Statistics and Probability Theory

A Short Refresher

Outline

1. Probability Distribution: Measures and Examples
2. Estimation
3. Inference

Sources

- Books on Statistics and Probability Theory:

  In German:

  ◇ Schira, J. (2016): *Statistische Methoden der VWL und BWL*, 5. Auflage, Pearson Studium, München

  In English:

  ◇ Stock, J. and Watson, M. (2015): *Introduction in Econometrics*, 3^{rd} Edition, Pearson, Boston

  ◇ Wackerly, D.D., Menderhall, W., Schaeffer, R.L. (2008): *Mathematical Statistics with Applications*, 7^{th} Edition, Brooks/Cole Cengage

- Refresher as appendix of introductory econometrics textbooks:

  ◇ Wooldridge, J.M. (2019) : *Introductory Econometrics: A Modern Approach*, 7^{th} Edition, South Western, Cengage Learning

  ◇ Green, W.H.(2008): *Econometrics Analysis*, 6^{th} Edition, Pearson Prentice Hall

  ◇ Pohlmeier, W. (2020): *Econometrics I, Lecture notes*, Department of Economics, University of Konstanz

1. Probability Distribution: Measures and Examples

1. Probability Distribution: Random Variable

- **Random Variable (RV):** a variable whose outcome is uncertain.
- *E.g. the number of people infected with Corona Virus (COVID-19), the monthly salary (earning) of an individual, the crime rate of a city, the GDP of a country, the price of the Apple stock (AAPL) on NASDAQ*
- Types of RV:
  - ◇ **Discrete**: the outcome is countable.
    E.g. the number of people infected with a virus, the number of car accidents, outcome of throwing a dice
  - ◇ **Continuous**: the outcome is infinitely divisible, and, thus, not countable.
    E.g. the monthly earning, the college grade point average (GPA), the price of a stock, the probability of getting infected with a virus

1. Probability Distribution: Definitions

- Denote by X a RV. The listing of the possible values x that X takes and the associated probabilities is denoted the **probability distribution** of X.
- Denote by $f_X(x)$ the probability distribution of X.

  → For discrete RV: $f_X(x)$ is denoted the **probability mass function** and it is defined as $f_X(x) = \text{Prob}(X = x)$, such that

  1. $0 \leq f_X(x) \leq 1$
  2. $\sum_x f_X(x) = 1$

  → For continuous RV: $f_X(x)$ is denoted the **probability density function** (**pdf**) and satisfies the following conditions:

  1. $\text{Prob}(X = x) = 0$
  2. $\text{Prob}(a \leq X \leq b) = \int_a^b f_X(x)dx \geq 0$
  3. $\int_{-\infty}^{+\infty} f_X(x)dx = 1$

  → The pdf of X is also known as the *marginal* pdf of X.

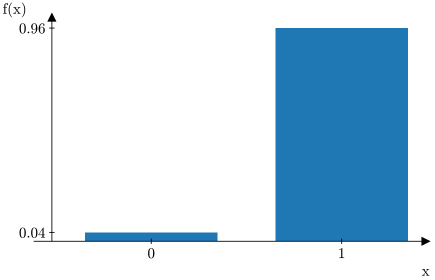
Example of a discrete RV

Assume X to be a binary variable that takes two values:

- value 1 if a person infected with the Corona virus has recovered from the virus.
- value 0 if a person infected with the Corona virus has died of the virus.

$$X = \begin{cases} 1, & \text{if recovered} \\ 0, & \text{if died} \end{cases}$$

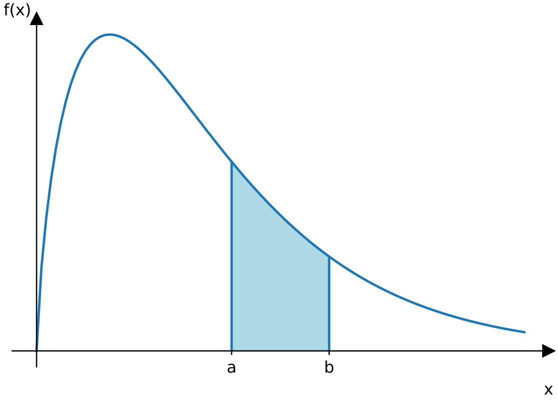
- Let $\text{Prob}(X = 1) = f_X(1) = 0.96$
- Then $\text{Prob}(X = 0) = f_X(0) = 1 - \text{Prob}(X = 1) = 0.04$



- X is Bernoulli distributed with parameter $p \equiv f_X(1)$.

