Econometric Methods

Introduction

Alberto Cappello

Department of Economics, Boston College

Spring 2025

Basic Idea of Causal Inference

- Social science (Economics) theories always ask causal questions.
- In general, a typical causal question is:

  The effect of a treatment (X) on an outcome (Y).

- **Outcome (Y):** A variable that we are interested in.
- **Treatment (X):** A variable that has the (causal) effect on the outcome of our interest.
- The best way to address this question is conducting a **Randomized Controlled Experiment**:
  - **Treatment group:** Receives a treatment.
  - **Control group:** Does not receive the treatment.

  the two groups are identically equal except for being treated or non-treated.

Outcome of Interest:

$$\Delta Y = \text{Outcome for treated individuals}(Y_1) - \text{Outcome for control individuals}(Y_0)$$

Example: Fertilizer and Crop Yield

Description:

- A randomized trial is conducted to evaluate the effect of a new fertilizer on crop yield.
- Farmers are randomly assigned to:
  - **Treatment group:** Use the new fertilizer.
  - **Control group:** Use traditional farming methods without the new fertilizer.

Findings:

- Average crop yield increased by 15% in the treatment group compared to the control group.

Why Randomization Works:

- Ensures that treated and control groups are similar in observed and unobserved characteristics (e.g., soil quality, farmer skills).
- Any difference in yield is attributable to the fertilizer.

Randomized Experiments in Econometrics

Randomized Experiments are often not feasible

- **Practical constraints**
- **Confounding Factors**

Quasi-Experimental Methods:

- **Difference-in-Differences (DiD):**
  - Compares pre- and post-treatment outcomes between treated and control groups.
- **Instrumental Variables (IV):**
  - Uses an external factor (instrument) that affects treatment assignment but not the outcome directly (e.g., weather patterns influencing fertilizer adoption).
- **Regression Discontinuity (RD):**
  - Exploits a cutoff rule for treatment assignment (e.g., subsidies based on farm size thresholds).
- **Propensity Score Matching (PSM):**
  - Matches treated and control units with similar observed characteristics (e.g., similar soil quality and farm size).

Example: Using Difference-in-Differences

Study: Impact of Fertilizer Subsidy on Crop Yield

- Government introduces a fertilizer subsidy for small farmers in one region (treatment group) but not in another region (control group).
- Compare crop yields before and after the subsidy.

Why This is Quasi-Experimental:

- No randomization of the subsidy.
- farmers may differ in observable and unobservable characteristics that can affect both treatment assignment and crop yield.

Potential Confounding Factors in Crop Yield Example

Potential Confounding Factors:

- **Soil Quality:**
  - Fields with better soil naturally have higher yields (Y).
  - Farmers with lower-quality soil may be more likely to invest in fertilizer (X).
- **Sunlight Exposure:**
  - Fields with better sunlight exposure have higher productivity (Y).
  - Sunlight exposure is often unobserved and may vary systematically across regions.
- **Technology and Farming Practices:**
  - Farmers who receive fertilizer may also have access to better technology or irrigation systems (Z).
  - Improved technology directly affects crop yield (Y), creating a spurious correlation between X and Y.

Why These Factors Matter:

- Ignoring these confounders leads to biased estimates of the fertilizer's effect.
- The observed increase in crop yield may not solely result from the fertilizer.

What is Econometrics?

- Econometrics combines economics, mathematics, and statistics.
- It aims to answer questions like:
  - Does an increase in education lead to higher earnings? (Causality)
  - How do changes in policy affect economic outcomes? (Policy Evaluation)
  - What factors predict a country's GDP growth? (Prediction)
- Ideally, we would like an Randomized Controlled Experiment, but almost always we only have observational (non-experimental) data.
- Issue to estimate causal effects with non-experimental data:

- confounding effects (omitted factors)
- selection-bias
- simultaneous causality
- "correlation does not imply causation"

Objective of this Course

Key Questions:

- Correlation vs Causation: How to determine causal relationships.
- Endogeneity: When explanatory variables are correlated with the error term.
- Model Selection: Choosing the appropriate model for analysis.
- Interpretation: Translating results into meaningful economic insights.

