Previous class exercise ...

$$\max_{x} \; R = a^T x - x^T A x$$

$$a^T = [5 \; 4 \; 2], \quad A = \begin{bmatrix} 2 & 1 & 3 \\ 1 & 3 & 2 \\ 3 & 2 & 5 \end{bmatrix}$$

$$\frac{\partial R}{\partial x} = \frac{\partial a^T x}{\partial x} - \frac{\partial x^T A x}{\partial x} = 0$$

$$= a - 2Ax = 0$$

$$a = 2Ax$$

$$(2A)^{-1} a = (2A^{-1})(2A)x$$

$$(2A^{-1}) a = Ix$$

$$\begin{bmatrix} 4 & 2 & 6 \\ 2 & 6 & 4 \\ 6 & 4 & 10 \end{bmatrix}^{-1} \begin{bmatrix} 5 \\ 4 \\ 2 \end{bmatrix} = \begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix}$$

Computing an Inverse

1) Find adjoint matrix (transpose of cofactor matrix)

2) Find determinant

determinant of the 2x2 matrices

① cofactor of A

$$\begin{bmatrix} \begin{vmatrix} 6 & 4 \\ 4 & 10 \end{vmatrix} & -\begin{vmatrix} 2 & 4 \\ 6 & 10 \end{vmatrix} & \begin{vmatrix} 2 & 6 \\ 6 & 4 \end{vmatrix} \\ -\begin{vmatrix} 2 & 6 \\ 4 & 10 \end{vmatrix} & \begin{vmatrix} 4 & 6 \\ 6 & 10 \end{vmatrix} & -\begin{vmatrix} 4 & 2 \\ 6 & 4 \end{vmatrix} \\ \begin{vmatrix} 2 & 6 \\ 6 & 4 \end{vmatrix} & -\begin{vmatrix} 4 & 6 \\ 2 & 4 \end{vmatrix} & \begin{vmatrix} 4 & 2 \\ 2 & 6 \end{vmatrix} \end{bmatrix} = \begin{bmatrix} 44 & 4 & -28 \\ 4 & 4 & -4 \\ -28 & -4 & 20 \end{bmatrix}$$

cofactor

10

$$\operatorname{adj}(A) = \begin{bmatrix} 44 & 4 & -28 \\ 4 & 4 & -4 \\ -28 & -4 & 20 \end{bmatrix}$$

(2) Det (A) := find the weighted sum of the elements of any row or column using their cofactors as weights

If we choose the first row...

$$\det(A) = 4(44) + 2(4) + 6(-28) = 16$$

(3) $A^{-1} = \dfrac{\operatorname{adj}(A)}{\det(A)}$

$$A^{-1} = \begin{bmatrix} 2.75 & 0.25 & -1.75 \\ 0.25 & 0.25 & -0.25 \\ -1.75 & -0.25 & 1.25 \end{bmatrix}$$

$$\Rightarrow A^{-1} \begin{bmatrix} 5 \\ 4 \\ 2 \end{bmatrix} = \begin{bmatrix} 11.25 \\ 1.75 \\ -7.25 \end{bmatrix} = \begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix}$$

Finally, to prove this is a global max we check the Hessian

$$\frac{\partial R}{\partial x} = a - 2Ax$$

$$\frac{\partial^2 R}{\partial x^2} = \frac{\partial\, a - 2Ax}{\partial x} = -2A = \begin{bmatrix} -4 & -2 & -6 \\ -2 & -6 & -4 \\ -6 & -4 & -10 \end{bmatrix}$$

which is negative definite ✓

Just feel the pain of all this calculations. Good thing we have computers today. Let's move to R for some matrix calculus.

Goodness of Fit

In the context of Linear Regression, a "good" model can be characterized by how much of the dependent variable it explains. In other words, we want a high ratio between the sample variance of the fitted values and the true values.

$$R^2 = \frac{\widehat{Var}(\hat{y}_i)}{\widehat{Var}(y_i)} \quad \left[\begin{array}{c} \text{Coefficient of} \\ \text{determination} \end{array} \right]$$

What are the bounds of R^2? $R^2 \in [0,1]$
At best, all of the true variation is explained. In time series, it is very common to get high R^2 values. In cross-sections anything between 0.2 and 0.6 is considered pretty good. Of course, granted you have a solid theorethical foundation.