Graph of *f(x)* for a continuous RV

$\text{Prob}(a \leq X \leq b) = \int_a^b f_X(x)dx$



1. Probability Distribution: Cumulative Distribution

- The **cumulative distribution function** (**cdf**), denoted by $F(x)$, gives the probability that X is less than or equal to value a:

$$F_X(a) = Prob(X \leq a) = \begin{cases} \sum\limits_{x \leq a} f_X(x), & \text{if } X \text{ is discrete} \\ \int\limits_{-\infty}^{a} f_X(x)dx, & \text{if } X \text{ is continuous} \end{cases}$$

- $f_X(x) = \dfrac{\partial F_X(x)}{\partial x}$ for continuous RV

  Properties:

  - $0 \leq F_X(x) \leq 1$
  - If $a > b, F_X(a) \geq F_X(b)$
  - $F_X(+\infty) = 1$
  - $F_X(-\infty) = 0$
  - $\text{Prob}(a < x \leq b) = F_X(b) - F_X(a)$

1. Probability Distribution: Mean or Expected Value

- **Expected Value:** measures the central tendency of the distribution of a RV.

$$\mathrm{E}[X] = \begin{cases} \sum\limits_{x} x\ f_X(x), & \text{if } X \text{ is discrete} \\ \int\limits_{-\infty}^{+\infty} x\ f_X(x)dx, & \text{if } X \text{ is continuous} \end{cases}$$

- *In the example of the discrete bivariate variable X describing an infected with Corona Virus being dead or recovered:* $E[X] = 1 \cdot Prob(X = 1) + 0 \cdot Prob(X = 0) = 1 \cdot 0.96 + 0 \cdot 0.04 = 0.96$

  Properties:

  - If c is a constant, then $\mathrm{E}[c] = c$
  - If a and b are constants, then $\mathrm{E}[aX + b] = a\mathrm{E}[X] + b$
  - If X and Y are two RVs, then $\mathrm{E}[X + Y] = \mathrm{E}[X] + \mathrm{E}[Y]$

1. Probability Distribution: Variance

- **Variance:** measures the dispersion of the distribution of a RV.

$$\mathrm{V}[X] = \mathrm{E}\left[(X-\mathrm{E}[X])^2\right] = \begin{cases} \sum_x (x-\mathrm{E}[X])^2 f_X(x), & \text{if } X \text{ is discrete} \\ \int_{-\infty}^{+\infty} (x-\mathrm{E}[X])^2 f_X(x) d(x), & \text{if } X \text{ is continuous} \end{cases}$$

- *E.g., The variance of the binary variable X from above is equal to:*
  $V[X] = (1-0.96)^2 \cdot 0.96 + (0-0.96)^2 \cdot 0.04 = 0.0348$
  Properties:
  - If c is a constant, then $\mathrm{V}[c] = 0$
  - If X is RV, then $\mathrm{V}[X] > 0$
  - $\mathrm{V}[X] = \mathrm{E}[X^2] - \mathrm{E}[X]^2$
  - If a is a constant, then $\mathrm{V}[aX] = a^2\mathrm{V}[X]$
  - If X and Y are RVs, then $\mathrm{V}[X+Y] = \mathrm{V}[X] + \mathrm{V}[Y] + 2\mathrm{Cov}[X,Y]$
  - $\mathrm{Cov}[X,Y] = \mathrm{E}[(X-\mathrm{E}[X])(Y-\mathrm{E}[Y])] = \mathrm{E}[XY] - \mathrm{E}[X]\mathrm{E}[Y]$
  - The standard deviation of X: $\mathrm{sd}[X] = \sqrt{\mathrm{V}(X)}$
  - $\mathrm{Corr}[X,Y] = \frac{\mathrm{Cov}[X,Y]}{\mathrm{sd}[X]\mathrm{sd}[Y]}$

