large then it may be far away from β and you probably don't learn so much from $\hat{\beta}$

Your Task **Simulated data set**

Create a new notebook in which you generate a sample with 400 observations where we know that the CEF is $y = 200 + 2x$ (put it all in one Colab cell):

- Create a variable which sets the number of observations:
  `n=400`
- Create DataFrame with n rows and columns x and y:
  `df=pd.DataFrame(index=range(n), columns=['x','y'])`
- Set x to a vector of n normally distributed random variables:
  `df['x']=np.random.normal(100,15,n)`
- Set y according to the above CEF and add some noise:
  `df['y']=200+2*df['x'] + np.random.normal(0,500,n)`
- Add a regression of `y` on `x`
- Run the script several times (each time a new sample is drawn) and write down (& compare) the estimated coefficients of `x`
- Save the notebook

3.1 Testing Hypotheses about a Parameter

Consider again the bivariate case where

$$\hat{\beta}_1 = \frac{\frac{1}{N}\sum_{i=1}^{N}(X_i - \bar{X})(Y_i - \bar{Y})}{\frac{1}{N}\sum_{i=1}^{N}(X_i - \bar{X})^2}$$

This can be rewritten (using that $Y_i = \beta_0 + \beta_1 X_i + e_i$) to become

$$\hat{\beta}_1 = \beta_1 + \frac{\sum_{i=1}^{N}(X_i - \bar{X})\, e_i}{\sum_{i=1}^{N}(X_i - \bar{X})^2}$$

→ The estimate is the sum of the population value and a function of the residuals

When the residuals are independent and follow $N(0, \sigma^2)$ the variance of $\hat{\beta}$ is

$$V\,[\hat{\beta} | X_1, X_2, \dots, X_N] = \frac{\sigma^2}{\sum_{i=1}^{N}(X_i - \bar{X})^2}$$

→ The estimated $\hat{\beta}$ fluctuate around the true β with variance $\frac{\sigma^2}{\sum_{i=1}^{N}(X_i - \bar{X})^2}$

We can use this to construct a test statistics for a coefficient

- Our Null hypothesis is $H_0: \beta = 0$
- The alternative hypothesis is $H_1: \beta \neq 0$

