Imperial College
Business School

Imperial means
Intelligent Business

Testing Times –

How to decide when to take an econometric result serious

by Ralf Martin (r.martin@imperial.ac.uk)

Objective for today

Understand the reliability of a regression result…

…assuming there is no bias or mis-specification of the model

We are talking about the known unknowns today

Go to **www.menti.com** and use the code **13 94 44 4**

Mentimeter

How would you decide decide if a dice is fair?

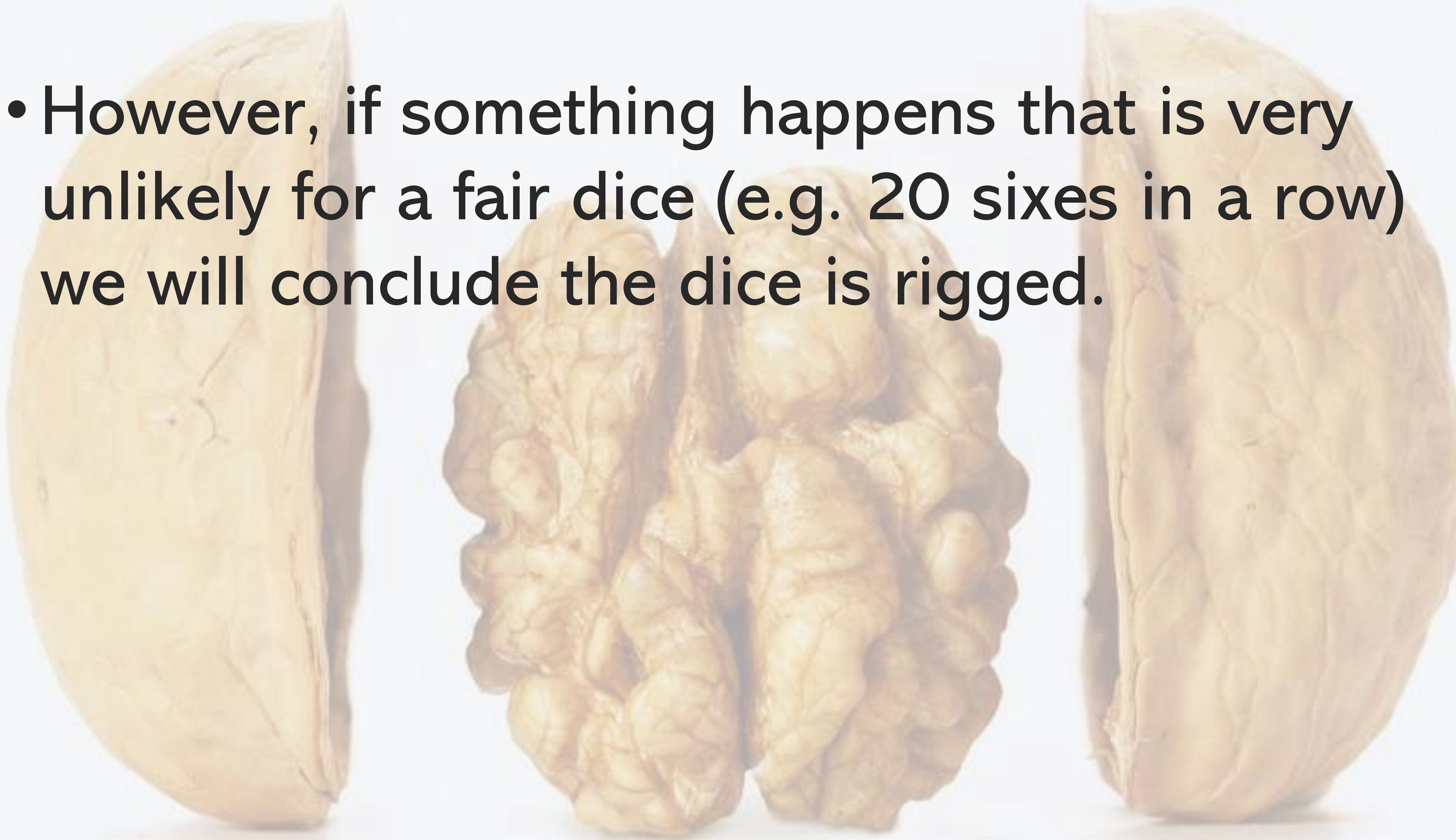
- I would throw a dice
- Roll it infinite number of times
- Probability of all 6 outcome should be the same
- If the probability for the outcome of each number is the same
- Experiment: roll it many many times and track results
- E(x) = 3.5
- Check the numbers on the dice, shape and weight
- Throw lots of times and compare to expected average of 3.5
- Take a large sample of rolls (100+) and look at the prob. distribution
- Run a lot of simulations
- repeat dice throws?
- Dice is always fair, because opportunities are based on luck !

Press S to show image

34

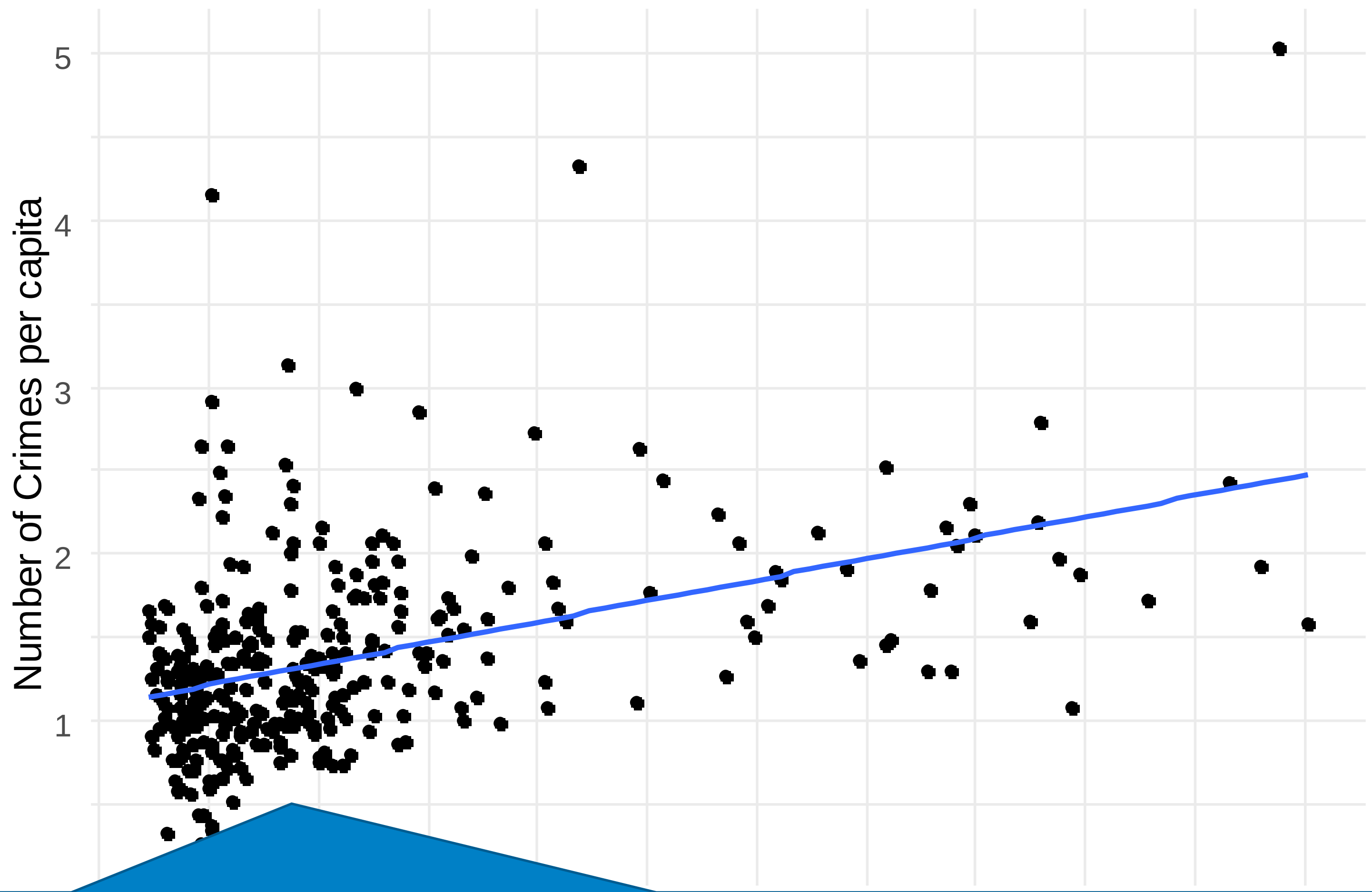
- We can never 100% certain if a dice is fair

- However, if something happens that is very unlikely for a fair dice (e.g. 20 sixes in a row) we will conclude the dice is rigged.



→ Hypothesis testing for dice in a nutshell

Hypothesis testing in for econometric models



$$\hat{Y} = \hat{\beta}_0 + \hat{\beta}_1 X$$

How likely is it to see a slope such as this…

- even if there is no relationship between foreigners and crime
- **and there is no endogeneity**

The distribution of our estimates

- To work out how likely a particular estimation outcome is given a hypothesis we need to know the distribution of the estimates

- For a distribution we need the notion of a random experiment (like throwing a dice)

- In the context of estimating an econometric model the random experiment is taking a random sample of a population

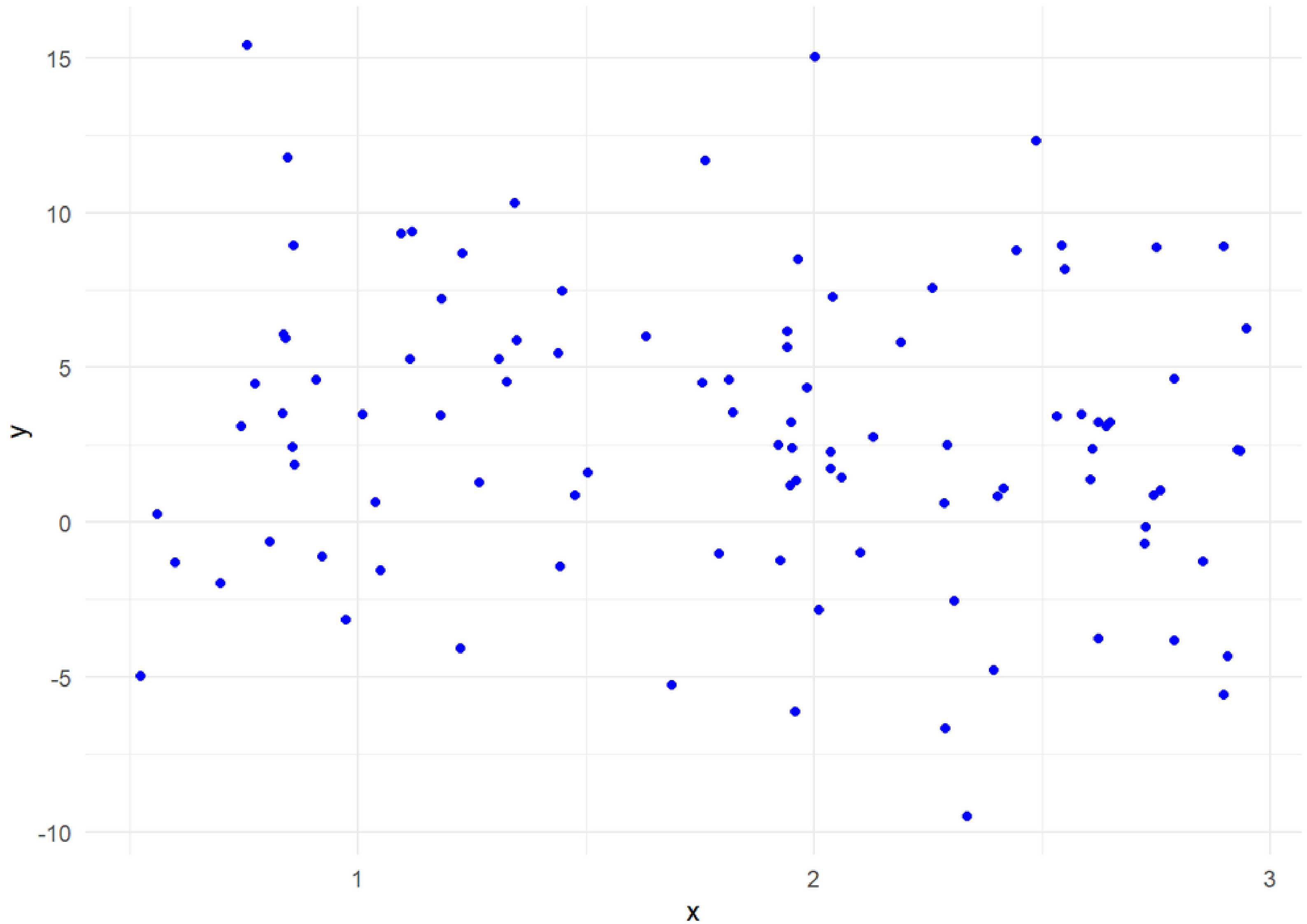
Monte Carlo Experiment

- Let's make the data ourselves
- E.g. suppose the true model is $Y_i = 2 + 0 \times X_i + \varepsilon_i$
- Here is how to do it in R

```
obs <- 100
x <- 0.5 + runif(obs)*2.5
sig=sqrt(5.5)*2
eps <- rnorm(obs,0,sig)
y <- 2 + x * 0 + eps

df=data.frame(x,y)

ggplot(df, aes(x, y))+geom_point(color="blue") +theme_minimal()
```



