Quantitative Methods in Social Research

Burak Sonmez

UCL

Feb 18, 2026 (UPF)

Day 1 – Outline

1. Introduction
2. Course Structure, Objectives, and Learning Outcomes
3. Drawing Inferences
4. Research Design and Formulation
5. Measurement, Descriptive Statistics, Sampling & Distributions
6. Hypothesis Testing

Day 1 – Outline

1. Introduction
2. Course Structure, Objectives, and Learning Outcomes
3. Drawing Inferences
4. Research Design and Formulation
5. Measurement, Descriptive Statistics, Sampling & Distributions
6. Hypothesis Testing

Who am I?

- Burak Sonmez
- Associate Professor in Quantitative Social Science
- Interested in social norms; trust; collective actions; experimental methods; computational social science
    - *University College London (2021-present), United Kingdom*
    - *London School of Economics (2020-2021), United Kingdom*
    - *University of Essex (2013-2020), United Kingdom*
    - *Bilkent University (2008-2012), Turkey*
- **Email:b.sonmez@ucl.ac.uk**
  Bluesky:@buraksonmez.bsky.social

Let us know each other

- Your name and educational background

Let us know each other

- Your name and educational background
- Research interests (very briefly)

Let us know each other

- Your name and educational background
- Research interests (very briefly)
- What motivates you to take this course?

Let us know each other

- Your name and educational background
- Research interests (very briefly)
- What motivates you to take this course?

Day 1 – Outline

Structure, Objectives, and Learning Outcomes

- Introducing social science researchers to and help them become more confident in quantitative research methods using R
- By the end of the course, students should be able to understand the introductory principles and practices of quantitative research methods, apply them to their research questions and evaluate their use in published research. Students should also be able to acquire a working knowledge of performing such statistical analyses using R.

Day 1 – Outline

The problem of inference

- "Inference is the process of using the facts we know to learn about facts we do not know." (King et al. 1994)
- "Statistical inference is used to learn from incomplete or imperfect data." (Gelman & Hill, 2007)

The problem of inference

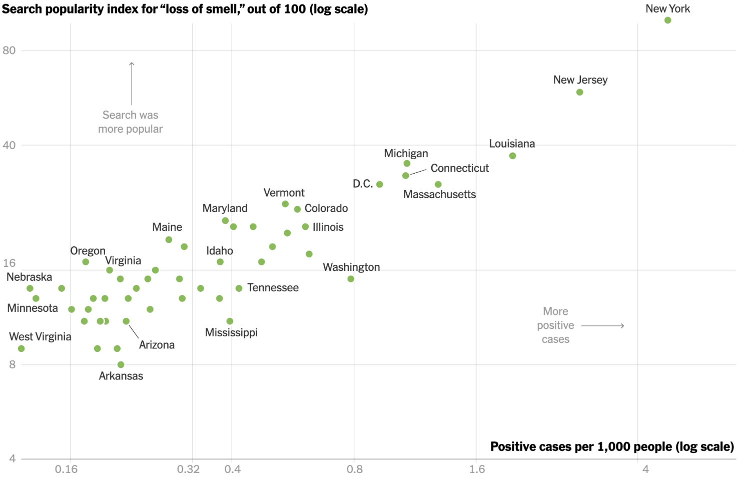
- "Inference is the process of using the facts we know to learn about facts we do not know." (King et al. 1994)
- "Statistical inference is used to learn from incomplete or imperfect data." (Gelman & Hill, 2007)
- Descriptive inference is an inference we make about how the world is/was
- We are interested in learning some characteristics of a population (for example, the mean and standard deviation of the heights of all women in the UK), which we must estimate from a sample, or subset, of that population (**sampling model**)
- We are interested in learning aspects of some underlying pattern or law (for example, $y = \alpha + \beta x + \epsilon$)

The problem of inference

- "Inference is the process of using the facts we know to learn about facts we do not know." (King et al. 1994)
- "Statistical inference is used to learn from incomplete or imperfect data." (Gelman & Hill, 2007)
- Descriptive inference is an inference we make about how the world is/was
- We are interested in learning some characteristics of a population (for example, the mean and standard deviation of the heights of all women in the UK), which we must estimate from a sample, or subset, of that population (**sampling model**)
- We are interested in learning aspects of some underlying pattern or law (for example, $y = \alpha + \beta x + \epsilon$)
- Causal inference is an inference we make about why something happens

Drawing Inferences

Evidence ⟶ Processing Filter ⟶ Conclusions ⟶ Claims



Source: Google | By The New York Times

Descriptive Inference vs Causal Inference

- Does increasing COVID-19 cases explain search popularity for loss of smell on Google?