Note that as $\hat{\beta} \sim N\left(\beta, V(\hat{\beta})\right)$ we have that

$$\frac{\hat{\beta} - \beta}{sd(\hat{\beta})} \sim N(0,1)$$

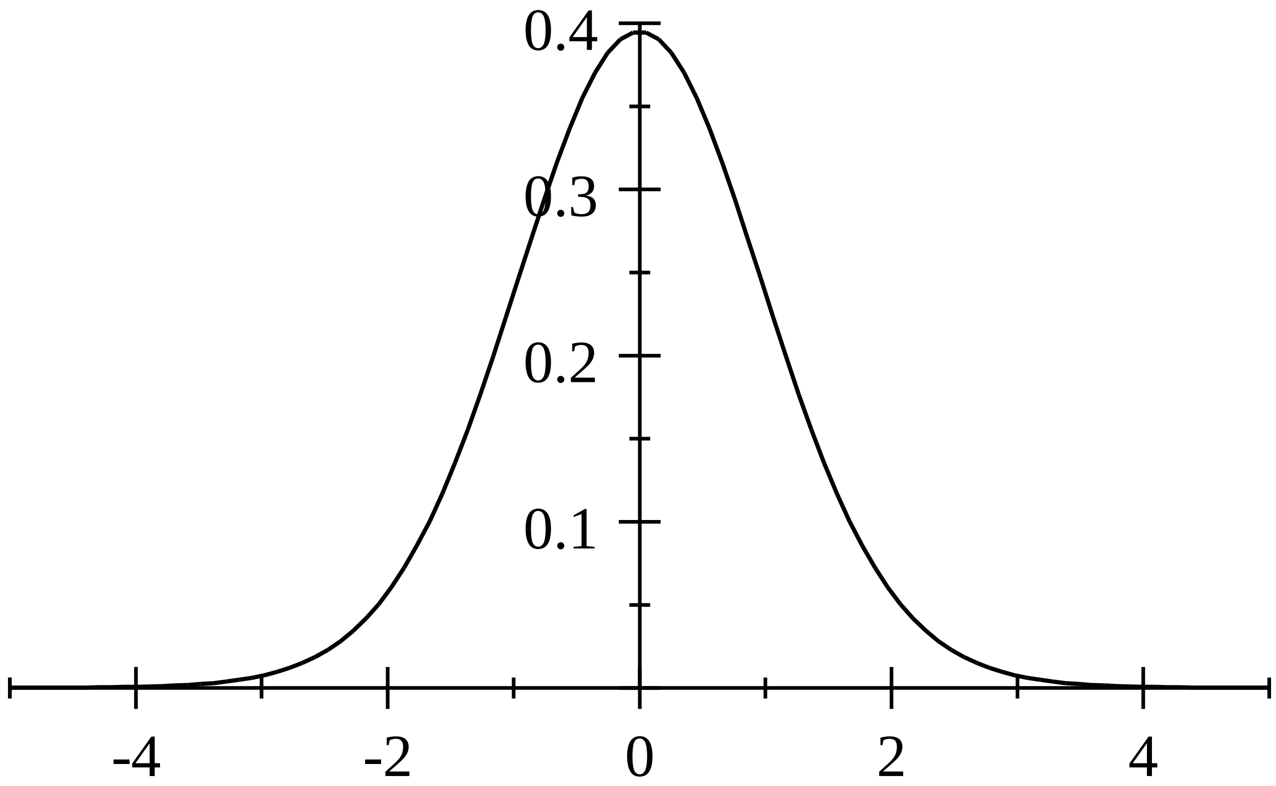
- As we do not know σ^2 and thus $sd(\hat{\beta})$ we have to estimate it from the data
- This estimate is the standard error $se(\hat{\beta})$ (→ see an econometrics textbook)
- Note: $se(\hat{\beta})$ is itself a random variable as it is an estimate based on the sample, but one can show:

  $\frac{\hat{\beta}-\beta}{se(\hat{\beta})}$ follows a Student‘s t-distribution with n-2 degrees of freedom:

$$\frac{\hat{\beta} - \beta}{se(\hat{\beta})} \sim t(N-2)$$

Note: The t-distribution is close to the standard normal distribution

Example: Density of $t(25)$



The Multivariate Case

- One can show analogously when β is a vector that

$$\hat{\beta} = \beta + \left[\sum_{i=1}^{N} X_i \, X_i'\right]^{-1} \sum_{i=1}^{N} X_i \, e_i$$

- In matrix notation

$$\hat{\beta} = \beta + (X'X)^{-1}X'e$$

- When the residuals are normally distributed & have the same variance

$$e \sim N\left(0, \sigma^2 I_N\right)$$

where I_N is the $N \times N$ identity matrix

- One can show that the vector of parameter estimates

$$\hat{\beta} \sim N\left(\beta, \sigma^2 (X'X)^{-1}\right)$$

- And (if there are k parameters to estimate, for each $j = 1, \ldots, k$)

$$\frac{\hat{\beta}_j - \beta_j}{se\left(\hat{\beta}_j\right)} \sim t(N - k - 1)$$

Hence, to test the Null hypothesis that $\beta_j = 0$ we can look at the $t\text{−}statistic$

