

# Workshop in experimental design and applied statistics

Educational materials developed by Maegen Simmonds

Agricultural Field Statistics Package App developed by Ian King  
Kyle, Jason Moore, and Maegen Simmonds

Funded by USAID

# Group introductions

Who are the instructors?

- Mark Bell, Professor, University of California, Davis, Email: [mark.andrew.bell@gmail.com](mailto:mark.andrew.bell@gmail.com)
- Miguel Macias, University of California, Davis, Email: [mmaciasgonzalez@ucdavis.edu](mailto:mmaciasgonzalez@ucdavis.edu)

# Workshop Goals

- To understand experimental design and principles of statistical analysis to produce trustworthy results and interpretation.
- To learn how to use the free Agricultural Field Statistics Package computer app to analyze and visualize your data

# Workshop Goals

- What do YOU want to learn?
- Write down 3 of your goals and we will create a list of your top priorities for the whole group

# What are we doing today?

- Instruction 9:00 – 11:00 AM
  - Introductions
  - Section 1: Experimental design, ANOVA, hypothesis testing
- Tea break 11:00 – 11:30 AM
- Instruction 11:30 – 1:00 PM
  - Section 2: Agricultural Field Statistics Package app
  - ANOVA and interpretation with 1 treatment and with 2 treatments, factorial
- Lunch 1:00 – 2:00 PM
- Instruction 2:00 – 2:30 PM
- Tea break 2:30 – 3:00 PM
- Instruction 3:00 – 4:00 PM

# What are basic experimental designs do you use in your research?

Examples:

1. Complete Randomized Design
2. Randomized Complete Block Design
3. Split-plot Design

Note: “Factorial” is a treatment structure that can occur within any of the experimental designs. We will also cover this.

# Section 1 Learning Objectives

Have basic understanding of:

1. Elements of an experiment
2. Complete Randomized Design (CRD)
3. Hypothesis testing
4. Statistical model for a CRD with one treatment
5. Principles of Analysis of Variance (ANOVA) for a CRD with one treatment
6. Elements in an ANOVA table for a CRD with one treatment

Research question: Is there an effect of milking practice (hand-milk 2x/day, hand-milk 3x/day, machine-milk 2x/day, machine-milk 3x/day) on milk production performance in Sahiwal cows?

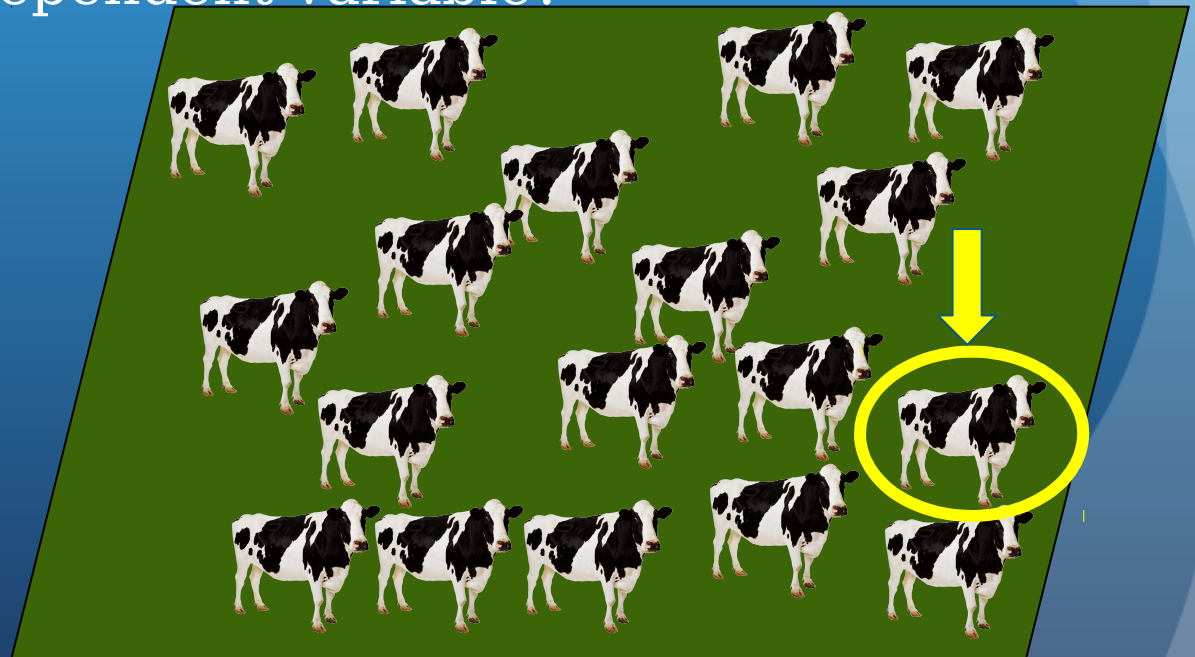
What is the experimental unit?  
What is the dependent variable?  
What is the independent variable?

Cow

Milk production

performance

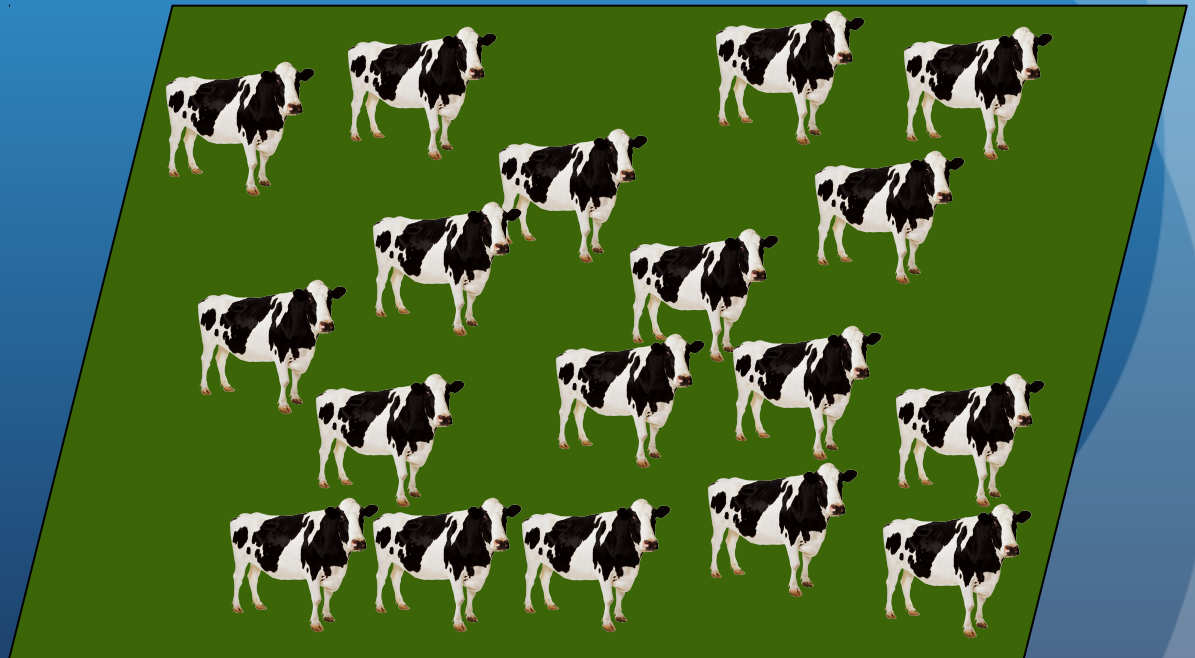
Milking practices





Research question: Is there an effect of milking practice (hand-milk 2x/day, hand-milk 3x/day, machine-milk 2x/day, machine-milk 3x/day) on milk production performance in Sahiwal cows?

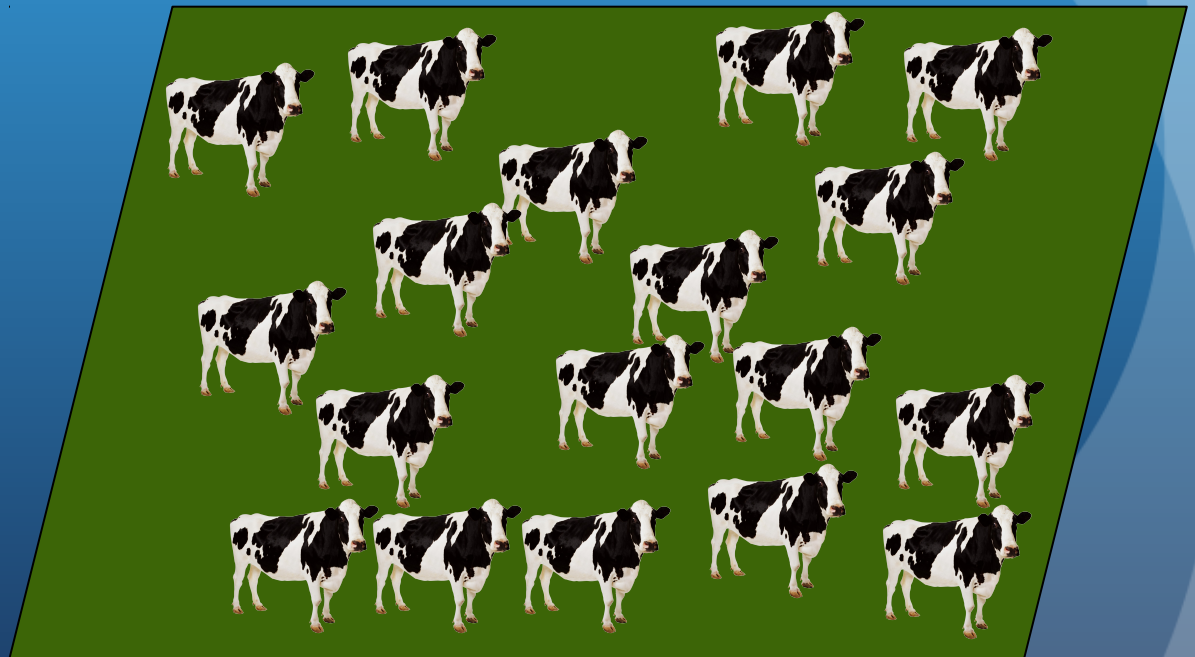
1. What experimental design would you use to determine if there is an effect of milking practice on milk production (kg)?



Research question: Is there an effect of milking practice (hand-milk 2x/day, hand-milk 3x/day, machine-milk 2x/day, machine-milk 3x/day) on milk production performance in Sahiwal cows?

1. What experimental design would you use to determine if there is an effect of milking practice on milk production (kg)?

If the experiment was at a *single* research station, using *cows with similar characteristics*, we may choose to apply the 4 treatments x 3 replications randomly to a subset of the cows.

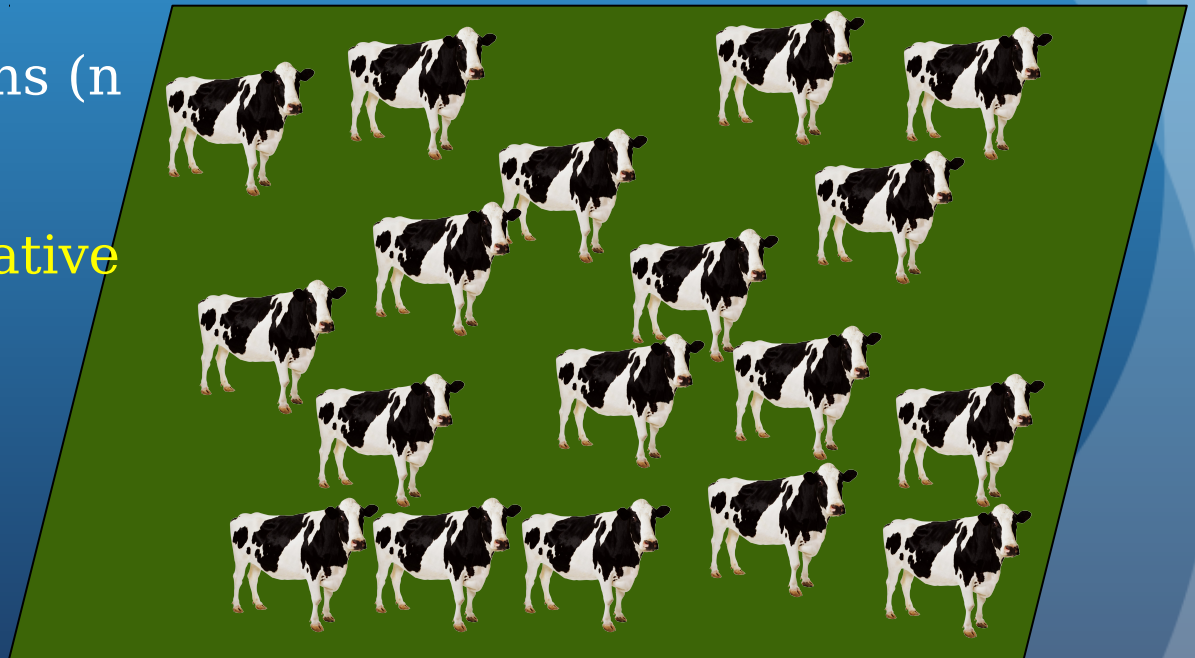


Research question: Is there an effect of milking practice (hand-milk 2x/day, hand-milk 3x/day, machine-milk 2x/day, machine-milk 3x/day) on milk production performance in Sahiwal cows?

This type of experimental design is called a Complete Randomized Design (CRD) (no blocking)

For 4 treatments with 3 replications ( $n = 12$ ),

1. 12 **representative** cows are randomly selected,

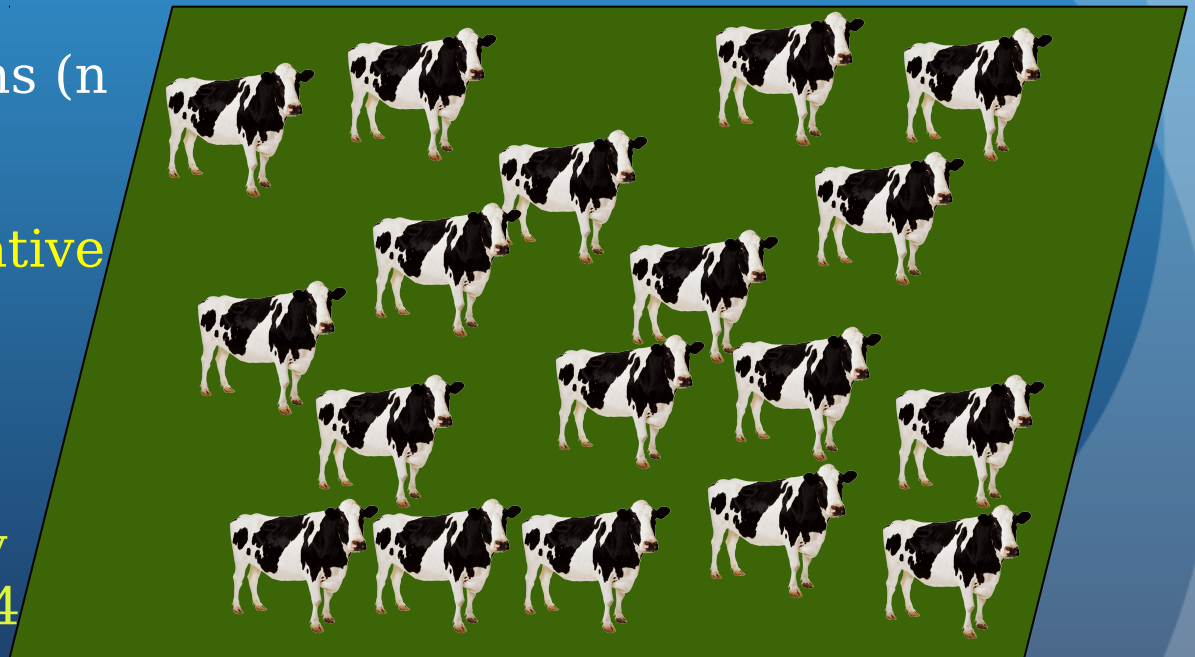


Research question: Is there an effect of milking practice (hand-milk 2x/day, hand-milk 3x/day, machine-milk 2x/day, machine-milk 3x/day) on milk production performance in Sahiwal cows?

This type of experimental design is called a Complete Randomized Design (CRD) (no blocking)

For 4 treatments with 3 replications ( $n = 12$ ),

1. 12 **representative** cows are randomly selected,
2. And **randomly** assigned the 4 treatments.

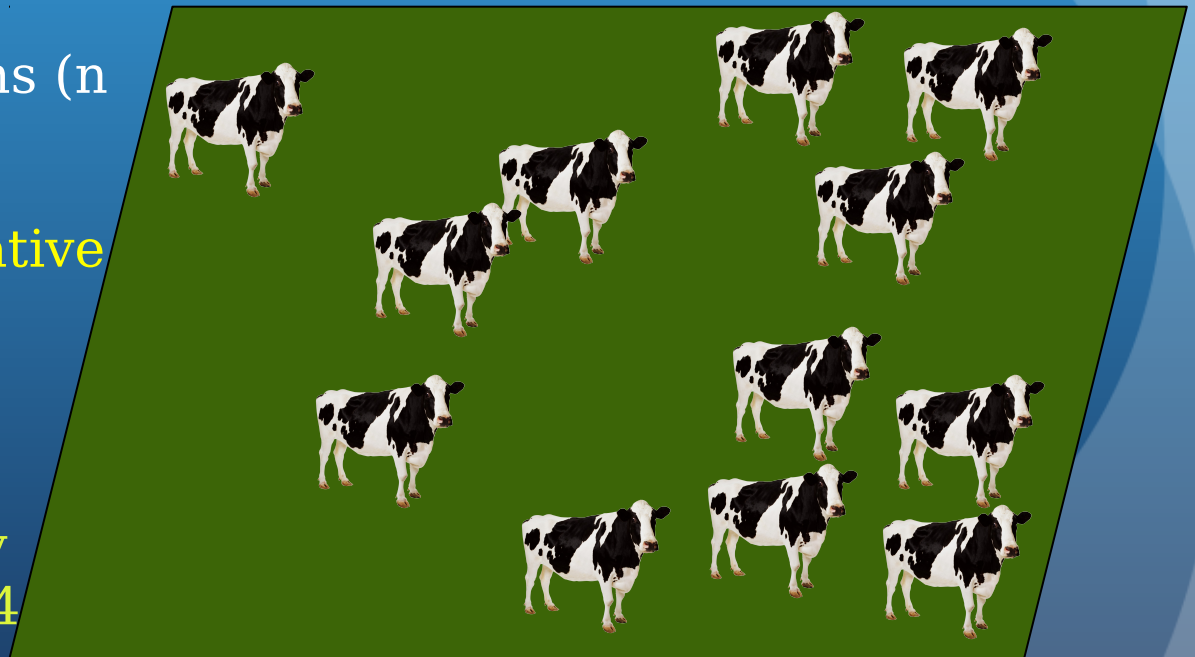


Research question: Is there an effect of milking practice (hand-milk 2x/day, hand-milk 3x/day, machine-milk 2x/day, machine-milk 3x/day) on milk production performance in Sahiwal cows?

This type of experimental design is called a Complete Randomized Design (CRD) (no blocking)

For 4 treatments with 3 replications ( $n = 12$ ),

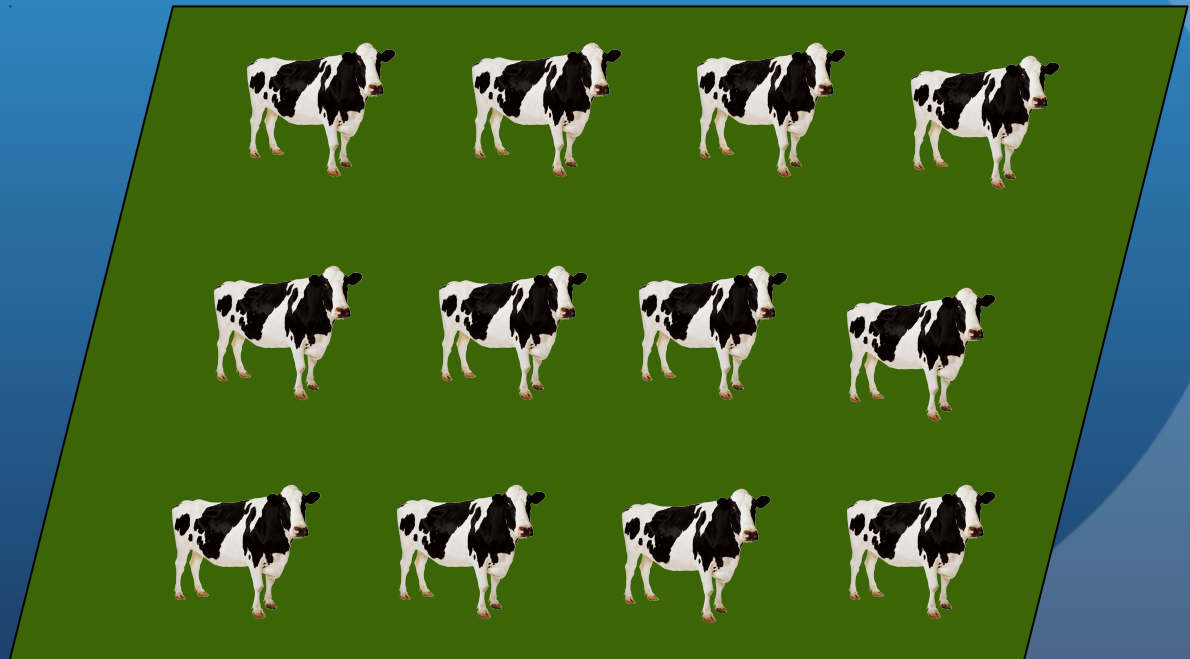
1. 12 **representative** cows are randomly selected,
2. And **randomly** assigned the 4 treatments.



# Replications versus subsamples: important reminders

1. What is your experimental unit?
2. Do you collect a sample from it more than once?  
If so, your data are subsamples. If not, they are replications.

For example, if you measured the **milk production of each cow** (experimental unit) **more than once**, would each observation be a replication or subsample?



# Replications versus subsamples: important reminders













1. What is your experimental unit?
2. Do you collect a sample from it more than once?  
If so, your data are subsamples. If not, they are replications.
3. If your data consists of subsamples, you must *average* the subsamples for each experimental unit *before* loading data into the app. Otherwise, your data analysis will not be correct.



Let's collect the milk  
production data (dependent  
variable)!

# Observed milk production (kg)













12 cows and 12 total observations ( $n = 12$ ). Each observation ( $Y_{ij}$ ) identified by  $i^{\text{th}}$  treatment and  $j^{\text{th}}$  replication.

	(1) Hand-milk 2x	(2) Hand-milk 3x	(3) Machine-milk 2x	(4) Machine-milk 3x
(1)	8.6 	11.2 	9.6 	11.5 
(2)	9.3 	10.8 	9.3 	12.0 
(3)	9.1 	11.7 	9.0 	11.8 
Treatment means:	9.0	11.2	9.3	11.8

# Observed milk production

Are these data  
replications or  
subsamples?













12 cows are randomly assigned to 4 treatments (n = 12). Each observation is a randomly selected cow by  $i^{\text{th}}$  treatment and  $j^{\text{th}}$  replication.

	(1) Hand-milk 2x	(2) Hand-milk 3x	(3) Machine-milk 2x	(4) Machine-milk 3x
(1)	8.6 	11.2 	9.6 	11.5 
(2)	9.3 	10.8 	9.3 	12.0 
(3)	9.1 	11.7 	9.0 	11.8 
Treatment means:	9.0	11.2	9.3	11.8













# Observed milk production

What was the  
milk production  
for  $Y_{3,1}$ ?

