Introduction to statistical inference

Austin Haffenden

Adapted from OpenIntro Statistics 2nd Edition
Diez, D.M., Barr, C.D. and Çetinkaya-Rundel, M.

Statistical inference

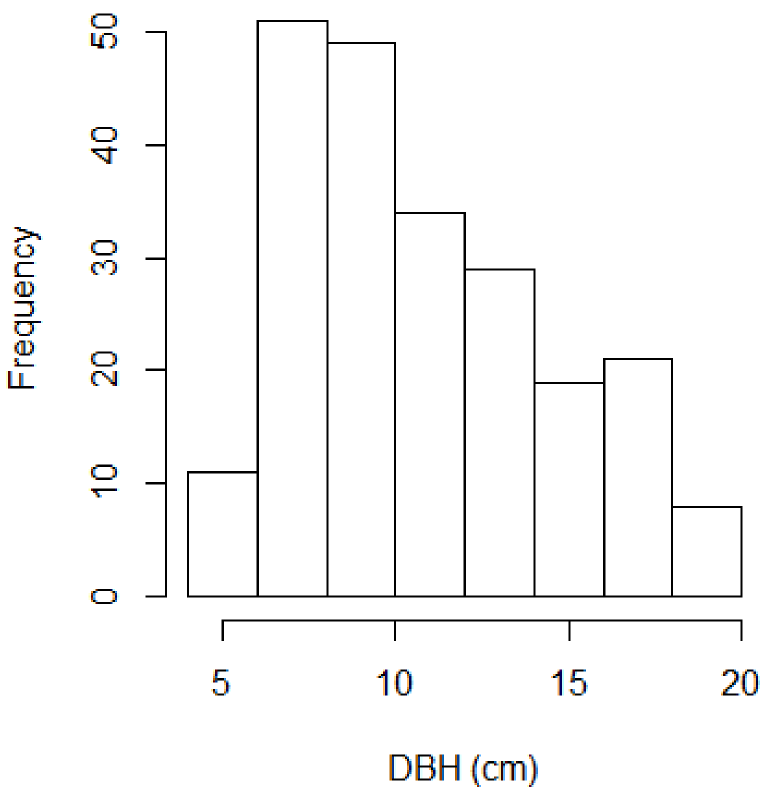
- Understanding quality of parameter estimates
  - e.g. “How sure are we that the estimated mean is representative of the population mean”
- We will discuss inference in relation to the population mean (μ)

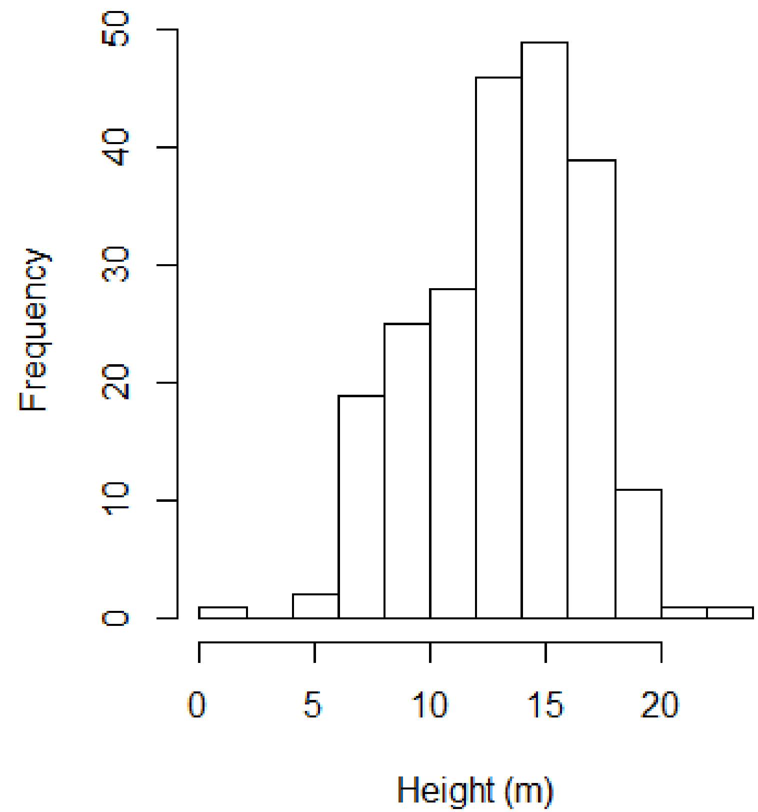
CyB survey data

Variable	**Description**
no	Tree number
DBH	Diameter at breast height (cm)
height	Tree height (m)
species	Tree species code

no	**DBH**	**height**	**species**
1	6.4	6.7	DF
2	5.1	6.9	DF
3	7.8	6.9	DF
.	.	.	.
222	10.3	8.7	DF

