Statistical Analysis and Decision Theory

Econ 6105, Fall 2024

Prof. Josh Abel

(CB chapters 7-9, 10.2)

Introduction

Probability theory is applied math.

- Distributions objectively result from model inputs and various rules/laws (i.e. math)

Statistics is the study of how to use observed data to infer the underlying model that produced it

- Primary challenge: the same dataset could be produced by many different models

So we want approaches that can identify the most "appealing" model, given the data.

- We're outside of pure math; we have some criteria/goals we strive for
- More like economics: optimizing something

A formalization of this is called "decision theory."

Loss functions

Let μ be some unknown object.
We need to decide what we think it is. Our decision will be denoted $\hat{\mu}$.
A **loss function**, $L(\mu, \hat{\mu})$ tells us how much disutility we suffer from that choice.
Two common example:

1. $L(\mu, \hat{\mu}) = (\mu - \hat{\mu})^2$
2. $L(\mu, \hat{\mu}) = |\mu - \hat{\mu}|$

Of course, we can never actually evaluate our loss because we don't know μ

- But this framework can allow us to properties of statistical approaches and their desirability

Risk functions

Suppose that our decision will be a function of data, $\hat{\mu}(X)$, where $X = \{x_1, x_2, ..., x_n\}$ is a sample drawn at random from some common distribution.

Because X is a RV, $\hat{\mu}(X)$ is a RV. Because it is an attempt to "estimate" μ, we call $\hat{\mu}(X)$ an **estimator**.

The **risk function** of an estimator is:

$$R(\mu, \hat{\mu}(X)) = E[L(\mu, \hat{\mu}(X))] \tag{1}$$

Analogous to the previous slide, two common examples are:

1. $R(\mu, \hat{\mu}) = E[(\mu - \hat{\mu}(X))^2]$
   - Known as "mean squared error" (MSE)
2. $R(\mu, \hat{\mu}) = E[|\mu - \hat{\mu}(X)|]$
   - We'll call this "mean absolute error" (MAE)

Prediction, MSE (population)

Suppose we will make a draw from some distribution F_X. Find the optimal prediction using the MSE criterion:

$$\hat{\mu} = \text{argmin}_a E[(X - a)^2] \tag{2}$$

Prediction, MSE (population)

Suppose we will make a draw from some distribution F_X. Find the optimal prediction using the MSE criterion:

$$\hat{\mu} = \text{argmin}_a E[(X - a)^2] \tag{2}$$

FOC:

$$-2 \cdot E[X - \hat{\mu}] = 0 \Rightarrow \hat{\mu} = E[X] \tag{3}$$

A new interpretation of the mean:

- MSE-minimizing prediction

Prediction, MSE (sample)

Suppose you have a sample of data, $\{x_1, x_2, ..., x_n\}$. We will draw one observation from the sample. Find the optimal prediction using the MSE criterion:

$$\hat{\mu}(X) = \text{argmin}_a \frac{1}{n} \cdot \sum_{i=1}^{n} (x_i - a)^2 \tag{4}$$

FOC:

$$-\frac{2}{n} \cdot \sum_{i=1}^{n} (x_i - a) = 0 \Rightarrow \hat{\mu} = \frac{1}{n} \cdot \sum_{i=1}^{n} x_i \tag{5}$$

The **sample mean**, $\frac{1}{n} \cdot \sum_{i=1}^{n} x_i$, is the best prediction of a draw from a sample, using the MSE criterion.

Prediction, MAE

Suppose you have a sample of data, $\{x_1, x_2, ..., x_n\}$. We will draw one observation from the sample. Find the optimal prediction using the MAE criterion:

$$\hat{\mu}(X) = \text{argmin}_a \frac{1}{n} \cdot \sum_{i=1}^{n} |x_i - a| \tag{6}$$

Prediction, MAE

Suppose you have a sample of data, $\{x_1, x_2, ..., x_n\}$. We will draw one observation from the sample. Find the optimal prediction using the MAE criterion:

$$\hat{\mu}(X) = \text{argmin}_a \frac{1}{n} \cdot \sum_{i=1}^{n} |x_i - a| \tag{6}$$

$$\hat{\mu}(X) = \{a : P(x \geq a) = 0.5\} = Med(X) \tag{7}$$

The **sample median** is the best prediction of a draw from a sample, using the MAE criterion.

Mean vs. median

Mean minimizes MSE; median minimizes MAE

- Kind of cool, but kind of frustrating
  - In some cases, you may really be able to say that you want to harshly penalize large errors (or not)
    - If so, this analysis tells you which estimator to use
  - But oftentimes you may not have a strong feeling about this

To help tailor a statistical approach, we can look more closely at how they behave. We'll focus on two features:

1. Sensitivity
2. Robustness

Sensitivity

Consider how the estimators respond to a change in the data:

$$\frac{d\bar{X}}{dx_i} = \frac{1}{n} \tag{8}$$

$$\frac{dMed(X)}{dx_i} = \begin{cases} 1 \text{ if } x_i = Med(X) \\ 0 \text{ if } x_i \neq Med(X) \end{cases} \tag{9}$$

This seems to be a point in favor of the mean

- $\bar{X}$ depends on every data point in a smooth way
- In contrast, changing a data point has no impact on the median, until/unless it becomes the median, at which point it has an enormous impact

Because any change to the data impacts the mean, we can say $\bar{X}$ is **sensitive**, where is $Med(X)$ is not.

Robustness

We may like that the mean responds to every data point.
But we also may want to make sure that 1 data point cannot have an outsized effect on the estimator.
Define an estimator's **breakdown value**, $s(\hat{\mu}(X))$, to be the largest share of the sample that can all simultaneously go to ∞ while $\hat{\mu}(X)$ remains finite.