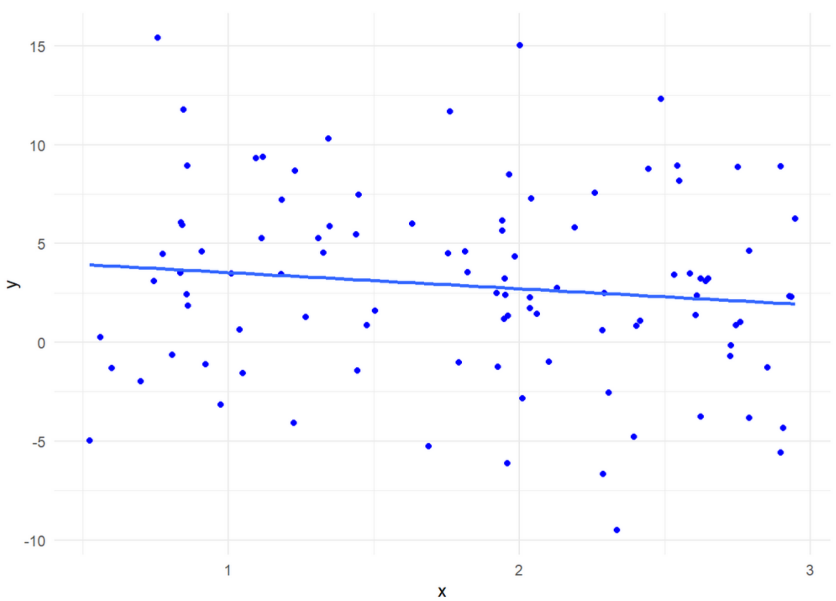
$\beta_1 = 0$

Let's run regression

```
monte1 <- lm(y ~ x , data = df)

summary(monte1)
```

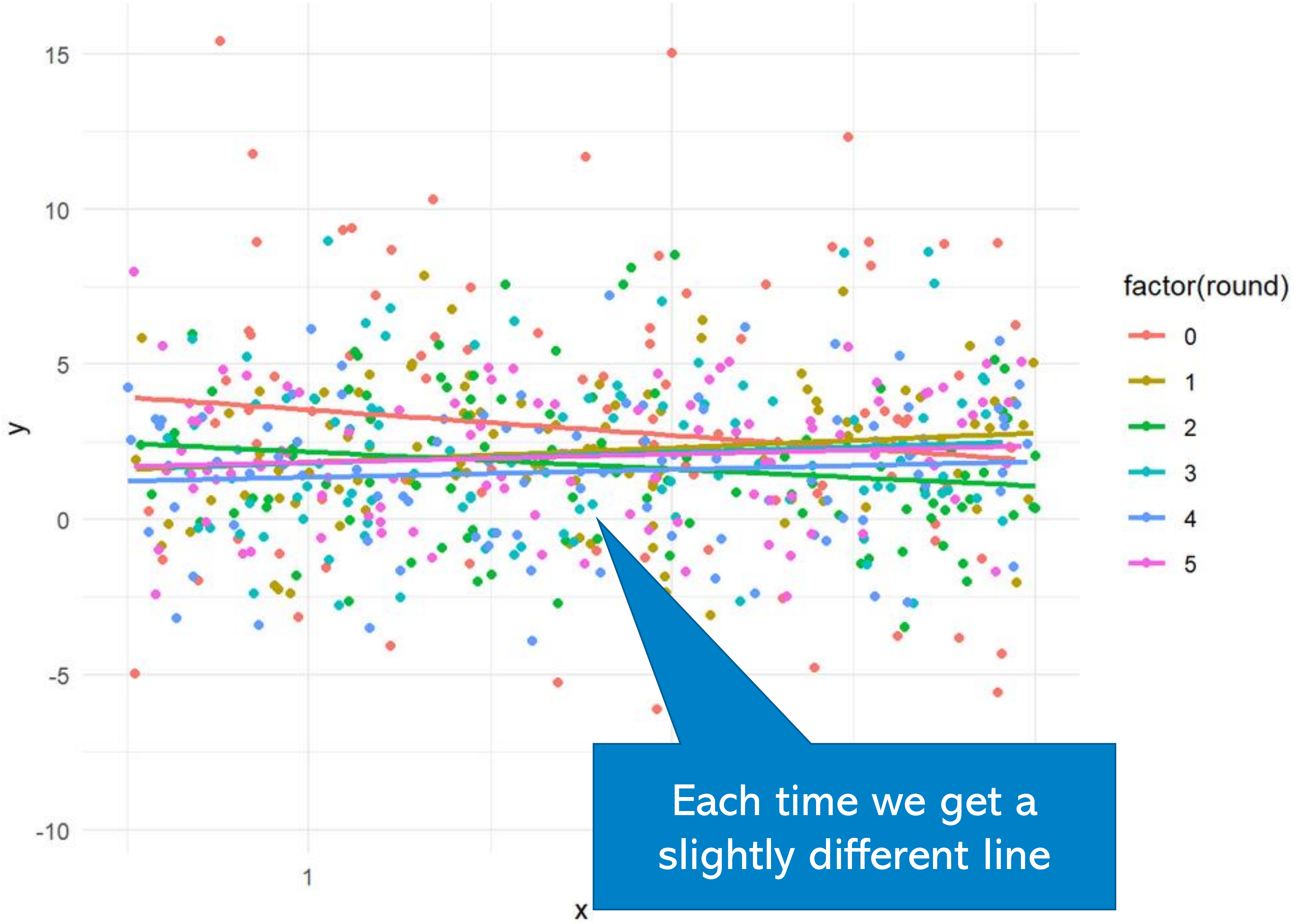
```
##
## Call:
## lm(formula = y ~ x, data = df)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -11.9561  -2.9585   0.0476   2.6857  12.3041
##
## Coefficients:
##             Estimate Std. Error t value Pr(>|t|)
## (Intercept)   4.3535     1.3215   3.294  0.00137 **
## x            -0.8188     0.6724  -1.218  0.22623
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 4.78[illegible] on 98 degrees of freedom
## Multiple R-squared:  0.01491,	Adjusted R-squared:  0.004855
## F-statistic: 1.483 on 1 and 98 DF,  p-value: 0.2262
```



How does it compare?

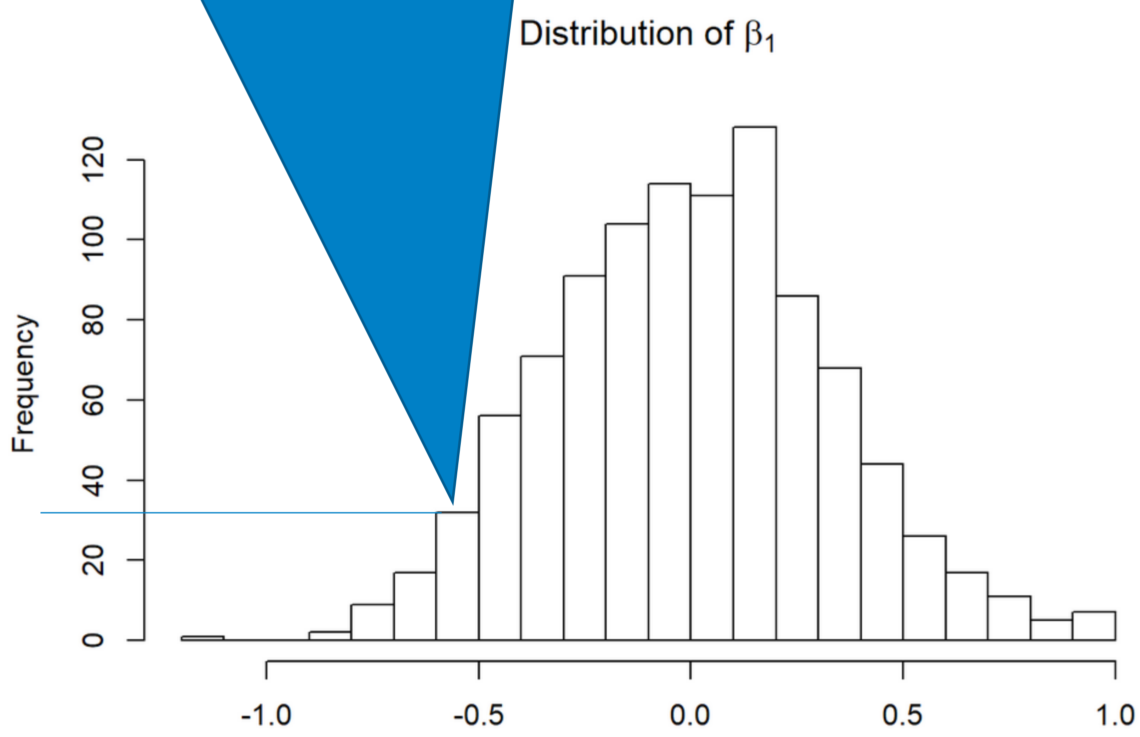
$$Y_i = 2 + 0 \times X_i + \varepsilon_i$$

