Lecture 4: Statistics For Transcriptomics

BIOINF3005/7160: Transcriptomics Applications

Dr Stephen Pederson

Bioinformatics Hub,
The University of Adelaide

March 23rd, 2020

Hypothesis Testing

Hypothesis Testing

Hypothesis Testing

In biological research we often ask:

"Is something happening?" or "Is nothing happening?"

We might be comparing:

- Cell proliferation in response to antibiotics in media
- mRNA abundance in two related cell types
- Methylation levels across genomic regions
- Allele frequencies in two populations

Hypothesis Testing

How do we decide if our experimental results are "significant"?

- Is it normal variability?
- What would the data look like if our *experiment had no effect?*
- What would our data look like if there was *some kind of effect?*

Every experiment is considered as a random sample from all possible repeated experiments.

Sampling

Most experiments involve measuring something:

- Discrete values e.g. read counts, number of colonies
- Continuous values e.g. Ct values, fluorescence intensity

Every experiment is considered as a random sample from all possible repeated experiments.

Sampling

Many data collections can also be considered as experimental datasets

Example 1

In the 1000 Genomes Project a risk allele for T1D has a frequency of $\pi = 0.07$ in European Populations.

Does this mean, the allele occurs in exactly 7% of Europeans?

Sampling

Example 2

In our in vitro experiment, we found that 90% of HeLa cells were lysed by exposure to our drug.

- Does this mean that exactly 90% of HeLa cells will always be destroyed?
- What does this say about in vivo responses to the drug?

Population Parameters

- Experimentally-obtained values represent an **estimate** of the true effect
- More formally referred to as *population-level parameters*
- Every experiment is considered a *random sample of the complete population*
- Repeated experiments would give a **different** *(but similar)* estimate