Descriptive Inference vs Causal Inference

- Does increasing COVID-19 cases explain search popularity for loss of smell on Google?
- Why do people search for loss of taste and smell more in certain states than others?

Day 1 – Outline

1. Introduction
2. Course Structure, Objectives, and Learning Outcomes
3. Drawing Inferences
4. Research Design and Formulation
5. Measurement, Descriptive Statistics, Sampling & Distributions
6. Hypothesis Testing

Fundamentals

- Formulation of research questions

Fundamentals

- Formulation of research questions
- Conceptualisation

Fundamentals

- Formulation of research questions
- Conceptualisation
- Operationalisation

Fundamentals

- Formulation of research questions
- Conceptualisation
- Operationalisation
- Measurement

Fundamentals

- Formulation of research questions
- Conceptualisation
- Operationalisation
- Measurement
- Hypothesising

Fundamentals

- Formulation of research questions
- Conceptualisation
- Operationalisation
- Measurement
- Hypothesising
- Data collection

Fundamentals

- Formulation of research questions
- Conceptualisation
- Operationalisation
- Measurement
- Hypothesising
- Data collection
- Data analysis

Fundamentals

- Formulation of research questions
- Conceptualisation
- Operationalisation
- Measurement
- Hypothesising
- Data collection
- Data analysis
- Draw inferences

Day 1 – Outline

From concepts to measurement

- For instance, Hardin (2002) considers trust as a set of interests existing between individuals that develop through life experiences. It can be changed by personal interactions or new social and political conditions.

From concepts to measurement

- For instance, Hardin (2002) considers trust as a set of interests existing between individuals that develop through life experiences. It can be changed by personal interactions or new social and political conditions.
- Trust in institutions: public attitudes towards politicians, governments, judiciary..

From concepts to measurement

- For instance, Hardin (2002) considers trust as a set of interests existing between individuals that develop through life experiences. It can be changed by personal interactions or new social and political conditions.
- Trust in institutions: public attitudes towards politicians, governments, judiciary..
- "Most people can be trusted or you can't be too careful" European Social Survey (ESS)

From concepts to measurement

- For instance, Hardin (2002) considers trust as a set of interests existing between individuals that develop through life experiences. It can be changed by personal interactions or new social and political conditions.
- Trust in institutions: public attitudes towards politicians, governments, judiciary..
- "Most people can be trusted or you can't be too careful" European Social Survey (ESS)
- Higher trust in government is associated with compliance with the law, pro-incumbent voting, support for domestic policy liberalism.

From concepts to measurement

- Trust is also operationalised as a behavioural construct (e.g. Trust games).

From concepts to measurement

- Trust is also operationalised as a behavioural construct (e.g. Trust games).

- Scholars consider the first mover as trusting if she sends more than the 'minimum' amount and define the second mover as trustworthy if he returns more than the 'minimum' amount or more than the amount sent to him (Chaudhuri & Gangadharant, 2007).

What is a variable?

- Anything that can vary across observations (e.g. across individuals, organizations, etc.). Many aspects of social life can be treated as variables.

What is a variable?

- Anything that can vary across observations (e.g. across individuals, organizations, etc.). Many aspects of social life can be treated as variables.
- For example, If I was interested in your household income as a variable, that would vary across students in the class. What is the unit of analysis here?

What is a variable?

- Anything that can vary across observations (e.g. across individuals, organizations, etc.). Many aspects of social life can be treated as variables.
- For example, If I was interested in your household income as a variable, that would vary across students in the class. What is the unit of analysis here?
- The unit of analysis is the major entity that you are analyzing in your study (e.g. individuals, cities, schools, countries, etc.).

What is a variable?

- Anything that can vary across observations (e.g. across individuals, organizations, etc.). Many aspects of social life can be treated as variables.
- For example, If I was interested in your household income as a variable, that would vary across students in the class. What is the unit of analysis here?
- The unit of analysis is the major entity that you are analyzing in your study (e.g. individuals, cities, schools, countries, etc.).
- Some variables change their attributes (e.g. sex, ethnicity, religion), while others change their numerical values (e.g. height, weight, income, grades, temperature).

Levels of measurement

- Nominal variables, as known as categorical variables, (e.g. gender, ethnic group, religious affiliation). They are unordered categories.

Levels of measurement

- Nominal variables, as known as categorical variables, (e.g. gender, ethnic group, religious affiliation). They are unordered categories.
- Interval-ratio variables (e.g. income, wealth, age). They have an order with respect to each other, equal distance apart.

Levels of measurement

- Nominal variables, as known as categorical variables, (e.g. gender, ethnic group, religious affiliation). They are unordered categories.
- Interval-ratio variables (e.g. income, wealth, age). They have an order with respect to each other, equal distance apart.
- Ordinal variables (e.g.levels of education as opposed to years of education, opinion scales). The distance between values is unequal.