Topics

- Ordinary Least Squares
- Issues with OLS: Omitted Variables, Heteroskedasticity, Simultaneity
- Instrumental Variables
- Bianary Dependent Variables (Logistic Regression) and Poisson Regression
- Generalized Method of Moments
- Panel Data Methods
- Treatment Effects (Difference in Difference and Regression Discontinuity)

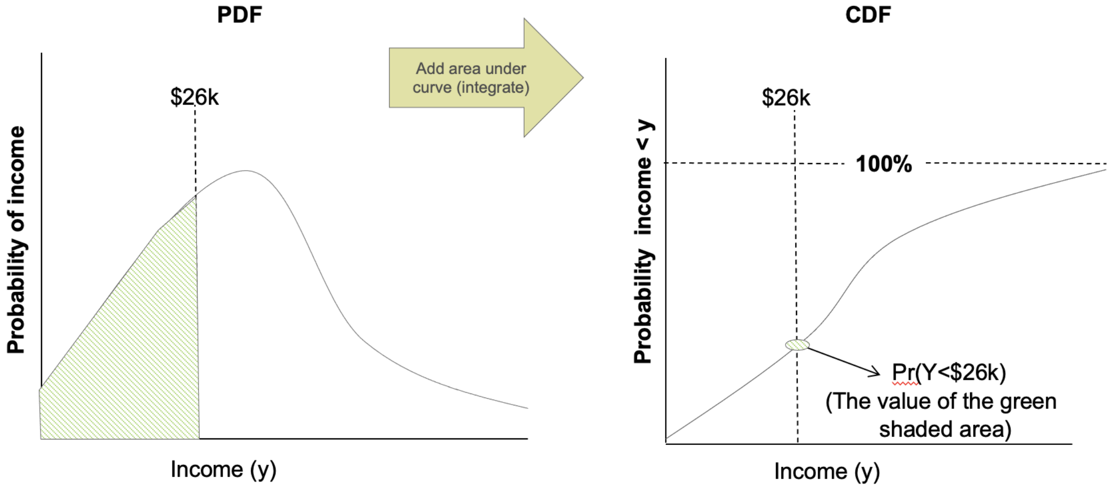
Review of Probability

- Random Variables and Distributions
- Expected Value and its Properties
- Variance, Covariance and Correlation
- Joint Distribution, Conditional Distribution
- Conditional Expectation

Random Variables

- A **random variable** is a variable whose value is determined by a random process.
- Types of random variables:
  - **Discrete**: Can take only discrete values (e.g., number of students in a class).
  - **Continuous**: Can take any value in a range (e.g., height, weight).
- Examples:
  - Discrete: Number of heads in 10 coin flips.
  - Continuous: Temperature in a city over a day.
- The Probability that a random variable X takes a specific value $(X = x)$ is defined by its probability distribution function (PDF) $f_X(x)$.
- The probability that a random variable X takes values below x is given by the Cumulative Probability Density Function $F_X(x)$ which corresponds to the area under the PDF.

Let Y be a random variable that represents the per-capita income



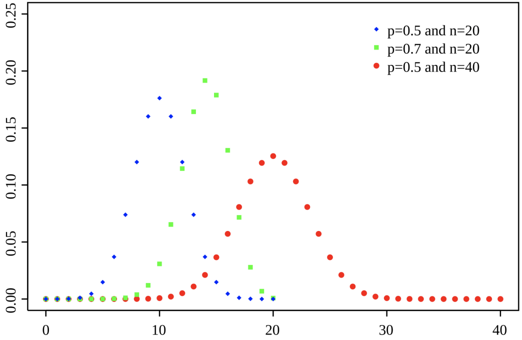
Discrete Distribution

- Example: the Binomial distribution models the number of successes (k) in a fixed number of independent Bernoulli trials (n), each with a probability p of success. For example: The number of heads when flipping 3 coins.
- The probability mass function (PMF) is:

$$P(X = k) = \binom{n}{k} p^k (1-p)^{n-k}, \quad k = 0, 1, \ldots, n$$

- **CDF**: The CDF is the cumulative sum of the PMF:

$$F_X(k) = P(X \leq k) = \sum_{i=0}^{k} P(X = i).$$



Continuous Distribution

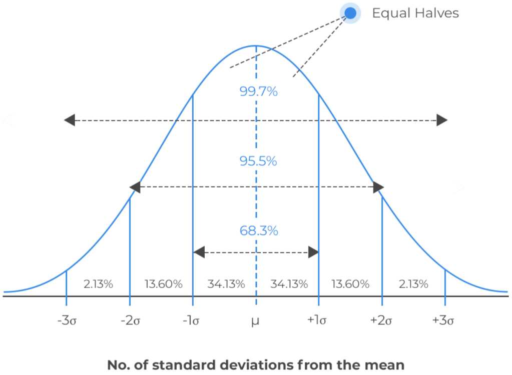
- The Normal distribution is commonly used to model natural phenomena, such as income or test scores, where data tends to cluster around a central value. **Example**: The income of HH in Boston area has mean $\mu = 95k$ and standard deviation $\sigma = 8k$.
- The probability density function (PDF) is:

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

- **CDF**: The CDF of the normal distribution is given by:

$$F_X(x) = \int_{-\infty}^{x} f_X(t)\,dt.$$

and represents the area under f_X for $X \leq x$.



