Lecture 2
Probability, Sampling and Statistical Significance

Dr Adam Mahdi

University of Oxford

October, 22, 2021

Last week:

- Course overview

- Statistical thinking
- Evidence triangle
- Descriptive statistics

Today's class:

- Probability (review)

- Inferential statistics
- Sampling distribution
- Central Limit Theorem

- Confidence intervals
- Hypothesis testing

Probability

Probability theory (one slide review)

- Experiment, sample, sample space, events, probability
- Conditional probability, Bayes' theorem, independence
- Discrete random variables

  Probability mass function (**PMF**): $P(X = x_i) = p(x_i)$
  $p(x_i) \geq 0 \quad$ and $\quad \sum p(x_i) = 1$

- Continuous random variables

  Cumulative distribution function (**CDF**): $F(x) = P(X < x)$ and
  Probability density function (**PDF**): $P(a \leq X \leq b) = \int_a^b f(x)dx$
  $f(x) \geq 0 \quad$ and $\quad \int_{\infty}^{\infty} f(x)dx = 1,$

- Expectation

  $E[X] = \sum x_i p_i$ (discrete), $\quad E[X] = \int_{\infty}^{\infty} xf(x)dx$ (continuous)
  $E(aX + bY) = aE(X) + bE(Y)$

- Variance

  $Var(X) = E[X - E[X]]^2 = E[X^2] - E[X]^2$
  $Var(aX + b) = a^2 Var(X)$

Sets and operations

Set: (e.g. A, B) is collection of objects

Empty set ($\varnothing$): set with no elements.

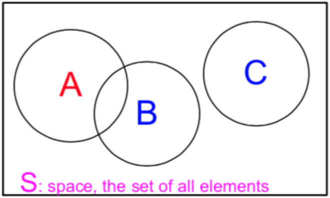
Subset: $A \subset B$ means that A is a subset of B

Union: $A \cup B$ is the set containing all the elements of A and B

Intersection: $A \cap B$ contains the elements common to both A and B

Complement: A^c is a set containing all the elements not in a particular set

Venn diagrams: a good way of visualising sets, and set operations



Laws of set algebra

Commutative property: $A \cup B = B \cup A$

Associative property: $(A \cup B) \cup C = A \cup (B \cup C)$

Distributive property: $A \cup (B \cap C) = (A \cup B) \cap (A \cup C)$

Additionally, if S is the space, i.e. the set that contains all possible elements, and $A \subset S$, then

$$A \cup \varnothing = A, \qquad A \cap \varnothing = \varnothing$$

$$A \cup S = S, \qquad A \cap S = A$$

$$A \cup A^c = S, \qquad A \cap A^c = \varnothing$$

$$A \cup A = A, \qquad A \cap A = A$$

Example (rolling a die - 1)

Consider rolling a conventionally numbered die once and getting a 6.

Experiment: rolling a die

Outcome (or **sample**): getting a 6

Sample space: $S = \{1, 2, 3, 4, 5, 6\}$

Often we are not interested in individual outcomes, but in whether an outcome belongs to a given subset (e.g. A) of S. These subsets are called **events**. In this example, we might consider two mutually exclusive events: throwing an even number or throwing an odd number; another event would be throwing a number which is an integer multiple of 3.

Example (rolling a die - 2)

If the die we considered above is *fair* or *unbiased*, each outcome is equally probable and we can say that the probability of getting any number is $1/6$.

$$P(\{1\}) = P(\{2\}) = P(\{3\}) = P(\{4\}) = P(\{5\}) = P(\{6\}) = \frac{1}{6}$$

A central idea in understanding probability calculations is the concept of **relative frequency**, i.e. the frequency with which we can expect a particular event to appear among all possible events. If all events are equally likely, we just need to count the number of possible results to assess probabilities.

Probability

Consider an experiment whose sample space is S. For each event E of the sample space S there exists a value $P(E)$ called the **probability** of E. The probabilities satisfy the following conditions (axioms):

$$\textbf{Axiom 1:} \qquad 0 \leq P(E) \leq 1$$

$$\textbf{Axiom 2:} \qquad P(S) = 1$$

Axiom 3: For any sequence of mutually exclusive events $E_1, E_2, \ldots$ (that is, events for which $E_i \cap E_j = \varnothing$ when $i \neq j$)

$$P(\bigcup_{i=1}^{\infty} E_i) = \sum_i P(E_i)$$

Example (jackpot)

What is the chance of winning the jackpot in the national lottery?

