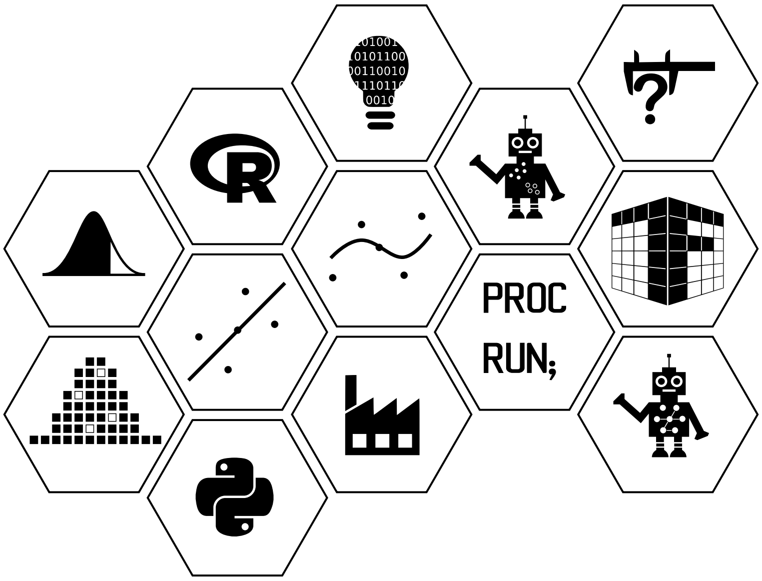
Learning from Data/Data Science Foundations

Week 3: Statistical learning



DATA ANALYTICS GLASGOW

Statistical learning

In the previous two weeks we have considered data, data sources and structures, ways to collect and collate data and ways to summarise, visualise and describe data. However, it is usually of interest to go beyond simply identifying data sources and using simple summaries and visualisations. There are two main reasons for this, and we'll re-visit our initial diagram from week 1 to describe these.

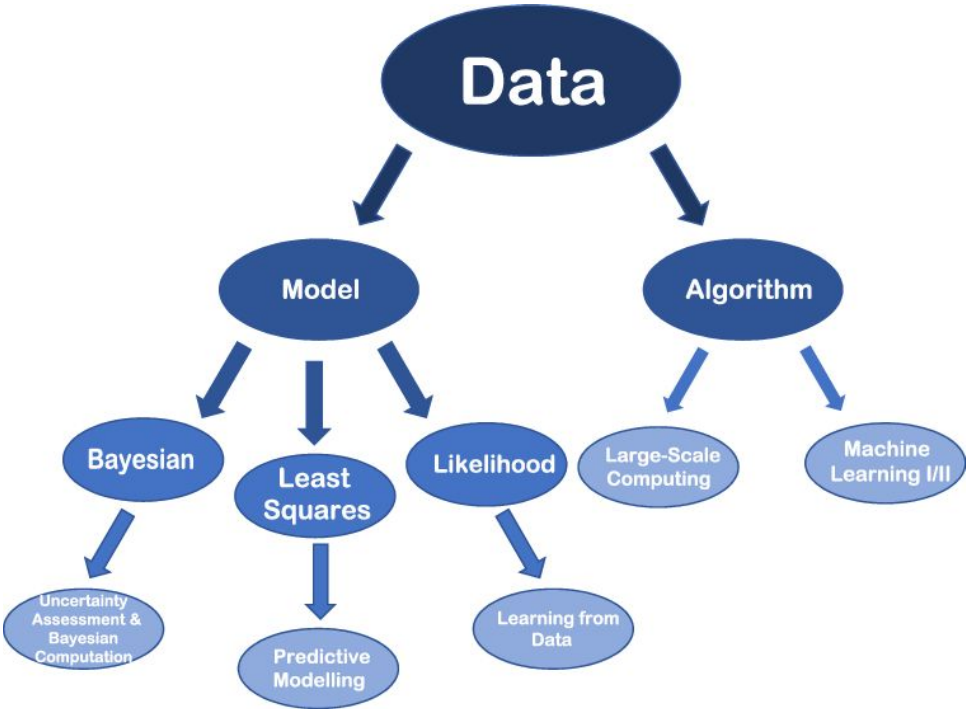


Figure 1: *Approaches to Learning from Data*

If we are able to source and collect data on all individual elements, objects or people that we are interested in, i.e. we have data from our whole *population* of interest, then simple summaries or visualisations may enable us to answer questions that we are interested in regarding the data. However, it maybe that the data dimension in such a situation has extremely high volume i.e. that it is highly dimensional and approaches are required in order to identify and extract patterns and relationships from the data. In such a situation one option would be to use algorithms in order to *learn* from the data, as in the right section of our diagram in Figure 1.

