Statistics

Why study Statistics?

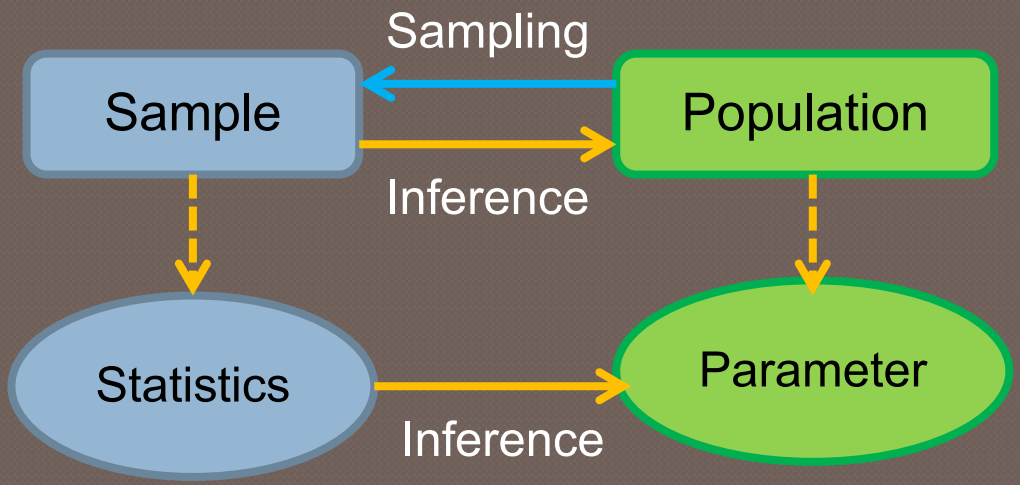
- Statistics is the science of collecting, analyzing and making inference from data.
- Statistical methods and analyses are often used to communicate research findings and to support hypotheses and give credibility to research methodology and conclusions.

Types of Statistics

- Descriptive Statistics
  - A simple quantitative summary of a data set that has been collected.
  - Mean ($\bar{x}$), Median (Me), Mode (Mo), Standard Deviation (S.D.), Variance (s^2), Maximum, Minimum, frequency, proportion, Spearman Rank Correlation (ordinal), Pearson Correlation (interval/ratio), and etc.

Types of Statistics

- Inferential Statistics
  - A study to apply the conclusions that have been obtained from one study to more general populations.

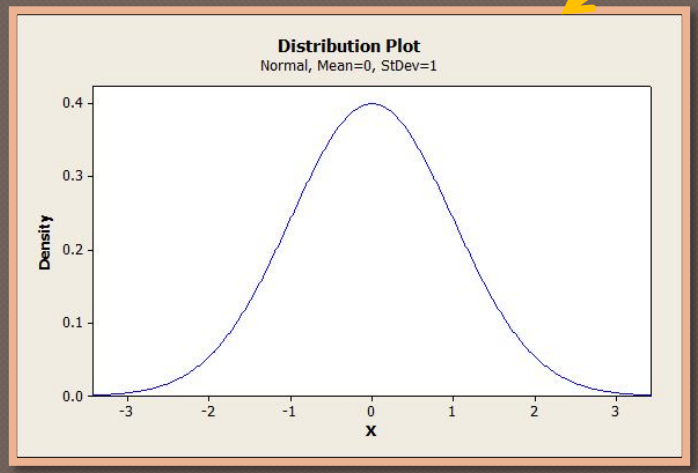


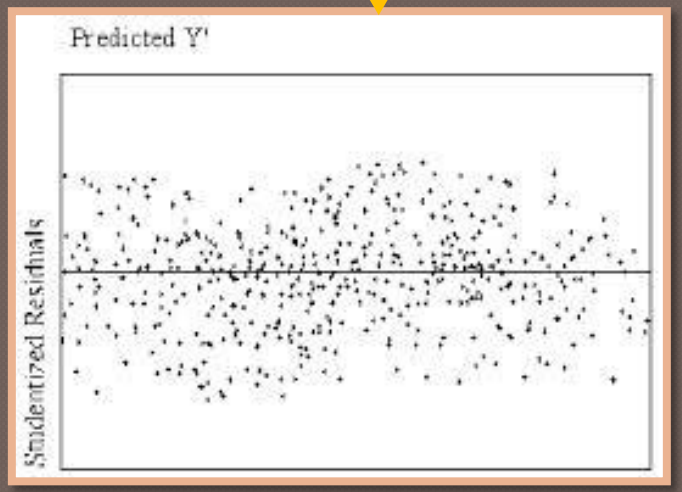
Types of Statistics

	Statistics	Population
Mean	$\bar{x}$	μ
Standard Deviation	S.D.	σ
Variance	S^2	σ^2

Statistical Tests

- Parametric Tests
  - Assuming that the data has come from a type of probability distribution and making inferences about the parameters of the distribution (Ex. Prerequisites: Normality, Homoscedasticity of variance)





Statistical Tests

- Non-Parametric Test
  - The tests do not require any specific form for the distribution of population.
  - The tests do not make inferences about the parameters of the distribution

Statistical tests

- Parametric Tests
  - ✓ Chi-square (χ^2) test
  - ✓ t-test
  - ✓ ANOVA, ANCOVA, MANOVA
  - ✓ Regression
  - ✓ Cluster Analysis
  - ✓ Discriminant Analysis
  - ✓ Logistic Regression Analysis
  - ✓ etc.
- Non-Parametric Tests
  - ✓ Fisher's exact test
  - ✓ Man-Whitney U test
  - ✓ Wilcoxon's rank sum test
  - ✓ Kruskall-Wallis H test
  - ✓ Wilcoxon's signed rank test
  - ✓ Friedman's ANOVA
  - ✓ etc.