$$t_{\hat{\beta}} = \frac{\hat{\beta}_j}{se(\hat{\beta}_j)}$$

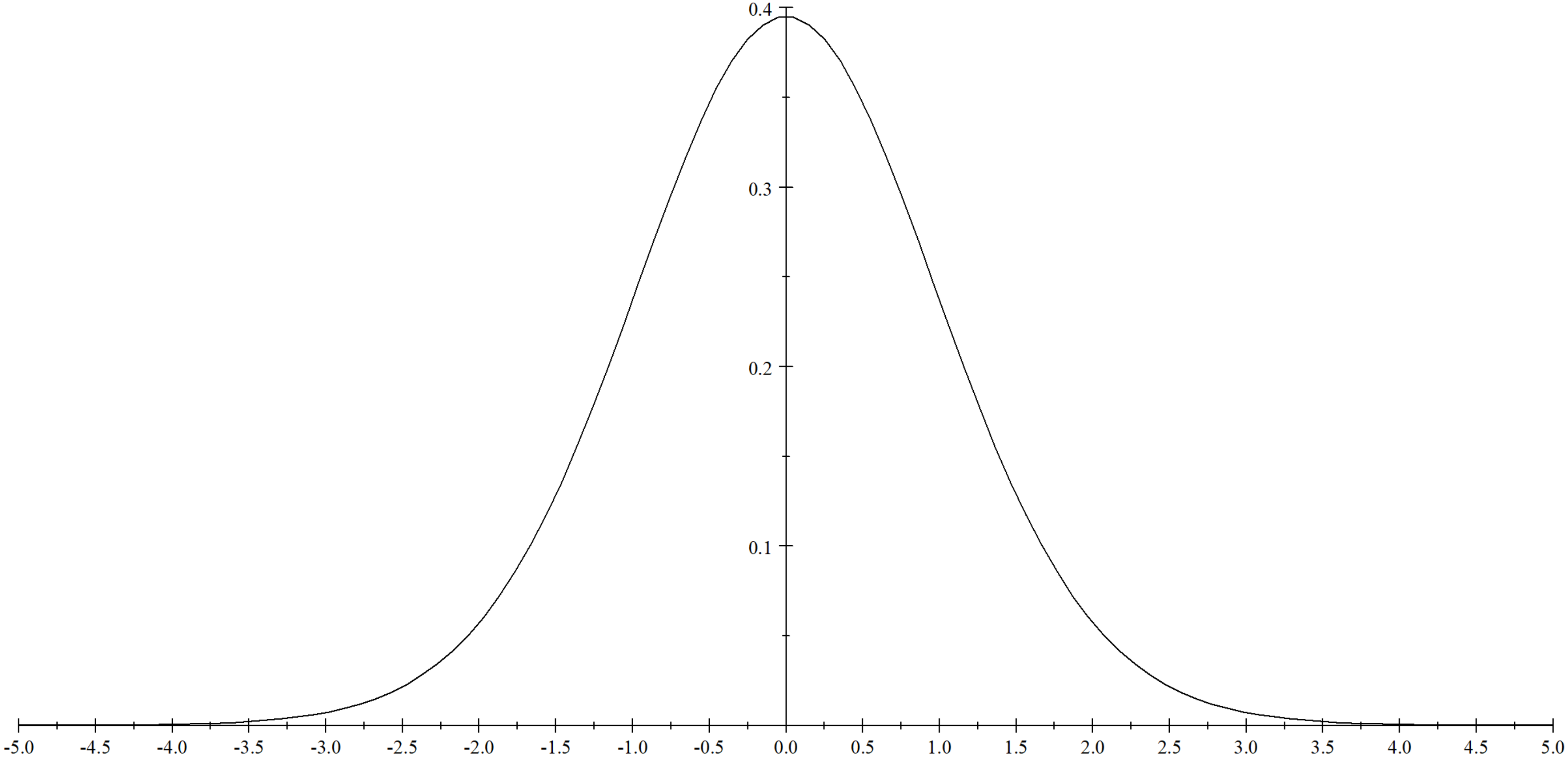
- When we consider a two-sided test,
  - we will reject H_0 whenever $t_{\hat{\beta}}$ is too large or too small
  - then it is unlikely that we would obtain an estimate $\hat{\beta}$ if the true β were 0
- Significance level α: likelihood that H_0 is rejected when it is in fact true
- Hence, we will reject H_0 at a significance level α if