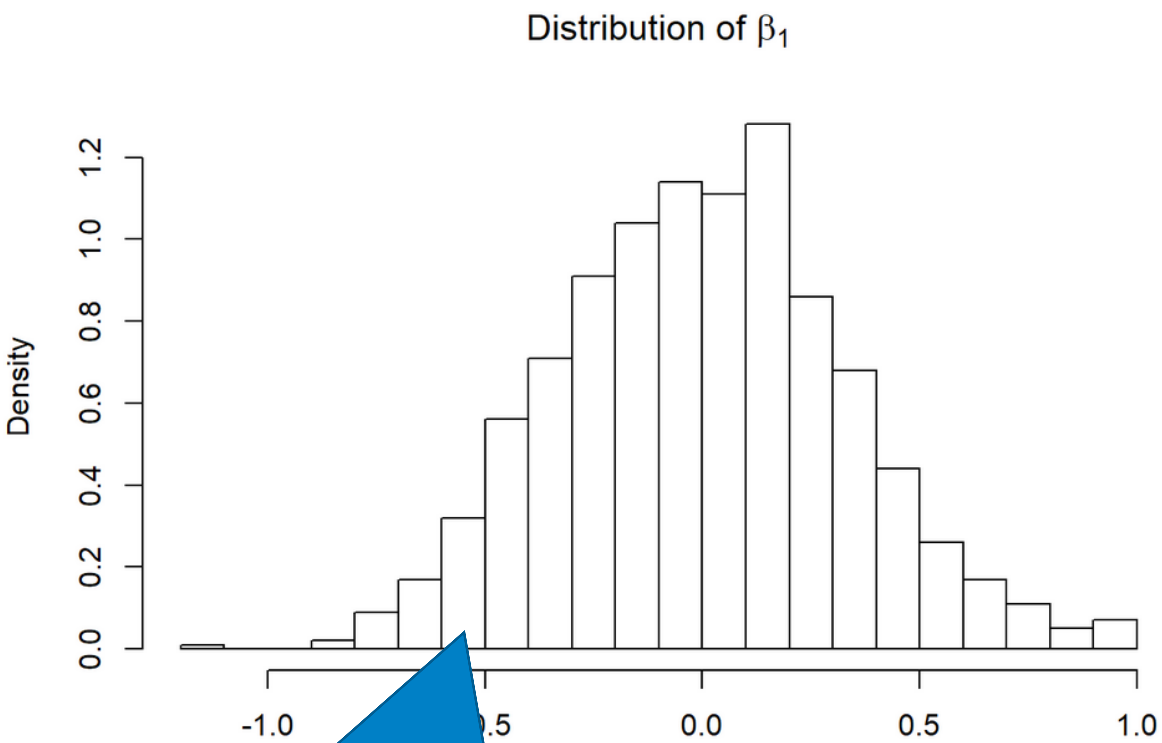
Let's do it many times



Let's look at histogram of β_1 for 1000 replications of drawing a sample

How many cases fall in a particular interval (e.g. about 30 between -0.6 & -0.5)





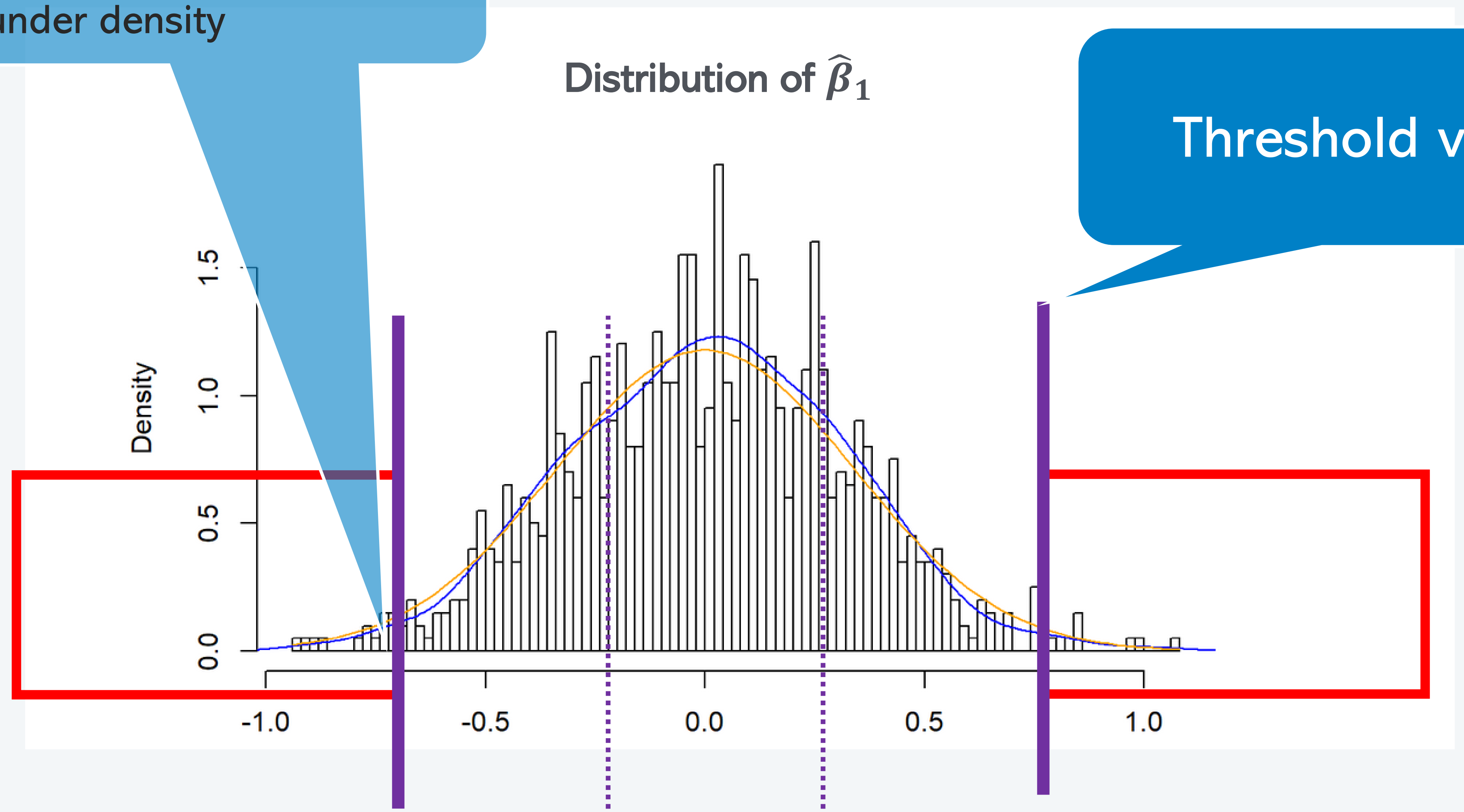
Surface area of bar represents what fraction of estimates fall between in a particular interval (e.g. about $0.3 \times 0.1 = 3\%$ between -0.6 & -0.5)

Let's look at histogram of β_1 1000 times

Distribution is very close to a normal distribution

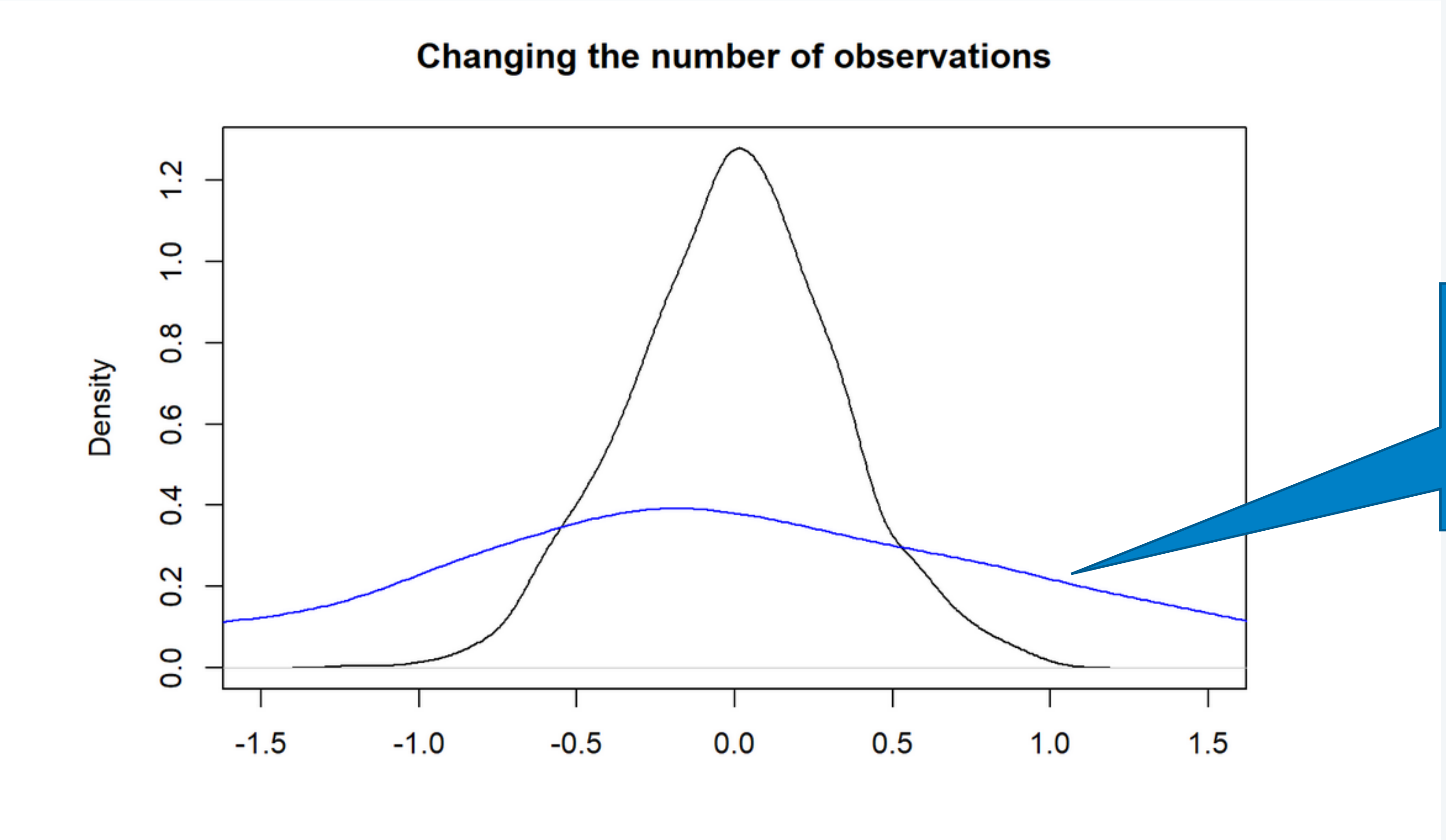
Probability that we have a $\hat{\beta}_1$ smaller than -0.75 is the sum of bars = area under density

Threshold value



Large vs small sample

The distribution is more dispersed for a sample of 10 (small sample) than for a sample of 100