However, we are often in the situation where even if we have a very large amount of data it is not possible to collect data on all elements, objects or individuals of interest, as we described in week 1. Therefore, in order to draw conclusions and make inferences using the data we can use a probability model (as in the left section of our diagram in Figure 1) and hence statistical inference to help extract information, and identify appropriate patterns and relationships. As we'll see in the material for this week, an important aspect of statistical inference is that it enables us to account for uncertainty in the estimated results that we obtain.

The week 3 material will formally introduce the key terminology that we will use throughout the remainder of the course to enable us to *learn from data* using statistical inference. We'll start off by formally defining and extending some of the terminology that has been introduced throughout weeks 1 & 2.

In weeks 1 & 2 we also introduced that in order to answer specific questions it's important to have an in-depth understanding of your data through understanding the context, identifying data features, summarising and visualising your data. This process is often referred to as learning using **descriptive statistics**. Numerical summaries and plots of your data can help you to identify an appropriate probability model for your data. Once we have estimates from a sample of data using summary statistics, if we assume that our data have arisen from a particular probability distribution then hypothesis testing and interval estimation are approaches used within statistical inference to enable us to draw generalisable conclusions from our data while accounting for varibility and hence uncertainty.

The second part of week 3 will introduce the ideas of **inferential statistics** using hypothesis testing and interval estimation.

Week 3 learning material aims

The material in week 3 covers:

- what is statistical inference?
- statistical terminology;
- a framework for hypothesis testing;
- stating appropriate null and alternative hypotheses;
- terminology for hypothesis testing;
- interpreting confidence intervals and p-values.

Introduction : What is statistical inference?

The science of Statistical Inference (or the science of 'Statistics' for short) provides concepts, approaches and methods for studying 'situations' subject to unexplained variability. Statistical inference is used in a wide range of fields and variety of applications. For example, covering challenges such as:

- investigating air pollution and its effect on human health;
- estimating level of risk for financial investment products;
- investigating brain response to external stimuli.

Variability

A key problem is that outcomes or data collected to answer our questions of interest typically cannot tell us about all *individuals* of interest. We typically only have information on a smaller number of *individuals* (that it was possible to investigate as part of the study). The data are therefore subject to elements of randomness and uncertainty based on the *individuals* that we have actually investigated.

Uncertainty can also be induced through the measurement process of the data or algorithm used to generate the data.

For example,

- Every subject in a study will return slightly different results under the same conditions.
- Two different groups of the British population are likely to contain a different proportion of people that will vote Labour at the next election (by chance).
- All things being equal, share prices of energy companies will be at different levels.
- Brain signals will be masked by noise from surrounding variation.

We must take account of this variability and uncertainty by making assumptions about it and using probability theory.

Probability models provide descriptions of such variability.

Statistical inference aims to describe (i) the typical behaviour of a group of *individuals*; (ii) the variability among the *individuals* in the group.

The basic approach to Learning from Data

In week 1 we started an outline for an approach to *learn from data*, and for us this leads us through the important steps for *statistical inference*.

1. Specify clearly each question of interest;
2. Design a suitable means of gathering or collating appropriate data to answer the question posed (this usually has to ensure that the data are representative of all *individuals* under study);
3. Plot or tabulate, and summarise the data (exploratory analysis) in an appropriate form as part of quality assuring the data, and to get a subjective answer to the question posed;

We are now going to focus on how we use probability distributions in order to describe how the data we have collected were actually generated i.e. accounting for variability, and this will enable us to answer questions of interest and provide appropriate measures of uncertainty attached to our estimated conclusions.

Therefore the remaining steps in the approach are:

4. Build a probability model to describe how the data actually collected were generated (use your detailed exploratory analysis to inform this);

5. Analyse the data using this model; and

6. Report, in simple English, the answers from such an analysis exploiting effective visualisation of the data (and the statistical modelling results) to allow ease of interpretation of the conclusions of your analysis.

