EDC459: Week 4
Inference

Dr. Calum Webb
Sheffield Methods Institute, the University of Sheffield.
c.j.webb@sheffield.ac.uk

Sign In

Learning Objectives

What will I learn? | How does this week fit into my course?

By the end of this week you will:

- Get to grips with the intuition behind using inferential statistics for hypothesis testing.
- Be able to interpret a p-value from a hypothesis test, in conjunction with a critical value.
- Be able to judge when the use of hypothesis testing for generalisation of findings is appropriate depending on the kind of sample our data is from.
- Understand the intuition of some common statistical tests for testing the relationship between two variables.

Learning Objectives

What will I learn? | How does this week fit into my course?

- Hypothesis testing is an essential skill for doing quantitative social research, and plays to the strengths of quantitative methods for identifying social patterns.
- In order to accurately assess the results from reading other social science research publications, you must be able to interpret the results from statistical significance tests (and p-values) — even if you don't do quantitative research yourself!
- You should have a good sense of the theory behind hypothesis testing to ensure that you use it responsibly and effectively in a research career, including if you are in a leadership role.

Inferential statistics for hypothesis testing

Variable Type	Nominal	Ordinal	Continuous
Nominal	Chi-squared Test of Association		
Ordinal	Chi-squared Test of Association	Chi-squared/Spearman Correlation t-test	
Continuous	ANOVA/t-test	ANOVA/t-test	Pearson/Spearman Correlation t-test

Over the last two week we have learned how to describe the different types of variables in data and relationships between them.

But how can we be confident that a relationship or pattern in our data applies to the entire population we are interested in, and isn't just an artefact of our specific sample?

We could…

- Collect data from the entire population (very expensive, often unfeasible)

We could…

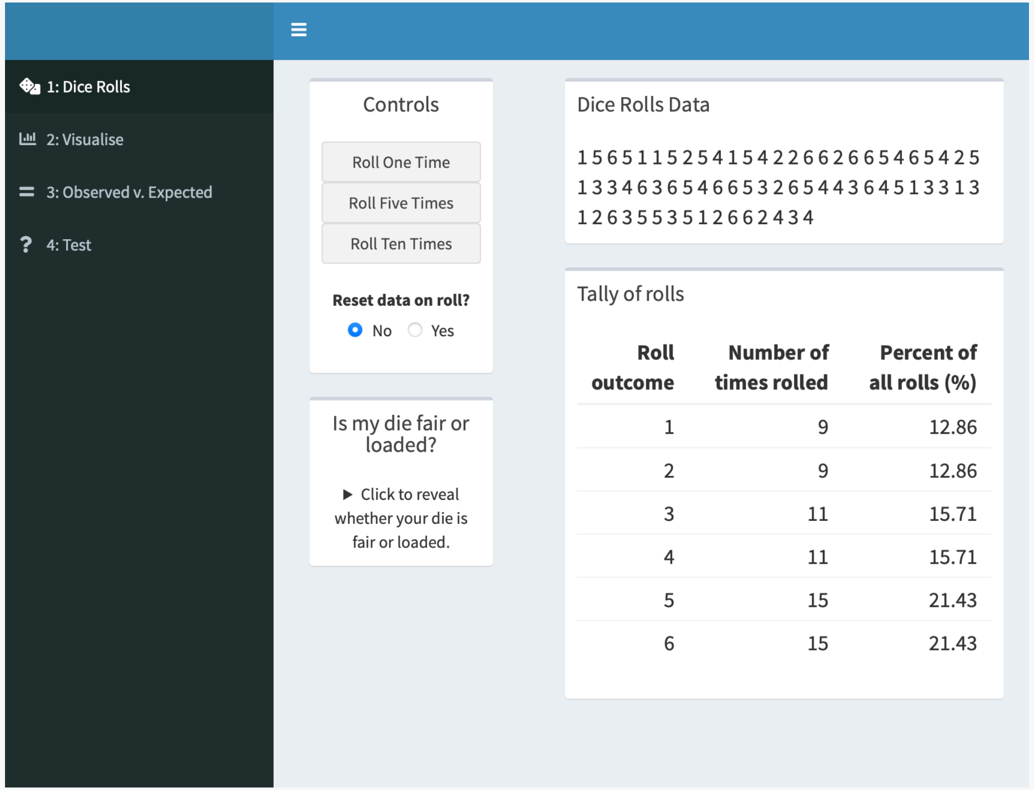
- Collect data from the entire population (very expensive, often unfeasible)
- Collect more random samples and see if we get the same results consistently (good option, but how many before we can be sure? When do we stop?)

We could…

- Collect data from the entire population (very expensive, often unfeasible)
- Collect more random samples and see if we get the same results consistently (good option, but how many before we can be sure? When do we stop?)
- Use inferential statistics

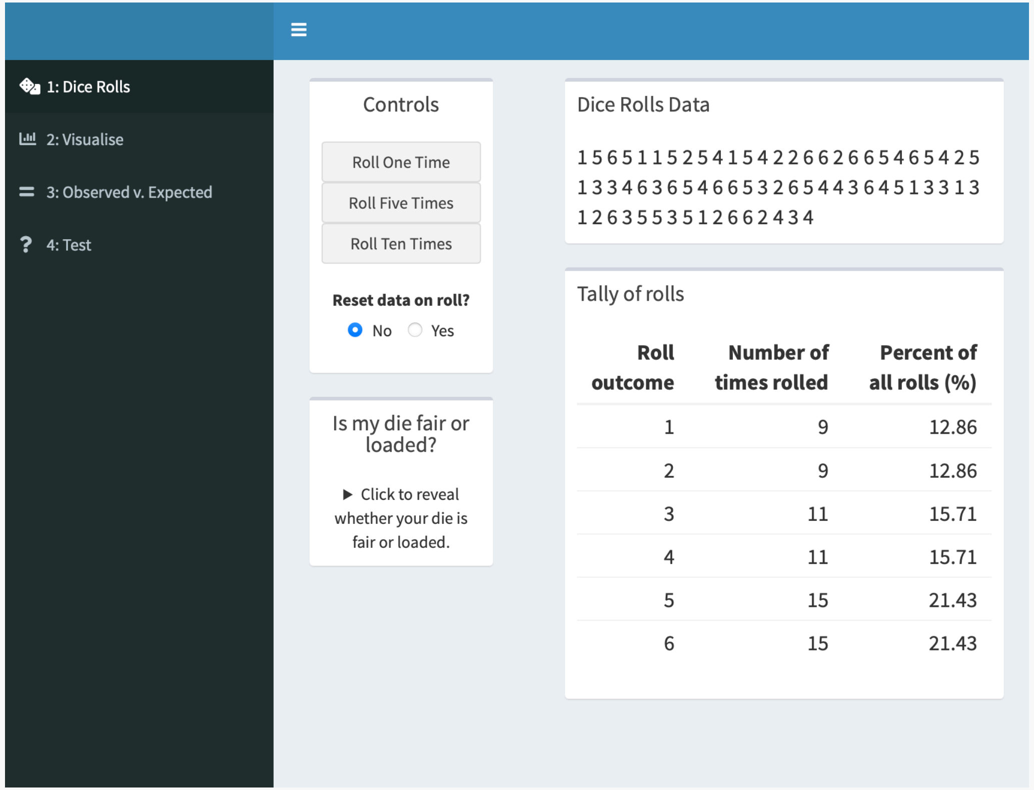
Getting a feel for inferential statistics

- Load the `chisq-sig` Shiny app in R (follow the handout if you didn't do this in advance), or load the online version by going here: https://webb.shinyapps.io/chisq-sig/



Getting a feel for inferential statistics

- Load the `chisq-sig` Shiny app in R (follow the handout if you didn't do this in advance), or load the online version by going here: https://webb.shinyapps.io/chisq-sig/
- We've been contracted by a Las Vegas casino who have discovered that half of their dice are loaded, and do not roll fairly. 🕵

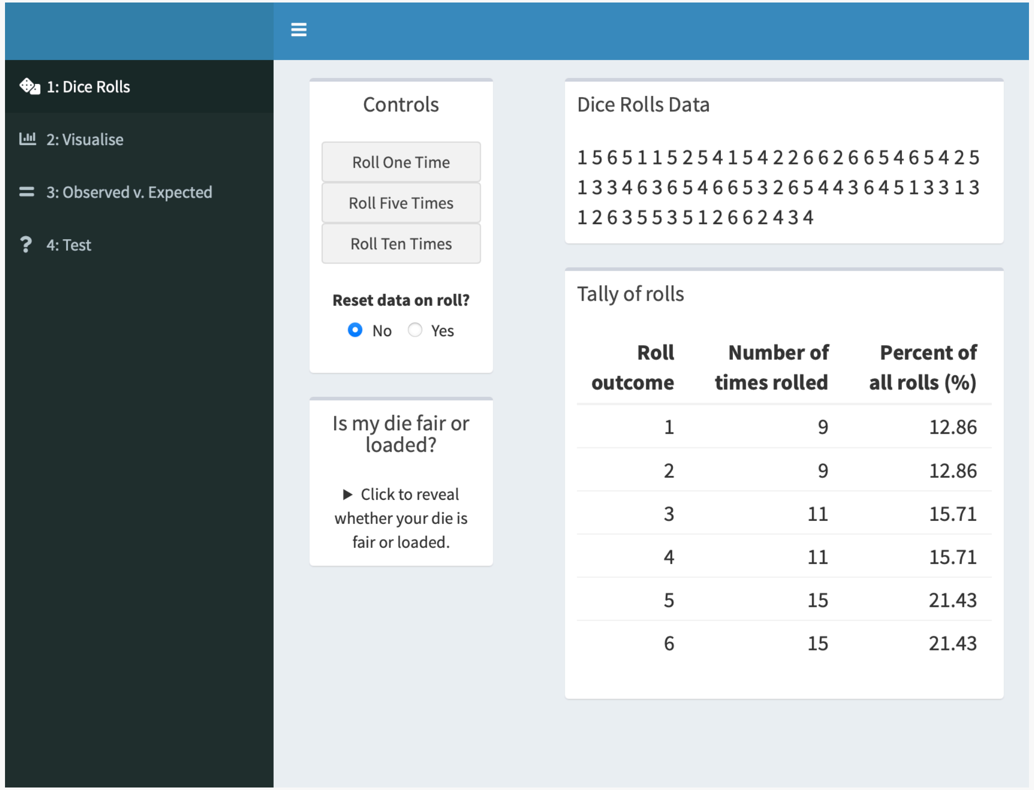


Getting a feel for inferential statistics

- Load the `chisq-sig` Shiny app in R (follow the handout if you didn't do this in advance), or load the online version by going here: https://webb.shinyapps.io/chisq-sig/
- We've been contracted by a Las Vegas casino who have discovered that half of their dice are loaded, and do not roll fairly. 🕵️

