Hypothesis Testing

Two Sample

Null Hypothesis

$H_0: \mu_1 = \mu_2$

$H_0: \mu_1 - \mu_2 = 0$ Not Different

$H_0: \delta = 0$

Alternative Hypothesis

$H_A: \mu_1 \neq \mu_2$

$H_A: \mu_1 - \mu_2 \neq 0$ Different

$H_A: \delta \neq 0$

Hypothesis Testing

Error Types

	Reject H_0	Accept H_0
H_0 True	Type I Error (False Positive)	Correct
H_0 False	Correct	Type II Error (False Negative)

Hypothesis Testing

Error Types

Decision	Truth		
	$\delta < 0$	$\delta = 0$	$\delta > 0$
$\delta < 0$	**Correct**	Type I	Type III
$\delta = 0$	Type II	**Correct**	Type II
$\delta > 0$	Type III	Type I	**Correct**

$\delta = \mu_1 - \mu_2$

Hypothesis Testing

Probabilities

Alpha

α = P(Type I Error)

α = P(Reject H_0 | H_0 is True)

Beta

β = P(Type II Error)

β = P(Accept H_0 | H_0 is False)

Power

= 1 - β

Comparing Two Sample Means

t-tests

$$t = \frac{\bar{x}_1 - \bar{x}_2}{s_{\bar{x}_1 - \bar{x}_2}} = \frac{\bar{d}}{s_{\bar{d}}}$$

$$S_{\bar{d}} = \sqrt{\frac{2s_p^2}{n}}$$

$$df = (n_1 - 1) + (n_2 - 1)$$

$$H_0 : \bar{x}_1 - \bar{x}_2 = \bar{d} = 0$$

Switchgrass

Variety Trial (Yield, Mg/ha)

	Cultivar	
Replication	A	B
1	10.8	8.9
2	12.7	8.9
3	11.4	8.6
4	13.6	8.8
5	13.9	8.5
6	13.0	6.1
Mean	12.6	8.3
Variance	1.46	1.24

Switchgrass Comparison

t-test

$$H_0 : \mu_A = \mu_B$$

$$t = \frac{12.571 - 8.300}{\sqrt{\frac{2(1.349)}{6}}} = \frac{4.271}{0.671} = 6.37$$

- The probability that $t_{1,10\ df} > 6.37$ is 0.000081.
- Therefore, the probability of the two means being from the same population is very low.

Comparing Two Sample Means

Standard Errors (of the difference)

if $s^2_1 = s^2_2$ and $n_1 = n_2$, then:

$$S_{\bar{d}} = \sqrt{\frac{2s_p^2}{n}}$$

if $s^2_1 = s^2_2$, then:

$$S_{\bar{d}} = \sqrt{s_p^2\left(\frac{1}{n_1} + \frac{1}{n_2}\right)}$$

Comparing Two Sample Means

Unequal Variances

if $s^2_1 \neq s^2_2$, then:

$$S_{\bar{d}} = \sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}}$$

Satterthwaite's Approximation:

$$df = \frac{\left(S_{\bar{x}_1}^2 + S_{\bar{x}_2}^2\right)^2}{\frac{\left(S_{\bar{x}_1}^2\right)^2}{n_1 - 1} + \frac{\left(S_{\bar{x}_2}^2\right)^2}{n_2 - 1}}$$

Hypothesis Testing

Relative Consequences of Type I & II Errors

	Low probability of false negative ↓ $\beta < 0.05$ Power high ↑	High probability of false negative ↑ $\beta > 0.05$ Power low ↓
Low probability of false positive $\alpha < 0.05$ ↓	**Ideal situation** Frequently not achievable, but can be done with large effect sizes	Typical of many published experiments in agronomy
High probability of false positive $\alpha > 0.05$ ↑	Exploratory research Positive effects are often re-evaluated	**Not an ideal situation** Precision may be improved by increasing sample size or reducing experimental complexity

Source: Kimberly Garland Campbell. 2018. Chapter 1 - Errors in Statistical Decision Making. Applied Statistics in Agricultural, Biological, and Environmental Sciences. ASA-CSSA-SSSA.