The key to analysing data using statistical inference is to be able to communicate the approach and results effectively and simply with your colleagues, stakeholders, the general public etc.

As we introduced in week 1, advances in technology have meant that we are surrounded by data, and there are millions of data elements being collected on all aspects of daily life every day. Often we are interested in empirical prediction and uncertainty quantification from noisy data rather than issues of study design or the measurement process. Reliable and robust analysis of such data requires both computational approaches for the analysis of such data and key knowledge of the underlying fundamental statistical principles required to extract scientific truth in the presence of uncertainty.

The basic vocabulary/jargon for statistical inference

We have already briefly introduced the ideas of collecting data from a **population**, and using a **sample** of data to give us information from the whole **population**. We'll now formally define terminology for these aspects to enable us to generalise to any study and to perform statistical inference and hence statistical learning.

Specifically, we'll consider the terms:

- variable
- population
- sample
- parameter
- statistic

This video describes the basic approach for learning from data using statistical inference and provides illustrations for the statistical vocabulary above:

Video

Statistical inference and terminology

Duration 6:47

A **Variable** is a single aspect or characteristic associated with each of a large group of *individuals* under consideration, and can take on different values for different *individuals*.

A **Population** is the large group of *individuals* under consideration.

A **Sample** is a subset of the population *selected* as being representative of the population and each of whom will have the variable of interest measured.

A **Parameter** is a single number summarising in some way the values of the variable of interest in the population (e.g. population mean, population proportion).

A **Statistic** is a single number summarising in some way the values of the variable of interest in the sample actually collected (e.g. sample mean, sample proportion). It provides an initial estimate (which we'll refer to as a *point estimate* for a *population parameter*).

Now let's consider the terminology above for a specific example:

Example 1

Cars

We might be interested in investigating the average car price (in U.S. dollars) in the United States of America in 2002.

Variable: The price of a car (in U.S. dollars).

Population: All cars in the United States of America in 2002 (approximately 253 million).

Sample: A random sample of 1000 cars from across the States in 2002.

Parameter: The mean (or median) car price of all types of cars in America in 2002.

Statistic: The mean (or median) car price of all types of cars in the sample in 2002.

An additional example for statistical terminology based on a blood pressure application is given in this video:

Video

Statistical terminology - an example

Duration 1:29

NOTE: we can never evaluate a parameter (unless we have data from our whole population); we shall only infer about the value of the parameter from the actual value of a suitable statistic (or statistics) evaluated in the observed sample.

Task 1

For the question below, try to define the terms: *variable*, *population*, *sample*, *parameter* and *statistic (estimate)*:

What percentage of the Scottish population will vote Labour in the next general election?

Summary - what is Statistical inference?

The major aim is to draw inferences (i.e. make conclusions) about the value of parameters in the population of interest based upon the observed values of the summary statistics from the actual sample obtained.

Inferential statistics - hypothesis testing and interval estimation

Once we have identified a specific question of interest for a study, collated our data, performed data quality checks and explored our data using graphical and numerical summaries, we can consider the type of statistical inference that we will perform.

The remainder of this course will develop theory to estimate population parameters. A point estimate (statistic) on its own is of little use - some measure of its precision is also necessary i.e. to account for variability and hence uncertainty in the point estimate produced. Hypothesis testing and interval estimation will be introduced as two approaches to make inferences about parameters for a population.

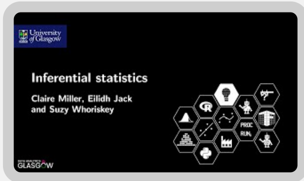
In order to form conclusions about a particular study, we might be interested in testing a hypothesis about a specific population parameter or calculating a range of plausible values for a population parameter given the data that we have observed and any assumptions that we have made about the distribution of the data.

This video introduces and illustrates the hypothesis testing framework for a water quality example.

Video

Inferential statistics

Duration 7:23



Hypothesis Testing

In hypothesis testing we start with some default theory, called a null hypothesis, and we ask if the data provide sufficient evidence to reject the theory. If not we do not reject the null hypothesis.

Well-conducted research studies often begin with a question about a population. Here are some examples of *questions of interest*:

- Is there a difference in the amount of water cats drink if the water is still or flowing?
- Is the average level of sulphur dioxide in the air in Europe less than the daily average limit of 125 $\mu g/l$?
- In 1972, the proportion of adults in the UK earning over 20K GBP per annum was 15%, this figure had risen to 45% by 1992. Has this figure risen further since 1992?

