Statistics Overview

Edicson Luna

University of Pennsylvania

Introduction

Basic Probability Theory

Random sampling

Hypothesis testing

Introduction

Variable

A variable is a characteristic, feature, or quantity that can take different values across individuals, objects, places, or time.

Examples:

- Age of a person.
- Income of a household.
- Education level of an individual.
- Number of children in a family.
- Whether a person is employed or unemployed.

Population vs Sample

An important distinction:

- Parameter: number describing the population (e.g. population mean).
- Statistic: number describing the sample (e.g. sample mean or average).

Usually, it is impossible to know the "parameter" with certainty. That is why we use statistics to deduce information about the whole population (although with errors).

Measures of central tendency

Consider a sample $x = (x_i)_{i=1}^N$, we define

Sample mean: $\bar{x} := \frac{\sum_{i=1}^N x_i}{N}$
This is probably the most used statistic. As a drawback, the sample mean is very sensitive to outliers.

Sample median: start by ranking the observation from the min to max $(x_1, x_2, ..., x_N)$. Then, if N is odd, take $med(x) := x_{\frac{N+1}{2}}$. Otherwise, define $med(x) := \frac{x_{\frac{N}{2}} + x_{\frac{N}{2}+1}}{2}$.

Measures of variability

Sample variance: $s^2 := \frac{1}{N-1} \sum_{i=1}^{N} (x_i - \bar{x})^2$
The $N - 1$ is done to obtain an "unbiased" estimator of the variance (we will talk about it later if we have time). Nonetheless, we will see that this is not as important as $N \to \infty$.

Sample standard deviation: $s = \sqrt{s^2}$
This statistic has the same units as the data (which is an advantage for data analysis).

Working with more than one variable

Sometimes we may have a sample of two variables, $(x, y) = (x_i, y_i)_{i=1}^N$, and we may want to analyze the relationship between the two. Define

Sample covariance: $s_{xy} := \frac{1}{N-1} \sum_{i=1}^N (x_i - \bar{x})(y_i - \bar{y})$
The sign of the covariance tells us about the direction of the dependence between x and y, but the value is difficult to analyze.

Sample correlation: $r_{x,y} = \frac{s_{xy}}{s_x s_y}$ (s_j refers to the standard deviation of j as defined above)
$r_{x,y} \in [-1, 1]$ which allows us to interpret the value.

Back to parameters

In the previous slides, I explicitly used "sample" to indicate that I was talking about statistics. Nonetheless, I will stop mentioning it since it should be obvious when we are working with sample values.

The analogous population parameters are:

- Mean: $\mu_x = E[x_i]$
- Median: $\text{Med}_x = \text{Q}_{0.5}[x_i]$
- Var: $\sigma^2 = E[(x_i - \mu)^2]$
- SD: $\sigma = \sqrt{\sigma^2}$
- Cov: $\sigma_{xy} = E[(x_i - \mu_x)(y_i - \mu_y)]$
- Corr: $\rho_{xy} = \frac{\sigma_{xy}}{\sigma_x \sigma_y}$

Back to parameters

Analyzing the sample (and keeping it like that) is perfectly fine. Nevertheless, we are sometimes (always in this class) interested in parameters.

Fortunately, we have some tools that allow us to go from statistics to parameters. But first, we need to go over some probability concepts.

Basic Probability Theory

Probability

Event (E): a set of basic outcomes in the sample space Ω (all possible outcomes that can result from an experiment).

e.g. when tossing a coin, the sample space is H (heads) and T (tails), and an event could be "getting heads". In this case, we may represent it as $\Omega = \{H, T\}$ and $E = \{H\}$.

The concept of **probability** is defined by events. What is the probability of getting heads when tossing a coin?

Probability Distribution

A **probability distribution** describes the distribution of probabilities across the possible outcomes of a random variable.

If we have a finite set of possible events that are equally likely, $\mathbb{P}$ is defined as

$$\mathbb{P}(E) := \frac{|E|}{N}$$

where $|E|$ represents the cardinality of E and N the total number of possible outcomes.

e.g. when flipping a coin, getting $\mathbb{P}(H) = \mathbb{P}(T) = 0.5$. When rolling a dice $\mathbb{P}(A) = \frac{1}{6}\ \forall A = 1, ..., 6$. For some events, the probability is not that simple.

