The purpose of Statistics is to **ANSWER QUESTIONS USING DATA**

Know the **type of question** and you can choose what **type of statistics...**

Aim: DESCRIBE

Type of question: What's going on?

Examples:

- How many chapters do novels have?
- What possibilities are there for body temperature after a meal with or without chilli?
- What sort of relationship might the amount of sleep a student gets have with their grades?
- What sorts of things might be related to whether a person does volunteer work?

Type of Statistics: Descriptive statistics: graphs and basic numbers

Aim: DECIDE

Type of question: Yes or no?

Examples:

- Is the median number of chapters in a novel 20?
- Is your body temperature higher after a meal if it has chilli in it?
- Does getting more sleep affect a students' grades?
- Are women more likely to participate in volunteer work than men?

Type of Statistics: Hypothesis tests (p-values)

Aim: ESTIMATE

Type of question: What's this number?

Examples:

- What is the median number of chapters in a novel?
- How much higher is your body temperature after a chilli meal compared to one without?
- On average, how much of an effect does 30 minutes more sleep have on a students' grades?
- How much more (or less) likely is a woman to participate in volunteer work than a man?

Type of Statistics: Confidence intervals

Aim: PREDICT / EXPLAIN

Type of question: What's the formula?

Examples:

- How can I explain a person's body temperature after a meal using their temperature before and the chilli content of the meal?
- How can I calculate a student's grade based on their number of hours of sleep during semester?
- How can I use a person's gender, age, income and religion to predict their chances of participating in volunteer work?

Type of Statistics: Modelling and regression

The purpose of Statistics is to **ANSWER QUESTIONS USING DATA**

Know more **about your data** and you can choose what **statistical method...**

HOW THE DATA IS COLLECTED

- what is done to the subjects?
- when is information recorded?
- how are the subjects chosen?

HOW MUCH DATA

- lots of things recorded per subject?
- lots of subjects?
- missing data?

VARIABLES IN THE DATA

- how to measure?
- outcome or explanatory?
- what type?
- what distribution?
- defining groups or measurements?

DATA COLLECTION

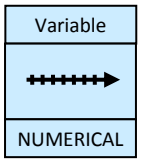
Randomness:

- *Random selection*: you choose the subjects randomly from a population, or at least they are independent.
- *Random allocation*: you choose which subject got what treatment randomly.

Type of study:

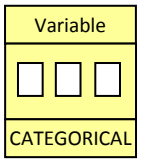
- *Observational study*: the only thing you did to the subjects while you were watching them was to record information about them.
- *Experiment*: You made a choice at least once to do something that might influence the outcome (possibly you made the choice randomly).

TYPES OF VARIABLES (things you record)



Numerical / Quantitative / Scale (numbers: how far apart has meaning)

- Continuous (measured)
- Discrete (counted)



Categorical / Qualitative (words: how far apart has no meaning)

- Nominal (names: more or less has no meaning)
- Ordinal (ordered: more or less has meaning)

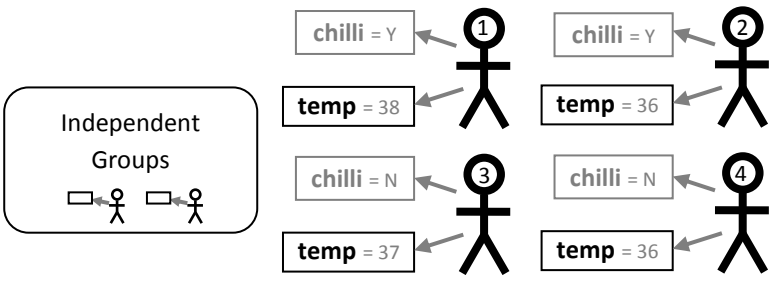
DISTRIBUTIONS OF NUMERICAL VARIABLES (how the possible values are spread out)

- Approximately normal
  – parametric tests will be fine

- Skewed or worse
  – non-parametric tests might be better

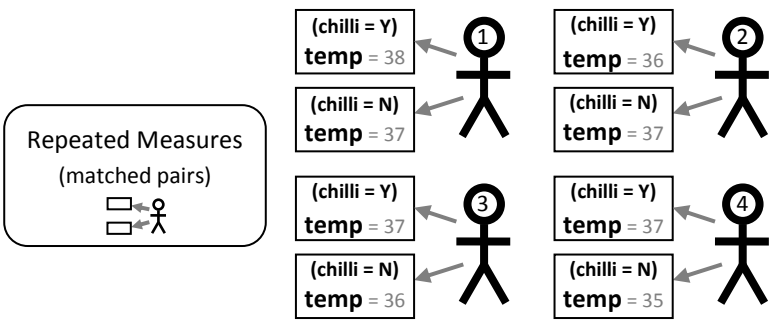
- Likert scale
  – might have to treat as categorical

WHAT EXPLANATORY CATEGORICAL VARIABLES DEFINE:



BECOMES...

	chilli	temp
1	Y	38
2	Y	36
3	N	37
4	N	36



BECOMES...

	(chilli = Y) temp	(chilli = N) temp
1	38	37
2	36	37
3	37	36
4	37	35