Hypothesis Testing

Relative Consequences of Type I & II Errors

Agronomy Example	Low probability of false negative ↓ $\beta < 0.05$ Power high ↑	High probability of false negative ↑ $\beta > 0.05$ Power low ↓
Low probability of false positive $\alpha < 0.05$ ↓	Growers adopt new variety and are happy.	Growers don't adopt new variety. Cost is associated with unrealized potential profit from increased yields of new variety.
High probability of false positive $\alpha > 0.05$ ↑	Growers adopt new variety, but it doesn't perform better than old variety. Cost differential depends on relative cost of seed. Or Type 3 error occurs, and new variety actually performs worse than old variety.	Experiment is too variable to make a decision. Resources for trialing wasted.

Source: Kimberly Garland Campbell. 2018. Chapter 1 - Errors in Statistical Decision Making. Applied Statistics in Agricultural, Biological, and Environmental Sciences. ASA-CSSA-SSSA.

Switchgrass Comparison

t-test

$$\bar{d}_{.05} = 2.008\sqrt{\frac{2(1.349)}{6}} = 3.41$$

- The lowest difference that can be detected at α=0.05 is 3.41.
- You can be confident that you have the power to detect larger differences since there is no possibility of making a Type II error.
- What if you wanted to be able to detect smaller differences?

Hypothesis Testing

Power

Power = 1 – β

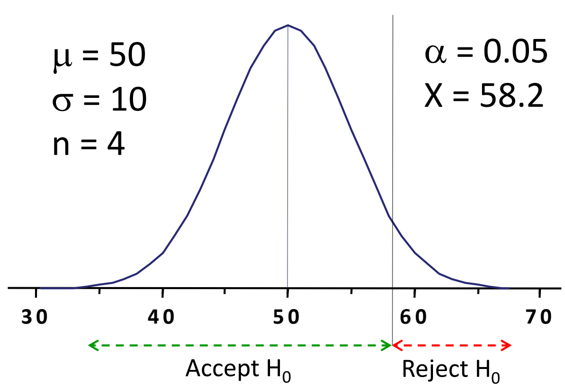
The ability to detect true differences

<u>Factors</u>:

- H_A is true
- Magnitude of mean difference
- Alpha – inversely related
- Variance
- Sample size

