ADVANCED QUANTITATIVE METHODS CLINIC

Master's in Sustainability Leadership, Cambridge Institute for Sustainability Leadership

---

Sreekumar Thaithara Balan

Monday 3rd August, 2015

Department of Physics and Astronomy,
University College London

OUTLINE

OUTLINE

- Software for data analysis (∼15mins)
- Data visualisation (∼20mins)
- Descriptive statistics (∼20mins)
- Inferential statistics (∼20mins)
- Regression (∼20mins)
- Discussion (∼10mins)

SOFTWARE FOR DATA ANALYSIS

SOFTWARE FOR DATA ANALYSIS

A large list can be found in `Wikipedia`. Some widely used ones are below.

- `Python`, `https://www.python.org/`
- `R`, `https://cran.r-project.org/`
- `Excel`, `https://products.office.com/en-us/excel`

I will demostrate the examples using `Python`. If you have no prior experience, no problem, there will be plenty of help.

SOFTWARE SET-UP

We need at least one of the statistical softwares mentioned in the previous slide. Please follow the instructions below

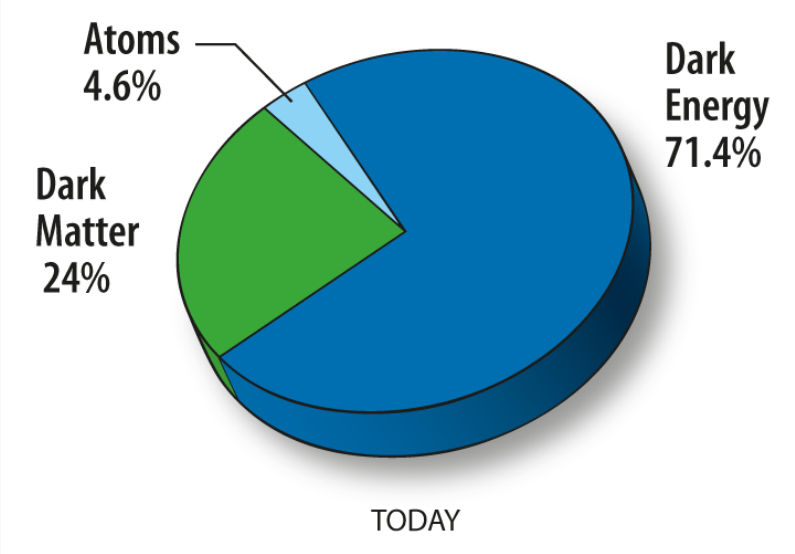
- `http://docs.continuum.io/anaconda/install`
- `https://cran.r-project.org/`
- `https://products.office.com/en-us/excel`

Has everyone installed one of the above?

DATA VISUALISATION

examples

- bar-charts
- histograms
- scatter-plots
- errobars
- pie-charts
- many more!



DATA VISUALISATION: EXAMPLES

- We have several examples in the repository.
- Please follow the instructions in `https://github.com/tbs1980/CISLQuantWorkshop/tree/master/AdvancedQuantitativeMethodsClinic`.
- Try to finish the first example.
- We have 15 mins for this session.
- Get your hands dirty!