$$\left|t_{\hat{\beta}}\right| > t_{\frac{\alpha}{2}}$$

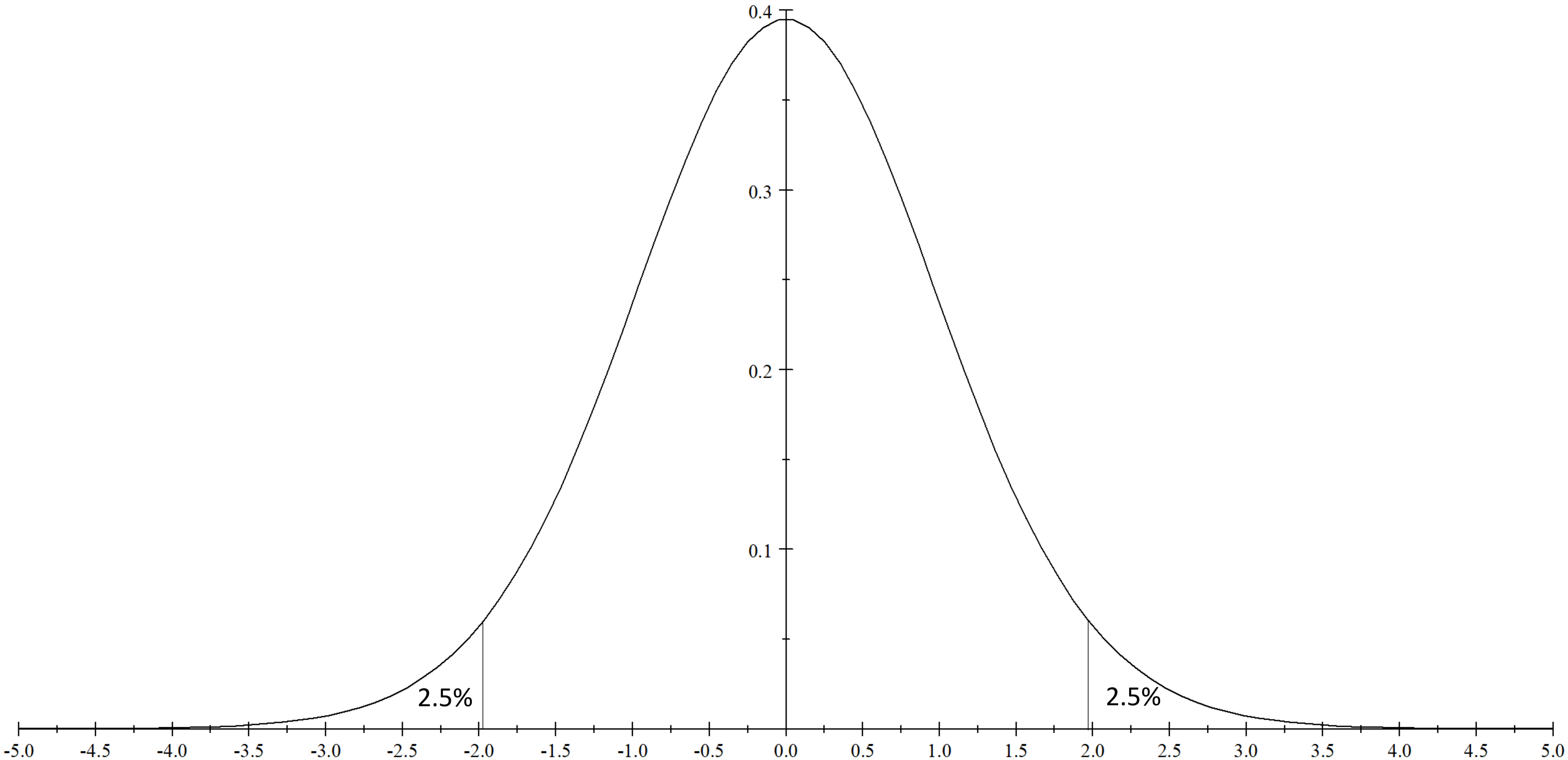
where $t_{\frac{\alpha}{2}}$ is the respective quantile of the t-distribution

Example: When $\alpha = 0.05$ we compare $\left|t_{\hat{\beta}}\right|$ with $t_{\frac{\alpha}{2}} = 1.962$ for $n - k - 1 = 1000$

