NAICC 2019 Annual Meeting and Ag Pro Expo
January 15, 2019
Savannah, GA

Trial Designs and Statistics to Obtain Reliable Results

Julien M. Beuzelin
Everglades Research and Education Center
Belle Glade, FL

UF IFAS
UNIVERSITY of FLORIDA

My background

- **I am an entomologist**
  - PhD in Entomology, minor in experimental statistics
  - I am not a statistician but I use experimental design and statistics to do my job

- **We all face comparable challenges**

A possible scenario

- **Efficacy evaluation**
  - 3 treatments + non-treated check
  - RBD with 4 replications
  - Apply treatments once
  - Collect data from 10 plants/plot 3, 7, and 14 DAT
  - Collect whole plot yield data

- **A potential approach**
  - Get protocol, enter protocol in ARM or Field Pro
  - Conduct trial
  - Enter data in software, run standard ANOVA and separate means with LSD
  - Look at mean separations on each date using letters

Are we using trial designs & statistics that will provide reliable results?

- Why do we have an RBD?
- Why do we have 4 reps?
- Why do we take 10 plants/plot?
- How do we consider the impact of date?
- How do we compare treatments?

- What does this mean?
- Should it be done differently?

Experimental design and data analysis are dependent

The goal of a trial is to compare experimental treatments

- **We measure variables of interest for this comparison** (e.g., yield for each treatment)
- **Variables can describe populations** (e.g., all fields treated by treatments)

- **These variables have a mean (location) and variance (variability)**

The goal of a trial is to compare experimental treatments

- **We cannot measure variables for all individuals in the populations** (e.g., yield in all fields treated by experimental treatments)
- **We work on subsets of populations (samples)** (e.g., trial plots and samples in trial plots)

- **Variables in samples also have means and variances, which are estimates of population means and variances: We can make inferences about populations**

The goal of a trial is to compare experimental treatments

- The sample means are also variables that have distributions with same means as populations and a bell shape