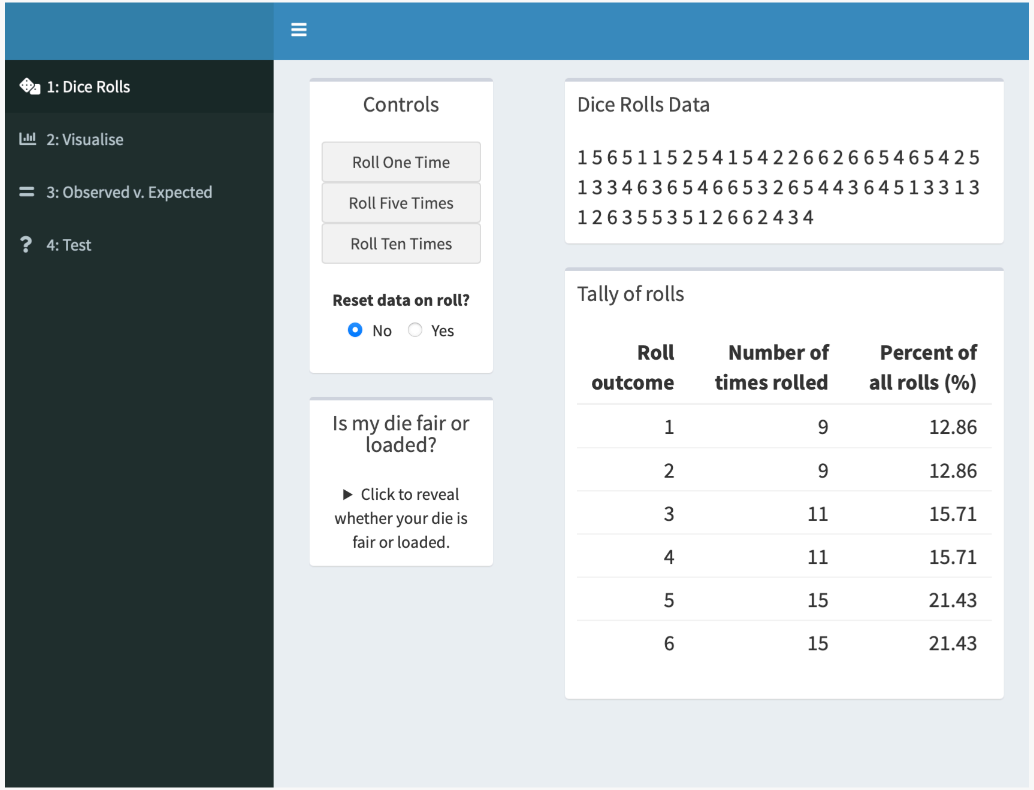
Getting a feel for inferential statistics

- Load the `chisq-sig` Shiny app in R (follow the handout if you didn't do this in advance), or load the online version by going here: https://webb.shinyapps.io/chisq-sig/
- We've been contracted by a Las Vegas casino who have discovered that half of their dice are loaded, and do not roll fairly. 🕵
- Each of you have been given a (virtual) die that you can roll as many times as you like. You don't know whether you have a **fair** die or a **loaded die**. A loaded die will roll some numbers more often than others. 🎲



Getting a feel for inferential statistics

- Load the `chisq-sig` Shiny app in R (follow the handout if you didn't do this in advance), or load the online version by going here: https://webb.shinyapps.io/chisq-sig/
- We've been contracted by a Las Vegas casino who have discovered that half of their dice are loaded, and do not roll fairly.
- Each of you have been given a (virtual) die that you can roll as many times as you like. You don't know whether you have a **fair** die or a **loaded die**. A loaded die will roll some numbers more often than others.
- Our task is to use our data analysis skills to determine whether we have a fair or loaded die.

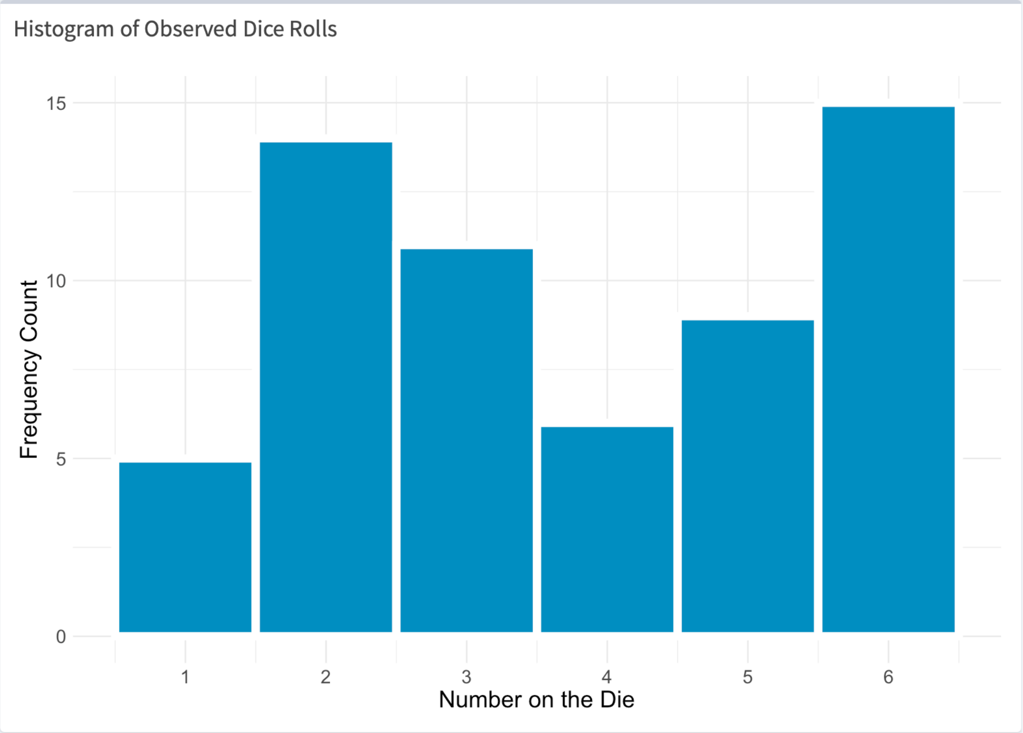


Inferential statistics help us quantify the confidence we have in a hypothesis based on how likely we would expect to see the results we got if it were accurate.

(e.g. that a die is fair, or that there is no relationship between two variables)

Hypothesis testing

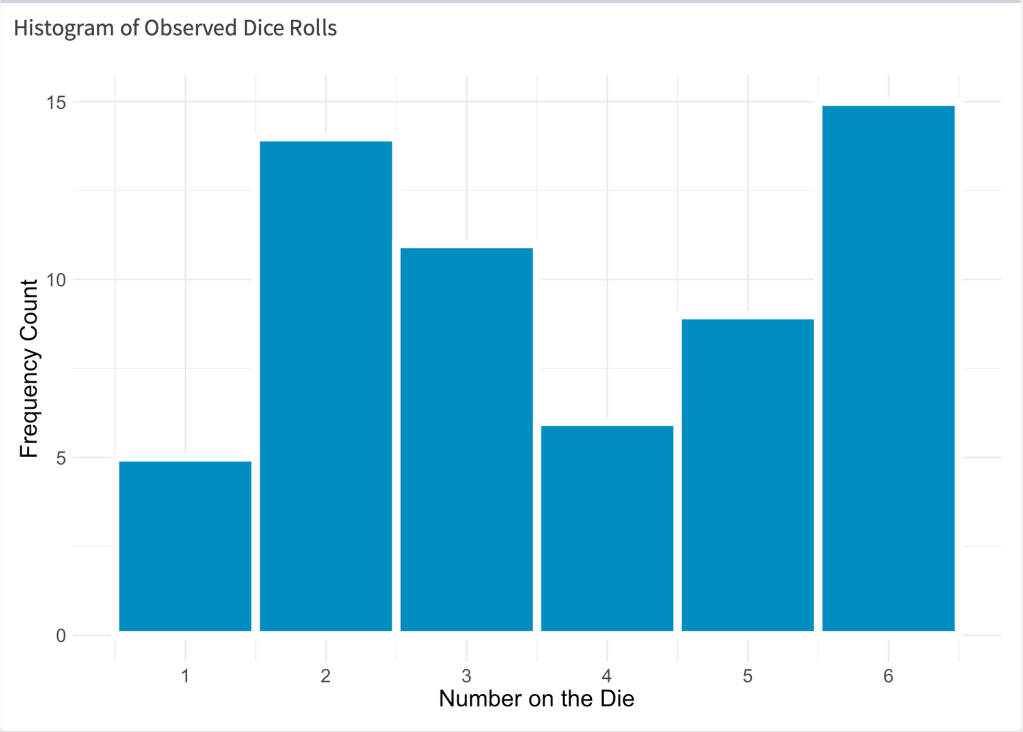
What are the chances we would see a sample of rolls like this...



(Observed)

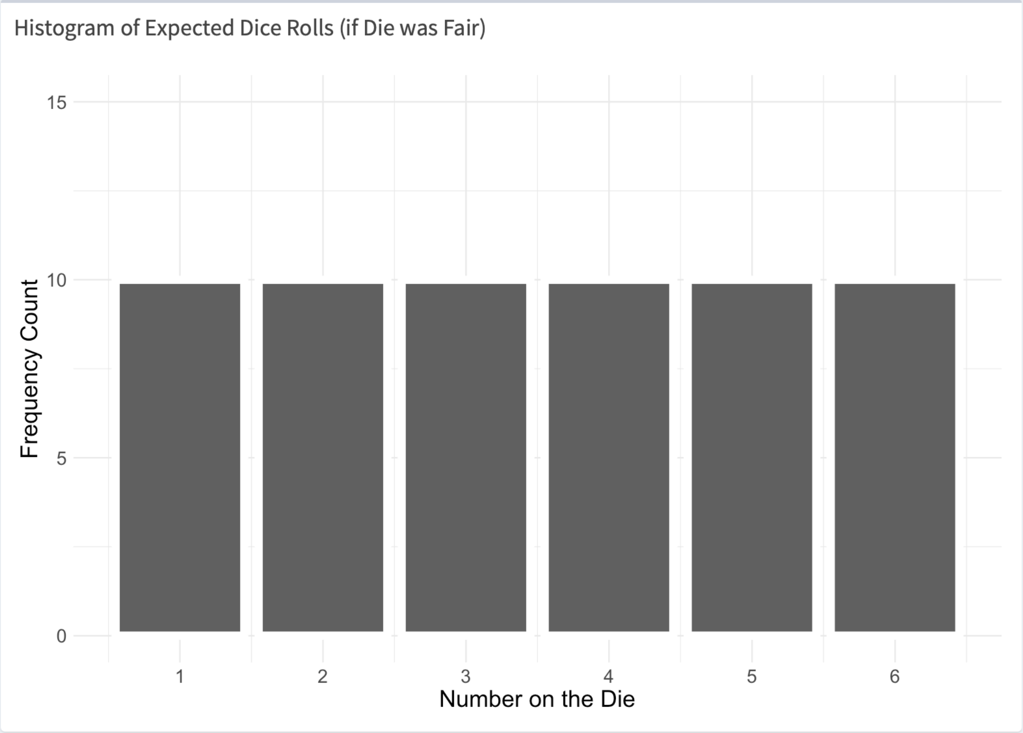
Hypothesis testing

What are the chances we would see a sample of rolls like this...



