Sampling distributions, Estimation, Hypothesis Testing
GEO1001.2020

Clara García-Sánchez, Stelios Vitalis

Resources adapted from:
- David M. Lane et al. (http://onlinestatbook.com)
- Allen B. Downey et al. (https://greenteapress.com/wp/think-stats-2e/)

Lesson A4
Sampling distributions, Estimation

Overview

- Introduction
- Estimation
- Guess the variance
- Degrees of freedom
- Characteristics of estimators
- Sampling Distribution of the mean
- Sampling Distribution of the difference between means
- Confidence interval for the mean
- T distribution
- Common issues

Overview

- **Introduction**
- Estimation
- Guess the variance
- Degrees of freedom
- Characteristics of estimators
- Sampling Distribution of the mean
- Sampling Distribution of the difference between means
- Confidence interval for the mean
- T distribution
- Common issues

Introduction

A critical part of inferential statistics involves determining how far sample statistics are likely to vary from each other and from the population parameter

Sample statistics are the sample means, and the population parameter is the population mean.

Overview

- Introduction
- **Estimation**
- Guess the variance
- Degrees of freedom
- Characteristics of estimators
- Sampling Distribution of the mean
- Sampling Distribution of the difference between means
- Confidence interval for the mean
- T distribution
- Common issues

Estimation

One of the major applications of statistics is estimating population parameters from sample statistics:

For example: a poll may seek to estimate how many adults in a city support a proposition, being 106 out of 200, 0.53 of people support —> this value is called 'point estimate' of the population proportion

Point estimates usually are supplemented by interval estimates called **confidence intervals**.

Confidence intervals —> are intervals constructed using a method that contains the population parameter a specified proportion of time.

For example: in the pollster used a method that contains the parameter 95% of the time it is used then the confidence intervals would be: [0.46,0.60]

Estimation

I think if a distribution, and you have to guess what it is. I tell you it is a normal distribution with a random sample drawn:

[-0.441, 1.774, -0.101, -1.138, 2.975, -2.138]

What do you think is the **μ** parameter of this distribution?

$$\overline{x} = \mathbf{0.155}$$

This process is called **estimation**, and the sample mean is an **estimator**.

Imagine the sample was instead:

Outlier!!!

[-0.441, 1.774, -0.101, -1.138, 2.975, -213.8]

Estimation

Which estimator is best depends on the circumstances, does your distribution have outliers? Are you trying to minimise errors or getting the correct answer?

If there are no outliers, the sample mean minimises the **mean squared error (MSE)** —> If we play the estimation game multiple times, the sample mean minimises:

$$MSE = \frac{1}{m} \sum \left(\overline{x} - \mu\right)^2$$

Where m is the number of times we play the estimation game.

Overview

- Introduction
- Estimation
- **Guess the variance**
- Degrees of freedom
- Characteristics of estimators
- Sampling Distribution of the mean
- Sampling Distribution of the difference between means
- Confidence interval for the mean
- T distribution
- Common issues

Guess the variance

What if we needed to compute the variance now?

[-0.441, 1.774, -0.101, -1.138, 2.975, -2.138]

$$S^2 = \frac{1}{n} \sum (x_i - \overline{x})^2$$

For large samples, S^2 is an adequate estimator, but for small samples it tends to be too low —> it is a biased estimator!

To unbias S^2:

$$S^2_{n-1} = \frac{1}{n-1} \sum (x_i - \overline{x})^2$$

Overview

- Introduction
- Estimation
- Guess the variance
- **Degrees of freedom**
- Characteristics of estimators
- Sampling Distribution of the mean
- Sampling Distribution of the difference between means
- Confidence interval for the mean
- T distribution
- Common issues

Degrees of Freedom

Degrees of freedom: of an estimate is the number of independent pieces of information on which the estimate is based

In general: the degrees of freedom for an estimate is equal to the number of values minus the number of parameters estimated en route to the estimate in question.

For the variance estimation in a sample we need the mean:

$$S_{n-1}^{2} = \frac{1}{n-1} \sum \left(x_i - \overline{x} \right)^2$$

The denominator of this formula is the degrees of freedom!

Overview

- Introduction
- Estimation
- Guess the variance
- Degrees of freedom
- **Characteristics of estimators**
- Sampling Distribution of the mean
- Sampling Distribution of the difference between means
- Confidence interval for the mean
- T distribution
- Confidence interval for the difference between means

Characteristics of estimators

Bias:

Refers to whether an estimator tends to either over or underestimate the parameter.

A statistic is biased if the long-term average value of the statistic is not the parameter it is estimating—> if the mean of the sampling distribution of the statistic is not equal to the parameter

Sampling variability/error:

Refers to how much the estimate varies from sample to sample.

