I) What counts as a difference - this is a statistical philosophical question of two parts
   A) How shall we compare measurements - statistical methodology
   B) What will count as a significant difference - philosophical question and as such subject to convention

II) Statistical Methodology - General
   A) Null hypotheses - H₀
      1) In most sciences, we are faced with:
         a) NOT whether something is true or false (Popperian decision)
         b) BUT rather the degree to which an effect exists (if at all) - a statistical decision.
      2) Therefore 2 competing statistical hypotheses are posed:
         a) H₀: there is no difference in effect between (usually posed as < or >)
         b) H₁: there is a difference in effect between

Statistical tests: a set of rules whereby a decision about hypotheses is reached (either do or don’t reject)
1) Associated with rules - some indication of the accuracy of the decisions -- that measure is a probability statement -- p-value
2) Statistical hypotheses:
   a) Critical p-values indicate what counts as an acceptable level of uncertainty – set by convention, usually 0.05
   b) Example: if critical p-value = 0.05, this means that we are unwilling to reject the Null hypothesis (i.e. no effect) unless:
      1) we are 95% sure that H₁ is correct, or equivalently that
      2) we are willing to accept an error rate of 5% or less that we are wrong when we do not reject H₀
We sample mussel settlement and produce two frequency distributions:

Are these different??

1) Assume the null hypothesis is true
   (i.e., if we don’t detect a difference, then null $H_0$ of no difference between distributions – inside vs. outside – is true)

2) Compare measurements
generally this means calculating the mean difference between two sample distributions (e.g. the numbers from the experiment or set of observations), summarized as a test statistic (e.g., t or F).

3) Determine the probability that the difference between our two distributions is due to chance alone, rather than the hypothesized effect (translate test statistic to calculated p-value)

4) Compare calculated p-value with a critical p-value to assign significance (i.e., whether accept or reject null hypothesis)
Are they different???

How do you compare distributions?

a) Generally assume (but test) that distributions are the same shape (usually normal – “bell” – “parametric”)

b) Compare means, taking into account the confidence you have in your estimate of the means…

By estimating the error (variation) associated with your estimate:

Standard Error of the Mean (SEM) = variation / sample size

Standard Deviation (sd) /\sqrt{n}

where, sd = \sqrt{\frac{\sum (X_i - \bar{X})^2}{n - 1}}

Then we calculate a standardized difference between the means of the two distributions (again, taking into account their error)

\frac{\bar{X}_{\text{inside}} - \bar{X}_{\text{outside}}}{\sqrt{\frac{2 \cdot S^2}{n}}}

S, is the variance, = (sd)^2

And ask…

“what is the likelihood that the calculated difference (given our error) is due to random chance (or, alternatively, reflects a real difference between the two sample distributions)?”

Probability that estimated difference is not due to random chance (i.e. no real difference, H_0 is true) is the calculated P-value, derived from a table or stats package

If \( P_{\text{calc}} > P_{\text{crit (0.05)}} \), then do not reject H_0

If \( P_{\text{calc}} < P_{\text{crit (0.05)}} \), then reject H_0
Example

$H_A =$ More mussel settlers inside adult distribution  
$H_o =$ No difference in mussel settlement inside vs outside adult distribution

<table>
<thead>
<tr>
<th>Calculated p-value ($\alpha$)</th>
<th>Critical p-value</th>
<th>Decision</th>
</tr>
</thead>
<tbody>
<tr>
<td>0.90</td>
<td>0.05</td>
<td>Do not reject $H_o$</td>
</tr>
<tr>
<td>0.50</td>
<td>0.05</td>
<td>Do not reject $H_o$</td>
</tr>
<tr>
<td>0.051</td>
<td>0.05</td>
<td>Do not reject $H_o$</td>
</tr>
<tr>
<td>0.049</td>
<td>0.05</td>
<td>Reject $H_o$, not $H_A$</td>
</tr>
</tbody>
</table>

But we still don’t know the probability of concluding that mussel settlement is greater inside the adult distribution when in fact it truly is greater inside distribution

Types of statistical error – Type I and II

Type I and Type II error:

1) By convention, the hypothesis tested is the null hypothesis (no difference between sample distributions)
   a) In statistics, assumption is made that the hypothesis is true (assume $H_o$ true = assume $H_A$ false)
   b) accepting $H_o$ (saying it is likely to be true) is the same as rejecting $H_A$ (falsification)
   c) Scientific method is to falsify competing alternative hypotheses (alternative $H_A$’s)
2) Errors in decision making

| Decision  | \begin{tabular}{l}
Truth
  
$H_o$ true \\
$H_o$ false
\end{tabular} | \begin{tabular}{l}
Accept $H_o$ \\
Type II error ($\beta$) \\
Type I error ($\alpha$)
\end{tabular} | \begin{tabular}{l}
Reject $H_o$ \\
no error (1-$\alpha$) \\
no error (1-$\beta$)
\end{tabular} |
|-----------|------------------|------------------|------------------|

Type I error - probability $\alpha$ that we mistakenly reject a true null hypothesis ($H_o$)  
Type II error - probability $\beta$ that we mistakenly fail to reject (accept) a false null hypothesis  
Power of Test - probability (1-$\beta$) of not committing a Type II error - The more powerful the test the more likely you are to correctly conclude that an effect exists when it really does (reject $H_o$ when $H_o$ false = accept $H_A$ when $H_A$ true).
Simply

Type I error = \( \alpha = \) Reject \( H_0 \) when \( H_0 \) true  
Accept \( H_A \) when \( H_A \) false 

Type II error = \( \beta = \) Accept \( H_0 \) when \( H_0 \) false  
Reject \( H_A \) when \( H_A \) true 

Power of Test = \( 1 - \beta = \) Reject \( H_0 \) when \( H_0 \) false  
Accept \( H_A \) when \( H_A \) true 

Control of Error

A) Type I error can be directly controlled.  
The acceptable level is set by the experimenter = critical p-value (often 0.05) 

B) Type II error is indirectly controlled by the design of the experiment (more about this later) 

1. Level of critical P-value (usually 0.05) 
2. Magnitude of difference (“effect”) between means 
3. Amount of natural variation in data 
4. Sample size (level of replication)
Statistical Power, effect size, replication and alpha

Statistical comparison of two distributions

Calculation of statistical distributions

Simulated distribution of mussel settlers - inside adult distribution

Samples = 100 cells (10 x 10)
Mean per sample = 25
Total settlers = 2500
Evaluate effect of **sample size** on calculation of Mean

Compare for sample size's of 5, 10, 20, 50, 99 cells
Iterate each sample size 50 times

Example: sample 10 sites to determine mean

<table>
<thead>
<tr>
<th>X₁</th>
<th>X₂</th>
</tr>
</thead>
<tbody>
<tr>
<td>21.5</td>
<td>25.8</td>
</tr>
</tbody>
</table>

... x fifty iterations

True (simulated) Mean = 25
Effect of number of observations on estimate of Mean

Frequency distributions of sample means

Notice how estimated mean approaches true mean with larger sample size! ("Central Limit Theorem")

Notice how variation around estimated mean increases with lower sample size!

Statistical Power, effect size, replication and alpha

Statistical comparison of two distributions

\[ \bar{y}_1 \quad \bar{y}_2 \]
Statistical Power, effect size, replication and alpha

Area under the curve = 1.00

Ho true: Distributions of means are truly the same

\[ t = \frac{\bar{y}_1 - \bar{y}_2}{s_p \sqrt{\frac{1}{n_1} + \frac{1}{n_2}}} \]
Ho true: Distributions of means are truly the same

1) Estimate of $\bar{y}_1 = \text{estimate of } \bar{y}_2$

$t = \frac{\bar{y}_1 - \bar{y}_2}{s_p \sqrt{\frac{1}{n_1} + \frac{1}{n_2}}}$

2) Estimate of $\bar{y}_1 \neq \text{estimate of } \bar{y}_2$

$t = \frac{\bar{y}_1 - \bar{y}_2}{s_p \sqrt{\frac{1}{n_1} + \frac{1}{n_2}}}$
Ho true: Distributions of means are truly the same

\[ t = \frac{\bar{y}_1 - \bar{y}_2}{s_p \sqrt{\frac{1}{n_1} + \frac{1}{n_2}}} \]

t distribution (df)