(Observed)

When we know if the die were fair we would expect to see something like this...? (Null hypothesis)

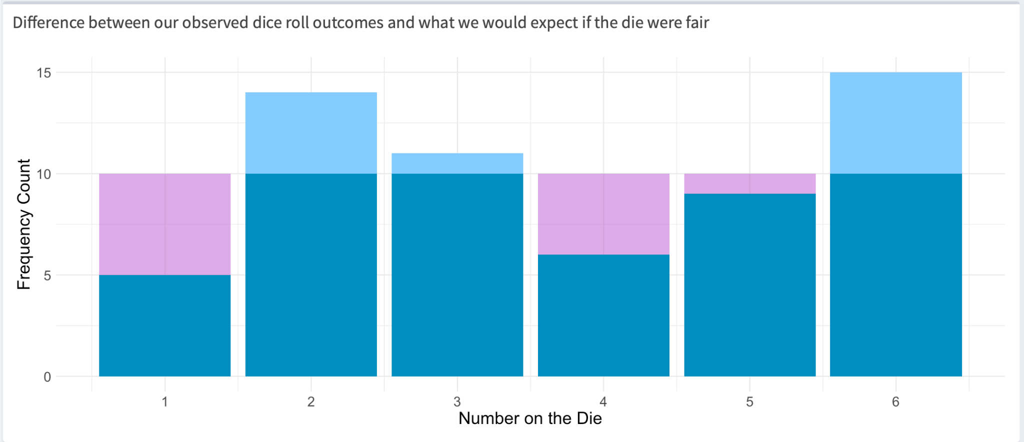


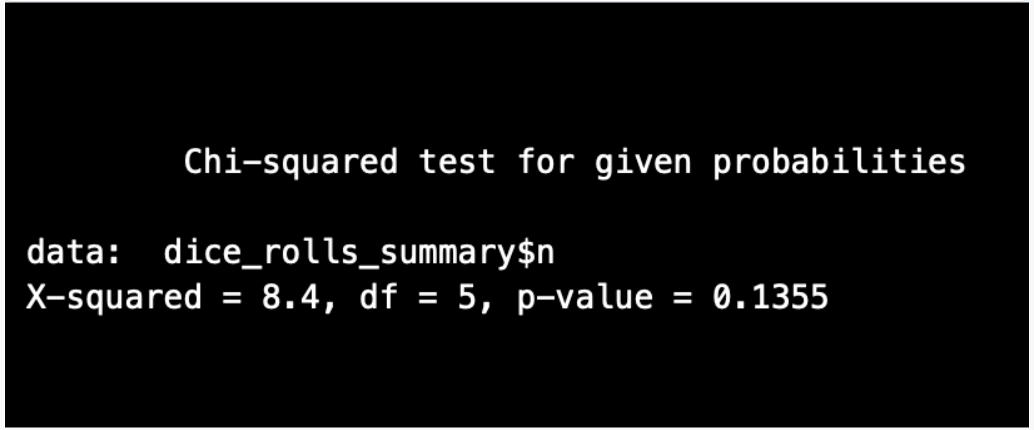
(Expected)

Hypothesis testing

We can express how unlikely we were to get results like this if the die was fair using a **p-value**.

There are many different kinds of inferential statistics and tests we can use for different hypotheses and kinds of relationships in data. The one we use here is called a **chi-squared goodness-of-fit test** but don't worry about how it's calculated at this point!



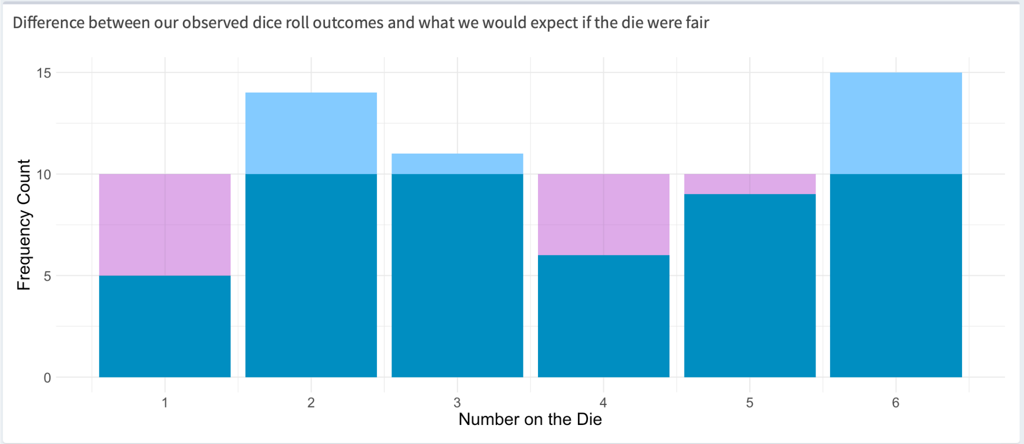


Hypothesis testing

We can express how unlikely we were to get results like this if the die was fair using a **p-value**.

There are many different kinds of inferential statistics and tests we can use for different hypotheses and kinds of relationships in data. The one we use here is called a **chi-squared goodness-of-fit test** but don't worry about how it's calculated at this point!

- An inferential statistic gives us a **p-value**.



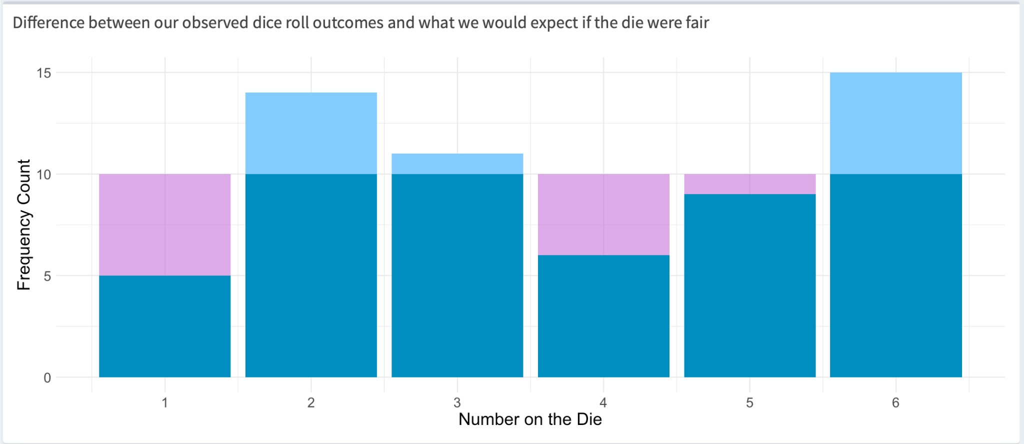

```
        Chi-squared test for given probabilities

data:  dice_rolls_summary$n
X-squared = 8.4, df = 5, p-value = 0.1355
```

Hypothesis testing

We can express how unlikely we were to get results like this if the die was fair using a **p-value**.

There are many different kinds of inferential statistics and tests we can use for different hypotheses and kinds of relationships in data. The one we use here is called a **chi-squared goodness-of-fit test** but don't worry about how it's calculated at this point!

- An inferential statistic gives us a **p-value**.
- The p-value tells us the probability of seeing the kind of results we got **if the null hypothesis** (that the die is fair) **is the best explanation for the distribution of the data**.



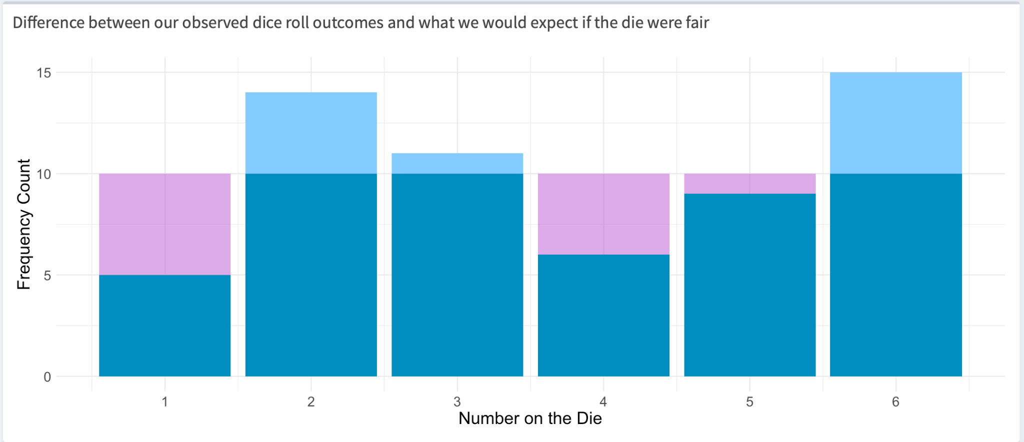

```
        Chi-squared test for given probabilities

data:  dice_rolls_summary$n
X-squared = 8.4, df = 5, p-value = 0.1355
```

Hypothesis testing

We can express how unlikely we were to get results like this if the die was fair using a **p-value**.

There are many different kinds of inferential statistics and tests we can use for different hypotheses and kinds of relationships in data. The one we use here is called a **chi-squared goodness-of-fit test** but don't worry about how it's calculated at this point!

- An inferential statistic gives us a **p-value**.
- The p-value tells us the probability of seeing the kind of results we got **if the null hypothesis** (that the die is fair) **is the best explanation for the distribution of the data**.
- For the above example, **our p-value was 0.1355**.



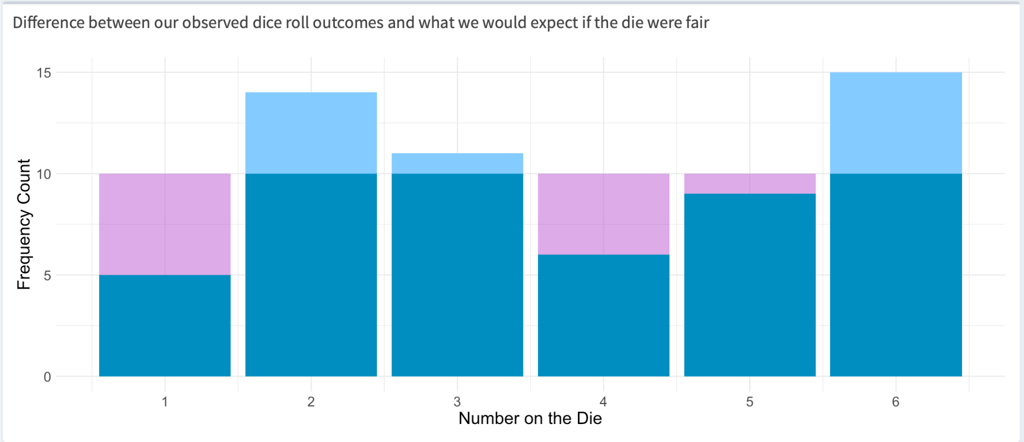

```
        Chi-squared test for given probabilities

data:  dice_rolls_summary$n
X-squared = 8.4, df = 5, p-value = 0.1355
```

Hypothesis testing

We can express how unlikely we were to get results like this if the die was fair using a **p-value**.