Levels of measurement

- Nominal variables, as known as categorical variables, (e.g. gender, ethnic group, religious affiliation). They are unordered categories.
- Interval-ratio variables (e.g. income, wealth, age). They have an order with respect to each other, equal distance apart.
- Ordinal variables (e.g.levels of education as opposed to years of education, opinion scales). The distance between values is unequal.
- Can you tell us a variable for each level?

Describing data

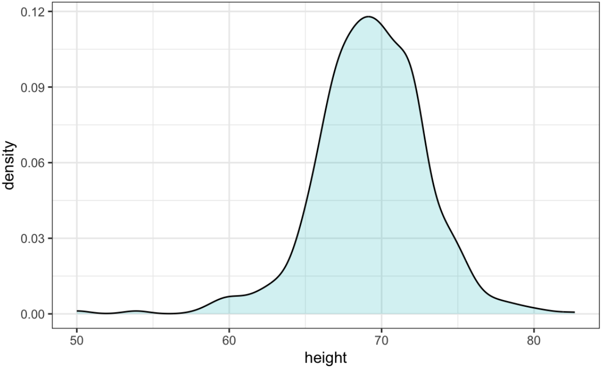
- There are lots of different ways to describe data.

Describing data

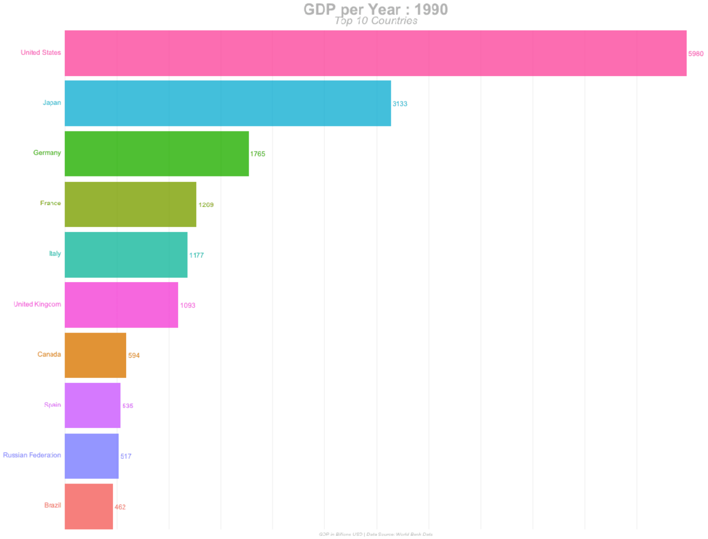
- There are lots of different ways to describe data.
- Describing data is necessary because there is usually too much of it, so it does not make any sense by itself.

Distributions

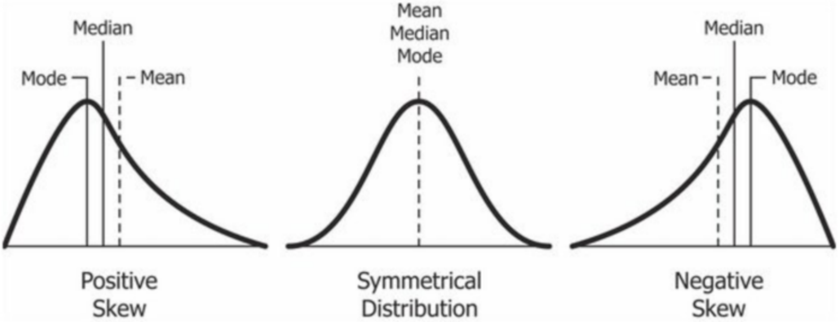
- In a population (or a sample), variables can be characterised by their frequency distribution (e.g. by the distribution of the frequency of their values). They can describe data as in the graph below.



Frequency distribution of GDP



Frequency distributions



Measures of the central tendency of distribution

- There are three standard measures: the mean, the mode, and the median.

Measures of the central tendency of distribution

- There are three standard measures: the mean, the mode, and the median.
- The mean is the conventional average calculated by adding all values and dividing by the number of cases. Mean can provide a distorted impression of central tendency when a variable contains a small number of extreme scores at one end of the distribution (e.g. outliers).

$$Mean = \bar{X} = \frac{\sum_{i=1}^{n} x_i}{N} \tag{1}$$

Measures of the central tendency of distribution

- There are three standard measures: the mean, the mode, and the median.
- The mean is the conventional average calculated by adding all values and dividing by the number of cases. Mean can provide a distorted impression of central tendency when a variable contains a small number of extreme scores at one end of the distribution (e.g. outliers).

$$Mean = \bar{X} = \frac{\sum_{i=1}^{n} x_i}{N} \tag{1}$$

- The mode is simply the most frequently occurring number in your measurement.