In terms of a regression $y = X\beta + \varepsilon$ we first estimate β with OLS to get

$$y = X\hat{\beta} + \varepsilon = \hat{y} + \varepsilon$$

Note that because of the FOC, the sum of the residuals is guaranteed to be zero

$$\frac{\partial\, \varepsilon^T \varepsilon}{\partial \beta} = (y - X\beta)^T (y - X\beta) = 0$$

$$X^T (y - X\beta) = 0$$

$$X^T \varepsilon = 0$$

If the first column of X is all ones, so that we have an intercept term, then

$$1^T \varepsilon = \sum \varepsilon_i = 0$$

The existance of an intercept term imposes this constraint.

The idea is that the intercept forces the regression line to "balance" around the mean. This also implies that $E[y] = E[\hat{y}] = E[X\beta]$ since $E[\varepsilon] = 0$ by the exogeneity assumption.

Let $\bar{y} = E[y] = E[\hat{y}] = \bar{x}^T \beta^{OLS}$

↳ Vector of column means

To get the variance, we want to compute deviations from the mean. Recall the idempotent matrix M^0 which computes deviations from the sample mean

$$M^0 = [I - \frac{1}{n} ii^T]$$

residuals have mean zero

$$M^0 y = M^0 X \beta_{OLS} + M^0 \varepsilon \quad , \quad M^0 \varepsilon = \varepsilon$$

then the total sum of squares is

$$y^T M^0 y = \beta_{OLS}^T X^T M^0 X \beta_{OLS} + \varepsilon^T \varepsilon$$

$$TSS = \underbrace{RSS}_{\text{regression}} + \underbrace{ESS}_{\text{error}}$$

Therefore, we can derive the R^2 coefficient of determination by dividing RSS over TSS

$$R^2 = \frac{RSS}{TSS} = \frac{\beta_{OLS}^T X^T M^0 X \beta_{OLS}}{\beta_{OLS}^T X^T M^0 X \beta_{OLS} + \varepsilon^T \varepsilon}$$

$$= \frac{RSS}{RSS + ESS} = \frac{RSS + ESS - ESS}{RSS + ESS}$$

$$= 1 - \frac{ESS}{RSS + ESS}$$

$$= 1 - \frac{\varepsilon^T \varepsilon}{y^T M^0 y} \in [0, 1]$$

A major problem with R^2 is that it will always increase if we add more regressors, even if these extra ones are uninformative.

The intuition is that additional variables increases the parameter space. In the larger model, the worst that can happen is that the new variables are irrelevant and hence have coefficient 0, which just gives us the original model. But if the new regressors explain even a bit of the dependent variable, then the fit improves.

You will get to prove this on the HW2.

To fix this, we define the adjusted R^2 which penalizes model complexity (i.e., larger models).

# of regressors

$$\bar{R}^2 = 1 - \frac{\varepsilon^T\varepsilon/(N-K)}{y^T M^0 y/(N-1)} = 1 - \frac{N-1}{N-K} \cdot \frac{\varepsilon^T\varepsilon}{y^T M^0 y}$$

# of observations

$$= 1 - \frac{N-1}{N-K}\left(1 - R^2\right)$$

Increases only if the contribution of the new regressor offsets the penalty.

Partitioned Regression

Often, we specify a multiple linear regression but we are really interested in the effect of one variable. For ex, recall our model of earn ~ educ

$earn_i = \beta_0 + \beta_1 educ_i + \beta_2 Age_i + \beta_3 Age_i^2 + \varepsilon_i$

How can we isolate the effect of educ on earnings? Or a subset of regressors more generally?

Suppose we have a regression with two sets of regressors $X_1 \in \mathbb{R}^{n \times m}$, $X_2 \in \mathbb{R}^{n \times q}$

$$y = X\beta + \varepsilon = X_1\beta_1 + X\beta_2 + \varepsilon$$

Say we are interested in finding β_2

Recall $\hat{\beta}^{OLS} = (X^T X)^{-1} X^T y$

The form $(X^T X)\beta = X^T y$ is called the normal equation. Splitting X into our two subsets yields

$$\begin{array}{c} (1) \\ (2) \end{array} \begin{bmatrix} X_1^T X_1 & X_1^T X_2 \\ X_2^T X_1 & X_2^T X_2 \end{bmatrix} \begin{bmatrix} \hat{\beta}_1 \\ \hat{\beta}_2 \end{bmatrix} = \begin{bmatrix} X_1^T y \\ X_2^T y \end{bmatrix}$$

To find $\hat{\beta}_2$ we can solve this system of equations. There are other methods but require more complex matrix algebra concepts ...