There are many different kinds of inferential statistics and tests we can use for different hypotheses and kinds of relationships in data. The one we use here is called a **chi-squared goodness-of-fit test** but don't worry about how it's calculated at this point!

- An inferential statistic gives us a **p-value**.
- The p-value tells us the probability of seeing the kind of results we got **if the null hypothesis** (that the die is fair) **is the best explanation for the distribution of the data**.
- For the above example, **our p-value was 0.1355**.
- This means we would see results at least this different to what we would expect around **13.55% of the time or less**, when a die is fair (when the null hypothesis is an accurate description).




```
        Chi-squared test for given probabilities

data:  dice_rolls_summary$n
X-squared = 8.4, df = 5, p-value = 0.1355
```

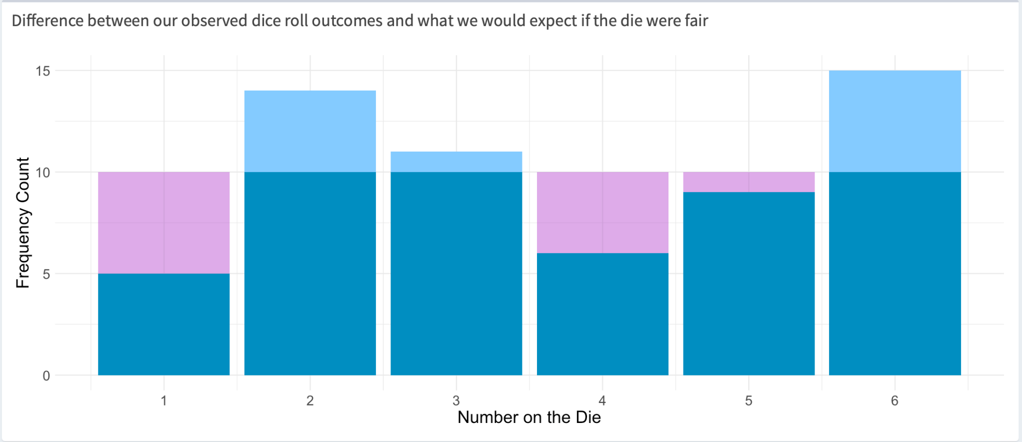
So, what do we think?

13.55% is quite a low probability of something happening.
Should we report this die as unfair or not?

Hypothesis testing

In applied statistics, we compare our p-value with a pre-chosen '**critical value**' (sometimes called *alpha*) below which we decide to reject the null hypothesis.

- Conventionally, our critical value, **below which we reject the null hypothesis**, is **0.05**.




```
        Chi-squared test for given probabilities

data:  dice_rolls_summary$n
X-squared = 8.4, df = 5, p-value = 0.1355
```

Hypothesis testing

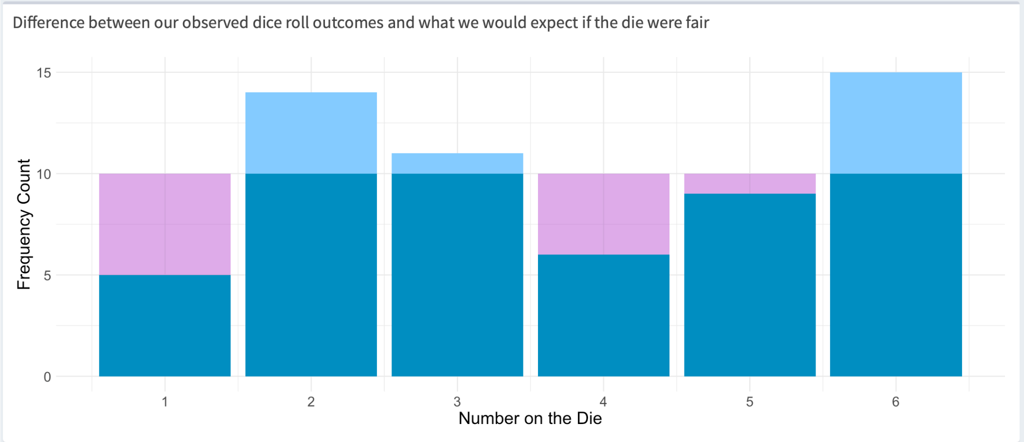
In applied statistics, we compare our p-value with a pre-chosen 'critical value' (sometimes called *alpha*) below which we decide to reject the null hypothesis.

- Conventionally, our critical value, **below which we reject the null hypothesis**, is **0.05**.

There is no strong reason why 5% is used in the social sciences, and sometimes 10%, 1% or 0.1% are used instead, but it can depend on the following:

- What are the risks if we set our critical value too high and incorrectly reject the null hypothesis? **(Type I error; false positive)**
- What are the risks if we set our critical value too low and incorrectly fail to reject the null hypothesis? **(Type II error; false negative)**

5%, or 0.05, is often seen as a good compromise between these two risks.



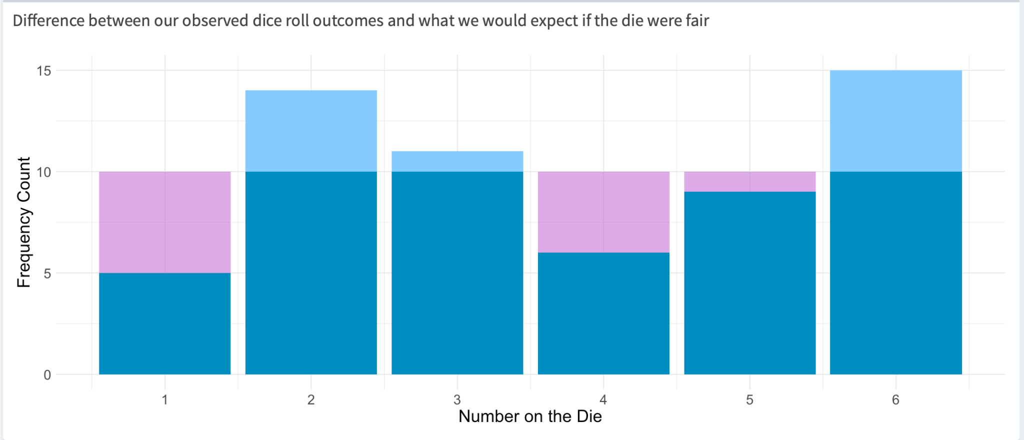

```
        Chi-squared test for given probabilities

data:  dice_rolls_summary$n
X-squared = 8.4, df = 5, p-value = 0.1355
```

Hypothesis testing

- Our p-value is **0.1355**
- Our critical value is **0.05**
- **0.1355 is greater than 0.05** (p > 0.05), and therefore we **should not reject our null hypothesis** (that the die is fair) based on this evidence.
- We conclude that **our data does not support the idea** that the die is unfair.

Don't worry if this is difficult to grasp immediately! No one is comfortable interpreting p-values the first time they come across them!

We will practice using them and interpreting them many many times over the next few weeks!




```
	Chi-squared test for given probabilities

data:  dice_rolls_summary$n
X-squared = 8.4, df = 5, p-value = 0.1355
```

Can you see how this statistic performs a similar function to our intuition when raising our hand when we feel confident that the die is or is not fair?

Now I want you to roll your dice as many times as you think would be a good sample, use the Chi-Square test calculated on the last tab to decide whether you think it is fair or not, and then check if you got it right!

When should we use inferential statistics in social research?

- When we wish to make generalisations beyond our sample of data to a wider population.

When can we use inferential statistics in social research?

- When our data **does not violate any of the assumptions made about it** by the tests (e.g. bivariate normal distrubution).

When can we use inferential statistics in social research?

- When our data **does not violate any of the assumptions made about it** by the tests (e.g. bivariate normal distrubution).
- When our sample is **proportionally representative** of the population we wish to generalise to.

When can we use inferential statistics in social research?

- When our data **does not violate any of the assumptions made about it** by the tests (e.g. bivariate normal distrubution).
- When our sample is **proportionally representative** of the population we wish to generalise to.
- The easiest way to know a sample is proportionally representative of the population is by finding out how the sample was collected. If it is **randomly selected** (a random sample), it is likely to be representative of the population because **every 'thing' in the sample had an equal chance of being selected**.

When can we use inferential statistics in social research?

- When our data **does not violate any of the assumptions made about it** by the tests (e.g. bivariate normal distrubution).
- When our sample is **proportionally representative** of the population we wish to generalise to.
- The easiest way to know a sample is proportionally representative of the population is by finding out how the sample was collected. If it is **randomly selected** (a random sample), it is likely to be representative of the population because **every 'thing' in the sample had an equal chance of being selected**.
- However, **this is quite difficult to achieve with human beings** — they are annoying and do some of the following things:
    - Ignore your invitations to join the sample
    - Refuse to answer questions
    - Withdraw from the study
    - Die
    - And other things.

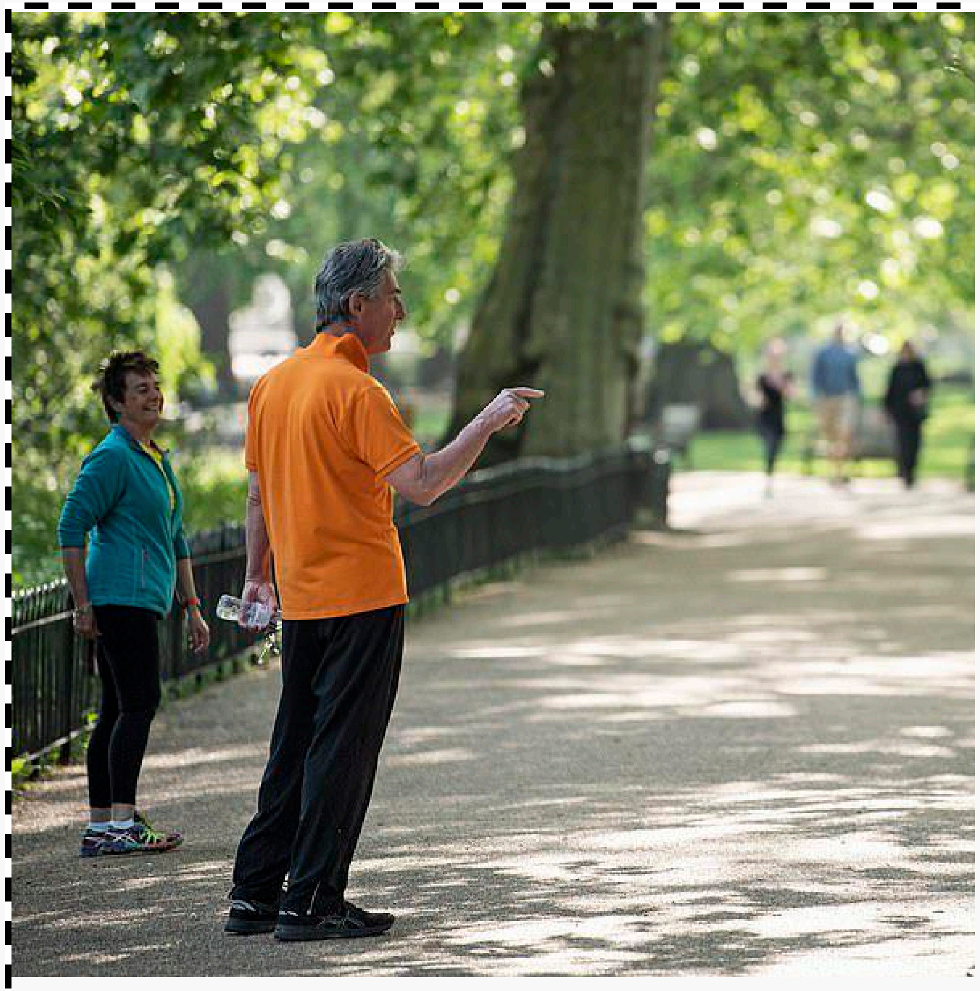
Sampling methods

Whether our data are a representative sample of a larger population often depends on the *sampling method*

Sampling methods

- **Volunteer or opportunity sampling**: the sample is chosen based on who is available to take part in the study (e.g. advertising an online survey; selecting people off the street)

Opportunity sampling (60% of pixels)



Population

Sampling methods

- **Volunteer or opportunity sampling**: the sample is chosen based on who is available to take part in the study (e.g. advertising an online survey; selecting people off the street)
  - Very unlikely to be representative of a population you want to generalise to — what about people without internet access? Or who aren't in close vicinity?