The smaller the standard error, the less the sampling variability. The standard error of the mean is a measure of the sampling variability of the mean:

$$\sigma^2_M = \frac{\sigma^2}{N}$$

The larger N, the smaller standard error —> lower sampling variability

Overview

- Introduction
- Estimation
- Guess the variance
- Degrees of freedom
- Characteristics of estimators
- **Sampling Distribution of the mean**
- Sampling Distribution of the difference between means
- Confidence interval for the mean
- T distribution
- Common issues

Sampling distribution of the mean

The mean of the sampling distribution of the mean is the mean of the population from which the scores were sampled.

If a population has a mean μ, then the mean of the sampling distribution of the mean is also μ.

$$\mu_M = \mu$$

Variance:

The variance of the sampling distribution of the mean:

$$\sigma^2_M = \frac{\sigma^2}{N}$$

σ^2: population variance
N: sample size

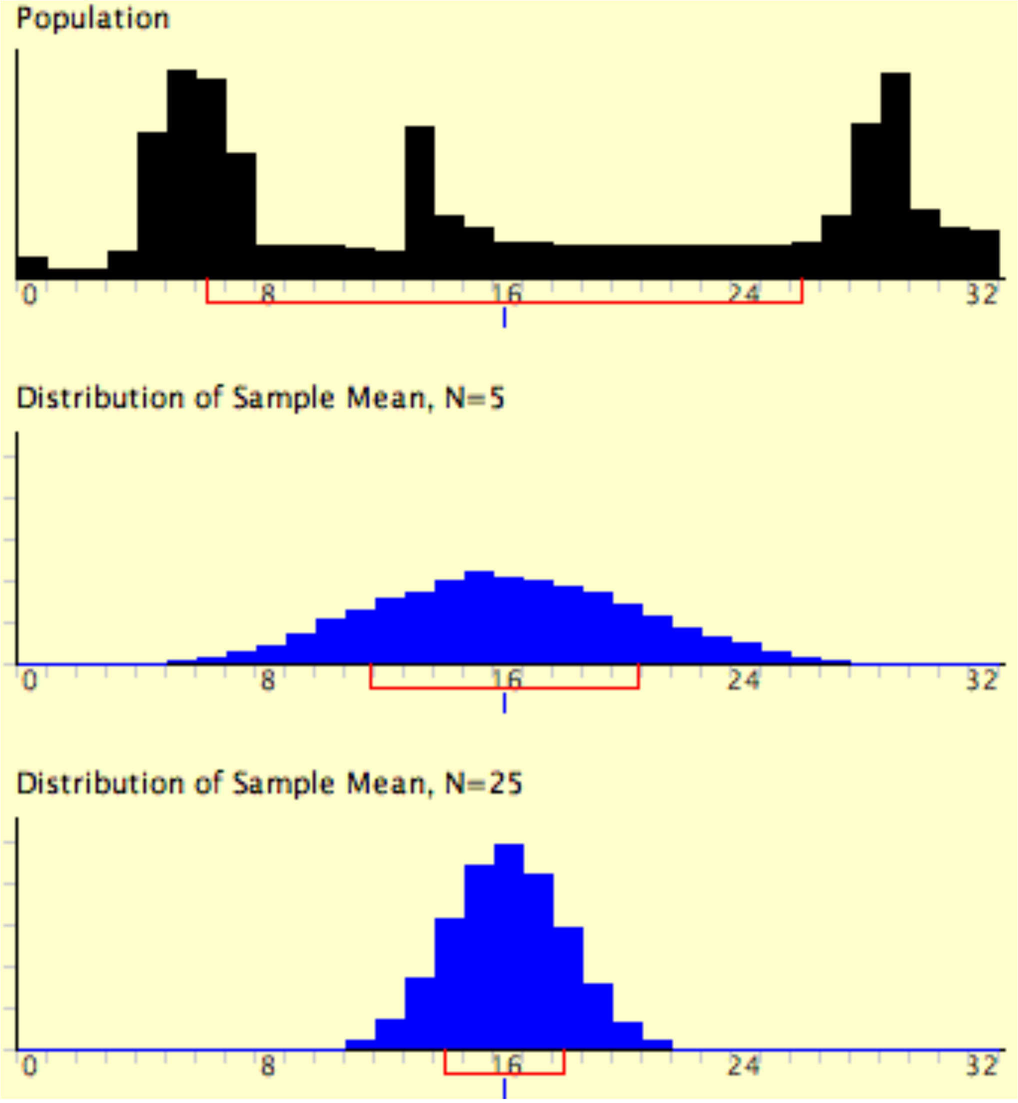
Standard error of the mean:

$$\sigma_m = \frac{\sigma}{\sqrt{N}}$$

Sampling distribution of the mean

Given a population with a finite mean μ and a finite non-zero variance σ^2, the sampling distribution of the mean approaches a normal distribution with a mean μ and a variance of σ^2/N as N, the sample size, increases.

This is regardless of the shape of the parent population!



Overview

- Introduction
- Estimation
- Guess the variance
- Degrees of freedom
- Characteristics of estimators
- Sampling Distribution of the mean
- **Sampling Distribution of the difference between means**
- Confidence interval for the mean
- T distribution
- Common issues

Sampling distribution of the difference between means

Statistical analysis are very often concerned with the *difference between means*.