There are 49 balls, numbered from 1 to 49, in the machine. Six balls are selected without replacement and the jackpot is won when all 6 of the selected balls are correctly identified. How many ways of choosing combinations of 6 balls from a set of 49 are there?

$$\binom{49}{6} = \frac{49!}{(49-6)!6!} = \frac{49!}{43!6!} = 13,983,816$$

Our chances: each combination is equally likely so the chance of selecting all the numbers is one in 13,983,816.

Conditional probability

- There are many situations in which we possess prior information, i.e. we already know something about the outcome.

As an example for our die we can ask "if we know that the outcome is even, what is the probability of it being larger than 3?"

- What we are trying to calculate here is a **conditional probability**, i.e. the probability of the event A given that the event B has happened, written as $P(A|B)$.

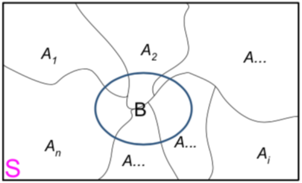
$$P(A|B) = \frac{P(A \cap B)}{P(B)}, \qquad P(B) \neq 0$$

$$P(A \cap B) = P(A|B)P(B) = P(B|A)P(A)$$

We read this: *the chance of two events happening simultaneously is the chance of one of them happening, multiplied by the chance of the second happening given that the first has happened.*

Total probability

If we partition the sample space S into a set of n disjoint sets A_i



the addition rule above that the probability of B is given by

$$P(B) = P(B|A_1)P(A_1) + P(B|A_2)P(A_2) + \ldots + P(B|A_n)P(A_n)$$

In particular

$$P(B) = P(B|A)P(A) + P(B|A^c)P(A^c)$$

Bayes' rule

Rearranging the equalities

$$P(A \cap B) = P(A|B)P(B) = P(B|A)P(A)$$

we obtain the **Bayes' rule**

$$P(A|B) = \frac{P(B|A)P(A)}{P(B)}$$

Or combined with total probability we obtain:

$$P(A_j|B) = \frac{P(B|A_j)P(A_j)}{\sum_{i=1}^{n} P(B|A_i)P(A_i)}$$

Example (picnic)

You are planning a picnic today, but the morning is cloudy. We know that 50% of all rainy days start off cloudy! But cloudy mornings are common (about 40% of days start cloudy). And this is usually a dry month (only 3 of 30 days tend to be rainy, or 10%) What is the chance of rain during the day?

- We will use Rain to mean rain during the day, and Cloud to mean cloudy morning.
- The chance of Rain given Cloud is written $P(Rain|Cloud)$
- Bayes' formula

$$P(Rain|Cloud) = \frac{P(Rain)P(Cloud|Rain)}{P(Cloud)}$$

$$P(Rain|Cloud) = \frac{0.1 \times 0.5}{0.4} = .125$$

Independence

If prior knowledge does not affect the probability of the second event, $P(A|B) = P(A)$ that is

$$P(A \cap B) = P(A|B)P(B) = P(B|A)P(A) = P(A)P(B)$$

then we say that the events A and B are **independent**.

So, if two events are independent, the probability that the two of them happen in the same experiment is the product of their individual probabilities. Note that statistical dependence does not require any causative link.

Rare events

Example

A test for a rare disease detects the disease with a probability of 99%, and has a false positive ratio (i.e. it tests positive even though the person is healthy) of 0.5%. We know that the percentage of the general population who have the disease is 1 in 10,000.

Suppose we chose a random subject and perform the test, which comes out positive. What is the probability of the person actually having the disease?

Example: Rare events

- The probability of D (having the disease) before the test is 1 in 10,000: $P(D) = 0.0001$ (i.e. $P(D^c) = 0.9999$)
- The conditional probabilities of getting a positive in the test when the person does / does not have the disease are $P(T|D) = 0.99$ (true positive ratio) and $P(T|D^c) = 0.005$ (false positive ratio).
- The probability of getting a positive result in the test is $P(T)$, which we can calculate using total probability:

$$P(T) = P(T|D)P(D) + P(T|D^c)P(D^c) = 0.005$$

- Now we just need to apply Bayes' theorem

$$P(D|T) = \frac{P(T|D)P(D)}{P(T)} = \frac{0.99 \times 0.0001}{0.005} = \mathbf{0.02}$$

Are you surprised???