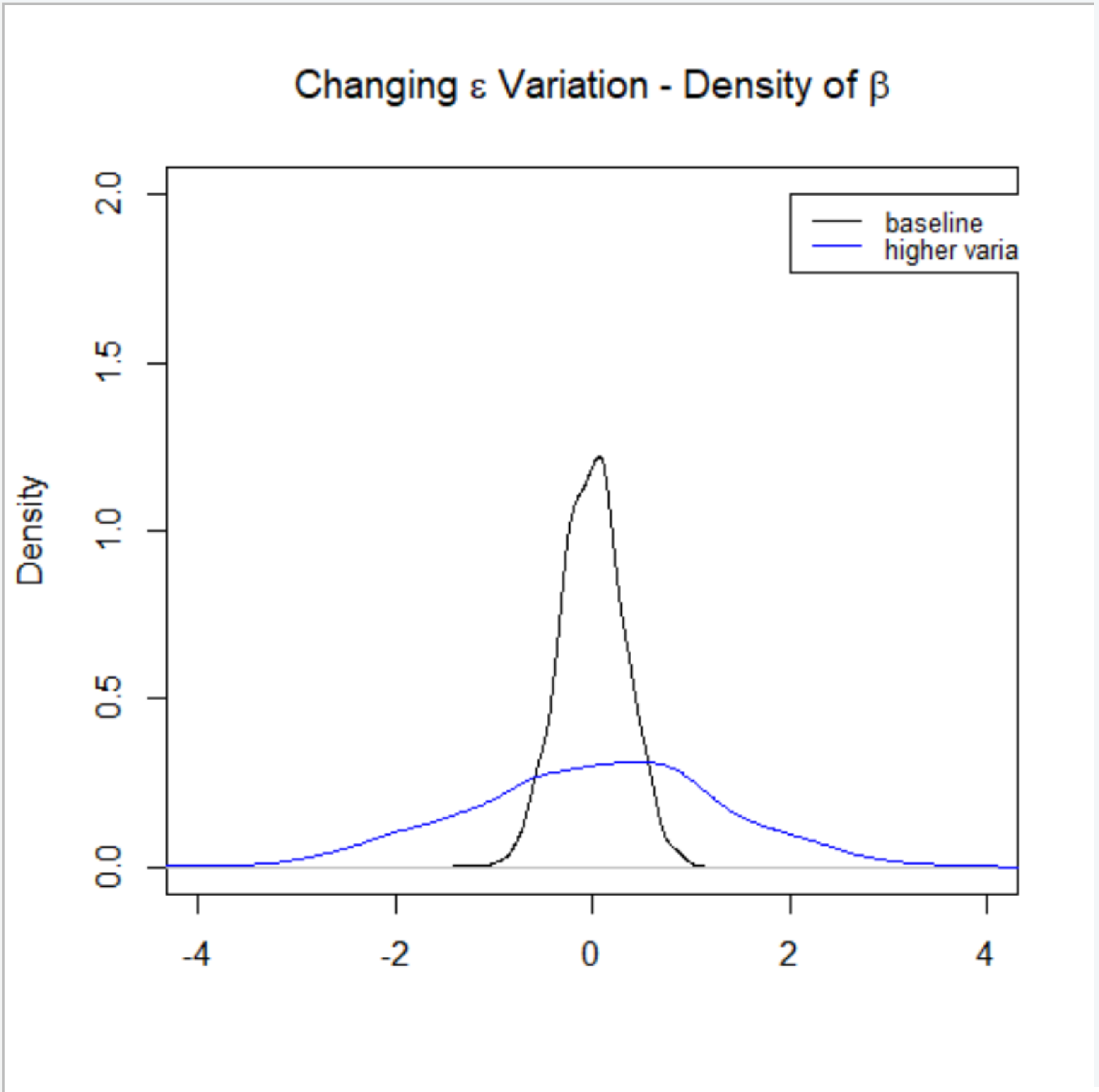


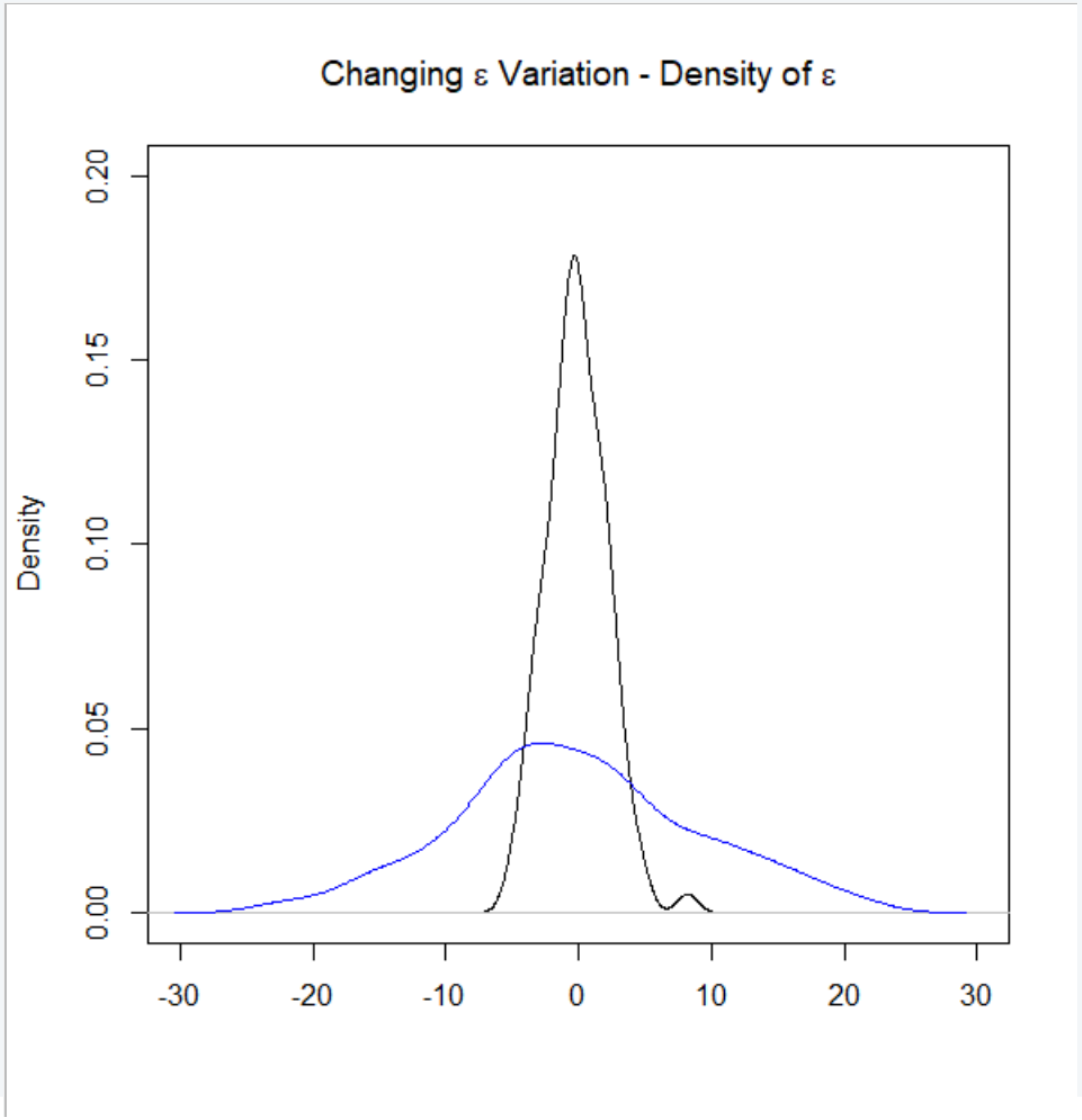
10 observations

More variation in $\boldsymbol{\varepsilon}$

$$Y_i = 2 + 0 \times X_i + \varepsilon_i$$

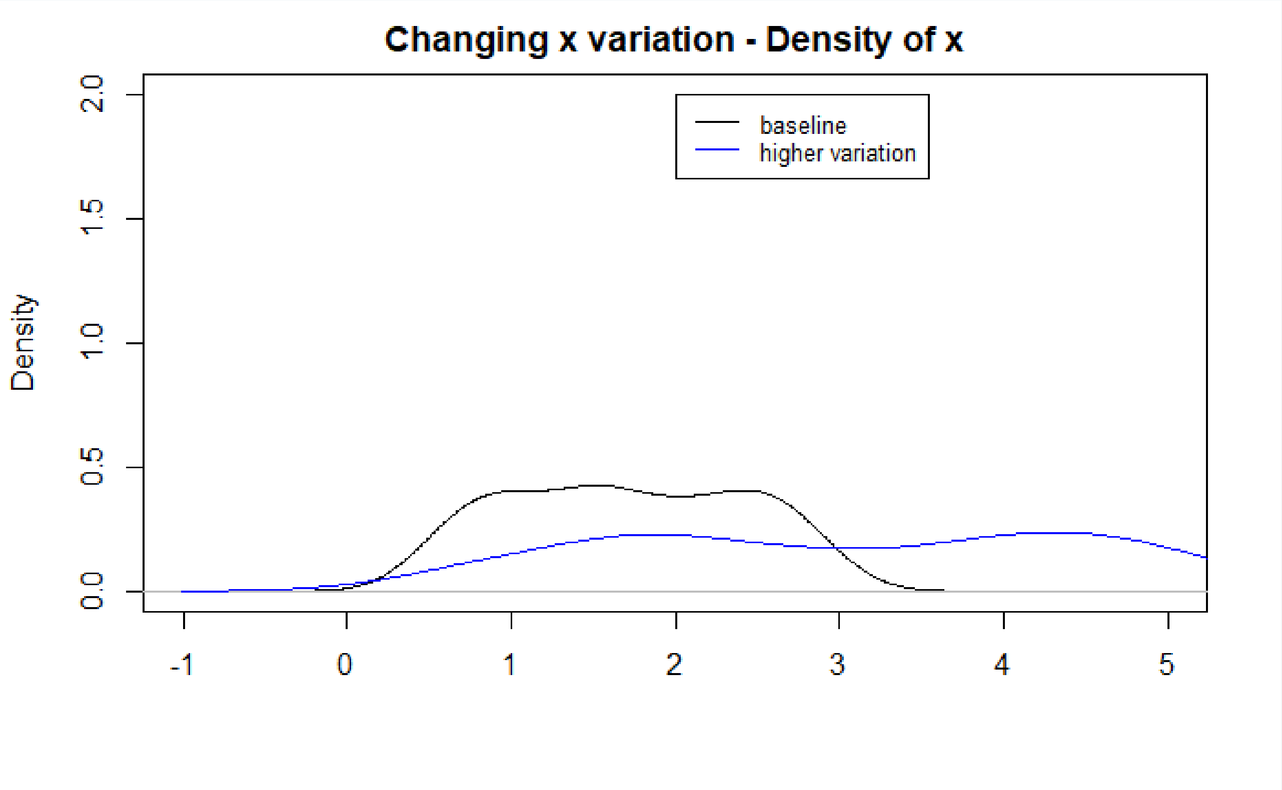
There is a lot more we don't know about Y