- $s(\bar{X}) = 0$
- $s(Med(X)) = 0.5$

The median is more **robust** to outliers, which can be comforting.

Huber Estimator

Huber (1964) presented a compromise loss function:

$$L(x_i, \hat{\mu}(X)) = \begin{cases} (x_i - \hat{\mu}(X))^2 & \text{if } |x_i - \hat{\mu}(X)| \leq k \\ 2 \cdot k \cdot |x_i - \hat{\mu}(X)| - k^2 & \text{if } |x_i - \hat{\mu}(X)| > k \end{cases} \tag{10}$$

- $|x_i| \leq k$: acts like MSE, $(x_i - a)^2$
- $|x_i| > k$: acts like MAE, $|x_i - a|$
- The function is smooth, even at $x_i = k$ and $x_i = -k$:
  - $2 \cdot k \cdot |k| - k^2 = k^2$ and $2 \cdot k \cdot |-k| - k^2 = k^2$

Has MSE's sensitivity to data points when $|x_i|$ is not too large...but also MAE's robustness to outliers ($|x_i|$ large).

- Small k makes the loss function more like MAE
- Large k makes the loss function more like MSE

Despite the cleverness of the Huber estimator, it is not a workhorse approach, at least in part due to more abstract interpretation

Tractability

In practice, MSE is a lot easier to work with than MAE

- $\frac{dMSE}{da} = 0$ exists
  - We found the solution with our standard optimization approach
- $\frac{dMAE}{da}$ is never equal to 0
  - Recall, there was no FOC!

When summarizing one variable, calculating a median is not too much harder than a mean

- Though if the dataset is large, sorting is a pain, even for a computer

But minimizing MSE in a multivariate context (with a **least squares** regression) is much easier than minimizing MAE
So in practice, we spend much more time thinking about means than medians

- But keep in mind that this implicitly assumes a particular loss function (MSE)

Trimming/Winsorizing

While we privilege the mean, the previous discussion highlighted a concern

- Lack of robustness to outliers

In practice, it is common to deal with outliers by either:

1. **Trimming**: Dropping observations with values that are greater than the 99th percentile or below the 1st percentile; or
2. **Winsorizing**: Replacing all values that were *above* the 99th percentile *equal* to the 99th percentile, and replacing all values that were *below* the 1st percentile *equal* to the 1st percentile

This should not be done blindly. Consider why are there outliers...

- If you suspect a data error, trimming makes a lot of sense
- If you think the data is true, winsorizing makes more sense
  - But in some cases, those outliers may be very important, so you don't want to lose/alter them
  - In other cases, you may want to find a general pattern, so winsorizing will allow you to make sure outliers don't obscure that

Bias-Variance Tradeoff

$$\text{MSE}(\mu, \hat{\mu}) \equiv E[(\mu - \hat{\mu}(X))^2] = \underbrace{\left(\mu - E[\hat{\mu}(X)]\right)^2}_{\text{Bias}^2} + \underbrace{E\left[(\hat{\mu} - E[\hat{\mu}(X)])^2\right]}_{\text{Variance}} \tag{11}$$

An estimator will have larger squared errors on average (i.e. MSE) if:

1. Bias: it is systematically too high or too low
2. Variance: it is volatile

BVT: Interpretation 1

Suppose we are trying to estimate the mean of a distribution, μ_X. Consider taking a random sample, $\{x_1, ..., x_n\}$, and calculating the sample average, $\bar{X}$.

$$E[\bar{X}] = E\left[\frac{1}{n}\sum x_n\right] = \frac{1}{n}\cdot n \cdot \mu = \mu \quad (12)$$

In words, the sample mean is **unbiased**, or:

$$E\left[E[\bar{X}] - \mu\right] = 0 \quad (13)$$

We already know that $\bar{X}$ minimizes MSE. Therefore:

- Because MSE $=$ Bias^2 $+$ Var...
- ...and Bias$(\bar{X}) = 0$...
- ...$\bar{X}$ must have the lowest variance of all unbiased estimators

$\bar{X}$ is **efficient**, or it is the **best unbiased estimator (BUE)**.

BVT: Interpretation 2

The Bias-Variance Tradeoff has a more provocative – and useful! – interpretation:

- **Introducing bias may be beneficial if it sufficiently reduces variance**

BVT example

Trying to estimate μ with data from 2 sources:

1. $x_i \sim N(\mu, 5)$ for $i = 1, ..., 5$
2. $x_i \sim N(\mu - 0.5, 5)$ for $i = 6, ..., 55$

Consider 2 estimators:

1. Unbiased: $\frac{1}{5} \cdot \sum_{i=1}^{5} x_i$
   - MSE $=$
2. Biased: $\frac{1}{55} \cdot \sum_{i=1}^{55} x_i$
   - MSE $=$

BVT example

Trying to estimate μ with data from 2 sources:

1. $x_i \sim N(\mu, 5)$ for $i = 1, ..., 5$
2. $x_i \sim N(\mu - 0.5, 5)$ for $i = 6, ..., 55$

Consider 2 estimators:

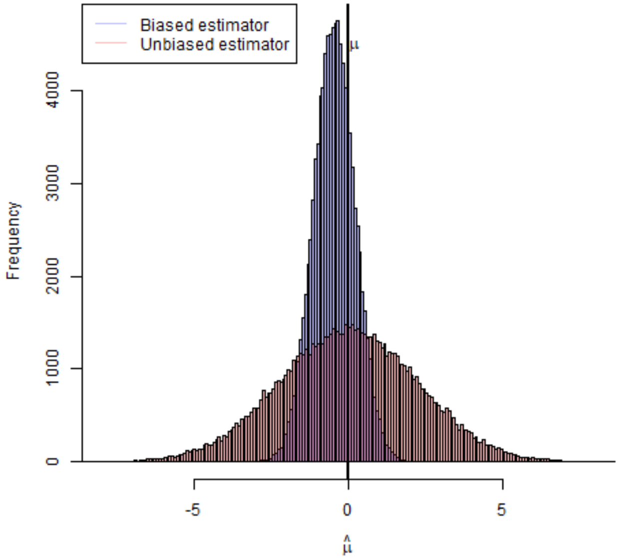
1. Unbiased: $\frac{1}{5} \cdot \sum_{i=1}^{5} x_i$
   - MSE $= 4.98$
2. Biased: $\frac{1}{55} \cdot \sum_{i=1}^{55} x_i$
   - MSE $= 0.66$

Estimator 2 does introduce some bias because it's using data that did not come from the "target" population...

...but it performs much better because it is much more stable! The bias was worth the reduction in variance.

If you wanted to predict the temperature on 11/30/2024, would you look at past data only from 11/30, or perhaps all days in late November?

BVT example, visualized



Contrast with axiomatic approach

Sometimes, statistical techniques are motivated with an "axiomatic" approach

- Let's use the efficient estimator, i.e. lowest-variance unbiased estimator, i.e. BUE
- This is often how regression is motivated. Gauss-Markov Theorem says OLS is BLUE.

(IMO), decision theory gives a much more compelling way to think about statistical analysis.

- Why restrict yourself to only unbiased estimators?
- Doing so ignores that you can likely get more precise estimates by admitting some bias
- That tradeoff should be taken seriously

When we get to regression later in the semester, I will motivate it as the result of decision theory, not from axioms as is often done.

Hypothesis Testing

A hypothesis test is a familiar use of decision theory.
Consider some parameter θ that we are interested in. If $\theta \in \Theta$ and $\Theta_0 \subset \Theta$, then a **null hypothesis** is the following statement:

$$H_0 : \theta \in \Theta_0. \tag{14}$$

The **alternative hypothesis** is:

$$H_A : \theta \in \Theta_0^c. \tag{15}$$

A **hypothesis test** is a procedure for using data to decide whether to reject H_0 in favor of H_A.

Familiar approach

Suppose you have data drawn from a Normal distribution with mean μ. You want to test the following hypotheses:

$$H_0 : \mu = \mu_0; \; H_A : \mu \neq \mu_0 \tag{16}$$

- I.e. "$\Theta_0 = \mu_0$," "$\Theta_0^c =$ the rest of R^1"

Further suppose you have data randomly sample from that distribution, $\{x_1, ..., x_n\}$, with sample mean $\bar{x}$.
Define:

$$z \equiv \frac{\bar{x} - \mu_0}{sd(x_i)/\sqrt{n}} \tag{17}$$

Note that if H_0 is true, $z \sim N(0, 1)$:

- $E[z] = \frac{\sqrt{n}}{sd(x_i)} \cdot (E[\bar{x}] - \mu_0) = 0$
- $var(z) = \frac{var(\bar{x})}{var(x_i)/n} = \frac{var(\sum x_i)/n^2}{var(x_i)/n} = \frac{n \cdot var(x_i)/n^2}{var(x_i)/n} = 1$

Familiar approach (2)

$$H_0 : \mu = \mu_0; \;\; H_A : \mu \neq \mu_0$$

$$z \equiv \frac{\bar{x} - \mu_0}{sd(x_i)/\sqrt{n}} \sim N(0,1)$$

Because $z \sim N(0,1)$, $P(|z| \geq 1.96) = 0.05$.

- Equivalently, $P(|\bar{x} - \mu_0| \geq 1.96 \cdot sd(x_i)/\sqrt{n}) = 0.05$

Therefore, if $|\bar{x} - \mu_0| \geq 1.96 \cdot sd(x_i)/\sqrt{n}$, either:

1. H_0 is wrong; or
2. Something very unusual just happened (i.e. less than 5% likelihood)

By definition, “very unusual” things are unlikely, so H_0 seems unlikely to be true

- We reject H_0 “at the 5% level.”
- If we had found $|\bar{x} - \mu_0| < 1.96 \cdot sd(x_i)/\sqrt{n}$, we would not reject H_0 at the 5% level

Likelihood Ratio Test

Suppose $x_i \sim f(x|\theta)$, with unknown parameter θ, and we draw a random sample from the population to form a dataset, $\{x_1, ..., x_n\}$.
For any hypothesized value of θ, denoted by $\hat{\theta}$, the dataset's likelihood functions is given by:

$$L(\mathbf{x}|\hat{\theta}) = \prod_{i=1}^{n} f(x_i|\hat{\theta}) \tag{18}$$

Given hypotheses $H_0 : \theta \in \Theta_0$ and $H_A : \theta \in \Theta_0^c$, the **test statistic of the LRT** is:

$$\lambda(\mathbf{x}) = \frac{\sup_{\hat{\theta} \in \Theta_0} L(\mathbf{x}|\hat{\theta})}{\sup_{\hat{\theta} \in \Theta} L(\mathbf{x}|\hat{\theta})} \tag{19}$$

Likelihood Ratio Test (2)

Given hypotheses $H_0 : \theta \in \Theta_0$ and $H_A : \theta \in \Theta_0^c$, the **test statistic of the LRT** is:

$$\lambda(\mathbf{x}) = \frac{\sup_{\hat{\theta} \in \Theta_0} L(\mathbf{x}|\hat{\theta})}{\sup_{\hat{\theta} \in \Theta} L(\mathbf{x}|\hat{\theta})} \tag{20}$$