1. Probability Distribution: Skewness and Kurtosis

- **Skewness:** measures the degree of symmetry of a distribution around its mean.

$$S(X) = \frac{\mathrm{E}\left[(X - \mathrm{E}[X])^3\right]}{\mathrm{sd}[X]^3}$$

  - ◇ If the distribution of X is symmetric, $S(X) = 0$
  - ◇ If the distribution of X is skewed, $S(X) \neq 0$
  - ◇ If $S(X) > 0$, then the distribution is shifted to the left (tail on the right side): right-skewed
  - ◇ If $S(X) < 0$, then the distribution is shifted to the right (tail on the left side): left-skewed

- **Kurtosis:** measures the shape a distribution; i.e. how fat (long, heavy) the tails of the distribution are.

$$K(X) = \frac{\mathrm{E}\left[(X - \mathrm{E}[X])^4\right]}{\mathrm{sd}[X]^4}$$

  - ◇ If $K(X) > 3$, the distribution of X has fat tails.

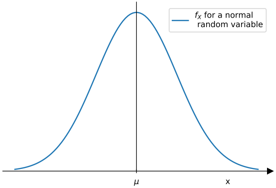
1. Probability Distribution: Normal or Gaussian distribution

Consider a RV X to be normally distribution with $\mathrm{E}[X] = \mu$ and $\mathrm{V}[X] = \sigma^2$, then

$$f(x \mid \mu, \sigma^2) = \frac{1}{\sigma\sqrt{2\pi}} e^{\frac{(x-\mu)^2}{2\sigma^2}} \tag{1}$$

Properties:

- $S(X) = 0$
- If $\mu = 0$ and $\sigma = 1$, then X is standard normally distributed: $S(X) = 0$, $K(X) = 3$
- It is "stable" under summation:
  – If a and b are constants, then $aX + b$ has also a normal distribution with $\mathrm{E}[aX + b] = a\mu + b$ and $\mathrm{V}[aX + b] = a^2\sigma^2$;
  – If X and Y are normally distributed, then $X + Y$ is also normally distributed with $\mathrm{E}[X + Y] \ = \mathrm{E}[X] + \mathrm{E}[Y]$ and $\mathrm{V}[X + Y] = \mathrm{V}[X] + \mathrm{V}[Y] + 2\mathrm{Cov}[X, Y]$



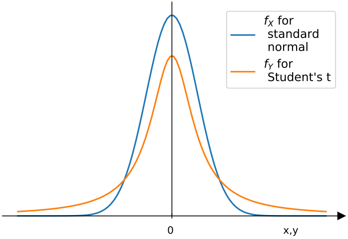
1. Probability Distribution: Student's t Distribution

If X is a Student's t distributed RV with υ degrees of freedom ($\upsilon > 0$), then:

$$f(x \mid \upsilon) = \frac{\Gamma\left(\frac{\upsilon+1}{2}\right)}{\sqrt{\upsilon\pi}\,\Gamma\left(\frac{\upsilon}{2}\right)} \left(1 + \frac{x^2}{\upsilon}\right)^{-\frac{\upsilon+1}{2}}, \text{ where } \Gamma(\cdot) \text{ is the gamma function}$$

Properties:

- $\mathrm{E}[X] = 0$ if $\upsilon > 1$
- $\mathrm{V}[X] = \frac{\upsilon}{\upsilon - 2}$ if $\upsilon > 2$
- $S[X] = 0$ if $\upsilon > 3$
- $K[X] = \frac{6}{\upsilon - 4} + 3 > 3$ if $\upsilon > 4$
- It is not "stable" under summation:
  – If a and b are constants, then $aX + b$ has a non-standardized Student's t distribution with another *pdf* than above
  – If X and Y are Student's t distributed, then $X + Y$ is usually not Student's distributed.