Sampling methods

- **Volunteer or opportunity sampling**: the sample is chosen based on who is available to take part in the study (e.g. advertising an online survey; selecting people off the street)
  - Very unlikely to be representative of a population you want to generalise to — what about people without internet access? Or who aren't in close vicinity?
- **Simple random sampling**: the sample is chosen truly at random from a population sampling frame (e.g. randomly mailing surveys to or visiting addresses on record)

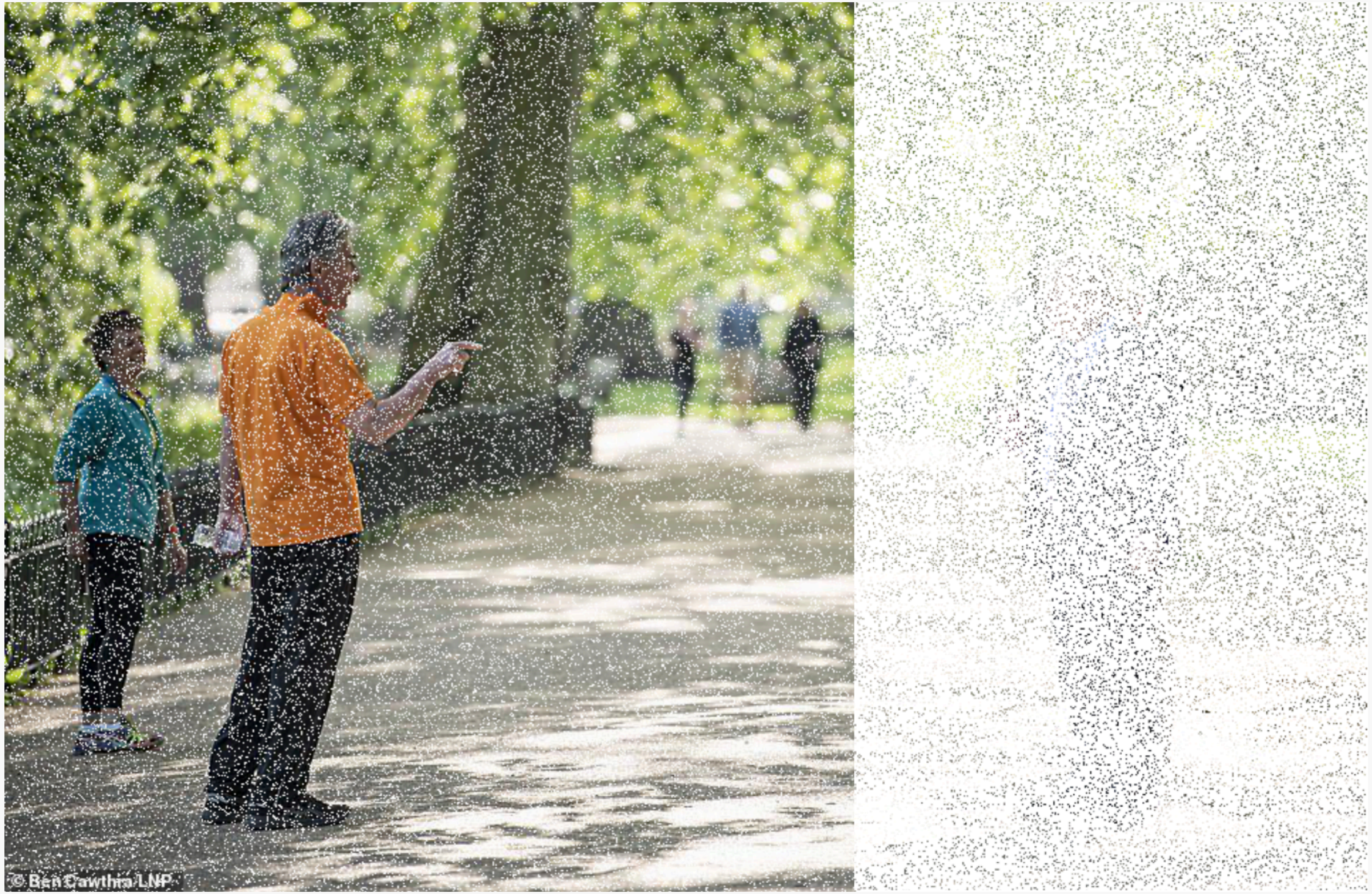
Random sampling (60% of pixels)

Random sampling (30% of pixels)

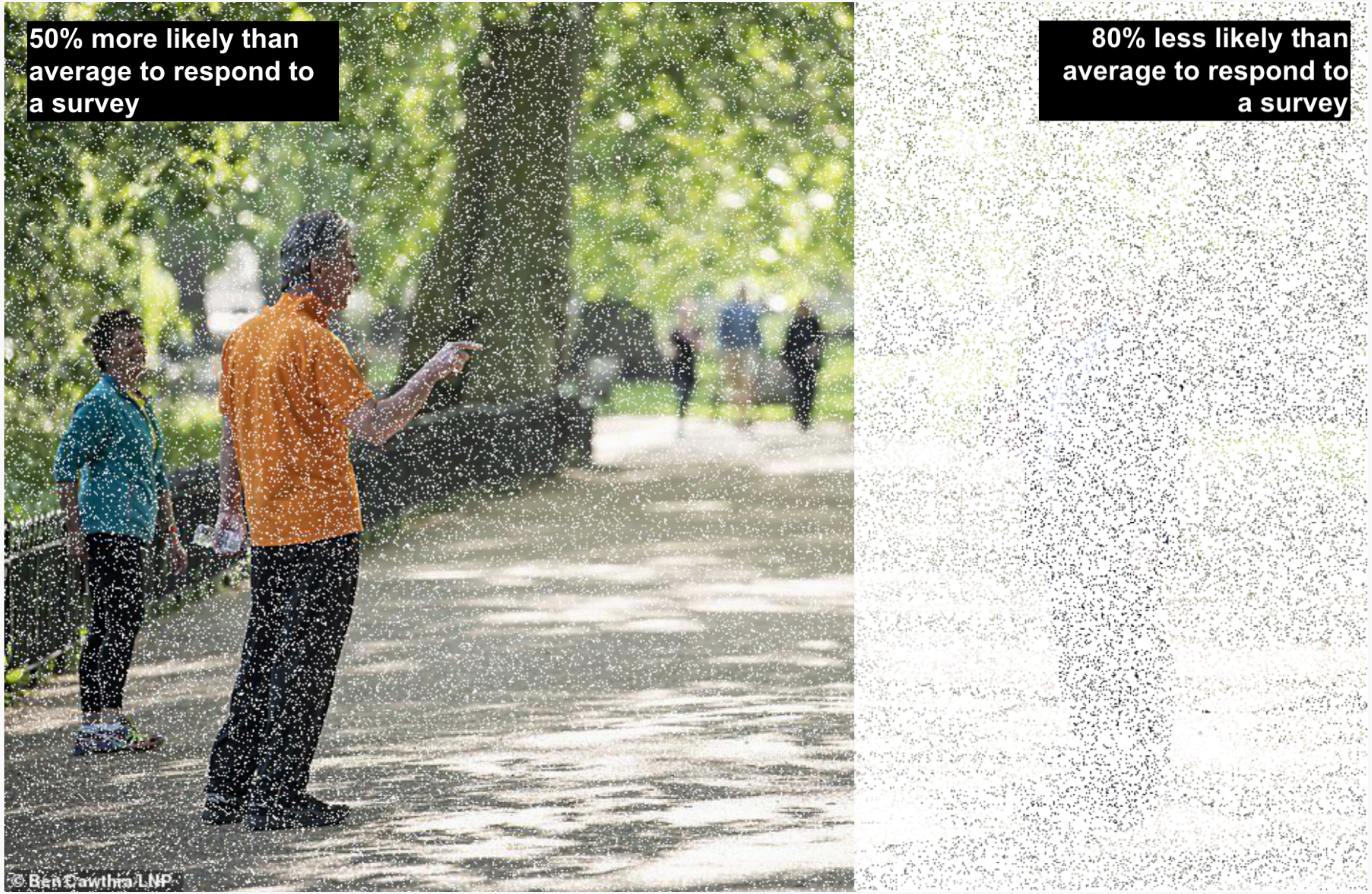
Sampling methods

- **Volunteer or opportunity sampling**: the sample is chosen based on who is available to take part in the study (e.g. advertising an online survey; selecting people off the street)
    - Very unlikely to be representative of a population you want to generalise to — what about people without internet access? Or who aren't in close vicinity?
- **Simple random sampling**: the sample is chosen truly at random from a population sampling frame (e.g. randomly mailing surveys to or visiting addresses on record)
    - Ignores the fact that some people in the population may be more prone to non-response than others/participation bias (Berg, 2010).

Random sampling (with non-response)



Random sampling (with non-response)



Sampling methods

- **Volunteer or opportunity sampling**: the sample is chosen based on who is available to take part in the study (e.g. advertising an online survey; selecting people off the street)
  - Very unlikely to be representative of a population you want to generalise to — what about people without internet access? Or who aren't in close vicinity?
- **Simple random sampling**: the sample is chosen truly at random from a population sampling frame (e.g. randomly mailing surveys to or visiting addresses on record)
  - Ignores the fact that some people in the population may be more prone to non-response than others/participation bias (Berg, 2010).
- **Stratified random sampling**: important demographic categories are first chosen (strata), and then participants are randomly sampled from within those categories (usually proportional to the percentage of the entire population they make up derived from, e.g. a census)

Stratified random sampling (with non-response adjustment)

Stratified random sampling (with non-response adjustment)

Stratified random sampling (with non-response adjustment)

Sampling methods

- **Volunteer or opportunity sampling**: the sample is chosen based on who is available to take part in the study (e.g. advertising an online survey; selecting people off the street)
    - Very unlikely to be representative of a population you want to generalise to — what about people without internet access? Or who aren't in close vicinity?
- **Simple random sampling**: the sample is chosen truly at random from a population sampling frame (e.g. randomly mailing surveys to or visiting addresses on record)
    - Ignores the fact that some people in the population may be more prone to non-response than others/participation bias (Berg, 2010).
- **Stratified random sampling**: important demographic categories are first chosen (strata), and then participants are randomly sampled from within those categories (usually proportional to the percentage of the entire population they make up derived from, e.g. a census)
    - Who decides which demographic categories are meaningful and should be strata and who decides which are unimportant?