12 cows are randomly assigned to 4 treatments (n = 12). Each observation is a milk production by  $i^{\text{th}}$  treatment and  $j^{\text{th}}$  replication.

Replication	(1) Hand-milk 2x	(2) Hand-milk 3x	(3) Machine-milk 2x	(4) Machine-milk 3x
(1)	8.6 	11.2 	9.6 	11.5 
(2)	9.3 	10.8 	9.3 	12.0 
(3)	9.1 	11.7 	9.0 	11.8 
Treatment means:	9.0	11.2	9.3	11.8

# Research question: Is there an effect of milking practice on milk production performance in Sahiwal cows?

Hand-milk 2x	Hand-milk 3x	Machine- milk 2x	Machine- milk 3x
8.6 	11.2 	9.6 	11.5 
9.3 	10.8 	9.3 	12.0 
9.1 	11.7 	9.0 	11.8 
9.0	11.2	9.3	11.8

Overall  
mean: 9.3

Treatment  
means:

In other words, does mean milk production differ among milking practice treatments?

In statistics, we actually test the *null* hypothesis ( $H_0$ ).

Which is what?

All treatment means are equal

$$H_0 : \mu_1 = \mu_2 = \mu_3 = \mu_4$$













We are actually interested in the alternative hypothesis ( $H_A$ ), which is what?

$H_A$  : *at least* two means differ

What linear model predicts milk production for an individual cow in this experiment?

$$Y_{ij} = ?$$

where  $i$  = treatment, and  $j$  = replication













Hand-milk 2x	Hand-milk 3x	Machine- milk 2x	Machine- milk 3x
8.6 	11.2 	9.6 	11.5 
9.3 	10.8 	9.3 	12.0 
9.1 	11.7 	9.0 	11.8 
9.0	11.2	9.3	11.8

Grand mean = 9.3

Treatment means:

$$Y_{ij} = \mu + \tau_i + \varepsilon_{ij}$$

$Y_{ij}$  is the dependent variable (what we are observing/measuring), which is equal to the grand mean ( $\mu$ ), plus the treatment effect ( $\tau_i$ ), plus random error ( $\varepsilon_{ij}$ ).

Hand-milk 2x	Hand-milk 3x	Machine- milk 2x	Machine- milk 3x
8.6 	11.2 	9.6 	11.5 
9.3 	10.8 	9.3 	12.0 
9.1 	11.7 	9.0 	11.8 
9.0	11.2	9.3	11.8













Grand  
mean =  
9.3

Treatment  
means:



$$Y_{ij} = \mu + \tau_i + \varepsilon_{ij}$$

In other words, each observation deviates from the mean of all the samples due to the **effect of the treatment** and due to unexplained **error**.













Hand-milk 2x	Hand-milk 3x	Machine- milk 2x	Machine- milk 3x
8.6 	11.2 	9.6 	11.5 
9.3 	10.8 	9.3 	12.0 
9.1 	11.7 	9.0 	11.8 
9.0	11.2	9.3	11.8

Grand  
mean =  
9.3

Treatmen  
t means:

$$Y_{ij} = \mu + \tau_i + \varepsilon_{ij}$$

What things can cause error in an experiment?

Hand-milk 2x	Hand-milk 3x	Machine- milk 2x	Machine- milk 3x
8.6 	11.2 	9.6 	11.5 
9.3 	10.8 	9.3 	12.0 
9.1 	11.7 	9.0 	11.8 
9.0	11.2	9.3	11.8













Grand  
mean =  
9.3

Treatment  
means:

Analysis of Variance (ANOVA) tests the validity of your statistical model using an **F-test**. The F-test determines if the **treatment variance** is sufficiently larger than the ***within variance***.

If it is, we reject  $H_0$  that all means are equal. If it's not, we fail to reject  $H_0$ .

ANOVA partitions the total variation between each observations and the mean (Total Sums of Squares) into the variation due to each factor in the statistical model.

Hand-milk 2x	Hand-milk 3x	Machine- milk 2x	Machine- milk 3x
8.6 	11.2 	9.6 	11.5 
9.3 	10.8 	9.3 	12.0 
9.1 	11.7 	9.0 	11.8 
9.0	11.2	9.3	11.8

Grand  
mean =  
9.3

Treatment  
means:

Let's fill in the ANOVA table together for the milking practices example, and then use the AIP App to do it for us.

## ANOVA table

Response ( $Y_{ij}$ ) = milk production

Source of variation	Degrees of freedom (df)	Sum of Squares (SS)	Mean Sq (MS)	F value	P value (Pr(>p))

First, we have the total variation among all samples.

## ANOVA table

Response ( $Y_{ij}$ ) = milk production

Source of variation	Degrees of freedom (df)	Sum of Squares (SS)	Mean Sq (MS)	F value	P value (Pr(>p))
Total					

The TOTAL variation is due to variation due the treatment (milking practice) and random error.

## ANOVA table

Response ( $Y_{ij}$ ) = milk production

Source of variation	Degrees of freedom (df)	Sum of Squares (SS)	Mean Sq (MS)	F value	P value (Pr(>p))
Total					

The TOTAL variation is due to variation due the treatment (milking practice) and random error.

## ANOVA table

Response ( $Y_{ij}$ ) = milk production

Source of variation	Degrees of freedom (df)	Sum of Squares (SS)	Mean Sq (MS)	F value	P value (Pr(>p))
Total					
Milking practice					
Error					



Next, let's calculate **Total Sum of Squares (TSS)**.

## ANOVA table

Response ( $Y_{ij}$ ) = milk production

Source of variation	Degrees of freedom (df)	Sum of Squares (SS)	Mean Sq (MS)	F value	P value (Pr(>p))
Total					
Milking practice					
Error					

# Total “Sums of squares” (TSS)


is the difference between each observation and the grand mean, squared, and all added up (i.e. sum of squared deviations).

$$TSS = \sum_{i=1}^t \sum_{j=1}^r (Y_{ij} - \bar{Y}_{..})^2$$

$i = i^{th}$  treatment  
 $j = j^{th}$  replication  
 $t = \text{total \# treatments}$   
 $r = \text{total \# reps}$

Grand mean  
(across all  
treatments and reps)

1<sup>st</sup> treatment, 1<sup>st</sup> rep      1<sup>st</sup> treatment, r<sup>th</sup> rep


$$TSS = (Y_{1,1} - \bar{Y}_{..})^2 + (Y_{1,2} - \bar{Y}_{..})^2 + \dots + (Y_{t,r} - \bar{Y}_{..})^2$$













$$TSS = \sum_{i=1}^t \sum_{j=1}^r (Y_{ij} - \bar{Y}_{..})^2$$

# Total Sums of Squares

1<sup>st</sup> treatment, 1<sup>st</sup>

Grand mean 4<sup>th</sup> treatment, 3<sup>rd</sup>

$$TSS = (8.6 - 9.3)^2 + (9.3 - 9.3)^2 + \dots + (11.8 - 9.3)^2 = 18.1$$

Hand-milk 2x	Hand-milk 3x	Machine- milk 2x	Machine- milk 3x
8.6 	11.2 	9.6 	11.5 
9.3 	10.8 	9.3 	12.0 
9.1 	11.7 	9.0 	11.8 
9.0	11.2	9.3	11.8

Grand  
mean =  
9.3

Treatment  
means:

# ANOVA table

Response ( $Y_{ij}$ ) = milk production

Source of variation	Degrees of freedom (df)	Sum of Squares (SS)	Mean Sq (MS)	F value	P value (Pr(>p))
Total		18.1			
Milking practice					
Error					

Degrees of freedom (df) = total number of independent observations

An easy way to remember:

Df is the total # of elements used to estimate the parameter(s) of interest, minus total # of parameters.

$$df = e - p$$

## ANOVA table













Response ( $Y_{ij}$ ) = milk production

Source of variation	Degrees of freedom (df)	Sum of Squares (SS)	Mean Sq (MS)	F value	P value (Pr(>p))
Total		18.1			
Milking practice					
Error					

“Within” variation of 12 elements relative to 1 estimated parameter, the grand mean

# Total Sums of Squares:

$$TSS = \sum_{i=1}^t \sum_{j=1}^r (Y_{ij} - \bar{Y}_{..})^2$$

Hand-milk 2x	Hand-milk 3x	Machine- milk 2x	Machine- milk 3x
8.6 	11.2 	9.6 	11.5 
9.3 	10.8 	9.3 	12.0 
9.1 	11.7 	9.0 	11.8 
9.0	11.2	9.3	11.8

Grand  
mean =  
9.3

Treatment  
means:

Thus, total df = 12 - 1 = 11

Grand mean was  
estimated based on 12  
individuals

$$TSS = \sum_{i=1}^t \sum_{j=1}^r (Y_{ij} - \bar{Y}_{..})^2$$

$t = t^{th}$  treatment  
 $j = j^{th}$  replication  
 $t = \text{total \# treatments}$   
 $r = \text{total \# reps}$

## ANOVA table

Response ( $Y_{ij}$ ) = milk production

Source of variation	Degrees of freedom (df)	Sum of Squares (SS)	Mean Sq (MS)	F value	P value (Pr(>p))
Total	11	18.1			
Milking practice					
Error					

Recall, TOTAL variation is due to variation due the treatment (milking practice) and random error.

Thus, Total SS = Treatment SS + Residual SS

## ANOVA table

Response ( $Y_{ij}$ ) = milk production

Source of variation	Degrees of freedom (df)	Sum of Squares (SS)	Mean Sq (MS)	F value	P value (Pr(>p))
Total	11	18.1			
Milking practice					
Error					



But how do we calculate Treatment SS and Residual SS?

Total SS = SS Treatment + SS Residual

## ANOVA table

Response ( $Y_{ij}$ ) = milk production

Source of variation	Degrees of freedom (df)	Sum of Squares (SS)	Mean Sq (MS)	F value	P value (Pr(>p))
Total	11	18.1			
Milking practice					
Error					

Let's start calculating SS of Treatment.

Total SS = SS Treatment + SS Residual

## ANOVA table

Response ( $Y_{ij}$ ) = milk production

Source of variation	Degrees of freedom (df)	Sum of Squares (SS)	Mean Sq (MS)	F value	P value (Pr(>p))
Total	11	18.1			
Milking practice					
Error					

Sums of squares of treatment (SST) is the difference between each treatment mean and the grand mean, squared, all added up, and multiplied by the total number of reps.

$$SST = r \sum_{i=1}^t (\bar{Y}_{i.} - \bar{Y}_{..})^2$$

$i = i^{\text{th}}$  treatment  
 $t = \text{total \# treatments}$   
 $r = \text{total \# reps}$

$\#$  reps      1<sup>st</sup> treatment mean      Grand mean      t<sup>th</sup> treatment mean

The diagram illustrates the components of the SST formula. A vertical yellow arrow points from '# reps' to the coefficient 'r'. A diagonal yellow arrow points from '1<sup>st</sup> treatment mean' to the first treatment mean term  $\bar{Y}_{1.}$ . A vertical yellow arrow points from 'Grand mean' to the grand mean term  $\bar{Y}_{..}$ . Another vertical yellow arrow points from 't<sup>th</sup> treatment mean' to the final treatment mean term  $\bar{Y}_t$ .

$$SST = r(\bar{Y}_{1.} - \bar{Y}_{..})^2 + r(\bar{Y}_{2.} - \bar{Y}_{..})^2 + \dots + r(\bar{Y}_t - \bar{Y}_{..})^2$$













$$SST = r \sum_{i=1}^t (\bar{Y}_{i.} - \bar{Y}_{..})^2$$

Sums of Squares of Treatment (SST):  
deviations due to milking practice

# reps    1<sup>st</sup> treatment mean

Grand mean

$$SST = 3(9 - 9.3)^2 + 3(11.2 - 9.3)^2 + 3(9.3 - 9.3)^2 + 3(11.8 - 9.3)^2 = 17.1$$

Hand-milk 2x	Hand-milk 3x	Machine- milk 2x	Machine- milk 3x
8.6 	11.2 	9.6 	11.5 
9.3 	10.8 	9.3 	12.0 
9.1 	11.7 	9.0 	11.8 
9.0	11.2	9.3	11.8

Grand  
mean =  
9.3

Treatment  
means:

# ANOVA table

Response ( $Y_{ij}$ ) = milk production

Source of variation	Degrees of freedom (df)	Sum of Squares (SS)	Mean Sq (MS)	F value	P value (Pr(>p))
Total	11	18.1			
Milking practice		17.1			
Error					

What are df for milking practice?

## ANOVA table

Response ( $Y_{ij}$ ) = milk production













Source of variation	Degrees of freedom (df)	Sum of Squares (SS)	Mean Sq (MS)	F value	P value (Pr(>p))
Total	11	18.1			
Milking practice		17.1			
Error					

Treatment variance of 4 elements relative to 1 estimated parameter, the grand mean.

Sums of Squares of Treatment (SST)

$$SST = r \sum_{i=1}^t \left( \bar{Y}_{i.} - \bar{Y}_{..} \right)^2$$

$t$  = # treatments  
 $t$  = total # treatments  
 $r$  = total # reps

Hand-milk 2x	Hand-milk 3x	Machine- milk 2x	Machine- milk 3x
8.6 	11.2 	9.6 	11.5 
9.3 	10.8 	9.3 	12.0 
9.1 	11.7 	9.0 	11.8 
9.0	11.2	9.3	11.8

Grand  
mean =  
9.3

Treatment means:

Thus, total df = 4 - 1 = 3

Grand mean was  
estimated based on 4  
elements

$$SST = r \sum_{i=1}^t (\bar{Y}_{i.} - \bar{Y}_{..})^2$$

$i = i^{\text{th}}$  treatment  
 $t = \text{total \# treatments}$   
 $r = \text{total \# reps}$

## ANOVA table

Source of variation	Degrees of freedom (df)	Sum of Squares (SS)	Mean Sq (MS)	F value	P value (Pr(>p))
Total	11	18.1			
Milking practice	3	17.1			
Error					



Once we know df and SS of TOTAL & TREATMENT, we can solve for df error and SS error.

Total SS = SS Treatment + SS Error  $\therefore$  SS Error = Total SS - SS Treatment

Total df = df Treatment + df Error  $\therefore$  df Error = Total df - df Treatment

## ANOVA table

Response ( $Y_{ij}$ ) = milk production

Source of variation	Degrees of freedom (df)	Sum of Squares (SS)	Mean Sq (MS)	F value	P value (Pr(>p))
Total	11	18.1			
Milking practice	3	17.1			
Error	8	1			

# ANOVA table

Response ( $Y_{ij}$ ) = milk production

What is causing the majority of the total variation in milk production?

Source of variation	Degrees of freedom (df)	Sum of Squares (SS)	Mean Sq (MS)	F value	P value (Pr(>p))
Total	11	18.1			
Milking practice	3	17.1			
Error	8	1			

The *majority* of total variation ( $TSS = 18.1$ ) is due to milking practice ( $SST = 17.1$ )! But we don't know yet if the treatment effect is statistically significant.

## ANOVA table

Response ( $Y_{ij}$ ) = milk production

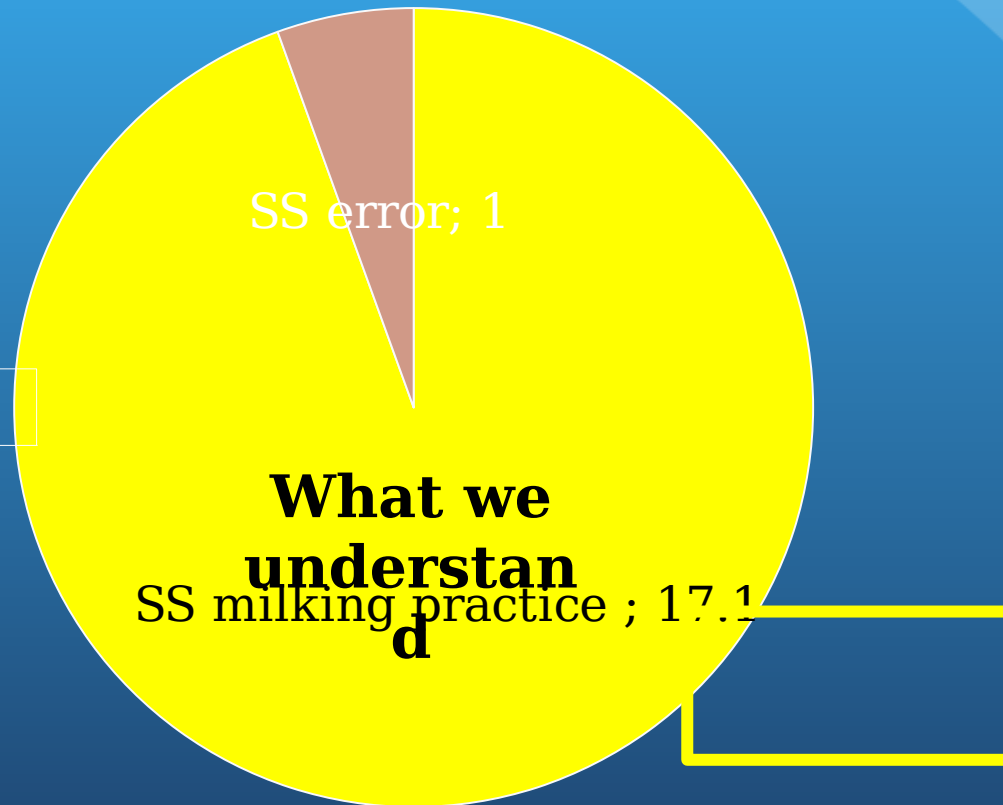
Source of variation	Degrees of freedom (df)	Sum of Squares (SS)	Mean Sq (MS)	F value	P value (Pr(>p))
Total	11	18.1			
Milking practice	3	17.1			
Error	8	1			

$$\text{SS Total} = \text{SS Treatment} + \text{SS Error}$$

$$\text{SS Total} = 18.1$$

**Total  
variation!**

$$SS \text{ Total} = SS \text{ Treatment} + SS \text{ Error}$$



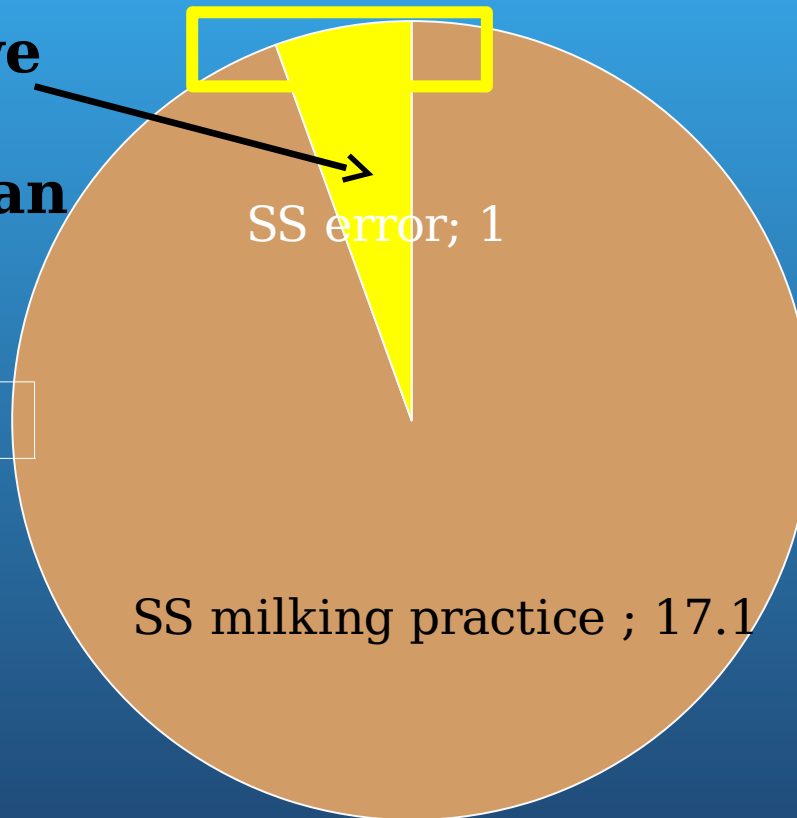
SS Total = 18.1

**What we  
understand**  
SS milking practice ; 17.1

$$\text{SS Total} = \text{SS Treatment} + \text{SS Error}$$

**What we  
don't  
understand**

SS Total = 18.1



Next, calculate the **mean squared deviations (MS)** (Also known as sample variance) for milking practice and error.