Expectated Value and Its Properties

- Discrete:

$$E[X] = \sum xP(x)$$

- Continuous:

$$E[X] = \int xf(x)dx$$

- **Properties of Expectation:**
  - Linearity: $E(aX + b) = aE(X) + b$
  - Additivity: $E(X + Y) = E(X) + E(Y)$
  - For independent random variables X and Y, $E(XY) = E(X)E(Y)$
- **Expectation of a Function:** For a function (non random) $g(X)$, the expectation is:

$$E(g(X)) = \int_{-\infty}^{\infty} g(x)f_X(x)\,dx \quad \text{(continuous)}$$

$$E(g(X)) = \sum_x g(x)P(X = x) \quad \text{(discrete)}$$

Probability as an Expectation

- **Indicator Function:** Define the indicator function I_A for an event A, where:

$$I_A = \begin{cases} 1 & \text{if event } A \text{ occurs} \\ 0 & \text{otherwise} \end{cases}$$

- **Expectation of Indicator Function:** The expected value of the indicator function I_A is:

$$E(I_A) = P(A)$$

This means the expectation of an indicator function is equal to the probability of the event occurring.

Variance

- **Variance** measures the spread of a random variable around its mean:

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

- **Properties of Variance:**
  - For any constant a and random variable X, $\text{Var}(aX + b) = a^2\text{Var}(X)$
  - For independent random variables X and Y, $\text{Var}(X + Y) = \text{Var}(X) + \text{Var}(Y)$
  - For non-independent random variables X and Y, $\text{Var}(X + Y) = \text{Var}(X) + \text{Var}(Y) + 2\text{Cov}(X, Y)$
- **Standard Deviation**: $\sigma_X = \sqrt{Var(X)}$
- **Coefficient of Variation**:

$$CV = \frac{\sigma_X}{\mu_X}$$

Example with a discrete random variable

- **Example**: Let X be a discrete random variable with the following probability mass function (PMF):

$$P(X = 0) = 0.2, \quad P(X = 1) = 0.5, \quad P(X = 2) = 0.3.$$

- **Mean (Expected Value)**: The mean is given by $E[X] = \sum_x x \cdot P(X = x)$ For this example:

$$E[X] = 0 \cdot 0.2 + 1 \cdot 0.5 + 2 \cdot 0.3 = 0 + 0.5 + 0.6 = 1.1.$$

- **Variance**: The variance is given by $\text{Var}(X) = E[X^2] - (E[X])^2$ where $E[X^2]$ is the expected value of the square of X:

$$E[X^2] = \sum_x x^2 \cdot P(X = x).$$

For this example:

$$E[X^2] = 0^2 \cdot 0.2 + 1^2 \cdot 0.5 + 2^2 \cdot 0.3 = 0 + 0.5 + 1.2 = 1.7.$$

Now, we can calculate the variance:

$$\text{Var}(X) = 1.7 - (1.1)^2 = 1.7 - 1.21 = 0.49.$$

Covariance

Covariance measures the joint variability of two random variables X and Y:

$$\text{Cov}(X, Y) = E[(X - E(X))(Y - E(Y))] = E(XY) - E(X)E(Y)$$

Properties of Covariance:

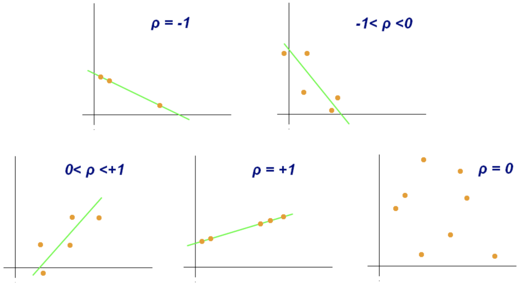
- Positive covariance indicates that X and Y tend to move in the same direction.
- Negative covariance indicates that X and Y tend to move in opposite directions.
- If X and Y are independent, $\text{Cov}(X, Y) = 0$, but $\text{Cov}(X, Y) = 0$ does not imply independence.
- $Cov(X, a) = 0$
- $Cov(X, X) = Var(X)$
- $Cov(X, Y) = Cov(Y, X)$
- $Cov(aX, bY) = abCov(X, Y)$
- $Cov(aX + bY, cW + dZ) = acCov(X, W) + adCov(X, Z) + bcCov(Y, W) + bdCov(Y, Z)$
- $Cov(X, Y) = 0 \rightarrow E[XY] = E[X]E[Y]$

Correlation

- **Correlation** is the normalized measure of the linear relationship between X and Y:

$$\mathrm{Corr}(X, Y) = \frac{\mathrm{Cov}(X, Y)}{\sqrt{\mathrm{Var}(X)\mathrm{Var}(Y)}} \in [-1, 1]$$

- Interpretation:
  - $\rho > 0$: Positive linear relationship.
  - $\rho < 0$: Negative linear relationship.
  - $\rho = 0$: No linear relationship.