Measures of the central tendency of distribution

- There are three standard measures: the mean, the mode, and the median.
- The mean is the conventional average calculated by adding all values and dividing by the number of cases. Mean can provide a distorted impression of central tendency when a variable contains a small number of extreme scores at one end of the distribution (e.g. outliers).

$$Mean = \bar{X} = \frac{\sum_{i=1}^{n} x_i}{N} \tag{1}$$

- The mode is simply the most frequently occurring number in your measurement.
- Median is the value that falls in the middle of the ordering. For an odd number of n observations, middle observation is (n+1)/2.

Dispersion

- The variance is the sum of the squared difference scores, where the difference scores are computed between each score and the mean.

$$s_N = \frac{1}{N} \sum_{i=1}^{N} (x_i - \bar{x})^2 \tag{2}$$

In English:

$$Variance = \frac{\text{Sum of squared difference scores}}{\text{Number of observations}}$$

Dispersion

- The variance is the sum of the squared difference scores, where the difference scores are computed between each score and the mean.

$$s_N = \frac{1}{N}\sum_{i=1}^{N}(x_i - \bar{x})^2 \tag{2}$$

In English:

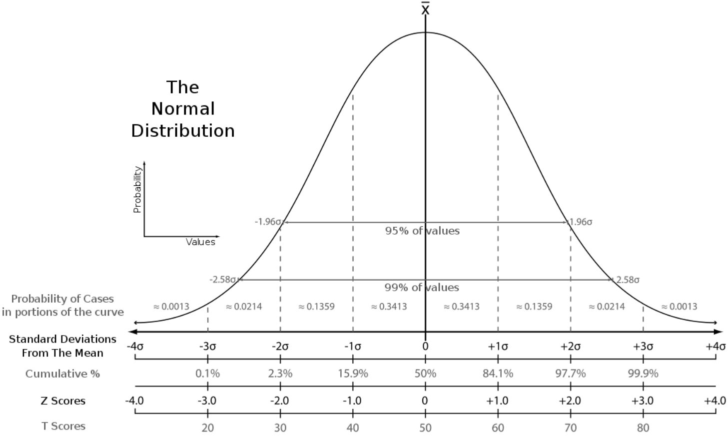
$$Variance = \frac{\text{Sum of squared difference scores}}{\text{Number of observations}}$$

- The standard deviation of a variable is the most common measure of its dispersion. It is the average departure of individual values from the mean value. In other words, it is the square root of the variance.

$$\text{Standard deviation} = \sqrt{\frac{\sum_i^n (x_i - \bar{x})^2}{N}} \tag{3}$$

The Normal Distribution

- Y is distributed according to a normal distribution with mean μ and variance δ^2 – $Y \sim N(\mu, \delta^2)$



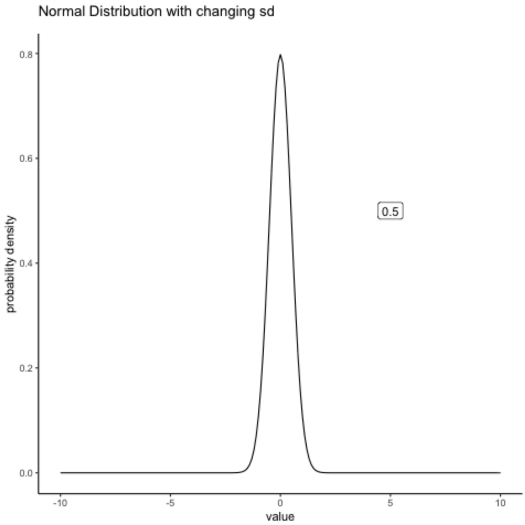
Standard Normal Distribution

- To calculate probabilities for a normal variable with a general mean and variance, we must standardise the variable by first subtracting the mean, then by dividing the result by the standard deviation

$$\mathrm{z} = \frac{X - \mu_x}{\delta_x} \tag{4}$$

- Z-scores indicate the number of standard deviation units a value is from the mean of a distribution
- The standard normal distribution is the normal distribution with mean $\mu = 0$ and variance $\delta^2 = 1$ and is denoted N(0, 1).