Population visualisation





Random sample of the data

- Take a random sample of 100 rows
- Visualise DBH and height of sample with histogram
- Will use this sample to draw conclusions about the entire population (full dataset)
- Broadly – this is statistical inference

Variability in estimates
Point estimates

- Will estimate:
  - Mean DBH (cm) of all trees
  - Mean height (m) of all trees
- Sample mean is a point estimate
- Estimates vary between samples
  - Create new sample: sample2
  - Compare samples 1, 2 and the population using summary
- Variation suggests point estimate not exact

Point estimates are not exact

- Estimates improve with more data
  - Plot the running mean for sample1
- Sample point estimates vary between samples
  - Plot running mean of sample2 in red on same plot
- It is useful to quantify variation of estimate between samples

Standard error of the mean

- Sample mean for sample1 and sample2 are different
- If we sample many times all will potentially be different but will fit a normal distribution around the population parameter
- Can build a **sampling distribution** by sampling many times from the same population

Sampling distribution

- Represents distribution of point estimates of fixed sample size from a population
- In this case
  - Point estimates: sample means
  - Fixed sample size: 100
  - Population: All DF DBHs in CyB_data
- Always think of the distribution when think about the sample mean
- Central to understanding statistical inference

Sampling distribution

- In R build and plot histogram of 1000 samples from the population data
- Sampling distribution:
  - Unimodal
  - Approximately symmetrical about mean
- Plot mean of sample1 and sample2 on the histogram
- **Always think of the distribution when think about the sample mean**

Standard error of sample mean

- Describes a point estimate's typical error or uncertainty
- Is the standard deviation associated with the point estimate
- Standard error is:

$$SE = \frac{\sigma}{\sqrt{n}}$$

- Where:
  - σ is the standard deviation of the Population
  - n is the sample size

Sample standard error

- Population standard deviation usually unknown
- Use point estimate of the standard deviation from the sample
  - Sample must be sufficiently large ≈ 30 cases
  - Without strong skew
    - If skewed need larger sample

Basic properties of point estimates

- Used to estimate population parameters
  - Single believable parameter value
- Not exact
  - Varies between samples
  - Usually some error
- Sensible to provide a plausible range of values for the parameter

Confidence intervals

- As discussed – point estimate is a single value for a parameter – associated error
- We can describe range of possible values
- Plausible range is called a ‘Confidence interval’

Approximate 95% confidence interval

- Built around point estimate
  - Mean in this case – as most plausible value
- Standard error – the uncertainty with the point estimate – guides size of confidence intervals
- Approx. 95% of time the point estimate will be within 1.96 standard errors of the parameter:

$$point\ estimate \pm 1.96\ * SE$$

(1.96 is the approximate value of the 97.5 percentile)

What this means

- We are 95% confident that the population parameter is between the confidence intervals
- This is not a probability – it is only how plausible it is that the parameter is between the CIs
- CIs only capture the population parameter – not about individual observations or point estimates
- Confidence intervals only attempt to capture population parameters

Let’s visualise this

- If take many samples and build confidence intervals from each one
- ≈ 95% will contain the parameter mean
- ≈ 5% will not
- R:
  - Take 25 random samples
  - Calculate mean and CI
  - Plot with parameter mean
  - Don’t panic all written for you
- Then try with 100 samples

Hypothesis testing

- Do the DBHs of trees from the DF species differ from those of the HWD species?
- Simplify the question into two competing **hypotheses**
  - *H0*: The average DBHs are the same for both species
  - *HA*: The average DBHs are different for both species
- *H0* is the null hypothesis
- *HA* is the alternate hypothesis

Null and alternate hypotheses

- The null hypothesis (*H0*):
  - Often represents a sceptical perspective or one of no difference
- The alternative hypothesis:
  - Represents an alternative claim to be considered. Often a new perspective or the possibility of change.
- The sceptic won't reject *H0* unless the evidence for *HA* is so strong that they have to

Testing hypotheses using confidence intervals

- Can start the hypothesis evaluation by comparing point estimates of the means
- Difference in means may be due to *sampling variation*
  - The variability between random samples
- Does the HWD mean DBH fall within the Confidence Intervals of the DF DBH?