Basic properties

- $\mathbb{P}(\Omega) = 1$
- $0 \leq \mathbb{P}(E) \leq 1 \; \forall E$
- $\mathbb{P}(E_1 \cup E_2) = \mathbb{P}(E_1) + \mathbb{P}(E_2)$ when E_1 and E_2 are two mutually exclusive events ($E_1 \cap E_2 = \emptyset$).

There are some more complicated characteristics required for the probability to be defined, but we will ignore them for this class.

Conditional probability

Conditional probability: Sometimes we want to know the probability of an event B considering that another event A has happened (having A a positive probability).

$$\mathbb{P}(B|A) = \frac{\mathbb{P}(A \cap B)}{\mathbb{P}(A)}$$

This immediately implies the **Bayes Rule**:

$$\mathbb{P}(A|B) = \mathbb{P}(A) \times \frac{\mathbb{P}(B|A)}{\mathbb{P}(B)}$$

Independence

Two events are **independent** if $\mathbb{P}(A \cap B) = \mathbb{P}(A)\mathbb{P}(B)$.
Notice that this is the same as saying $\mathbb{P}(A|B) = \mathbb{P}(A)$.

Consider an event A such that a set of mutually exclusive events (E_i for $i = 1, ..., n$) satisfy $A = E_1 \cup ... \cup E_n$, the **Law of Total Probability** is defined as

$$\mathbb{P}(A) = \mathbb{P}(A|E_1)\mathbb{P}(E_1) + \mathbb{P}(A|E_2)\mathbb{P}(E_2) + ... + \mathbb{P}(A|E_n)\mathbb{P}(E_n)$$

Random variable

A **random variable** (RV), usually written X, is a function defined from the set of outcomes of an experiment (Ω) to numerical values. In reality, the definition is from Ω to a measurable space ($\mathcal{F}$), but we will encompass them in $\mathbb{R}$ (which is a measurable space and is enough for this class).

e.g. in the coin example, a random variable would be given by $X : \{H, T\} \to \mathbb{R}$, where $X(T) = 0$ and $X(H) = 1$.

There are two types of random variables: discrete (e.g. flipping a coin) and continuous (e.g. age of a person).

Random variable

We use uppercase letters, such as X, to denote random variables, and lowercase letters, such as x, to denote observed values or realizations.

The **support** of a RV X is the set of values that X can take with positive probability, or that are arbitrarily close to values with positive probability. Formally, a value x_0 belongs to the

support of X if, for every $\varepsilon > 0$,

$$\Pr\left(X \in (x_0 - \varepsilon, x_0 + \varepsilon)\right) > 0.$$

That is, every small interval around x_0 has positive probability.

Discrete RV

As its name says, it is defined in discrete points in $\mathbb{R}$.

The **probability mass function** (pmf) is a function mapping from $\mathbb{R}$ to $[0, 1]$ that only takes positive values if the events are in the support of X.

e.g. when rolling a dice $\mathbb{P}(A) = \frac{1}{6}$ for $A = 1, ..., 6$, and $\mathbb{P}(A) = 0$ otherwise.

They follow the Bernoulli distribution (or its generalization).

Continuous RV

Defined in continuous points in $\mathbb{R}$. Individual points have zero probability.

We need intervals to be able to talk about positive probabilities.

In Economics, continuous RVs are usually normally or uniformly distributed.

The normal distribution is the most important for this class since it is the one resulting when we apply the Central Limit Theorem (we will cover this later).

Continuous RV

The **probability density function** (pdf) is any function $f : \mathbb{R} \to \mathbb{R}_+$ satisfying

$$\int_a^b f(x)dx := \mathbb{P}(a \leq X \leq b)$$

Even if the probability of an individual point is zero, the pdf of a point in the support is positive. Even more, it can be greater than 1.

CDF

The **cumulative density function** (cdf) - usually represented as F- is defined as the integration from $-\infty$ to a point x of the pdf (sum in the case of discrete variables). That is,

$$F(x) = \mathbb{P}(-\infty < X \leq x) := \int_{-\infty}^{x} f(z)dz$$

It is easy to observe that the pdf is the first derivative of the cdf (in the continuous world).

CDF

Notice that the CDF is a probability, then it has the same properties previously discussed. In addition to that, it is

- Non-decreasing: $x_0 < x_1 \rightarrow F(x_0) \leq F(x_1)$
- Zero at negative infinity: $\lim_{x_0 \rightarrow -\infty} F(x_0) = 0$
- One at positive infinity: $\lim_{x_0 \rightarrow \infty} F(x_0) = 1$

From now on, I will focus on definitions for continuous variables.