How to choose the right statistical test?

1) Compare Means

Analysis Type	Parametric	Non-Parametric
Compare means between two distinct/independent groups	T-test (interval/ratio)	Wilcoxon rank-sum test
Compare two quantitative measurements taken from the same individual	Paired t-test (interval/ratio)	Wilcoxon signed-rank test

How to choose the right statistical test?

1) Compare Means

Analysis Type	Parametric	Non-Parametric
Compare means between three or more distinct/independent groups	ANOVA ANCOVA MANOVA (dependent v. = interval/ratio Independent v. = at least 1 nominal)	Kruskal-Wallis test
Estimate the degree of association between two quantitative variables	Pearson correlation (interval/ratio)	Spearman's rank correlation

How to choose the right statistical test?

2) Explanation/Prediction of Relationships

Scale of Dependent variable	Scale of Independent variable	Test Statistics
Nominal	Nominal	Chi-square test
Interval/Ratio	All	Simple Regression Multiple Regression
Dichtomous	All	Probit/Logit
Polytomous	All	Logistic

How to choose the right statistical test?

3) Structure Analysis of Variables

Scale of Variables	Test Statistics
All	✓Cluster Analysis ✓Component Analysis ✓Exploratory Factor Analysis ✓Confirmatory Factor Analysis

T-test

- Compare means between two groups
- H_o: $\mu_1 = \mu_2$

  H_A: $\mu_1 \neq \mu_2$

$$t = \frac{\bar{x}_1 - \bar{x}_2}{\sqrt{\frac{{s_1}^2}{n_1} + \frac{{s_2}^2}{n_2}}}$$

ANOVA

- Compare means among many groups (> 2 groups)
- Assumptions: $e_{ij} \sim N(0, \sigma^2)$, random sampling
- $H_o: \mu_1 = \mu_2 = \mu_i$,
  $H_A: \mu_1 \neq \mu_2 \neq \mu_i$

THE LOGIC OF ANOVA (ANALYSIS OF VARIANCE)

$$F = \frac{\text{variance of the group means}}{\text{mean of the within-group variances}}$$

$$F^* = \frac{\text{MSTR}}{\text{MSE}}$$

where

$$\text{MSTR} = \frac{\text{SSTR}}{I - 1}$$, I = number of treatments

and

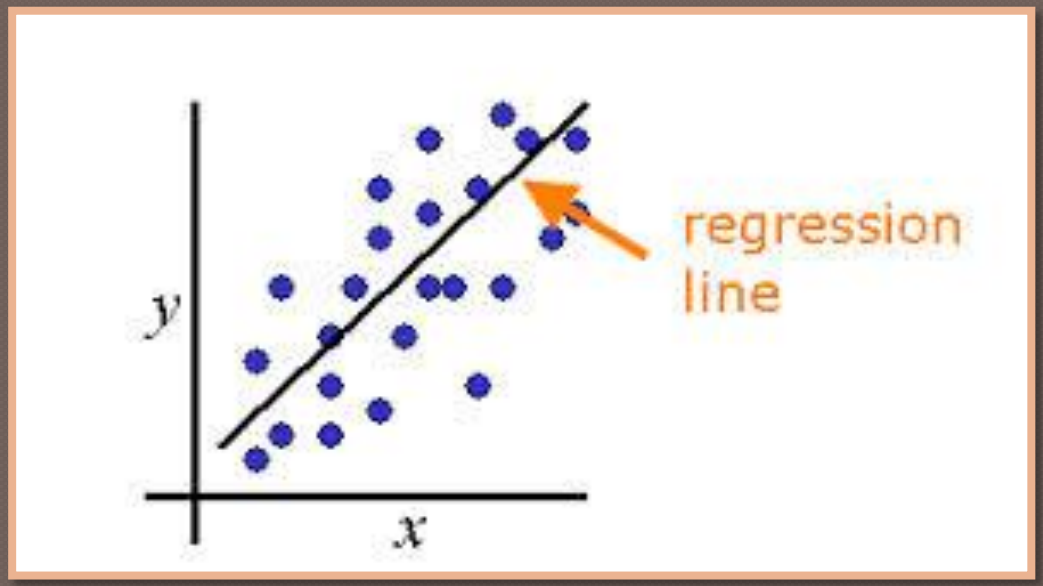
$$\text{MSE} = \frac{\text{SSE}}{n_T - I}$$, n_T = total number of cases

Regression Analysis

- Test causal relationship
- <u>Assumptions:</u> 1) Linearity 2) No perfect/high multicollinearity 3) No autocorrelation 4) $e_{ij} \sim N(0, \sigma^2)$ 5) no outlier
- <u>Equation:</u> Simple regression: $Y = a + bx + e$

  Multiple regression: $Y = a + b_1x_1 + b_ix_i + e$

Regression Analysis



$$y = a + bx$$

$$a = \frac{\sum y - b\sum x}{n}$$

$$b = \frac{n\sum(xy) - (\sum x)(\sum y)}{n\sum x^2 - (\sum x)^2}$$

x	=	The current time period
n	=	The total number of time periods


Q&A