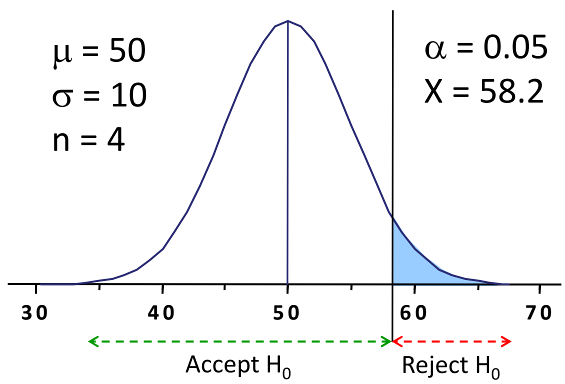
Hypothesis Testing

Alpha



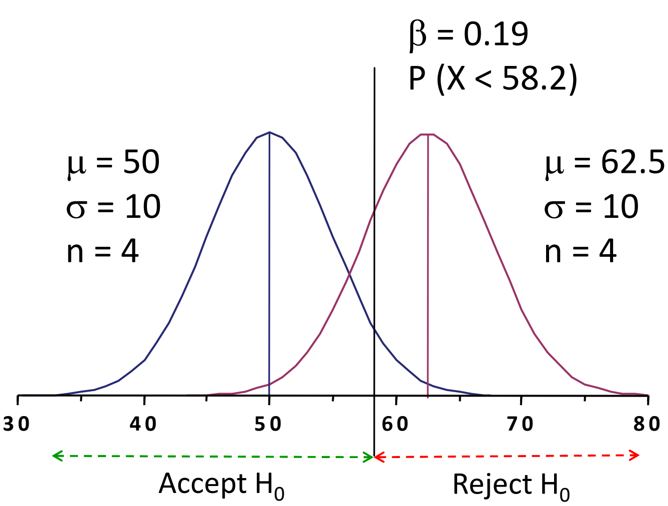
Hypothesis Testing

Alpha



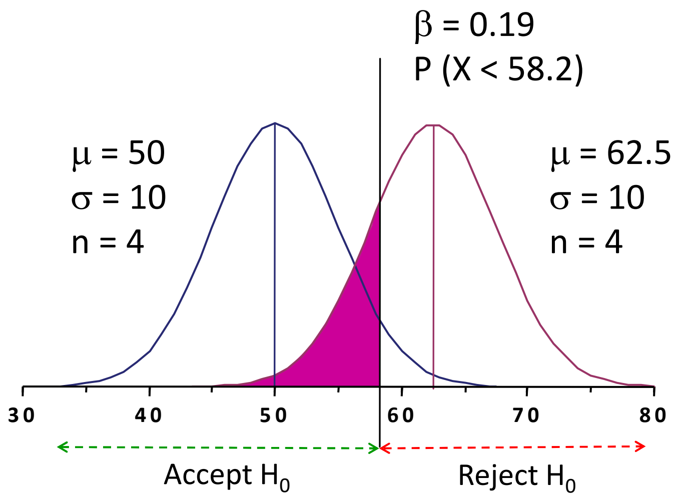
Hypothesis Testing

Beta



Hypothesis Testing

Beta



Improving Power

The power to detect a mean difference increases when:

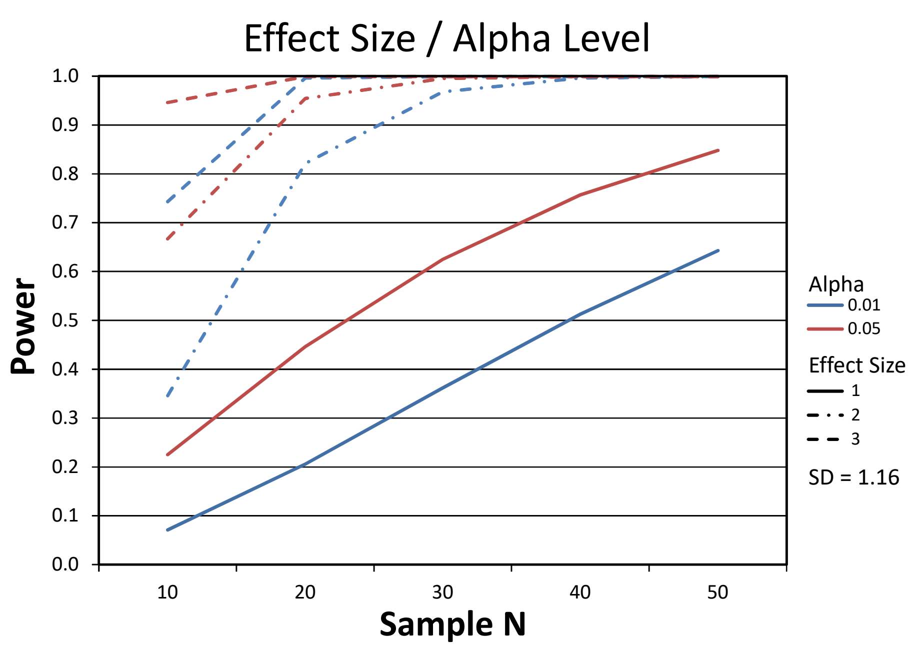
- The magnitude of the difference increases
- The alpha level is relaxed (increased)
- The number of replicates increases
- The precision ($1/\sigma^2$) increases (i.e. σ^2 decreases)

Effect Size

Statistical significance does not equate to biological or economic significance.