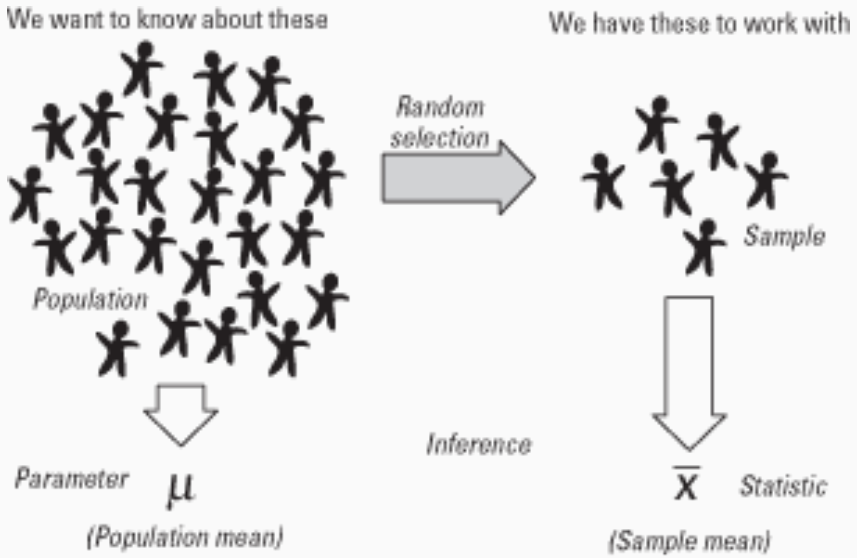
DESCRIPTIVE STATISTICS

DESCRIPTIVE STATISTICS

- Definitions
- Freqency distributions
- Central tendency and variability

Glossary

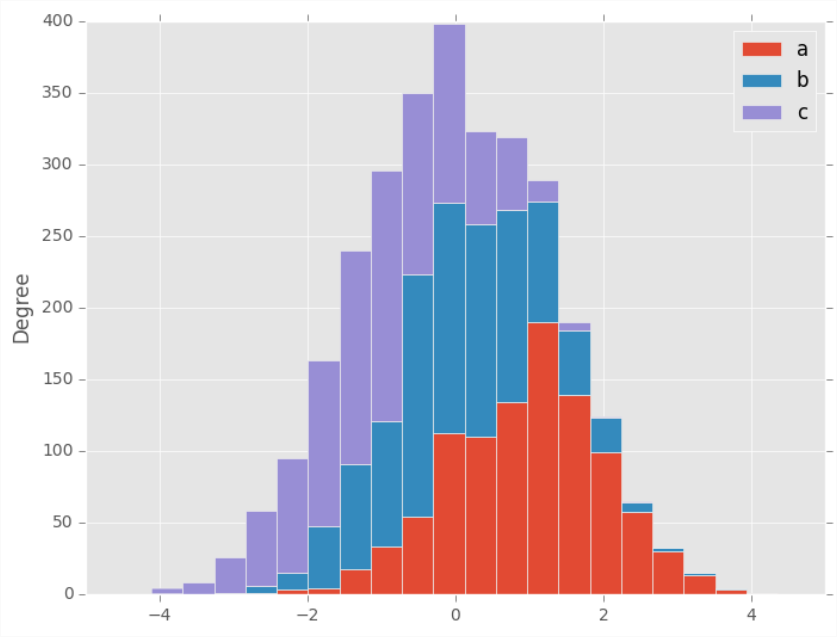
- Population
- Samples
- Variable
- Data
- Parameter
- Statistic



FREQENCY DISTRIBUTIONS

Defined by

- Size
- Range
- Bins-size
- Normalisation



How to characterise a distribution?