The sampling distribution of the difference between means can be thought of as the distribution result if we repeat the following steps over and over again:

1. Sample n_1 scores from population 1 and n_2 scores from population 2;
2. Compute the means from the two samples (M_1,M_2)
3. Compute the difference between means M_1-M_2

The distribution of the differences between means is the sampling distribution of the difference between means.

The mean of the sampling distribution of the difference between means is:

$$\mu_{M_1-M_2} = \mu_1 - \mu_2$$

Sampling distribution of the difference between means

The **variance** of the sampling distribution of the difference between means is (variance sum law):

$$\sigma^2_{M_1 - M_2} = \sigma^2_{M_1} + \sigma^2_{M_2} = \frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}$$

The standard error of a sampling distribution is the standard deviation of the sampling distribution, the standard error of the difference between means is:

$$\sigma_{M_1 - M_2} = \sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}$$

Overview

- Introduction
- Estimation
- Guess the variance
- Degrees of freedom
- Characteristics of estimators
- Sampling Distribution of the mean
- Sampling Distribution of the difference between means
- **Confidence interval for the mean**
- T distribution
- Common issues

Confidence interval for the mean

Point estimates usually are supplemented by interval estimates called **confidence intervals**.

Confidence intervals —> are intervals constructed using a method that contains the population parameter a specified proportion of time.

For example: a poll may seek to estimate how many adults in a city support a proposition, being 106 out of 200, 0.53 of people support —> this value is called 'point estimate' of the population proportion

For example: in the pollster used a method that contains the parameter 95% of the time it is used then the confidence intervals would be: [0.46,0.60]

Confidence interval for the mean

- Confidence intervals provide more information than point estimates. They can be computed for diverse parameters.
- Confidence intervals for means are intervals constructed using a procedure that will contain the population mean a specified proportion of the time (typically either 95% or 99% of the time)
- 95% CI == interval with a 0.95 probability of containing the population mean
- Two properties:
  1. Each interval is symmetric about the point estimate
  2. Each interval is contiguous

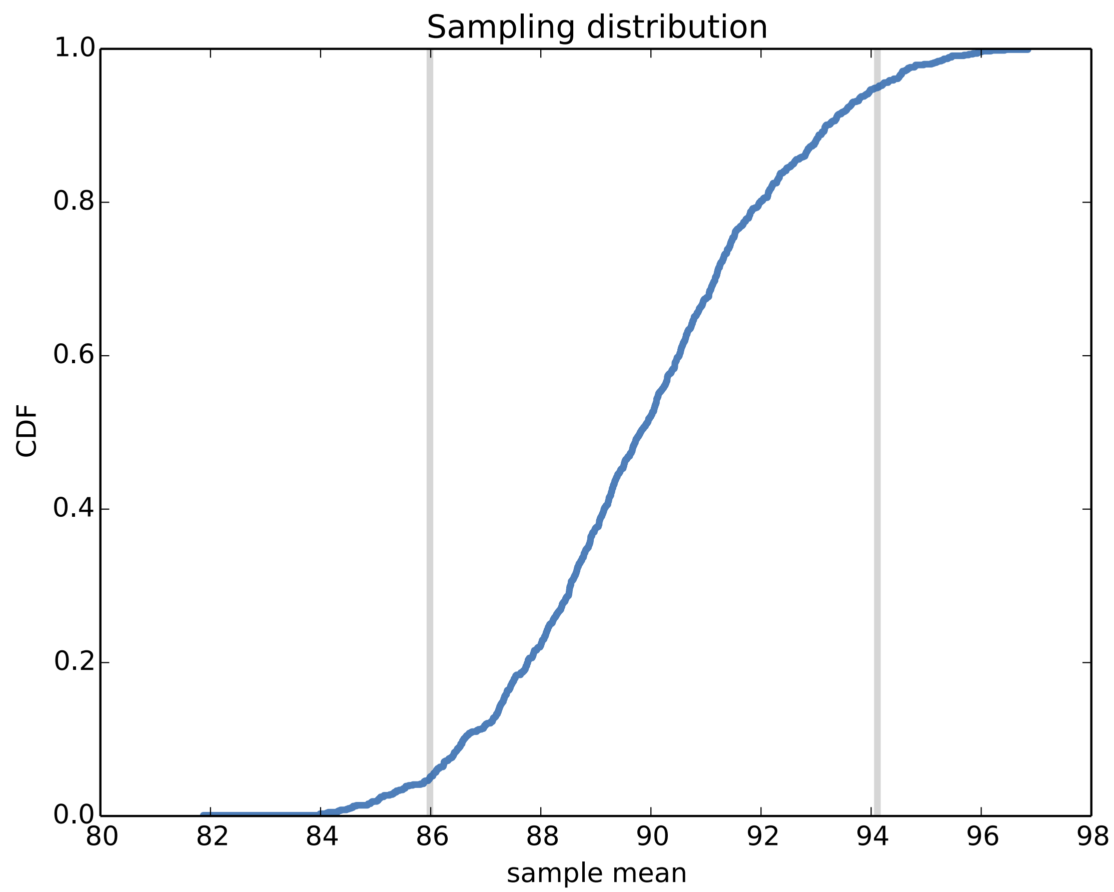


Figure 8.1: Sampling distribution of $\bar{x}$, with confidence interval.