## ANOVA table

Response ( $Y_{ij}$ ) = milk production

Source of variation	Degrees of freedom (df)	Sum of Squares (SS)	Mean Sq (MS)	F value	P value (Pr(>p))
Total	11	18.1			
Milking practice	3	17.1			
Error	8	1			

Mean Squared Deviation (sample variance):

$$s^2 = \frac{SS}{df}$$



Mean Squared Treatment (MST) represents the *treatment* variance,

$$\text{Mean Squared Treatment} = MST = \frac{SST}{df_t}$$

while Mean Squared Error (MSE) is “*within*” sample variance

$$\text{Mean Squared Error} = MSE = \frac{SSE}{df_e}$$

Calculate the MS of treatment (MST):

$$MST = \frac{SST}{df_t} = \frac{17.1}{3} = 5.7$$

## ANOVA table

Response ( $Y_{ij}$ ) = milk production

Source of variation	Degrees of freedom (df)	Sum of Squares (SS)	Mean Sq (MS)	F value	P value (Pr(>p))
Total	11	18.1			
Milking practice	3	17.1	5.7		
Error	8	1			

Calculate the MS of Error (MSE):

$$MSE = \frac{SSE}{df_e} = \frac{1}{8} = 0.125$$

## ANOVA table

Response ( $Y_{ij}$ ) = milk production

Source of variation	Degrees of freedom (df)	Sum of Squares (SS)	Mean Sq (MS)	F value	P value (Pr(>p))
Total	11	18.1			
Milking practice	3	17.1	5.7		
Error	8	1	0.125		

“F test”:  $H_0$  = All treatment means are equal  
Compares the probabilities associated with (1) F-calculated and (2) F-critical

## ANOVA table

Response ( $Y_{ij}$ ) = milk production

Source of variation	Degrees of freedom (df)	Sum of Squares (SS)	Mean Sq (MS)	F value	P value (Pr(>p))
Total	11	18.1			
Milking practice	3	17.1	5.7		
Error	8	1	0.125		

“F test”:  $H_0 =$  All treatment means are equal

- **F-calc**: ratio of treatment sample variance (MST) to within sample variance (MSE)
- **F-critical**: the F-value associated your *chosen* acceptable level of probability of making a type I error (significance level). \*We will use  $p=0.05$

## ANOVA table

Response ( $Y_{ij}$ ) = milk production

Source of variation	Degrees of freedom (df)	Sum of Squares (SS)	Mean Sq (MS)	F value	P value (Pr(>p))
Total	11	18.1	Treatment variance		
Milking practice	3	17.1	5.7		F-calc
Error	8	1	0.125	“within” variance	

“F test”:  $H_0 =$  All treatment means are equal

1. Calculate F statistic:  $F_{\text{calc}} = \frac{MST}{MSE}$

which is really  $\frac{\text{Treatment variance}}{\text{Within variance}}$

## ANOVA table

Response ( $Y_{ij}$ ) = milk production

Source of variation	Degrees of freedom (df)	Sum of Squares (SS)	Mean Sq (MS)	F value	P value (Pr(>p))
Total	11	18.1			
Milking practice	3	17.1	5.7		
Error	8	1	0.125		

**Treatment variance** (points to MS for Milking practice)

**“within” variance** (points to MS for Error)

**F-calc** (points to the empty F value cell for Milking practice)

“F test”:  $H_0 =$  All treatment means are equal

1. Calculate F statistic:  $F\text{-calc} = \frac{MST}{MSE} = \frac{5.7}{0.125} = 45.6$

## ANOVA table

Response ( $Y_{ij}$ ) = milk production

Source of variation	Degrees of freedom (df)	Sum of Squares (SS)	Mean Sq (MS)	F value	P value (Pr(>p))
Total	11	18.1	Treatment variance ↓		
Milking practice	3	17.1		45.6	← F-calc
Residual	8	1	0.125 ↑ “within” variance		

“F test”:  $H_0 =$  All treatment means are equal

1. Calculate F statistic:  $F_{\text{-calc}} = \frac{MST}{MSE} = \frac{5.7}{0.125} = 45.6$
2. Determine P value

## ANOVA table

Response ( $Y_{ij}$ ) = milk production

Source of variation	Degrees of freedom (df)	Sum of Squares (SS)	Mean Sq (MS)	F value	P value (Pr(>p))
Total	11	18.1	Treatment variance		
Milking practice	3	17.1	5.7	45.6	F-calc
Residual	8	1	0.125	“within” variance	



“F test”:  $H_0$  = All treatment means are

equal  
● **p-value**: probability of getting as extreme an F statistic by chance alone

⚙️ Smaller the p-value, the more confident we are that the treatment effect is real (truth) and not a mistake (type I error)

## ANOVA table

Response ( $Y_{ij}$ ) = milk production

Source of variation	Degrees of freedom (df)	Sum of Squares (SS)	Mean Sq (MS)	F value	P value (Pr(>p))
Total	11	18.1			
Milking practice	3	17.1	5.7	45.6	
Error	8	1	0.125		

**p-value** ↓

“F test”:  $H_0$  = All treatment means are equal

Probability Density

‘F Distribution’ for 3 df in numerator and 8 degrees of freedom in denominator

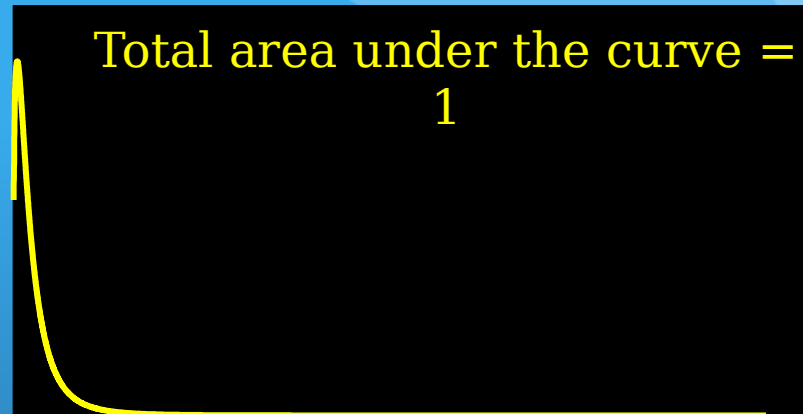
## ANOVA table

Response ( $Y_{ij}$ ) = milk production

Source of variation	Degrees of freedom (df)	Sum of Squares (SS)	Mean Sq (MS)	F value	P value (Pr(>p))
Total	11	18.1			
Milking practice	3	17.1	5.7	45.6	
Error	8	1	0.125		

“F test”:  $H_0$  = All treatment means are equal

Probability  
Density



## ANOVA table

Response ( $Y_{ij}$ ) = milk production

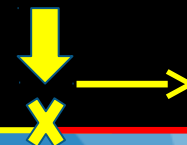
Source of variation	Degrees of freedom (df)	Sum of Squares (SS)	Mean Sq (MS)	F value	P value (Pr(>p))
Total	11	18.1			
Milking practice	3	17.1	5.7	45.6	p-value ↓
Error	8	1	0.125		

Probability  
Density

“F test”:  $H_0 =$  All treatment means are equal

P-value = area under the curve to the right of F-value (i.e. probability of false positive)

F-calc = 45.6



## ANOVA table

Response ( $Y_{ij}$ ) = milk production

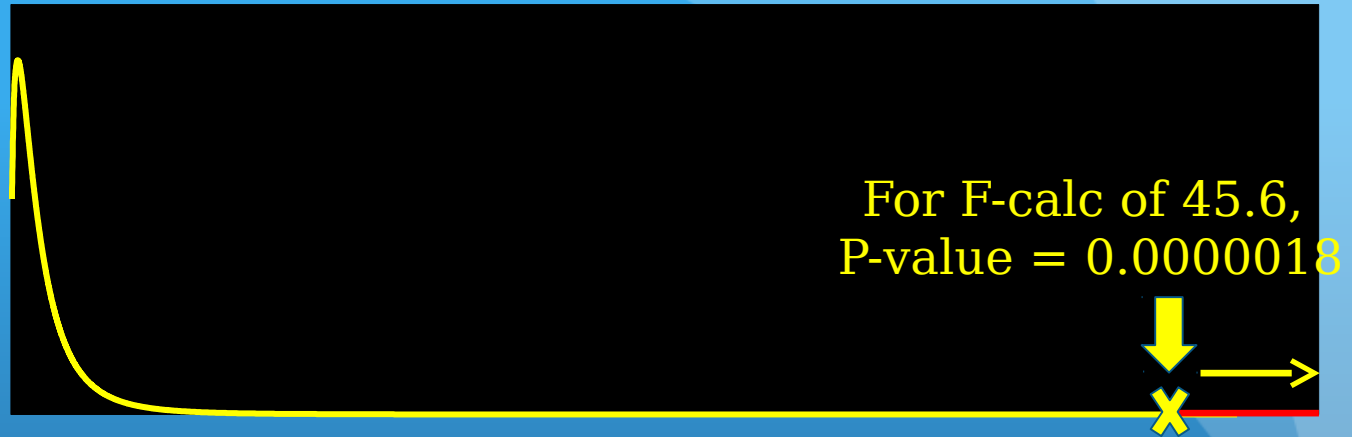
Source of variation	Degrees of freedom (df)	Sum of Squares (SS)	Mean Sq (MS)	F value	P value (Pr(>p))
Total	11	18.1			
Milking practice	3	17.1	5.7	45.6	
Error	8	1	0.125		

p-value



“F test”:  $H_0$  = All treatment means are equal

Probability  
Density



## ANOVA table

Response ( $Y_{ij}$ ) = milk production

Source of variation	Degrees of freedom (df)	Sum of Squares (SS)	Mean Sq (MS)	F value	P value (Pr(>p))
Total	11	18.1			
Milking practice	3	17.1	5.7	45.6	p-value ↓ <0.001
Error	8	1	0.125		

“F test”:  $H_0$  = All treatment means are equal

Probability  
Density

We typically choose significance level  
of 0.05, or 5% probability of type I  
error

For F-calc of 45.6,  
P-value = 0.0000018



## ANOVA table

Response ( $Y_{ij}$ ) = milk production

Source of variation	Degrees of freedom (df)	Sum of Squares (SS)	Mean Sq (MS)	F value	P value (Pr(>p))
Total	11	18.1			
Milking practice	3	17.1	5.7	45.6	<b>p-value</b> ↓ <b>&lt;0.001</b>
Error	8	1	0.125		

“F test”:  $H_0$  = All treatment means are equal

Probability  
Density

Compare the p-value for F-calc to the  
significance level of  $p=0.05$

For F-calc of 45.6,  
P-value = 0.0000018



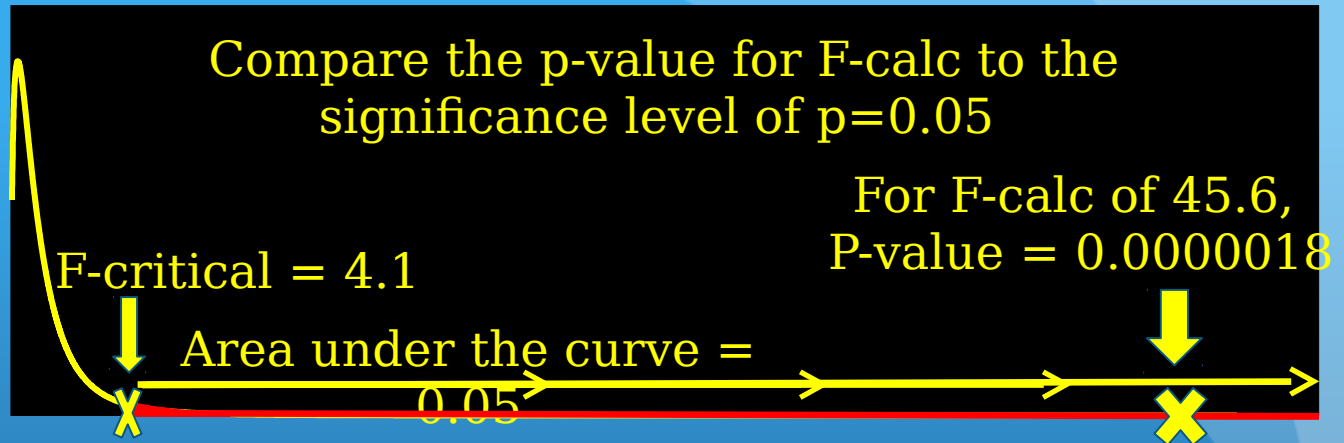
## ANOVA table

Response ( $Y_{ij}$ ) = milk production

Source of variation	Degrees of freedom (df)	Sum of Squares (SS)	Mean Sq (MS)	F value	P value (Pr(>p))
Total	11	18.1			
Milking practice	3	17.1	5.7	45.6	<b>p-value</b> <b>&lt;0.001</b>
Error	8	1	0.125		

“F test”:  $H_0 =$  All treatment means are equal

Probability  
Density



## ANOVA table

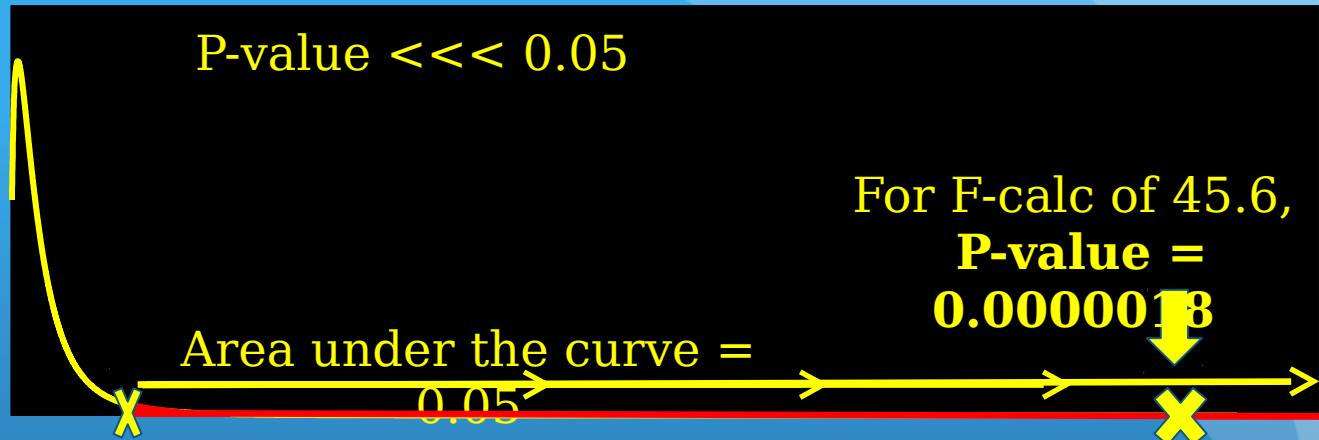
Response ( $Y_{ij}$ ) = milk production

Source of variation	Degrees of freedom (df)	Sum of Squares (SS)	Mean Sq (MS)	F value	P value (Pr(>p))
Total	11	18.1			
Milking practice	3	17.1	5.7	45.6	p-value ↓ <0.001
Error	8	1	0.125		



“F test”:  $H_0$  = All treatment means are equal

Probability  
Density



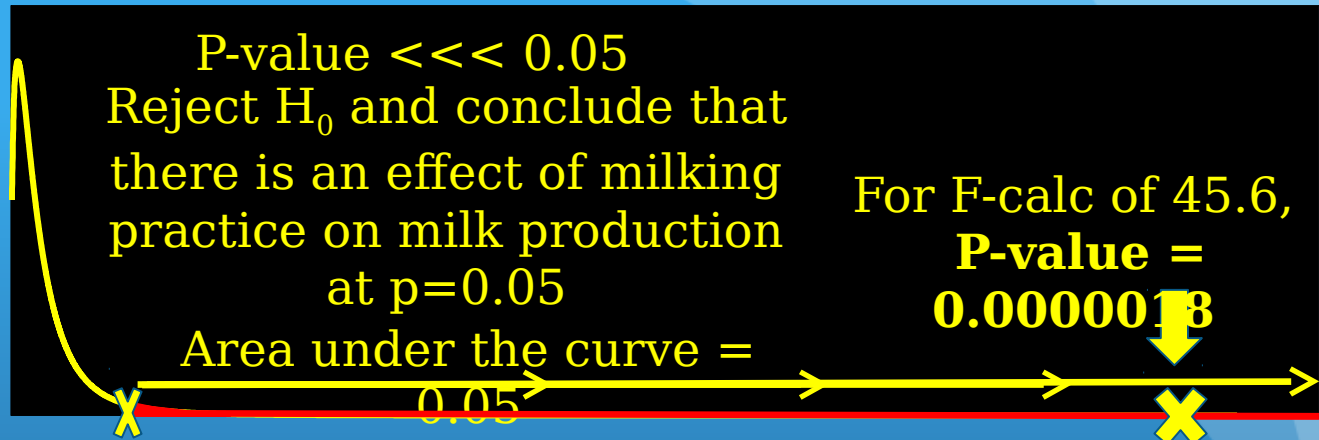
## ANOVA table

Response ( $Y_{ij}$ ) = milk production

Source of variation	Degrees of freedom (df)	Sum of Squares (SS)	Mean Sq (MS)	F value	P value (Pr(>p))
Total	11	18.1			p-value
Milking practice	3	17.1	5.7	45.6	<0.001
Error	8	1	0.125		

"F test":  $H_0 =$  ~~All treatment means are equal~~

Probability  
Density



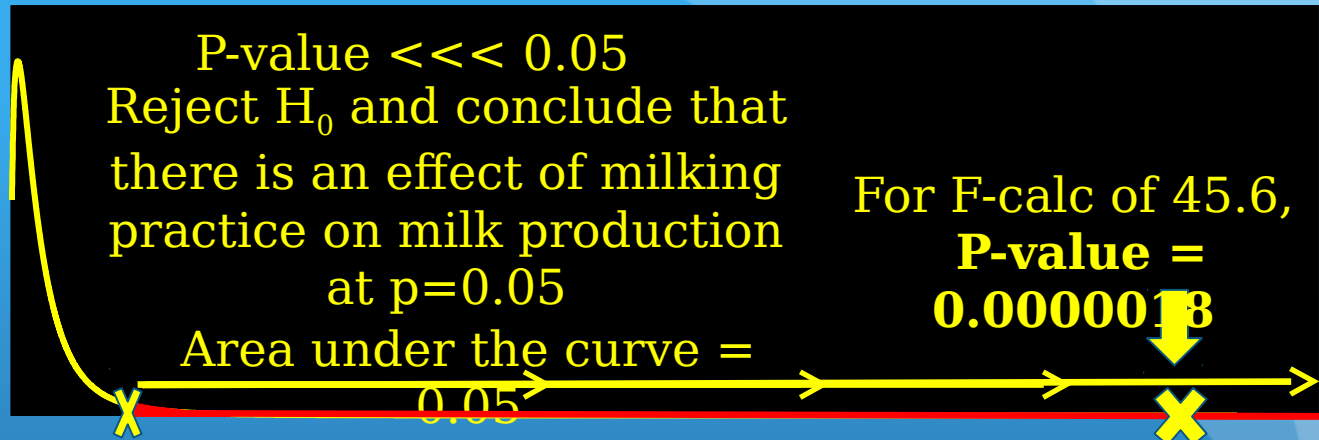
## ANOVA table

Response ( $Y_{ij}$ ) = milk production

Source of variation	Degrees of freedom (df)	Sum of Squares (SS)	Mean Sq (MS)	F value	P value (Pr(>p))
Total	11	18.1			p-value
Milking practice	3	17.1	5.7	45.6	<0.001
Error	8	1	0.125		

"F test":  $H_0 =$  ~~All treatment means are equal~~

Probability  
Density



## ANOVA table

Response ( $Y_{ij}$ ) = milk production

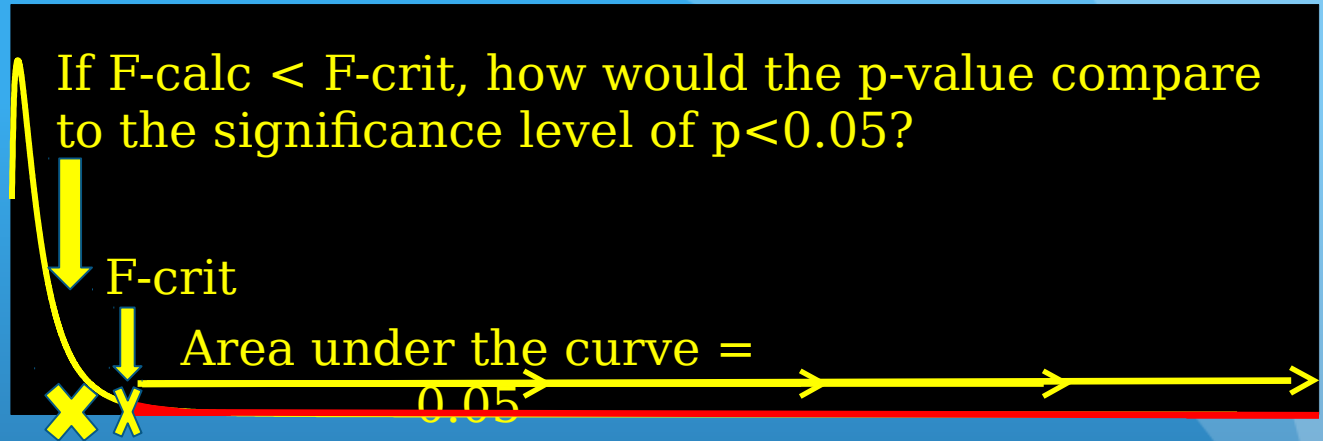
Source of variation	Degrees of freedom (df)	Sum of Squares (SS)	Mean Sq (MS)	F value	P value (Pr(>p))
Total	11	18.1			
Milking practice	3	17.1	5.7	45.6	<0.001*
Error	8	1	0.125		

\*\*\* = highly significant

<0.001\*  
\*\*

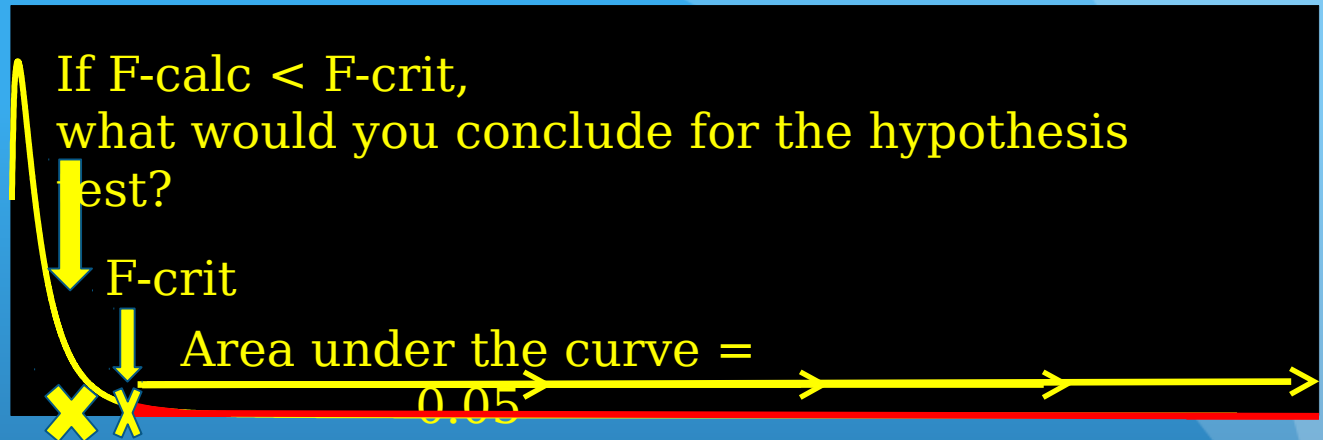
“F test”:  $H_0 =$  All treatment means are equal

Probability  
Density



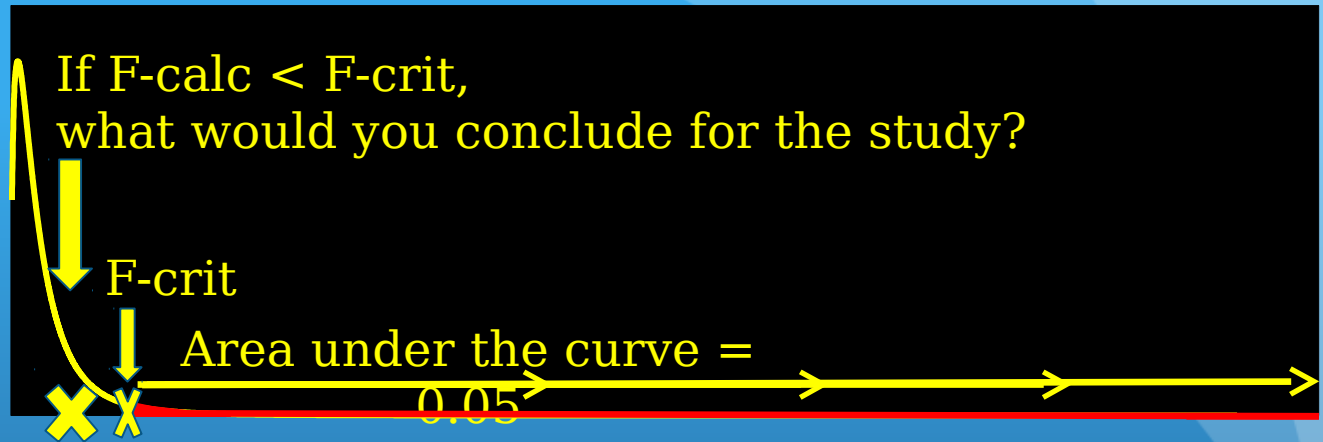
“F test”:  $H_0 =$  All treatment means are equal

Probability  
Density



“F test”:  $H_0 =$  All treatment means are equal

Probability  
Density



# Section 1 Learning Objectives













You should now have a basic understanding of:

1. Elements of an experiment
  - Experimental unit, independent variable, dependent variable, randomization
2. Complete Randomized Design (CRD)
3. Statistical model for a CRD with one treatment
4. Principles of Analysis of Variance (ANOVA) for a CRD with one treatment
  - Elements in an ANOVA table for a CRD with one treatment (Sources of variation, Sums of Squares, Mean Squares, F-value, P-value)
  - Hypothesis testing using F-test

We concluded that milking practice has an effect on milk production ( $p < 0.05$ ) 🐄

$$Y_{ij} = \mu + \tau_i + \varepsilon_{ij}$$

But *which* milking practices differ?