1. Probability Distribution: Joint Distribution

- **Joint Distribution:** describes the occurrence of events involving more than one RV.
  E.g. It gives the joint probability that an individual is 22 years old (age is given by the variable X) and earns monthly 2200 Euros (salary is given by the variable Y)
- The joint distribution function is formally denoted by $f_{XY}(x, y)$.
- If X and Y are discrete: $f_{XY}(x, y) = \text{Prob}(X = x, Y = y)$
- If X and Y are continuous: $\text{Prob}(a \leq X \leq b, c \leq Y \leq d) = \int_a^b \int_c^d f_{XY}(x, y)dydx$
- X and Y are **independent** if and only if:
  $f_{XY}(x, y) = f_X(x) \cdot f_Y(y)$, where $f_X(x)$ is the marginal *pdf* of X and $f_Y(y)$ is the marginal *pdf* of Y.
- If X and Y are two independent RVs, then X and Y are uncorrelated, i.e. $\text{Cov}[X, Y] = \text{Corr}[X, Y] = 0$
- If X and Y are two uncorrelated RVs, then they are not necessarily independent.
- Only exception: If X and Y are normally distributed and uncorrelated, then they are also independent.

1. Probability Distribution: Conditional Distribution

- **Conditional Distribution:** describes the occurrence of an event involving a RV X given the occurrence of another event involving another RV, Y.
  E.g. It gives the probability of earning monthly 2200 Euros (salary described by the variable Y) given that one is 22 years old (age described by the variable X).
- The conditional distribution of Y given X is formally denoted by $f_{Y|X}(y \mid x)$.
- $f_{Y|X}(y \mid x) = \frac{f_{XY}(x,y)}{f_X(x)}$, where $f_X(x)$ is the (marginal) pdf of X.
- If X and Y are independent: $f_{Y|X}(y \mid x) = f_Y(y)$
  I.e. the occurrence of an event for X plays no role for the occurrence of an event for Y.

1. Probability Distribution: Conditional Mean

- The conditional mean of Y given X:

$$\mathrm{E}[Y \mid X] = \begin{cases} \sum_{y} y \; f_{Y|X}(y \mid x), & \text{if } Y \text{ is discrete} \\ \int_{-\infty}^{+\infty} y \; f_{Y|X}(y \mid x) dy, & \text{if } Y \text{ is continuous} \end{cases}$$

Properties:

1. $\mathrm{E}[Y \mid X]$ is a function of X: $\mathrm{E}[Y \mid X] = h(X)$
2. $\mathrm{E}[g(X) \mid X] = g(X)$
3. If X and Y are independent, $\mathrm{E}[Y \mid X] = \mathrm{E}[Y]$
4. The Law of Iterated Expectations:
   $\mathrm{E}[Y] = E_X[\mathrm{E}[Y \mid X]]$, where $E_X[\cdot]$ indicates the expectation over the values of X.

2. Estimation

2. Estimation: An Example

- *Assume a sample of 5 individuals and 2 RVs that take the values in the table below:*

Person	Age (x)	Monthly gross earning in EUR (y)
1	20	1900
2	21	2000
3	20	1700
4	25	2100
5	22	2200

- *The mean of age $E[X]$ is estimated by the average of x's in the table:*
$$\overline{x} = \frac{\sum_{i=1}^{5} x_i}{5} = \frac{20+21+20+25+22}{5} = 21.6$$
- *The mean of earnings $E[Y]$ is estimated by the average of y's in the table:* $\overline{y} = 1980$
- *The variance of earnings $V[Y]$ is consistently estimated by the sample variance given the values y in the table:*

$$s^2 = \frac{1}{5-1}\sum_{i=1}^{5}(y_i - \overline{\mathrm{y}})^2 = \frac{1}{4}\cdot\left[(1900-1980)^2 + \cdots + (2200-1980)^2\right] = 37000$$