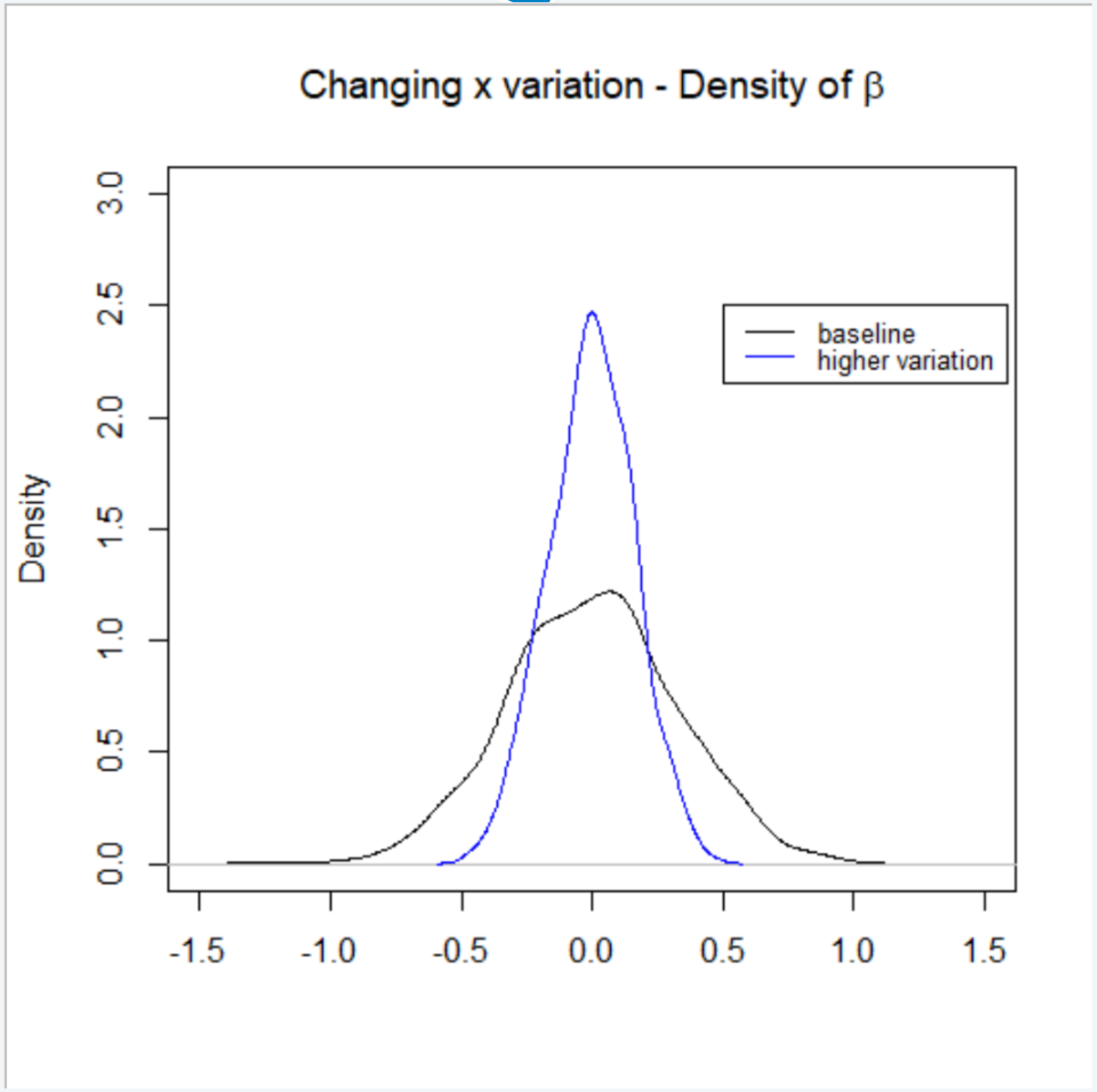




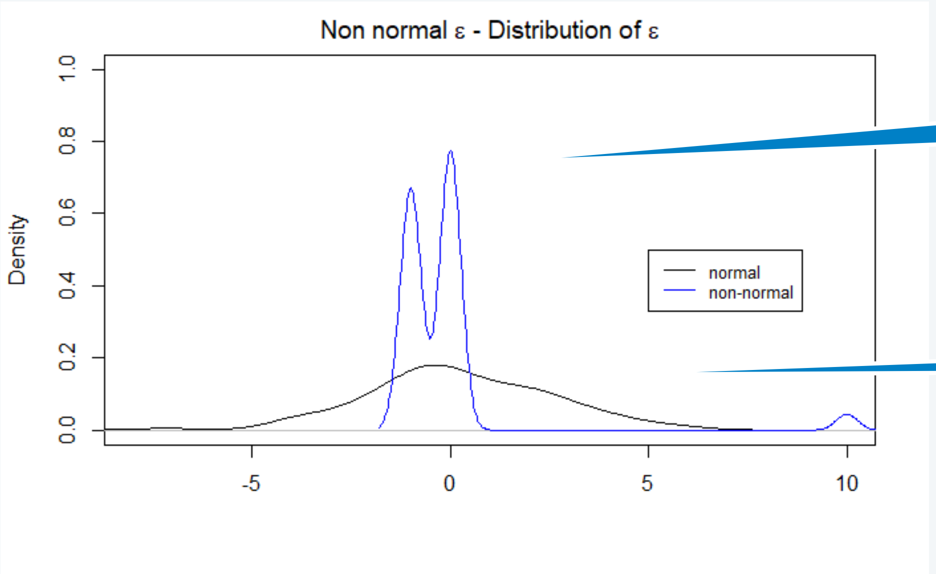
Dispersed X vs not so dispersed X

If X varies more our estimate of β becomes more precise





Non normal ϵ

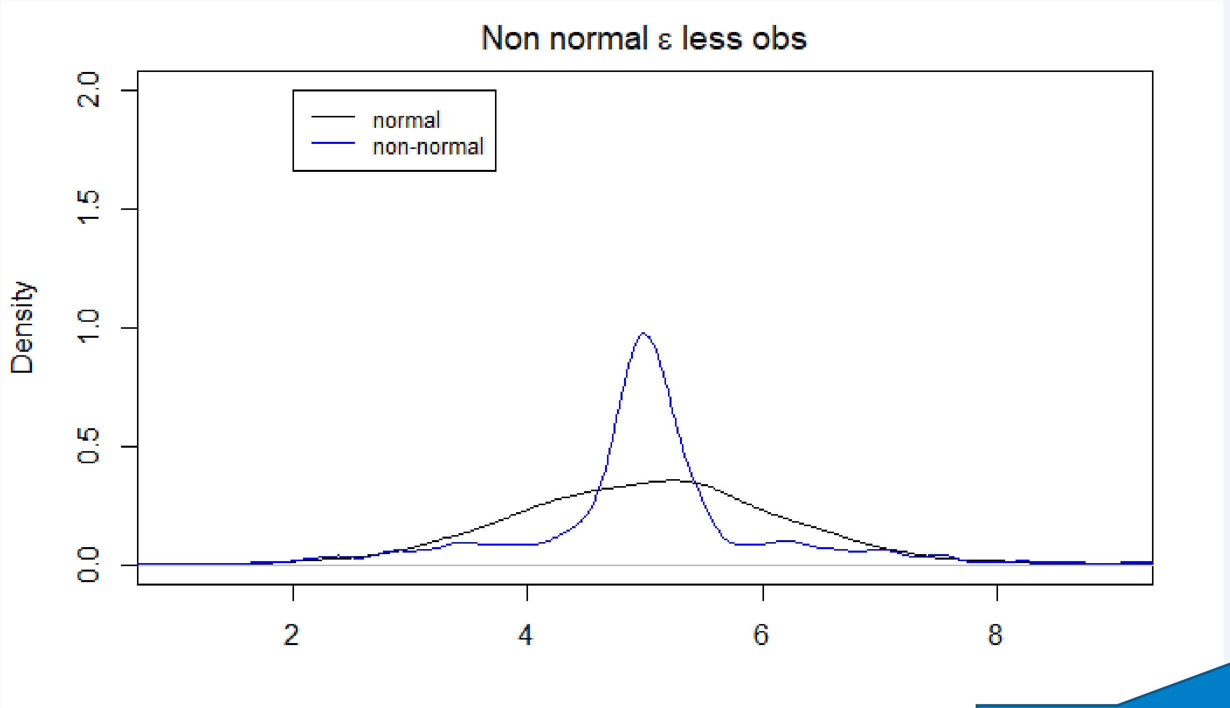


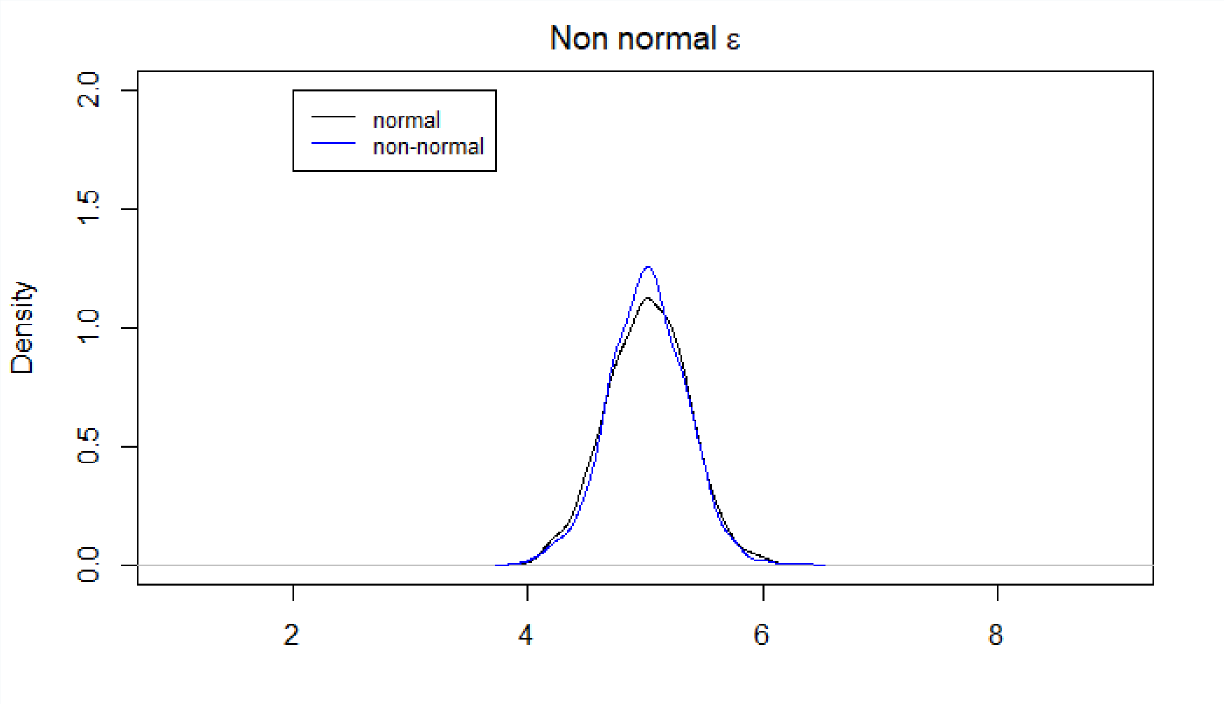
Non normal ϵ with same variance

Normal ϵ

Small sample (10 obs)

Large sample (100 obs)





Central limit theorem

The variance of the estimator

Standard Error of ϵ
We can estimate from $\hat{\epsilon}$

Standard Error of estimate

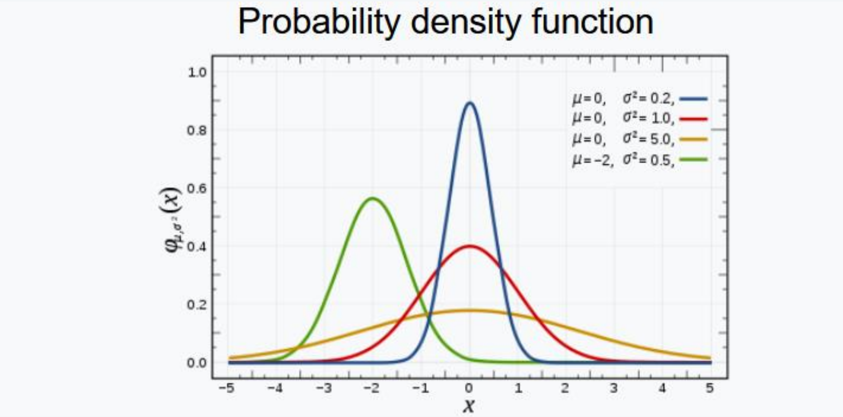
$$\sigma^2_{\widehat{\beta}_1} = \frac{\sigma^2_{\epsilon}}{nVAR(X)}$$

- Hence we also see in the formula that a larger number of observations means a lower variance of the estimated parameter.
- Moreover a larger variance of the of X (relative to the variance of ϵ) will imply a smaller variance of the estimate of β . Intuition: with bigger changes in X it will be easier to detect it's effect on Y.