Random Variables

When an experiment is performed, we are frequently interested mainly in some function of the outcome as opposed to the actual outcome itself. For instance, in tossing dice, we are often interested in the sum of the two dice and are not really concerned about the separate values of each die. That is, we may be interested in knowing that the sum is 7 and may not be concerned over whether the actual outcome was $(1, 6)$, $(2, 5)$, $(3, 4)$, $(4, 3)$, $(5, 2)$, or $(6, 1)$.

These quantities of interest, or, more formally, these real-valued functions defined on the sample space, are known as **random variables.**

If the sample space is a set of discrete points then the variable is **discrete**, otherwise it is **continuous**.

Discrete random variables

Given a sample space containing a discrete number of values x_i, we say that the probability that the random variable X equals that number x_i, is $p(x_i)$ or

$$P(X = x_i) = p(x_i)$$

- The function, $p(x_i)$, is a **discrete probability density function** (pdf). The function must obey the basic rules of probability, i.e.

$$p(x_i) \geq 0 \qquad \sum_i p(x_i) = 1.$$

- For historical reasons the discrete probability density function is also known as the **probability mass function.**
- **Cumulative Distribution Function (CDF)** is given by

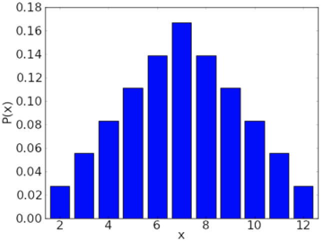
$$F(x) = P(X < x)$$

Distributions

What is the probability distribution for a random variable X that is given by the sum of the result of rolling two six-sided dice?

- X = sum of the results of rolling two six-sided dice
- $X = \{2, 3, \ldots, 12\}$
- The probability distribution of X are all the values $P(X = x_i) =?$ when x_i is $2, 3, \ldots, 12$.
- For example what is $P(X = 2)$, that is what is the probability of rolling two 1s?

$$P(x) = \begin{cases} \frac{1}{36} & \text{if } x \in \{2, 12\} \\ \frac{2}{36} = \frac{1}{18} & \text{if } x \in \{3, 11\} \\ \frac{3}{36} = \frac{1}{12} & \text{if } x \in \{4, 10\} \\ \frac{4}{36} = \frac{1}{9} & \text{if } x \in \{5, 9\} \\ \frac{5}{36} & \text{if } x \in \{6, 8\} \\ \frac{6}{36} = \frac{1}{6} & \text{if } x = 7 \\ 0 & \text{otherwise.} \end{cases}$$



Expectation and Variance (discrete)

The **expected value** of X is a weighted average (by the probability) of the possible values that X can take on:

$$E[X] = \sum_i x_i p(x_i)$$

The **variance** gives information about how much the values of a random variable X are likely to vary between tests

$$Var[X] = E\left[(X - E[X])^2\right] = E[X^2] - (E[X])^2$$

Properties:

- $E[aX + b] = aE[X] + b$
- $Var[aX + b] = a^2 Var[X]$

Bernoulli distribution

Question: What is the probability of getting a 6 when rolling a die?

In a **Bernoulli distribution** we have only two possible outcomes in a single trial: 1 (success) and 0 (failure).

So the random variable X which has a Bernoulli distribution can take value 1 with the probability of success, say p, and the value 0 with the probability of failure, say q or $1-p$.

$$P(X=k) = \begin{cases} \text{1-p} & \text{k=0} \\ \text{p} & \text{k=1} \end{cases}$$

If X is a Bernoulli distribution, show that:

- $E(X) = p$
- $Var(X) = p(1-p)$

Binomial distribution

Question: At a particular junction 10% of cars turn left. Five cars approach the junction, what is the probability that exactly 3 will turn left?

In a **Binomial distribution** we are counting the number of times a condition, which can either be success or a failure, is met in n identical trials.

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

- Each trial is independent
- Only two possible outcomes in a trial
- A total number of n identical trials are conducted
- The probability of success and failure is same for all trials

If X is a Binomial distribution, show that:

- $E(X) = np$
- $Var(X) = np(1-p)$

Poisson distribution

Question: If the number of accidents on a particular road is 2 per month, what is the probability that there are no accidents during a given month?

$$P(X = k) = \frac{\lambda}{k!} e^{-\lambda}$$

The requirements are:

- each event is independent of each other
- only one event can happen at a time
- the mean rate of events is constant.