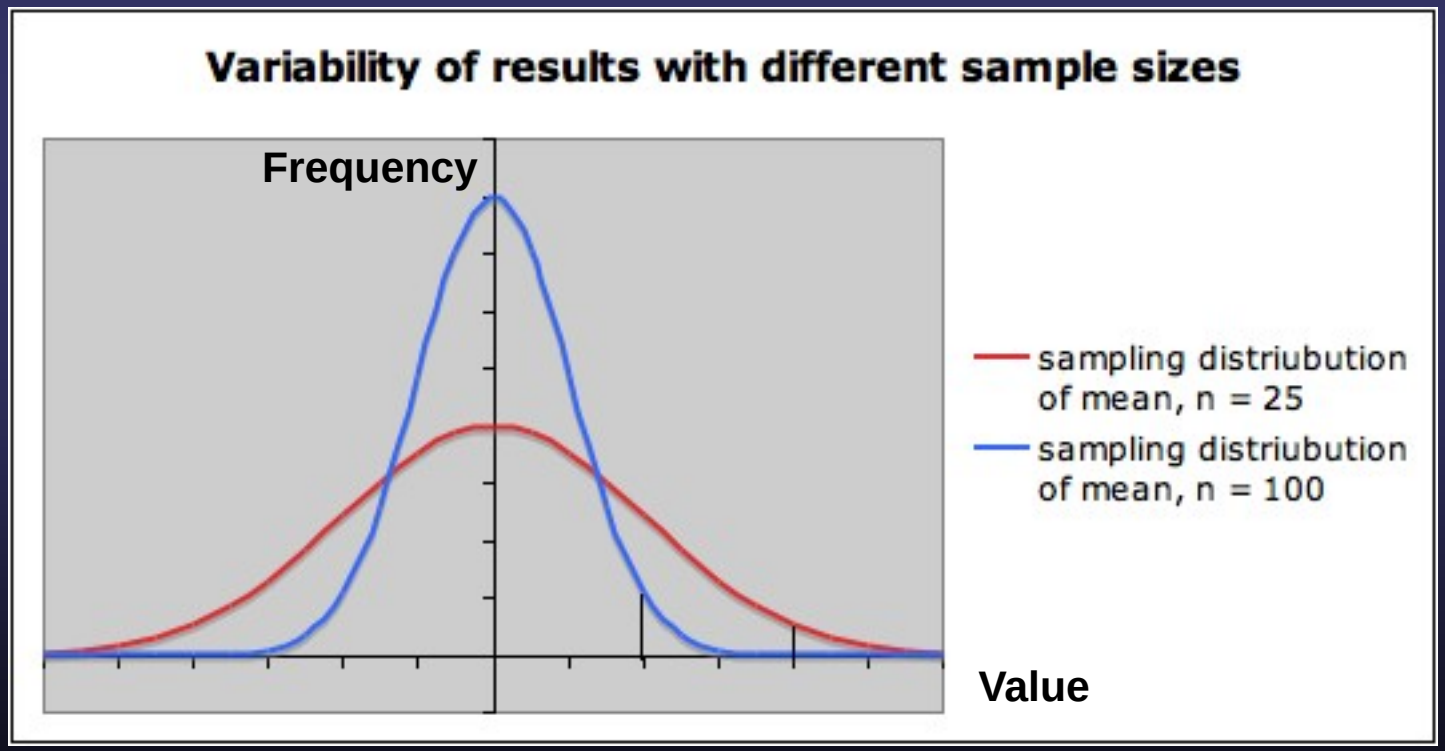
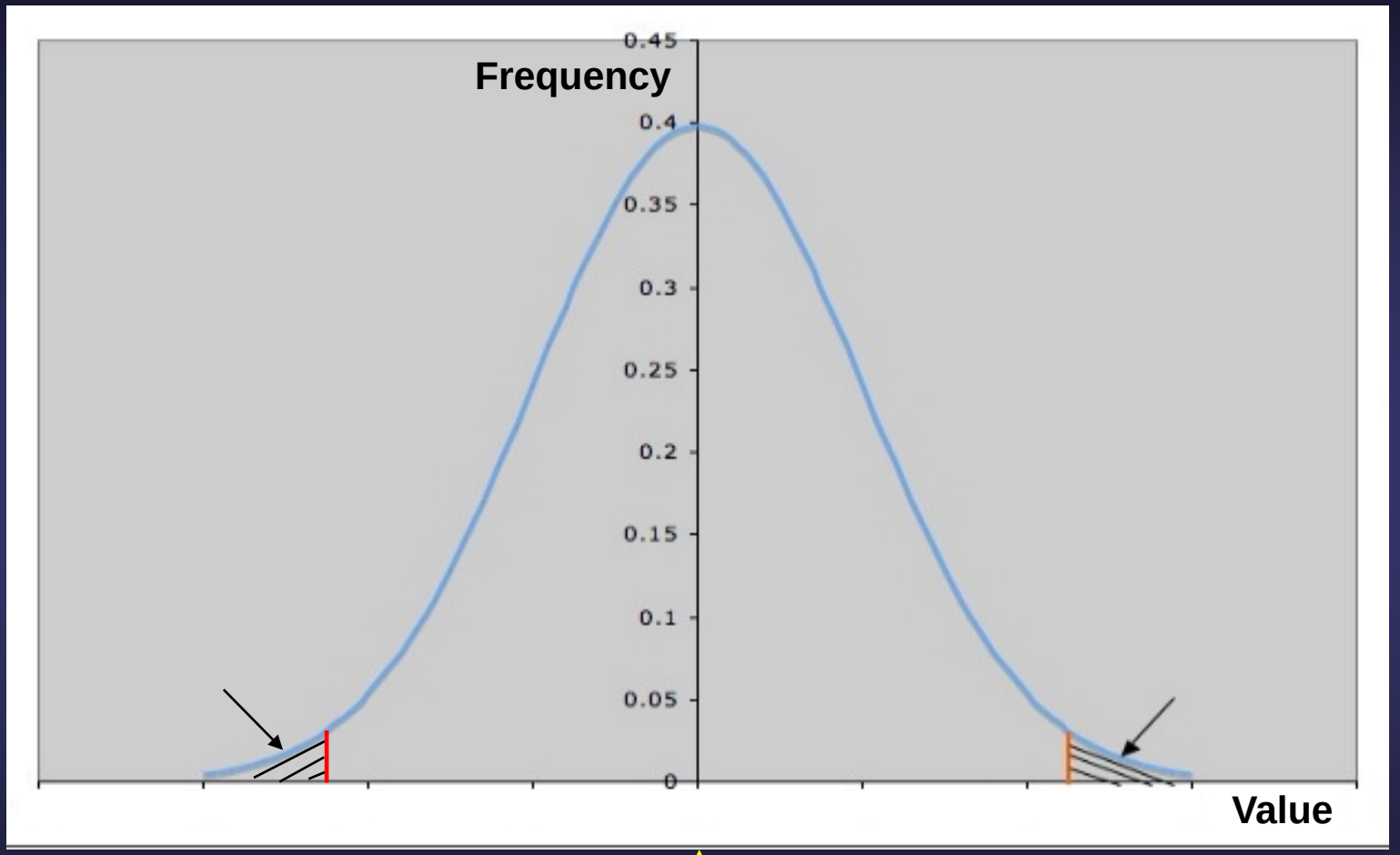