Start by finding an expression for $\hat{\beta}_1$

$$(X_1^T X_1)\beta_1 + (X_1^T X_2)\beta_2 = X_1^T y$$

$$\hat{\beta}_1 = (X_1^T X_1)^{-1} X_1^T (y - X_2 \beta_2) \qquad (A)$$

$\hat{\beta}_1$ is the set of coefficients in the regression $y \sim X_1$ minus a correction vector

Now insert (A) into the second normal equation

$$X_2^T X_1 \beta_1 + X_2^T X_2 \beta_2 = X_2 y$$

$$(X_2^T X_1)(X_1^T X_1)^{-1} X_1^T (y - X_2 \beta_2)$$

$$+ (X_2^T X_2)\beta_2 = X_2 y$$

Recall the projection matrix

$$P = X(X^T X)^{-1} X^T$$

$$X_2^T P_1 y - X_2^T P_1 X_2 \beta_2 + X_2^T X_2 \beta_2 = X_2 y$$

$$[X_2^T X_2 - X_2^T P_1 X_2]\beta_2 = [X_2^T - X_2^T P_1] y$$

$$[X_2^T (X_2 - P_1 X_2)]\beta_2 = X_2^T [I - P_1] y$$

$$[X_2^T (I - P_1) X_2]\beta_2 = X_2^T [I - P_1] y$$

Recall the annihilator matrix

$$M = I - P$$

$$[X_2^T M_1 X_2]\beta_2 = X_2^T M_1 y$$

$$\hat{\beta}_2 = (X_2^T M_1 X_2)^{-1} (X_2^T M_1) y$$

Since M_1 is the annihilator matrix for X_1, we know that $M_1 X_2$ is the vector of residuals in the regression of the corresponding column of X_2 on the variables/columns in X_1. We can simplify notation by defining the following vars.

$$X_2^* = M_1 X_2 \Rightarrow X_2^{*T} = (M_1 X_2)^T = X_2^T M_1^T$$

and $y^* = M_1 y$. So that

$$\hat{\beta}_2 = (X_2^{*T} X_2^*)^{-1} X_2^{*T} y^*$$

Recall that M is idempotent so $M^T = M$,

$M^2 = M$

Frisch-Waugh Theorem:

In the linear least squares regression of vector y on two sets of variables

X_1, and X_2, the subvector $\hat{\beta}_2$ is the set

of coefficients obtained when the residuals from a regression of y on X_1 $(y \sim X_1)$ alone are regressed on the set of residuals obtained when each column of X_2 is regressed on X_1.

Applied to our earnings and education example, we can get the effect (partial regression coefficient) of education by running earn$_i$ ~

Age + Age^2 and educ ~ Age + Age^2 and then using these residuals in a simple regression.

$$\text{earn}_i = \beta_0 + \beta_1 \text{educ}_i + \beta_2 \text{Age} + \beta_3 \text{Age}^2 + \varepsilon_i$$

$$y = X_1 \beta_1 + X_2 \beta_2 + \varepsilon$$

$$X_1 = [1, \text{educ}]^T, \quad X_2 = [\text{Age}, \text{Age}^2]^T$$

Define residual maker for the variables we want to "control" for.

$$M_2 = I - P_2$$

Use M_2 to get the residuals

$\tilde{y} = M_2 y$ (earnings "residualized" on age and age^2)

and

$$\widetilde{\text{educ}} = M_2 \,\text{educ}$$

Finally, run $\tilde{y} = \widetilde{educ} \cdot \beta_{educ} + \varepsilon$

β_{educ} gives the effect of education on earnings after controlling for Age and Age^2.

Stats Review II

Data: Observations from a sample

Statistic: Any function computed from the sample data

Parameter: Unknown variable to be estimated (e.g., coefficient in regression or moment of a distribution)

Estimator: A function or rule used to
estimate the parameter from the
data

Population distribution: The true
distribution
of the data generating process. Most
often we don't know this, so we
estimate its moments or assume a
form.