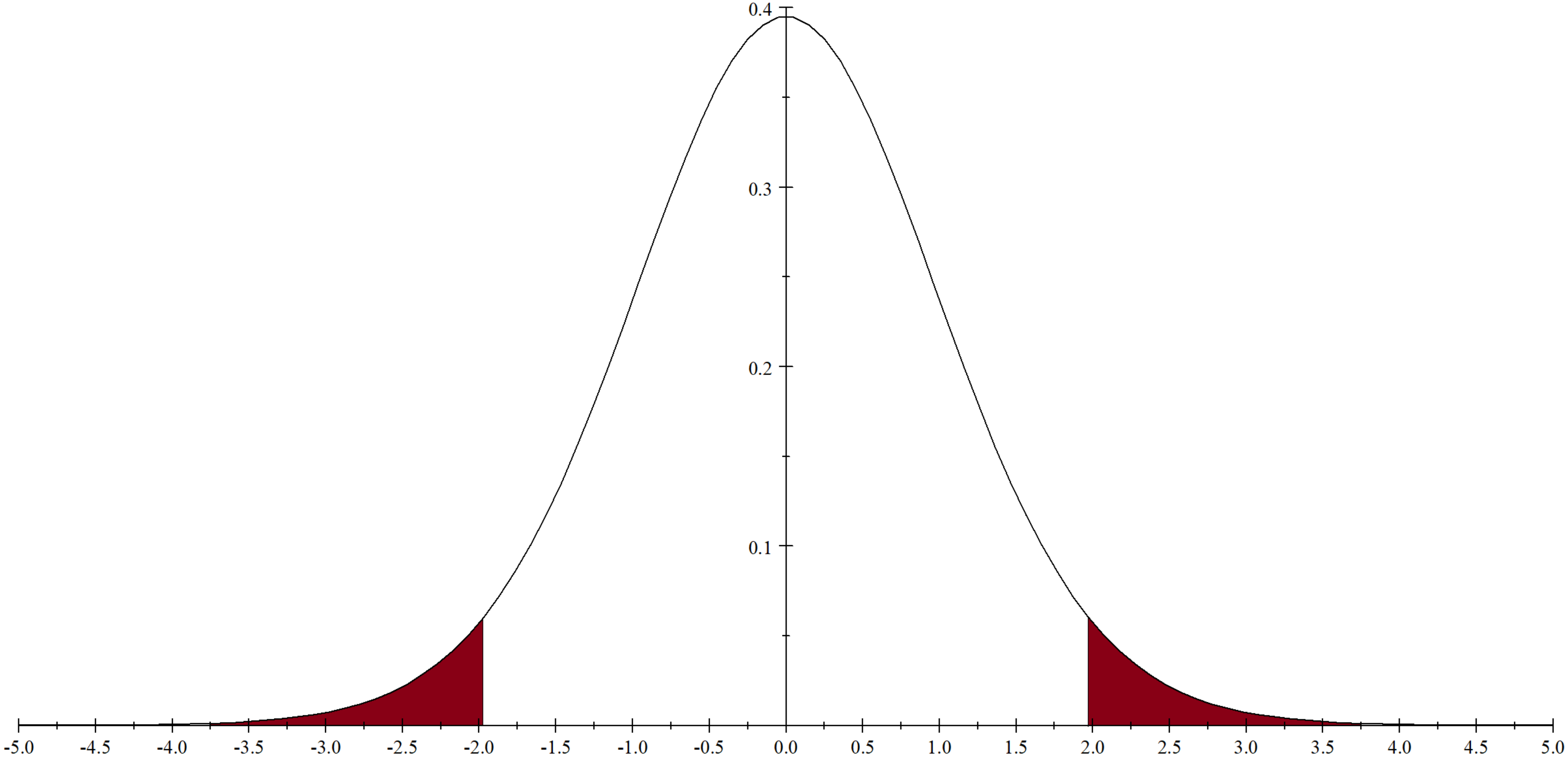
Consider the t-distribution under the Null hypothesis (if the true $\beta = 0$)



The likelihood that we will observe a t-value above 1.96 or below -1.96 is 5%:



We reject the Null hypothesis at a significance level of 5% if we observe a t-value above 1.96 or below -1.96



- Standard errors reported in regression tables yield the (estimated) standard deviation of the estimated $\hat{\beta}$

- The standard errors are used to construct the t-statistics

- From that we can compute the p-values (reported by statsmodels automatically)

Intuition: This gives us the following

- If I would draw different samples (of the given size) I would obtain different estimators $\hat{\beta}$
- What is the standard deviation of these $\hat{\beta}$?
- When this is small: we are close to the true β
- When this is large: there is much noise and therefore it is likely that the estimated $\hat{\beta}$ is further away from the true β