Chart: https://web.ma.utexas.edu

Observed sample mean in a trial



Observed sample mean

Population mean

Sample means

Chart: https://web.ma.utexas.edu

Inference using hypothesis testing

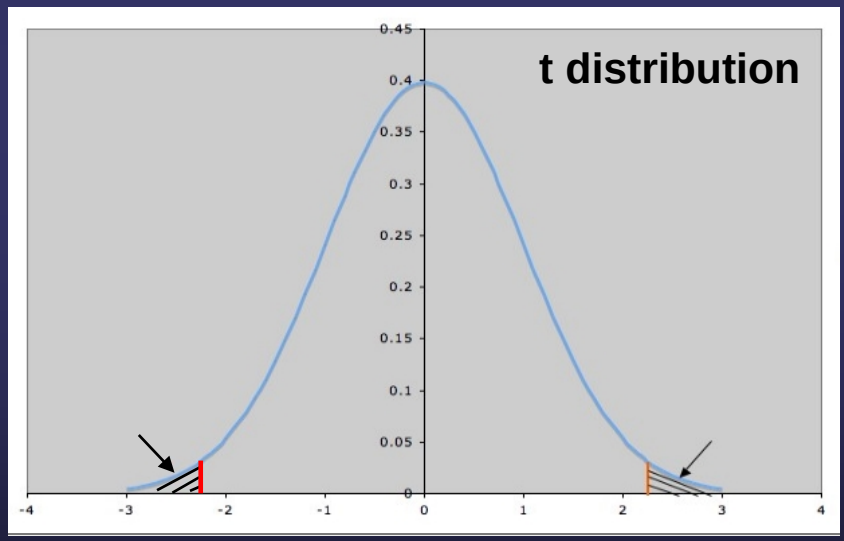
- **Is the population mean different from a hypothesized mean value (e.g, average yield of standard)?**

Null hypothesis H_0: **Mean yield of experimental treatment = Historical mean yield of standard treatment**

t value: a "t value" can be computed using the sample mean and variance, and the hypothesized mean

$$\text{t value} = \frac{\text{sample mean - hyp. mean}}{\sqrt{(\text{variance/n})}}$$

One sample t test: under some assumptions, t value follows a t distribution with n-1 degrees of freedom



t value under H_0

Chart: https://web.ma.utexas.edu

Inference using hypothesis testing

- **Is the population mean different from a hypothesized mean value (e.g, yield of standard)?**

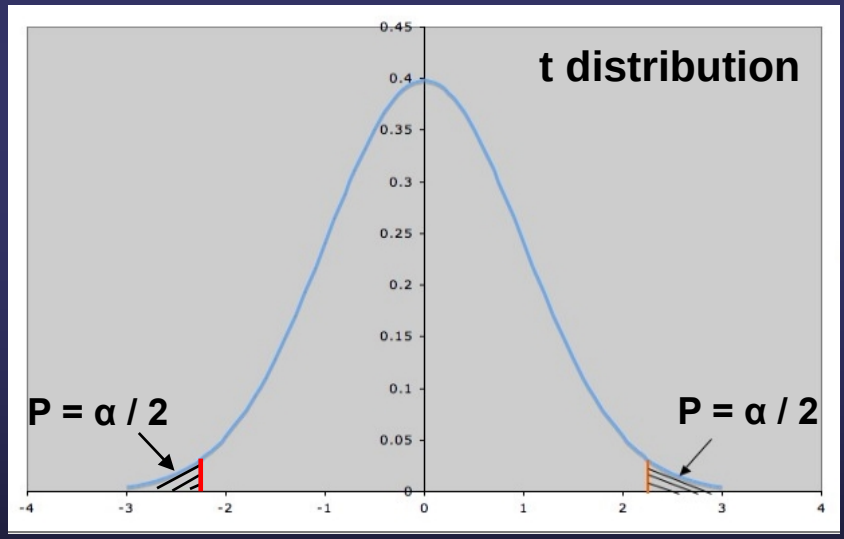
Null hypothesis H_0: **Mean yield of experimental treatment = Historical yield of standard treatment**

Type 1 error: probability of rejecting the null hypothesis while it's true

α: significance level, often 0.05. Probability of type 1 error is ≤ 0.05

P-value: probability of observing the sample mean if H_0 is true

If P-value ≤ α, we reject H_0
Otherwise, we cannot reject.



t value under H_0

Chart: https://web.ma.utexas.edu

Inference using hypothesis testing

- **Is the population mean different from a hypothesized mean value (e.g, yield of standard)?**

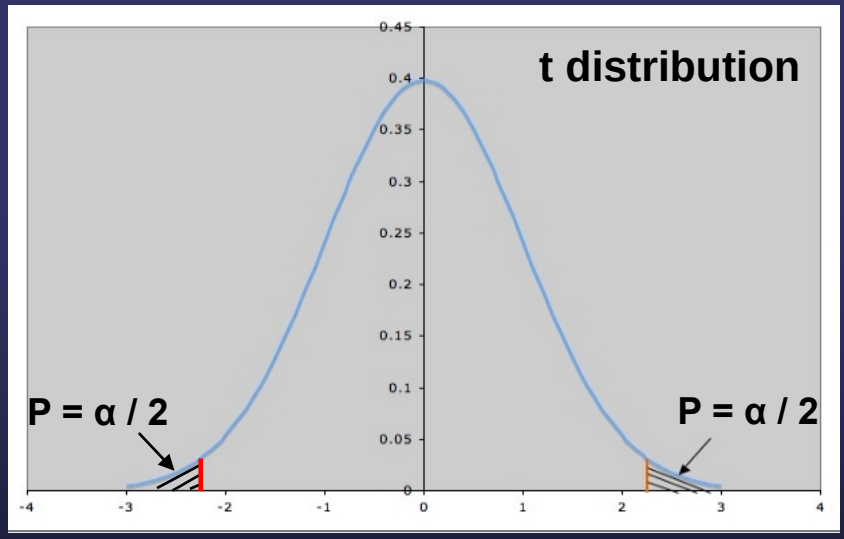
Null hypothesis H_0: **Mean yield of experimental treatment = Historical yield of standard treatment**

Type 1 error: probability of rejecting the null hypothesis while it's true

α: significance level, often 0.05. Probability of type 1 error is ≤ 0.05

P-value: probability of observing the sample mean if H_0 is true

If P-value ≤ α, we reject H_0
Otherwise, we cannot reject.