Sampling distribution: The observed or
sampled
distribution of your statistic or
estimator

There is only one true population
distribution, but each statistic
considered will have its own
sampling distribution. We want the

mean of this sample to be as close as possible to the true population mean, and its variance to be as small as possible.

There are two general ways to get an estimate:

- Point Estimate: The estimator returns
  a single estimated value for the parameter. OLS is a point estimator.

Regardless of the properties of an estimator, the estimate obtained will vary from sample to sample since it's a function of a random variable. Hence, there is a probability that the estimate is wrong by chance. A point estimate will not provide any information on the likely range of error. Capturing these "error bounds" is the purpose of interval estimation. Generally, it will look like

Point Estimate ± Measure of Sampling error

Ex: $\hat{\theta} \pm SD(\hat{\theta})$

- Interval Estimate: Use the sample data to construct an interval $[T_1, T_2]$ s.t. we capture the true parameter with some desired level of confidence (i.e., probability of success).

Pivotal Quantity Method

Suppose $X \sim f_\theta(x)$, where $\theta \in \Theta$ is a vector of parameters that completely specifies the probability density function $f_\theta(x)$. Say we have N observations and that we want to be 95% sure that θ lies inside our interval estimate.

Let's define two statistics (functions of the sample data) to capture the bounds of this interval

Lower bound : $T_1(X)$
Upper bound : $T_2(X)$

Two conditions must apply:

1) $T_1 \leq T_2 \quad \forall x \in X$

2) $\text{Prob}(T_1 \leq \theta \leq T_2) = 95\%$
confidence

Here is where the idea of a "confidence level" comes in. Usually denoted α. For a 95% confidence we want $1-\alpha = 0.95$ such that

$$\text{Prob}(T_1 \leq \theta \leq T_2) = 1-\alpha$$

If these two conditions hold, then we call the random interval (T_1, T_2) our "confidence interval" with confidence level equal to $(1-\alpha)$.

A Pivotal Quantity is a function of both the parameter and a point estimate that has known distribution. In practical terms this will be the distribution of your data and point estimator (e.g., OLS) chosen. We can denote it $g(X_1, \dots, X_n; \theta)$, so that X is the data and θ the parameter (or vector of ...) The problem we are trying to solve is

$$\text{Prob}(a \le g(X_1, \dots, X_n; \theta) \le b) = 1-\alpha, \quad \forall \theta$$

find a and b given α. Then solve for θ $\text{Prob}(T_1(x) \le \theta \le T_1(x)) = 1-\alpha, \ \forall \theta$

- $g(x;\theta)$ must be continuous and invertible.

OLS application (univariate)

Let $y_i = \beta_0 + \beta_1 x_i + \varepsilon_i$ be our regression so the OLS estimates are

$$\hat{\beta}_0 = \bar{y} - \hat{\beta}_1 \bar{x}, \quad \hat{\beta}_1 = \frac{\hat{\text{Cov}}(x,y)}{\text{Var}(x)}$$

$\hat{\beta}_0$ and $\hat{\beta}_1$ are our point estimates. Let's construct the confidence interval, s.t.

$$\text{Prob}\left(T_1(X) \leq \beta_1 \leq T_2(X)\right) = 1-\alpha$$

Three concrete criteria must be met for

1) $g(X;\beta_1)$ is a function of the sample data and the parameter(s) we want to estimate (but not other unknown parameter).

2) It's distribution depends only on the data (not on any of the parameters, Known or unknown.)

3) Is continuous and invertible wrt θ

To continue our example we could pick

$$t = \frac{\hat{\beta}_1 - \beta_1}{\sqrt{\frac{\hat{\sigma}^2}{\sum (x_i - \bar{x})^2}}} = \frac{\hat{\beta}_1 - \beta_1}{SE(\hat{\beta}_1)}$$