- What is a measure of central tendency?
- Mean, median and mode

The mean μ of samples $\{x_1, x_2, \cdots, x_n\}$ can be computed as

$$\mu = \frac{\sum_i x_i}{\sum_i} \qquad (1)$$

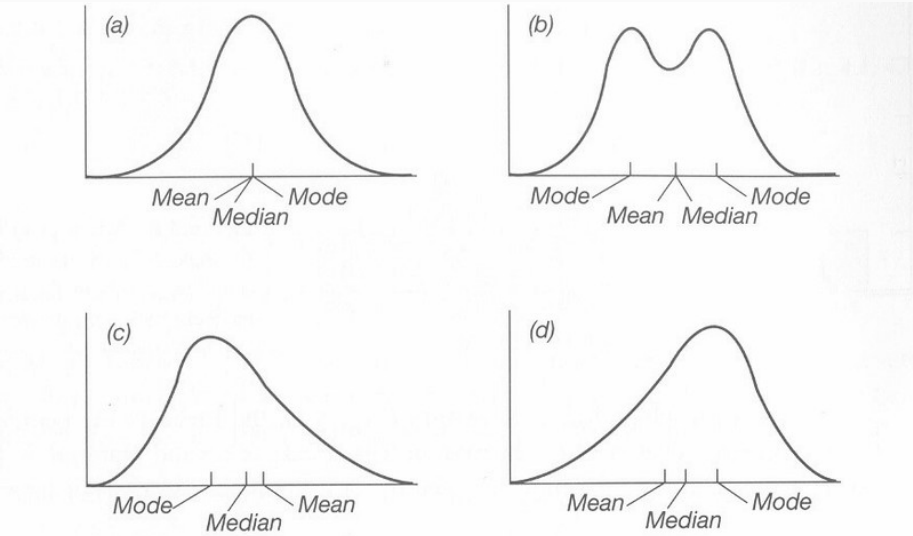


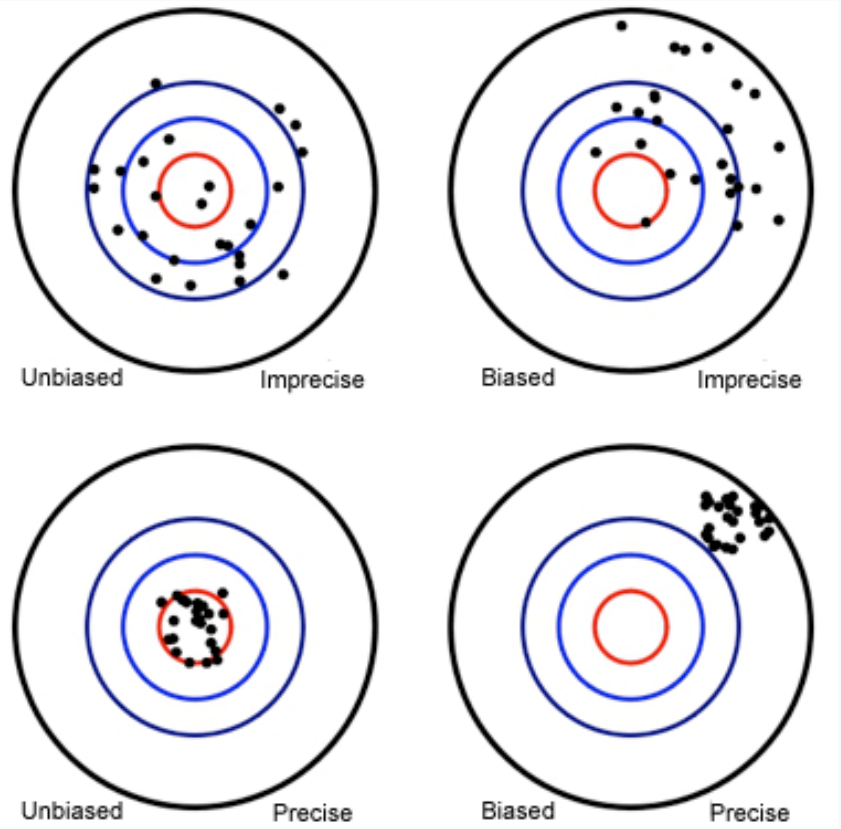
Figure 3.2 Frequency distributions showing measures of central tendency. Values of the variable are along the abscissa (horizontal axis), and the frequencies are along the ordinate (vertical axis). Distributions (a) and (b) are symmetrical, (c) is positively skewed, and (d) is negatively skewed. Distributions (a), (c), and (d) are unimodal, and distribution (b) is bimodal. In a unimodal asymmetric distribution, the median lies about one-third the distance between the mean and the mode.*

How to measure variations?

- Are you a good shooter?
- Vairance and standard deviation
- Population and samples

The (biased) samples variance is defined as

$$\sigma^2 = \frac{\sum_i (x_i - \mu)^2}{\sum_i} \qquad (2)$$



EXAMPLES/DISCUSSION

- How do we characterise skewed distributions?
- Concept of moments
- Distributions outside law of large numbers
- Examples can be found at `https://github.com/tbs1980/CISLQuantWorkshop/tree/master/AdvancedQuantitativeMethodsClinic/examples`
- Use the rest of the time for examples/discussion.