- The numerator is the best you can do if you restrict yourself to only the hypothesized region of Θ, Θ_0
- The denominator is the best you can do when you have access to the entire parameter space

Note that $\lambda(\mathbf{x}) \in [0, 1]$. A **likelihood ratio test** involves picking some critical level, c, such that:

1. H_0 is rejected if $\lambda(\mathbf{x}) \leq c$
2. H_0 is not rejected if $\lambda(\mathbf{x}) > c$

LRT for Mean of a Normal RV

Suppose you have data drawn from a Normal distribution with mean μ. You want to test the following hypotheses:

$$H_0 : \mu = \mu_0; \ H_A : \mu \neq \mu_0$$

It can be shown that $L(\mathbf{x}, \hat{\theta})$ is maximized at $\theta = \bar{x}$. Therefore:

$$\begin{aligned}
\lambda(\mathbf{x}) &= \frac{L(\mathbf{x}|\mu_0)}{L(\mathbf{x}|\bar{x})} \\
&= \frac{exp\big(-\sum(x_i - \mu_0)^2/2\big)}{exp\big(-\sum(x_i - \bar{x})^2/2\big)} \\
&= exp\Big(\big(\sum(x_i - \bar{x})^2 - \sum(x_i - \mu_0)^2\big)/2\Big) \\
&= exp\Big(-n \cdot (\bar{x} - \mu_0)^2/2\Big)
\end{aligned} \tag{21}$$

LRT for Mean of a Normal RV (2)

$$H_0 : \mu = \mu_0; \; H_A : \mu \neq \mu_0$$

$$\lambda(\mathbf{x}) = exp\Big(- n \cdot (\bar{x} - \mu_0)^2/2\Big)$$

Rejecting when $\lambda(\mathbf{x}) \leq c$ means we will reject when:

$$|\bar{x} - \mu_0| \geq \sqrt{-2 \cdot \ln(c)/n} \equiv z^* \tag{22}$$

Since $c \in (0, 1)$, $z^* \in (0, \infty)$, so the LRT amounts to saying:

- If the sample mean is close enough to our hypothesized value (above or below), we will not reject...
- If the sample mean is far enough to our hypothesized value (above or below), we will reject...

So the familiar test of comparing $|\bar{x} - \mu_0|$ to $1.96 \cdot sd(x_i)/\sqrt{n}$ is a specific LRT.

- $z^* = 1.96 \cdot sd(x_i)/\sqrt{n}$

Type I and Type II Errors

	Accept H_0	Reject H_0
H_0 is true	✓	Type I Error
H_A is true	Type II Error	✓

There is a tradeoff between errors of Types I and II.

A permissive test could set c low, e.g. z^* high

- High threshold for rejection

Type I and Type II Errors

	Accept H_0	Reject H_0
H_0 is true	✓	Type I Error
H_A is true	Type II Error	✓

There is a tradeoff between errors of Types I and II.
A permissive test could set c low, e.g. z^* high

- High threshold for rejection
  - Type I relatively rare
  - Type II relatively common

A strict test could set c high, e.g. z^* low

- Low threshold for rejection
  - Type I relatively common
  - Type II relatively rare

Power and Size

To operationalize these errors, we define the following:

- A test's **size** is the highest probability of a Type I Error among all $\theta \in \Theta_0$
  - If $\Theta_0 = \theta_o$ is a single value as in our prior example, then the size is equal to $P(\text{Reject } H_0 | \theta = \theta_0)$
  - In your homework, you'll consider a test in which Θ_0 is a set
- A test's **power function**, $\beta(\theta)$, shows the probability of a Type I Error for all values of $\theta \in \Theta_0^c$

Size and power function, example

Hypotheses:

$$H_0 : \mu = 1; \; H_A : \mu \neq 1$$

Define $z \equiv \frac{\bar{x}-1}{\sigma_{x_i}/\sqrt{n}}$. Note that if $\mu = 1$, $z \sim N(0,1)$.

Consider two hypothesis testing approaches:

1. Reject if $|z| \geq z^* = 1.96$
   - Size: $P(z \leq -1.96 | \mu - 1 = 0) + P(z \geq 1.96 | \mu - 1 = 0) = 5\%$
2. Reject if $|z| \geq z^* = 2.58$
   - Size: $P(z \leq -2.58 | \mu - 1 = 0) + P(z \geq 2.58 | \mu - 1 = 0) = 1\%$

Note that the size of the test does not depend on the details of data!

- σ_{x_i} and n are not part of the calculation; only z^* matters
- Such a test is designed so that if H_0 is true, we have a fixed probability (1%, 5%, etc) of incorrectly rejecting it

Power functions are more complicated...