We might say that we wish to investigate the plausibility of a hypothesis (or statement) about the population. Here are some possible hypotheses.

- There is a difference in the median amount of water cats drink when the water is flowing and the water is still.
- The mean level of sulphur dioxide in air in Europe is less than the environmental daily average limit of 125 $\mu g/l$.
- The proportion of UK adults earning over 20K GBP is now greater than 45%.

These are all statements about what researchers expect to find out. They are usually saying that something has changed, or is different, or can be made different. These are study hypotheses, which are often denoted H_1. Hypotheses are usually written in terms of population values (i.e. parameters). We will use η as a symbol for the population median, μ as a symbol for the population mean and θ as the symbol for a population proportion.

- $H_1 : \eta \neq 0$ where η is the population median difference between the amount of flowing and still water drunk by cats.

- $H_1 : \mu < 125$ where μ is the population mean level of sulphur dioxide ($\mu g/l$) in the air in Europe.

- $H_1 : \theta > 0.45$ where θ is the population proportion of UK adults who earn more than 20K GBP.

We use the data collected from a sample to assess the evidence in favour of the study hypothesis. It is necessary also to consider what would be true if the study hypothesis were not true. We can frame this statement about the population as a hypothesis, called the **null hypothesis** (H_0). H_0 usually includes the possibility that nothing has changed, or two things are equal.

- $H_0 : \eta = 0$ (the median difference in the amount of water)

- $H_0 : \mu \geq 125$ (the environmental daily average limit)

- $H_0 : \theta \leq 0.45$ (the proportion of adults who earn more than 20K GBP)

On the basis of the evidence provided by the sample data, we must decide which of the null and study hypotheses is the more plausible. For this reason, the study hypothesis (H_1) is often called the **alternative hypothesis**.

It might seem as though we ought to treat the null and alternative hypotheses equally, but we do not. Null hypotheses usually, by their nature, cannot be proven to be true unless we have access to data from the whole population.

When assessing the evidence from a sample we have just two possibilities. Either we reject H_0 in favour of H_1 or we do not reject H_0. We can never prove H_0 from a sample, although often we do not have sufficient evidence to reject it.

Generally hypotheses fall into two main classes: one-sided and two sided alternative hypotheses.

One-sided

- H_0: $\mu \geq 125$, H_1: $\mu < 125$

- H_0: $\theta \leq 0.45$, H_1: $\theta > 0.45$

Two-sided

- H_0: $\eta = 0$, H_1: $\eta \neq 0$

Assessing the evidence against H_0

For each of the questions of interest identified above we are interested in calculating information about a *population* i.e.

- The median difference in the amount of water drunk.
- The mean level of SO_2 in air.
- The proportion of working adults in the UK that earn over 20K GBP.

In each case we can summarise the evidence from our sample by computing summary statistics e.g. we can compute the median difference in the amount of flowing and still water drunk by a representative sample of cats. Such summary information can form the basis of our **test statistics** for each question that we are interested in. We saw in the first part of this week that a statistic is a function of the data in the sample and that this value will change from sample to sample. The observed value for each of the examples above (i.e. median difference in amount of water) in our samples of data can form the basis of our **observed test statistics**.

A test statistic enables us to carry out a **hypothesis test**, which is a formal procedure to decide between the null and alternative hypotheses. The strength of the evidence against H_0 (and for H_1) is judged using a **p-value**, which is determined on the basis of the observed value of the test statistic (i.e. using the information contained in the data and an assumed probability model).

The **p-value** is the probability of obtaining a value for your test statistic that is at least as extreme as the observed value of the test statistic (*assuming H_0 is true*).

The p-value is an attempt to measure the consistency of the data with the null hypothesis.

1. Low p-values imply that the data are improbable if H_0 is true. Since the data actually happened, this suggests rejecting H_0 in favour of H_1.
2. High p-values suggest that the data are quite probable under H_0. Since the data and hypothesis appear consistent, H_0 is not rejected.