Hand-milk 2x	Hand-milk 3x	Machine- milk 2x	Machine- milk 3x
8.6 	11.2 	9.6 	11.5 
9.3 	10.8 	9.3 	12.0 
9.1 	11.7 	9.0 	11.8 
9.0	11.2	9.3	11.8

Grand  
mean =  
9.3

Treatment  
means:



# We will learn how to do all of this using the AIP App after the break!

Hand-milk 2x	Hand-milk 3x	Machine- milk 2x	Machine- milk 3x
8.6 	11.2 	9.6 	11.5 
9.3 	10.8 	9.3 	12.0 
9.1 	11.7 	9.0 	11.8 
9.0	11.2	9.3	11.8

Grand  
mean =  
9.3

Treatment  
means:

# Section 2 Learning Objectives

1. Know how to use the AIP App to analyze a CRD
  - Formatting and loading CSV file
  - Testing assumptions (later), ANOVA, LSD mean separation
2. Know how to interpret results

# How to format data file for AIP inp

The screenshot shows a Microsoft Excel spreadsheet titled "Mile Production Data for Screenshots.xlsx". The spreadsheet is formatted for AIP input with the following data:

	A	B	C	D
1	Hand-milk 2x	Hand-milk 3x	Machine-milk 2x	Machine-milk 3x
2	8.6	11.2	9.6	11.5
3	9.3	10.8	9.3	12
4	9.1	11.7	9	11.8
5				
6				
7				
8				
9				
10				
11				
12				
13				
14				
15				

# First note, column versus row.

	A	B	C	D
1	Hand-milk 2x	Hand-milk 3x	Machine- milk 2x	Machine- milk 3x
2	8.6	11.2	9.6	11.5
3	9.3	10.8	9.3	12
4	9.1	11.7	9	11.8
5				
6				
7				
8				
9				

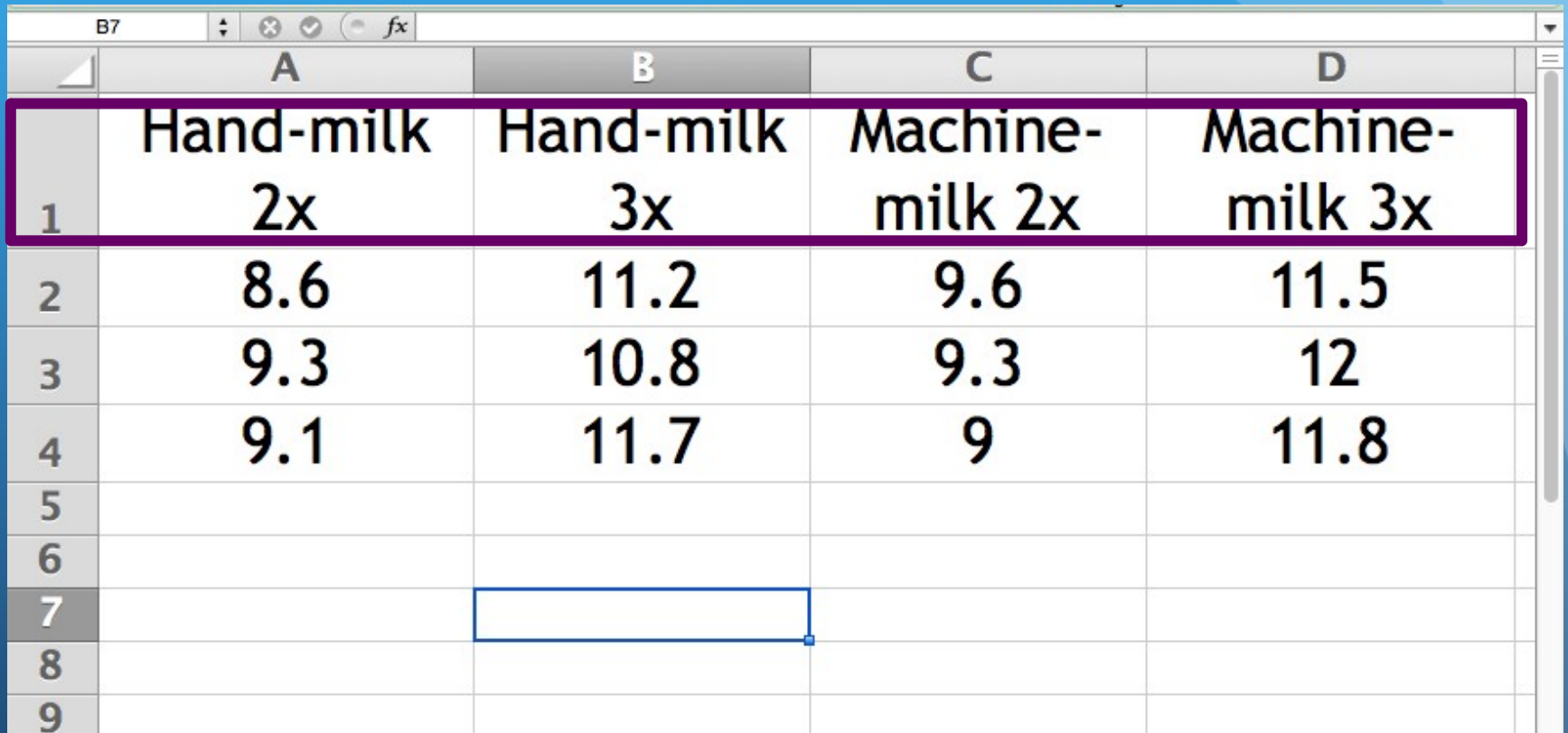
# Four columns

	A	B	C	D
1	Hand-milk 2x	Hand-milk 3x	Machine- milk 2x	Machine- milk 3x
2	8.6	11.2	9.6	11.5
3	9.3	10.8	9.3	12
4	9.1	11.7	9	11.8
5				
6				
7				
8				
9				

# Four rows

	A	B	C	D
1	Hand-milk 2x	Hand-milk 3x	Machine- milk 2x	Machine- milk 3x
2	8.6	11.2	9.6	11.5
3	9.3	10.8	9.3	12
4	9.1	11.7	9	11.8
5				
6				
7				
8				
9				

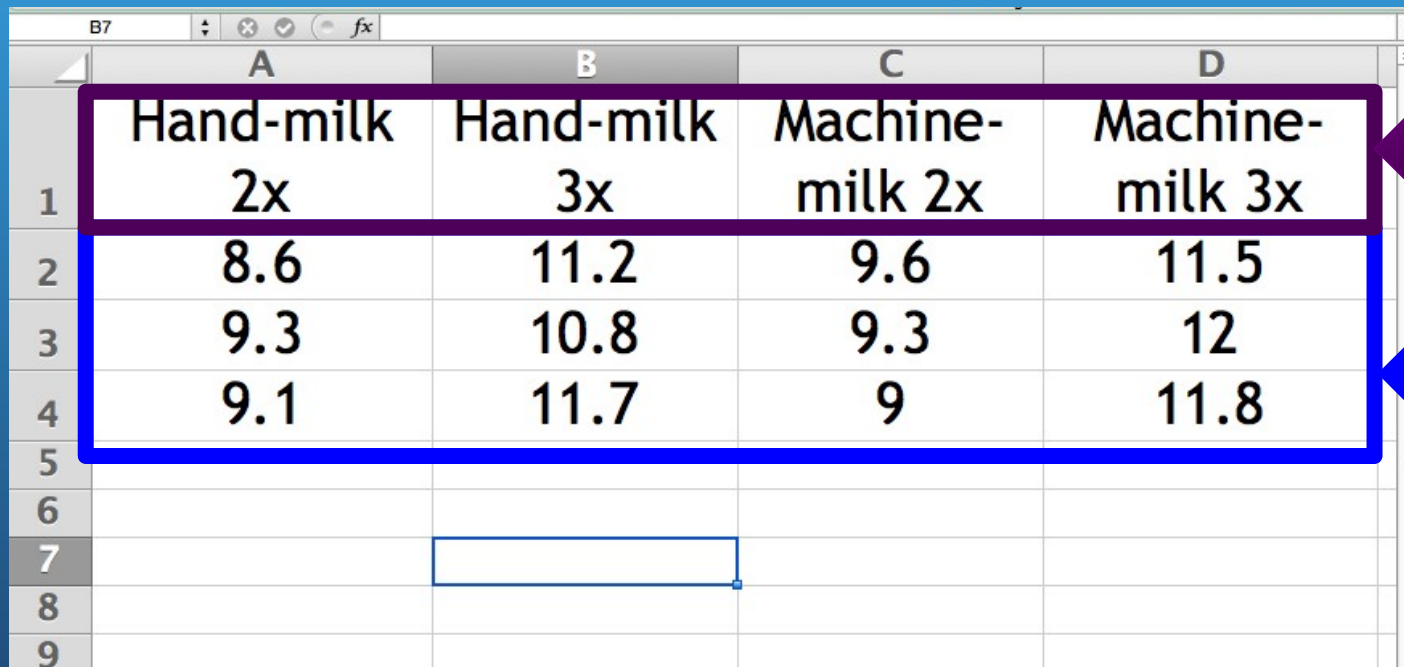
# Header row shows the titles of each column



	A	B	C	D
1	Hand-milk 2x	Hand-milk 3x	Machine-milk 2x	Machine-milk 3x
2	8.6	11.2	9.6	11.5
3	9.3	10.8	9.3	12
4	9.1	11.7	9	11.8
5				
6				
7				
8				
9				

Formatting data file for App:

Dependent data (response variable) must be cut and paste to a single column and independent data (treatments) in a separate column.



	A	B	C	D
1	Hand-milk 2x	Hand-milk 3x	Machine-milk 2x	Machine-milk 3x
2	8.6	11.2	9.6	11.5
3	9.3	10.8	9.3	12
4	9.1	11.7	9	11.8
5				
6				
7				
8				
9				

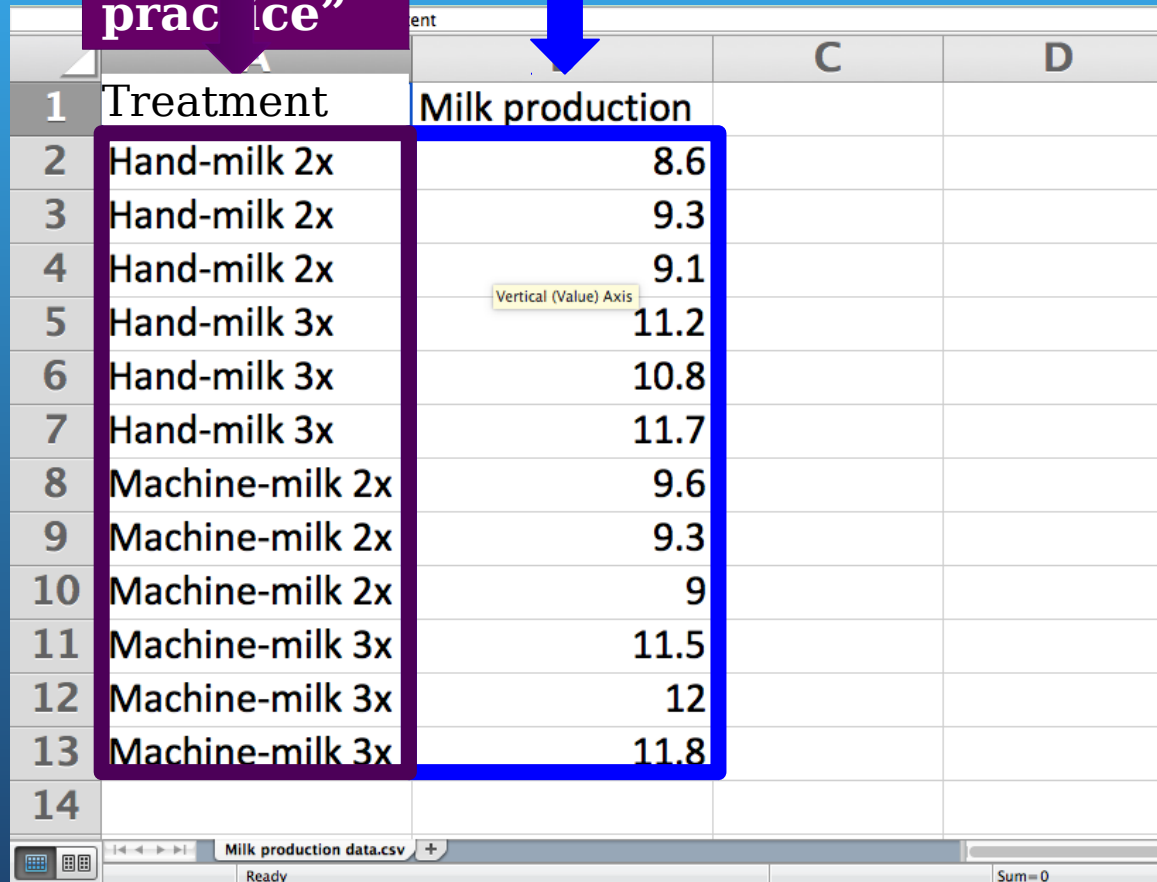
**Independent variable = Treatment , “milking practice”**

**Dependent variable = Milk production**



**Independent  
variable =  
Treatment  
, “milking  
practice”**

**Dependent  
variable =  
Milk  
production**

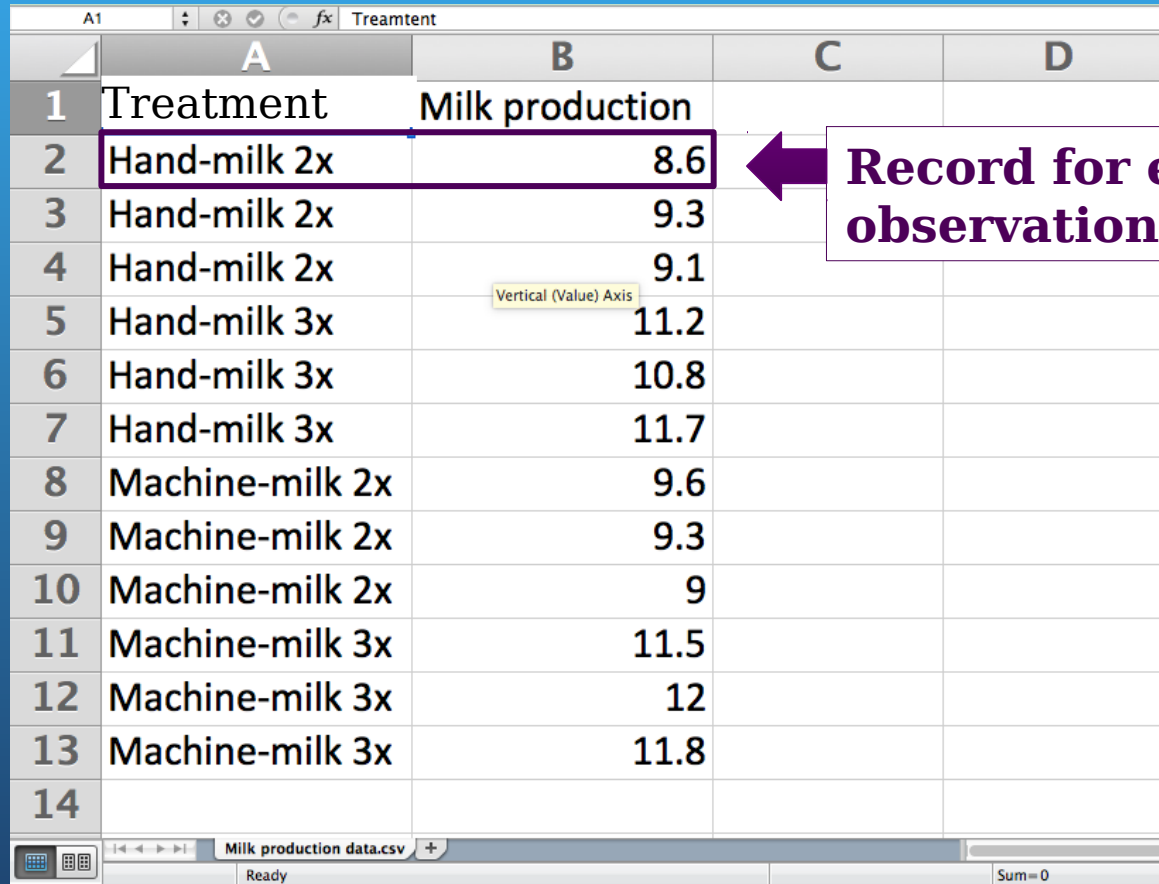


The image shows an Excel spreadsheet with the following data:

	Treatment	Milk production	C	D
1	Treatment	Milk production		
2	Hand-milk 2x	8.6		
3	Hand-milk 2x	9.3		
4	Hand-milk 2x	9.1		
5	Hand-milk 3x	11.2		
6	Hand-milk 3x	10.8		
7	Hand-milk 3x	11.7		
8	Machine-milk 2x	9.6		
9	Machine-milk 2x	9.3		
10	Machine-milk 2x	9		
11	Machine-milk 3x	11.5		
12	Machine-milk 3x	12		
13	Machine-milk 3x	11.8		
14				

The data is highlighted with a blue box. A label 'Vertical (Value) Axis' points to the 'Milk production' column.

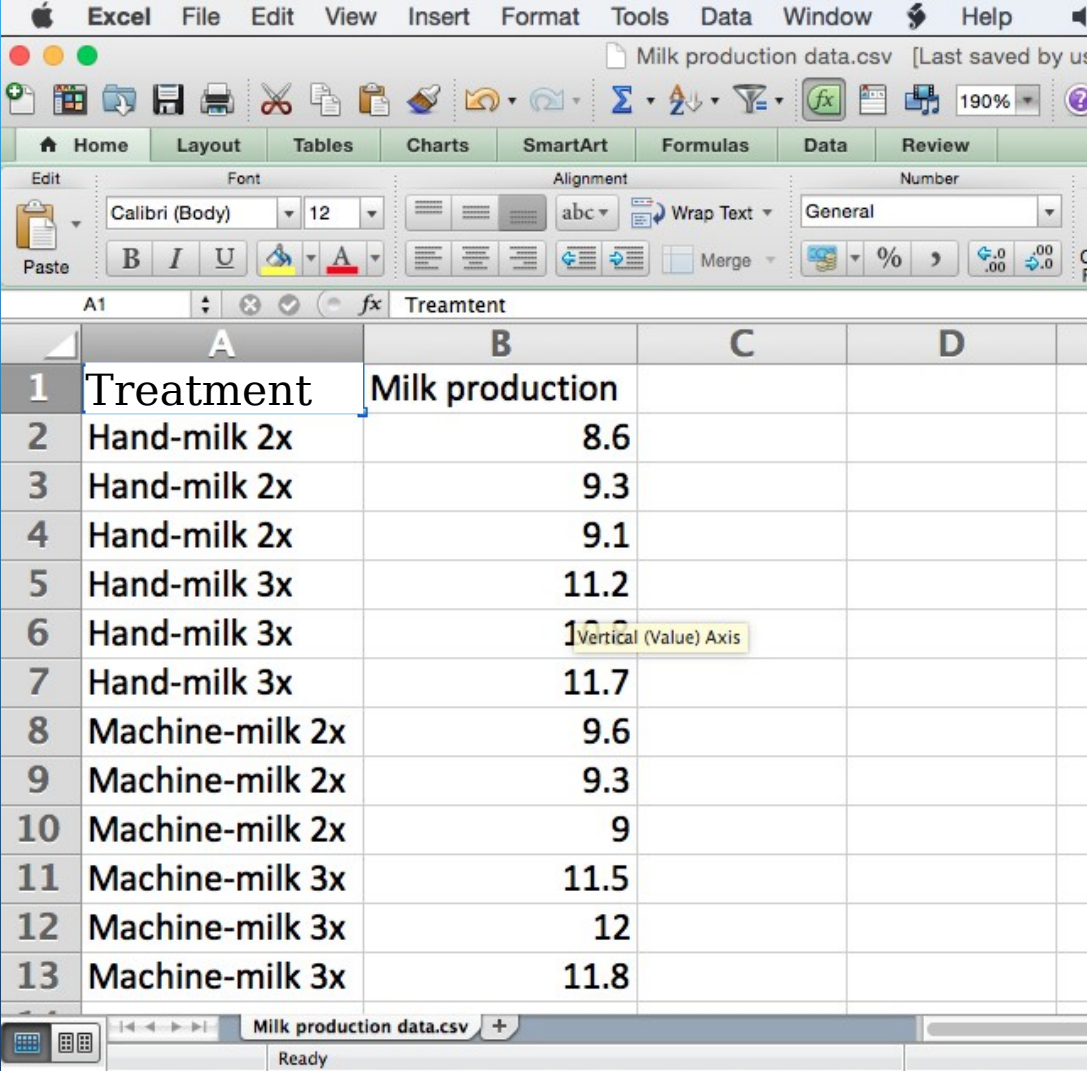
includes the (1) treatment (independent variable) and the (2) observation (dependent variable). If there were other independent variables (e.g. block, more treatments), they would have their own columns to identify all information for each record.



