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 - Ho
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) HA: there is a difference in effect between (usually posed as < or >)
   b) H0: there is no difference in effect between

Specific hypothesis HA
Number of mussel settlers is greater in areas inside adult distribution than outside it

Assessment - Deductive reasoning

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 Ha 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 Ha.
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) = \[ \sqrt{\frac{\sum (X_i - \bar{X})^2}{n - 1}} \]

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

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

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 net-due to random chance (i.e. no real difference, H0 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 H0

If \( P_{\text{calc}} < P_{\text{crit}} (0.05) \), then reject H0

Example

\( H_0 = \) More mussel settlers inside adult distribution
\( H_0 = \) No difference in mussel settlement inside vs outside adult distribution

Calculated p-value (α) | Critical p-value | Decision
--- | --- | ---
0.90 | 0.05 | Accept H0
0.50 | 0.05 | Accept H0
0.051 | 0.05 | Accept H0
0.049 | 0.05 | Reject H0

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 error - probability of a Type I error (α)

Type II error - probability of a Type II error (β)

Power of Test - probability (1-β) of 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 H0 when H0 false – accept H0 when H0 true).
More simply

Type I error = $\alpha = \text{Reject } H_0 \text{ when } H_0 \text{ true}$
$\text{Accept } H_A \text{ when } H_A \text{ false}$

Type II error = $\beta = \text{Accept } H_0 \text{ when } H_0 \text{ false}$
$\text{Reject } H_A \text{ when } H_A \text{ true}$

Power of Test = $1-\beta = \text{Reject } H_0 \text{ when } H_0 \text{ false}$
$\text{Accept } H_A \text{ when } H_A \text{ true}$

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 -

Calculation of statistical distributions

Simulated distribution of mussel settlers - inside adult distribution

Statistical comparison of two distributions

Samples = 100 cells (10 x 10)
Mean per sample = 25
Total settlers = 2500
SAMPLE SIZE can affect amount of error in measurements

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

\[
\begin{align*}
X_1 &= 21.5 \\
X_2 &= 25.8
\end{align*}
\]

\[
\text{... x fifty iterations}
\]

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!

Type 1 and Type 2 error

H₀ true: Distributions of means are truly the same

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

<table>
<thead>
<tr>
<th>Decision</th>
<th>Truth</th>
<th>Type 1 error</th>
</tr>
</thead>
<tbody>
<tr>
<td>Accept H₀</td>
<td>H₀ true</td>
<td>no error</td>
</tr>
<tr>
<td>Reject H₀</td>
<td>H₀ false</td>
<td>Type I error (1-alpha)</td>
</tr>
</tbody>
</table>

If distributions are truly different then: area to left of critical \( t \): represents the Type II error rate (Red); any calculated \( t \ < t_c \) will cause incorrect conclusion that distributions are same

<table>
<thead>
<tr>
<th>Decision</th>
<th>Truth</th>
<th>Type II error</th>
</tr>
</thead>
<tbody>
<tr>
<td>Accept H₀</td>
<td>H₀ false</td>
<td>no error</td>
</tr>
<tr>
<td>Reject H₀</td>
<td>H₀ true</td>
<td>Type II error (1-beta)</td>
</tr>
</tbody>
</table>

assign critical \( p \) (assume = 0.05)
How to control Type II error (distributions are truly different)

This will maximize statistical power to detect real impacts

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

- Make critical P more stringent (smaller)
- Make critical P (larger values)

2) Vary magnitude of effect (vary distance between means)

- Decrease replication
- Increase replication

3) Vary replication (which controls estimates of error)

- Note Type 1 error is constant and t_c is allowed to vary
How to estimate optimal sample size (and control error rates)

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