Recap

- Regression estimates are (approximately) normally distributed
- We can work out the variance
- Normal distribution is fully characterized by standard error and mean

$$f(x) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2}$$



- To work out the likelihood of that a value of a particular value arises we can work out the area under the density

Significance level

- We can define how much risk of being wrong we are willing to accept and then work out a critical threshold

The foreigners cause crime hypothesis

P value: Probability that we have values more extreme than what we estimated

Small P means we can reject that the coefficient is 0 <u>(with little risk of being wrong)</u>

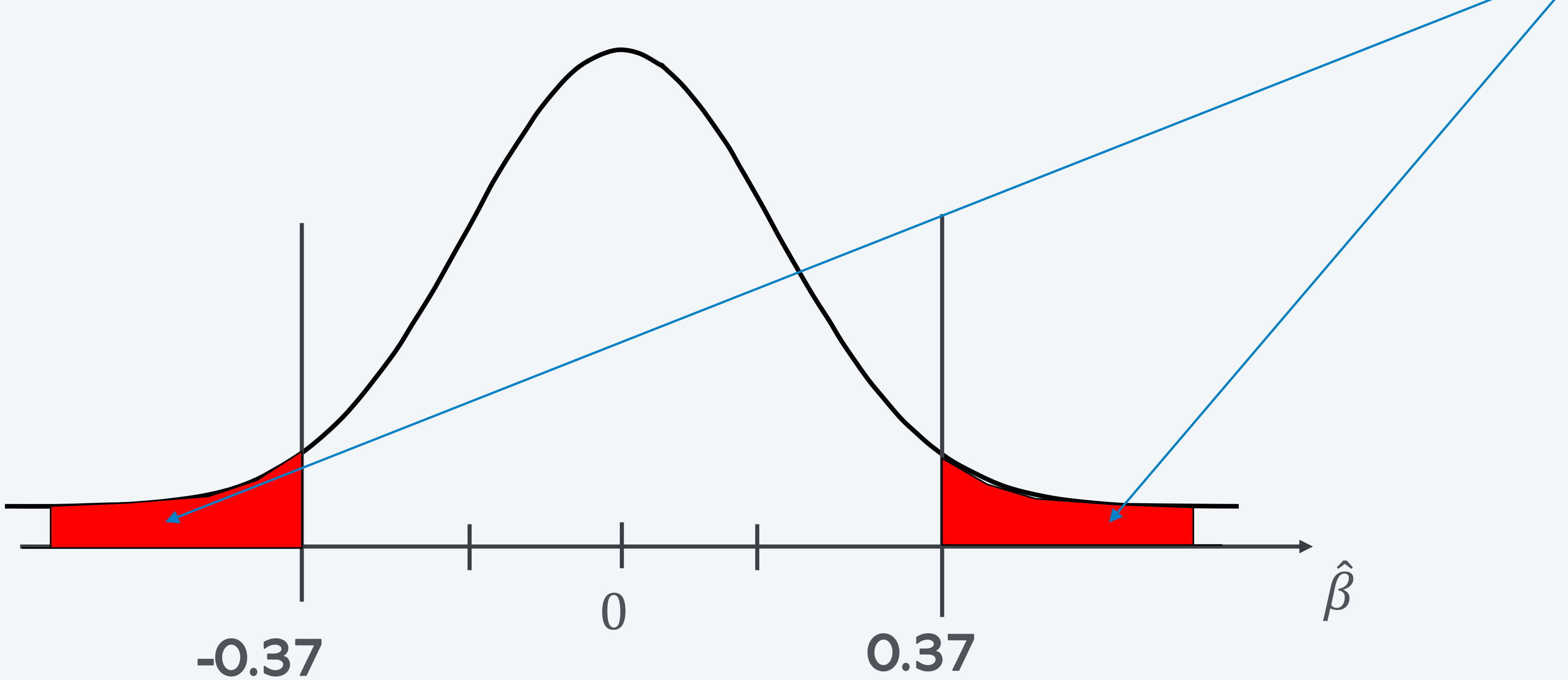
```
df=read_dta("../data/foreigners.dta")

 df['crimesPc']=df$crimes11/df$pop11
 reg1=lm(crimesPc~b_migr11,df)
 summary(reg1)

##
## Call:
## lm(formula = crimesPc ~ b_migr11, data = df)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -1.5886 -0.3789 -0.1038  0.2046 14.0988
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept) 0.992957   0.079387  12.508  < 2e-16 ***
## b_migr11    0.037630   0.005088   7.396 1.23e-12 ***
## ---
```

P-value

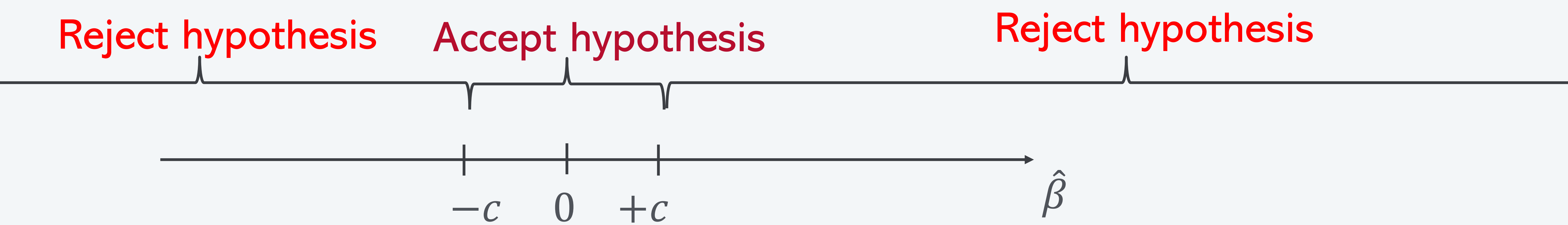
```
##
## Coefficients:
##             Estimate Std. Error t value Pr(>|t|)
## (Intercept) 0.992957   0.079387  12.508  < 2e-16 ***
## b_migr11    0.037630   0.005088   7.396 1.23e-12 ***
## ---
```



The P-value tells us how likely it is that we get an estimate that is smaller that is further away from 0 than the estimated value

Significance levels

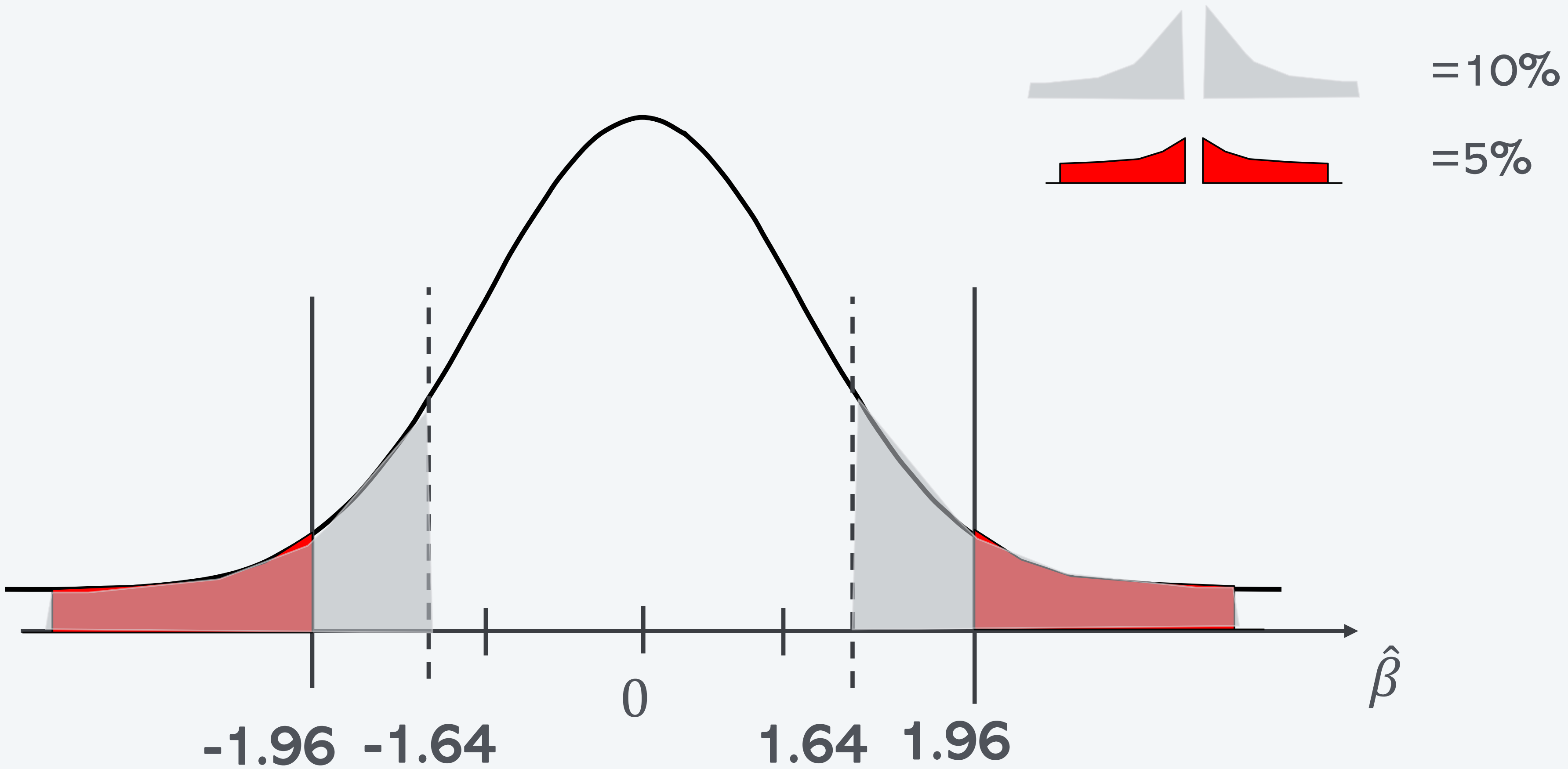
- Define how much risk of being wrong we are willing to accept and Work out a critical threshold value for $\hat{\beta}$ (call it c)
- If we find $\hat{\beta}$>c or $\hat{\beta} < -$c we know to reject that it is 0.



Null Hypothesis H0: $\beta = 0$

Alternative Hypothesis H1: $\beta \neq 0$

Finding c: Standard Normal ($\sigma = 1$)

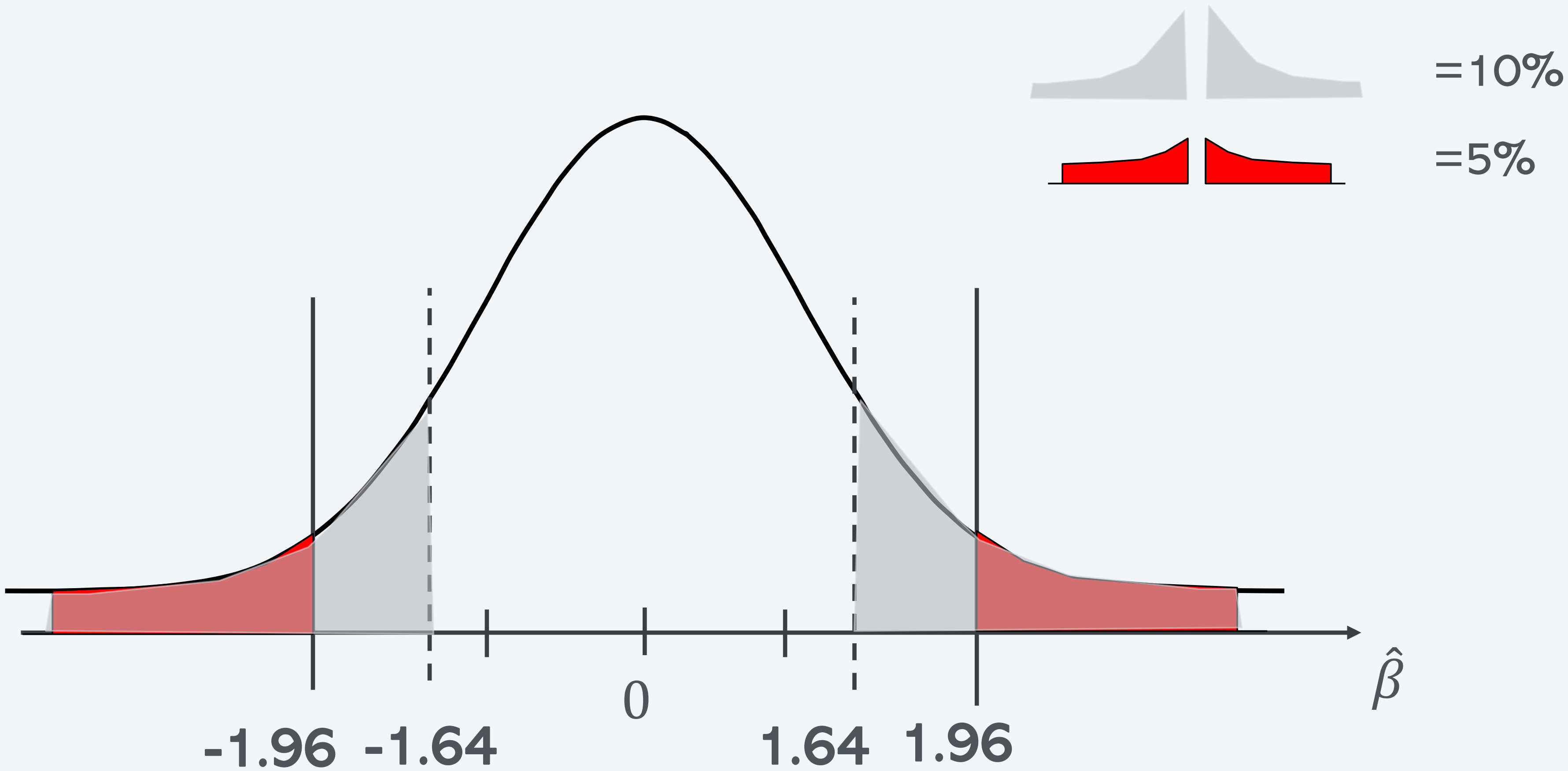



```r
pnorm(-1.644854)
## [1] 0.04999996
pnorm(-1.959964)
## [1] 0.025
```

- Say we willing to accept a higher risk
- Would we have a lower or higher threshold than c=1.96?
- E.g. what about 10% Type I risk?