Joint Distributions

- **Joint PMF (Probability Mass Function)**: The joint PMF $p(x, y)$ gives the probability that two discrete random variables (X, Y)take specific values (x, y) simultaneously.
- **Example (Discrete Case)**: Consider the following joint PMF for X and Y:

$$p(x, y) = \begin{cases} 0.2 & \text{if } (x, y) = (1, 1) \\ 0.3 & \text{if } (x, y) = (1, 2) \\ 0.1 & \text{if } (x, y) = (2, 1) \\ 0.4 & \text{if } (x, y) = (2, 2) \end{cases}$$

- **Finding the Marginal PMF of X**: To find $p_X(1)$, sum the joint probabilities for all values of y where $x = 1$:

$$p_X(1) = p(1, 1) + p(1, 2) = 0.2 + 0.3 = 0.5$$

  Similarly, for $x = 2$:

$$p_X(2) = p(2, 1) + p(2, 2) = 0.1 + 0.4 = 0.5$$

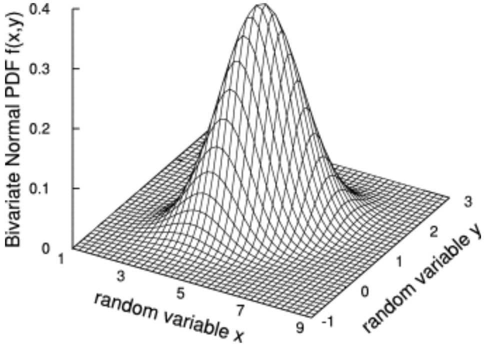
Bivariate Normal Distribution

The joint pdf of a bivariate normal random variable (X, Y)

$$f(x,y) = \frac{1}{2\pi\sigma_X\sigma_Y\sqrt{1-\rho^2}} \exp\left(-\frac{1}{2(1-\rho^2)}\left[\left(\frac{x-\mu_x}{\sigma_X}\right)^2 - 2\rho\frac{(x-\mu_x)(y-\mu_y)}{\sigma_x\sigma_y} + \left(\frac{y-\mu_Y}{\sigma_Y}\right)^2\right]\right)$$

where:

- μ_X, μ_Y are the means of X and Y,
- σ_X, σ_Y are the standard deviations of X and Y,
- ρ is the correlation coefficient between X and Y,
- $f(x, y)$ is the joint PDF of X and Y.



Conditional Distribution

- **Conditional PDF**: The conditional probability density function (PDF) of a continuous random variable Y given $X = x$ is defined as:

$$f_{Y|X}(y|x) = \frac{\text{Joint pdf of (X,Y)}}{\text{Marginal pdf of X}} = \frac{f_{X,Y}(x,y)}{f_X(x)}$$

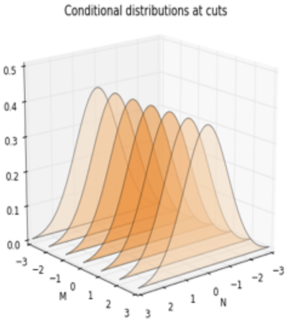
- **Conditional Normal Distribution of Y given $X = x$:**

$$\mu_{Y|X} = E[Y|X = x] = \mu_Y + \rho\frac{\sigma_Y}{\sigma_X}(x - \mu_X)$$

$$\sigma^2_{Y|X} = \text{Var}(Y|X = x) = \sigma^2_Y(1 - \rho^2)$$

Thus, the conditional PDF of Y given $X = x$ is:

$$f_{Y|X}(y|x) = \frac{1}{\sqrt{2\pi\sigma^2_{Y|X}}} \exp\left(-\frac{(y - \mu_{Y|X})^2}{2\sigma^2_{Y|X}}\right)$$



Conditional Expectation

- **Definition of Conditional Expectation:** The conditional expectation is the expected value of a random variable Y given another random variable $X = x$ is:

$$E(Y|X = x) = \sum_{y} yPr(Y = y|X = x)$$

  for discrete random variables, or

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

  for continuous random variables.

- **Conditional Expectation as a Random Variable:**

$$h(x) = E(Y|X = x)$$

  As x changes, the conditional distribution of Y given $X = x$ typically changes as well, and so might the conditional expectation of Y given $X = x$. So we can view $E[Y|X = x]$ as a function of x.