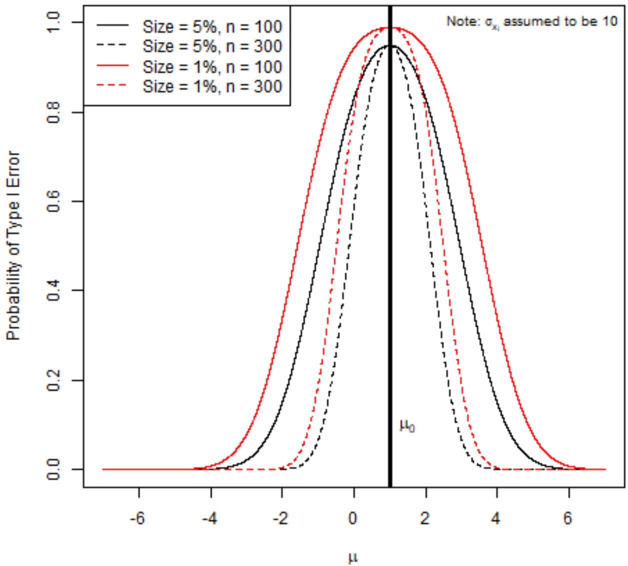
Size and power function, example (2)

$$\begin{aligned}\beta(\mu) &= P\left(\left|\frac{\bar{x}-1}{\sigma_{x_i}/\sqrt{n}}\right| < z^* \mid E[x_i]=\mu\right)\\ &= P\left(\left|\frac{\bar{x}-\mu}{\sigma_{x_i}/\sqrt{n}} + \frac{\mu-1}{\sigma_{x_i}/\sqrt{n}}\right| < z^* \mid E[x_i]=\mu\right)\\ &= P\left(\left|\tilde{z} + \frac{\mu-1}{\sigma_{x_i}/\sqrt{n}}\right| < z^* \mid E[x_i]=\mu\right)\\ &= P\left(\tilde{z} + \frac{\mu-1}{\sigma_{x_i}/\sqrt{n}} < z^* \text{ AND } \tilde{z} + \frac{\mu-1}{\sigma_{x_i}/\sqrt{n}} > -z^* \mid E[x_i]=\mu\right)\end{aligned} \tag{23}$$

So the probability of a Type II Error is below, where $\tilde{z} \sim N(0,1)$:

$$P\left(\tilde{z} \in \left(-z^* - \frac{\mu-1}{\sigma_{x_i}/\sqrt{n}}, z^* - \frac{\mu-1}{\sigma_{x_i}/\sqrt{n}}\right) \mid E[x_i]=\mu\right) \tag{24}$$

Size and power function example, graph



Size and power function, intuition

A tradeoff exists been size and power

- You can avoid Type II Errors by setting a high rejection threshold (low size)...
  - ...but then you open yourself up to a high probability of Type I Errors
- You could set a low rejection threshold (high size), which will lead to more Type II Errors...
  - ...but the benefit will be to reduce Type I Errors

Increasing sample size improves the tradeoff:

- Can lower one type of error without increasing the other one (or lower both)

Power calculation, setup

These concepts can guide data collection/experimental design.
Suppose you want to evaluate the following hypothesis:

$$H_0 : \theta \equiv \mu_1 - \mu_2 = 0; \; H_A : \theta \equiv \mu_1 - \mu_2 \neq 0 \tag{25}$$

- μ_1 is the population average in group 1
- μ_2 is the population average in group 2
  - Perhaps you are comparing earnings between people who receive job training and those who do not

You will collect data from random samples of the two groups and estimate:

$$\hat{\theta} = \hat{\mu}_1 - \hat{\mu}_2 = \bar{x}_1 - \bar{x}_2 = \frac{1}{n_1} \sum_{\text{Group 1}} x_i - \frac{1}{n_2} \sum_{\text{Group 2}} x_i \tag{26}$$

If data collection is costly, you need to think about how many observations to get

Power calculation, setup (2)

Perhaps, like most economists, you want your size to be 5%

- I.e., if $\theta = 0$, you want to incorrectly reject H_0 only 5% of the time

Let's also assume $\sigma_{x_i} = 30$

- Maybe you get this from a small initial sample, or from another dataset on earnings

Finally, assume that based on you economic setting, you think it's very important to reject H_0 if $|\theta| > 10$.

- I.e. if $\theta = 1$ and you incorrectly accept that $\theta = 0$, maybe that's not a big deal...
- But if $\theta = 10$ would lead an incorrect conclusion that $\theta = 0$ to be "bad"

Power calculation

Since size is 5%, will reject if $|z| > 1.96$. Therefore:

$$\begin{aligned} P(\text{Type I Error}|\theta = 10) &= P\left(\left|\frac{\hat{\theta} - 0}{sd(\hat{\theta})}\right| < 1.96|\theta = 10\right) \\ &= P\left(\left|\frac{\hat{\theta} - 10}{sd(\hat{\theta})} + \frac{10}{sd(\hat{\theta})}\right| < 1.96|\theta = 10\right) \\ &= P\left(\left|\tilde{z} + \frac{10}{sd(\hat{\theta})}\right| < 1.96\right), \end{aligned} \tag{27}$$

where $\tilde{z} \sim N(0, 1)$. Further note:

$$\begin{aligned} sd(\hat{\theta}) &= \sqrt{var(\hat{\theta})} \\ &= \sqrt{var(\bar{x}_1) + var(\bar{x}_2)} \\ &= \sqrt{\sigma^2_{x_i}/n_1 + \sigma^2_{x_i}/n_2} \end{aligned} \tag{28}$$

Power calculation (2)

So:

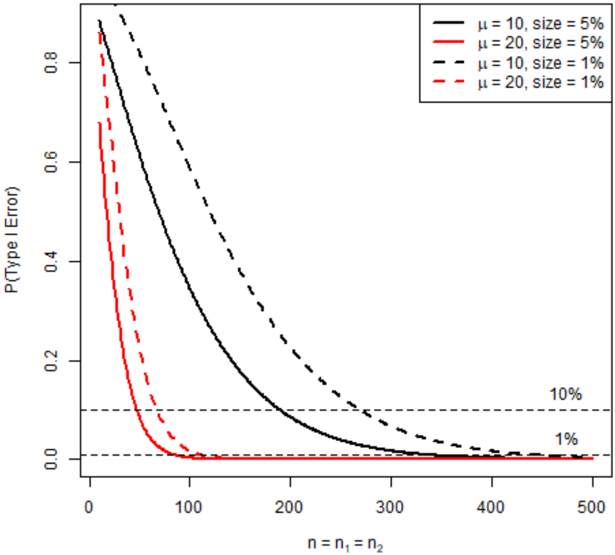
$$
\begin{aligned}
P(\text{Type I Error}|\theta = 10) &= P\left(\left|\tilde{z} + \frac{10}{sd(\hat{\theta})}\right| < 1.96\right) \\
&= P\left(\tilde{z} < 1.96 - \frac{10}{\sqrt{30^2/n_1 + 30^2/n_2}}\right) \\
&- P\left(\tilde{z} < -1.96 - \frac{10}{\sqrt{30^2/n_1 + 30^2/n_2}}\right)
\end{aligned} \tag{29}
$$