If X is a Poisson distribution, show that:

- $E(X) = \lambda$
- $Var(X) = \lambda$

Continuous random variables

Previously we considered discrete random variables - variables that can take one of a countable set of distinct values. In other cases the variable can vary continuously over a range: take for example the distribution of student heights within your class.

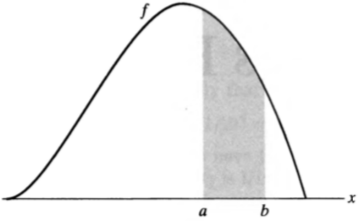
- We say that X is a **continuous random variable** if there exists a nonnegative function f, called **probability density function (PDF)** defined for all real $x \in (-\infty, +\infty)$ having the property that for any set B of real numbers

$$P[X \in B] = \int_B f(x)dx$$

- All probability statements about X can be answered in terms of f!!!

Example (height)

What is the probability that the height of randomly selected Londoner will be between $1.65m$ and $1.75m$? Assume we know the population (i.e. probability density function is known).



$P(a \leq X \leq b)$ = area of shaded region

$$P(1.65 \leq X \leq 1.75) = \int_{1.65}^{1.75} f(x)dx$$

Expectation and Variance (continuous)

The **expected value** of a continuous random variable X is given by

$$E[X] = \int_{-\infty}^{+\infty} xf(x)dx$$

Similarly as in the discrete case, the **variance** gives information about how much the values of a X are likely to vary between tests

$$Var[X] = E\left[(X - E[X])^2\right] = E[X^2] - (E[X])^2$$

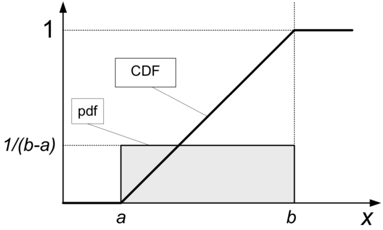
where

$$E[X^2] = \int_{-\infty}^{+\infty} x^2 f(x)dx$$

Uniform distribution

One of the simplest continuous distributions is the uniform distribution, where outcomes are equally probable inside a range $[a, b]$.

$$f(x) = \begin{cases} 1/(b-a) & \text{for } a \leq x \leq b \\ 0 & \text{elsewhere} \end{cases}$$



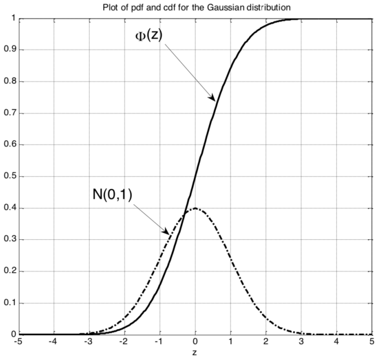
If X has a uniform distribution $U(a, b)$, show that:

- $E(X) = (a+b)/2$
- $Var(X) = (b-a)^2/12$

Normal distribution

The most important in probability is the normal (or Gaussian) distribution:

$$f(x) = N(m, \sigma^2) = \frac{1}{\sigma\sqrt{2\pi}} exp\Big[-\frac{(x-m)^2}{2\sigma^2} \Big]$$



Probability theory (one slide review)

- Experiment, events, sample space, probability of events
- Conditional probability, Bayes' theorem, independence
- Discrete random variables

  Probability mass function (**PMF**): $P(X = x_i) = p(x_i)$
  $p(x_i) \geq 0 \quad$ and $\quad \sum p(x_i) = 1$

- Continuous random variables

  Cumulative distribution function (**CDF**): $F(x) = P(X < x)$ and
  Probability density function (**PDF**): $P(a \leq X \leq b) = \int_a^b f(x)dx$
  $f(x) \geq 0 \quad$ and $\quad \int_{\infty}^{\infty} f(x)dx = 1,$

- Expectation

  $E[X] = \sum x_i p_i$ (discrete), $\qquad E[X] = \int_{\infty}^{\infty} xf(x)dx$ (continuous)
  $E(aX + bY) = aE(X) + bE(Y)$

- Variance

  $Var(X) = E[X - E[X]]^2 = E[X^2] - E[X]^2$
  $Var(aX + b) = a^2 Var(X)$

Tutorial sheet

Please submit the solutions by **Thursday Week 4**.

APPLIED ANALYTICAL STATISTICS

PROBLEM SET FOR LECTURE 2

A Mahdi, GY Qian, AE Zarebski

Set theory

1. Let A and B be events with probability $P(A) = \frac{2}{3}$ and $P(B) = \frac{2}{5}$. Show that $\frac{1}{15} \leq P(A \cap B) \leq \frac{2}{5}$. Find the corresponding bounds for $P(A \cup B)$.