How low should a p-value be in order to cause rejection of H_0? There is no *correct* answer to this, and many statisticians, data analysts or data scientists prefer not to make one up, simply quoting the p-value instead. If a value, α, is chosen, such that H_0 will be rejected if the p-value is $< \alpha$ then α is known as the **significance level** of the test. Traditional choices for α are 0.05 (5%) and 0.01 (1%), with 0.05 being the most common. (Note: in practice p-values around α are borderline and treated with caution -

for examples 0.049 and 0.051, there is very little difference in these values - should you reject or not reject?)

The usefulness of p-values as a measure of the evidence against H_0 stems from their general applicability: a p-value of 0.04 has the same meaning whatever hypothesis is being tested and whatever method is being used for the test, and that meaning is well defined.

Supplement 1

p-values are a useful source of evidence. However, they should be interpreted in the correct manner and should be used in conjuction with other evidence (after carefully considering what can be seen from the data) before drawing conclusions.

p-values in the news!: See the supplementary material below for recent discussion on the use/misuse of p-values.....

Psychology journal bans p-values: **p-values1**

p-values tip of iceberg: **p-values2**

The p-value statement 5 years on: **p-values3**

The data for the tasks below are contained in the Rdata object for week 3, which can be found at: RData, under the data objects, **IQ** and **blooddata** respectively.

Task 2

We have IQ measurements on a random sample of 39 people and we are interested in assessing if there is any evidence from the data to suggest that the average IQ of the general population of all people is different from 100. Carry out your initial impressions and state your hypotheses.

Two sided test:

H0 = 100
H1 != 100

Task 3

Phenylthiocarbamide (PTC) is a chemical that tastes bitter to some people ("tasters") and bland to everyone else ("non-tasters"), a response which is controlled by two forms of the TAS2R38 gene. A sample of individuals from across the world had their tasting status determined.

Initially, we are interested in answering the following question:

- What percentage of living humans are tasters?

The dataset, **blooddata**, contains the tasting status in column "Taster" ("Yes" = taster; "No" = non-taster).

Carry out your initial impressions and state your hypotheses, where we are interested in assessing if the population proportion of tasters is greater than 0.5.

Confidence intervals

Interval estimation is the use of sample data to calculate an interval of possible (or probable) values for an unknown population parameter e.g. population mean or population proportion. The most prevalent forms of interval estimation are:

- confidence intervals (a Frequentist method - see below); and
- credible intervals (a Bayesian method, which will be introduced very briefly as supplementary material for this course, and developed more fully in courses to follow e.g. **Uncertainty assessment and Bayesian computation**, see Figure 1) [1].

Here we will focus initially on *Frequentist* statistics.

Frequentist statistics considers the 'limit' of the relative frequency of an event in a large number of trials. For example, a 95% confidence interval is interpreted thus:

Under repeated sampling and recalculation, 95% of confidence intervals would contain the true population value.

A 95% confidence interval will not always contain the true value of the parameter. In fact, on average only 95% of such intervals will do so.

On 95% of the occasions on which a 95% confidence interval is calculated from sample data, it will contain the true value of the parameter.

To illustrate this, the figure below shows 2 examples of 100 realisations of a 95% confidence interval for a given population mean. In each plot in the figure, the 95% confidence intervals have been constructed from 100 random samples each of size 50 from the same population. If we randomly choose one realisation, the probability is 95% that we choose an interval that contains the true population mean of 20. The plots in this figure demonstrate the coverage property of confidence intervals. Averaging over many samples, 95% of the 95% confidence intervals constructed will capture the true population mean. And 5% will not!

Don't worry about the details of how the confidence interval is being computed, this will be covered in detail in week 4.