<table>
<thead>
<tr>
<th>Decision</th>
<th>Accept H.</th>
<th>Reject H.</th>
</tr>
</thead>
<tbody>
<tr>
<td>Truth</td>
<td></td>
<td></td>
</tr>
<tr>
<td>H. true</td>
<td>no error</td>
<td>Type I error</td>
</tr>
<tr>
<td></td>
<td>(1-alpha)</td>
<td>(alpha)</td>
</tr>
</tbody>
</table>

If distributions are truly the same then: area to right of critical \( t_c \) represents the Type 1 error rate (Blue); any calculated \( t > t_c \) will cause incorrect conclusion that distributions are different.

assign critical \( p \) (assume = 0.05)
Ho true: Distributions of means are truly the same

assign critical p (assume = 0.05)

$t = \frac{\bar{y}_1 - \bar{y}_2}{s_p \sqrt{\frac{1}{n_1} + \frac{1}{n_2}}}$

t distribution (df)

If distributions are truly the same then: area to right of critical $t_c$ represents the Type 1 error rate (Blue); any calculated $t > t_c$ will cause incorrect conclusion that distributions are different

Ho false: Distributions of means are truly different
$t = \frac{\bar{y}_1 - \bar{y}_2}{s_p \sqrt{\frac{1}{n_1} + \frac{1}{n_2}}}$

Ho false: Distributions of means are truly different
Ho false: Distributions of means are truly different

\[
t = \frac{\bar{y}_1 - \bar{y}_2}{s_p \sqrt{\frac{1}{n_1} + \frac{1}{n_2}}}
\]

Central t distribution
Assumes similarity of distributions

Non-central t distribution

If distributions are truly different then: area to left of critical \( t_c \) represents the Type II error rate (Red); any Calculated \( t < t_c \) will cause incorrect conclusion that distributions are same
How to control Type II error (distributions are truly different)
This will maximize statistical power to detect real impacts.

1) Vary critical P-Values
2) Vary Magnitude of Effect
3) Vary replication

<table>
<thead>
<tr>
<th>Decision</th>
<th>Truth</th>
<th>Accept H</th>
<th>Reject H</th>
</tr>
</thead>
<tbody>
<tr>
<td>H₁-true</td>
<td>Type I error (α)</td>
<td>(1-α)</td>
<td>Type II error (β)</td>
</tr>
<tr>
<td>H₁-false</td>
<td>Type II error (1-β)</td>
<td>Type I error (α)</td>
<td>(1-α)</td>
</tr>
</tbody>
</table>

Central t distribution

How to control Type II error (distributions are truly different)

1) Vary critical P-Values (change blue area)

Reference

A) Make critical P more stringent (smaller)
Type II error increases
Power decreases

A) Relax critical P (larger values)
Type II error decrease
Power increases
How to control Type II error (distributions are truly different)

2) Vary magnitude of effect (vary distance between $\bar{y}_1$ and $\bar{y}_2$ which affects non-central t distribution

- Critical $p$

<table>
<thead>
<tr>
<th>A) Make distance smaller</th>
<th>A) Make distance greater</th>
</tr>
</thead>
<tbody>
<tr>
<td>Type II error increases</td>
<td>Type II error decreases</td>
</tr>
<tr>
<td>Power decreases</td>
<td>Power increases</td>
</tr>
</tbody>
</table>

How to control Type II error (distributions are truly different)

3b) Vary replication (which controls estimates of error)

- Note Type 1 error is constant and $t_c$ is allowed to vary

<table>
<thead>
<tr>
<th>A) Decrease replication</th>
<th>A) Increase replication</th>
</tr>
</thead>
<tbody>
<tr>
<td>Type II error increases</td>
<td>Type II error decreases</td>
</tr>
<tr>
<td>Power decreases</td>
<td>Power increases</td>
</tr>
</tbody>
</table>
How to estimate optimal sample size

1) Do a preliminary study of variables that will be evaluated in project

2) Plot the mean and some estimate of variability of data as a function of sample size

3) Look for sample size where estimates of mean and variance (or standard deviation) converge on a stable value

4) Calculate a “bang for buck” relationship to determine if a robust design (sufficient sampling) can be paid for