2. Estimation: An Example

Person	Age (x)	Monthly gross earning in EUR (y)
1	20	1900
2	21	2000
3	20	1700
4	25	2100
5	22	2200

- *The conditional mean of earnings given that the age is 20* $E[Y \mid X = 20]$ *is estimated by the average of y's in the Table for the persons 1 and 3:* $\overline{y}_{|(X=20)} = \dfrac{1900 + 1700}{2} = 1800$

2. Estimation: Estimator vs. Estimate

- Let θ be a parameter describing the distribution of X;
  E.g., $\mu = E[X]$, $\sigma^2 = V[X]$.
- To estimate θ from a random sample of n observations drawn from the population, i.e., $x = (x_1, ..., x_n)'$, define an estimator for θ.
- An **estimator** for θ, denoted $\hat{\theta}$, is a <u>*known*</u> function of x: $\hat{\theta} = T_\theta(x)$.
  E.g., We are looking for an estimator of the mean: $\mu = \mathrm{E}[X]$:
  - ◇ *The average (sample mean) of the observed sample* $x_1, ...x_n$
    I.e. $\hat{\mu} = \bar{x} = \frac{\sum_{i=1}^{n} x_i}{n}$ *or*
  - ◇ *an alternative estimator* $\tilde{\mu} = \frac{\sum_{i=1}^{n} x_i}{n-1}$

  Which of the two estimators, $\hat{\mu}$ *vs* $\tilde{\mu}$*, is* ***better*** *to estimate* μ*?*
- An estimator is a RV.
- The realization (the value) of an estimator computed at the values $x_1, ..., x_n$ is denoted the **estimate**.

2. Estimation: Criteria to Judge Estimators

- Based on their:
  - Statistical Properties:
    1. unbiasedness
    2. efficiency
    3. mean squared error
    4. consistency, ...
  - Robustness to outliers
  - Computational burden

2. Estimation: Bias, Efficiency, Mean Squared Error

1.) An estimator $\hat{\theta} = T_\theta\left(x_1,...x_n\right)$ of θ is **unbiased** if and only if, for any n (or sample),

$$\mathrm{E}\left[\hat{\theta}\right] = \mathrm{E}\left[T_\theta\left(x_1,...x_n\right)\right] = \theta$$

E.g.

- *The empirical average is an unbiased estimator of the mean of a RV.* $E[\hat{\mu}] = \mu$, *however* $E[\tilde{\mu}] \neq \mu$, *where* $\hat{\mu}$ *and* $\tilde{\mu}$ *are defined on slide 21.*
- *Let* $s^2 = \dfrac{1}{n-1}\sum_{i=1}^{n}(x_i - \bar{x})^2$ *and* $\hat{\sigma}^2 = \dfrac{1}{n}\sum_{i=1}^{n}(x_i - \bar{x})^2$ *be two estimators of the variance of* X, $\sigma^2 = V[X]$. *Then* $E\left[s^2\right] = \sigma^2$, *but* $E\left[\hat{\sigma}^2\right] \neq \sigma^2$.

The **bias** is defined as: $B\left(\hat{\theta} \mid \theta\right) = \mathrm{E}\left[\hat{\theta} - \theta\right] = \mathrm{E}\left[\hat{\theta}\right] - \theta$

2.) **Efficiency:** Let $\hat{\theta}$ and $\tilde{\theta}$ be two unbiased estimators of θ; I.e., $\mathrm{E}\left[\hat{\theta}\right] = \mathrm{E}[\tilde{\theta}] = \theta$.

$\hat{\theta}$ is more **efficient** than $\tilde{\theta}$ if $\mathrm{V}\left[\hat{\theta}\right] < \mathrm{V}\left[\tilde{\theta}\right]$.