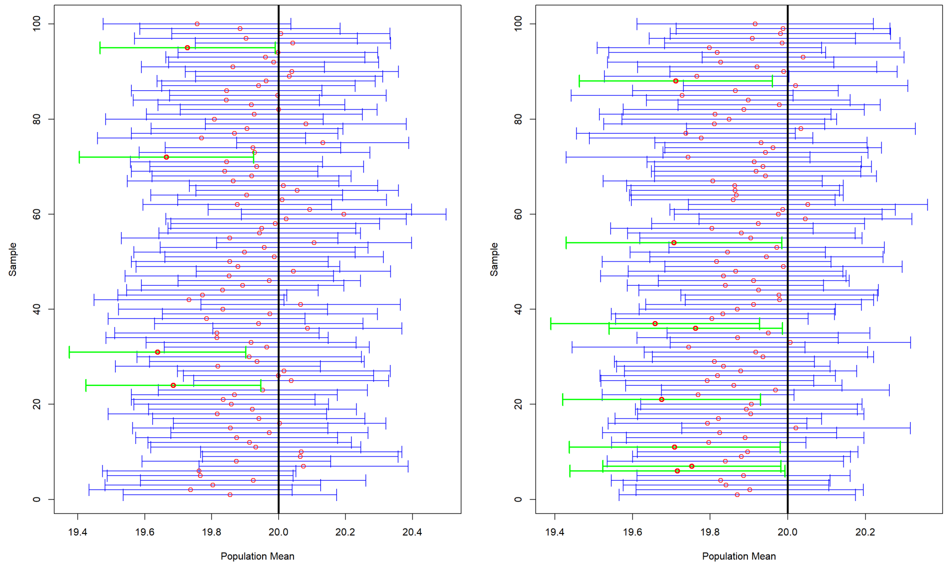


Figure 3

In the plot on the left, 4 of the intervals do not contain the true population mean (and are also highlighted in green). In the plot on the right, 8 of the intervals do not contain the true population mean (and are also highlighted in green).

Supplement 2

A full visualisation to help with the interpretation of confidence intervals is available at: confidence intervals

Let's look at a simple example here to start to introduce the interpretation for confidence intervals and p-values.

Example 2

Let's return to our example on IQ measurements and firstly test the following hypotheses:

H_0: $\mu = 100$, H_1: $\mu \neq 100$, where μ is the population mean IQ.

Using the IQ data here, and assuming that our data have arisen from a normal probability distribution, we obtain a p-value of < 0.001, and a 95% confidence interval for the population mean of (90.38, 93.97), with a point estimate for μ of 92.

When we have a very small p-value we often represent it as $<$ to indicate that the probability is very small. The accuracy produced here will depend on the software used and hence it's often more desirable to not report it exactly. Here, since the p-value is very small, if we use the rule-of-thumb of 5% then we would reject the null hypothesis. Here we conclude that there is evidence that the population mean is not equal to 100.

We can go further than this by producing a confidence interval i.e. a range of plausible values for the population mean accounting for uncertainty in our estimate. In this case a 95% confidence interval for the population mean is (90.38, 93.97), with an estimate for the population mean of 92. This interval tells us that (using these data and under our assumption of a normal probability distribution), the population mean IQ is highly likely to lie between 90 and 94 with a best guess of 92. Since 100 is not contained in the interval, then it is not a plausible value for the population mean IQ, given our data and assumed normal distribution.

If these results were not what we would have expected for the IQ of the general population then we would now need to think again about the context of our data and the data collection process. Perhaps the data were not representative of the general population, perhaps the sample size was too small.

This final example here has provided an introduction to illustrating the ideas and interpretation of results from hypothesis testing and confidence intervals. In week 4 we will consider how to construct hypothesis tests and confidence intervals in detail for simple scenarios for population means and proportions. We'll do this from a theoretical point of view and also through using functions within `R`, and we'll use specific data examples to illustrate further interpretation of the results.

Learning outcomes for week 3

By the end of week 3, you should be able to:

- define and contrast the terms population, sample, parameter and statistic;
- formulate null and alternative hypotheses;
- distinguish between hypotheses that require one and two-sided tests;
- define the following terms in the context of hypothesis testing: null and alternative hypotheses, test statistic, observed test statistic, significance level, p-value;
- define a 95% confidence interval.

Review exercises, selected video solutions and written answers to all tasks/review exercises are provided overleaf.

Review exercises

Task 4

For each of the following state the population of interest, a possible sample, the population parameter of interest and the sample statistic of interest.

1. What is the average height of adult males in France?

2. What proportion of people have debt of more than 10,000 GBP in the UK?

3. What is the median income in Scotland?

Task 5

Write down appropriate null and alternative hypotheses corresponding to each of the following research questions. Write your hypotheses in words and symbols.

1. A certain brand of chocolate is sold in bars that nominally weigh 200g. On average, do packs weigh less the nominal amount?

2. The median Intelligence Quotient (IQ) in the general population is 100. Is the median IQ of prisoners who have been convicted of theft different from this normal value?

3. Historical records show that in a certain region only 10% of business start ups survive their first year. Substantial government funds are invested in the area. Does the proportion of start up businesses which survive after their first year improve following the investment?