Overview

- Introduction
- Estimation
- Guess the variance
- Degrees of freedom
- Characteristics of estimators
- Sampling Distribution of the mean
- Sampling Distribution of the difference between means
- Confidence interval for the mean
- **T distribution**
- Common issues

T-distribution

- In the early 20th century, by W.S. Gosset, the t-distribution (Student's t distribution) was defined as the **distribution of a mean divided by its estimate of the standard error**.
- It is very similar to the normal distribution when the estimate of variance is based on many degrees of freedom. However, it has more scores in its tails when there are fewer degrees of freedom
- The percentage of the distribution within 1.96σ of the mean is less than 95% for the normal distribution.

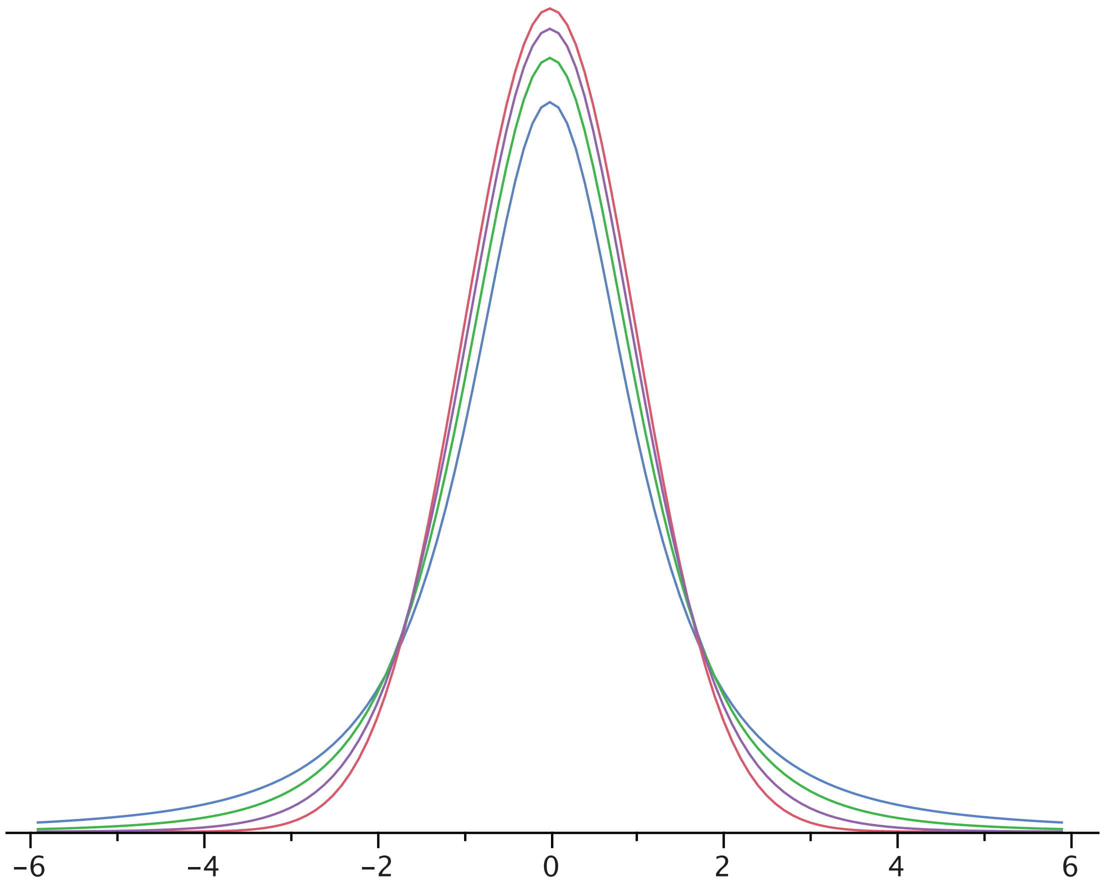


Figure 1. A comparison of t distributions with 2, 4, and 10 df and the standard normal distribution. The distribution with the highest peak is the 2 df distribution, the next highest is 4 df, the highest after that is 10 df, and the lowest is the standard normal distribution.

Overview

- Introduction
- Estimation
- Guess the variance
- Degrees of freedom
- Characteristics of estimators
- Sampling Distribution of the mean
- Sampling Distribution of the difference between means
- Confidence interval for the mean
- T distribution
- **Common issues**

Common issues

- People often **confuse standard error and standard deviation**.
  Remember that standard deviation describes variability in a measured quantity. Standard error describes variability in an estimate.
  One way to remember the difference is that, as sample size increases, standard error gets smaller; standard deviation does not.