	A	B	C	D
1	Treatment	Milk production		
2	Hand-milk 2x	8.6		
3	Hand-milk 2x	9.3		
4	Hand-milk 2x	9.1		
5	Hand-milk 3x	11.2		
6	Hand-milk 3x	10.8		
7	Hand-milk 3x	11.7		
8	Machine-milk 2x	9.6		
9	Machine-milk 2x	9.3		
10	Machine-milk 2x	9		
11	Machine-milk 3x	11.5		
12	Machine-milk 3x	12		
13	Machine-milk 3x	11.8		
14				

**Record for each observation**

# How to save as .csv file?

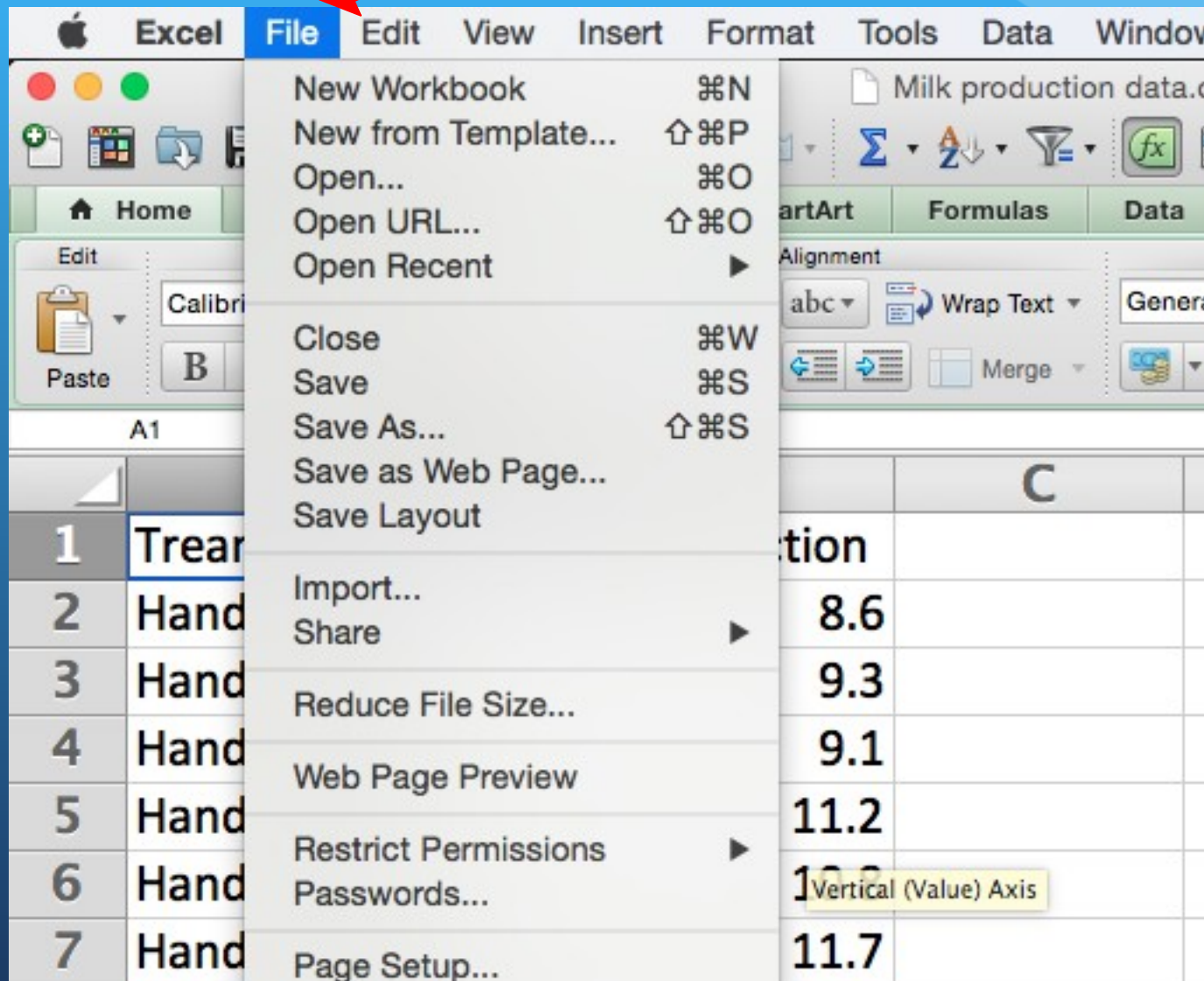


The screenshot shows the Microsoft Excel interface with the following data:

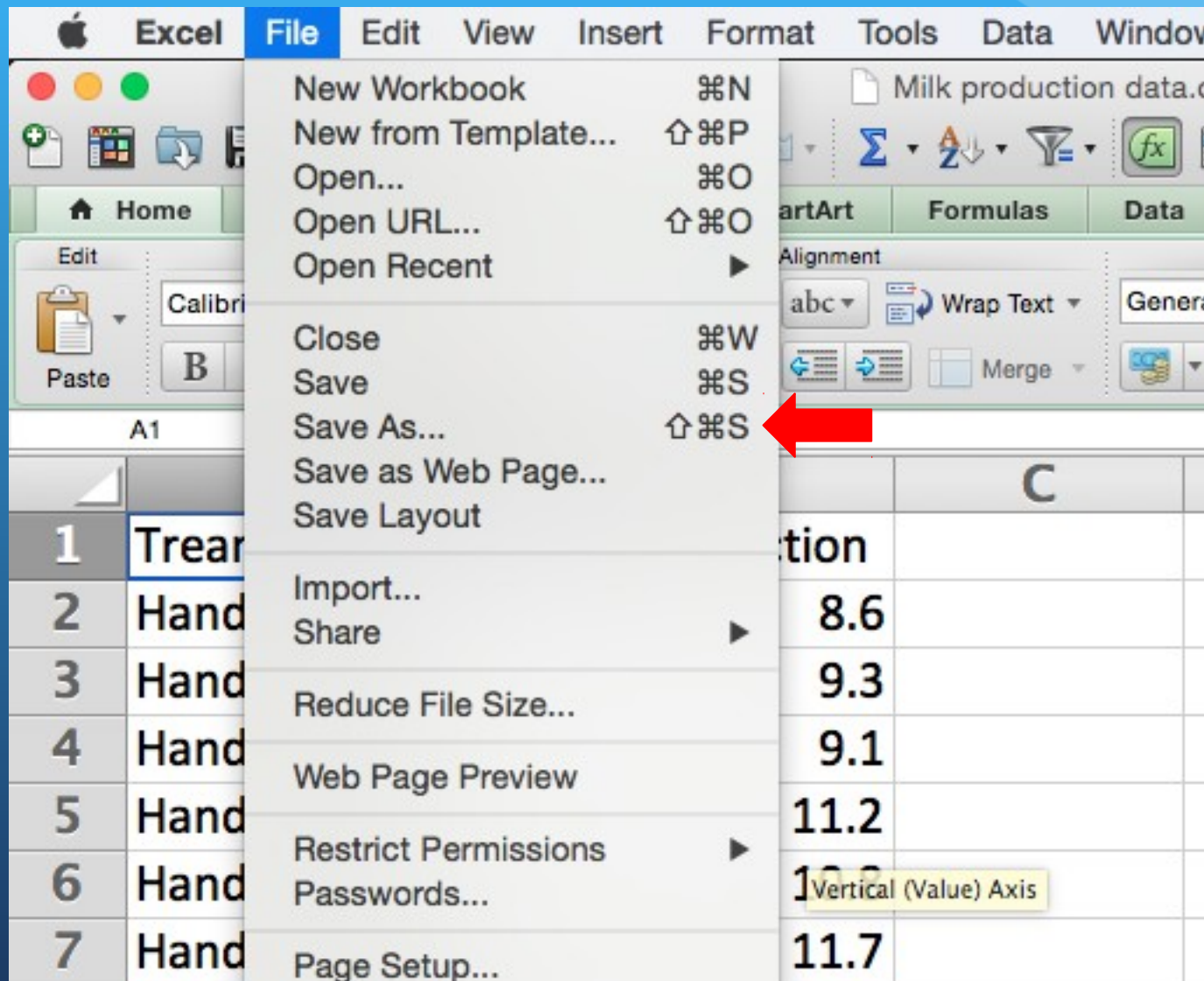
	A	B	C	D
1	Treatment	Milk production		
2	Hand-milk 2x	8.6		
3	Hand-milk 2x	9.3		
4	Hand-milk 2x	9.1		
5	Hand-milk 3x	11.2		
6	Hand-milk 3x	10.2		
7	Hand-milk 3x	11.7		
8	Machine-milk 2x	9.6		
9	Machine-milk 2x	9.3		
10	Machine-milk 2x	9		
11	Machine-milk 3x	11.5		
12	Machine-milk 3x	12		
13	Machine-milk 3x	11.8		

The status bar at the bottom indicates the file is named 'Milk production data.csv' and the status is 'Ready'.