3.) **Mean Squared Error:** gives the trade-off between bias and variance.

$$MSE\left[\hat{\theta} \mid \theta\right] = \mathrm{E}\left[(\hat{\theta} - \theta)^2\right] = B\left(\hat{\theta} \mid \theta\right)^2 + \mathrm{V}\left[\hat{\theta}\right]$$

2. Estimation: How to get an Estimator?

- *A special case: the location parameter model (LPM):*

$$Y_i = \mu + \varepsilon_i \ , \qquad i = 1,...,n \tag{2}$$

$$with \quad \varepsilon_i \quad i.i.d, \quad E[\varepsilon_i] = 0 \quad and \quad V[\varepsilon_i] = 1 \quad \forall i$$

  - ◇ *LPM is a regression equation, consisting of:*
    1. the observable variable, Y_i
    2. the unobservable error term, ε_i
    3. the unknown parameter that need to be estimated from the sample $y_1,....,y_n$*, namely* μ
  - ◇ *E.g., if* Y_i *gives the score from an IQ test for a person* i*, then* $\mu = E[Y_i]$ *in Equation (2) is the expected value of the score.*

2. Estimation: Least Squares (LS)

- *Assume* $\tilde{\mu}$ *to be a generic estimator of* μ.
- *Define* $\tilde{e}_i = y_i - \tilde{\mu}$ *to be the* <u>*estimated error*</u> *or the* <u>*residual*</u> *of the regression in Equation (2).*
- *The LS estimator of* μ*, denoted by* $\hat{\mu}$*, is the* $\tilde{\mu}$ *that minimizes the sum of squared residuals:* $S = \sum_{i=1}^{n} \tilde{e}_i^2$*, I.e.* $\hat{\mu} = \underset{\tilde{\mu}}{argmin}\ S$
- *To compute* $\hat{\mu}$*, we solve the first order condition:*

$$\frac{\partial S}{\partial \tilde{\mu}} = -2\sum_{i=1}^{n}(y_i - \tilde{\mu}) \overset{!}{=} 0 \qquad (3)$$

- *From Equation (3), we get that the LS estimator of* μ*:* $\hat{\mu} = \bar{y}$
- *The LS estimator* $\hat{\mu}$ *is unbiased:* $E[\hat{\mu}] = \mu$
- *The variance of the LS estimator* $\hat{\mu}$ *is equal to:* $V[\hat{\mu}] = V[\bar{y}] = 1/n \Rightarrow sd[\bar{y}] = \sqrt{1/n}$

2. Estimation: Point Estimation

- So far we talked about **point estimation**: the realisation of an estimator $\hat{\theta}$ based on a random sample $x_1;...,x_n$.
- The distribution of the estimator $\hat{\theta}$, which is a RV per se, is usually a continuous one.
- Thus, the probability that it may become equal to the true value θ is equal to zero: $Prob\left(\hat{\theta} = \theta\right) = 0$
- We hope, however, that in expectation $\hat{\theta}$ is equal to θ (I.e., $\hat{\theta}$ is unbiased): $\mathrm{E}\left[\hat{\theta}\right] = \theta$

2. Estimation: Confidence Interval

- Therefore it is reasonable to find the range of plausible values that $\hat{\theta}$ may take such that the true value θ is within this range in a specified proportion of the sample (i.e. with a certain confidence level)
- This range is known as the **confidence interval** and it has a lower (c_1) and a upper (c_2) value.
- If the confidence level is $1 - \alpha$, then c_1 and c_2 are chosen s.t. $Prob\,(\theta \in [c_1, c_2]) = 1 - \alpha$.
- c_1 and c_2 can easily be derived based on the distribution of $\hat{\theta}$.

  E.g., Assume $\hat{\theta}$ is standard normally distributed, $N(0,1)$. For a given value of α, c_1 and c_2 can be obtained from textbook tables with the critical values of the standard normal distribution; I.e. if $\alpha = 0.05 \Rightarrow c_1 = -1.645$ *and* $c_2 = 1.645$
  Thus, the 95% confidence interval of $\hat{\theta}$ is given by $[-1.645, 1.645]$.