Task 6

For each of the examples in Task 5 state whether the alternative hypothesis is one or two-sided.

Task 7

Let's return to our example on IQ and test the following hypotheses:

H_0: $\mu = 100$, H_1: $\mu \neq 100$, where μ is the population mean IQ.

Suppose that we have a new sample of data from the population, and using it (and assuming a normal probability distribution for the data) we obtain a p-value of 0.323 when testing the above hypotheses and a confidence interval of (98.32, 102.44), with a point estimate for μ of 100.

What would the conclusions be this time?

Answer 1

What percentage of the Scottish population will vote Labour in the next general election?

- Variable: whether or not a person will vote for Labour
- Population: all people eligible to vote in Scotland
- Sample: a representative selection of the Scottish population that we can interview e.g. 200 people.
- Parameter: the proportion of people in Scotland that will vote for Labour.
- Statistic (point estimate): the proportion of people in the sample that will vote for Labour.

Each sample will give a slightly different estimate for the proportion of people that will vote for Labour, which induces variability and hence uncertainty around our point estimate. We can account for this by using an appropriate probability model to report the uncertainty in our estimate.

Answer 2

IQ:

```
IQ <- read.csv("IQ.csv")
library(ggplot2)
library(gridExtra)
```

```
boxplot <- ggplot(IQ, aes(x="", y=iq))+
  geom_boxplot()+
  theme(axis.title.x=element_blank(),axis.text.x =
element_blank(),
        axis.ticks.x = element_blank())+
  ylab("IQ measurements")

hist<-ggplot(IQ, aes(x=iq))+
  geom_histogram(binwidth=5, fill="white", color="black")+
  xlab("IQ measurements")



grid.arrange(hist, boxplot, ncol=2)
```

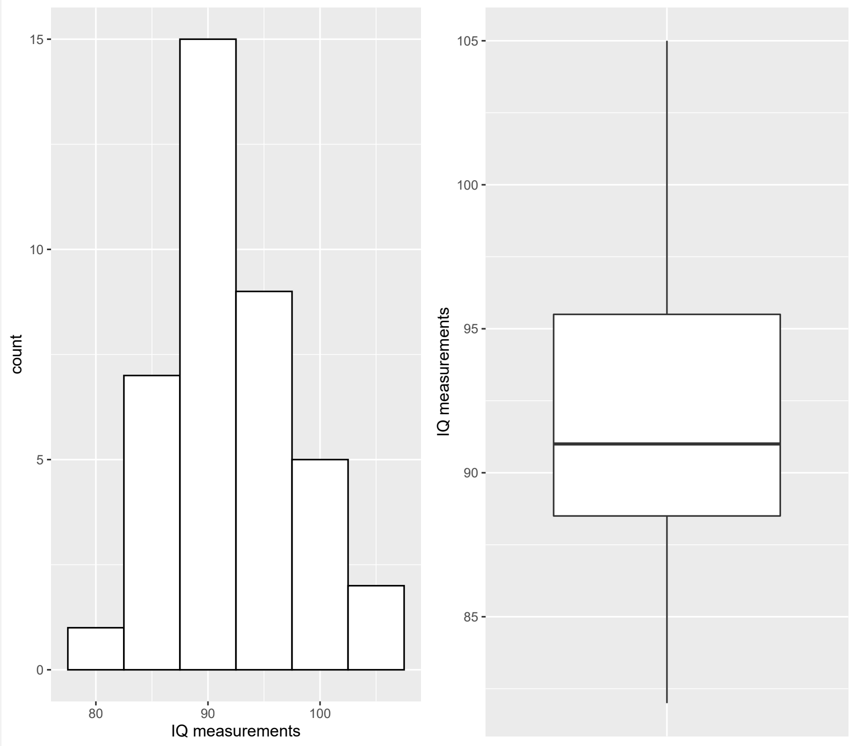


Figure 2

```
summary(IQ$iq)
```

```
R Console
Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
82.00   88.50   91.00   92.18   95.50  105.00
```

Firstly, the histogram shows that the IQ levels are generally lower than 100. There appear to be one or two observations greater than 100. This is emphasised in the summary statistics. Here, we get a numerical summary and see that the median IQ level of our sample is 91, and the mean 92.18. (Side note: the fact that our mean and median are almost equal indicates that our data are not highly skewed). Looking at the boxplot, we see that, again, the IQ levels seem to be below 100. This will need further investigation.