Result

- If the HWD mean DBH falls between the Confidence Intervals of the mean DF DBH:
  - we ‘fail to reject the null hypothesis’

- This double negative language is used to communicate that while we are not rejecting a position, we are also not saying that it is correct

Differences in tree height

- Create a null and alternative hypotheses to test for differences in mean height between DF and HWD species
- Test those hypotheses using confidence intervals as you did for DBH

BREAK

Decision Errors

- Hypothesis tests are not flawless
- Humans make wrong decisions in statistical hypothesis tests
- However we can quantify how often we make these mistakes

Decision Errors

		Test conclusion	
		Don't reject *H0*	Reject *H0* for *HA*
Truth	*H0* true	OK	Type 1 Error
	HA true	Type 2 Error	OK

- Type 1 Error: rejecting the null hypothesis when *H0* is true
- Type 2 Error: failing to reject the null hypothesis when *HA* is true

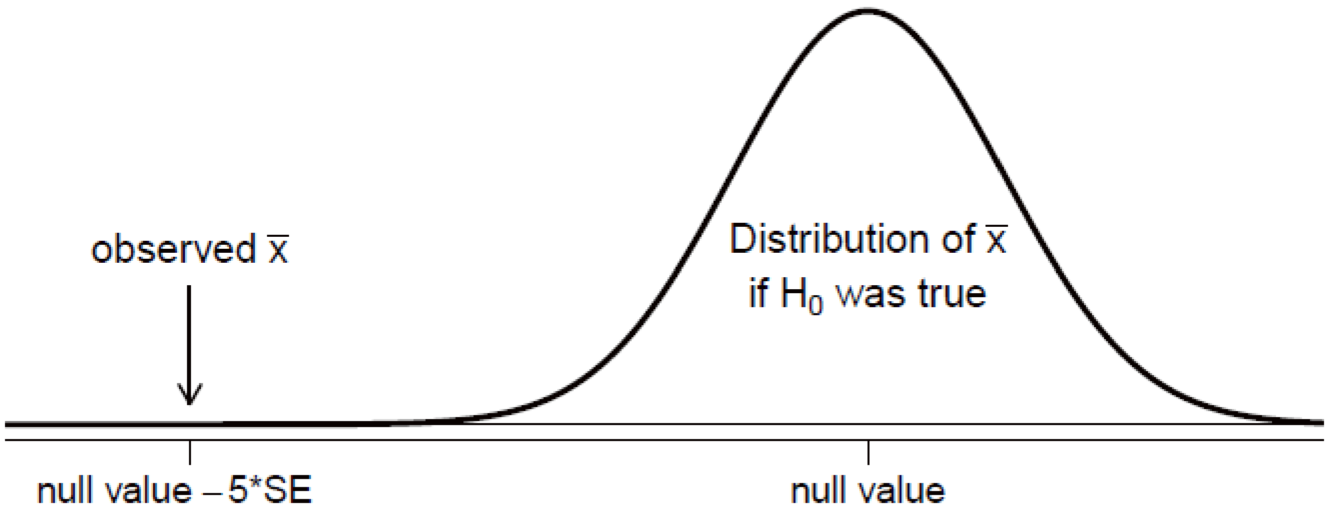
Significance Levels

- Hypothesis testing involves rejecting or failing to reject the null hypothesis
- Don’t want to incorrectly reject *H0* more than 5% of the time.
- Significance level of 5% ; $\alpha = 0.05$
- With 95% confidence interval we make an error when point estimate is 1.96 SE from the population parameter
  - Approx. 5% of the time; 2.5% in each tail

Simplistic

- Confidence interval is simplistic hypothesis test
- Can use the **p-value** to extend hypothesis tests
  - Can quantify similarity or difference
  - Useful when confidence intervals can't be constructed
- Way to quantify strength of evidence against the null hypothesis and in favour of the alternative – is a conditional probability