3. Inference

3. Inference: Testing

- **Test**: A decision problem with respect to (w.r.t.) an unknown parameter or w.r.t. a relation between different unknown parameters.
  E.g. The mean μ is:
  (A) different,
  (B) larger or
  (C) smaller than a certain value μ_0.
- In order to undergo a test, one has to decide what is the null (H_0) and what is the alternative (H_A) hypothesis
- H_0 is assumed to be true until the data suggest otherwise (like the innocence of a defendant of a jury trial).
- H_A is associated with the theory one would like to prove.

(A) Two-sided test	*(B) One-sided test*	*(C) One-sided test*
H_0: $\mu = \mu_0$	$H_0 : \mu \leq \mu_0$	$H_0 : \mu \geq \mu_0$
$H_A : \mu \neq \mu_0$	$H_A : \mu > \mu_0$	$H_A : \mu < \mu_0$

3. Inference: Testing

- In order to be able to take a decision about H_0, one needs to choose a *test statistic*.
- A *test statistic* is a **known** function of the random sample of observations at hand and it is a RV per se.
- A test statistics is chosen s.t. its distribution is **known** under H_0.
- Its distribution under H_A is different from the one under H_0 and it is usually not known to the econometrician.

3. Inference: Testing

- *Back to the LPM model:*

$$Y_i = \mu + \varepsilon_i \ , \qquad i = 1,\ldots,n \tag{4}$$

$$\text{with} \quad \mathrm{E}[\varepsilon_i] = 0 \quad \text{and} \quad \mathrm{V}[\varepsilon_i] = 1 \quad \forall i$$

- *Assume that* $\varepsilon_i \overset{i.i.d}{\sim} N(0,1)$. Thus, $Y_i \overset{i.i.d}{\sim} N(\mu, 1)$
- *The test statistic for (A)* H_0: $\mu = \mu_0$, (B) $H_0 : \mu \leq \mu_0$ and (C) $H_0 : \mu \geq \mu_0$ is given by:

$$T = \frac{\hat{\mu} - \mu_0}{\sqrt{\mathrm{V}(\hat{\mu})}} = \frac{\bar{Y} - \mu_0}{sd[\bar{Y}]} \sim t_{(n-1)}, \tag{5}$$

where t_{n-1} is Student's t distributed with $n-1$ degrees of freedom.

- *The t-statistic* T *from Equation (5) measures the distance from* $\bar{Y}$ *to* μ_0 *relative to the standard deviation of* $\bar{Y}$, $sd[\bar{Y}]$.
- *Let* t *be the estimate of* T *computed from the sample* $y_1,\ldots,y_n$.

3. Inference: Significance Level & Critical Values

- Choose a significance level (decision rule of the test), denoted α.
- The significance level is the probability of rejecting H_0 even though H_0 is true; I.e., it is the probability of committing a Type I error:

$$\text{Prob}(reject\ H_0 \mid H_0) = \text{Prob}(Type\ I\ error) = \alpha$$

- Usually α is chosen to be small: $1\%, 5\%, 10\%$.
- Based on the choice of the α, find the critical value c, corresponding to the α based on the distribution of the test statistic T under H_0.

(A) Two-sided test	*(B) One-sided test*	*(C) One-sided test*
$H_0\colon\ \mu = \mu_0$	$H_0 : \mu \leq \mu_0$	$H_0 : \mu \geq \mu_0$
$\text{Prob}(\lvert T\rvert > c \mid H_0) = \alpha$	$\text{Prob}(T > c \mid H_0) = \alpha$	$\text{Prob}(T < -c \mid H_0) = \alpha$

3. Inference: Significance Level & Critical Values

Take the decision of rejecting or <u>not</u> rejecting H_0 at the level α by comparing t to c.