Post-hoc adjustment for representativeness

- **Sample weighting**: researchers calculate 'survey weights' which can be used to 're-balance' each observation so that the overall sample proportionally matches the population of interest.

Post-hoc adjustment for representativeness

- **Sample weighting**: researchers calculate 'survey weights' which can be used to 're-balance' each observation so that the overall sample proportionally matches the population of interest.
  - For example, **imagine we wanted a representative sample of England & Wales.**

Post-hoc adjustment for representativeness

- **Sample weighting**: researchers calculate 'survey weights' which can be used to 're-balance' each observation so that the overall sample proportionally matches the population of interest.
    - For example, **imagine we wanted a representative sample of England & Wales.**
    - **Approximately 5%** of people in England & Wales live in Wales.

Post-hoc adjustment for representativeness

- **Sample weighting**: researchers calculate 'survey weights' which can be used to 're-balance' each observation so that the overall sample proportionally matches the population of interest.
    - For example, **imagine we wanted a representative sample of England & Wales.**
    - **Approximately 5%** of people in England & Wales live in Wales.
    - We used a stratified sampling method to **survey 1,000 people** in England and Wales. Ideally, we want 950 English respondents and 50 Welsh respondents.

Post-hoc adjustment for representativeness

- **Sample weighting**: researchers calculate 'survey weights' which can be used to 're-balance' each observation so that the overall sample proportionally matches the population of interest.
  - For example, **imagine we wanted a representative sample of England & Wales.**
  - **Approximately 5%** of people in England & Wales live in Wales.
  - We used a stratified sampling method to **survey 1,000 people** in England and Wales. Ideally, we want 950 English respondents and 50 Welsh respondents.
  - However, we ended up with 100 Welsh respondents and 900 English.

Post-hoc adjustment for representativeness

- **Sample weighting**: researchers calculate 'survey weights' which can be used to 're-balance' each observation so that the overall sample proportionally matches the population of interest.
    - For example, **imagine we wanted a representative sample of England & Wales.**
    - **Approximately 5%** of people in England & Wales live in Wales.
    - We used a stratified sampling method to **survey 1,000 people** in England and Wales. Ideally, we want 950 English respondents and 50 Welsh respondents.
    - However, we ended up with 100 Welsh respondents and 900 English.
    - We could count each of those Welsh respondents as only 0.5 of a respondent, and every one of the English respondents as 1.055 of a respondent to "re-balance" our sample to be proportionate to the population.

Post-hoc adjustment for representativeness

- **Sample weighting**: researchers calculate 'survey weights' which can be used to 're-balance' each observation so that the overall sample proportionally matches the population of interest.
  - For example, **imagine we wanted a representative sample of England & Wales.**
  - **Approximately 5%** of people in England & Wales live in Wales.
  - We used a stratified sampling method to **survey 1,000 people** in England and Wales. Ideally, we want 950 English respondents and 50 Welsh respondents.
  - However, we ended up with 100 Welsh respondents and 900 English.
  - We could count each of those Welsh respondents as only 0.5 of a respondent, and every one of the English respondents as 1.055 of a respondent to "re-balance" our sample to be proportionate to the population.

In reality this process is far more complicated but that's the basic jist of it! In reality, survey data will come with weights already calculated. You can apply them in analysis using the `survey` package in `R`, but this is more something to worry about for a PhD project. For now, don't worry about weighting data.

Post-hoc adjustment for representativeness (Non-response)

We can also have a scenario where we get a proportionally representative sample from our population, but then **not all of this sample respond to all questions or have data for all variables** (e.g. some might refuse). This would mean that some analyses will end up 'unbalanced' due to this **missing data** (usually coded as `NA` in `R`) when it is removed.

Post-hoc adjustment for representativeness (Non-response)

We can also have a scenario where we get a proportionally representative sample from our population, but then **not all of this sample respond to all questions or have data for all variables** (e.g. some might refuse). This would mean that some analyses will end up 'unbalanced' due to this **missing data** (usually coded as `NA` in `R`) when it is removed.

- Multiple ways of dealing with missing data: most common are *listwise* and *pairwise* deletion — where entire observations are deleted if they have missing data in a variable of interest. **This unbalances our sample**.

Post-hoc adjustment for representativeness (Non-response)

We can also have a scenario where we get a proportionally representative sample from our population, but then **not all of this sample respond to all questions or have data for all variables** (e.g. some might refuse). This would mean that some analyses will end up 'unbalanced' due to this **missing data** (usually coded as `NA` in `R`) when it is removed.

- Multiple ways of dealing with missing data: most common are *listwise* and *pairwise* deletion — where entire observations are deleted if they have missing data in a variable of interest. **This unbalances our sample**.
- However, we can try to **impute** missing data — e.g. fill in all of the missing values with our 'best guess' of what it would have been based on responses from respondents who are the most similar to them.

Post-hoc adjustment for representativeness (Non-response)

We can also have a scenario where we get a proportionally representative sample from our population, but then **not all of this sample respond to all questions or have data for all variables** (e.g. some might refuse). This would mean that some analyses will end up 'unbalanced' due to this **missing data** (usually coded as `NA` in `R`) when it is removed.

- Multiple ways of dealing with missing data: most common are *listwise* and *pairwise* deletion — where entire observations are deleted if they have missing data in a variable of interest. **This unbalances our sample**.
- However, we can try to **impute** missing data — e.g. fill in all of the missing values with our 'best guess' of what it would have been based on responses from respondents who are the most similar to them.
- Some methods have specific procedures for handling missing data (e.g. FIML).

Post-hoc adjustment for representativeness (Non-response)

We can also have a scenario where we get a proportionally representative sample from our population, but then **not all of this sample respond to all questions or have data for all variables** (e.g. some might refuse). This would mean that some analyses will end up 'unbalanced' due to this **missing data** (usually coded as `NA` in `R`) when it is removed.

- Multiple ways of dealing with missing data: most common are *listwise* and *pairwise* deletion — where entire observations are deleted if they have missing data in a variable of interest. **This unbalances our sample**.
- However, we can try to **impute** missing data — e.g. fill in all of the missing values with our 'best guess' of what it would have been based on responses from respondents who are the most similar to them.
- Some methods have specific procedures for handling missing data (e.g. FIML).

Imputation and **maximum likelihood** is too complex a topic to cover here, but is something to be aware of if you are doing a quantitative PhD.

Inference in experimental designs

Inference in experimental designs

If your research uses an **experimental** design (e.g. a survey experiment or a randomised controlled trial), you are usually aiming to test the significant differences between the groups *within* your sample. You might try and make your sample representative, but this is often difficult due to cost.

University of Sheffield | Sheffield Methods Institute

Inference in experimental designs

If your research uses an **experimental** design (e.g. a survey experiment or a randomised controlled trial), you are usually aiming to test the significant differences between the groups *within* your sample. You might try and make your sample representative, but this is often difficult due to cost.

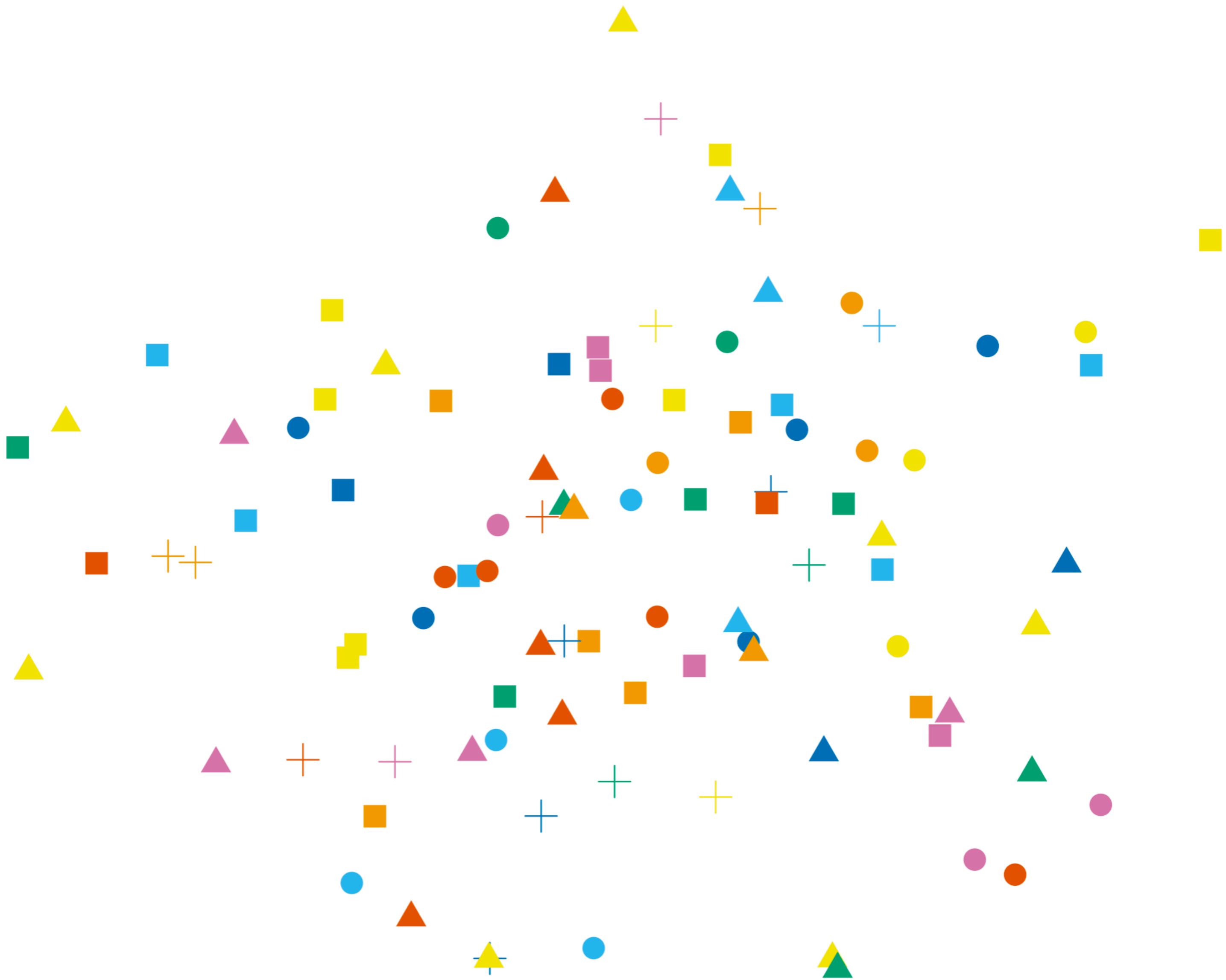
This is handled through **a priori randomisation** of conditions (randomly assigning participants to either 'treatment' or 'control' conditions). Because the **assignment is random**, statistical significance/inferential statistics can be used to generalise to the group who participated in the study as a whole.