Observed t value

Chart: https://web.ma.utexas.edu

Inference using hypothesis testing

- **Is the population mean different from a hypothesized mean value (e.g, yield of standard)?**

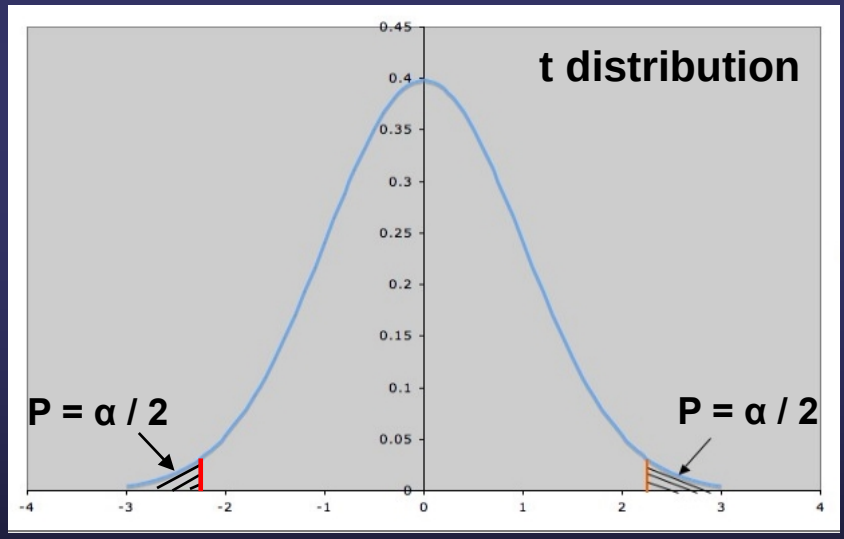
Null hypothesis H_0: **Mean yield of experimental treatment = Historical yield of standard treatment**

Type 1 error: probability of rejecting the null hypothesis while it's true

α: significance level, often 0.05. Probability of type 1 error is ≤ 0.05

P-value: probability of observing the sample mean if H_0 is true

If P-value ≤ α, we reject H_0
Otherwise, we cannot reject.



Observed t value

Chart: https://web.ma.utexas.edu

Inference using hypothesis testing

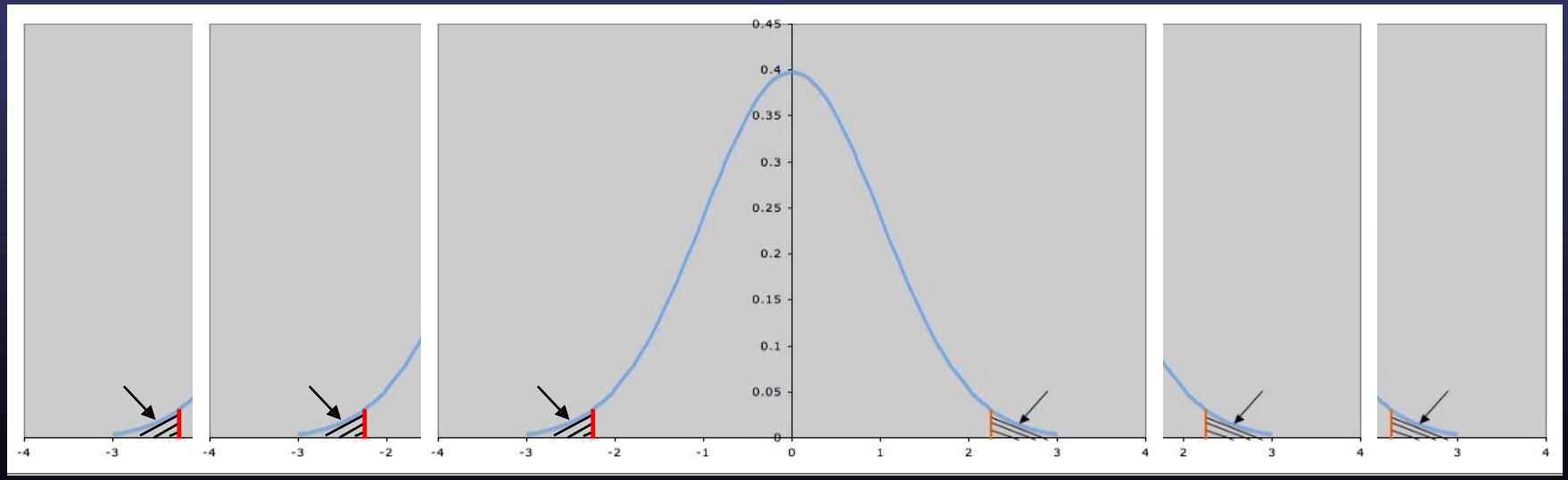
- **Is the population mean different from another population's mean (e.g, yield of standard)?**

Null hypothesis: **Mean yield of experimental treatment = Mean yield of standard**

We test if the difference between the means is equal to 0. We pool the variances between the two samples.