Expectation

The **expectation** of the RV X, $\mathbb{E}[X]$ or μ, with pdf f is defined by

$$\mu = \mathbb{E}[X] := \int_{-\infty}^{\infty} x f(x) dx$$

It is also known as the (population) mean or the first moment of the distribution of X.

This definition is also extended to functions defined over X. Take the function $h : \mathbb{R} \to \mathbb{R}$, then $h(X)$ is another RV and its expectation is given by

$$\mathbb{E}[h(X)] = \int_{-\infty}^{\infty} h(x) f(x) dx$$

Variance

The variance of a RV X is defined by

$$\text{Var}(X) := \mathbb{E}[(X-\mu)^2] = \mathbb{E}[X^2] - \mu^2$$

We can see that $(X-\mu)^2$ is a function of X. Then, we can rewrite the $\text{Var}(X)$ as

$$\text{Var}(X) = \int_{-\infty}^{\infty} (x-\mu)^2 f(x) dx$$

The standard deviation is given by $\text{sd}(X) = \sqrt{\text{Var}(X)}$

Basic expectation properties

Consider a continuous RV X, two real value constants a and b, and two real value functions g and h. It is true that

- $\mathbb{E}[g(X) + h(X)] = \mathbb{E}[g(X)] + \mathbb{E}[h(X)]$
- $\mathbb{E}[aX + b] = a\mathbb{E}[X] + b$
- $\text{Var}(aX + b) = a^2\text{Var}(X)$

Multiple RVs

When working with two RVs, we must consider their possible relationship. The "joint" pdf $f_{XY}(x, y)$ is any function satisfying

$$\mathbb{P}(a \leq X \leq b, c \leq Y \leq d) = \int_a^b \int_c^d f_{XY}(x, y) dy dx$$

Additional properties

- Marginal pdf of X: $f_X(x) = \int_{-\infty}^{\infty} f_{XY}(x, y) dy$
- Marginal pdf of Y: $f_Y(y) = \int_{-\infty}^{\infty} f_{XY}(x, y) dx$
- Expectation of $g(X, Y)$:
  $\mathbb{E}[g(X, Y)] = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} g(x, y) f_{XY}(x, y) dy dx$

Conditional distribution

The conditional pdf of Y given $X = x$ is given by

$$f_{Y|X}(y|x) = \frac{f_{XY}(x, y)}{f_X(x)}$$

In case of independence, $f_{XY}(x, y) = f_X(x)f_Y(y)$ or $f_{Y|X}(y|x) = f_Y(y)$. It immediately follows the definition of conditional expectation

$$\mathbb{E}[Y|X = x] = \int_{-\infty}^{\infty} y \; f_{Y|X}(y|x)dy$$

Law of Iterated expectations

A useful tool in Econometrics is the **Law of Iterated expectations**. This is given by

$$\mathbb{E}_X[\mathbb{E}[Y|X=x]] = \mathbb{E}[Y]$$

The proof is straightforward using the marginal pdf definition

$$\begin{aligned}\mathbb{E}_X[\mathbb{E}[Y|X=x]] &= \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} y\ f_{Y|X}(y|x)dy f_X dx \\ &= \int_{-\infty}^{\infty} y \int_{-\infty}^{\infty} f_{XY}(x,y)dxdy \\ &= \int_{-\infty}^{\infty} y\ f_Y(y)dy = \mathbb{E}[Y]\end{aligned}$$

Where we used $f_{XY}(x,y) = f_{Y|X}(y|x)f_X$ and the definition of marginal pdf of y.

Covariance

Consider X and Y two RV with mean μ_X and μ_Y, respectively. Then, the covariance is given by

$$\begin{aligned}\text{Cov}(X,Y) :=& \mathbb{E}[(X-\mu_X)(Y-\mu_Y)] = \mathbb{E}[XY] - \mu_X\mu_Y \\ &= \int_{-\infty}^{\infty}\int_{-\infty}^{\infty}(x-\mu_X)(y-\mu_Y)f_{XY}(x,y)dxdy\end{aligned}$$

The correlation is given by

$$\text{Corr}(X,Y) := \frac{\text{Cov}(X,Y)}{\text{sd}(X)\text{sd}(Y)}$$

Normal distribution

Probably the most important distribution.