What a p-values tell us:

- What is the probability of obtaining an estimate that is at least as "extreme" (distant from 0) as the value of $\hat{\beta}_j$ I have estimated when the true value of were $\beta = 0$
- When this probability is smaller (computed using the estimated standard errors) we can be more certain that the true β is not zero
- For instance, when $p < 0.05$ for a certain coefficient, we say that the coefficient is *statistically significant at the* 5% *level*

- We mark this in a typical regression table in a paper with
  - * if $p < 0.10$
  - ** if $p < 0.05$ and
  - *** if $p < 0.01$

Your Task — Simulated data set

- Open the `SimulateData` notebook
- Run the script again several times
- Compare the regression estimates
  - Compare the estimated coefficients
  - Also look at the standard errors, t-values and p-values
- Increase the number of observations to 10.000 (`n=10000`)
- Repeat the exercise

Your Task

Simulated data set

- Change your do file to generate „pure noise“

```
df['y']=200 + np.random.normal(0,500,n)
```

- Run the do file 20 times
(each time is like drawing a new sample from the population)

- Count the number of times you obtain a p-value for the coefficient of `x` that is smaller than *0.1*

3.2 (Robust) Standard Errors

- We made the assumption that
  - the residuals are normally distributed
  - they have the same variance for all observations
- In general we cannot be sure that this assumption that $e \sim N\left(0, \sigma^2 I_N\right)$ will hold
  - Implies that residuals are normally distributed and
  - that variance of residuals is constant → so-called *Homoscedasticity*
- But we can check to what extent this seems plausible
  - Graphically: plotting the residuals
  - Statistically: test the Null hypothesis that the residuals are homoscedastic (Breusch-Pagan test)

- But it always holds that

$$\hat{\beta} = \beta + \left[\sum_{i=1}^{N} X_i X_i'\right]^{-1} \sum_{i=1}^{N} X_i e_i$$

  That is, $\hat{\beta}$ is the sum of the true β plus a function of the residual
- One can now show (see Angrist/Pischke, p. 45) that $\hat{\beta}$ is asymptotically normally distributed
  - with probability limit β (i.e. sample size grows → $\hat{\beta}$ comes closer to β)
  - and a covariance matrix that can be estimated from X_i and the residuals
- So called "Robust" standard errors follow from this covariance matrix
- Such robust standard errors are reported in StatsModels if you use

```
reg= smf.ols('y~X', data=df).fit(cov_type='HC1')
```

- Are called *robust* because
  - they are derived without assuming that the variance of the residuals is independent of X_i (i.e. they allow for heteroscedasticity)
  - and in large enough samples they provide accurate hypothesis testing without further distributional assumptions

Interdependent Observations & Clustered Standard Errors

- Standard errors are estimated under the assumption that
  - data are *independent observations*
  - or in other words each observation is a random draw from the population
- Very often this is not the case, for instance when
  - We observe several employees that come from the same firm
  - Or we observe the same employee at different dates
- The residuals will be correlated when the observations come from the same person or colleagues in the same firm
- Standard errors that are estimated assuming independence of employee observations are then biased
  - They are typically too small
  - p-Values are then smaller than they should be
  - We run into the danger of rejecting the null hypothesis too often

Two possible solutions:

- Use only firm level observations
  - For instance build a data set that includes one observation per firm
  - (in the above example we would have 1.000 observations)
  - And use the average job satisfaction in the firm as dependent variable
  - Hint: Pandas `groupby()` helps to build these aggregated data sets

- Use a different method for estimating standard errors
  - There are procedures that account for the interdependence of observations within groups or clusters
  - With StatsModels use
  - `reg= smf.ols('y~X', data=df).fit(cov_type='cluster', cov_kwds={'groups’: df['groupvar’]})`

Your Task — Clustered Standard Errors

- Open your ManagementPractices.py file
- Run the script
- Inspect the DataFrame
- Note:
  - For each firm (account_id) there are observations from different years
  - These observations will not be independent and thus standard errors will be biased
- Copy your regression commands to have them three times in the cell
- In regression 2 estimate heteroscedasticity robust standard errors adding