Conditional Expectation: Example (Continued)

- By definition of conditional expectation:

$$E[Y|X=1] = 1 \cdot Pr(Y=1|X=1) + 2 \cdot Pr(Y=2|X=1)$$

- First, compute the conditional PMF $Pr(Y=y|X=1)$ for $X=1$:

$$Pr(Y=1, X=1|X=1) = \frac{Pr(Y=1, X=1)}{Pr(X=1)} = \frac{0.2}{0.5} = 0.4$$

$$Pr(Y=2, X=1|X=1) = \frac{Pr(Y=2, X=1)}{Pr(X=1)} = \frac{0.3}{0.5} = 0.6$$

Now, compute the conditional expectation $E[Y|X=1]$:

$$E[Y|X=1] = 1 \cdot Pr(Y=1|X=1) + 2 \cdot Pr(Y=2|X=1) = 1 \cdot 0.4 + 2 \cdot 0.6 = 1.6$$

Properties of Conditional Expectation

- **Linearity:** Conditional expectation is linear. For random variables X and Y, and constants $a, b \in \mathbb{R}$:
$$E[aX + bY|Z] = aE[X|Z] + bE[Y|Z]$$
- **Taking out what is known:** For any non-random function $h(\cdot)$:
$$E[h(Z)X|Z] = h(Z)E[X|Z]$$
- **Independence:** If X and Y are independent, then:
$$E[Y|X] = E[Y]$$
This means knowing X provides no additional information about Y.
- **Law of Iterated Expectations (Adam's Law):** The expectation of a conditional expectation equals the unconditional expectation:
$$E_Y\left[E_X\left[X|Y\right]\right] = E_X[X]$$

Properties of Conditional Expectation

- **Projection Theorem:** For any non-random function $h : \mathcal{X} \to \mathbb{R}$:

$$E[(Y - E[Y|X])h(X)] = 0 \quad \text{for any function } h(X)$$

  This shows that the residual $Y - E[Y|X]$ is uncorrelated with any function of X.

- **Conditional Expectation is the Best Predictor:** $E[Y|X]$ is the best predictor of Y given X in terms of minimizing mean squared error (MSE):

$$E[Y|X] = \arg\min_h \underbrace{E[(Y - h(X))^2]}_{\text{MSE}}$$

- **Decomposition**:

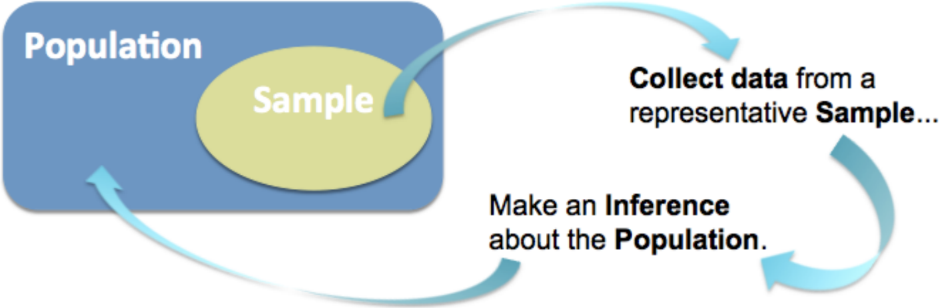
$$Y = \underbrace{E[Y|X]}_{\text{best prediction}} + \underbrace{Y - E[Y|X]}_{\text{residual}}$$

  Note that the two terms on the RHS are uncorrelated, by the projection theorem.

Review of Statistical Inference

- Sample Statistics
- Law of Large Numbers, Central Limit Theorem
- Hypothesis tests
- Confidence Intervals.

Statistical Inference



Population and Sample

- We are analyzing the relationship between the number of hours studied X and the test scores Y.
- From the students population we collect a sample

Student	Hours Studied (X)	Test Score (Y)
1	2	50
2	3	60
3	5	75
4	6	80
5	8	90

- the population variable X and Y have some distribution f_X and f_Y, and they are related to each other according to some joint distribution $f_{X,Y}$.
- We want to use the sample to infer the value of some population parameter: mean, variance, correlation.

Sample Statistics

- **Sample Average (Mean)** of X to estimate the population mean μ_X :

$$\bar{X} = \frac{1}{n}\sum_{i=1}^{n} X_i = \frac{2+3+5+6+8}{5} = \frac{24}{5} = 4.8$$

- **Sample Variance** of X to estimate the population variance σ_X^2:

$$S_X^2 = \frac{1}{n-1}\sum_{i=1}^{n}(X_i - \bar{X})^2 = \frac{(2-4.8)^2 + (3-4.8)^2 + (5-4.8)^2 + (6-4.8)^2 + (8-4.8)^2}{4} = 5.04$$

- **Sample Correlation** between X and Y to estimate the correlation $\rho_{X,Y}$:

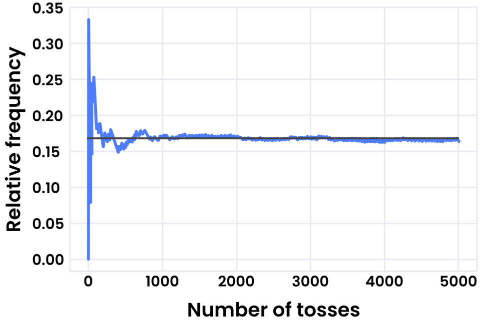
$$r_{XY} = \frac{Cov(X,Y)}{\sqrt{S_X^2}\sqrt{S_Y^2}} = \frac{\frac{1}{n}\sum_{i=1}^{n}(X_i - \bar{X})(Y_i - \bar{Y})}{\sqrt{\frac{1}{n-1}\sum_{i=1}^{n}(X_i - \bar{X})^2 \frac{1}{n-1}\sum_{i=1}^{n}(Y_i - \bar{Y})^2}} = 0.874$$