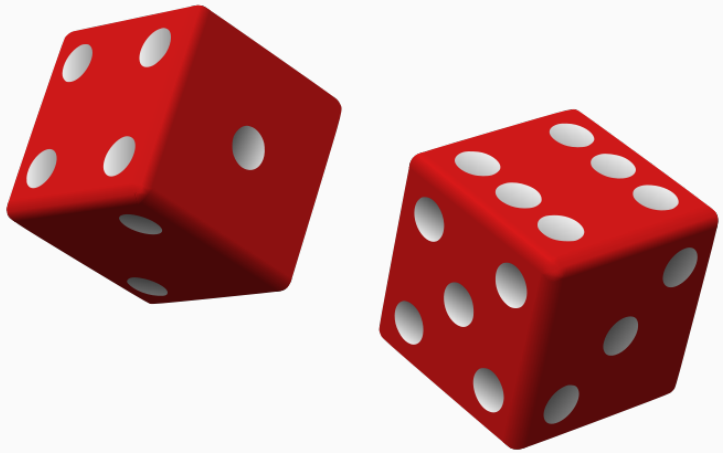
INFERENTIAL STATISTICS

INFERENTIAL STATISTICS

- Probability
- The Normal distribution
- Sample means and their distribution
- Introduction to hypothesis testing

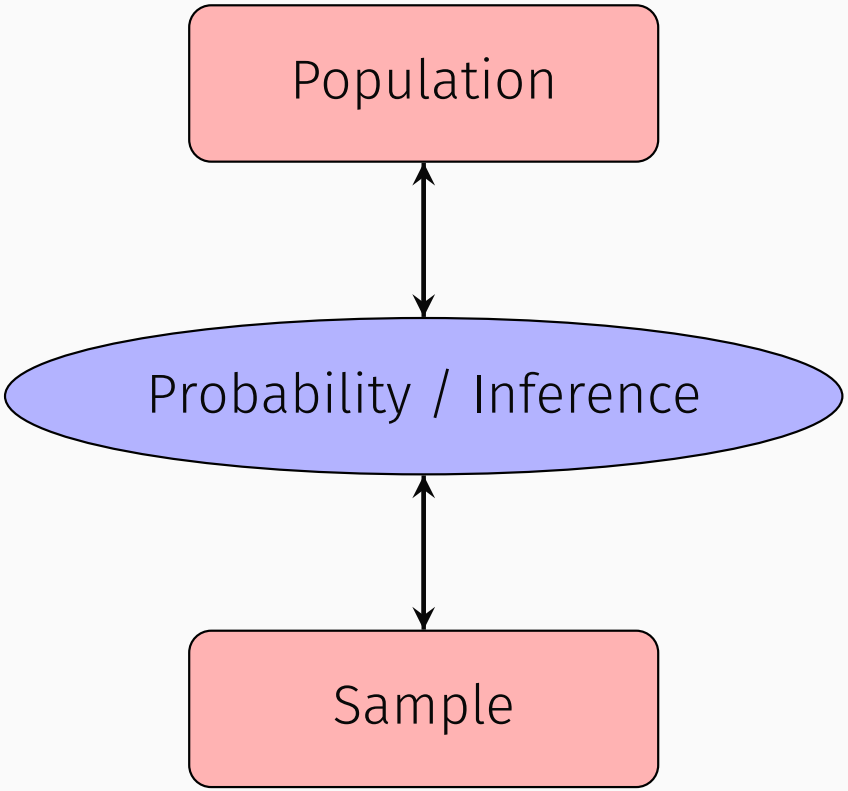
Frequency or degree of belief?