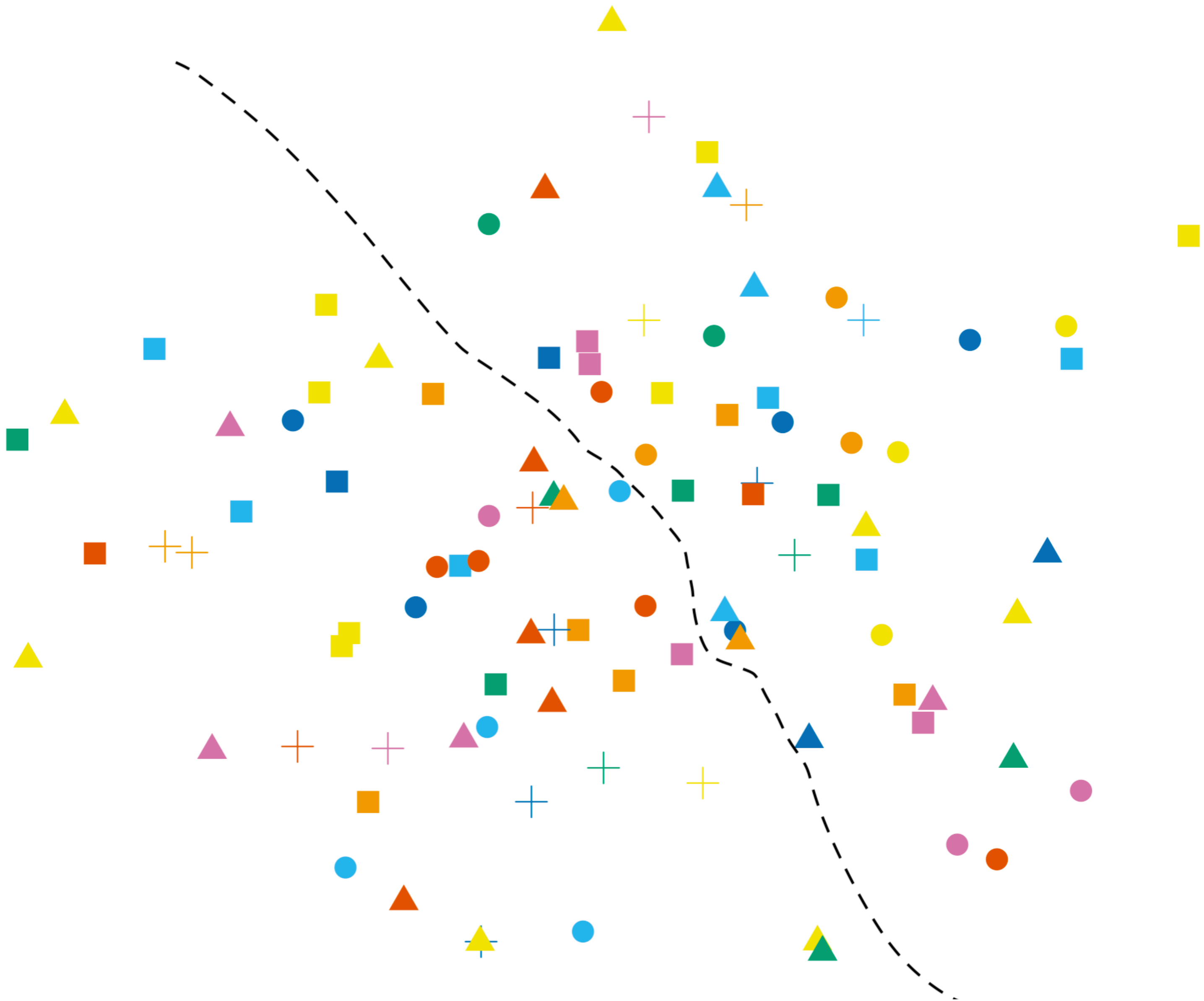
Inference in experimental designs

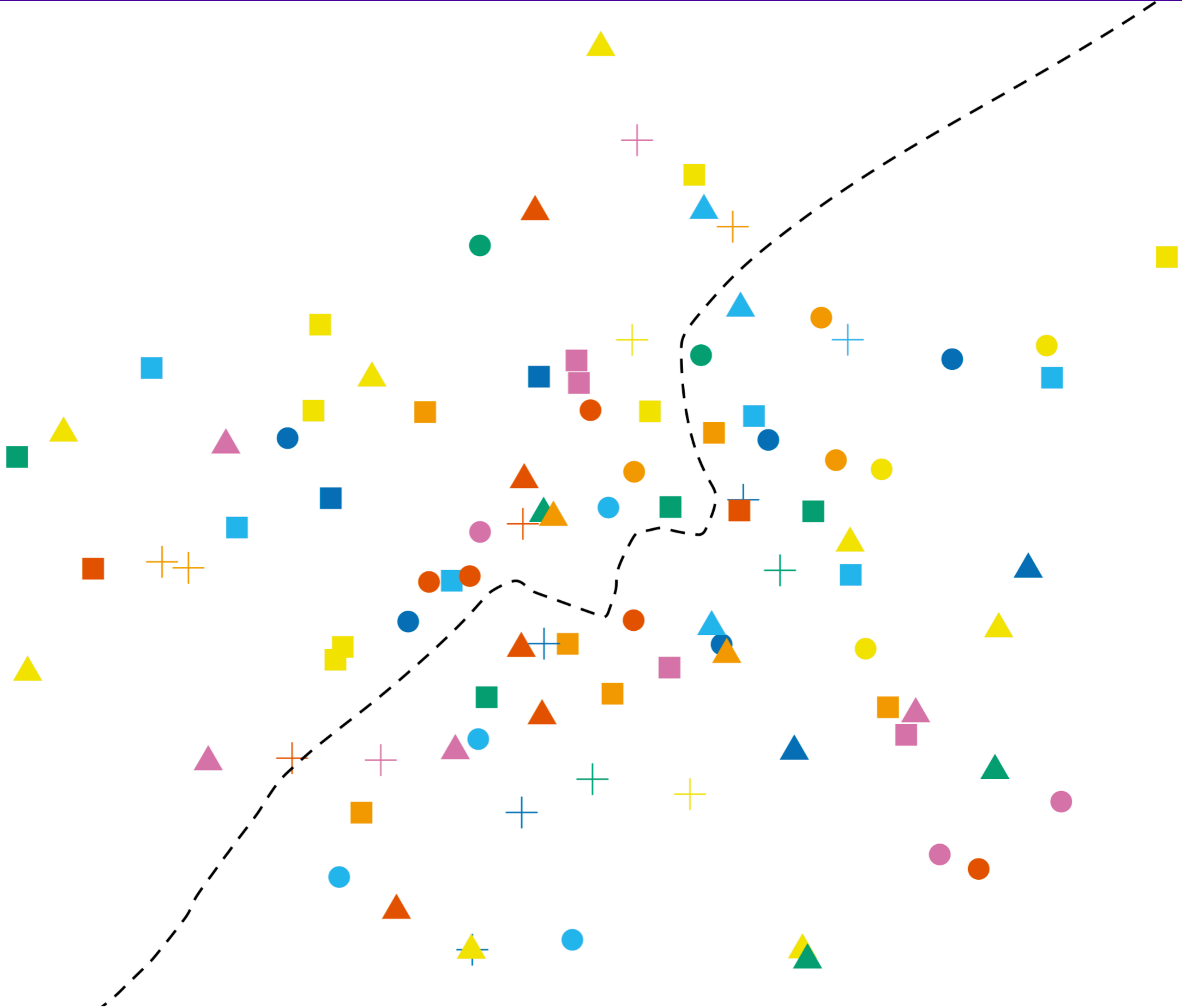
If your research uses an **experimental** design (e.g. a survey experiment or a randomised controlled trial), you are usually aiming to test the significant differences between the groups *within* your sample. You might try and make your sample representative, but this is often difficult due to cost.

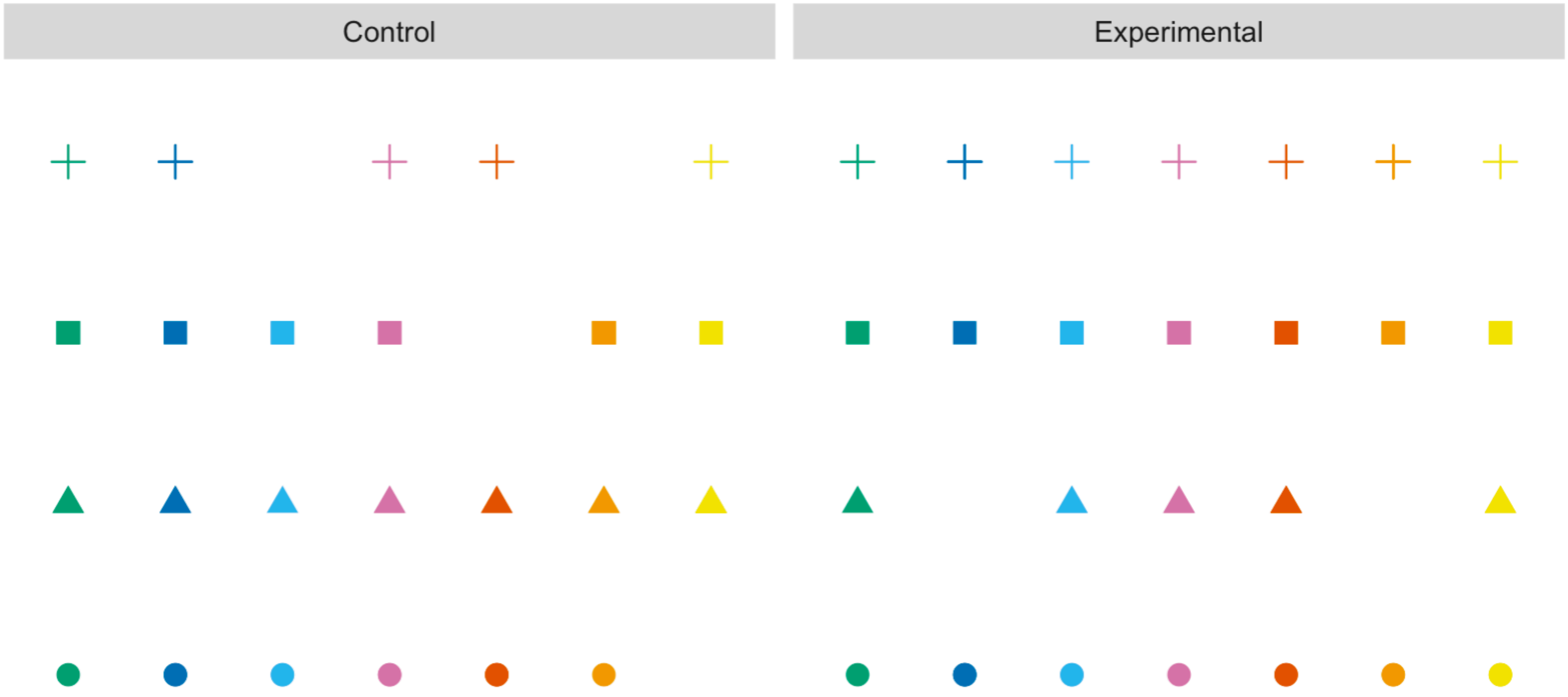
This is handled through **a priori randomisation** of conditions (randomly assigning participants to either 'treatment' or 'control' conditions). Because the **assignment is random**, statistical significance/inferential statistics can be used to generalise to the group who participated in the study as a whole.