(A) Two-sided test

H_0: $\mu = \mu_0$

- if $|t| \leq c \Rightarrow$ **do not reject** H_0 at α
- if $|t| > c \Rightarrow$ **reject** H_0 at α

(B) One-sided test

$H_0 : \mu \leq \mu_0$

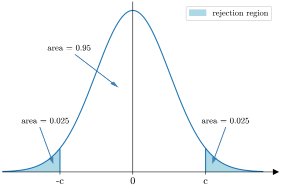
- if $t \leq c \Rightarrow$ **do not reject** H_0 at α
- if $t > c \Rightarrow$ **reject** H_0 at α

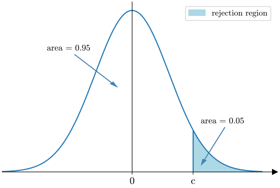
(C) One-sided test

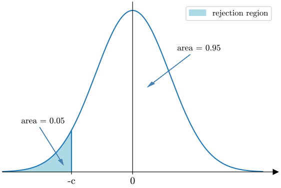
$H_0 : \mu \geq \mu_0$

- if $t \geq -c \Rightarrow$ **do not reject** H_0 at α
- if $t < -c \Rightarrow$ **reject** H_0 at α

In the following graphs, let $\alpha = 0.05$.







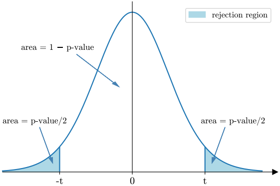
3. Inference: P-Value

- Based on the choice of the significance level α, one provides only limited results for the testing (up to the choice of α).
- One can provide the complete set of results for testing by reporting the **p-value**.
- **p-value**: is the smallest significance level at which H_0 can be rejected or the largest significance level at which H_0 can not be rejected.

(A) Two-sided test

$$H_0\colon \mu = \mu_0$$

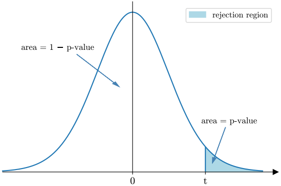
$$\text{p-value} = \text{Prob}(|T| > t \mid H_0)$$



(B) One-sided test

$$H_0 : \mu \leq \mu_0$$

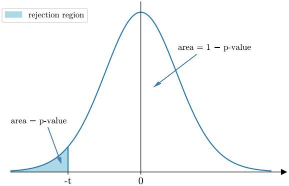
$$\text{p-value} = \text{Prob}(T > t \mid H_0)$$



(C) One-sided test

$$H_0 : \mu \geq \mu_0$$

$$\text{p-value} = \text{Prob}(T < -t \mid H_0)$$



- If p-value $\leq \alpha$, then reject H_0 at α; I.e. if p-value $= 0$, then reject H_0 $\forall \alpha$.

3. Inference: Sum-Up the Testing Steps

Step 1 Set up the H_0 and H_A.

Step 2 Compute the test statistic T.

Step 3 Find the distribution of T under the condition of H_0.

Step 4 Decide on the significance level α.

Step 5 (Usually) read the critical value c for the distribution at 3 for the level α from textbook tables.

Step 6 Compare the estimate of T, namely t to c.

Step 7 Take a decision about H_0 at the level α.

3. Inference: Testing Based on Confidence Interval

- *As already stated above, the test-statistic T for the mean μ in LPM, has the distribution:*

$$T = \frac{\hat{\mu} - \mu_0}{\sqrt{1/n}} \sim t_{(n-1)}$$

- *Choose the confidence level $1 - \alpha$.*
- *Then the $(1 - \alpha)$ confidence interval of T is equal to*

$$\left[T - t^*_{(n-1,1-\frac{\alpha}{2})}, T + t^*_{(n-1,1-\frac{\alpha}{2})}\right],$$

where $t^*_{(n-1,1-\frac{\alpha}{2})}$ is the critical value corresponding to the Student's t distribution with $n - 1$ degree of freedom computed at the level of $1 - \frac{\alpha}{2}$. This critical value can be obtained from the textbook tables with the critical values for the Student's t distribution.

- *Thus, the $(1 - \alpha)$ confidence interval of μ is equal to:*

$$\left[\bar{y} - t^*_{(n-1,1-\frac{\alpha}{2})} 1/\sqrt{n}, \bar{y} + t^*_{(n-1,1-\frac{\alpha}{2})} 1/\sqrt{n}\right] \tag{6}$$

- *The null hypotheses H_0: $\mu = \mu_0$ is rejected at the significance level α if μ_0 is not in the confidence interval given in Equation (6).*