Normal Distribution with High and Low Standard Deviation



Random Sampling and Sampling Distribution

- Sampling theory plays a huge role in specifying the assumptions upon which your statistical inferences rely. We draw inferences from (the sample) and what it is that we're drawing inferences about (the population)
- Most statistical methods rely on averages or weighted averages of a sample of data
- The values of the variables in our sample are random variables, and so any average we take of those values will also be random variables

Most samples are not simple random samples: What if we don't have a simple random sample

- **Simple random sample** every unit has equal probability of being selected. It simply involves selecting units from a population without replacement

Most samples are not simple random samples: What if we don't have a simple random sample

- **Simple random sample** every unit has equal probability of being selected. It simply involves selecting units from a population without replacement
- **Stratified random sample**, population is divided into strata and a simple random sample selected within each strata

Most samples are not simple random samples: What if we don't have a simple random sample

- **Simple random sample** every unit has equal probability of being selected. It simply involves selecting units from a population without replacement
- **Stratified random sample**, population is divided into strata and a simple random sample selected within each strata
- **Convenience sample**, The samples are chosen in a way that is convenient to the researcher, and not selected at random from the population of interest

The Sampling Distribution of the Sample Average

- In a simple random sample, n objects are drawn at random from a population and each object is equally likely to be drawn

$$\bar{Y} = \frac{Y_1 + Y_2 + ... + Y_n}{n} = \frac{\sum_{i=1}^{N} Y_i}{n} \tag{5}$$

The Sampling Distribution of the Sample Average

- In a simple random sample, n objects are drawn at random from a population and each object is equally likely to be drawn

$$\bar{Y} = \frac{Y_1 + Y_2 + ... + Y_n}{n} = \frac{\sum_{i=1}^{N} Y_i}{n} \tag{5}$$

- The value of $\bar{Y}$ will differ from one random sample to the next

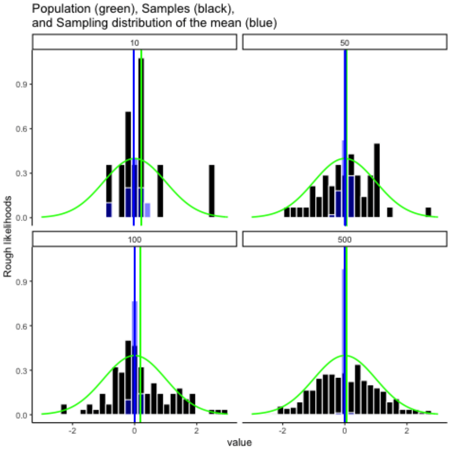
The Sampling Distribution of the Sample Average

- In a simple random sample, n objects are drawn at random from a population and each object is equally likely to be drawn

$$\bar{Y} = \frac{Y_1 + Y_2 + ... + Y_n}{n} = \frac{\sum_{i=1}^{N} Y_i}{n} \tag{5}$$

- The value of $\bar{Y}$ will differ from one random sample to the next
- As $\bar{Y}$ is random, it has a probability distribution. This is called the sampling distribution of $\bar{Y}$.

The Sampling Distribution of the Sample Average



- It gives an idea of how much the values of the estimate (mean) might vary from sample to sample
- This tells us something about the uncertainty in the estimate obtained from the one sample we actually observe

Day 1 – Outline

Hypothesis testing

- A **theory** is a tentative conjecture about the causes of some phenomenon of interest. Once a theory has been developed, we can restate it into one or more testable hypotheses.

Hypothesis testing

- A **theory** is a tentative conjecture about the causes of some phenomenon of interest. Once a theory has been developed, we can restate it into one or more testable hypotheses.
- A **hypothesis** is a theory-based statement about a relationship that we expect to observe.

Hypothesis testing

- A **theory** is a tentative conjecture about the causes of some phenomenon of interest. Once a theory has been developed, we can restate it into one or more testable hypotheses.
- A **hypothesis** is a theory-based statement about a relationship that we expect to observe.
- For every hypothesis there is a corresponding null hypothesis (e.g. there is no association between Y and X in the population).

Hypothesis testing

- A **theory** is a tentative conjecture about the causes of some phenomenon of interest. Once a theory has been developed, we can restate it into one or more testable hypotheses.
- A **hypothesis** is a theory-based statement about a relationship that we expect to observe.
- For every hypothesis there is a corresponding null hypothesis (e.g. there is no association between Y and X in the population).
- A **null hypothesis** is also a theory-based statement but it is about what we would expect to observe if our theory was incorrect.

Key elements in hypothesis testing

1. Assumptions
2. Hypothesis
3. Calculate a test-statistic
4. Calculate p-value
5. Drawing a conclusion

Assumptions

- Type of data/variables

Assumptions

- Type of data/variables
- Population distribution

Assumptions

- Type of data/variables
- Population distribution
- Sampling technique (e.g. simple random)