In other words, inferential statistics can be used to determine **whether the difference between the 'treatment' group and 'control' group would have been different to what would be expected under the null hypothesis if the groups were reverse, or if the random assignment was different**.








Control
Experimental

Inferential statistics for hypothesis testing

Which hypothesis tests should we use for each combination of variables?

Inferential statistics for hypothesis testing

Variable Type	Nominal	Ordinal	Continuous
Nominal	Chi-squared Test of Association		
Ordinal	Chi-squared Test of Association	Chi-squared/Spearman Correlation t-test	
Continuous	**ANOVA/t-test**	**ANOVA/t-test**	Pearson/Spearman Correlation t-test

ANOVA/t-test

Use case:

- One 'grouping' **nominal/categorical/ordinal** variable and one **continuous** variable.
- For t-test, 'grouping' variable must only have two groups. For ANOVA, grouping variable may have any number of groups.

ANOVA/t-test

Use case:

- One 'grouping' **nominal/categorical/ordinal** variable and one **continuous** variable.
- For t-test, 'grouping' variable must only have two groups. For ANOVA, grouping variable may have any number of groups.

Null hypothesis:

- **H_0**: The mean value of all groups is equal. (There are no significant differences between group averages).

ANOVA/t-test

Use case:

- One 'grouping' **nominal/categorical/ordinal** variable and one **continuous** variable.
- For t-test, 'grouping' variable must only have two groups. For ANOVA, grouping variable may have any number of groups.

Null hypothesis:

- **H_0**: The mean value of all groups is equal. (There are no significant differences between group averages).

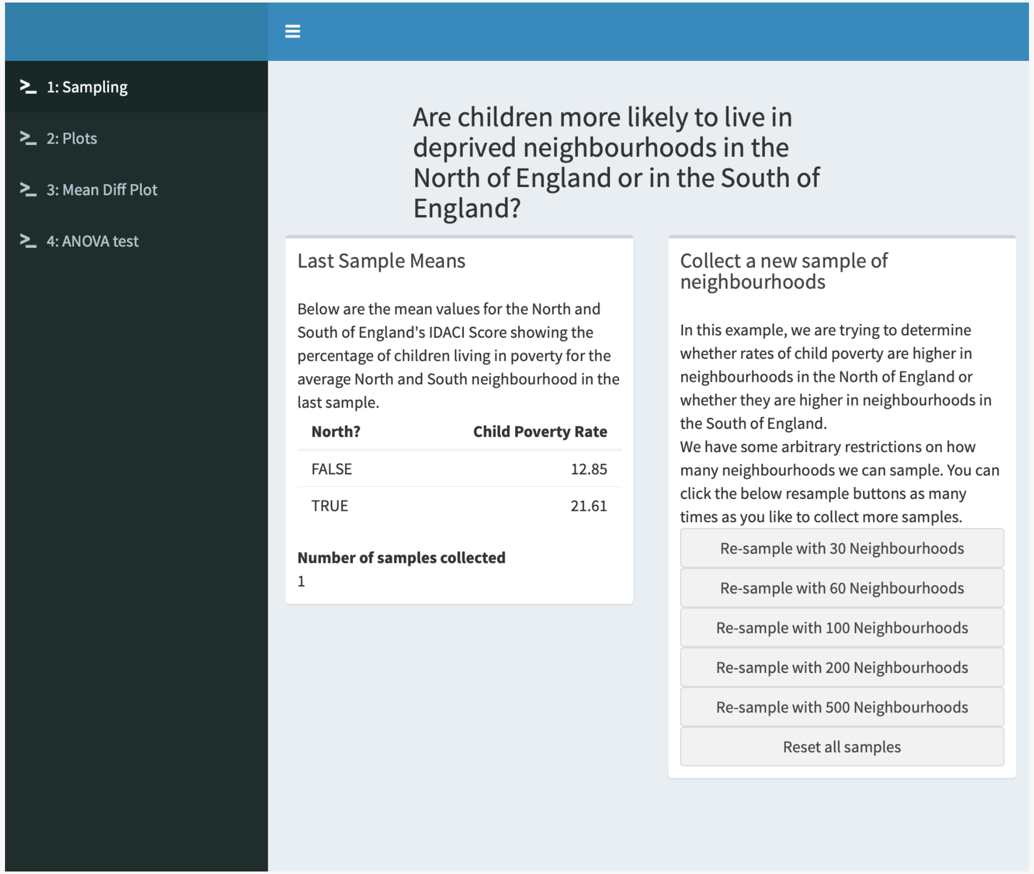
Assumptions:

- **Independence of observations**: Each observation has no bearing on the value of other observations (e.g. if there were multiple observations of the same person, this assumption would be violated)
- **Normality**: Normality of *residuals*; in reality, the means from multiple resamples from each group should be normally distributed in the population (Glass et al. 1972, Harwell et al. 1992, Lix et al. 1996).
- **Homogeneity of variances**: the variance of the continuous variable should be approximately the same in all groups.

ANOVA/t-test Example

Exercise

- Load up the **`anova-sig`** `R` Shiny App following the hand-out steps (hopefully you did this in advance!)
- If you can't get this working, you can use the online version: https://webb.shinyapps.io/anova-sig/



Inferential statistics for hypothesis testing

Variable Type	Nominal	Ordinal	Continuous
Nominal	Chi-squared Test of Association		
Ordinal	Chi-squared Test of Association	Chi-squared/Spearman Correlation t-test	
Continuous	ANOVA/t-test	ANOVA/t-test	**Pearson/Spearman Correlation t-test**

Correlation Coefficient Significance Tests

Use case:

- Testing the significance of an association between two continuous variables.

Correlation Coefficient Significance Tests

Use case:

- Testing the significance of an association between two continuous variables.

Null hypothesis:

- **H_0**: The correlation coefficient for the association between the two variables is equal to zero. (That there is no relationship between them).

Correlation Coefficient Significance Tests

Use case:

- Testing the significance of an association between two continuous variables.

Null hypothesis:

- **H_0**: The correlation coefficient for the association between the two variables is equal to zero. (That there is no relationship between them).

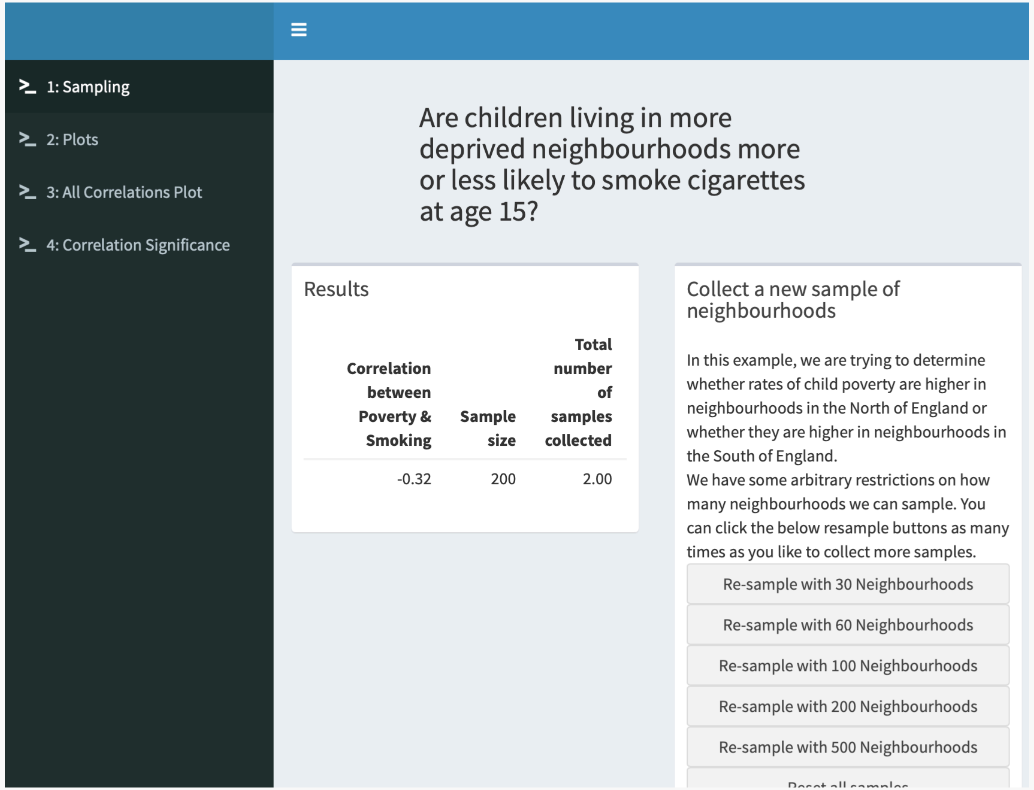
Assumptions:

- **Independence of observations**.
- **Linearity**: there is a linear association between the two variables. In other words, if you were to draw a line of best fit through a scatterplot of them, the best fit would be a straight line.
- **No significant outliers**: any large outliers should be identified and removed.
- **Bivariate normal distribution**: the variables should have a bivariate normal distribution. This is always the case if both variables are normally distributed and their relationship is linear, but can also be the case if one or both are non-normally distributed if, for example, the residuals around a line of best fit between them are normally distributed.

Correlation Coefficient Significance Tests

Exercise

- Load up the **`cor-sig`** R Shiny App following the hand-out steps (hopefully you did this in advance!)
- If you can't get this working, you can use the online version: https://webb.shinyapps.io/cor-sig/



Summary

- Hypothesis tests are one important way we can make broader **generalisations about our findings** from the data. This can be very powerful.
- *p-values* and *critical/alpha values* can be used to make judgements about whether we have enough evidence to reject a given null hypothesis; **if our p-value is lower than our critical value (usually 0.05), we can reject the null hypothesis**.
- The **results of hypothesis tests are determined by sample size** and by the **strength of the association** between variables.
- However, their use requires considerable caution to ensure that:
    - We select **the correct kind of hypothesis test** for our data.
    - p-values are **interpreted correctly**.
    - Hypothesis tests are **used appropriately** (on a suitable kind of sample or within an appropriate study design).
    - We report any possible **violations of assumptions**.

R Exercise

There is no R exercise this week, instead (if we have any time remaining) you should:

- Look at the **assessment 1 details** that are now on Blackboard
- Go back and finish any exercises you've not been able to finish
- Next week we will practice doing these tests in R