Quantify Strength of evidence against null hypothesis

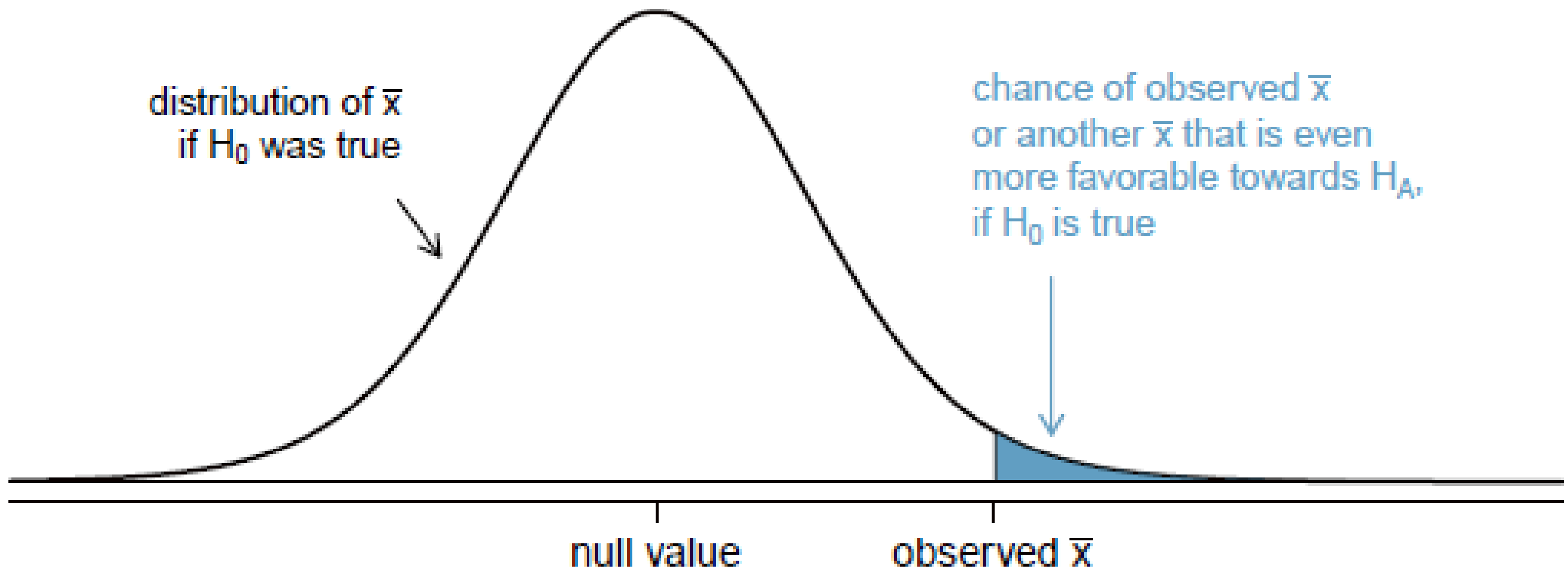


Testing using p-values

- p-value definition:

The p-value is the probability of observing data at least as favourable to the alternative hypothesis as our current data set, if the null hypothesis is true.

We typically use a summary statistic of the data, here the sample mean, to help compute the p-value and evaluate the hypotheses.



- p-value constructed so can compare with the significance level
- To identify p-value the distribution of the sample mean is considered as if the null hypothesis was true
- The p-value is then defined as the probability of the observed sample mean or a sample mean even more favourable to the alternate hypothesis under this distribution

- If the null hypothesis is true, how often should the p-value be < 0.05%

Ideas behind evaluating hypotheses with p-values

- The null hypothesis represents a sceptic's position or a position of no difference. We reject this position only if the evidence strongly favours HA.
- A small p-value means that if the null hypothesis is true, there is a low probability of seeing a point estimate at least as extreme as the one we saw.
  - We interpret this as strong evidence in favour of the alternative.
- We reject the null hypothesis if the p-value is smaller than the significance level, which is usually 0.05. Otherwise, we fail to reject H0.

Conclusions

- Point estimates are not exact – they come from a **sampling distribution** of possible values
- Standard Error describes a point estimates uncertainty
- Confidence Intervals provide a range of plausible values for the **population** parameter

Conclusions (cont.)

- Failing to reject the null hypothesis is not the same as accepting it
- Type 1 Error:
  - Rejecting the null hypothesis when *H0* true
- Type 2 Error:
  - Failing to reject the null hypothesis when *HA* is true
- p-value quantifies strength of evidence against the null hypothesis