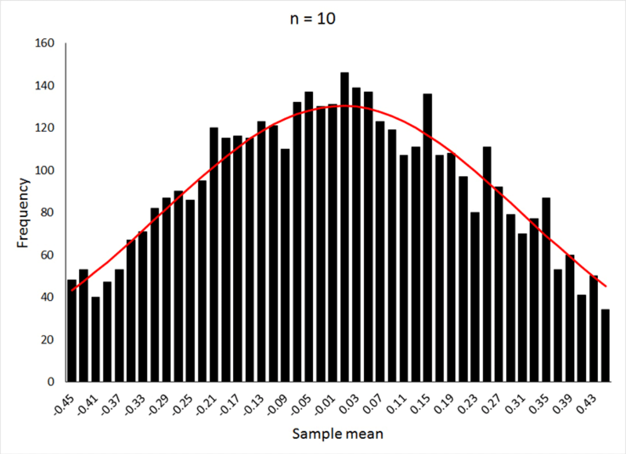
Assumptions

- Type of data/variables
- Population distribution
- Sampling technique (e.g. simple random)
- Sample size (small/large sample)

Assumptions

- If we have a simple random sample, then the best estimate of the population mean (μ) is the sample mean ($\bar{x}$)
- Central Limit Theorem: the entire distribution gets arbitrarily close to the standard normal distribution as n gets larger. ($N \geq 30$).
- Let $\{Y_1, Y_2, ..., Y_N\}$ be a random sample with mean and variance. $Z_n = (\frac{\bar{Y} - \mu}{\sigma/\sqrt{n}})$
- If we take repeated samples from a population, the means from those samples will fall in a normal distribution around the population mean.

Assumptions



- We know (through Central Limit Theorem) that repeated sampling will produce a normal distribution around the population value.

Hypothesis

- Hypothesis testing means determining whether or not we should reject or fail to reject the null hypothesis.
- $H_0 : \mu = \mu_0$ where μ_0 is some particular value.
  $H_A : \mu \neq \mu_0$
- We reject the null when there is a low probability that the claim made by the null is true in the population of interest.
- This probability is called the $\alpha - level$ and conventionally takes a value of **0.05**, **0.01**, or **0.001**.

Types of errors in hypothesis testing

- Type-I error (false-positive): rejecting H_0 when it is true.
- Type-II error (false-negative): not rejecting H_0 when it is false.
- There is a trade-off between minimizing type-I and type-II errors.
  $\downarrow$
- As we increase our α-level, we are more likely to commit a Type-I error, and less likely to commit a Type-II error.
- Hypothesis testing rules are constructed to make the probability of committing a Type-I error fairly small. We define the significance level (simply the α-level) of a test as the probability of a Type-I error.
- E.g. an $\alpha - level$ of 0.05 implies that, in the process of repeated sampling, we will incorrectly reject the null hypothesis 5% of the time.

$$\alpha = P(Reject H_0 | H_0)$$

Test statistics

- Point estimate or a number calculated from the sample that is used to test the null hypothesis.

Test statistics

- Point estimate or a number calculated from the sample that is used to test the null hypothesis.
- Single number that quantifies the difference between the observed data, and the expected data under the null hypothesis.

Test statistics

- Point estimate or a number calculated from the sample that is used to test the null hypothesis.
- Single number that quantifies the difference between the observed data, and the expected data under the null hypothesis.
- Examples of test statistics are: Chi-square, F-test, T-test, correlation coefficient r.

Test statistics

- Point estimate or a number calculated from the sample that is used to test the null hypothesis.
- Single number that quantifies the difference between the observed data, and the expected data under the null hypothesis.
- Examples of test statistics are: Chi-square, F-test, T-test, correlation coefficient r.
- The test statistic that should be used depends on the type of analysis.

T-statistic

William Sealy Gosset (1876-1937) was a chemist and statistician as Head Experimental Brewer of Guinness

Gosset figured out the solution to a specific problem: what's the fewest amount of samples we need to take to know that would mean Guinness could test fewer beers for quality, sell more beers for profit, and make the product testing time shorter? And he invented the t-test (1908).

T-test

- Because we must estimate the standard deviation by using the standard error, we must use the t distribution rather than the standard normal (z) distribution.

T-test

- Because we must estimate the standard deviation by using the standard error, we must use the t distribution rather than the standard normal (z) distribution.
- T-distribution controls for the fact that the estimate of standard deviation is going to be a bit o for smaller samples.

T-test

- Because we must estimate the standard deviation by using the standard error, we must use the t distribution rather than the standard normal (z) distribution.
- T-distribution controls for the fact that the estimate of standard deviation is going to be a bit o for smaller samples.
- Will be flatter in smaller samples, but approximates z as size increases.