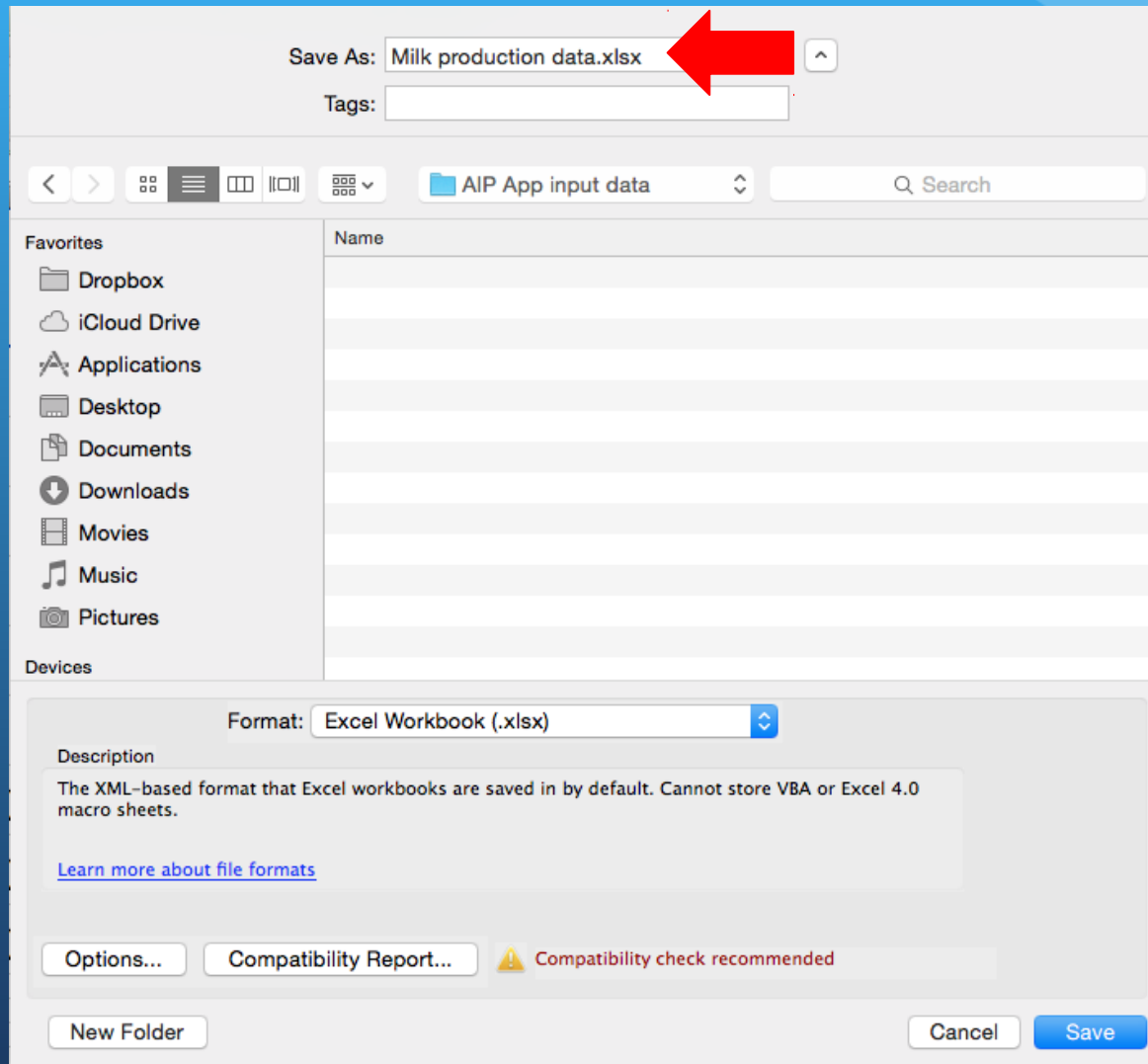
# 1. Click File



## 2. Click Save As...

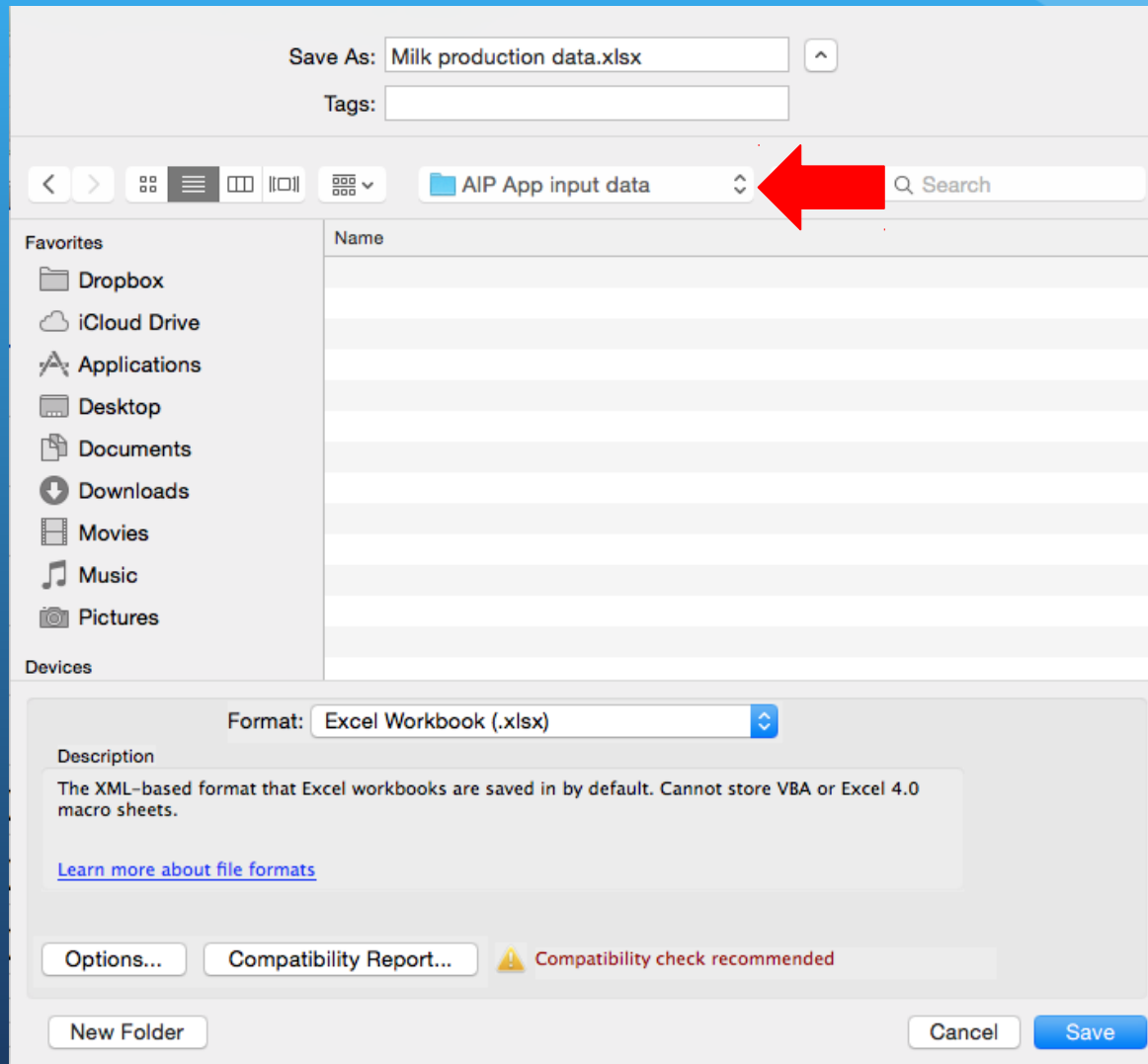


# 3. Enter filename

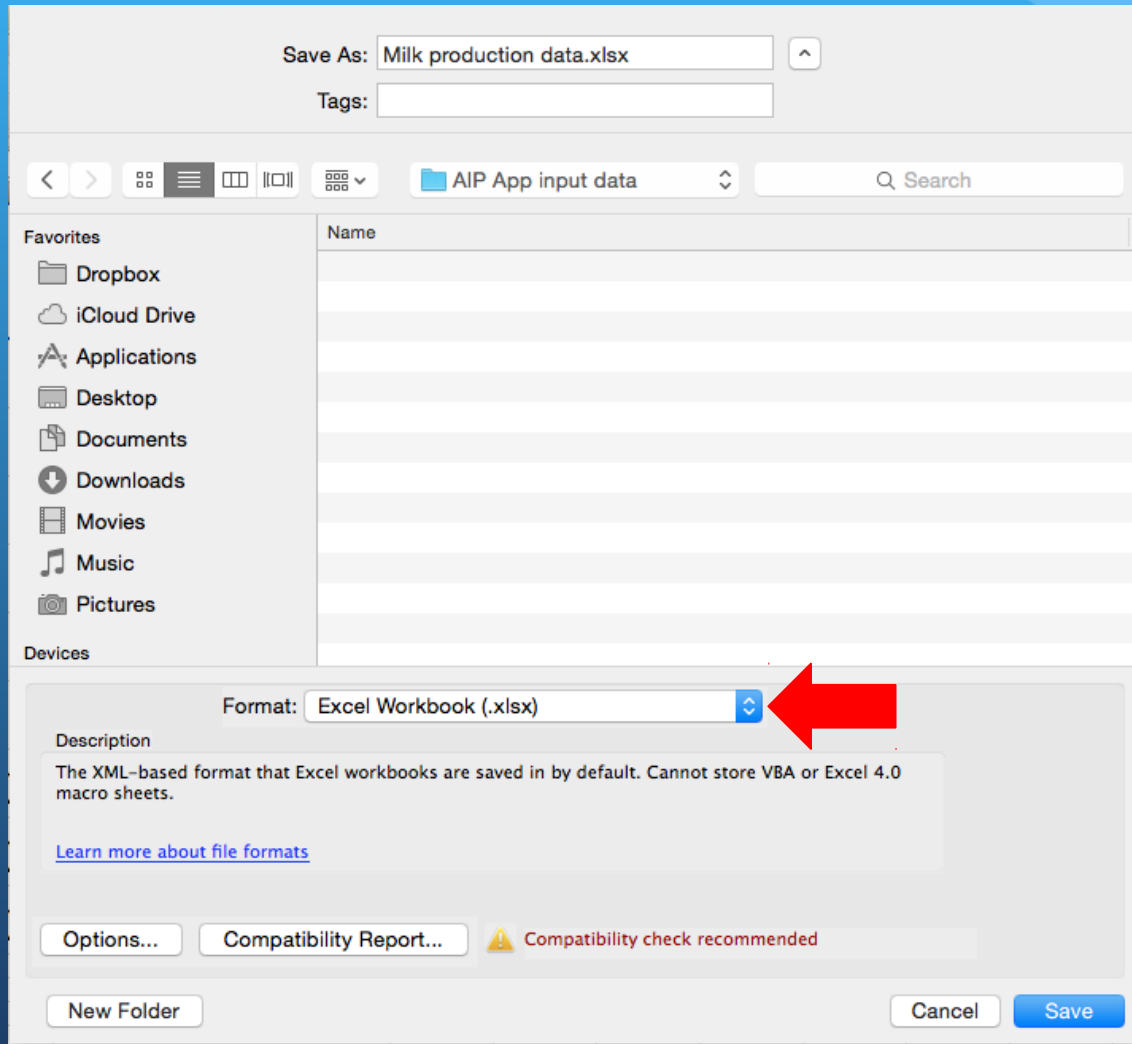




4. Note *where* you are saving the file.  
You will need to find it when you use the  
AIP app.

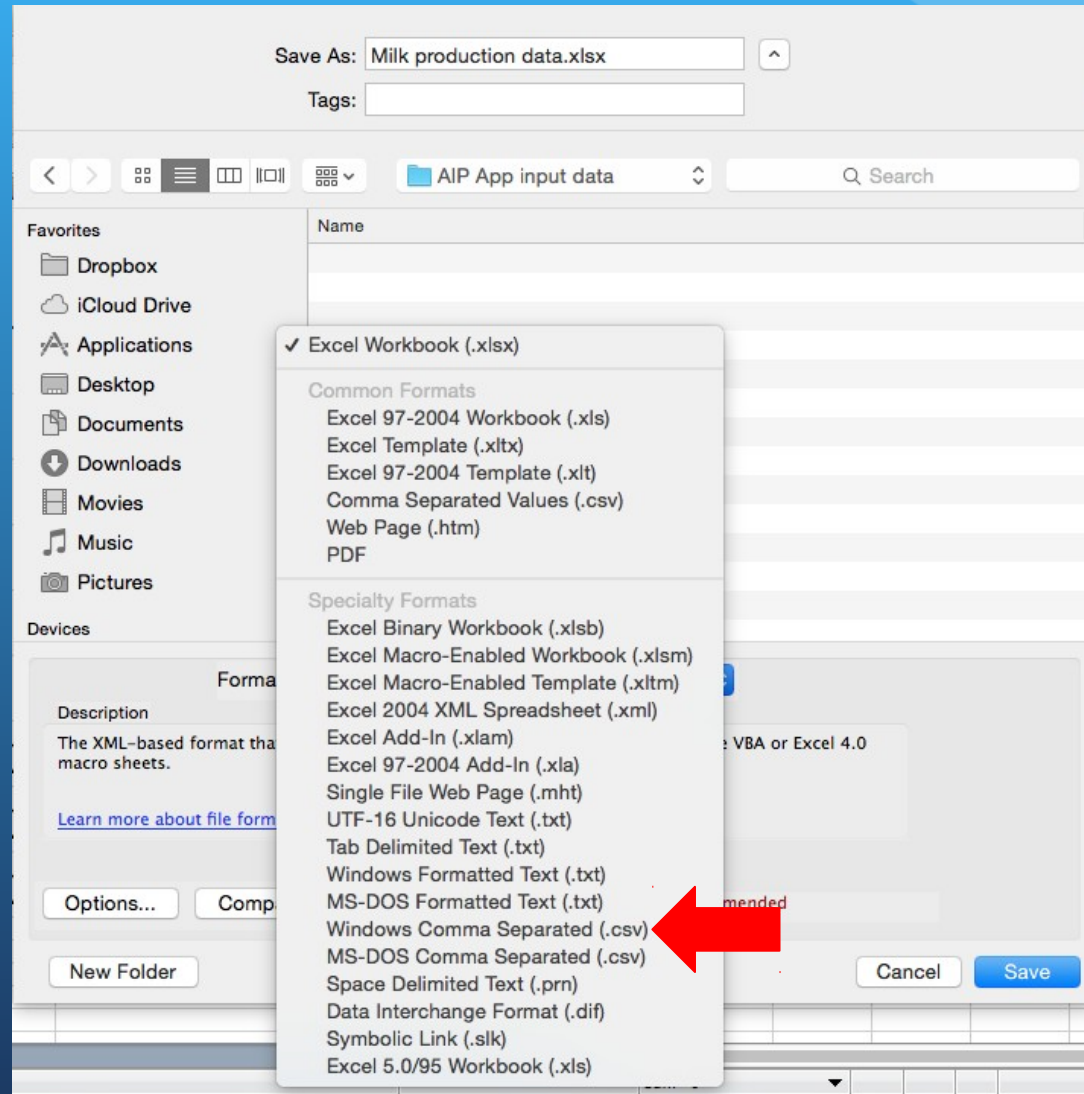


## 5. Click Format dropdown to select file type

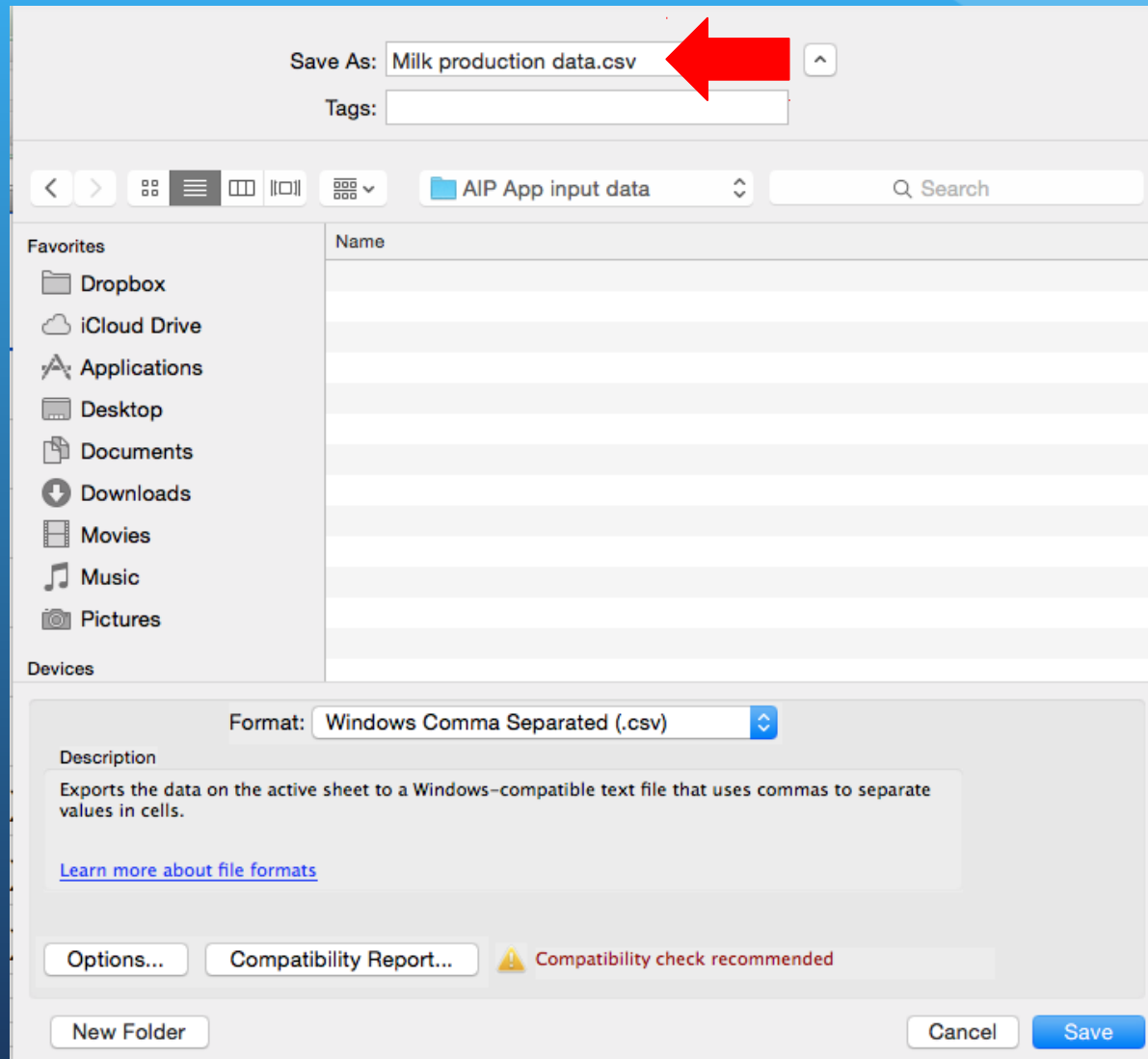




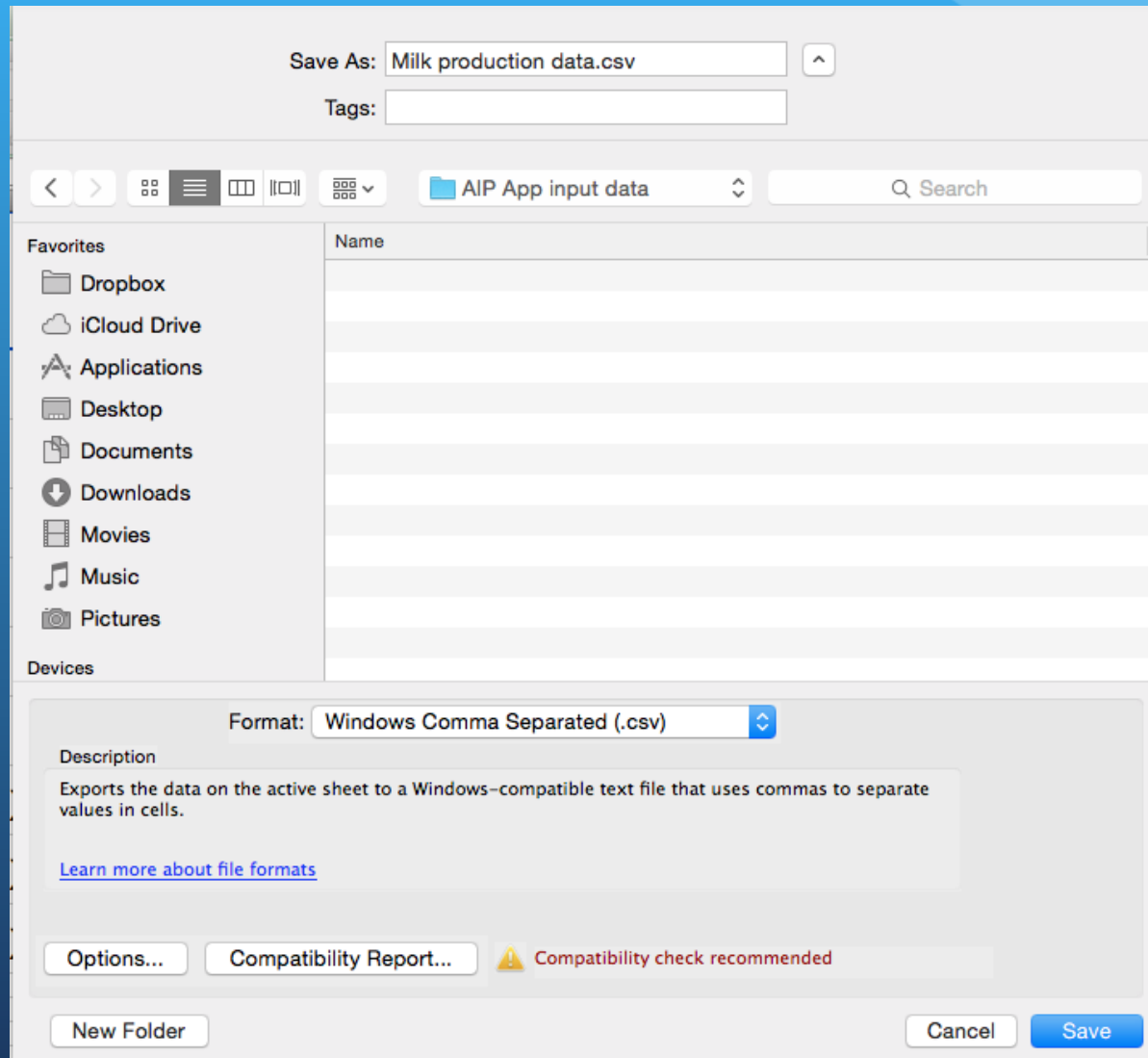
## 6. Click Windows Comma Separated (.csv)



## 7. Verify that your filename reads .csv



## 9. Lastly, click Save



# AIP App

- Simple and free
- We will learn how to:
  - How to format data in CSV file
  - Upload data to AIP app
  - Identify variables from CSV file and choose statistical model
  - Test assumptions of ANOVA
  - Post hoc tests
  - Plot results
  - Download statistics report and graphs

# How to load data into AIP App

AIP Analysis Interface

1. Load data

2. Data analysis

3. Post-hoc tests

4. Plots ▾

5. Report

Help ▾

Upload your CSV file by pressing  
"Load data" below

Your data should appear to the right. If this data  
is correct, please move to tab 2: "Data analysis"

Load data

Choose File No file chosen

Code used to read in data:

```
1 # code to read in your data
```

☐ Use sample data instead

Loaded Data

# 1. Click Choose File and select your csv file

AIP Analysis Interface

1. Load data

2. Data analysis

3. Post-hoc tests

4. Plots ▾

5. Report

Help ▾

Upload your CSV file by pressing "Load data" below

Your data should appear to the right. If this data is correct, please move to tab 2: "Data analysis"

Load data

Choose File

Import

Code used to read in data:

1

# code to read in your data

☐ Use sample data instead

Loaded Data

# 1. Click Choose File and select your csv file

AIP Analysis Interface

1. Load data   2. Data analysis   3. Post-hoc tests   4. Plots ▾   5. Report   Help ▾

**Upload your CSV file by pressing "Load data" below**

Your data should appear to the right. If this data is correct, please move to tab 2: "Data analysis"

**Load data**

No file chosen

Code used to read in data:

```
1 my_data <- read.csv(file = "Milk production data.csv")
```

☐ Use sample data instead

**Loaded Data**

Show  entries

Search:

Treamtent	Milk.production
Hand-milk 2x	8.6
Hand-milk 2x	9.3
Hand-milk 2x	9.1
Hand-milk 3x	11.2
Hand-milk 3x	10.8
Hand-milk 3x	11.7
Machine-milk 2x	9.6
Machine-milk 2x	9.3
Machine-milk 2x	9.0
Machine-milk 3x	11.5
Machine-milk 3x	12.0
Machine-milk 3x	11.8

Treamtent   Milk.production

Your data are loaded here

## 2. Click Data Analysis

AIP Analysis Interface

1. Load data

2. Data analysis

3. Diagnostic tests

4. Plots

5. Report

Help

Upload your CSV file by pressing "Load data" below

Your data should appear to the right. If this data is correct, please move to tab 2: "Data analysis"

Load data

Choose File

No file chosen

Upload complete

Code used to read in data:

```
1 my_data <- read.csv(file = "Milk
  production data.csv")
```

☐ Use sample data instead

Loaded Data

Show 25 entries

Search:

Treamtent	Milk.production
Hand-milk 2x	8.6
Hand-milk 2x	9.3
Hand-milk 2x	9.1
Hand-milk 3x	11.2
Hand-milk 3x	10.8
Hand-milk 3x	11.7
Machine-milk 2x	9.6
Machine-milk 2x	9.3
Machine-milk 2x	9.0
Machine-milk 3x	11.5
Machine-milk 3x	12.0
Machine-milk 3x	11.8

Treamtent

Milk.production



## 2. Click Data Analysis

AIP Analysis Interface

1. Load data

2. Data analysis

3. Post-hoc tests

4. Plots ▾

5. Report

Help ▾

1. Variable types

2. Type of analysis

3. Dependent variable

4. Independent variables

5. Interactions

6. Model check

Run analysis

1

# code to run analysis

Your data  
analysis  
options are  
here

# 3. Click Variable types

AIP Analysis Interface

1. Load data   2. Data analysis   3. Post-hoc tests   4. Plots ▾   5. Report   Help ▾

1. Variable types

2. Type of analysis

3. Dependent variable

4. Independent variables

5. Interactions

6. Model check

Run analysis

1

# code to run analysis

# 3. Click Variable types

AIP Analysis Interface

1. Load data

2. Data analysis

3. Post-hoc tests

4. Plots ▾

5. Report

Help ▾

1. Variable types

variable\_type\_panel

Indicate your variable types below

We have made guesses at the variable types in your data, but change the variable types below if they are incorrect.

Treatment

☐ Numeric ☒ Grouping

Milk.production

☒ Numeric ☐ Grouping

Information on variable type

2. Type of analysis

3. Dependent variable

4. Independent variables

5. Interactions

6. Model check

Run analysis

The column titles from your CSV file appear here

# 4. Choose the type of variable for each data column

AIP Analysis Interface   1. Load data   **2. Data analysis**   3. Post-hoc tests   4. Plots ▼   5. Report   Help ▼

**1. Variable types**

variable\_type\_panel

Indicate your variable types below

We have made guesses at the variable types in your data, but change the variable types below if they are incorrect.

**Treatment**

☐ Numeric   ☒ Grouping

**Milk.production**

☒ Numeric   ☐ Grouping

Information on variable type

**2. Type of analysis**

**3. Dependent variable**

**4. Independent variables**

**5. Interactions**

**6. Model check**

Run analysis

The column titles from your CSV file appear here

# 5. Click Type of Analysis

**AIP Analysis Interface**1. Load data2. Data analysis3. Post-hoc tests4. Plots ▼5. ReportHelp ▼

**1. Variable types**

variable\_type\_panel

Indicate your variable types below

We have made guesses at the variable types in your data, but change the variable types below if they are incorrect.

**Treatment**

☐ Numeric ☒ Grouping

**Milk.production**

☒ Numeric ☐ Grouping

Information on variable types

**2. Type of analysis**

**3. Dependent variable**

**4. Independent variables**

**5. Interactions**

**6. Model check**

Run analysis

# 6. Select your statistical model

AIF Analysis Interface

1. Load data   2. Data analysis   3. Post-hoc tests   4. Plots ▾   5. Report   Help ▾

1. Variable types

2. Type of analysis

select\_analysis\_panel  
**Select your statistical model**  
t-test ▾  
Analysis information

3. Dependent variable

4. Independent variables

5. Interactions

6. Model check

Run analysis

```
1 # code to run analysis
```

# 6. Select your statistical

Model

Analysis Interface 1. Load data 2. Data analysis 3. Post-hoc tests 4. Plots 5. Report Help

1. Variable types

2. Type of analysis

select\_analysis\_panel  
Select your statistical model

t-test

t-test  
ANOVA  
Linear or Generalized Linear Model  
Randomized Complete Block Design

4. Independent variables

5. Interactions

6. Model check

Run analysis

```
1 # code to run analysis
```

The experimental design for the milk dataset was a Complete Randomized Design 🏗️ Choose ANOVA

# 7. Click Dependent variable

AIP Analysis Interface

1. Load data

2. Data analysis

3. Post-hoc tests

4. Plots ▾

5. Report

Help ▾

1. Variable types

2. Type of analysis

select\_analysis\_panel

Select your statistical model

ANOVA ▾

Analysis information

3. Dependent variable

4. Independent variables

5. Interactions

6. Model check

Run analysis

1 # code to run analysis



# 8. Select your dependent

**Air Analysis Interface**   1. Load data   2. Data analysis   3. Post-hoc tests   4. Plots ▾   5. Report   Help ▾

1. Variable types

2. Type of analysis

3. Dependent variable

select\_variables\_panel  
**Select your dependent variable**

Milk.production ▾

More info


4. Independent variables

5. Interactions

6. Model check

Run analysis

```
1 # code to run analysis
```



# 9. Click Independent variables

AIP Analysis Interface

1. Load data

2. Data analysis

3. Post-hoc tests

4. Plots ▾

5. Report

Help ▾

1. Variable types

2. Type of analysis

3. Dependent variable

select\_variables\_panel

Select your dependent variable

Milk.production ▾

More info

4. Independent variables

5. Interactions

6. Model check

Run analysis

1 # code to run analysis



# 10. Select your independent

AIP Analysis Interface

1. Load data   2. Data analysis   3. Post-hoc tests   4. Plots ▾   5. Report   Help ▾

1. Variable types

2. Type of analysis

3. Dependent variable

4. Independent variables

iv\_info\_panel

Select your independent variable(s)


Treatment

5. Interactions

6. Model check

Run analysis

1 # code to run analysis



# 11. Click interactions

AIP Analysis Interface

1. Load data

2. Data analysis

3. Post-hoc tests

4. Plots ▾

5. Report

Help ▾

1. Variable types

2. Type of analysis

3. Dependent variable

4. Independent variables

iv\_info\_panel

Select your independent variable(s)

Treatment

Independent variable info

5. Interactions

6. Model check

Run analysis

1 # code to run analysis

# 11. Click interactions

AIP Analysis Interface   1. Load data   **2. Data analysis**   3. Post-hoc tests   4. Plots ▾   5. Report   Help ▾

1. Variable types

2. Type of analysis

3. Dependent variable

4. Independent variables

5. Interactions

**Error:** argument of length 0

6. Model check

Run analysis

```
1 # code to run analysis
```

Recall there is only 1 factor in this statistical (i.e. milking practice treatment), so there is no interaction to choose

# 12. Click Model check

AIP Analysis Interface

1. Load data   2. Data analysis   3. Post-hoc tests   4. Plots ▾   5. Report   Help ▾

1. Variable types

2. Type of analysis

3. Dependent variable

4. Independent variables


5. Interactions

Error: argument of length 0

6. Model check

Run analysis

```
1 # code to run analysis
```



# 13. Click Run analysis

AIP Analysis Interface

1. Load data   2. Data analysis   3. Post-hoc tests   4. Plots ▾   5. Report   Help ▾

1. Variable types

2. Type of analysis

3. Dependent variable


4. Independent variables

5. Interactions

6. Model check

Run analysis

1   `# code to run analysis`



# 13. Click Run analysis

AIP Analysis Interface

1. Load data

2. Data analysis

3. Post-hoc tests

4. Plots ▾

5. Report

Help ▾

1. Variable types

2. Type of analysis

3. Dependent variable

4. Independent variables

5. Interactions

6. Model check

Run analysis

Here is  
the R  
code

```
1 # convert categorical variables to factors
2 my_data$Treamtent <- as.factor(my_data$Treamtent)
3
4 # run ANOVA
5 model.fit <- aov(formula = Milk.production ~ Treatment,
6                   data = my_data)
```

Here are  
your  
ANOVA  
results!

Analysis of Variance Table

Response: Milk.production

	Df	Sum Sq	Mean Sq	F value	Pr(>F)
Treatment	3	17.1292	5.7097	46.929	2.012e-05 ***
Residuals	8	0.9733	0.1217		

Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1



# Same as our calculations

ANOVA table

Response ( $Y_{ij}$ ) = milk production

Source of variation	Degrees of freedom (df)	Sum of Squares (SS)	Mean Sq (MS)	F value	P value (Pr(>p))
Total	11	18.1025			
Milking practice	3	17.1292	5.7097	46.929	<0.001** *
Residual	8	0.9733	0.1217		

## Analysis of Variance Table

Response: Milk.production

	Df	Sum Sq	Mean Sq	F value	Pr(>F)
Treatm ent	3	17.1292	5.7097	46.929	2.012e-05 ***
Residuals	8	0.9733	0.1217		

---

Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1

# Note the significance level for our p-value

ANOVA table

Response ( $Y_{ij}$ ) = milk production

Source of variation	Degrees of freedom (df)	Sum of Squares (SS)	Mean Sq (MS)	F value	P value (Pr(>p))
Total	11	18.1025			
Milking practice	3	17.1292	5.7097	46.929	<0.001** *
Residual	8	0.9733	0.1217		

## Analysis of Variance Table

Response: Milk.production

	Df	Sum Sq	Mean Sq	F value	Pr(>F)
Treatment	3	17.1292	5.7097	46.929	2.012e-05 ***
Residuals	8	0.9733	0.1217		

---

Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1

\*\*\* = Significant at the  $p < 0.001$  level

Now let's use the App to determine which milking practices are significantly different, and display the results using the plot function

- [will add slides of screenshots and explanations once app is ready]

# Section 2

You should now be able to:

1. Use the AIP App to analyze a CRD, including:
  - Format and load CSV file
  - Customize data analysis for CRD with a single treatment
  - Run an ANOVA and multiple mean comparison using LSD
  - Display results using the Plots tab
2. Interpret results

# Section 3 Learning Objectives

Have basic understanding of:

1. Randomized Complete Block Design (RCBD)
  - Reasons to block
  - How to block and randomize treatments
  - Statistical model for RCBD
2. How to analyze data, display results, and interpret results from an RCBD experiment using the AIP App

# Research question: Is there a difference in yield among 4 new wheat varieties?

What is the null hypothesis?  $H_0$ : all varieties have equal mean yield

What is the alternative hypothesis?  $H_A$ : at least two varieties differ in mean yield

What is the independent variable? Wheat variety

What is the dependent variable? Yield

Research question: Is there a difference in yield among 4 new wheat varieties?

To test  $H_0$  using 4 replications in a field where there is a known gradient in soil nitrogen (shown below), what would you choose as:

(1) the experimental design (identify the experimental unit)

(2) the layout/orientation of your plots?

And (3) how would you assign treatments to plots?

*Experimental  
Field*

*Discuss and draw the  
layout in groups*

(low  
)



(high)

Soil Nitrogen

Research question: Is there a difference in yield among 4 new wheat varieties?

1. What is the experimental design and experimental unit?

*Experimental Field*

1	2	3	4
5	6	7	8
9	10	11	12
13	14	15	16

Randomized Complete Block Design (RCBD) with 4 trts and 4 blocks.

4 trts x 4 blocks = 16 plots

Plots are called **EXPERIMENTAL UNITS**, defined as the “object” that the treatments are applied to.

(low  
)

Soil Nitrogen

(high)



Research question: Is there a difference in yield among 4 new wheat varieties?

Why do we “block” in this field rather than randomly assign the treatments to any of the 16 plots?

1	2	3	4
5	6	7	8
9	10	11	12
13	14	15	16

Randomized Complete Block Design (RCBD) with 4 trts and 4 blocks.

4 trts x 4 blocks = 16 plots

Plots are called **EXPERIMENTAL UNITS**, defined as the “object” that the treatments are applied to.

(low  
)

Soil Nitrogen

(high)

Research question: Is there a difference in yield among 4 new wheat varieties?