- Frequency
- Desired outcome
- Random sample



ROLE OF PROBABILITY

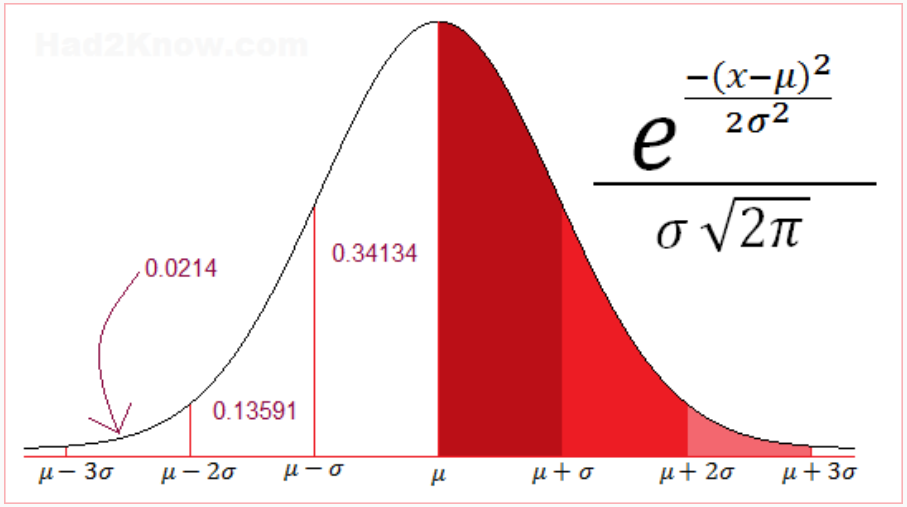
- What kind of samples are likeliy to obtained from the population?
- What can we say about the population given a sample?



Characteristics

- Mean μ
- Standard deviation σ
- Why is it important?
- Distribution of sample means

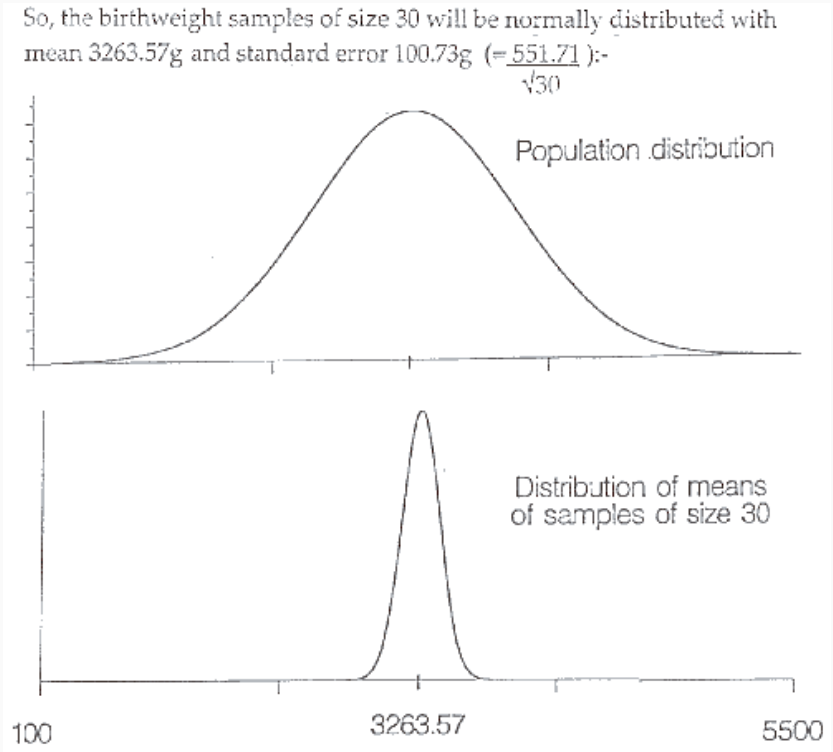
$$\Pr(x|\mu,\sigma) = \frac{1}{\sigma\sqrt{2\pi}} \exp\left(-\frac{1}{2}\frac{(x-\mu)^2}{\sigma^2}\right) \tag{3}$$