Recall $Var(\hat{\beta}_1) = Var\left(\frac{\sum (x_i - \bar{x})(y_i - \bar{y})^{*}}{\sum (x_i - \bar{x})^2} \right)$

$$= Var\left(\frac{\sum (x_i - \bar{x})^{*} \cdot y_i}{\sum (x_i - \bar{x})^2} \right);$$ * Prove in HW

$$= Var\left(\frac{\sum (x_i - \bar{x})(\beta_0 + \beta_1 x_i + \varepsilon_i)}{\sum (x_i - \bar{x})^2} \right)$$

constant [only depends on data]

$$= \frac{1}{(\sum (x_i - \bar{x})^2)^2} Var\left(\sum (x_i - \bar{x})(\beta_0 + \beta_1 x_i + \varepsilon_i) \right)$$

$$= \frac{1}{(\sum (x_i - \bar{x})^2)^2} Var\left(\sum (\beta_0 + \beta_1 x_i)(x_i - \bar{x}) + \sum \varepsilon_i (x_i - \bar{x}) \right)$$

constant β_0, β_1 are fixed true parameters, and x_i is the data. So there's no randomness hence $Var(\cdot) = 0$

$$= \frac{1}{\left(\sum (x_i - \bar{x})^2\right)^2} Var\left(\sum \varepsilon_i (x_i - \bar{x})\right)$$

constant

$$= \frac{1}{\left(\sum (x_i - \bar{x})^2\right)^2} \sum (x_i - \bar{x})^2 \, Var(\varepsilon_i)$$

homoskedasticity
$Var(\varepsilon_i) = \sigma^2$

we often still have to estimate it with the sample variance

$$= \frac{\sigma^2 \cancel{\sum (x_i - \bar{x})^2}}{\left(\sum (x_i - \bar{x})^2\right)^{\cancel{2}}} = \frac{\hat{\sigma}^2}{\sum (x_i - \bar{x})^2} = Var(\hat{\beta}_1)$$

$$\Rightarrow SE(\hat{\beta}_1) = \sqrt{Var(\hat{\beta}_1)}$$

So, t depends only on the data and the parameter β_1. Condition 1 ✓

It is T-distributed, which is a distribution that depends only on the degrees of freedom $(N-K)$. Since we know both ($K=2$ in this example), we know the distribution. Condition 2 ✓

$$\frac{\hat{\beta}_1 - \beta_1}{SE(\hat{\beta}_1)} \sim T_{N-2} \left[\frac{N(0, Var(\hat{\beta}_1 - \beta_1))}{\chi^2(n)} \right. =$$

T_{N-K}

Finally, t is continuous and invertible wrt β_1. Condition 3 ✓

So t is a pivotal quantity

We must solve

$$\text{Prob}\left(a \le t_{N-2} \le b\right) = 1-\alpha$$

Let t_{N-2} be some generic T_{N-2} distributed variable.

We use the properties of the T-distribution to get a and b. If we want the interval to be symmetric we add the condition $a+b=0$.

$$\Rightarrow a = t_{N-2;\frac{\alpha}{2}} \quad , \; b = t_{N-2;1-\frac{\alpha}{2}}$$

where $\frac{\alpha}{2}$ and $1-\frac{\alpha}{2}$ are the probabilities of a number being to the left of a and b, respectively That is, $P\left(t_{N-2} < t_{N-2;\frac{\alpha}{2}}\right) = \frac{\alpha}{2}$ and

$P\left(t_{N-2} < t_{N-2;1-\frac{\alpha}{2}}\right) = 1-\frac{\alpha}{2}$

Hence, $1-\frac{\alpha}{2}+\frac{\alpha}{2} = 1-\alpha$, so

$$P\left(t_{N-2;\frac{\alpha}{2}} \le \frac{\hat{\beta}_1-\beta_1}{SE(\hat{\beta}_1)} \le t_{N-2;1-\frac{\alpha}{2}}\right) = 1-\alpha$$

$$\Rightarrow \left(\hat{\beta}_1 - t_{N-2;\frac{\alpha}{2}} \cdot SE(\hat{\beta}_1) \le \beta_1 \le \hat{\beta}_1 - t_{N-2;1-\frac{\alpha}{2}} \cdot SE(\hat{\beta}_1)\right)$$

By symmetry $a = -b$

$$P\left(\hat{\beta}_1 - b \cdot SE(\hat{\beta}_1) \le \beta_1 \le \hat{\beta}_1 + b \cdot SE(\hat{\beta}_1)\right) = 1-\alpha$$

$\left(\hat{\beta}_1 - b \cdot SE(\hat{\beta}_1), \hat{\beta}_1 + b \cdot SE(\hat{\beta}_1)\right)$ is the interval estimate with confidence level α for OLS slope $\hat{\beta}_1$.