2. If A, B and C are subsets of S, show that:

$$P(A \cup B \cup C) = P(A) + P(B) + P(C) - P(A \cap B) - P(B \cap C)$$
$$-P(A \cap C) + P(A \cap B \cap C)$$

Random variables

3. To prepare them for marking students' test papers, some volunteer practised by marking a paper that the examiners had originally graded at 97.0%. The random distribution of marks given by these volunteers are as follows:

Mark (%)	96.8	96.9	97.0	97.1	97.2
Proportion	0.03	0.07	0.80	0.07	0.03

What are the mean and standard deviation of the marks? Give your answer to 3 significant figures.

4. The following year, another group of markers practised on that same paper, producing a mean mark of 97.1% with a standard deviation of 0.15%. What are the mean and standard deviation of the average of two marks, one from each group of markers?

Tutorial sheet

Please include self-assessment sheet as the cover:

Tutorial Self Assessment Form

Tutor Dr Adam Mahdi, University of Oxford

Paper ___________________________

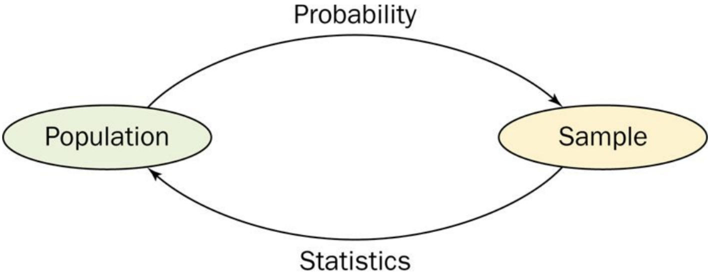
Student_ ________________________

Please indicate which questions you have completed successfully. If you have got stuck, please indicate where the problems are in each question. If there is a particular topic you want to go over in a tutorial, then please indicate this.

Question	Comments	Self Assessment Mark (out of 10)
1		
2		
3		

Statistics

Probability vs Statistics



Sampling

- Population and Samples
- Statistics (inferential) use a random sample of data taken from a population to describe the population
- These generalisations are usually made in spite of the fact that our dataset (while theoretically 'large' in absolute terms) is often a trivial portion of the whole population (the group of people who we want to generalise to)

Sampling and estimation

Estimates: Once we have a sample we can make guesses (inferences) about the population. E.g. if the mean income of our sample is £$22,000$, then we could guess (i.e. estimate) that the mean income of the population is the same.

Uncertainty: We also want to know how certain we can be of that guess. Because the process of sampling means there is a chance you will pick a group of people who don't represent the population.

Logic of inferential statistics

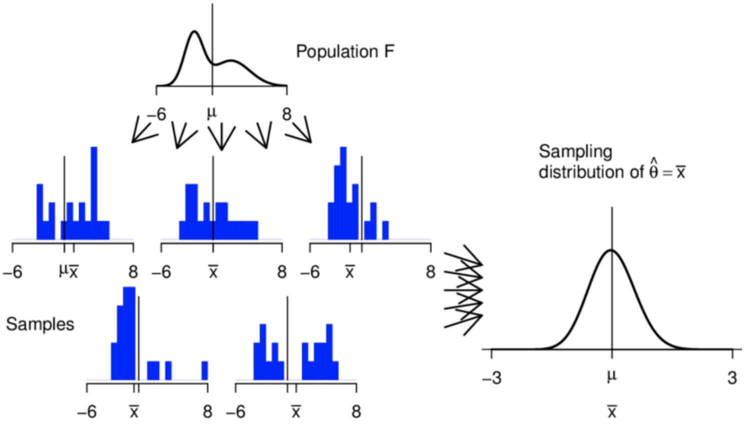
What we can infer depends on the *quality* of the sample set:

- How big was it?
- How varied was it?

- Think of this:
  - If we ask 10,000 people we will be more confident about our guess than if we ask 10 (sample size)
  - If everyone in the sample gives a similar answer we will be more confident than if we get lots of different answers (sample standard deviation)

- But how much more confident?
  - Inferential statistics provides us with the tools to formalise how our level of confidence should increase or decrease as sample size and sample standard deviation vary.

Central Limit Theorem: what is it?