If:

- $n_1 = n_2 = 50$, this is 61.5%
- $n_1 = n_2 = 100$, this is 34.6%

If you deem that a 10% probability of this error is acceptable, need $n_1 = n_2 >= 190$

Power calculation, visualize



Power calculation, intuition

Need more data when:

1. You want to keep the probability of Type I Error low (1% vs. 10%)
2. When you want to keep the probability of a Type II error low (1% vs. 5%)
3. You want to rule out values closer to the null hypothesis (10 vs. 20)

This relatively simple application of statistics tells you how much data you need to gather!

p-values

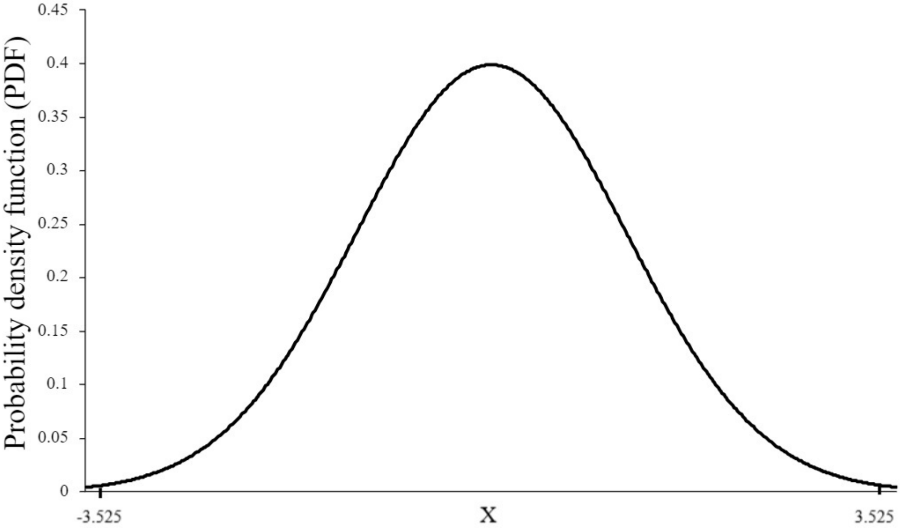
A hypothesis test gives a binary result.

- E.g. "If $|z| > 1.96$ I will reject H_0; otherwise I will not."

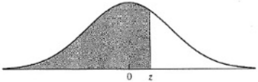
A **p-value** of a null hypothesis is the lowest tolerated size for which the test would be rejected.

- Recall, size is $P(\text{Reject } H_0 | \ H_0 \text{ true})$.
  - So if size $= 5\%$, you are willing to reject H_0 5% of the time when it is true
- If you observe $z = 3.525 > 1.96$, you would reject for size $= 5\%$
  - But you would also reject for size $= 1\%$, since $z = 3.525 > 2.58$
- Taking this to its limit, $P(|z| > 3.525) = 0.04\%$
  - Intuitive interpretation: "If H_0 is true, the data I observed ($z = 3.525$) is so extreme that it only had a 0.04% chance of occurring. Therefore, if my tolerance for a Type II Error is greater than 0.04%, I will reject H_0."
- So the p-value $= 0.04\%$.