We can use the cumulative (normal) distribution function (pnorm()) in R work these out

Finding c: Standard Normal ($\sigma = 1$)



- Say we willing to accept a higher risk
- Would we have a lower or higher threshold than c=1.96?
- E.g. what about 10% Type I risk?

Working out the threshold yourself

- The higher the significance level the smaller the threshold
- Higher significance level means we are less worried about an error of type I (reject even if true)
- Hence we are happy to reject in more cases

$\frac{1\%}{2}$

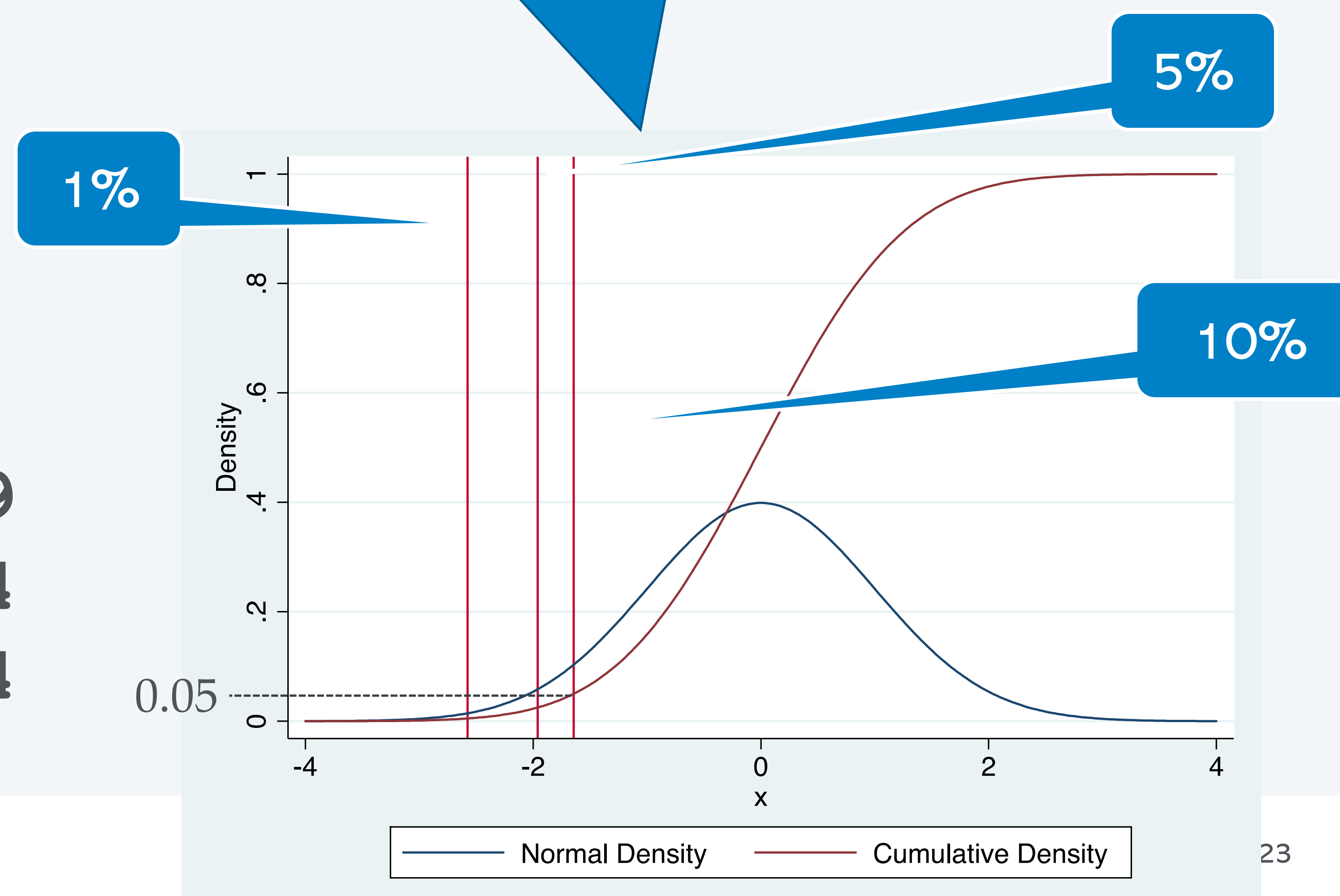
```
qnorm(0.005)   =   -2.575829
qnorm(0.005)   =   -2.575829
qnorm(0.025)   =   -1.959964
qnorm(0.05)    =   -1.644854
```

$\frac{10\%}{2}$

Inverse of the cumulative distribution function

To help with probability distributions check out this really cool probability distribution viewer (done in R)

```
qnorm(0.995) = 2.575829
qnorm(0.975) = 1.959964
qnorm(0.95)  = 1.644854
```



What if β is not standard normal (i.e. $\sigma_\beta \neq 1$)?

t statistic: ratio between estimate and standard error

$$t = \frac{\hat{\beta}}{\hat{\sigma}_{\hat{\beta}}} \sim N(0,1)$$

- Approximately standard normal
- **Therefore we can compare t statistic to the thresholds of the standard normal distribution**

The foreigners cause crime hypothesis

Standard error = 0.005

$$\frac{0.037}{0.005} = 7.4 > 1.96$$

```
df=read_dta("../data/foreigners.dta")

 df['crimesPc']=df$crimes11/df$pop11
 reg1=lm(crimesPc~b_migr11,df)
 summary(reg1)

##
## Call:
## lm(formula = crimesPc ~ b_migr11, data = df)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -1.5886 -0.3789 -0.1038  0.2046 14.0988
##
## Coefficients:
##             Estimate Std. Error t value Pr(>|t|)
## (Intercept) 0.992957   0.079387  12.508  < 2e-16 ***
## b_migr11    0.037630   0.005088   7.396 1.23e-12 ***
## ---
```

$$\sigma^2_{\hat{\beta}_1} = \frac{\sigma^2_\epsilon}{nVAR(X)}$$

t-value = Estimate/Standard Error

More or less significant estimates

- If we have a lower significance level (e.g. 1%) we are less likely to reject a hypothesis
- This is to avoid making the Type I error

- If we still reject the β=0 on the basis of an estimate $\hat{\beta}$ we say that **the estimate is highly significant**

- If we would only reject the hypothesis with a much higher significance level (e.g. 10% instead of 5%) we say that the estimate is only **weakly significant**

Another example

eaef <- read.csv("https://www.dropbox.com/s/9n0k7bs20z7qkv9/eaef21.csv?dl=1")

```
> mod_earn_exp <- lm(EARNINGS ~ EXP , data = eaef)
>   summary(mod_earn_exp)

Call:
lm(formula = EARNINGS ~ EXP, data = eaef)

Residuals:
    Min      1Q  Median      3Q     Max 
-17.140  -8.876  -3.723   3.869  99.986 

Coefficients:
            Estimate Std. Error t value Pr(>|t|)    
(Intercept)  15.5553     2.4425   6.369 4.09e-10 ***
EXP           0.2415     0.1398   1.727   0.0847 .  
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 14.39 on 538 degrees of
Multiple R-squared:  0.005515,   Adjusted R-sq
```

EXP= years of job experience

Your turn: What do you conclude from this regression? (multiple options can be correct)

(a) EXP coefficient is significantly different from 0 at 1%
(b) EXP coefficient is significantly different from 0 at 5%
(c) EXP coefficient is significantly different from 0 at 10%

Extra Slides

More general hypothesis tests

Previously we had H0: $\beta = 0$

$$t = \frac{\hat{\beta} - \beta}{\hat{\sigma}_{\hat{\beta}}}$$

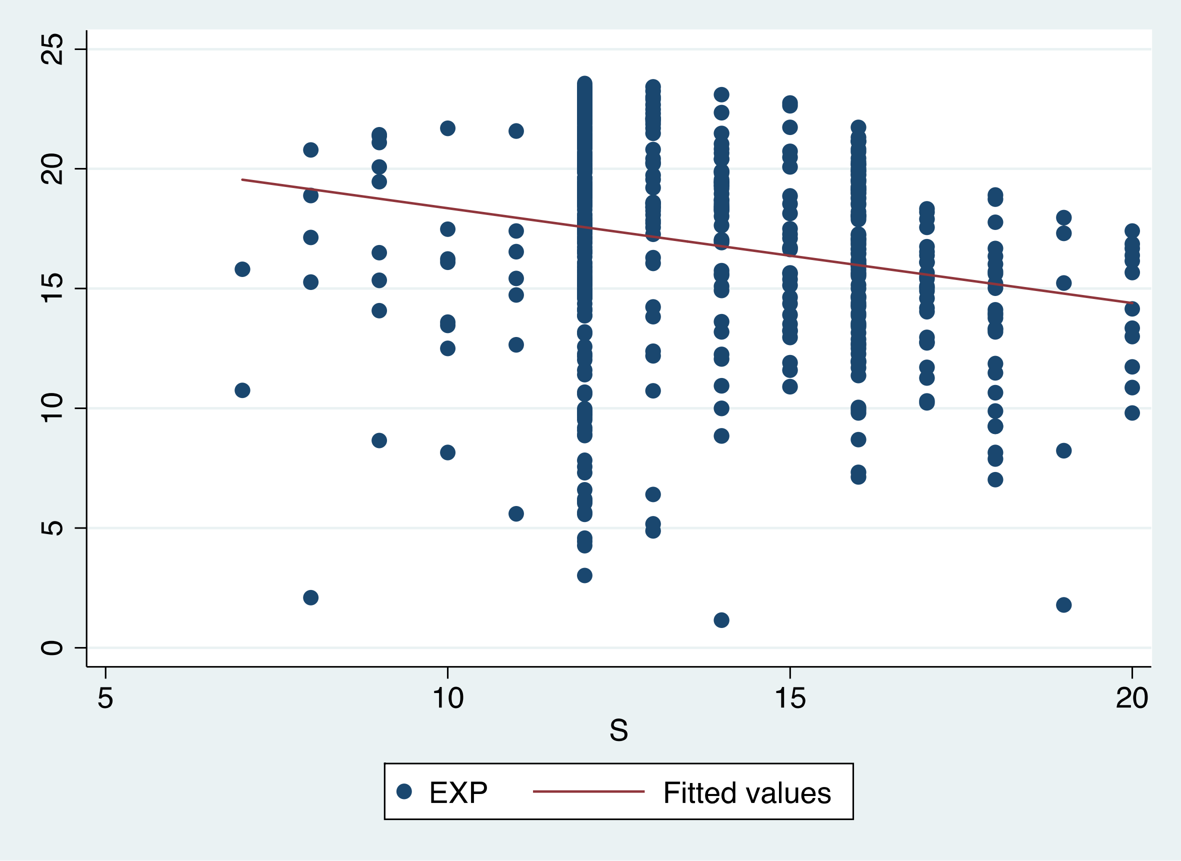
As before we can compare the t statistic with the critical values c for the standard normal distribution

Expected value of estimate under H0

More general tests example

Testing $\beta = 0$ is probably the most common test
However, many other could be of interest.