SAMPLE MEANS AND THEIR DISTRIBUTION

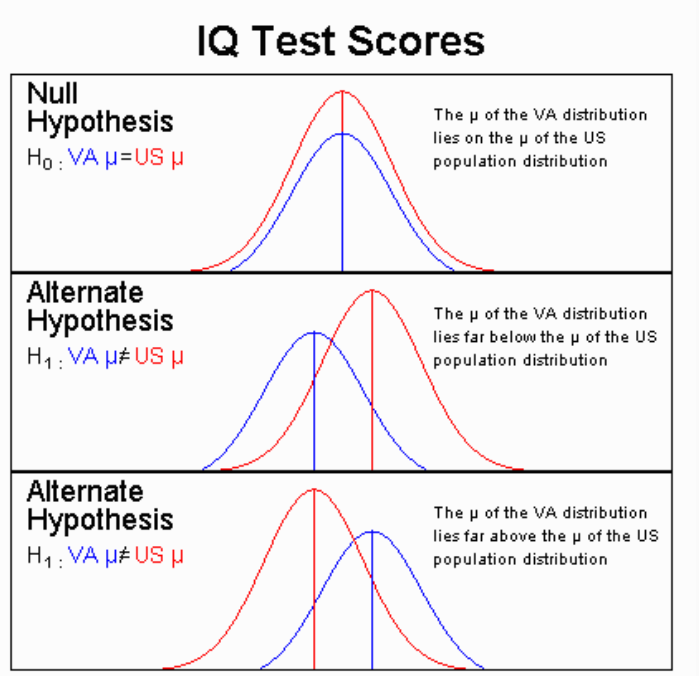
Characteristics

- Sampling error
- Distribution of sample means
- Expected value
- Standard error
- Law of large numbers



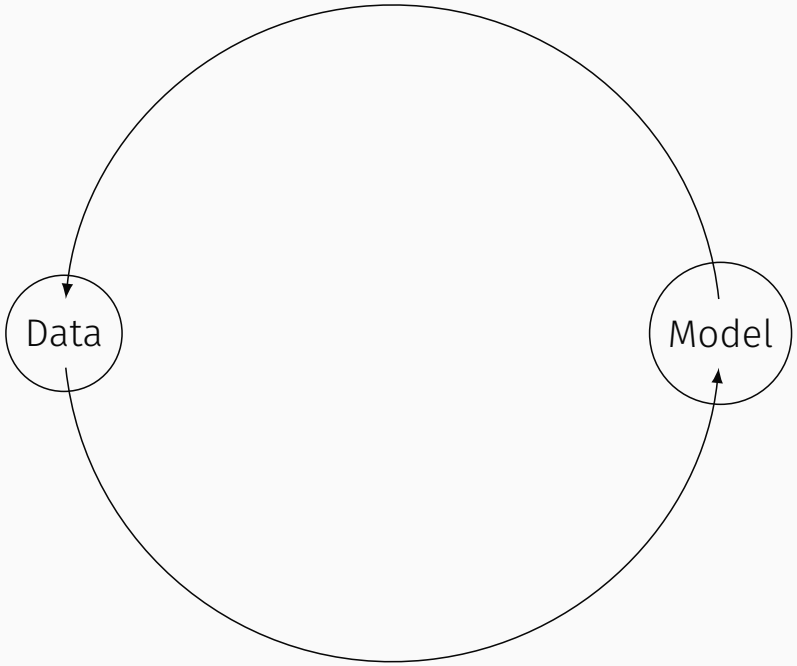
Baic idea

- Known versus unknown
- Null versus alternative hypothesis
- Decision crieteria
- Level of significance
- Critical region
- Uncertanity and errors
- Statistical significance



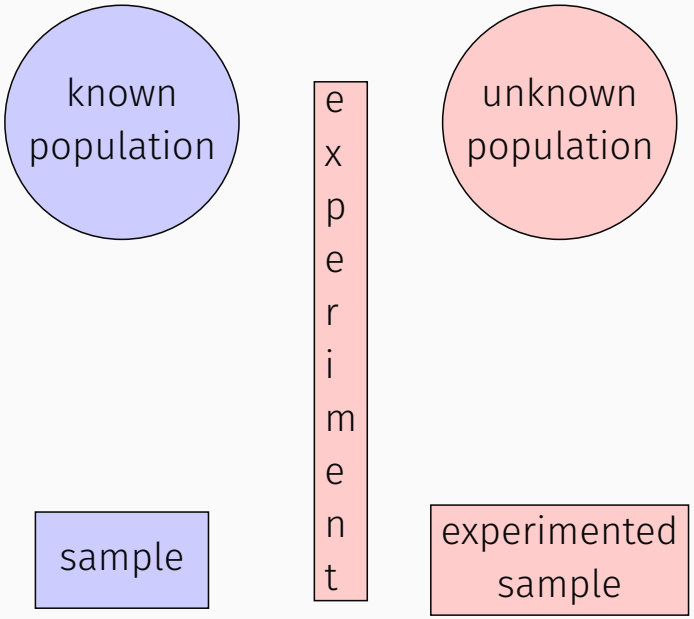
Questions

- Can we observe meaningful patterns in the data
- Are the findings statistically significant?
- Does the model adeqately describe the data?
- Is there evidence for an alternative hypothesis?