- People often think that there is a 90% probability that the actual parameter, μ, falls in the 90% confidence interval. Sadly, that is not true. If you want to make a claim like that, you have to use Bayesian methods.
  The sampling distribution answers a different question: it **gives you a sense of how reliable an estimate is by telling you how much it would vary if you ran the experiment again**.

Practice

We want to determine the average height of students in Geomatics. For that we randomly chose 4, with heights: 160,154, 170, 182 cm.

1) What is the sample standard deviation?
2) What is the standard error of the mean?

Lesson A4
Hypothesis Testing

Overview

- Introduction
- Significance testing
- Type I and II errors
- One- and two- tailed tests
- Interpreting results
- Steps in hypothesis testing
- Significance testing and confidence intervals

Overview

- **Introduction**
- Significance testing
- Type I and II errors
- One- and two- tailed tests
- Interpreting results
- Steps in hypothesis testing
- Significance testing and confidence intervals

Introduction

Hypothesis testing is a statistical procedure for testing whether **chance** is a plausible explanation of an **experimental finding**. Given a sample and an apparent effect, what is the probability of seeing such an effect by chance? Some steps to answer:

- Quantify the size of the apparent effect by choosing a test statistic. For example the *difference in means between two groups*.
- Define a *null hypothesis*, which is a model of the system based on the assumption that the apparent effect is not real. Null hypothesis is that there is no difference between both groups.
- Compute a *p-value*, which is the probability of seeing the apparent effect if the null hypothesis is true.
- Interpret the result. If the **p-value is low**, the **effect is said to be statistically significant, which means that it is unlikely to have occurred by chance**. In that case we infer that the effect is more likely to appear in the larger population.

<u>James Bond example:</u>

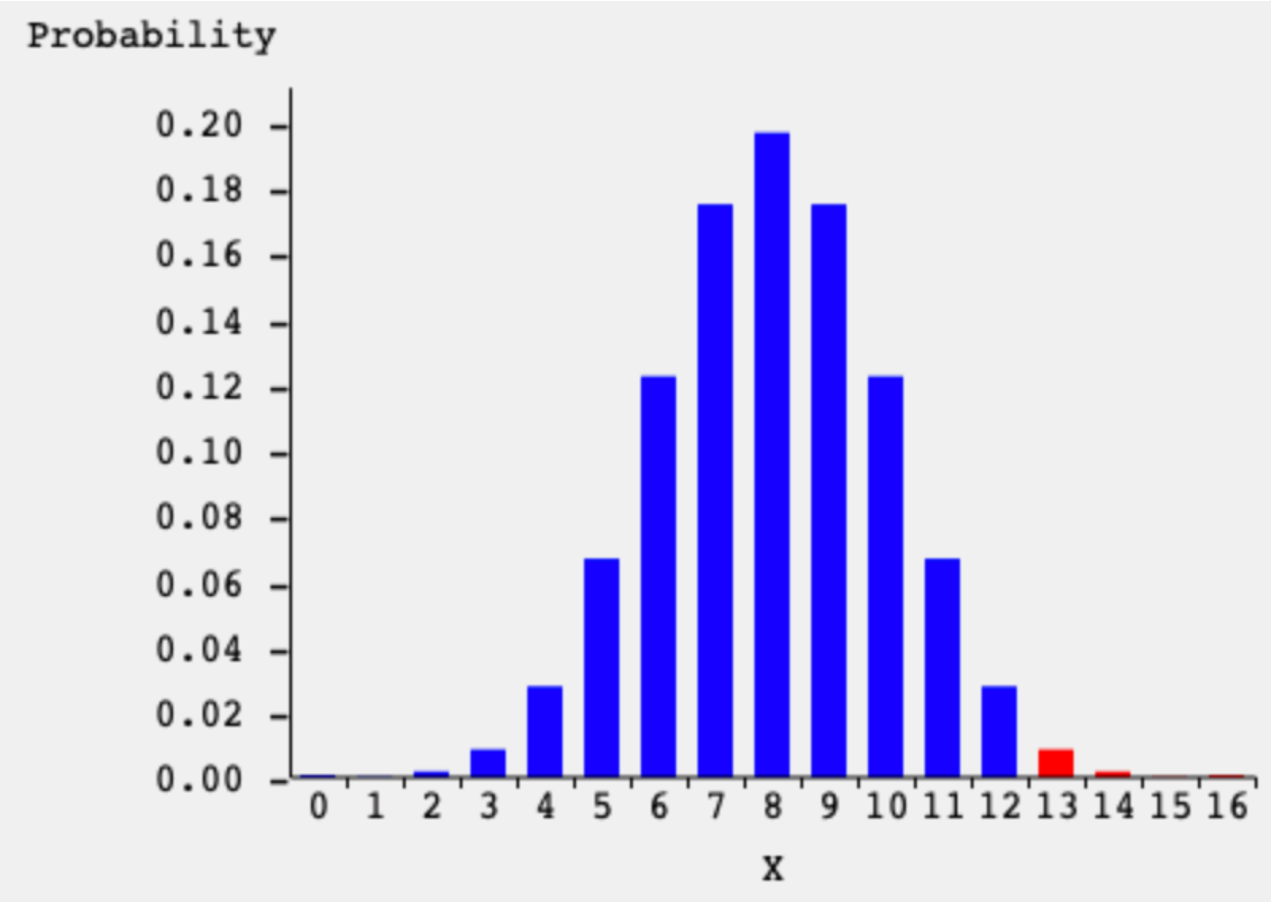
Determine whether Mr. Bond can tell the difference between a shaken and a stirred martini. Suppose we gave Mr. Bond a series of 16 taste tests. In each test, we flipped a fair coin to determine whether to stir or shake the martini. Then we presented the martini to Mr. Bond and asked him to decide whether it was shaken or stirred. Let's say Mr. Bond was correct on 13 of the 16 taste tests. *Does this prove that Mr. Bond has at least some ability to tell whether the martini was shaken or stirred*?