Standard normal distribution: $Z \sim \mathcal{N}(0, 1)$. Its support is $\mathbb{R}$. It is symmetric and "bell-shaped".

- pdf: $\phi(z) := \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}z^2} \quad \forall z \in \mathbb{R}$
- cdf: $\Phi(z) = \int_{-\infty}^{z} \phi(u) du$

We now see the general representation, but in the end, any normal distribution can be transformed into a standard normal.

Normal distribution

More generally, we work with any normal distribution. In this case, $X \sim \mathcal{N}(\mu, \sigma^2)$. As we can see, X may be thought of as a linear transformation of Z.

Something convenient about working with normal distributions is that we can easily sum two RVs. Take $X \sim \mathcal{N}(\mu_X, \sigma_X^2)$ and $Y \sim \mathcal{N}(\mu_Y, \sigma_Y^2)$. Then

$$X + Y \sim \mathcal{N}(\mu_X + \mu_Y, \sigma_X^2 + \sigma_Y^2 + 2\sigma_{XY})$$

where $\sigma_{XY} = \text{Cov}(X, Y)$.

Other probability distributions

The fact that the sum of two normal distributions is normal is a desirable feature. If we do not work with them, we must work with "convolutions", which may be nasty math.

In general, in Econometrics we use a couple of other distributions. For instance, the logistic distribution will be useful if we have time to cover dummy variables as outcomes.

Random sampling

Random sampling

We will start considering these sample statistics as RVs.

That is, we will think of the sample mean, $\bar{X}_n$, and the sample variance, $\bar{S}^2$, as RVs. Why can we do this? Simple, we are doing basic operations over random variables which are also random variables.

In this case, the subscript will usually represent the sample size (although it may also represent that we are just working with different samples).

Independence

The concept of independence is crucial in Econometrics. For this class, we will usually assume that the observations of our sample are "independent".

What does it mean? It means that the probability of observing one realization of a RV is not affected by the realization of another RV. For example, each time we toss a coin the first result is independent of the second result.

Technically, if two observations come from the two RV X_1 and X_2, we say that they are independent if
$\mathbb{P}(X_1 = x|X_2 = y) = \mathbb{P}(X_1 = x)\ \forall x, y$.

IID RVs

In addition to independence, we also consider RVs that are identically distributed. This means that each RV has the same distribution.

Getting back to our favorite example, each time we toss a coin we know that the probability of getting heads or tails is always 50%.

We will say that a list of n RVs, $X_1, X_2, ..., X_n$, satisfy random sampling if they are independently and identically distributed (i.i.d.). That is, random sampling and i.i.d. are the same concept.

Drawing with replacement

We can prove that random sampling in a finite population is also similar to drawing unit names i_1, i_2,..., i_n with equal probabilities, independently with replacement, and record x_{i_1}, ..., x_{i_n} (sample realizations).

Nonetheless, we will focus on i.i.d. RVs since they work for an infinite population.

Bias

Let $\hat{\theta}$ be an estimator of an unknown population parameter θ_0. The bias of $\hat{\theta}$ is the difference between the expected value of the estimator and the true parameter:

$$\text{Bias}(\hat{\theta}) = E[\hat{\theta}] - \theta_0.$$

We say that $\hat{\theta}$ is an unbiased estimator of θ_0 if:

$$E[\hat{\theta}] = \theta_0.$$

Intuitively, this means that if we repeatedly drew random samples and computed $\hat{\theta}$ each time, the average value of $\hat{\theta}$ would equal the true parameter θ_0.

The sampling distribution of the mean

Let $(X_i)_{i=1}^n$ be an i.i.d. sequence of RVs s.t. $\mathbb{E}[X_i] = \mu$ and $\text{Var}(X_i) = \sigma^2$. Define

$$\bar{X}_n := \frac{1}{n}\sum_{i=1}^{n} X_i$$

(As it was mentioned above, n stands for the sample size.) We can observe

- The sample mean is well-behave: $\mathbb{E}[\bar{X}_n] = \mathbb{E}\left[\frac{1}{n}\sum_{i=1}^n X_i\right] = \frac{1}{n}\sum_{i=1}^n \mathbb{E}\left[X_i\right] = \frac{n\mu}{n} = \mu$
- The variance shrinks to zero: $\text{Var}(\bar{X}_n) = \frac{\sigma^2}{n}$ (interestingly, the variance of $\sqrt{n}\bar{X}_n$is a constant).