T-test

- Because we must estimate the standard deviation by using the standard error, we must use the t distribution rather than the standard normal (z) distribution.
- T-distribution controls for the fact that the estimate of standard deviation is going to be a bit o for smaller samples.
- Will be flatter in smaller samples, but approximates z as size increases.
- T distribution looks like normal distribution but has fatter tails (and degrees of freedom)

T-test

- Because we must estimate the standard deviation by using the standard error, we must use the t distribution rather than the standard normal (z) distribution.
- T-distribution controls for the fact that the estimate of standard deviation is going to be a bit o for smaller samples.
- Will be flatter in smaller samples, but approximates z as size increases.
- T distribution looks like normal distribution but has fatter tails (and degrees of freedom)
- Formula for t-statistic similar to z-statistic, except uses the standard error:

$$t = \frac{(\bar{x} - \mu_0)}{\frac{s_x}{\sqrt{n}}}$$

$$t = \frac{\text{Sample Mean - Population Mean}}{\text{Sample Standard Error}}$$

Inference from two means (T-test)

- Because we assume equal variances between groups, we pool the information on variability (sample variances) to generate an estimate of the variability in the population. Pooled sample variance:

$$S_p^2 = \frac{(n-1)S_X^2 + (m-1)S_Y^2}{n+m-2}.$$

Let $X_1, \cdots, X_n$ and $Y_1, \cdots, Y_m$ be two independent samples from normal distributions with the same population variance. The test statistic for testing:

$$H_0 : \mu_X = \mu_Y \text{ vs } H_1 : \mu_X \neq \mu_Y$$

$$T = \frac{\bar{X} - \bar{Y} - (\mu_X - \mu_Y)}{S_p\sqrt{1/n + 1/m}} \sim t_{n+m-2}.$$

Reject H_0 if $|T| > t_{\alpha/2,n+m-2}$

Example

- Race and job market discrimination
- This hypothetical example is based on an experiment conducted to examine the effects of racial feature of job candidates on how recruiters make their decisions on hiring.
- The subjects of the study were 133 people. The researcher sent job applications to administrative roles in the UK, the attached CVs in the job applications randomly either include ethnic-sounding names, or exclude names of candidates, thus creating two experimental conditions.

Example

- The response variable Y is a measure of the recruiter's pre-screening decision for hiring, measured on a 10-point scale where large values indicate high levels of being interested in the job candidate.

Example – Data

- Samples from two groups, one with the experimental condition where CVs exclude names, with sample size $n_1 = 67$, mean $\bar{Y}_1 = 8.23$ and standard deviation $s_1 = 2.39$, and the second with the experimental condition where CVs include ethnic-sounding names, with $n_2 = 66$, $\bar{Y}_2 = 6.49$ and $s_2 = 2.01$.

Example – Assumptions

- The observations are random samples of statistically independent observations from two populations, one with mean μ_1 and standard deviation σ_1, and the other with with mean μ_2 and the same standard deviation σ_2, where the standard deviations are equal, with value $\sigma = \sigma_1 = \sigma_2$. The sample sizes n_1 and n_2 are sufficiently large for the sampling distribution of the test statistic under the null hypothesis to be approximately standard normal.

Example – Hypotheses

- These are about the difference of the population means $\Delta = \mu_2 - \mu_1$, with null hypothesis $H_0 : \Delta = 0$.
- Although the two-sided alternative hypothesis, $H_a : \Delta \neq 0$, is considered in this example, you can also test one of the one-sided alternative hypotheses, $H_a : \Delta > 0$ or $H_a : \Delta < 0$.

Example – Calculate t-statistic

- Test statistic: the two-sample t-statistic] $t = \frac{\hat{\Delta}}{\hat{\sigma}_{\hat{\Delta}}} = \frac{-1.74}{0.383} = -4.55$

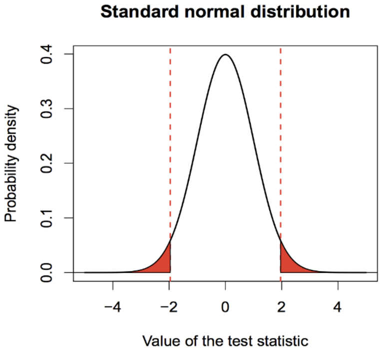
  Where $\hat{\Delta} = \bar{Y_2} - \bar{Y_1} = 6.49 - 8.23 = -1.74$

  and $\hat{\sigma}_{\hat{\Delta}=\hat{\sigma}\sqrt{\frac{1}{n_2}+\frac{1}{n_1}}} = 2.210 \text{ x } \sqrt{\frac{1}{66} + \frac{1}{67}} = 0.383$

  with $\hat{\sigma} = \sqrt{\frac{(n_2-1)s_2^2+(n_1-1)s_1^2}{n_1+n_2-2}} = \sqrt{\frac{65x2.01^2+66x2.39^2}{131}} = 2.210$

P value

- If in the population there was no association between Y and X, what are the chances of obtaining results from the sample data that suggest otherwise?
- In other words, the P-value is the probability of observing a sample result as extreme as by pure chance given that the null hypothesis is actually true.