$$P(x) = \frac{N!}{x!(N-X)!}\pi^{x}(1-\pi)^{N-x}$$

$$P(>13) = P(13) + P(14) + P(15) + P(16) = 0.0106$$

<u>James Bond example:</u>

Determine whether Mr. Bond can tell the difference between a shaken and a stirred martini. Suppose we gave Mr. Bond a series of 16 taste tests. In each test, we flipped a fair coin to determine whether to stir or shake the martini. Then we presented the martini to Mr. Bond and asked him to decide whether it was shaken or stirred. Let's say Mr. Bond was correct on 13 of the 16 taste tests. Does this prove that Mr. Bond has at least some ability to tell whether the martini was shaken or stirred?



Introduction

The probability value:

It is important to understand that the computed probability is the probability James Bond will be correct 13 out of 16 times if he were just guessing

The null hypothesis:

The hypothesis that an apparent effect is due to chance is called “the null hypothesis”, which can be written as:

$$\mu_{population1} - \mu_{population2} = 0$$

The null hypothesis is typically the opposite of the researcher’s hypothesis. If the null hypothesis is rejected, the alternative hypothesis is accepted

Overview

- Introduction
- **Significance testing**
- Type I and II errors
- One- and two- tailed tests
- Interpreting results
- Steps in hypothesis testing
- Significance testing and confidence intervals

Significance testing

A low probability value casts doubt on the null hypothesis, but how low must the probability value be in order to conclude that the null hypothesis is false?

Most common values: **0.05** or **0.01**, and it is called "significance level (α)" —> then we can say that the effect is "statistically significant"

NOTE: keep in mind that statistical significance means only that the null hypothesis of exactly no effect is rejected; it doesn't mean that the effect is important, which is what "significant" usually means.

Statistical significance ≠ Practical significance

Overview

- Introduction
- Significance testing
- **Type I and II errors**
- One- and two- tailed tests
- Interpreting results
- Steps in hypothesis testing
- Significance testing and confidence intervals

Type I and II errors

Type I:

Type I error occurs when a significance test results in the rejection of a true null hypothesis.

This type of error is affected by the α level —> the lower the level the lower the Type I error rate.

Type II:

Type II error is failing to reject a false null hypothesis. Thus, it is not really an error, because when a statistical test is not significant, it means that the data do not provide strong evidence that the null hypothesis is false.

The researcher should consider the test inconclusive.

If the null hypothesis is false, the probability of a Type II error is called β.

Overview

- Introduction
- Significance testing
- Type I and II errors
- **One- and two- tailed tests**
- Interpreting results
- Steps in hypothesis testing
- Significance testing and confidence intervals

One- and two-tailed tests

Going back to our James Bond case, we saw a binomial distribution where the red bars show the values greater or equal to 13.