Normal distribution

In the case of X_1, X_2, ..., $X_n \sim_{i.i.d.} \mathcal{N}(\mu, \sigma^2)$, we have

$$\bar{X}_n \sim \mathcal{N}\left(\mu, \frac{1}{n}\sigma^2\right)$$

We can also build the standard normal distribution as

$$\sqrt{n}\left(\frac{\bar{X}_n - \mu}{\sigma}\right) \sim \mathcal{N}(0, 1)$$

We will see that for non-normally distributed RVs if $n \to \infty$ we will still get the normality of the mean in the limit under certain conditions.

LLN

In Econometrics, we will usually think of $n \to \infty$ (this is called asymptotic analysis). A key theorem is the following

- **Law of large numbers:** Take $X_1, X_2, ...$ a sequence of i.i.d. RVs whose mean μ is finite. Then, as $n \to \infty$, it is true that

$$\bar{X}_n \xrightarrow{p} \mu$$

where $\xrightarrow{p}$ means convergence in probability. To be more specific, we also have $\xrightarrow{a.s.}$ (almost surely convergence), but $\xrightarrow{p}$ is the one we will be using often. We can also say that as $n \to \infty$, $\bar{X}_n$ degenerates into the constant μ.

CLT

Now, we study the most important result for this class.

- **Lindeberg-Lévy CLT:** Take $X_1, X_2, ...$ a sequence of i.i.d. RVs whose mean μ and σ^2 are finite. Then, as $n \to \infty$, it is true that

$$\sqrt{n}(\bar{X}_n - \mu) \xrightarrow{d} \mathcal{N}(0, \sigma^2)$$

where $\xrightarrow{d}$ means convergence in distribution.

Finite-sample vs asymptotic

For the mean estimator $\bar{X}_n$ of a sequence of normally distributed RVs, we have the following properties

Finite-sample properties:

- $\mathbb{E}[\bar{X}_n] = \mu$
- $\text{Var}(\bar{X}_n) = \frac{1}{n}\sigma^2$
- $\bar{X}_n \sim \mathcal{N}(\mu, \frac{1}{n}\sigma^2)$

Asymptotic properties:

- $\bar{X}_n \xrightarrow{p} \mu$
- $\text{Var}(\bar{X}_n) \to 0$
- $\sqrt{n}(\bar{X}_n - \mu) \xrightarrow{d} \mathcal{N}(0, \sigma^2)$

Finite-sample vs asymptotic

If we always (or at least in 99.9% of cases) work with finite samples, why should we care about asymptotic properties? The reason is that finite-sample properties are not always as trivial as in the case of $\bar{X}_n$. Since working out asymptotic properties is feasible (thanks to LLN and CLT), it is usually the case that we prefer to analyze asymptoticity in Econometrics.

Although some articles still focus on finite-sample properties, Econometrics papers usually analyze the asymptotic properties of the estimators. We will try to focus a little more on these kinds of properties as well.

CMT

We keep working with asymptotic properties.

Take $\hat{\theta}_n$ as an estimator of θ_0.
Define a continuous function $g : \mathbb{R} \to \mathbb{R}$.

- **Continuous mapping theorem (CMT):**

$$\hat{\theta}_n \xrightarrow{p} \theta_0 \Rightarrow g(\hat{\theta}_n) \xrightarrow{p} g(\theta_0)$$

This will be a very useful tool in the next classes.

Hypothesis testing

Confidence intervals

When we get the estimate $\bar{x}_n$ of the RV $\bar{X}_n$ for the mean μ how certain are we that the estimate is correct? Confidence intervals (CI) can help us with that

- A **CI** of **confidence level** $1-\alpha$ is a random interval $[\hat{A}_n, \hat{B}_n]$ computed from the sample data such that the probability of this interval containing the population parameter θ_0 is equal to $1-\alpha$:

$$\mathbb{P}(\hat{A}_n \leq \theta_0 \leq \hat{B}_n) = 1-\alpha$$

If we set $\alpha = 0.05$, we say that the confidence level is 0.95.

CI for normal mean

Given the symmetry of the normal distribution, we work with a symmetric (wrt the mean) CI.