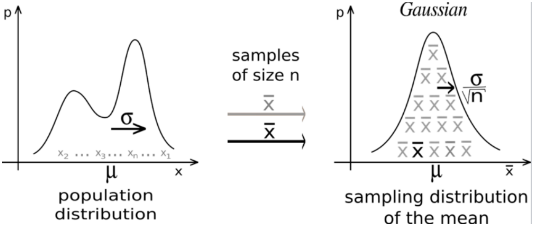
Tim Hesterberg, "What Teachers Should Know About the Bootstrap: Resampling in the Undergraduate Statistics Curriculum"

Central Limit Theorem: what is it?

Consider n random samples taken from a population with mean μ and variance σ^2, and take a mean

$$\bar{x} = (x_1 + x_2 + \ldots + x_n)/n.$$

As n increases, the sampling distribution of $\bar{x}$ is increasingly concentrated around μ and becomes closer and closer to a Gaussian distribution.



Rouaud, Mathieu (2013). Probability, Statistics and Estimation.

Central Limit Theorem: how to use it?

Confidence intervals: We can construct a confidence interval, which gives us a range of values around our estimate for a population parameter, thus expressing how uncertain we are about the estimate

Hypothesis testing: We can ask hypothetical questions about how likely it is that we would get our sample if the population mean is equal to a given value of interest

Confidence Intervals

Confidence Intervals: introduction

- We may be 95% confident that μ (population mean) lies in the interval $(-0.8, 2.5)$.
- We may be 99% confident that σ (population standard deviation) lies in the interval $(1.0, 11.2)$.
- We never construct confidence intervals for statistics (e.g. $\bar{x}$) only for parameters!

Confidence Intervals: introduction

Example

We want to know the average salaries of voters for a given political party.

A survey from 100 randomly selected voters for this party report an average salary of £28,000.

- The average salary of $\bar{x} = 28,000$ is an estimate of the population mean μ (unknown)
- **How close** is the statistics $\bar{x}$ likely to be to μ?

Point estimate $\pm$ margin of error

Confidence intervals for μ (known σ)

- $\bar{x}$ is a sample mean and a point estimate of μ
- Confidence Interval: $\bar{x} \pm$ margin of error

A $(1-\alpha)100\%$ confidence interval for μ (known σ):

$$\bar{x} \pm z_{\alpha/2} \cdot \frac{\sigma}{\sqrt{n}}$$

where

- $z_{\alpha/2}$ taken from the standard normal distribution
- standard error $SE(\bar{x}) = \sigma/\sqrt{n}$

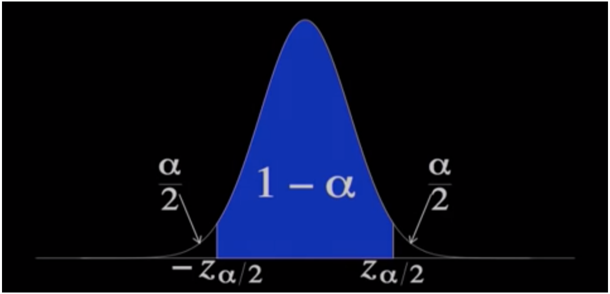
Assumptions:

- simple random sample (very important!)
- normally distributed population (not so important for large samples)
- population standard deviation σ must be known

Confidence intervals for μ (known σ)

A $(1-\alpha)100\%$ confidence interval for μ:

$$\bar{x} \pm z_{\alpha/2} \cdot \frac{\sigma}{\sqrt{n}}$$

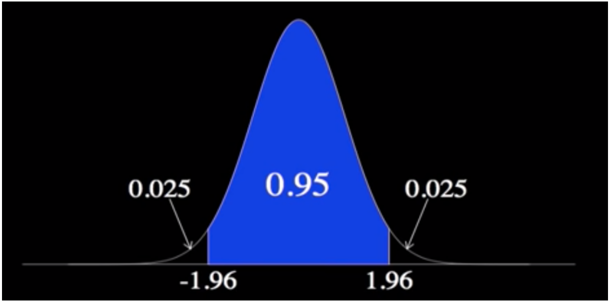


By jbstatistics

Confidence intervals for μ (known σ)

A 95% (i.e. $\alpha = 0.05$) confidence interval for μ:

$$\bar{x} \pm 1.96 \cdot \frac{\sigma}{\sqrt{n}}$$



By jbstatistics

Confidence intervals for μ (unknown σ)

