**A primer on using parametric statistics in biological research**

Problem: How to determine if two population means differ. (e.g., Does the relative growth rate of cotton bollworms differ on corn vs. tobacco?)

1. Sample two populations.
2. Estimate true population means ($\mu_1$ and $\mu_2$) from sample means ($\bar{x}_1$ and $\bar{x}_2$).
3. Estimate precision of sample means (i.e., how accurately $\bar{x}$ approximates $\mu$) from variability among replicate samples.

$$SD = \text{Standard deviation}$$

= average amount by which a sample differs from sample mean. (In a normal distribution, 68% of population lies within ±1 SD of the mean, and 96% lies within ±2 SDs of mean.)

$$SD = \sqrt{\frac{\sum_{i=1}^{n} (x_i - \bar{x})^2}{n - 1}}$$

where: $x_i$ = each sample observation
$\bar{x}$ = sample mean
$n$ = sample size

$$SE = \text{Standard error of the mean}$$

= average amount by which a sample mean ($\bar{x}$) differs from true population mean ($\mu$). (Given a normal distribution and random, independent samples, ±68% of the possible sample means lie within ±1 SE of the true mean, and ±96% lies within ±2 SE of mean.)

$$SE = \frac{SD}{\sqrt{n}}$$

4. Construct null hypothesis
   (e.g., growth rate on corn = growth rate on tobacco or $\mu_1 = \mu_2$)
5. Reject null hypothesis if sample means and standard errors indicate less than a 5% chance it is true (i.e., if $P < 0.05$).
   As an informal criterion, reject null hypothesis if two sample means differ by more than two standard errors
   e.g., if $|\bar{x}_1 - \bar{x}_2| > (SE_1 + SE_2)$
   Else “fail to reject null hypothesis”.
   Formal statistical tests = t-test with two samples or analysis of variance (ANOVA) with >2 samples.

Assumptions of this statistical procedure

1. Sampled population has a normal distribution
2. Samples were collected randomly (i.e., all possible samples within the population had an equal probability of being selected)
3. Replicate samples were independent (e.g., deviations from $\mu_i$ in the first replicate sample at 1 m were unrelated to deviations from $\mu_i$ in the second, third, fourth, fifth and sixth replicate sample at 1 m).