Independent sample t test.

If P-value ≤ α, we reject H_0. Otherwise, we cannot reject.

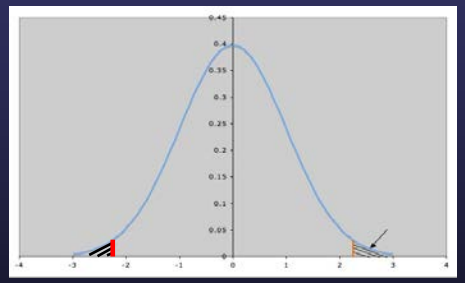
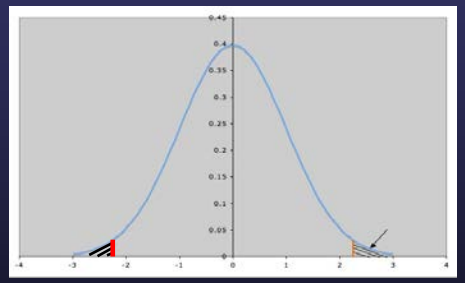
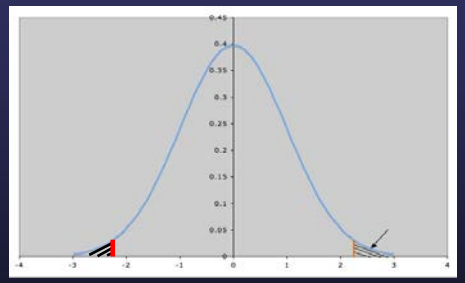
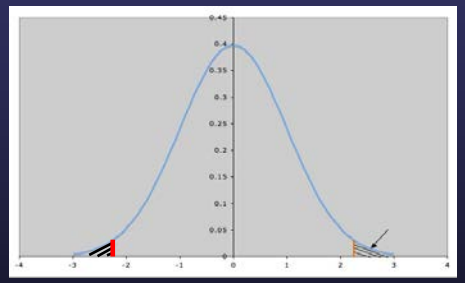


Inference using hypothesis testing

- **Are means from several populations different from each other?**

Null hypothesis: **Mean yields of all treatments evaluated are equal**

We cannot do several pairwise comparisons using independent sample t tests because it substantially increases the overall α

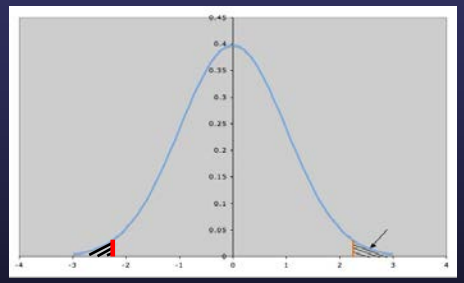
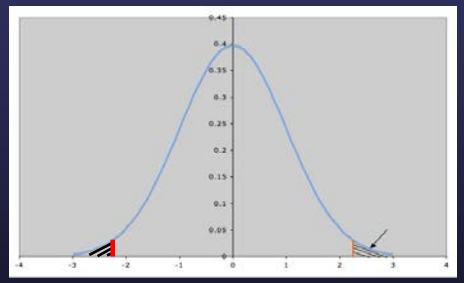
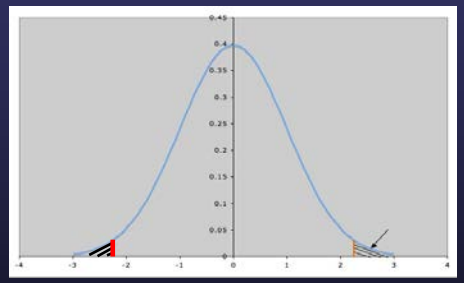
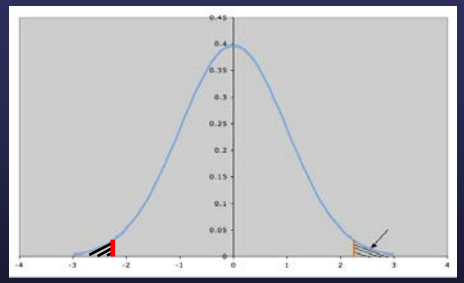