The Law of Large Numbers (LLN)

- **Definition:** Given an i.i.d. sample $(X_1, X_2, ..., X_n)$ with $\mathbb{E}[X_i] = \mu$ and $Var[X_i] = \sigma^2 < \infty$, the sample mean **converges in probability** to the expected value μ

$$\lim_{n\to\infty} \frac{1}{n} \sum_{i=1}^{n} X_i \xrightarrow{p} \mu$$

- **Intuition**: Suppose that we are interested in the probability of obtaining 6 when rolling a dice. Let $X_i = 1$ if we get 6 and $X_i = 0$ otherwise. Consider the sample mean $\overline{X} = \frac{1}{n}\sum_{i=1}^{n} X_i$. If you perform an experiment (roll a dice) with a large number of trials, the value of $\overline{X}$ converges to the expected value of X_i (which is $\frac{1}{6}$).



You can see from the figure that after approximately 1500 throws, the blue relative frequency has stabilized very close to the actual probability in black.

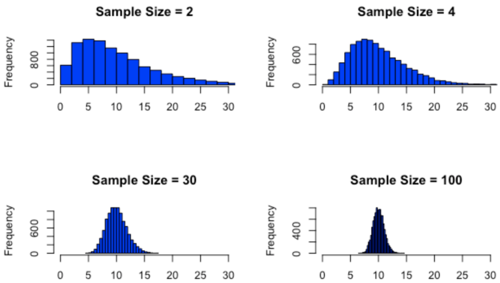
The Central Limit Theorem (CLT)

- **Definition:** Let $(X_1, X_2, ..., X_n)$ be an i.i.d. sample of size n from a population with with $\mathbb{E}[X_i] = \mu$ and $Var[X_i] = \sigma^2 < \infty$. The Central Limit Theorem states that the distribution of the sample mean $\bar{X}_n$ will approach a Normal distribution as the sample size n increases, regardless of the original population's distribution. In other words the sample mean **converges in distribution** to a Normal distribution

$$\bar{X}_n \xrightarrow{d} \mathcal{N}\left(\mu, \frac{\sigma^2}{n}\right) \quad \text{as} \quad n \to \infty$$

or equivalently

$$Z = \frac{\bar{X}_n - \mu}{\sigma/\sqrt{n}} \xrightarrow{d} \mathcal{N}(0,1) \quad \text{as} \quad n \to \infty$$



Hypothesis Testing

- A **statistical hypothesis** is a claim (null hypothesis H_0) about the value of a population parameter.
- The objective of hypothesis testing is to decide, based on sample information, if the alternative hypotheses is actually supported by the data.
- The burden of proof is placed on those who believe in the alternative claim (alternative hypothesis H_a). In other words the null hypothesis H_0 is assumed to be true.
- This initially favored claim (H_0) is rejected in favor of the alternative claim (H_a) if the sample evidence provides significant support for the alternative assertion.
- If the sample does not strongly contradict H_0, we continue to believe in the plausibility of the null hypothesis.
- The two possible conclusions:
  1) Reject H_0.
  2) Fail to reject H_0.

Hypothesis Testing Procedure

A statistical hypothesis test is a method of statistical inference used to decide whether the data sufficiently supports a particular hypothesis. The general steps in hypothesis testing are:

1. **State the hypotheses**:
   - Null hypothesis (H_0): assumed to be true
   - Alternative hypothesis (H_a): invalidates the null hypothesis
2. **Choose the significance level (α)**:
   - The significance level (α) represents the probability of rejecting the null hypothesis when it is true. A commonly used value is 0.05.
3. **Compute the test statistic**:
   - The test statistic z is a numerical value calculated from the sample data that measures the degree of agreement between the null hypothesis and the sample data.
4. **Compare with the critical value**: If the test statistic exceeds the critical value, reject H_0, otherwise, fail to reject H_0.