P value

- Given we set α at 0.05, the critical value of the test statistic under the normal distribution is 1.96.
- Under the standard normal, the probability of $|t| \geq 1.96$ is 0.05.
- The p-value will always be less than 0.05 when t is less than -1.96 or greater than 1.96.
- We can actually reject the null without calculating the p-value directly.

Example – Drawing conclusion(s)

- The probability that a randomly selected value from the standard normal distribution is at most -4.55 or at least 4.55, which is about 0.000005 (reported as $P < 0.001$).
- Therefore, a two-sample t-test indicates very strong evidence that the average level of pre-screening decision on hiring is different when the CVs include ethnic-sounding names than when they don't ($P < 0.001$).

From hypothesis test to confidence interval

- We would also like to know the range of plausible values of our statistic in the population. We use a confidence interval to quantify the uncertainty of the inference.
- To achieve this, we can calculate the confidence intervals of our statistic.
- The width of confidence interval depends on standard error (SE) of the estimator, for the given sample size.
- The smaller the SE, the narrower confidence interval, and the more accurate (efficient) the estimate.
- The most commonly utilised confidence levels are 90%, 95%, and 99%.

Confidence interval

- Confidence intervals for the differences Δ are also of the familiar form where $z_{\frac{\alpha}{2}}$ is the appropriate multiplier from the standard normal distribution to obtain the required confidence level, e.g. $z_{.025} = 1.96$ for 95% confidence intervals. The multiplier is replaced with a slightly different one if the t-distribution is used as the sampling distribution.

$$\hat{\Delta} \pm z_{\frac{\alpha}{2}} \hat{\sigma}_{\hat{\Delta}}$$

Confidence interval

- A confidence interval for the mean difference $\Delta = \mu_1 - \mu_2$ is obtained by substituting appropriate expressions into the general formula in the previous slide.
- Specifically here $\Delta = \bar{Y_2} - \bar{Y_1}$ and a 95% confidence interval for Δ is:
  $(\bar{Y_2} - \bar{Y_1}) \pm 1.96\hat{\sigma}\sqrt{\frac{1}{n_2} + \frac{1}{n_1}}$
- The validity of this again requires that the sample sizes n_1 and n_2 from both groups are reasonably large.
- Let's calculate the confidence intervals for our example!
  -1.74 $\pm$ 1.96 x 0.383 = -1.74 $\pm$ 0.75 = $(-2.49; -0.99)$.

The interpretation of confidence intervals

- Based on the data in our example, we are thus 95% confident that recruiter's hiring tendency is lower when CVs include ethnic-sounding names compared to the condition where CVs exclude names, by between 0.99 and 2.49.

The interpretation of confidence intervals

- Based on the data in our example, we are thus 95% confident that recruiter's hiring tendency is lower when CVs include ethnic-sounding names compared to the condition where CVs exclude names, by between 0.99 and 2.49.
- The confidence does not include the null difference of 0: we reject the null.

The interpretation of confidence intervals

- Based on the data in our example, we are thus 95% confident that recruiter's hiring tendency is lower when CVs include ethnic-sounding names compared to the condition where CVs exclude names, by between 0.99 and 2.49.
- The confidence does not include the null difference of 0: we reject the null.
- Confidence intervals will always give the same conclusions as a hypothesis test at the same α-level.

The interpretation of confidence intervals

- Based on the data in our example, we are thus 95% confident that recruiter's hiring tendency is lower when CVs include ethnic-sounding names compared to the condition where CVs exclude names, by between 0.99 and 2.49.
- The confidence does not include the null difference of 0: we reject the null.
- Confidence intervals will always give the same conclusions as a hypothesis test at the same α-level.
- If we continuously resample from the population, the confidence intervals will include μ_Y in $1 - \alpha$ of our samples.

Moving beyond hypothesis testing

- Analyze all your data: Use all of your data to provide yourself and your readers with all the relevant information. Better to anticipate potential criticisms than to hide your data and fear for the eventual exposure.

Moving beyond hypothesis testing

- Analyze all your data: Use all of your data to provide yourself and your readers with all the relevant information. Better to anticipate potential criticisms than to hide your data and fear for the eventual exposure.
- Present all your comparisons: It is recommended to display more of your data rather than focusing on comparisons that happen to reach statistical significance.

Moving beyond hypothesis testing

- Analyze all your data: Use all of your data to provide yourself and your readers with all the relevant information. Better to anticipate potential criticisms than to hide your data and fear for the eventual exposure.
- Present all your comparisons: It is recommended to display more of your data rather than focusing on comparisons that happen to reach statistical significance.
- Make your data public (subject to any confidentiality restrictions). If the topic is worth studying, you should want others to be able to make rapid progress.