Inference using hypothesis testing

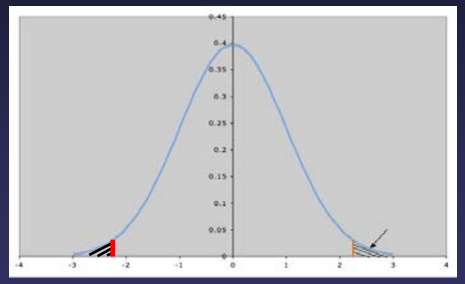
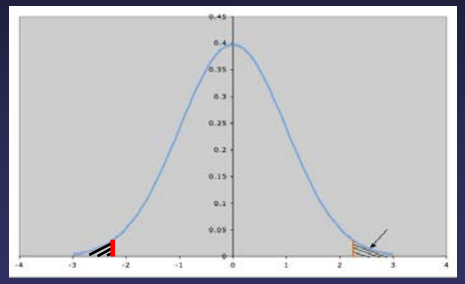
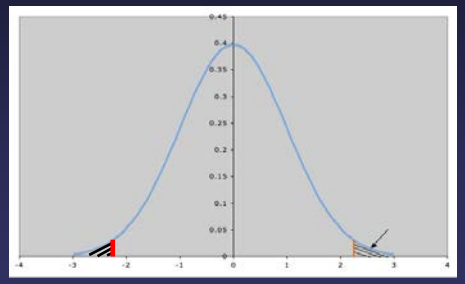
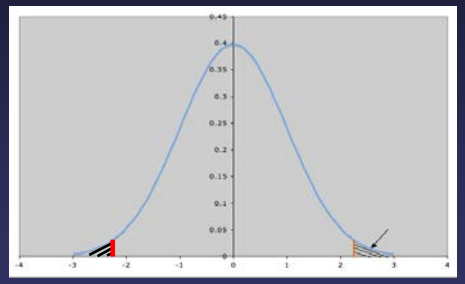
- **Are means from several populations different from each other?**

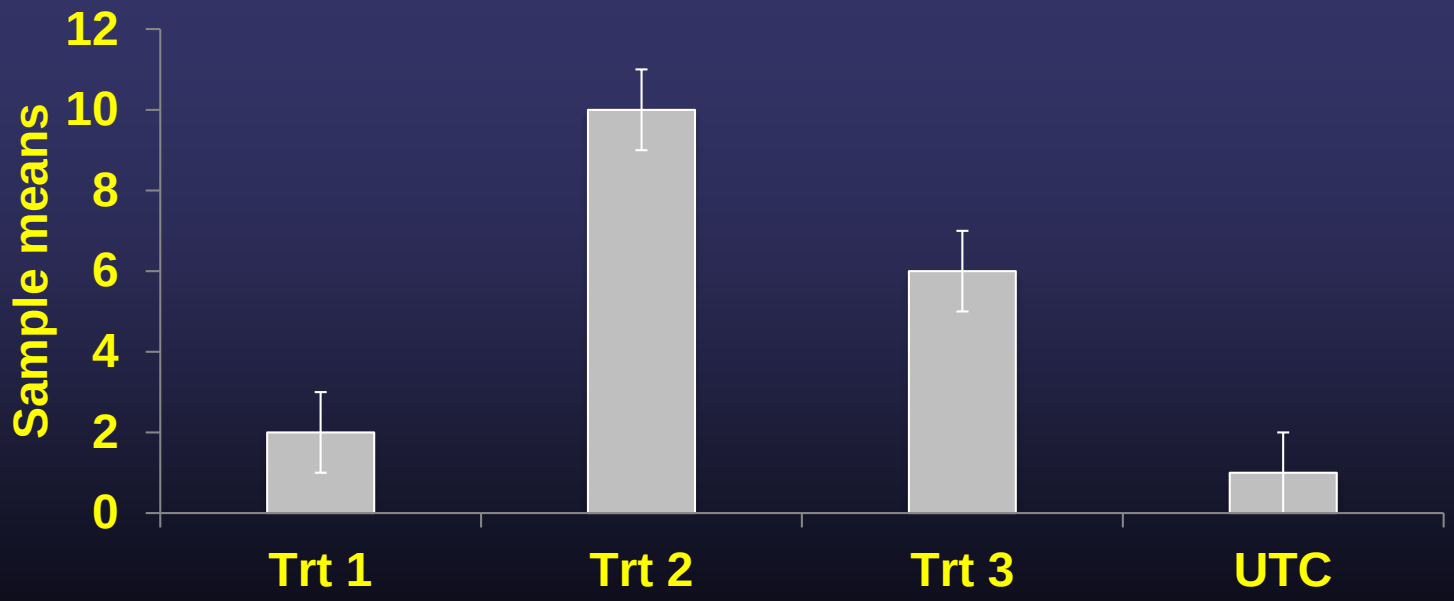
Null hypothesis: **Mean yields of all treatments evaluated are equal**

ANOVA. We compare variances.