  `.fit(cov_type='HC1')`
- In regression 3 estimate cluster robust standard errors adding

  `.fit(cov_type='cluster', cov_kwds={'groups': df['account_id']})`
- Compare the standard errors in the three regressions

3.3 Confidence Intervals

- It is often useful to construct so called **confidence intervals** to illustrate the uncertainty we still have when estimating a parameter
- Again we have to define a level of significance first
- When constructing a 95% confidence band, we want to find k such that

$$Pr\left(\hat{\beta}_j - k \leq \beta \leq \hat{\beta}_j + k\right) = 0.95$$

or equivalently

$$Pr\left(-k \leq \hat{\beta}_j - \beta \leq k\right) = 0.95$$

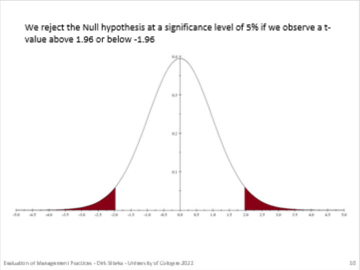
Recall

$$\frac{\hat{\beta}_j - \beta_j}{se\left(\hat{\beta}_j\right)} \sim t(N-k-1)$$

which is equivalent to

$$\hat{\beta}_j - \beta_j \sim se\left(\hat{\beta}_j\right) \cdot t(N-k-1)$$

Consider $Pr\left(-k \leq \hat{\beta}_j - \beta \leq k\right) = 0.95$



- Recall the percentiles $t_{0.025} \approx -1.96$ and $t_{0.975} \approx 1.96$
- Hence,

$$Pr\left(-1.96 \cdot se\left(\hat{\beta}_j\right) \leq \hat{\beta}_j - \beta \leq 1.96 \cdot se\left(\hat{\beta}_j\right)\right) = 0.95$$

- The confindence interval is thus $\left[\hat{\beta}_j - 1.96 \cdot se\left(\hat{\beta}_j\right), \hat{\beta}_j + 1.96 \cdot se\left(\hat{\beta}_j\right)\right]$
- The respective 90% interval is $\left[\hat{\beta}_j - 1.65 \cdot se\left(\hat{\beta}_j\right), \hat{\beta}_j + 1.65 \cdot se\left(\hat{\beta}_j\right)\right]$

Importantly: This means

- When we draw different samples, in 95% of the samples the 95% confidence interval will include the true value β
- This does not mean that we can be sure that the true value is in the interval
- But: it will be in the interval in about 19 of 20 samples drawn
- A 95% confidence interval also represents the set of values that are not statistically significantly different from the point estimate at the 5% level

RCT in Retailing: A Bonus for raising the average receipt

Manthei/Sliwka/Vogelsang (Management Science, 2021):

- Two experiments in one large region of a retail chain on a bonus for raining the average receipt

- Experiment 1: Bonus for district managers Nov 2015 – Jan 2016
  - 25 district managers (152 stores) randomly assigned to the treatment (Norm. Bonus)
  - 24 district managers (148 stores) in the control group

- Experiment 2: Bonus for *store managers* Nov 2016 – Jan 2017
  - 99 store managers assigned to the treatment Norm. Bonus
  - 95 store managers assigned to the treatment Simple Bonus
  - 95 store managers assigned to the control group

- Average receipt as KPI for bonus assignment in both experiments, experiment on a lower hierarchical level & includes a simplified bonus formula

Table 1. Main Effects of Experiments I and II

	Experiment I—District level			Experiment II—Store level		
	(1)	(2)	(3)	(4)	(5)	(6)
	Sales per Customer	*Sales per Customer*	CI 90%	*Sales per Customer*	*Sales per Customer*	CI 90%
Treatment effect						
Norm. Bonus	0.0020	–0.0240	[–0.1037; 0.0556]	–0.0162	–0.0099	[–0.0902; 0.0703]
	(0.0464)	(0.0475)		(0.0437)	(0.0478)	
Simple Bonus				0.0328	0.0347	[–0.0649; 0.1343]
				(0.0504)	(0.0594)	
Time FE	Yes	Yes		Yes	Yes	
Store/district FE	Yes	Yes		Yes	Yes	
District manager FE	No	Yes		No	Yes	
Store manager FE	No	No		No	Yes	
No. of observations	637	637		3,822	3,473	
Level of observations	District	District		Store	Store	
No. of districts/stores	49	49		294	294	
Cluster	49	49		50	50	
Within R^2	0.9427	0.9478		0.8473	0.8476	
Overall R^2	0.1043	0.1185		0.0497	0.0327	