This means that, in repeated sampling, the true value β_1 will be contained in this interval 95% of the time

Hypothesis Testing

The idea is to use our estimated

statistic, confidence interval, and sample data to determine if our parameter falls into one of (usually) two subsets of the parameter space.

The "classical" or Neyman-Pearson approach consists of defining two regions: Rejection and Acceptance.

Let Θ be the parameter space (i.e., the set

of all possible values of our paremeter). Note if we have a vector of parameters this would be multi-dimensional.

Partition Θ into two mutually exclusive sets: Θ_0 and Θ_1, $\Theta_0 \cap \Theta = \emptyset$

Here, Θ_0 and Θ_1 are our hypotheses!

$$\begin{cases} H_0 : \theta \in \Theta_0 \text{ (null) } H_1 : \\ \theta \in \Theta_1 \text{ (alternative)} \end{cases}$$

A statistical hypothesis is therefore a claim on the value of parameter θ.

A test is a procedure or set of rules to decide wether to accept or reject H_0 given our sample data.

Since our sample is random, the test is random as well. Hence we will find errors with positive probability. There are two types we will encounter in this class:

Type I : Reject H_0 but is actually true

-> Aka false positive

Type II : Accept H_0 but is false
-> Aka false negative.

There are two main properties of a test we care about:

Size: Significance level, α, value.
It gives the Prob (type I) = α
We get to choose this one.

Power: Probability of correctly rejecting H_0. power = 1 - Prob (type II) = 1 - λ

There is a natural trade-off here.

We can make α arbitrarily small so that we guarantee no Type I errors. But this means the chances of Type II error goes up.

Ideally, we want both α and λ to be as small as possible. So, given a value of α we want to choose, from all the tests that satisfy this condition, the one with the greatest power

Neyman - Pearson (Regression example)

Suppose we have a regression

$$y_i = \beta_0 + \beta_1 x_i + \varepsilon_i$$

and we want to verify that x_i is

actually relevant for explaining y_i.

$$\left\{\begin{array}{l} H_0 : \beta_1 = 0 \\ H_1 : \beta_1 \neq 0 \end{array}\right.$$

"2-sided t-test"

Now we use the pivotal quantity method to construct a rejection region, $\Theta_0 = \{0\}$. The idea is to fix a threshold, say h, such that if our OLS estimate $\hat{\beta}_1$ is less or greater than h we reject H_0 with α confidence.

$|\hat{\beta}_1| > h = \hat{\beta}_1 < -h \lor \hat{\beta}_1 > h$ "or"

Recall our t-statistic $t_{N-2} = \dfrac{\hat{\beta}_1 - \beta_1}{SE(\hat{\beta}_1)} \sim T_{N-2}$

$$\Rightarrow \frac{\hat{\beta}_1 - \beta_1}{SE(\hat{\beta}_1)} < \frac{-h - \beta_1}{SE(\hat{\beta}_1)} \vee \frac{\hat{\beta}_1 - \beta_1}{SE(\hat{\beta}_1)} > \frac{h - \beta_1}{SE(\hat{\beta}_1)}$$

Under the null $\beta_1 = 0$

$$\underset{\beta_1 = 0}{\text{Prob}} \left(\frac{\hat{\beta}_1}{SE(\hat{\beta}_1)} < \frac{-h}{SE(\hat{\beta}_1)} \vee \frac{\hat{\beta}_1}{SE(\hat{\beta}_1)} > \frac{h}{SE(\hat{\beta}_1)} \right) = \alpha$$

Since the T-distribution only depends on the number of parameters and sample size, we know everything about it. There's nothing else to estimate. Hence we can easily solve for h.

Let t_{N-2} be some generic T-distributed var.

$$\text{Prob} \left[t_{N-2} < \frac{-h}{SE(\hat{\beta}_1)} \vee \right.$$

$$t_{N-2} > \frac{h}{SE(\hat{\beta}_1)} \Big] = \alpha$$

By symmetry $t_{N-2;\frac{\alpha}{2}} = -t_{N-2;1-\frac{\alpha}{2}} = -\frac{h}{SE(\hat{\beta}_1)}$

$$\Rightarrow h = t_{N-2;1-\frac{\alpha}{2}} \cdot SE(\hat{\beta}_1) = -t_{N-2;\frac{\alpha}{2}} \cdot SE(\hat{\beta}_1)$$