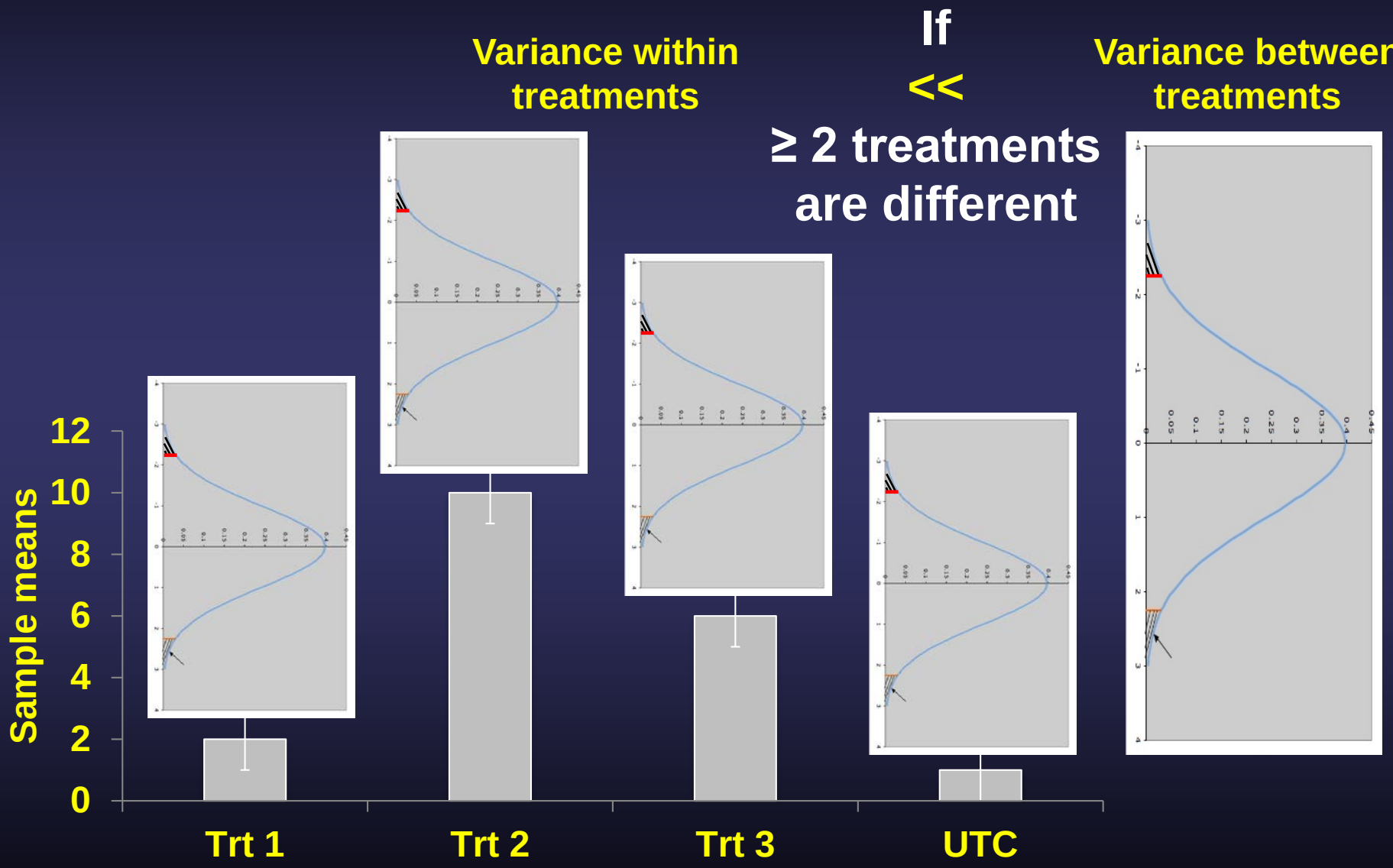





ANOVA

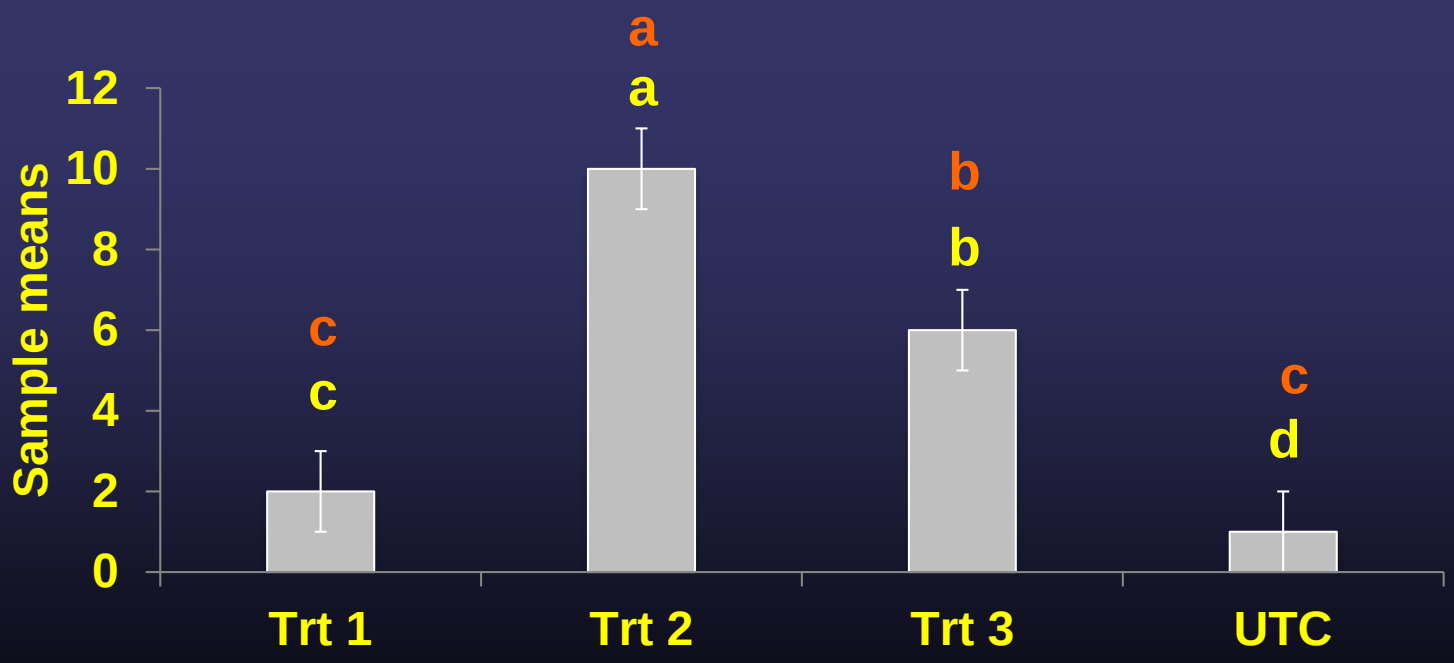


ANOVA


Variance within treatments
If
<<
≥ 2 treatments are different
Variance between treatments
Sample means
12
10
8
6
4
2
0
Trt 1
Trt 2
Trt 3
UTC

ANOVA

- **Variance between treatments / variance within treatments: This is the F value used to determine the P-value**
- **If P-value ≤ α, H_0 is rejected**
- **If H_0 rejected, means are separated using pairwise tests (means with the same letters are not different, P > 0.05)**



Tukey's HSD

LSD

Significant or not significant? That is the question

- **A note on “significance”**
  - **Probability of type 1 error is acceptably low**
  - **α is up to you and your peers**
- **Even though differences are significant, they may not have any biological or agronomic relevance**
- **If differences are not significant, it just means that we cannot detect differences regardless of numerical trends.** **In reality, differences may or may not exist.**
- **Expertise is important**

To compare experimental treatments…

- **We need to estimate means but also variability**
  - **Justification for replications**
- **We need to control variability as much as possible so that variability observed is primarily associated with known factors, including our experimental treatments**
  - **Justification for experimental design and professional experience**

Experimental design

- **Design units**

  - Experimental units: experimental material to which treatments are assigned using randomization (e.g., plots)

  - Sampling units (subsets of the experimental unit) (e.g., rows, plants, fruits within plots)

  - Observation units (unit on which variables are measured)

- **The number of experimental units per treatment is the number of replications**

Number of replications (sample size)

- **How many replications do we need?**
  - Are 4 replications OK? 4 is better than 3. 5 is better than 4
  - There are conventions, practical considerations, and previous experiences
- **Formula based on LSD**