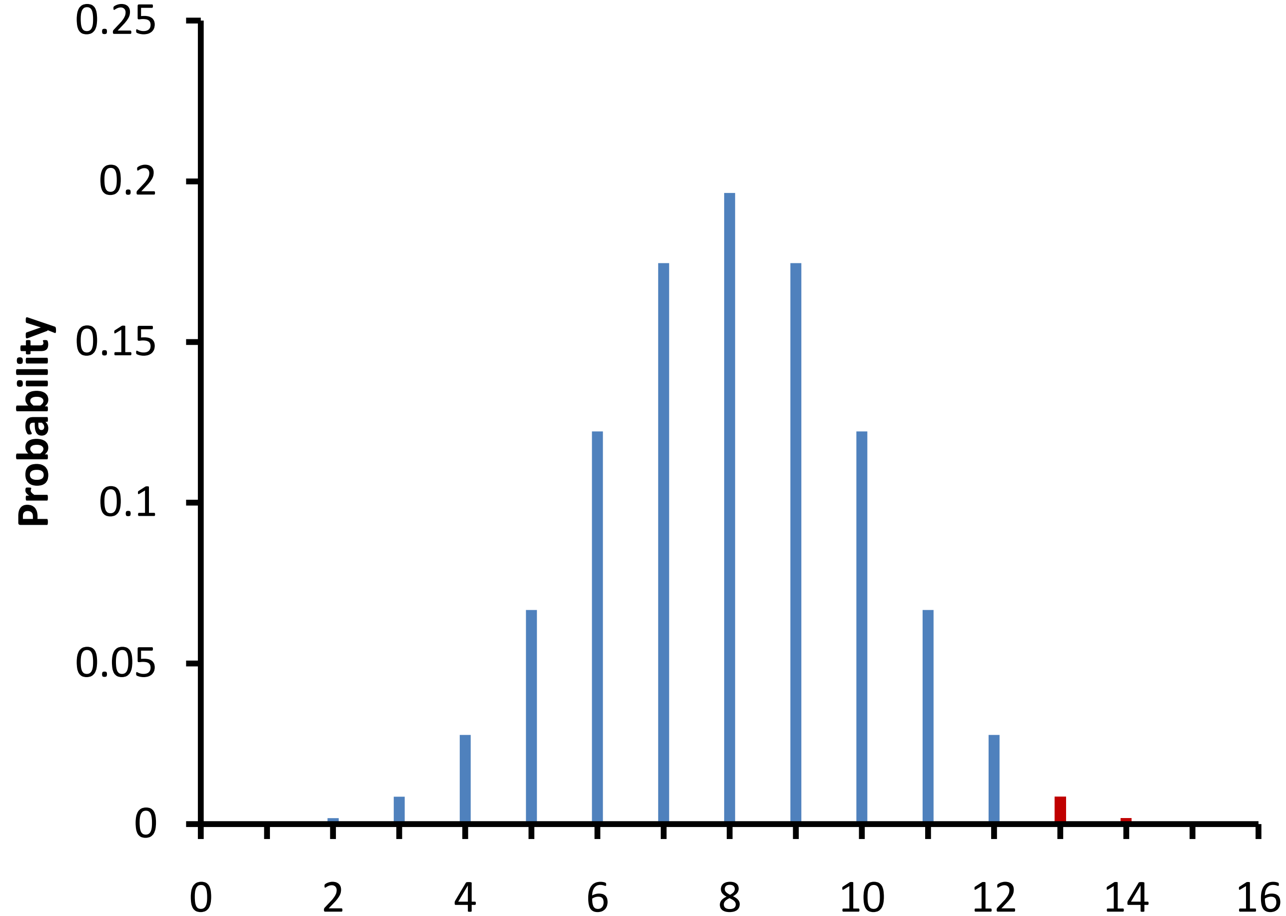


Figure 1. The binomial distribution. The upper (right-hand) tail is red.

A probability calculated in only one tail of the distribution is called a "one-tailed probability".

One- and two-tailed tests

A slight different question can be asked of the data: What is the probability of getting a result as extreme or more extreme than the one observed? What is the probability of 3/13?