Notes. The table reports results from a fixed effects regression with the sales per customer on the district/store level as the dependent variable. The regression accounts for time and store district fixed effects and adds fixed effects for district managers in column (2) and fixed effects for district and store managers in column (5). For experiment I, the regressions compare pretreatment observations (January 2015–October 2015) with the observations during the experiment (November 2015–January 2016). For experiment II, the regressions compare pretreatment observations (January 2016–October 2016) with the observations during the experiment (November 2016–January 2017). "Treatment effect" thus refers to the difference-in-difference estimator. All regressions control for possible refurbishments of a store. Observations are excluded if a store manager switched stores during the treatment period. Robust standard errors are clustered on the district level of the treatment start and displayed in parentheses. Columns (3) and (6) display 90% confidence intervals of the specification in columns (2) and (5), respectively. CI, confidence interval; FE, fixed effects.

* $p < 0.1$; ** $p < 0.05$; *** $p < 0.01$.

Appendix

Consider

$$\hat{\beta}_1 = \frac{\frac{1}{N}\sum_{i=1}^{N}(X_i - \bar{X})\,(Y_i - \bar{Y})}{\frac{1}{N}\sum_{i=1}^{N}\left(X_i - \bar{X}\right)^2}$$

We can rewrite this as

$$\frac{\sum_{i=1}^{N}(X_i - \bar{X})\,(\beta_0 + \beta_1 X_i + e_i - (\beta_0 + \beta_1 \bar{X} + \bar{e}))}{\sum_{i=1}^{N}\left(X_i - \bar{X}\right)^2}$$

$$= \frac{\sum_{i=1}^{N}(X_i - \bar{X})\,\beta_1 (X_i - \bar{X})}{\sum_{i=1}^{N}\left(X_i - \bar{X}\right)^2} + \frac{\sum_{i=1}^{N}(X_i - \bar{X})\,e_i}{\sum_{i=1}^{N}\left(X_i - \bar{X}\right)^2} - \frac{\sum_{i=1}^{N}(X_i - \bar{X})}{\sum_{i=1}^{N}\left(X_i - \bar{X}\right)^2}\,\bar{e}$$

and as $\sum_{i=1}^{N}(X_i - \bar{X}) = 0$ we obtain

$$\hat{\beta}_1 = \beta_1 + \frac{\sum_{i=1}^{N}(X_i - \bar{X})\,e_i}{\sum_{i=1}^{N}\left(X_i - \bar{X}\right)^2}$$

→ The estimate is the sum of the population value and a function of the residuals

Take

$$\hat{\beta}_1 = \beta_1 + \frac{\sum_{i=1}^{N}(X_i - \bar{X})\, e_i}{\sum_{i=1}^{N}\left(X_i - \bar{X}\right)^2}$$

When the residuals are independently normally distributed and have the same variance $e \sim N\left(0, \sigma^2\right)$

$$V\,[\hat{\beta}_1 \big| X_1, X_2, \ldots, X_N] = \frac{1}{\left(\sum_{i=1}^{N}\left(X_i - \bar{X}\right)^2\right)^2} V\left[\sum_{i=1}^{N}(X_i - \bar{X})\, e_i\right]$$

$$= \frac{\sum_{i=1}^{N}\left(X_i - \bar{X}\right)^2 V[e_i]}{\left(\sum_{i=1}^{N}\left(X_i - \bar{X}\right)^2\right)^2} = \frac{\sigma^2}{\sum_{i=1}^{N}\left(X_i - \bar{X}\right)^2}$$

Hence

$$\hat{\beta}_1 \sim N\left(\beta_1, \frac{\sigma^2}{\sum_{i=1}^{N}\left(X_i - \bar{X}\right)^2}\right)$$