- It means that the ratio of the treatment variance to error variance is large enough to be considered unlikely to occur randomly.
- A difference (effect size) that is considered statistically important may not be practically so.
- Conversely, a difference that cannot be detected statistically may be quite important biologically or economically.
- ∴ It is important to think about what effect sizes you need to be able to detect in the design process.

Power Curve



Hypothesis Testing

Power and Replication

$$n = \frac{(t_{\alpha/2} + t_{\beta})^2 \sigma_D^2}{\delta^2}$$

Where:

- n = minimum number of replicates
- $t_{\alpha/2}$ = alpha
- t_{β} = beta
- σ^2_D = variance of differences ($2\sigma^2$)
- δ^2 = minimum difference

Power and Replication

Number of reps required to detect mean differences

δ	**CV**				
% Mean	**5**	**10**	**15**	**20**	**25**
5	16				
10	4	16			
15	2	7	16		
20	1	4	9	16	
25	1	3	6	10	16

$\alpha = 0.05$, $\beta = 0.2$, n = 6*

* Conservative

Mean Comparisons

Improving Sensitivity

$$t = \frac{\bar{x}_1 - \bar{x}_2}{s_{\bar{x}_1 - \bar{x}_2}} = \frac{\bar{d}}{s_{\bar{d}}} \qquad S_{\bar{d}} = \sqrt{\frac{2s_p^2}{n}}$$

The sensitivity of an experiment for detecting treatment differences depends on the magnitude of the variance (S^2) and the number of replications (n).

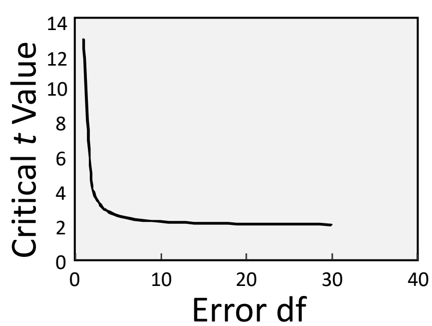
There are two approaches for improving sensitivity:

1. Increase the number of reps used to estimate S^2
2. Decrease the S^2 by using design control

Mean Comparisons

Improving Sensitivity - Replication

$$LSD = t_{.05} \times S_{\bar{d}} = t_{.05} \times \sqrt{\frac{2MS_E}{r}} = t_{.05} \times \sqrt{2} \times \frac{RMSE}{\sqrt{r}}$$



General Rule – the sensitivity of an experiment doubles for each 4-fold increase in the number of replications.

Mean Comparisons

Improving Sensitivity – Plot Size

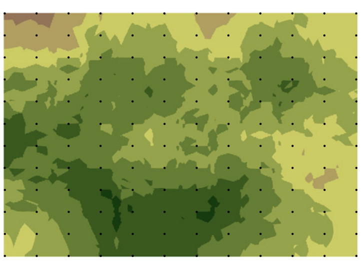
Smith's Formula:

$$V_s = \frac{V}{s^b} \qquad \ln(V_n) = \ln(V_1) - b\ln(n)$$

where:
V_S = variance among plots of size *s*
b = soil heterogeneity index

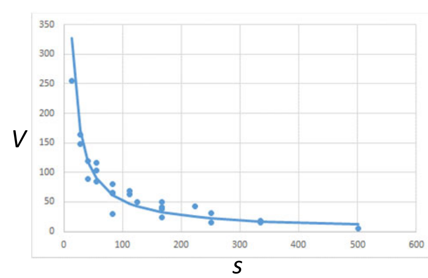
Mean Comparisons

Improving Sensitivity – Plot Size



Corn yield (bu/acre)