p-values, visualized

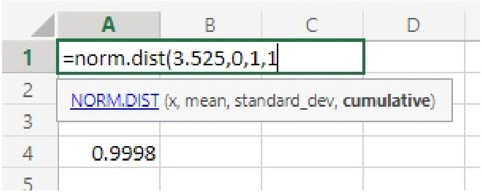


p-values, visualized



z	0.00	0.01	0.02	0.03	0.04	0.05	0.06	0.07	0.08	0.09
0.0	.5000	.5040	.5080	.5120	.5160	.5199	.5239	.5279	.5319	.5359
0.1	.5398	.5438	.5478	.5517	.5557	.5596	.5636	.5675	.5714	.5753
0.2	.5793	.5832	.5871	.5910	.5948	.5987	.6026	.6064	.6103	.6141
0.3	.6179	.6217	.6255	.6293	.6331	.6368	.6406	.6443	.6480	.6517
0.4	.6554	.6591	.6628	.6664	.6700	.6736	.6772	.6808	.6844	.6879
0.5	.6915	.6950	.6985	.7019	.7054	.7088	.7123	.7157	.7190	.7224
0.6	.7257	.7291	.7324	.7357	.7389	.7422	.7454	.7486	.7517	.7549
0.7	.7580	.7611	.7642	.7673	.7704	.7734	.7764	.7794	.7823	.7852
0.8	.7881	.7910	.7939	.7967	.7995	.8023	.8051	.8078	.8106	.8133
0.9	.8159	.8186	.8212	.8238	.8264	.8289	.8315	.8340	.8365	.8389
1.0	.8413	.8438	.8461	.8485	.8508	.8531	.8554	.8577	.8599	.8621
1.1	.8643	.8665	.8686	.8708	.8729	.8749	.8770	.8790	.8810	.8830
1.2	.8849	.8869	.8888	.8907	.8925	.8944	.8962	.8980	.8997	.9015
1.3	.9032	.9049	.9066	.9082	.9099	.9115	.9131	.9147	.9162	.9177
1.4	.9192	.9207	.9222	.9236	.9251	.9265	.9279	.9292	.9306	.9319
1.5	.9332	.9345	.9357	.9370	.9382	.9394	.9406	.9418	.9429	.9441
1.6	.9452	.9463	.9474	.9484	.9495	.9505	.9515	.9525	.9535	.9545
1.7	.9554	.9564	.9573	.9582	.9591	.9599	.9608	.9616	.9625	.9633
1.8	.9641	.9649	.9656	.9664	.9671	.9678	.9686	.9693	.9699	.9706
1.9	.9713	.9719	.9726	.9732	.9738	.9744	.9750	.9756	.9761	.9767
2.0	.9772	.9778	.9783	.9788	.9793	.9798	.9803	.9808	.9812	.9817
2.1	.9821	.9826	.9830	.9834	.9838	.9842	.9846	.9850	.9854	.9857
2.2	.9861	.9864	.9868	.9871	.9875	.9878	.9881	.9884	.9887	.9890
2.3	.9893	.9896	.9898	.9901	.9904	.9906	.9909	.9911	.9913	.9916
2.4	.9918	.9920	.9922	.9925	.9927	.9929	.9931	.9932	.9934	.9936
2.5	.9938	.9940	.9941	.9943	.9945	.9946	.9948	.9949	.9951	.9952
2.6	.9953	.9955	.9956	.9957	.9959	.9960	.9961	.9962	.9963	.9964
2.7	.9965	.9966	.9967	.9968	.9969	.9970	.9971	.9972	.9973	.9974
2.8	.9974	.9975	.9976	.9977	.9977	.9978	.9979	.9979	.9980	.9981
2.9	.9981	.9982	.9982	.9983	.9984	.9984	.9985	.9985	.9986	.9986
3.0	.9987	.9987	.9987	.9988	.9988	.9989	.9989	.9989	.9990	.9990
3.1	.9990	.9991	.9991	.9991	.9992	.9992	.9992	.9992	.9993	.9993
3.2	.9993	.9993	.9994	.9994	.9994	.9994	.9994	.9995	.9995	.9995
3.3	.9995	.9995	.9995	.9996	.9996	.9996	.9996	.9996	.9996	.9997
3.4	.9997	.9997	.9997	.9997	.9997	.9997	.9997	.9997	.9997	.9998
3.5	.9998	.9998	.9998	.9998	.9998	.9998	.9998	.9998	.9998	.9998
3.6	.9998	.9998	.9999	.9999	.9999	.9999	.9999	.9999	.9999	.9999

p-values, visualized

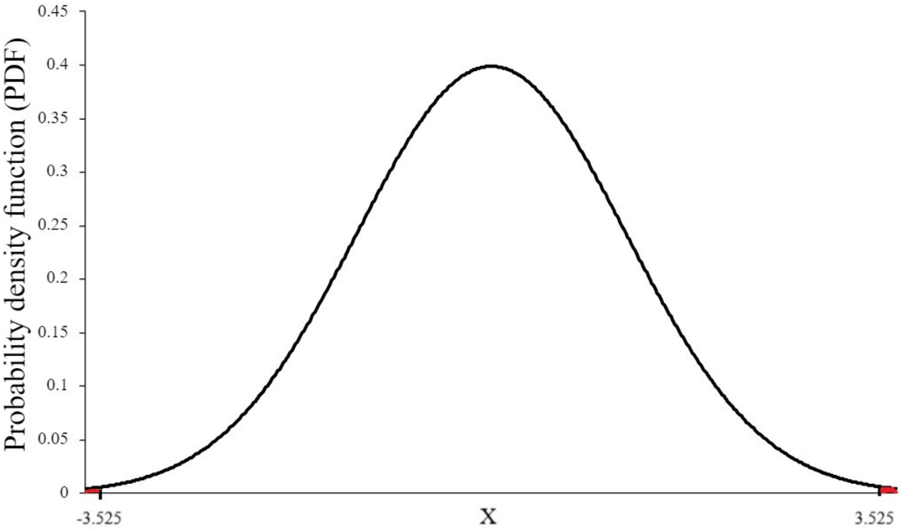


p-values, visualized

R Console

```
R version 4.2.1 (2022-06-23 ucrt) -- "Funny-Looking Kid"
Copyright (C) 2022 The R Foundation for Statistical Computing
Platform: x86_64-w64-mingw32/x64 (64-bit)

R is free software and comes with ABSOLUTELY NO WARRANTY.
You are welcome to redistribute it under certain conditions.
Type 'license()' or 'licence()' for distribution details.

  Natural language support but running in an English locale

R is a collaborative project with many contributors.
Type 'contributors()' for more information and
'citation()' on how to cite R or R packages in publications.

Type 'demo()' for some demos, 'help()' for on-line help, or
'help.start()' for an HTML browser interface to help.
Type 'q()' to quit R.

> pnorm(3.525,0,1)
[1] 0.9997883
> 
```

p-values, visualized



Set Estimation

So far we have focused on point estimation

1. Produce a single "best guess" of a parameter
2. Test whether that parameter is equal to a certain hypothesized value

But perhaps we need not be constrained to giving a single "best guess"

- Maybe, a set of "reasonable guesses" would suffice
  - E.g. "I think μ is between 1.5 and 2.3," rather than "I think μ is 1.9."