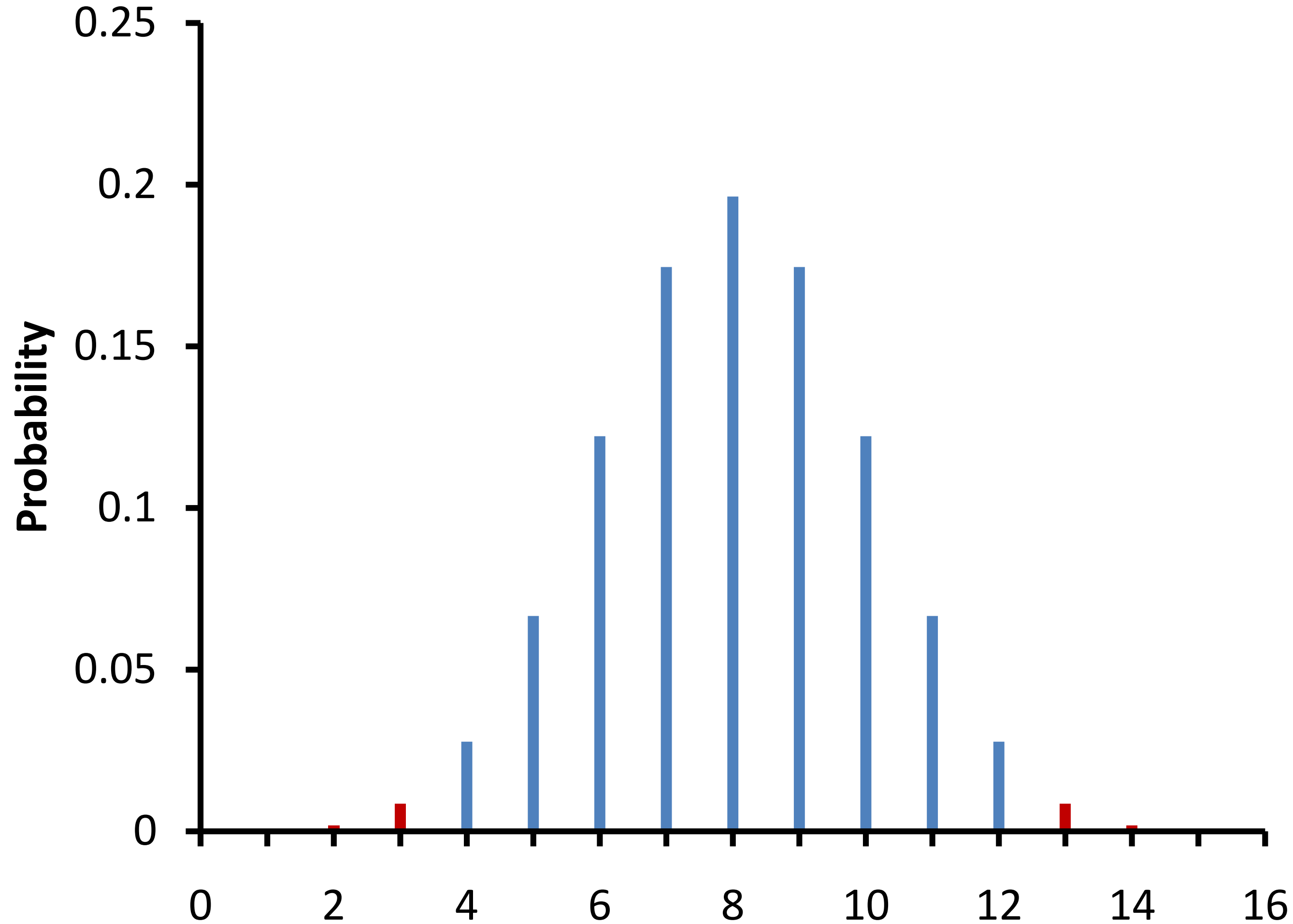


Figure 2. The binomial distribution. Both tails are red.

A probability calculated in two tails of the distribution is called a "two-tailed probability". This probability is exactly the double:

$$0.0106*2=0.0212$$

One- and two-tailed tests

Should the one-tailed or two-tailed probability be used to assess Mr. Bond's performance? It depends on the question:

1. Can Mr. Bond tell the difference between shaken or stirred Martinis?

    —> We would conclude he could if he performed either much better than chance or much worse than chance. If he performed much worse than chance, we would conclude he can tell the difference but he doesn't know which is which —> **two-tailed**

2. Is Mr. Bond better than chance at determining whether a Martini is shaken or stirred?

    —> In this case we will use **one-tailed** probability. If Mr. Bond guess only 3 out 16 correct, this will give high probability and the null hypothesis will be rejected

One- and two-tailed tests

Which one is more common in science?

Two-tailed tests

—> Because an outcome signifying that something other than chance is operating is usually worth noting.

Overview

- Introduction
- Significance testing
- Type I and II errors
- One- and two- tailed tests
- Interpreting results
- **Steps in hypothesis testing**
- Significance testing and confidence intervals

Steps in hypothesis testing

1. Specify a null hypothesis:

—> two-tailed test: μ_1-μ_2=0

—> one-tailed test: μ_1-$\mu_2 \geqslant 0$

2. Specify the significance level: α=0.05 or α=0.01

3. Compute the probability value (p-value): the probability of obtaining a sample statistic as different or more different from the parameter specified in the null hypothesis given that the null hypothesis is true.

4. Compare the p-value with the α-level:

- p-value < α —> reject the null hypothesis
- p-value > α —> findings are inconclusive, we can't reject, and we don't have support for the null hypothesis either

Overview

- Introduction
- Significance testing
- Type I and II errors
- One- and two- tailed tests
- Interpreting results
- Steps in hypothesis testing
- **Significance testing and confidence intervals**

Significance testing and confidence intervals

There is a close relationship between confidence intervals and significant tests:

1. If a statistic is significantly different from 0 at the **0.05 level** then the **95% confidence interval** will not contain 0.

2. Similarly for **99% confidence interval** and significance at the **0.01 level**.

If the 95% CI contains zero (more precisely, the parameter value specified in the null hypothesis) —> then the effect will not be significant at the 0.05 level.

Misconceptions

<u>Misconception 1:</u> The probability value is the probability that the null hypothesis is false.

> The probability value is the probability of a result as extreme or more extreme given that the null hypothesis is true. It is the probability of the data given the null hypothesis, not the probability that the null hypothesis is false.

<u>Misconception 2:</u> A low probability value indicates a large effect.

> A low probability value indicates that the sample outcome would be very unlikely if the null hypothesis were true. A small probability value can occur with small effect sizes, particularly if the sample size is large.

<u>Misconception 3:</u> A non-significant outcome means that the null hypothesis is probably true.

> A non-significant outcome means that the data do not conclusively demonstrate that the null hypothesis is false.

Practice

Testing a difference in means with the NSFG data.
Evaluate the hypothesis:
1) if the mean pregnancy length for the first babies is slightly longer;
2) the mean birth weight is slightly smaller.

Use the tools inside scipy.stats library from python.

https://3d.bk.tudelft.nl/courses/geo1001/

Questions?