Hypothesis Test Outcomes

	Decision	
	Retain H_0	Reject H_0
H_0 true	$\surd$	Type I error (false positive)
H_1 true	Type II error (false negative)	$\surd$

Errors in Hypothesis Testing:

- A type I error is when H_0 is rejected, but it is true. Let $\alpha = \Pr(\text{reject } H_0 \mid H_0 \text{ is true})$
- A type II error is not rejecting H_0 when H_0 is false. Let $\beta = \Pr(\text{fail to reject } H_0 \mid H_0 \text{ is false})$

Size and Power of a Test:

- If $\alpha \downarrow$ then $\beta \uparrow$, and viceversa. In other words, No rejection region can be changed to simultaneously make both α and β smaller.
- α is also called *significant level* of a test. Typical levels are .10, .05, and .01.
- $\pi = 1 - \beta$ is called *power* of a test

Alternative Hypothesis and Rejection Region

Given the Null hypothesis:

$$H_0 : \mu = \mu_0$$

and the i.i.d. sample $(x_1, x_2, ..., x_n)$, compute the test statistics z. Reject H_0 if the test statistics z falls in the Rejection Region.

$$z \in \text{Rejection Region}$$

The size and shape of the Rejection Region is determined by the alternative hypothesis H_a and the significance level α.

Alternative Hypothesis:

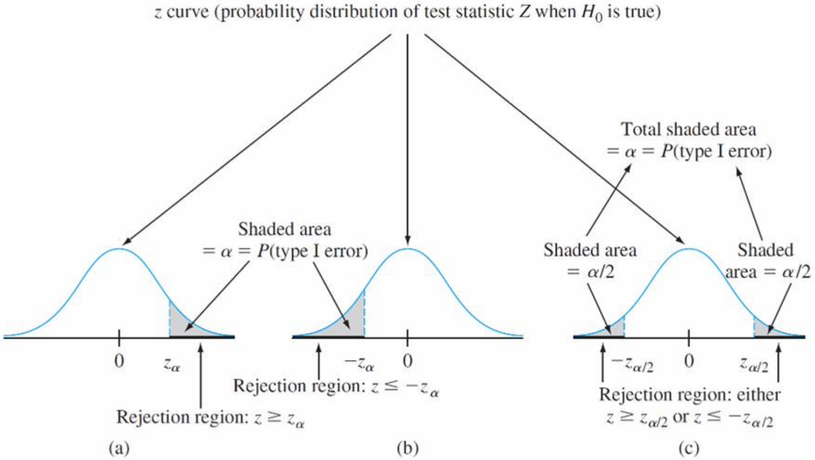
- $H_a : \mu > \mu_0$
- $H_a : \mu < \mu_0$
- $H_a : \mu \neq \mu_0$

Rejection Region for Level α Test:

- $z \geq z_\alpha$ (upper tailed test)
- $z \leq -z_\alpha$ (lower tailed test)
- $z \leq -z_{\alpha/2}$ or $z \geq z_{\alpha/2}$ (two tailed test)

Rejection Regions

The test statistics z is a function of the sample (which is a set of random variables) and therefore is itself a random variable with some distribution.



Example

- An inventor has developed a new, energy-efficient lawn mower engine.
- The leading brand lawnmower engine runs for 300 minutes on 1 gallon of gasoline
- He claims that the engine will run continuously for more than 5 hours (300 minutes) on a single gallon of regular gasoline.

$$H_0 : \mu = 300\text{min} \quad vs. \quad H_a : \mu > 300\text{min}$$

- From his stock of engines, the inventor selects a simple random sample of 50 engines for testing.
- The engines run for an average of 305 minutes $\rightarrow \overline{X} = 305$. Suppose that the standard deviation is known ($\sigma = 30$). Suppose that the run times of the engines are normally distributed

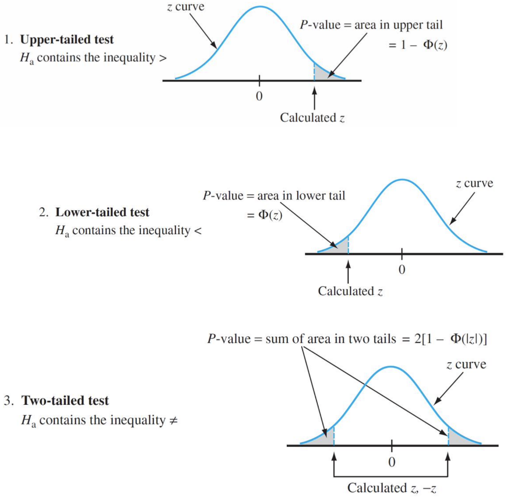
$$z = \frac{\overline{X} - \mu_0}{\sigma/\sqrt{n}} \sim N(0, 1)$$