For simplicity, assume we know σ. We can work with the standardized distribution. Take $\alpha = 0.05$, in any standard distribution Z, it is well known that $c = 1.96$ in $\mathbb{P}(-c \leq Z \leq c) = 0.95$. Hence,

$$\mathbb{P}\left(-1.96 \leq \frac{\sqrt{n}(\bar{X}_n - \mu)}{\sigma} \leq 1.96\right) = 0.95$$

$$\mathbb{P}\left(\bar{X}_n - 1.96\frac{\sigma}{\sqrt{n}} \leq \mu \leq \bar{X}_n + 1.96\frac{\sigma}{\sqrt{n}}\right) = 0.95$$

CI for normal mean

Then, in this case

$$\hat{A}_n = \bar{X}_n - 1.96 \text{ se}(\bar{X}_n)$$
$$\hat{B}_n = \bar{X}_n + 1.96 \text{ se}(\bar{X}_n)$$

Where $\text{se}(X_n) = \frac{\sigma}{\sqrt{n}}$.

Formal meaning: If we estimate many sample means, each with different confidence intervals, approximately 95% of them will contain μ.

Sample mean for CI

In the end, we build one interval using one sample, but the interpretation is not "with 95% of probability, μ is in the CI we estimated", but rather "if we do this same process many times, in 95% of times we will get μ in the interval". Actually, the μ is or is not in each calculated CI.

If we don't know σ (which is usually the case), we use the sample variance.

Recall that using the CLT we can say $\bar{X}_n$ is normally distributed in large n (even if each X_i is not). We can also work with other types of distributions.

Hypothesis Testing

A hypothesis test starts with two competing statements:

- Null hypothesis (H_0): the default claim or status quo.
- Alternative hypothesis (H_1): the claim we consider if the evidence against H_0 is strong enough.

Example: in a trial,

- H_0: the accused is innocent.
- H_1: the accused is guilty.

At the end of the test, we either reject H_0 or fail to reject H_0. We do not say that we "accept" H_0.

Type I and Type II Errors

Because decisions are based on evidence, hypothesis tests can make mistakes.

- Type I error: rejecting H_0 when H_0 is true.
- Type II error: failing to reject H_0 when H_0 is false.

Example: in a trial,

- Type I error: convicting an innocent person.
- Type II error: not convicting a guilty person.

The goal is not to eliminate all errors, but to control the probability of making them.

Hypothesis testing

5 steps for hypothesis testing

- Step 1: Declare H_0 (null hypothesis) and H_1 (alternative hypothesis)
- Step 2: Specify an acceptable level of Type 1 error, α, normally 0.05 or 0.01
- Step 3: Select a test statistic
- Step 4: Identify the values of the test statistic that lead to rejection of the null hypothesis
- Step 5: Use data to determine whether to reject H_0

Hypothesis testing

In this class, the null hypothesis will usually be of the form $\theta = 0$ (θ is our parameter of interest).

We want to avoid rejecting that $\theta = 0$ when it is true (this is the type 1 error where we will set a limit).

We will use the T-statistic. When n is sufficiently large ($n \geq 30$), the distribution of this test takes the same shape as the standard normal distribution.

Given the previous indication, if $\alpha = 0.05$, we know that we reject the null hypothesis if $|T_n| > 1.96$.

CI and HT

There is a duality between CI and HT. Take $X_1, ..., X_n \sim i.i.d.$ with $E[X_i] = \mu$ and $\text{Var}(X_i) = \sigma^2$. $H_0 : \bar{X}_n = \mu$ and $\alpha = 0.05$ as an example

$$\begin{aligned}
\text{reject } H_0 &\iff |T_n| := \frac{\sqrt{n}|\bar{X}_n - \mu_0|}{\hat{\sigma}} > 1.96 \\
&\iff \mu_0 \notin [\bar{X}_n - 1.96 \text{ se}(\bar{X}_n), \bar{X}_n + 1.96 \text{ se}(\bar{X}_n)] \\
&\iff \mu_0 \notin \text{CI}
\end{aligned}$$

This duality does not always apply. For instance, if X_i's are dummy variables and we want to analyze proportions, the CI is similarly defined, but the hypothesis test differs.

P-Value

Notice that rejecting the null hypothesis depends completely on the α we choose in the sense that larger αs make it easier to reject.

Then, we will look for the dividing line between not rejecting and rejecting (basically looking for the α that barely fails to reject).

P-value is the significance α such that the critical value c exactly equals the observed value of the test statistic.