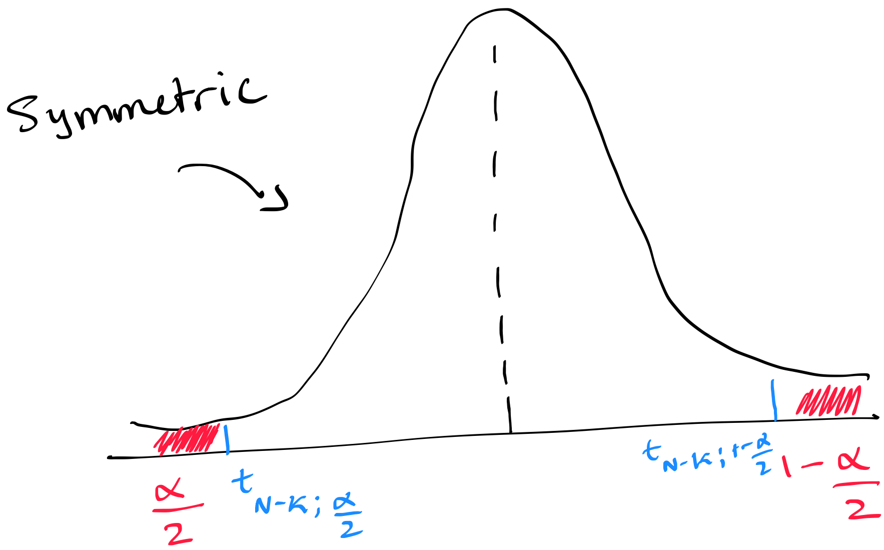
So, after estimating our regression we reject $H_0: \beta_1 = 0$ if $|\hat{\beta}_1| > t_{N-2;1-\frac{\alpha}{2}} \cdot SE(\hat{\beta}_1)$

Suppose our estimate passes this test. That is, $\hat{\beta}_1 \neq 0$ with α confidence. Now, we want to know if its value is greater than some number γ

$$\begin{cases} H_0 : \beta_1 > \gamma \quad \text{"1-sided} \\ \end{cases}$$

t-test" $H_1 : \beta_1 \leq \gamma$

T-distribution



If we want to test hypothesis about the parameter beign in a given range we look at both tails. If we fix one of the bounds of the range then we look at only one of them

So the probability of Type I error is

$$\text{Prob}_{\beta_1 > \gamma}\left(\hat{\beta}_1 \leq \gamma\right) = \alpha$$

Using the t-statistic

$$\text{Prob}_{\beta_1 > \gamma}\left(\frac{\hat{\beta}_1 - \beta_1}{SE(\hat{\beta}_1)} \leq \frac{h - \beta_1}{SE(\hat{\beta}_1)}\right) = \alpha$$

Under the null $H_0: \beta_1 > \gamma$ we can replace β_1 with γ

$$\text{Prob}\left(\frac{\hat{\beta}_1 - \gamma}{SE(\hat{\beta}_1)} \leq \frac{h - \gamma}{SE(\hat{\beta}_1)}\right) = \alpha$$

Again, since $\dfrac{\hat{\beta}_1 - \gamma}{SE(\hat{\beta}_1)} \sim T_{N-2}$, we have

$$\text{Prob}\left(t_{N-2} \leq \frac{h - \gamma}{SE(\hat{\beta}_1)} \right) = \alpha$$

From where we can solve for h

$$h = \gamma + t_{N-2;\,\alpha} \cdot SE(\hat{\beta}_1)$$

$\hookrightarrow$ Not $\frac{\alpha}{2}$ because it is 1-side test

Hence we reject H_0 with α confidence if $|\hat{\beta}_1| > \gamma + t_{N-2;\,\alpha} \cdot SE(\hat{\beta}_1)$

Ex: Apply these steps to show that for the test

$$\begin{cases} H_0 : \beta_1 < \gamma \\ H_1 : \beta_1 \geq \gamma \end{cases}$$

the rejection condition is

$$|\hat{\beta}_1| > \gamma + t_{N-2;\,1-\alpha} \cdot SE(\hat{\beta}_1)$$

Q: What distribution does T approach as $N \to \infty$? $\mathcal{N}(0,1)$

If so, then we use the z-statistic or z-scores for our tests.

$$t_{N-2;\,\alpha} \longrightarrow z_\alpha = \Phi^{-1}(\alpha)$$