- Objective: Test hypothesis that the mean run time is more than 300 minutes. Use a 0.05 level of significance. Reject $\mu = 300min$ if

$$z = \frac{305 - 300}{30/\sqrt{50}} > z_{0.05}$$

The p-value of a Test

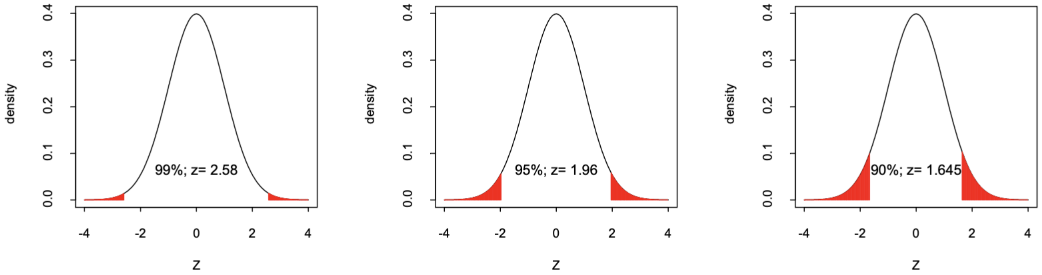
- **p-value**: The probability of observing a test statistic at least as extreme as the one computed, under the assumption that H_0 is true.
- **Interpretation**: p-value in the area under the pdf above the above (if upper tailed) or below (if lower tailed) the observed value of the test-statistic.
- **Decision Rule** can also be written as:
  - p-value $< \alpha$: Reject H_0.
  - p-value $\geq \alpha$: Fail to reject H_0.
- The p-value can be thought of as the smallest significance level at which H_0 can be rejected.
- So, the smaller the p-value, the more evidence there is in the sample data against the null hypothesis and in favor of the alternative hypothesis.



Confidence Intervals

- Recall that, by the CLT, the sample mean, $\overline{X}$, can be regarded as being normally distributed with mean μ and standard deviation $\frac{\sigma}{\sqrt{n}}$ when n is large enough.
- **Confidence Interval (CI)**: A range of values where the true parameter is expected to lie with a certain probability.
$$CI = \text{Point Estimate} \pm \text{Margin of Error}.$$
- Suppose we want to construct a $100(1-\alpha)\%$ CI for the mean run time. This is equivalent to finding a value δ such that
$$\Pr(|\overline{X} - \mu| \leq \delta) = 1 - \alpha.$$
which is equivalent to
$$\Pr(Z < \frac{\delta}{\sigma/\sqrt{n}}) - \Pr(Z < -\frac{\delta}{\sigma/\sqrt{n}}) = 1 - \alpha.$$
where $Z = \frac{\overline{X}-\mu}{\sigma/\sqrt{n}}$ is the standard normal distribution.
- The point that satisfies this condition is denoted as $z_{1-\alpha/2} \rightarrow \delta = z_{1-\alpha/2}\frac{\sigma}{\sqrt{n}}$

$z_{1-\alpha/2}$ is the $1-\alpha/2$th quantile of Z. Because Z is symmetric about 0, we immediately know that $1-\alpha/2$ of the area under the standard normal curve lies to the left of - $z_{1-\alpha/2}$, and $1-\alpha/2$ of the area under the standard normal curve lies to the right of $z_{1-\alpha/2}$ (For example, if $1-\alpha=95\%$, then $-z_{0.975}=z_{0.025}$)

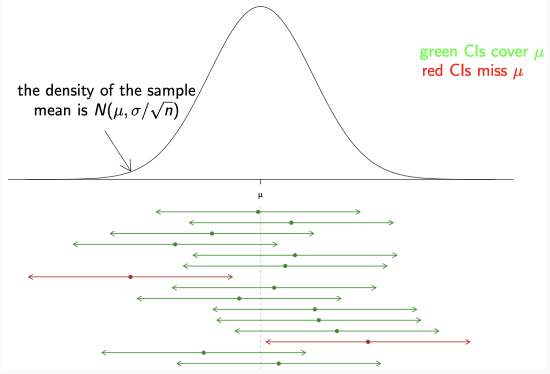


and the CI is given by

$$\left[\overline{X}-z_{1-\alpha/2}\frac{\sigma}{\sqrt{n}},\overline{X}+z_{1-\alpha/2}\frac{\sigma}{\sqrt{n}}\right]$$

Confidence Levels

Suppose that $100(1-\alpha)\% = 95\%$. What is the thing that has a 95% chance to happen?



About 95% of the intervals constructed following the procedure (taking a SRS and then calculating $\bar{X} \pm Z_{\alpha/2}\frac{\sigma}{\sqrt{n}}$) will cover the true population mean μ.

Remarks

Notice that

- $$n = \left(\frac{z_{1-\alpha/2}}{\delta}\right)^2 \sigma^2$$

  thus the sample size n must increase if:
  - δ decreases (i.e. we require greater precision); or
  - σ increases (i.e. there is more dispersion in the population); or
  - α decreases (i.e. we require greater accuracy).
- if σ is unknown, we need to estimate it using the the sample std S. The procedure to construct the CI is the same, but we nee to replace σ with S, and

  $$T = \frac{\overline{X} - \mu}{S/\sqrt{n}}$$

  no longer has a standard normal distribution, but rather a Student's t distribution with $(n-1)$ degrees of freedom.