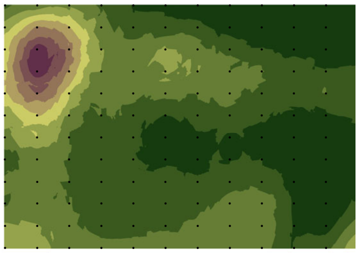
- 239.3000031 - 245
- 245.0000001 - 250
- 250.0000001 - 255
- 255.0000001 - 260
- 260.0000001 - 265
- 265.0000001 - 270
- 270.0000001 - 275
- 275.0000001 - 280
- 280.0000001 - 285



$$V_s = \frac{3733.1}{s^{0.9257}}$$

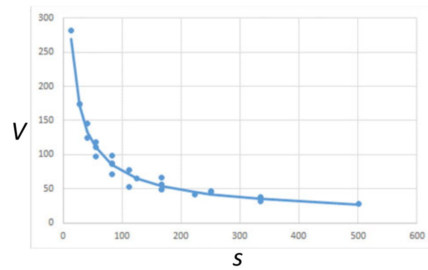
Mean Comparisons

Improving Sensitivity – Plot Size



Corn yield (bu/acre)

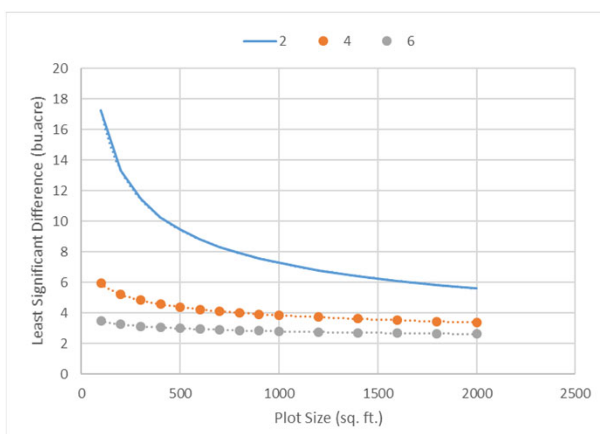
210.8999939 - 225
225.0000001 - 230
230.0000001 - 235
235.0000001 - 240
240.0000001 - 245
245.0000001 - 250
250.0000001 - 255
255.0000001 - 260
260.0000001 - 265



$$V_s = \frac{1453.9}{s^{0.6415}}$$

Mean Comparisons

Improving Sensitivity – Plot Size & Replication



$$V_s = \frac{2269.0}{s^{0.7506}}$$

Mean Comparisons

Factors Effecting Plot Size

Adapted from Petersen, 1994

Factor	Small plots →→→ Large Plots
Soil variability	Uniform →→→→→→→ Heterogeneous
Crop	Turf → Cereals → Row crops → Pasture
Research phase	Basic →→ Developmental →→ Adaptive
Experiment type	Breeding → Fertility → Tillage → Irrigation
Machinery	None →→→ Research →→→ Farm scale

Mean Comparisons

Improving Sensitivity – Design Control

Design controls are used to decrease the value of S^2 by accounting for some of the unexplained variation in the measured response and partitioning it out of the error mean square (S^2). The net effect is to make S^2 smaller.

There are two common approaches for doing this:

- Blocking methods that partition some of the error SS to a blocking factor
- Using covariates to explain some of the variation not accounted for by treatments

One-Factor ANOVA

Switchgrass Example

Source	DF	Sum of Squares	Mean Square	F Value	Pr > F
Cultivar	9	355.5271	39.50301	4.57	0.0002
Error	50	432.5709	8.651419		
Total	59	788.098			

Cultivar	Mean Yield (Mg/ha)
A	12.6
B	8.3
C	14.2
D	12.7
E	9.4
F	6.7
G	13.9
H	9.6
I	9.4
J	9.1

We know from the ANOVA that there is at least one significant difference among pairs of means. How do we find out which pairs actually differ?