$$LSD = t\sqrt{(2*variance/n)}$$

- If a difference of predetermined value should be detected. **Predetermined difference = LSD**
- If comparable experiments have been conducted in the past. **Variance = variance from past experiments**

$$n = 2*t^2*variance/LSD^2$$

$$t \approx 2.3$$

Experimental design: Blocks

- **CRDs and RBDs**
  - **What are blocks?**
    Groups of experimental units (not independent)
  - **Group can be in space or time**
- **Why using blocks in RBDs?**
  - **Control of variability**
- **Other designs use blocks**
  - **Latin square design (same no. treatments and blocks, 2 block structures)**

Experimental design: Split plot designs

- **It's all about experimental units**
  - Split plots can be CRDs, RBDs, or Latin squares
  - Used when 2 factors are studied (e.g., irrigation and variety, 3 levels each)

- 1 factor randomized to experimental units (main plots)
- 1 factor randomized to subsets of experimental units (sub plots)
- Alternative: conventional factorial design in which combinations of 2 factors are randomized to experimental units (plots)

- Why using split plot designs?
  Practical considerations

Experimental design: Repeated measures

- **Exactly same variable measured on exactly the same piece of experimental material over time or space**

- **Reduced number of experimental material** (no need for different material over time or space)
- **Reduction of variability associated with experimental material** (same used throughout study)

- **Observations are not independent** (correlated within experimental material)

The ANOVA assumptions

- ANOVA assumes normal distribution, homogeneity of variance, independence of samples
- Assumptions are often violated
- What do we do?
  - Keep it simple
  - Have a good experimental design, ANOVA model

The skills and experiences of research staff are extremely important

Trial Designs, Statistics to Obtain Reliable Results

Julien Beuzelin

Phone: 561-993-1559, **E-mail:** jbeuzelin@ufl.edu

UF IFAS UNIVERSITY of FLORIDA