2. How do we arrange the blocks in the field?

Experimental  
Field

<b>1</b>	<b>2</b>	<b>3</b>	<b>4</b>
5	6	7	8
9	10	11	12
13	14	15	16

(low  
)

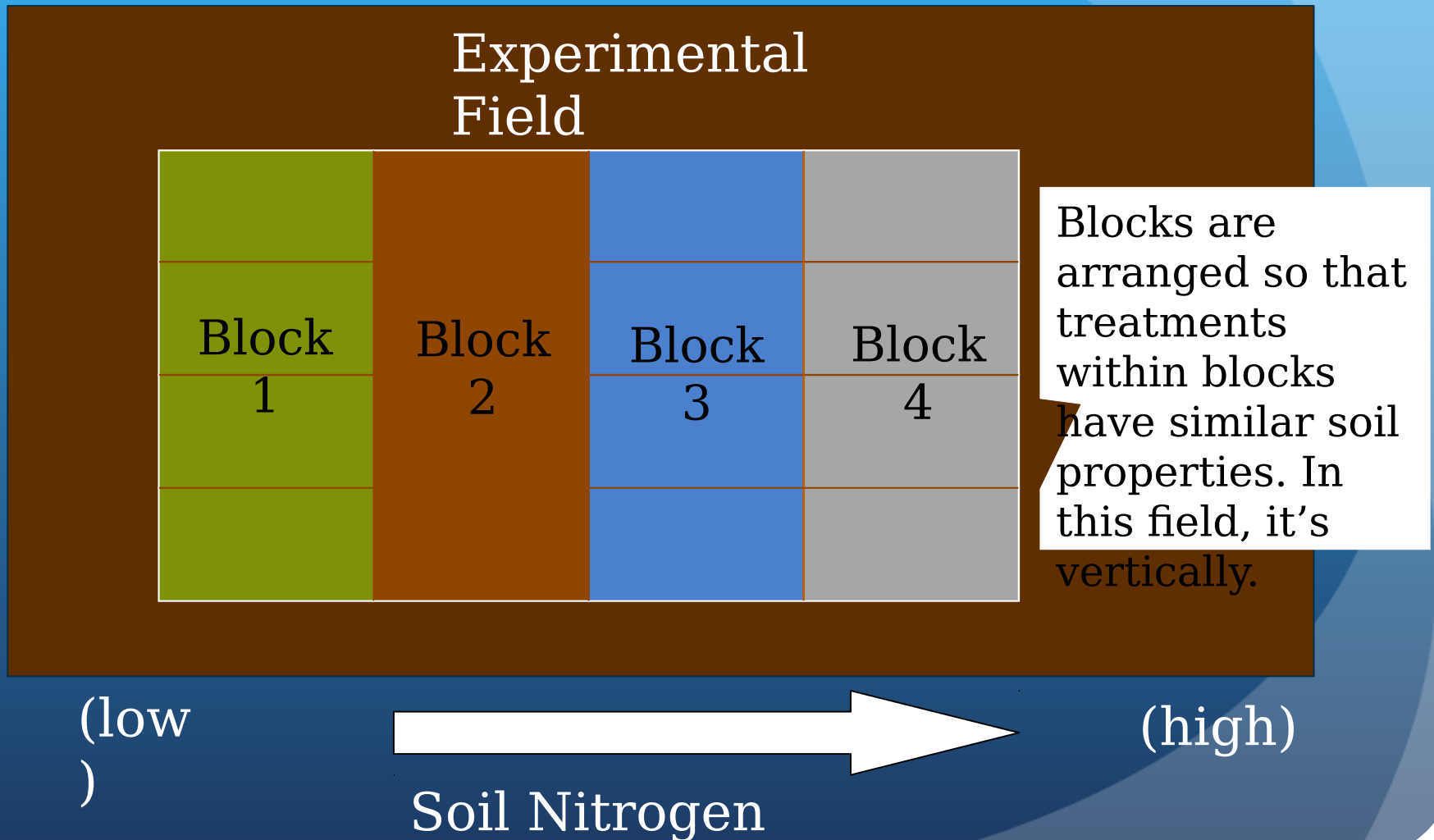


(high)

Soil Nitrogen

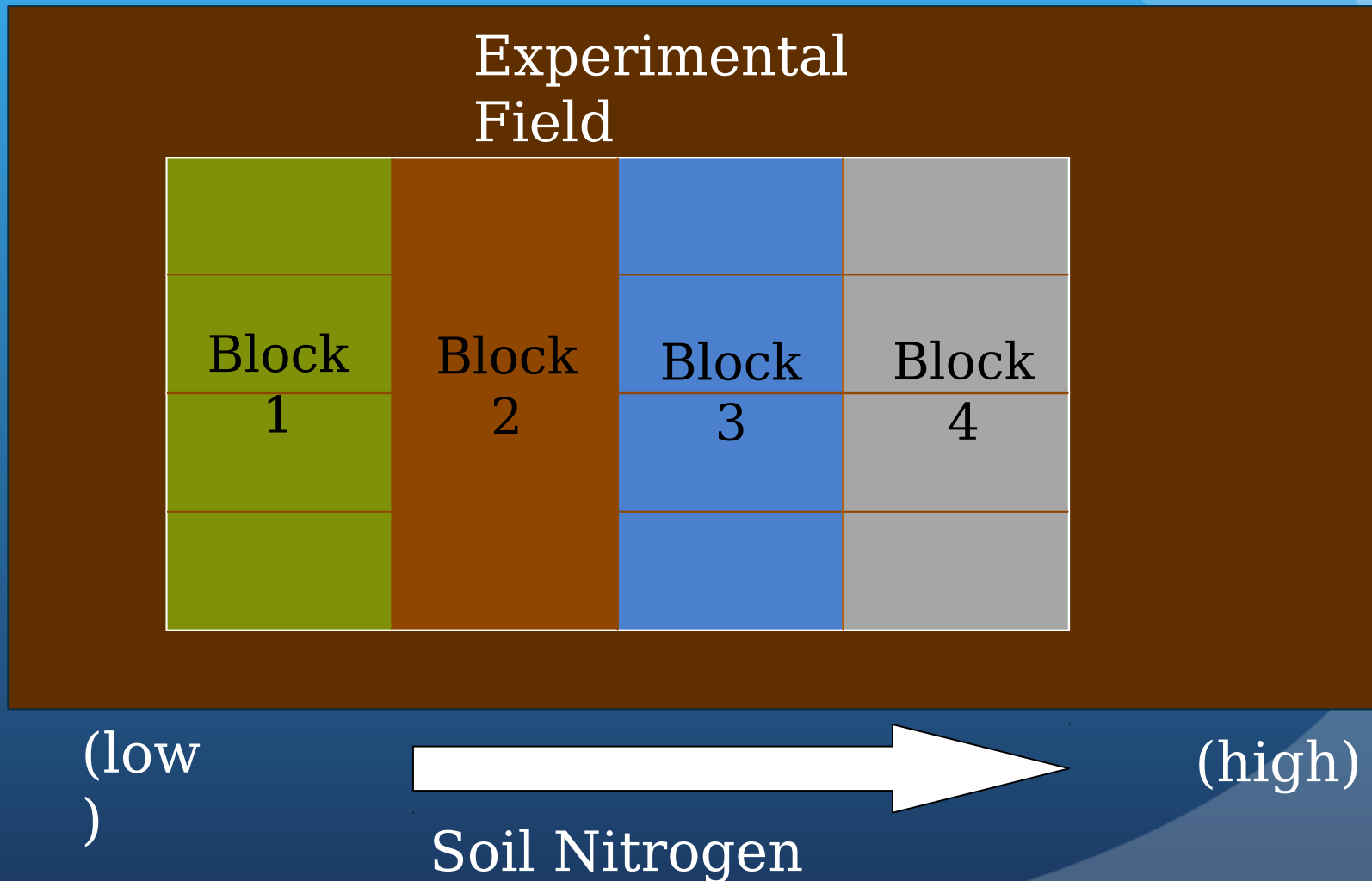
Research question: Is there a difference in yield among 4 new wheat varieties?

2. How do we arrange the blocks in the field?

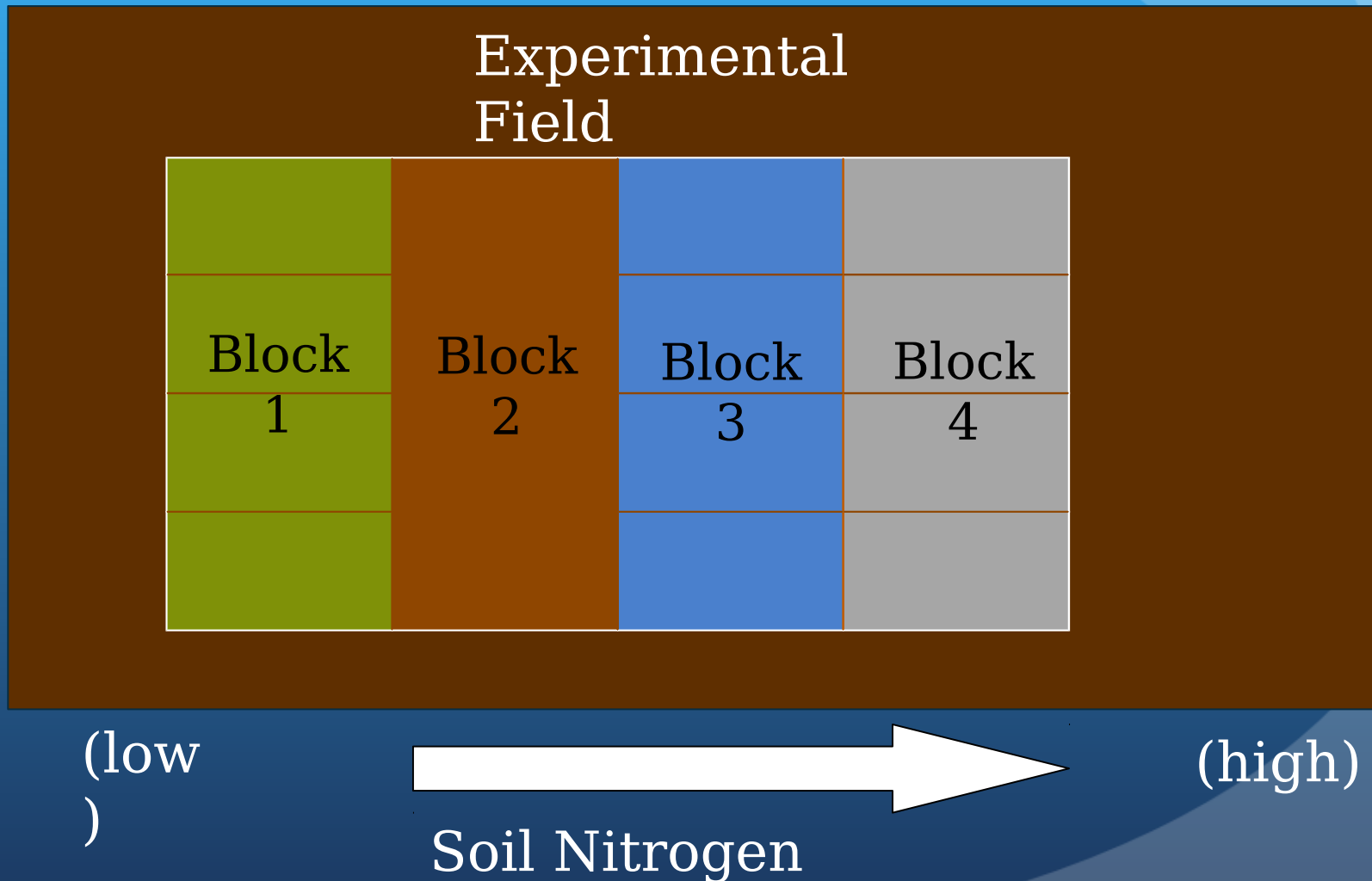


Research question: Is there a difference in yield among 4 new wheat varieties?

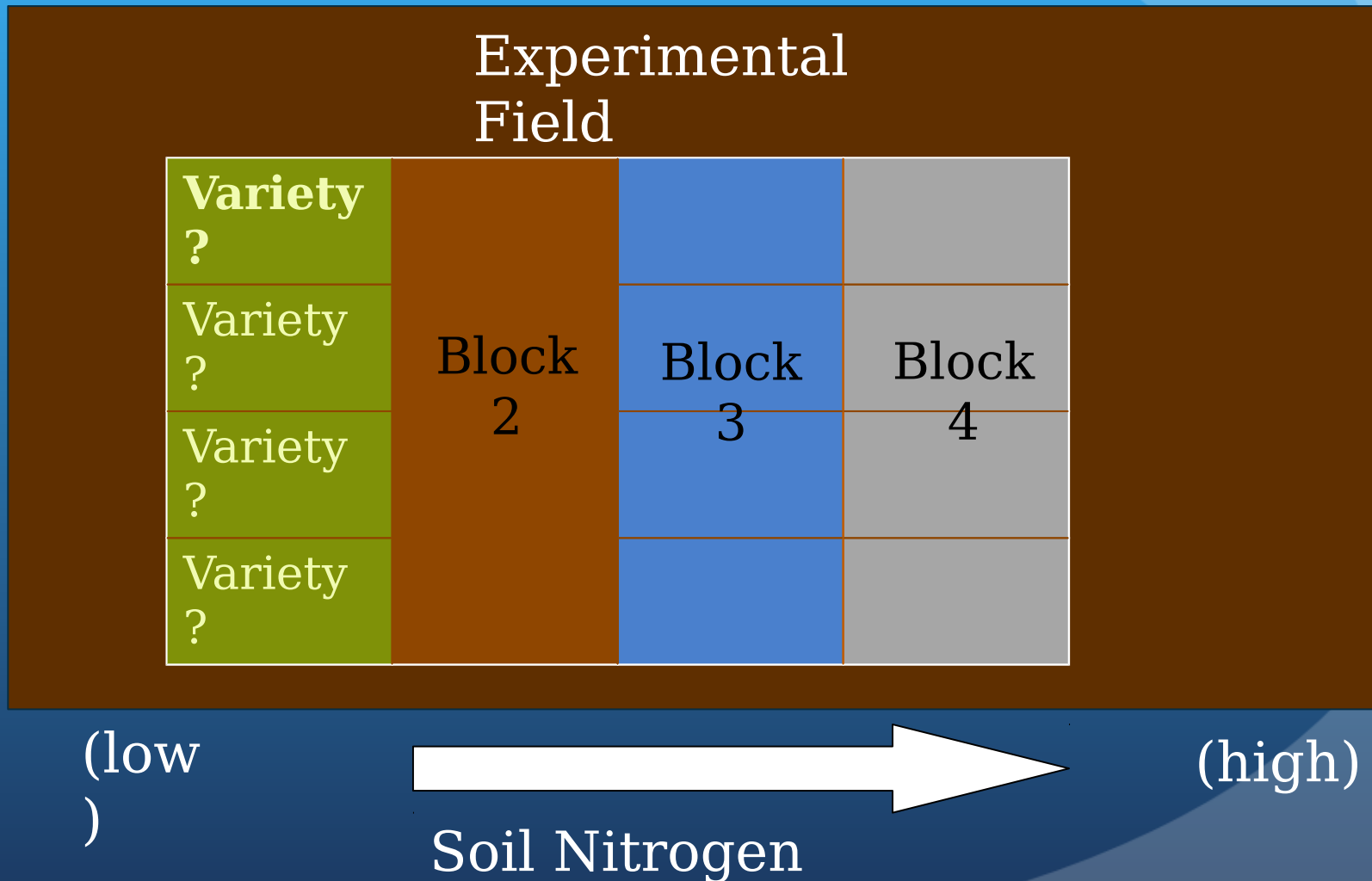
3. How are treatments assigned to plots?



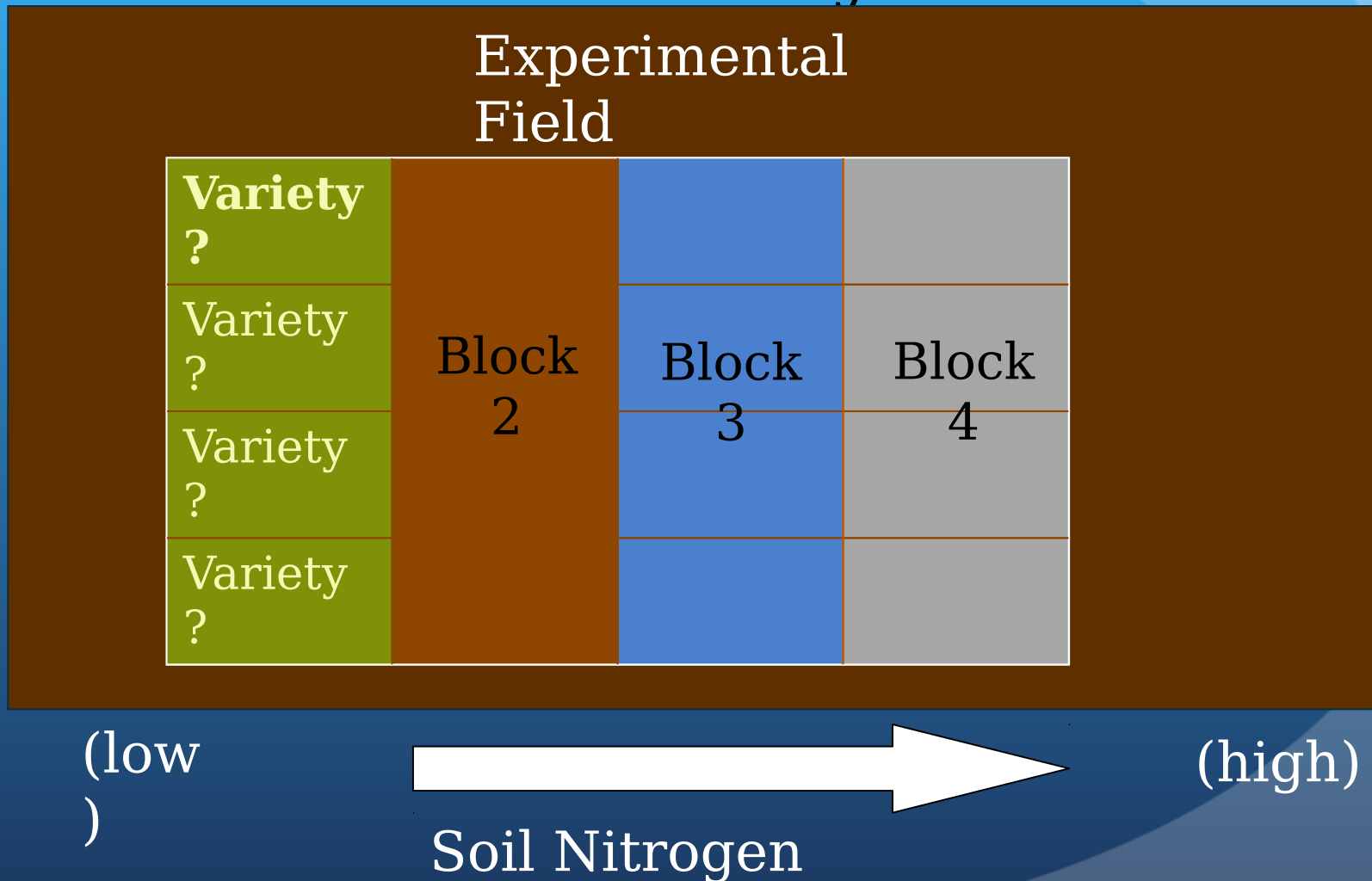
All treatments are **RANDOMLY** assigned to plots within each block.



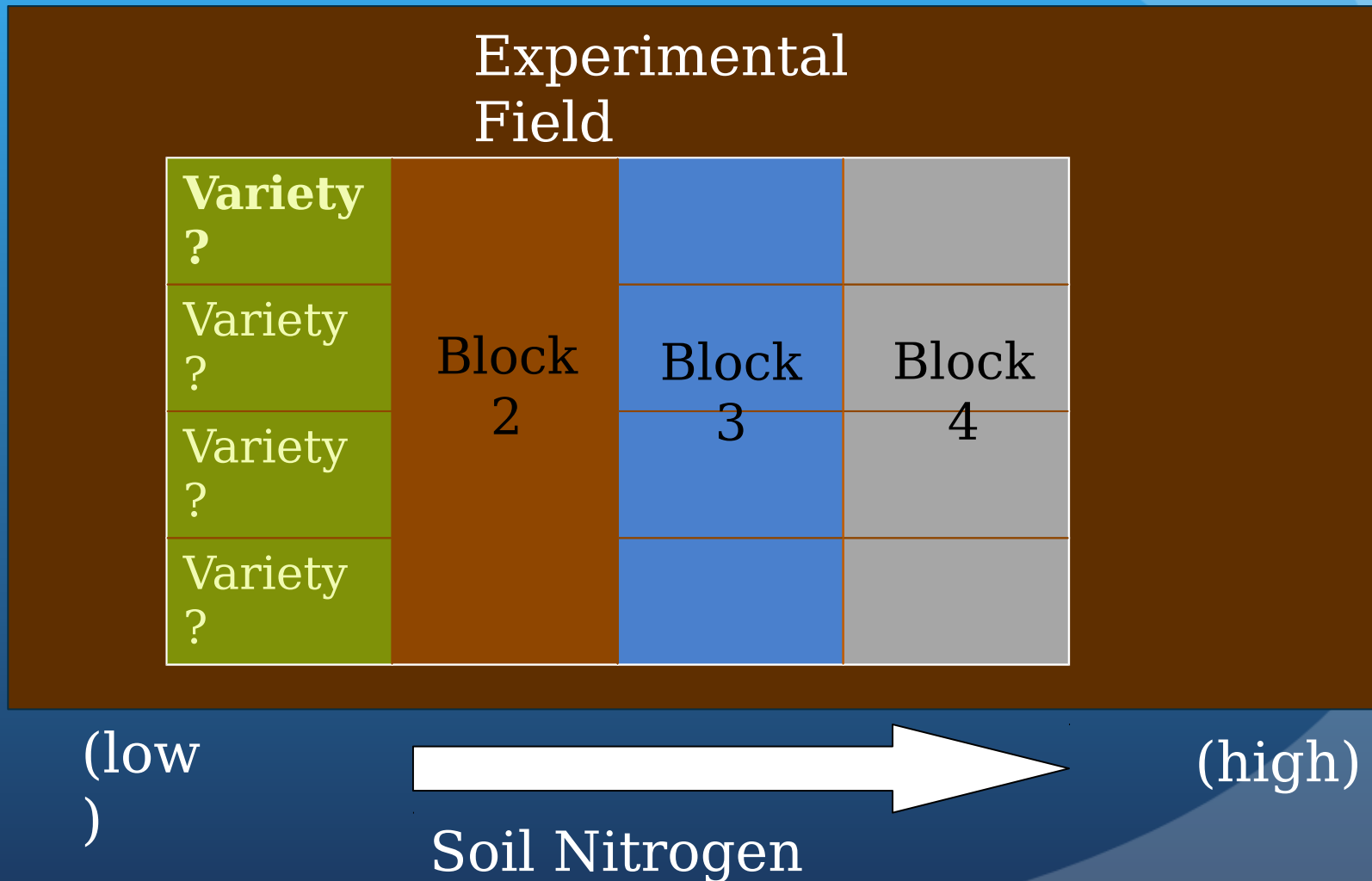
Starting with block 1, randomly assign the 4 wheat varieties to the 4 plots.



# But how do we do it randomly?



Using a **random number generator**, or drawing numbers from a hat without replacement. Any other ideas?





Example of free random number generator online

## Tools for Statistics

### Probability Calculators

Basic Probability  
Bayes Rule Calculator

### Mathematical Analysis

Random Number Table  
Combination-Permutation  
Factorial Calculator  
Event Counter

## Browse

### Tutorials

AP Statistics  
Statistics and Probability  
Matrix Algebra

### AP Statistics

Test Preparation  
Practice Exam  
Study Guide Review  
Approved Calculators  
AP Statistics Formulas  
FAQ: AP Statistics

### Stat Tables

Binomial  
Chi-Square  
f Distribution  
Hypergeometric  
Multinomial  
Negative Binomial  
Normal  
Poisson  
t Distribution

### Sample Planning Wizard

Description  
Capabilities

## Random Number Generator

Use the Random Number Generator to create a list of random numbers, based on your specifications. The numbers you generate appear in the [Random Number Table](#).

For help in using the Random Number Generator, read the [Frequently-Asked Questions](#) or review the [Sample Problems](#).

- Enter a value in each of the first three text boxes.
- Indicate whether duplicate entries are allowed in the table.
- Click the **Calculate** button to create a table of random numbers.

**Note:** The seed value is optional. Leave it blank to generate a new set of numbers. Use it to repeat a previously-generated set of numbers.

How many random numbers?

Minimum value

Maximum value

Allow duplicate entries

Seed (optional)

Enter number of treatments (4 varieties)

And the min and max values (1

Choose "False" for duplicate entries so you only get 1, 2, 3 and 4 as possibilities.