The main manifestation of this in applied econometrics is confidence intervals

Inverting a Hypothesis Test

Recall our standard test of $H_0 : \mu = \mu_0$

$$\text{Reject } H_0 \text{ if } \left|\frac{\bar{x} - \mu_0}{\sigma_{x_i}/\sqrt{n}}\right| > z^*$$

While this can be used to test a specific μ_0, we can also rearrange to find all values of μ_0 such that the test will not lead to rejection:

$$\begin{aligned}
\text{Do not reject } H_0 & \text{ if } \left|\frac{\bar{x} - \mu_0}{\sigma_{x_i}/\sqrt{n}}\right| < z^* \\
& \text{ if } \frac{\bar{x} - \mu_0}{\sigma_{x_i}/\sqrt{n}} < z^* \text{ AND } \frac{\bar{x} - \mu_0}{\sigma_{x_i}/\sqrt{n}} > -z^* \\
& \text{ if } \mu_0 > \bar{x} - z^* \cdot \frac{\sigma_{x_i}}{\sqrt{n}} \text{ AND } \mu_0 < \bar{x} + z^* \cdot \frac{\sigma_{x_i}}{\sqrt{n}} \\
& \text{ if } \mu_0 \in \left[\bar{x} - z^* \cdot \frac{\sigma_{x_i}}{\sqrt{n}}, \bar{x} + z^* \cdot \frac{\sigma_{x_i}}{\sqrt{n}}\right]
\end{aligned}$$

Confidence Intervals

$$\text{CI}(x_1, ..., x_n) \equiv \left[\bar{x} - z^* \cdot \frac{\sigma_{x_i}}{\sqrt{n}}, \bar{x} + z^* \cdot \frac{\sigma_{x_i}}{\sqrt{n}}\right] \tag{30}$$

By choosing z^* to correspond to a hypothesis test of size α, the set CI is called a **(1-α) confidence interval**

- E.g. by setting $z^* = 1.96$, we get a 95% confidence interval

Note that $\text{CI}(\mathbf{x})$ is a RV, since it is a function of randomly sampled data. Furthermore, $\mu \in CI(\mathbf{x})$ with probability $1 - \alpha$.

- Suppose $\mu = \tilde{\mu}$
- Drawing a $\bar{x}$ so far away from $\tilde{\mu}$ that $\tilde{\mu} \in \text{CI}(\mathbf{x})^c$...
- ...is the same event as rejecting $H_0 : \mu = \tilde{\mu}$, even though it's true.
- The latter occurs with probability α, so the former must as well
- So $P(\tilde{\mu} \in \text{CI}(\mathbf{x})) = 1 - \alpha$

Confidence interval, example

$\bar{x} = 25$, $\sigma_{x_i} = 15$
95% CI:

1. $n = 10$: $[15.7, 34.3]$
2. $n = 30$: $[19.6, 30.4]$

Alternative confidence sets

$\bar{x} = 25$, $\sigma_{x_i} = 15, n = 10$

- $P(\mu \in [15.7, 34.3]) = 95\%$.

Note also:

1. $P(\mu \in [-\infty, 24.7]$ U $[25.3, \infty]) = 95\%$
   -
2. $P(\mu \in [-\infty, 32.8]) = 95\%$
   -
3. $P(\mu \in [17.2, \infty]) = 95\%$
   -

What do you think of these alternatives?

-

Alternative confidence sets

$\bar{x} = 25$, $\sigma_{x_i} = 15$, $n = 10$

- $P(\mu \in [15.7, 34.3]) = 95\%$.

Note also:

1. $P(\mu \in [-\infty, 24.7]$ U $[25.3, \infty]) = 95\%$
   - Seems awful. We've ruled out only a tiny interval, [24.7,25.3], and it's the interval that contains the point estimate, 25!
2. $P(\mu \in [-\infty, 32.8]) = 95\%$
   -
3. $P(\mu \in [17.2, \infty]) = 95\%$
   -

What do you think of these alternatives?

-

Alternative confidence sets

$\bar{x} = 25$, $\sigma_{x_i} = 15$, $n = 10$

- $P(\mu \in [15.7, 34.3]) = 95\%$.

Note also:

1. $P(\mu \in [-\infty, 24.7]$ U $[25.3, \infty]) = 95\%$
   - Seems awful. We've ruled out only a tiny interval, [24.7,25.3], and it's the interval that contains the point estimate, 25!
2. $P(\mu \in [-\infty, 32.8]) = 95\%$
   - Much wider than the original, but if you really don't want to miss too high, this is better.
3. $P(\mu \in [17.2, \infty]) = 95\%$
   -

What do you think of these alternatives?

-

Alternative confidence sets

$\bar{x} = 25$, $\sigma_{x_i} = 15, n = 10$

- $P(\mu \in [15.7, 34.3]) = 95\%$.

Note also:

1. $P(\mu \in [-\infty, 24.7] \text{ U } [25.3, \infty]) = 95\%$
   - Seems awful. We've ruled out only a tiny interval, [24.7,25.3], and it's the interval that contains the point estimate, 25!
2. $P(\mu \in [-\infty, 32.8]) = 95\%$
   - Much wider than the original, but if you really don't want to miss too high, this is better.
3. $P(\mu \in [17.2, \infty]) = 95\%$
   - Much wider than the original, but if you really don't want to miss too low, this is better.

What do you think of these alternatives?

- The original is the shortest, which seems desirable. But options 2 and 3 are the inversions of the tests of $H_0 : \mu < \mu_0$ and $H_0 : \mu > \mu_0$, respectively, so could be defensible depending on your goals.