Consider Experience vs Schooling



Possible hypothesis: one year of schooling leads to one year less of experience

Can we reject this?

How to find out?

Experience vs schooling

```
mod_earn_exp <- lm(EXP ~ S , data = eaef)
```

```
Call:
lm(formula = EXP ~ S, data = eaef)

Residuals:
     Min       1Q   Median       3Q      Max 
-17.0512  -2.3320   0.8564   3.1391   6.3756 

Coefficients:
            Estimate Std. Error t value Pr(>|t|)    
(Intercept)  22.3165     1.0624  21.006  < 2e-16 ***
S            -0.3961     0.0765  -5.178 3.17e-07 ***
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1
```

Coefficient is negative but smaller than 1.
But is it small enough to reject that $\beta = -1$?

Experience vs schooling

```
mod_earn_exp <- lm(EXP ~ S , data = eaef)
```

```
Call:
lm(formula = EXP ~ S, data = eaef)

Residuals:
     Min       1Q   Median       3Q      Max
-17.0512  -2.3320   0.8564   3.1391   6.3756

Coefficients:
            Estimate Std. Error t value Pr(>|t|)
(Intercept)  22.3165     1.0624  21.006  < 2e-16 ***
S            -0.3961     0.0765  -5.178 3.17e-07 ***
---
Signif. cod                   001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

$t = \frac{-0.3961446-(-1)}{0.0765003}$ =7.894>1.96, hence we reject the hypothesis

Note: .

```
disp qt(0.025,538)
-1.9643832
```

Experience vs schooling

```
mod_earn_exp <- lm(EXP ~ S , data = eaef)
```

```
Call:
lm(formula = EXP ~ S, data = eaef)

Residuals:
     Min       1Q   Median       3Q      Max
-17.0512  -2.3320   0.8564   3.1391   6.3756

Coefficients:
            Estimate Std. Error t value Pr(>|t|)
(Intercept)  22.3165     1.0624  21.006  < 2e-16 ***
S            -0.3961     0.0765  -5.178 3.17e-07 ***
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1
```

```
Linear hypothesis test

Hypothesis:
S = - 1

Model 1: restricted model
Model 2: EXP ~ S

  Res.Df   RSS Df Sum of Sq      F    Pr(>F)
1    539 11260
2    538 10091  1    1168.7 62.307 1.658e-14 ***
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1
```

Alternative way to implement this test in R:

```
library("car")
linearHypothesis(mod_earn_exp, c( "S = -1") )
```

A note of caution

An estimate can be significant and biased
Or non-significant and non-biased (or vice versa)

- Significance is separate from bias

- We don't necessarily prefer an over another because one is significant.

- We need to ask for underlying reasons why one estimate is significant and the other one not.

Quick test: we have 2 estimates of the same parameter. Which would you prefer?

- Estimate 1 is biased and significant, estimate 2 is not significant but not biased?

Estimation of $\boldsymbol{\sigma_{\widehat{\beta}}}$

$$\sigma_{\widehat{\beta}}^2 = \frac{\sigma_\epsilon^2}{nVAR(X)}$$

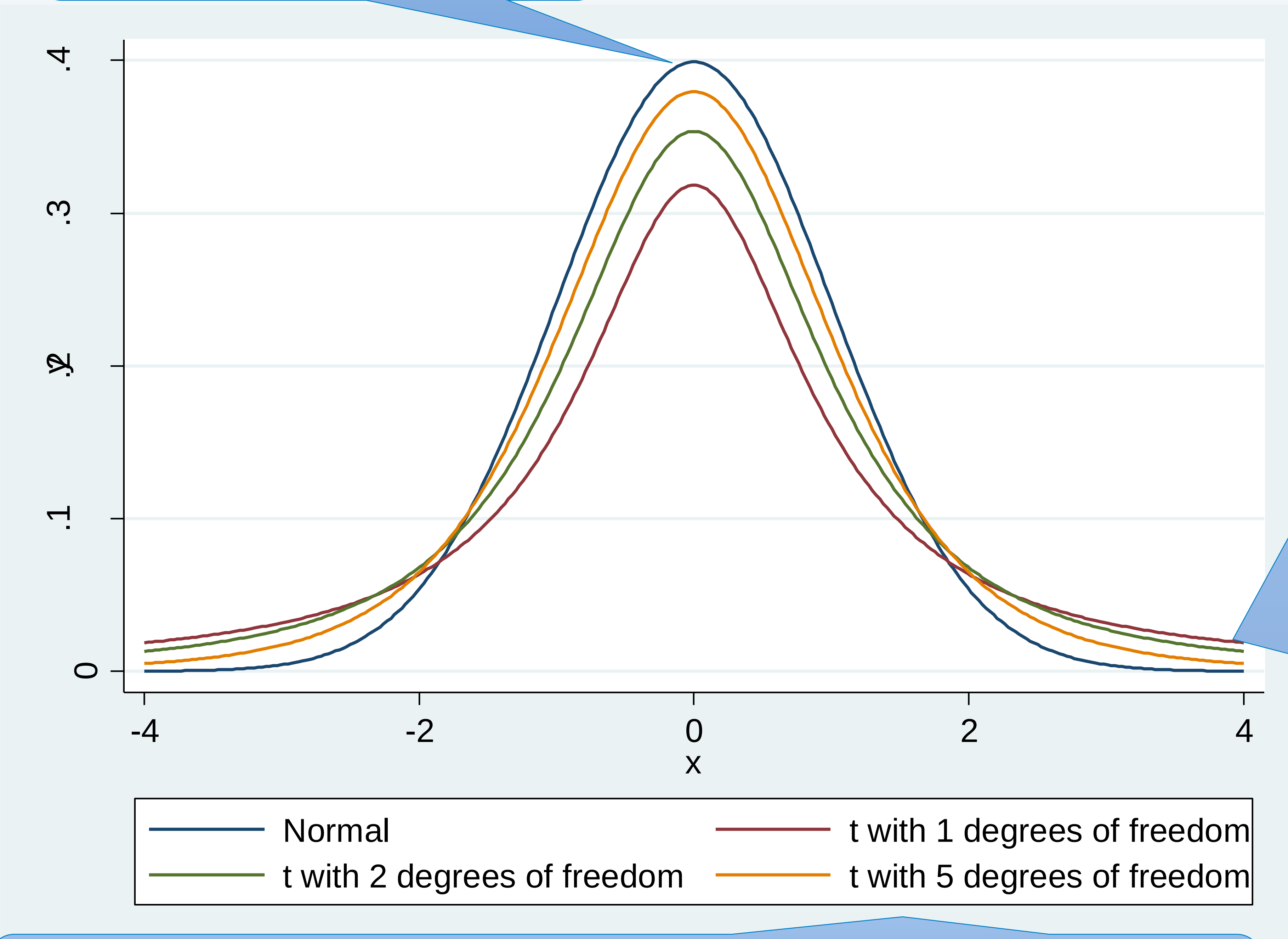
$$\hat{\sigma}_{\widehat{\beta}}^2 = \frac{\hat{\sigma}_\epsilon^2}{nVAR(X)}$$

Estimate using $VAR(\hat{\epsilon}^2)$

Student’s t-Distribution

Cheers

Standard normal

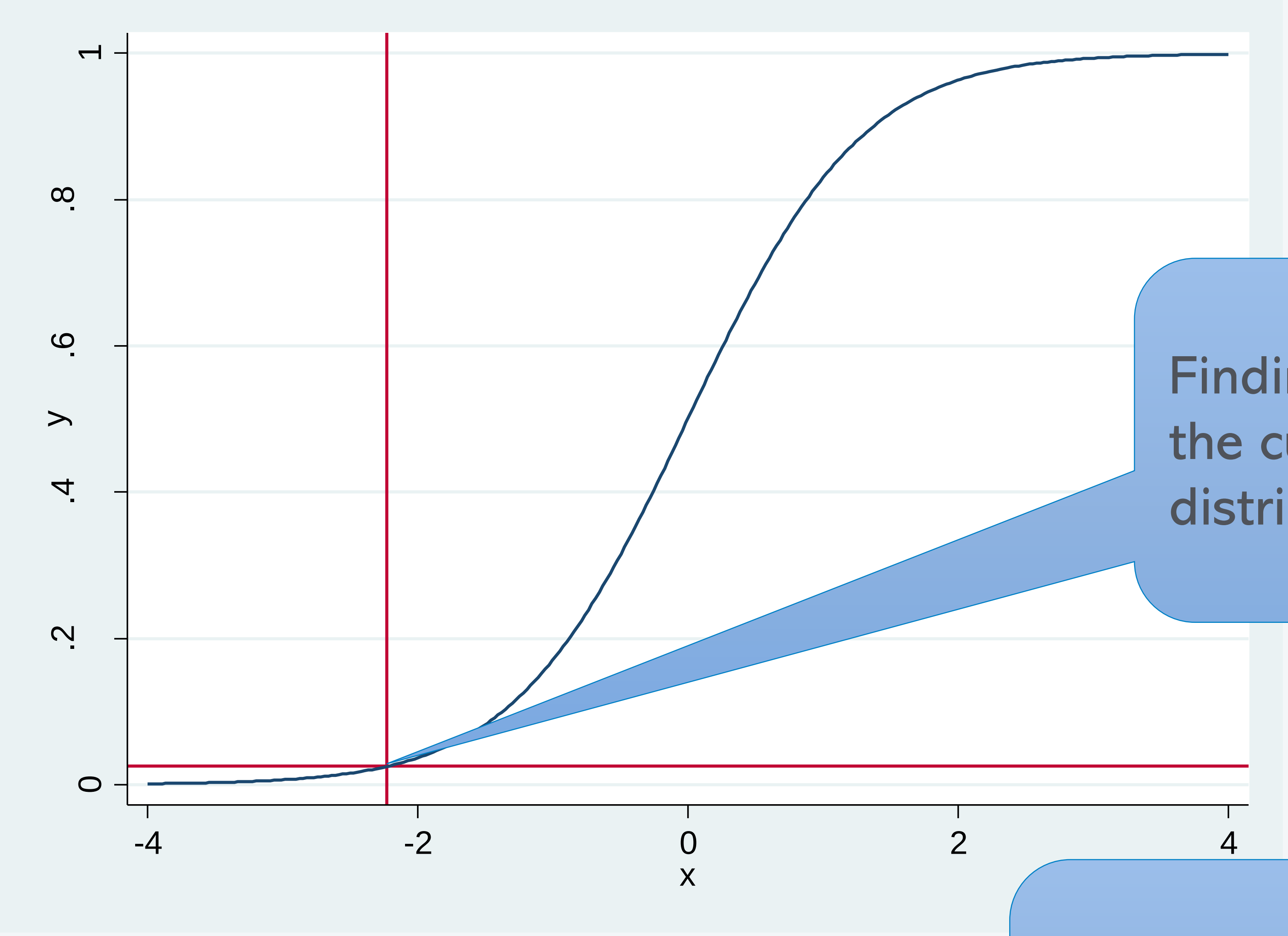


William Sealy Gosset
AKA Student

- t is a bit more dispersed than the normal
- Converges to Normal for large n
- We only need to worry about t for really small samples (<50)

Degrees of Freedom (DoF): observations – parameters we need to estimate before we can estimate ϵ

Critical values t distribution



Finding the value where the cumulative distribution is 0.025

```
qt(0.025, 10)
-2.228139
```

- 2.228>1.96
- i.e. to have the same level of risk of making error I we reject fewer values
- more probability weight in the tails