- $\bar{x}$ is the sample mean
- s is the sample standard deviation

A $(1-\alpha)100\%$ confidence interval for μ (unknown σ):

$$\bar{x} \pm t_{\alpha/2} \cdot \frac{s}{\sqrt{n}}$$

where

- $t_{\alpha/2}$ taken from the t-distribution with $n-1$ degrees of freedom
- standard error $SE(\bar{x}) = s/\sqrt{n}$

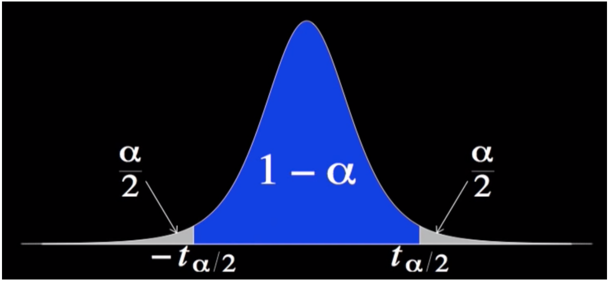
Assumptions:

- simple random sample (very important!)
- normally distributed population (not so important for large samples)

Confidence intervals for μ (known σ)

A $(1-\alpha)100\%$ confidence interval for μ (unknown σ):

$$\bar{x} \pm t_{\alpha/2} \cdot \frac{s}{\sqrt{n}}$$

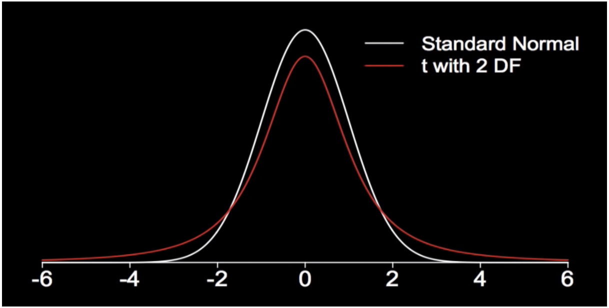


By jbstatistics

Confidence intervals for μ (known σ)

A $(1-\alpha)100\%$ confidence interval for μ (unknown σ):

$$\bar{x} \pm t_{\alpha/2} \cdot \frac{s}{\sqrt{n}}$$

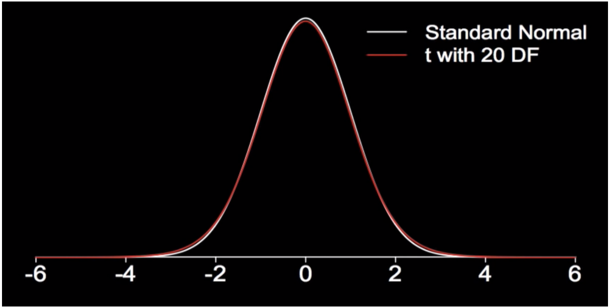


By jbstatistics

Confidence intervals for μ (known σ)

A $(1-\alpha)100\%$ confidence interval for μ (unknown σ):

$$\bar{x} \pm t_{\alpha/2} \cdot \frac{s}{\sqrt{n}}$$



By jbstatistics

Confidence intervals for μ

A 95% confidence interval for μ:

$$\text{sample mean} \pm 1.96 \cdot \text{standard error}$$

- Standard error (SE), i.e. standard deviation of the sampling distribution is typically unknown, but can be estimated with:

$$SE \approx \frac{sample\ \ standard\ \ deviation}{\sqrt{sample\ \ size}} = \frac{s}{\sqrt{n}}$$

- SE measures how much error there is in the sampling distribution
- $s = \sqrt{\sum(x - \bar{x})^2/(n-1)}$, (why $n-1$ and not just n?)

Confidence Intervals: example

Example

We want to know the average salary of voters for a given political party; A survey from 100 randomly selected voters for this party report an average salary of £28,000. Randomly selected 100 voters for this party report an average salary of £28,000 with standard devation £5,000. How can we specify a 95% confidence interval around this value?

- Compute standard error SE

$$SE \approx \frac{sample \;\; standard \;\; deviation}{\sqrt{sample \;\; size}} = \frac{5,000}{\sqrt{100}} = 500$$

- 95% CI is

$$sample \;\; mean \pm 1.96 \cdot SE = 28,000 \pm 980$$

Hypothesis Test

Hypothesis testing

In hypothesis testing, we turn a research question into hypothesis about the value of a parameter(s).