To investigate this, our hypotheses are: H_0: $\mu = 100$, H_1: $\mu \neq 100$, where μ is the population mean IQ.

Answer 3

Taste test:

```
table(blooddata$Taster)
```

R Console

```
No Yes
57 156
```

```
prop.table(table(blooddata$Taster))
```

R Console

```
No        Yes
0.2676056 0.7323944
```

The table of raw counts tells us that in our sample of 213 individuals, 57 cannot taste PTC and 156 can. Turning this into a much more informative proportion, we can say that of our sample, 26.8% are non-tasters and 73.2% are tasters. A more formal analysis will need to be carried out to apply this sample to the worldwide population of PTC tasters. We will come to this later in the course.

To do so, we state our hypothesis.

To investigate this, our hypotheses are: H_0: $\theta \leq 0.5$, H_1: $\theta > 0.5$, where θ is the population proportion of people that are tasters.

Answer 4

Describing population/sample/parameter/statistic:

Scenario 1

Population: All adult males in France.

Sample: A representative sample of, say, 2000 adult males from France.

Parameter: The (population) mean height of all adult males in France.

Statistic: The (sample) mean height of the 2000 adult males in the sample.

Scenario 2

Population: Everyone in the UK that is eligible to have credit (assuming we are considering legal debt).

Sample: A representative sample of, say, 2000 people in the UK that are eligible to have credit and are representative of all social classes.

Parameter: The (population) proportion of everyone in the UK that is eligible to have credit that has more than 10,000 GBP of debt.

Statistic: The (sample) proportion that has more than 10,000 GBP of debt in our sample.

Scenario 3

Population: Everyone of working age in Scotland.

Sample: A representative sample of, say, 2000 people from Scotland that are of working age.

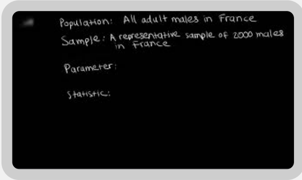
Parameter: The (population) median income of everyone in Scotland of working age.

Statistic: The (sample) median income of our representative sample.

Video

Video model answers for Task 4 Scenario 1

Duration 2:06



Answer 5

Hypotheses:

(a) Let μ be the population mean weight (g) of these bars of chocolate. Then:

$\mathrm{H}_0 : \mu \geq 200$ $\mathrm{H}_1 : \mu < 200$

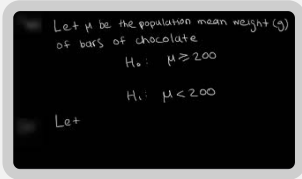
(b) Let η be the population median IQ of prisoners convicted of theft. Then:

$\mathrm{H}_0 : \eta = 100$ $\mathrm{H}_1 : \eta \neq 100$

(c) Let θ be the true proportion of start up businesses surviving beyond their first year. Then:

$\mathrm{H}_0 : \theta \leq 0.1$ $\mathrm{H}_1 : \theta > 0.1$

Video



Video model answers for Task 5 parts (a) and (b)

Duration 2:23

Answer 6

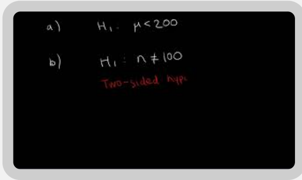
The alternative hypothesis is:

(a) One-sided

(b) Two-sided

(c) One-sided

Video



Video model answers for Task 6

Duration 2:37

Answer 7

Since the p-value is greater than 0.05 then we would not reject H_0 (at 5% significance). Therefore, we conclude that there is insufficient evidence that the population mean is different from 100. Note: we can only 'reject' or 'do not reject' H_0. We cannot prove that H_0 is true from a sample of data.

The confidence interval tells us that it is highly likely that the population mean IQ lies between 98 and 102, with a best guess of 100. Therefore, here the value of 100 is consistent with what we have from the data and is a plausible value.

Footnotes

1. Bayesian inference is a method of statistical inference in which evidence is used to estimate parameters and predictions. Additional available information can be accounted for besides that provided by the data e.g. from previous studies. It is a statistical method for using data to update prior beliefs. Bayesian inference is a modern statistical approach typically requiring vast amounts of computer power. ↩