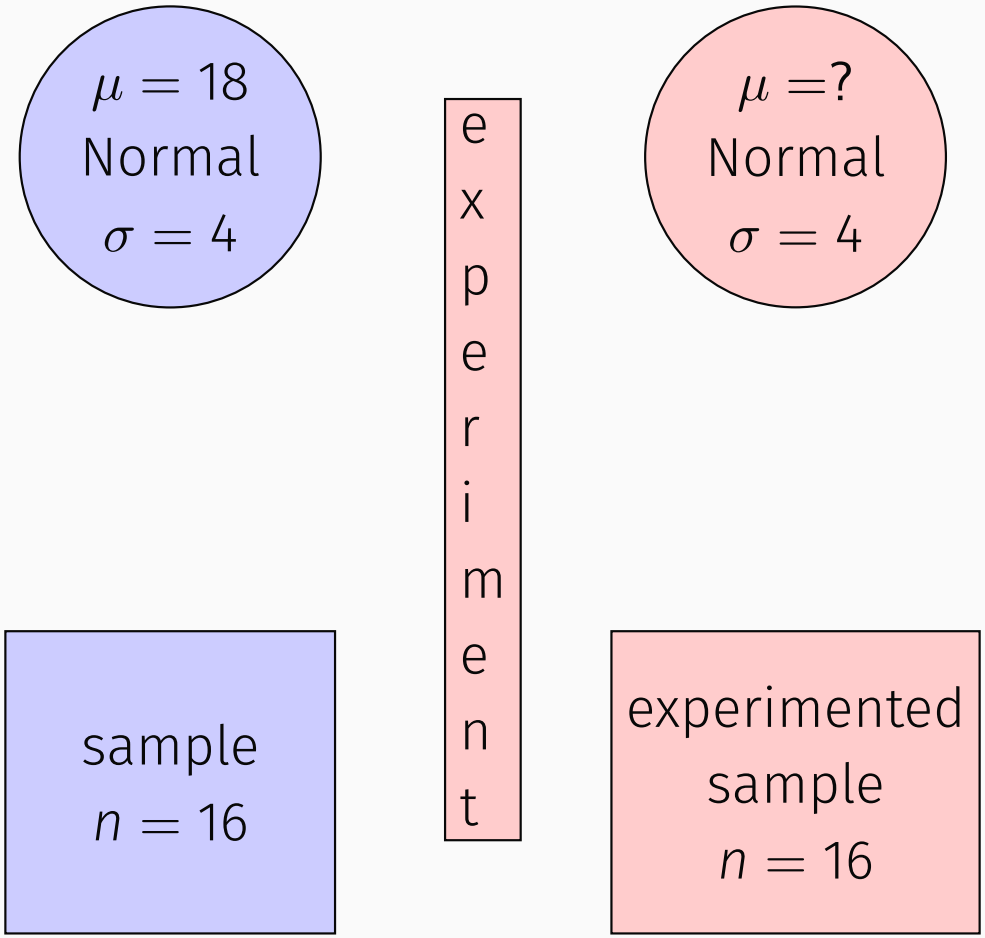


Questions

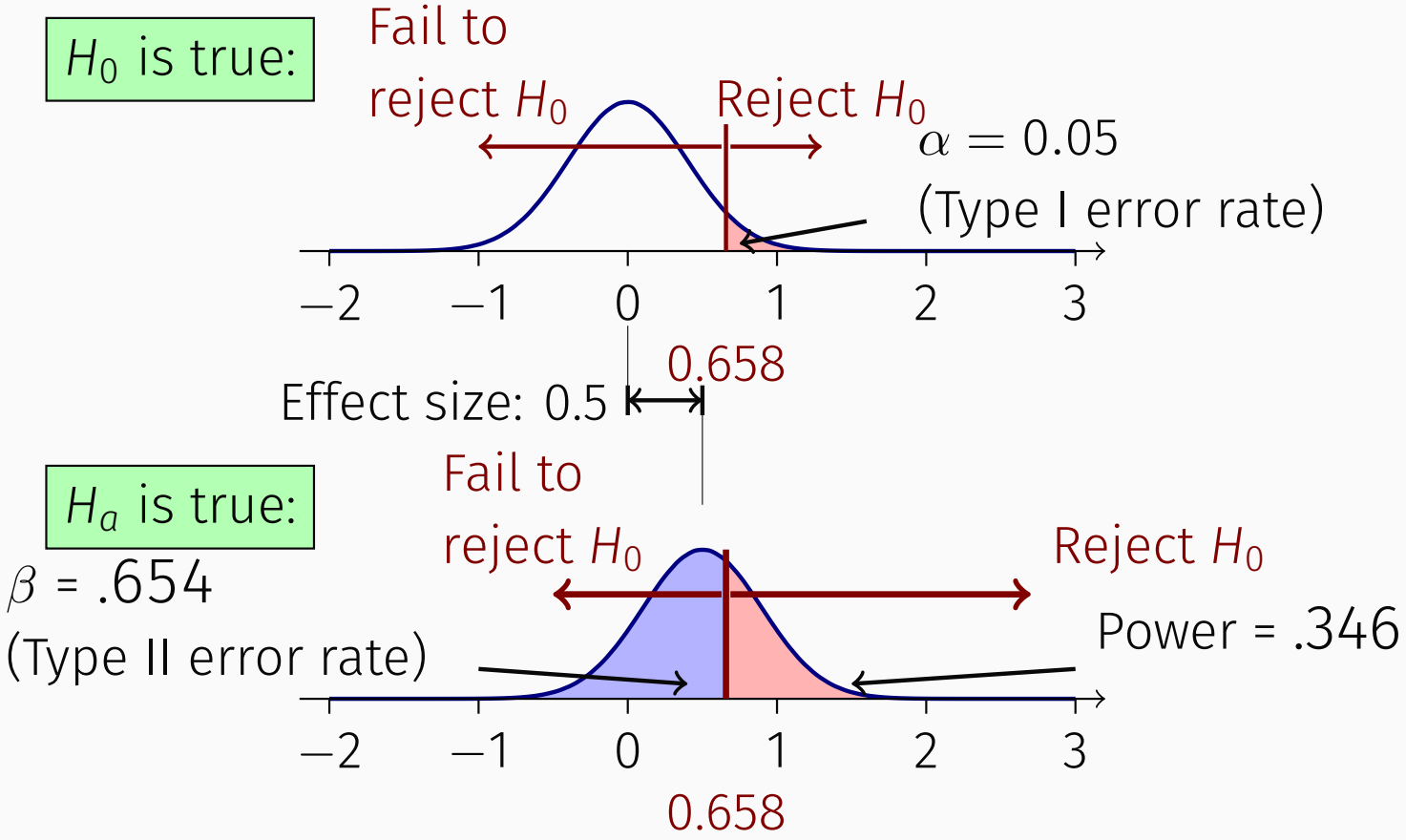
- Known characterstics of a population
- Selected sample for research
- Characterstics of the sample after experiment
- How do they compare?



COMPARING CHANGES IN μ



COMPARING CHANGES IN μ: STATISTICAL ODDS



EXAMPLES/DISCUSSION

- How accurate is the σ invariance assumption?
- How will we choose the level of significance?
- Examples can be found at `https://github.com/tbs1980/CISLQuantWorkshop/tree/master/AdvancedQuantitativeMethodsClinic/examples`
- Use the rest of the time for examples/discussion.

INFERENCES ABOUT POPULATION MEANS

INFERENCES ABOUT POPULATION MEANS

- *t*-statistic
- Analysis of Variacne (ANOVA)

REGRESSION

REGRESSION

- Parametric
- Correlation
- Non-parametric

DISCUSSION

DISCUSSION AND SUMMARY

- Descriptive and inferential statistics
- Hypotheis testing
- Regression
- Best practices