- H_0: **a null hypothesis** (also status quo hypothesis)
- H_a: **an alternative hypothesis** (also research hypothesis)

- Hypothesis testing involves asking a hypothetical question: if the true population value is x, how likely would I be to observe the data in my sample?
- We can measure how far away our sample is from this hypothetical value thanks to the Central Limit Theorem.
- We perform hypothesis test about parameters not statistics. Don't write $(\bar{x}_M = \bar{x}_F)$!

Hypothesis testing: questions

Example: Is the average weight of certain type of chocolate different than the desired 250 grams?

- $H_0 : \mu = 250$
- $H_a : \mu \neq 250$ (or $\mu > 250$)

Example: Do men and women in Oxford have, on average, different hight?

- $H_0 : \mu_M = \mu_F$
- $H_a : \mu_M \neq \mu_F$

Hypothesis testing: enough evidence?

- H_0: **a null hypothesis** (also status quo hypothesis)
- H_a: **an alternative hypothesis** (also research hypothesis)

- The more 'unlikely' our sample is (given our hypothesis), the stronger evidence we have against the null hypothesis.
- If we decide there is enough evidence to reject the null hypothesis, then our results can be labelled **statistically significant**.
- We typically express evidence for or against the null hypothesis as a **p-value**.

Hypothesis testing: what are the p values?

- **What are the p values?**

- p values have a universal meaning across tests.

- p values are often denoted with stars in research papers:

$$^{*}p < 0.05, \qquad ^{**}p < 0.01, \qquad ^{***}p < 0.001$$

- In social science research (and other areas), a p value of less than 0.05 is often the benchmark for rejecting the null hypothesis. Why?

- Types of error

  Type I: falsely rejected the null hypothesis
  Type II: falsely accepted the null hypothesis

Hypothesis testing: example

Example (dehydration)

Does mild dehydration affect reaction time?

Suppose that young women have an average reaction time of 0.95 seconds on certain type of test. The reaction time of 25 dehydrated young women (volunteers) is 1.00 second, with a standard deviation of 0.18 seconds.

Is there an evidence that the true mean reaction time for dehydrated young women is different from 0.95 seconds?

$$H_0 : \mu = 0.95, \qquad H_a \neq 0.95, \quad \alpha = 0.05 \text{ (significance level)}$$

- $\bar{x} = 1.0, \quad s = 0.18, \quad n = 25$
- $t = (\bar{x} – \mu_0)/SE(\bar{x}) = (\bar{x} - \mu_0)/s\sqrt{n} = (1.0–0.95)/(0.18/\sqrt{25}) = 1.39$
- The corresponding p-value=0.1776 (using software)

Hypothesis testing: example

Example (dehydration)

Does mild dehydration affect reaction time?

Suppose that young women have an average reaction time of 0.95 seconds on certain type of test. The reaction time of 25 dehydrated young women (volunteers) is 1.00 second, with a standard deviation of 0.18 seconds.

Is there an evidence that the true mean reaction time for dehydrated young women is different from 0.95 seconds?

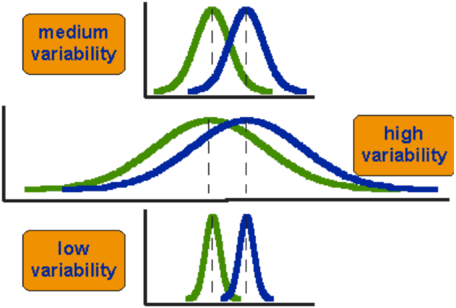
- There isn't strong evidence ($p \approx 0.18$) that the true mean reaction time for dehydrated young women differs from 0.95 seconds.
- Note 1: Young women in this study were not a random sample (they were volunteers)
- Note 2: drawing conclusion about the general cohort of young women is very dubious.

Hypothesis testing: example

Example (video games)

Suppose we want to know whether video game use causes violence. One way of addressing this would be to get a sample of individuals, and work out whether they play video games and measure how 'violent' they are.

- Our estimate of the 'effect' of playing video games would be the size of the difference in mean violent scores between the two groups.
- Our hypothesis test is about establishing whether the difference we observe is statistically significant.
- Is the mean difference far enough away from zero to give me a p value < 0.05?



- In order to express the difference in means between two groups in terms of standard errors, we need to define the **standard error**!
- t-test statistic is the size of the difference in means, measured in standard errors, i.e. (Mean1-Mean2)/SE.
- Larger test statistics are greater evidence against the null hypothesis