## Random Number Table

[Random Number Generator](#) | [Frequently-Asked Questions](#) | [Sample Problems](#)

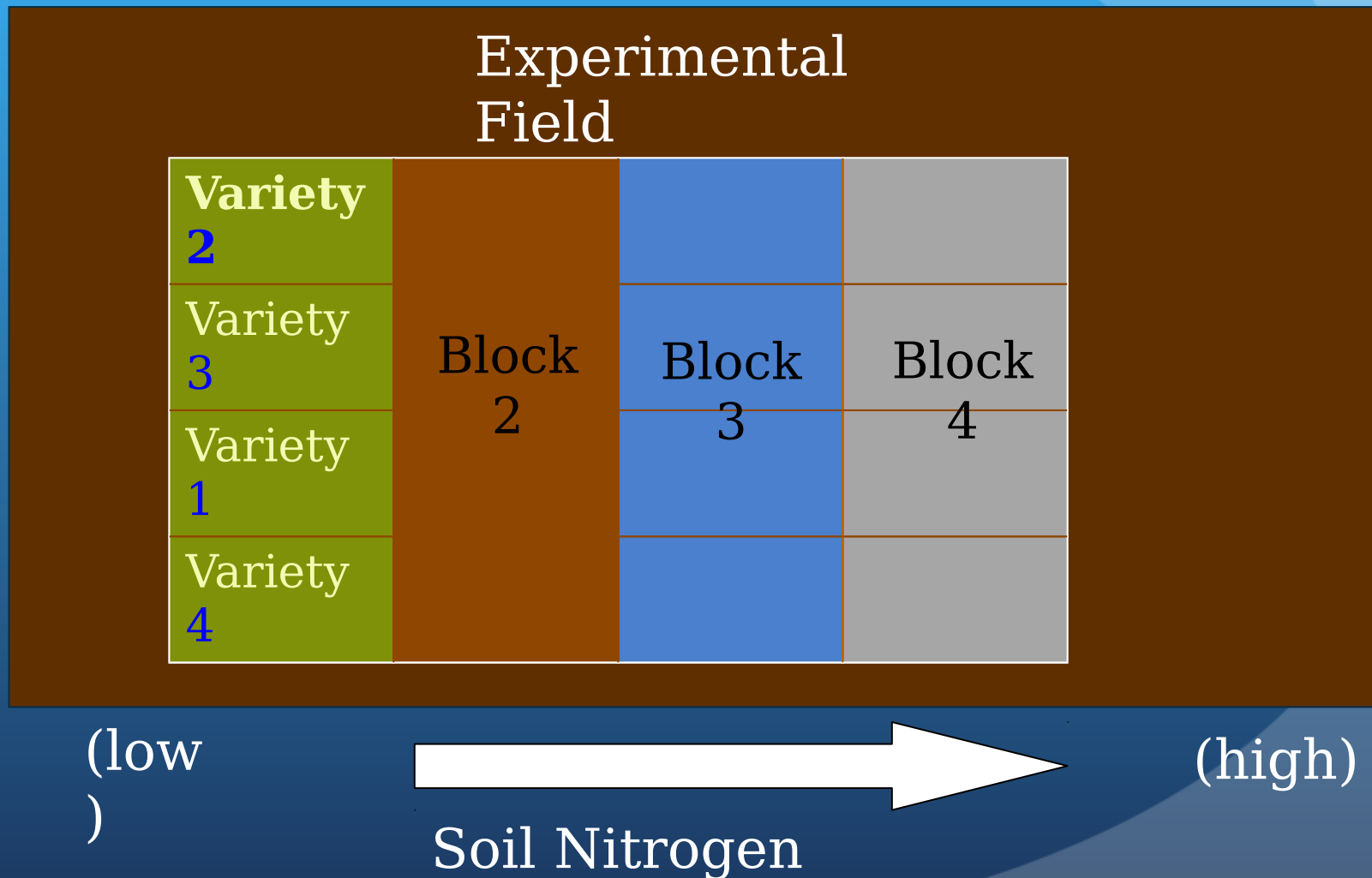
## 4 Random Numbers

2 3 1 4

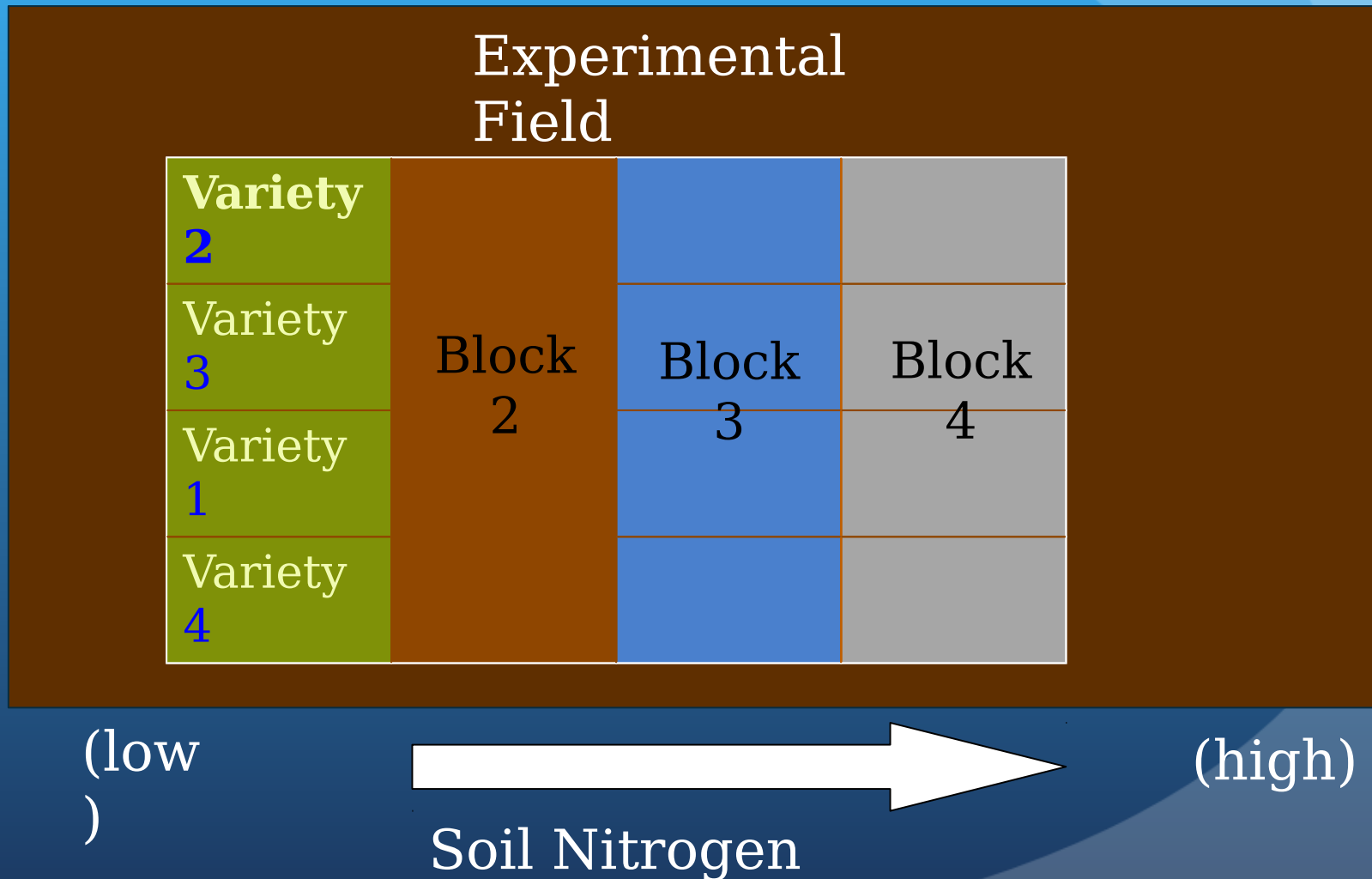
**Specs:** This table of 4 random numbers was produced according to the following specifications: Numbers were randomly selected from within the range of 1 to 4. Duplicate numbers were not allowed. This table was generated on 4/21/2015.

The order of your treatments for block 1! This time it was

# Randomly assign treatments to block 1!



Repeat for each block by refreshing website, or by replacing numbers into hat and drawing them again for each block.



Repeat for each block by refreshing website, or by replacing numbers into hat and drawing them again for each block.

Experimental  
Field Example,  
For example,

Variety 2	Variety 1	Variety 3	Variety 2
Variety 3	Variety 3	Variety 4	Variety 1
Variety 1	Variety 4	Variety 1	Variety 3
Variety 4	Variety 2	Variety 2	Variety 4

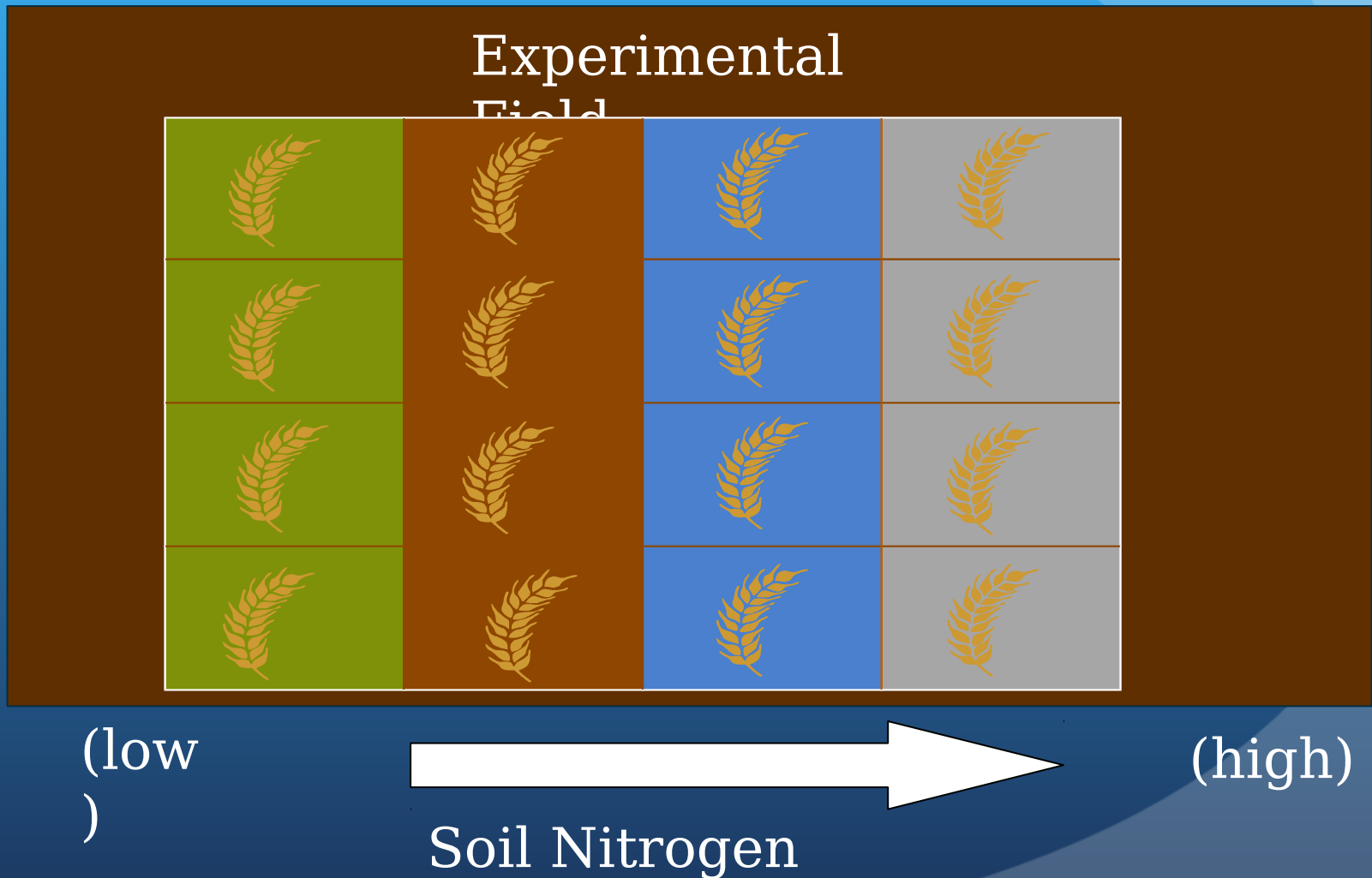
(low  
)



(high)

Soil Nitrogen

Our field trial is underway!



**Let's review, why use  
blocking variable?**

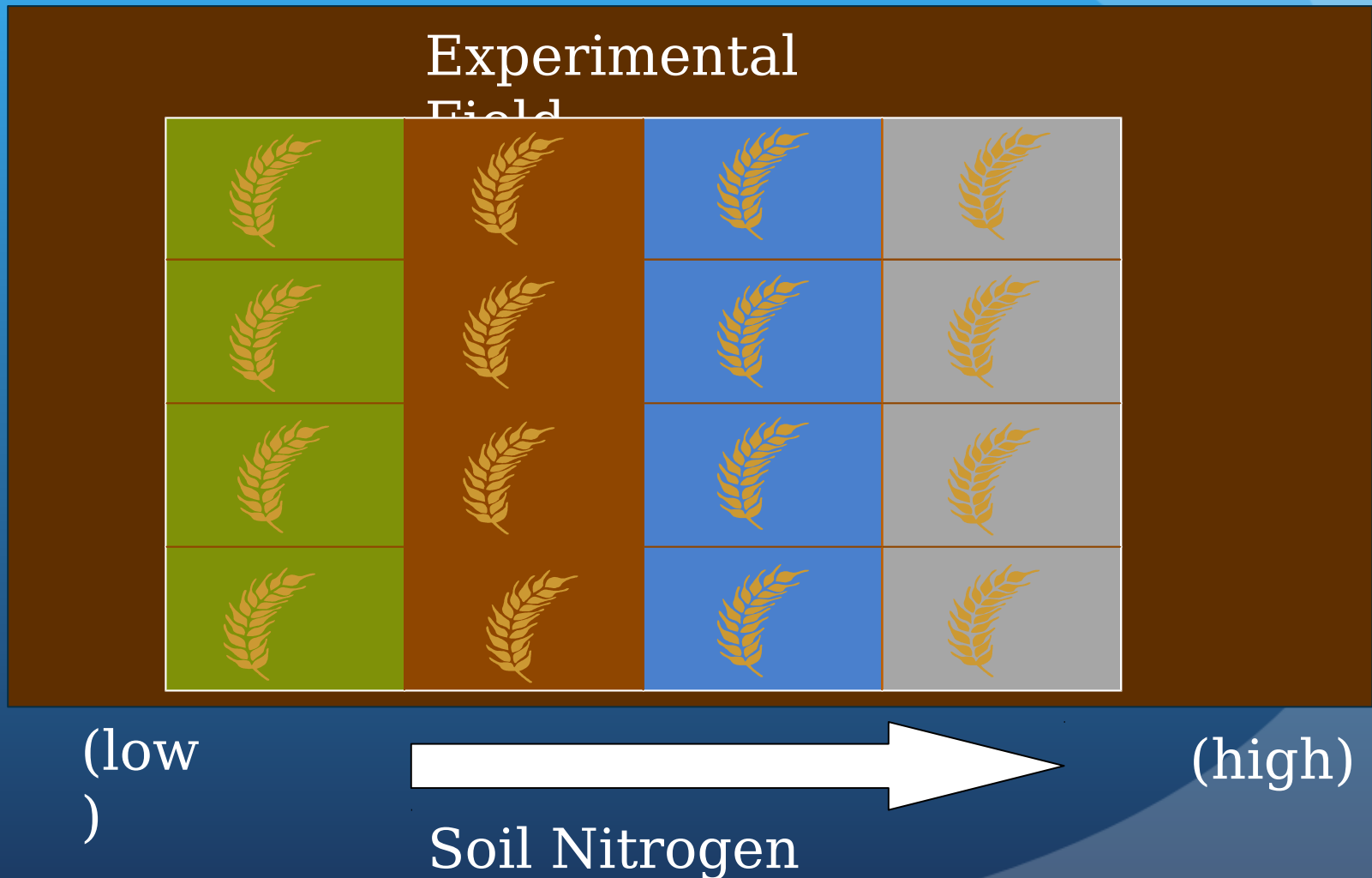
- Blocking is done when the **experimental units** (plot, animal, potted plant, etc.) **are not uniform** due to some naturally occurring differences prior to applying a treatment.

## Examples??

- A sloping agricultural field where plots are the experimental unit and the upper, middle, and lower hill-slope are the blocks.
- Multiple cattle ranches where the cows are the experimental unit, and the block is the cattle ranch.
- **Blocks can account for differences that may affect the dependent variable, and therefore reduce unexplained variation (error), which improves our ability to detect differences among treatments!**

What statistical model predicts yield for an individual plot in this experiment?

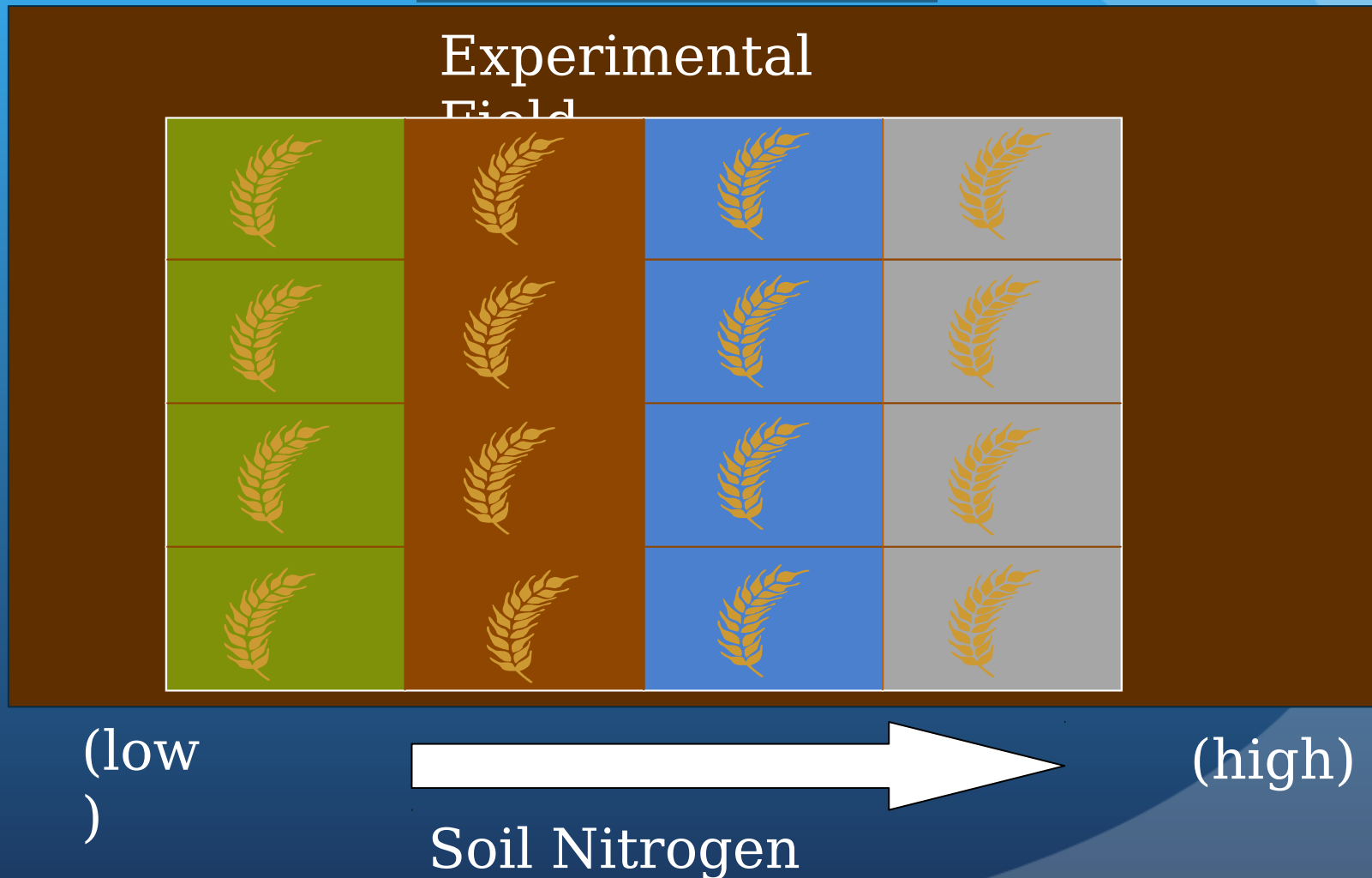
$Y_{ij} = ?$  where  $i$  is  $i^{\text{th}}$  treatment, and  $j$  is the  $j^{\text{th}}$  block





What statistical model predicts yield for an individual plot in this experiment?

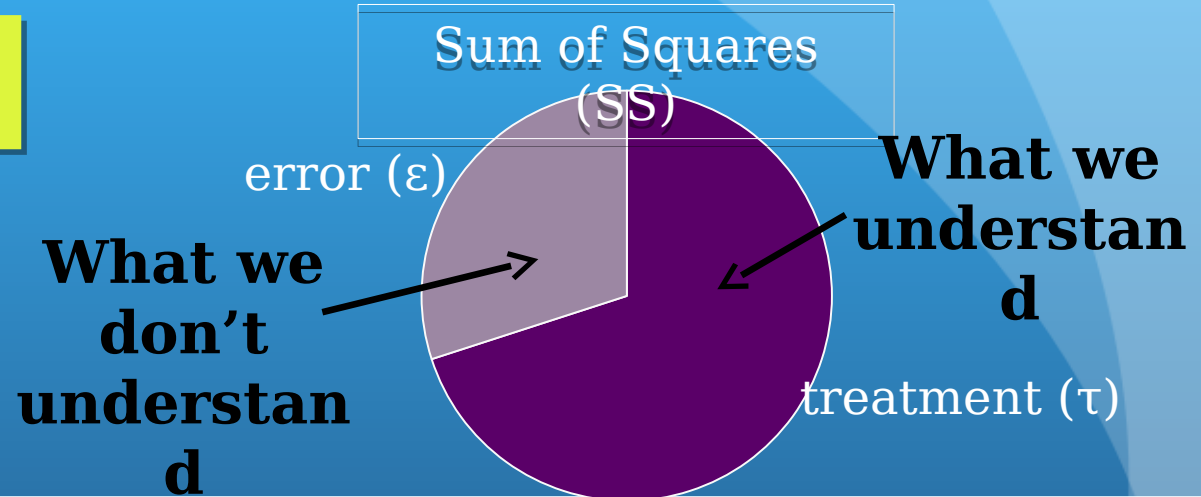
$$Y_{ij} = \mu + \tau_i + \nu_j + \varepsilon_{ij}$$



# Recall

## Complete Randomized Design (CRD):

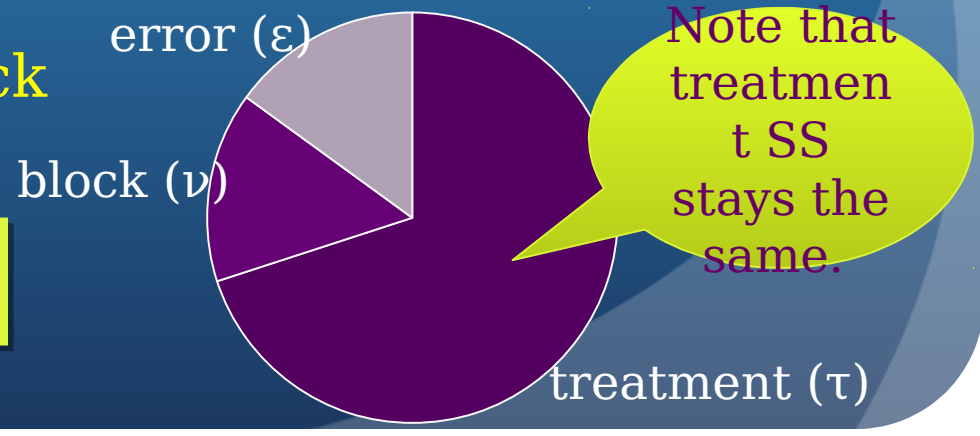
$$Y_{ij} = \mu + \tau_i + \epsilon_{ij}$$



If we can explain more of the variation than 'treatment' alone, add it to the experimental design and to the model  $\mu$  lower the error  $\mu$  increase power!

## Randomized Complete Block Design (RCBD):

$$Y_{ij} = \mu + \tau_i + \nu_j + \epsilon_{ij}$$



# Exercise 2

In groups, analyze the 'wheat variety' dataset using the AIP App:

1. Determine if there is an effect of wheat variety on yield, and if so, what are the relative differences among the wheat varieties? Use Protected LSD.
2. Write main conclusions using biological and statistical interpretations.

- [Finished up to this point...
- Will add screenshots once app is ready: loading RCBD wheat data, running ANOVA, LSD mean separation, display results]

# Other resources for statistical software:

- Free (coding): R
  - Download for Windows:  
<https://cran.r-project.org/bin/windows/base/>
  - Download for Mac:  
<https://cran.r-project.org/bin/macosx/>
- Cost (GUIs): JMP, Sigma Plot (graphics)