

What Do We Mean When We Say ‘p-value’?

**Peter Claussen
Gylling Data Management**

On Farm Strip Trial

- Suppose we have an agronomic treatment that we expect will improve yield, relative to untreated crops.
- We provide a farmer with product to be applied to part of his fields, and we hope to measure crop yield over treated and untreated areas of the field.
- At a minimum, we hope to determine if the treated crop out-yielded the untreated crop. Ideally, we hope to state that the product increases yield by τ bushels per acre.

On Farm Strip Trial

- By 'we', I mean :

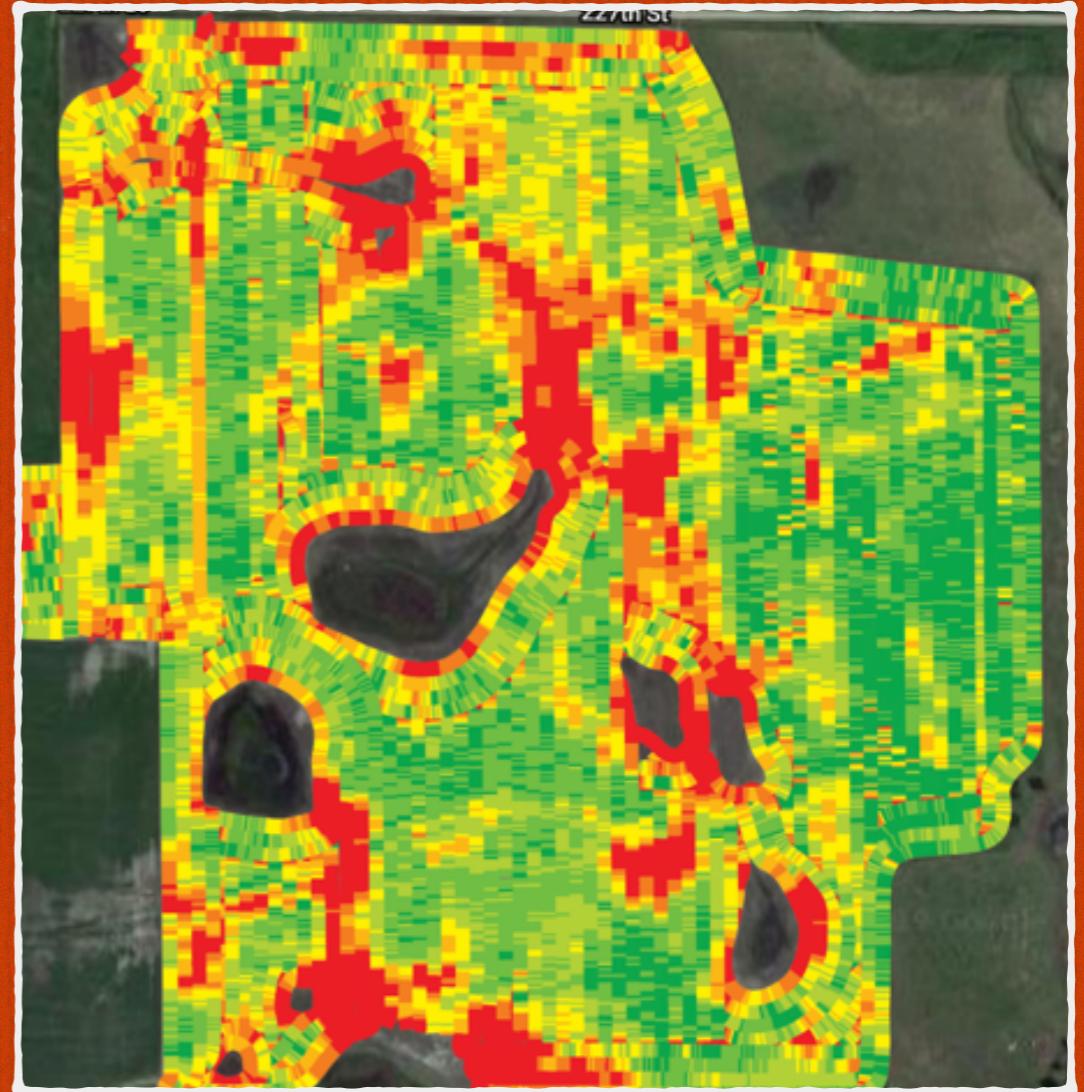


SOUTH DAKOTA SOYBEAN
ON-FARM RESEARCH
PROGRAM



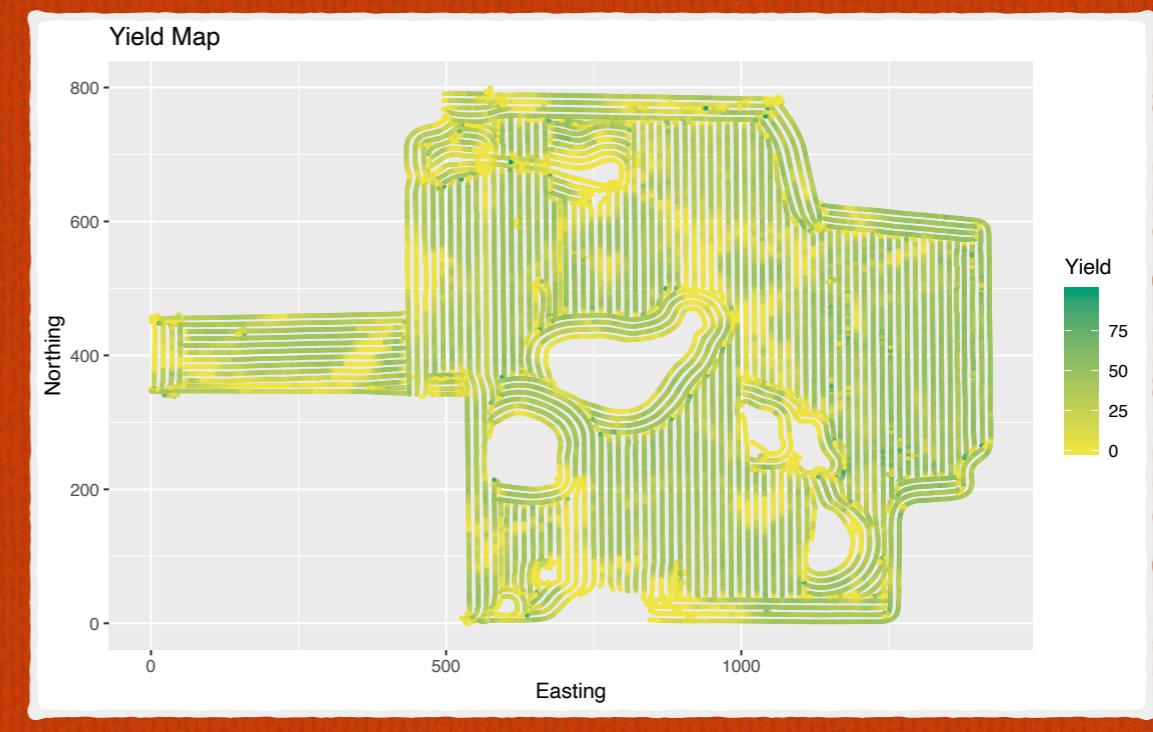
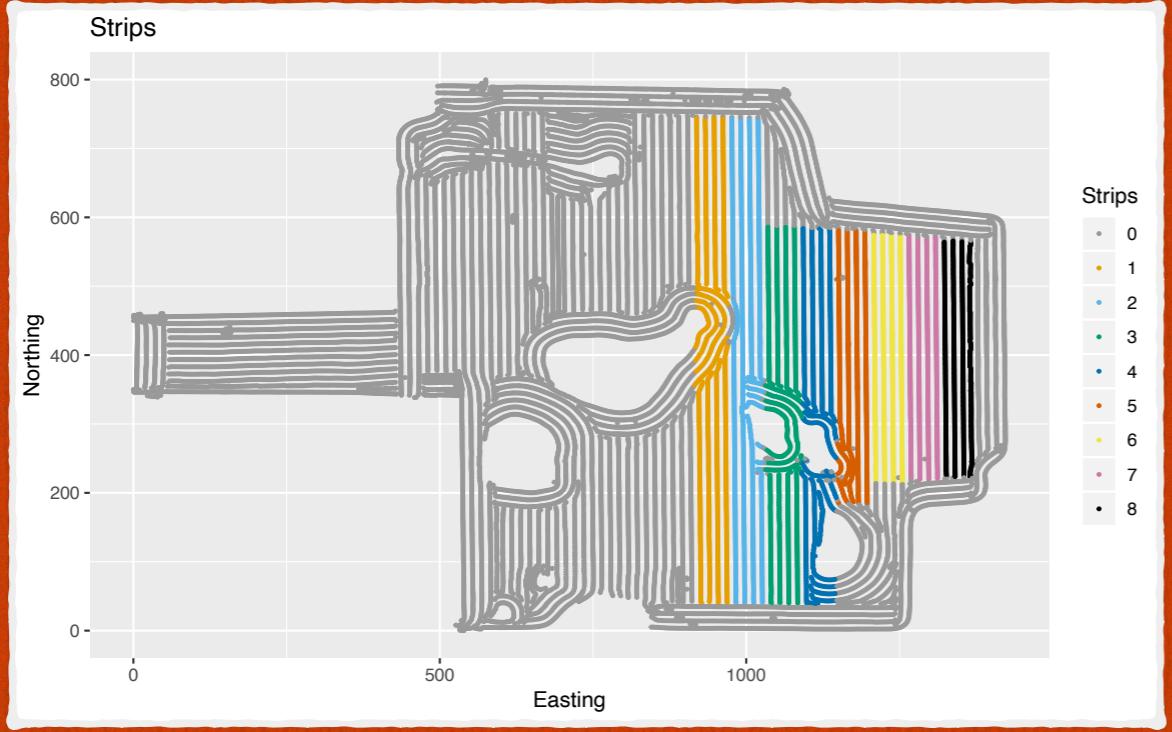
- By 'farmer', I mean :





Spray and Yield Maps

From the data providers



Spray and Yield Maps

As analyzed
Strips 2,4,6 and 8 are sprayed

Naive Analysis

- Each treated strip covers 4 harvest passes.**
- Pair each treated strip with an adjacent strip of untreated harvest passes.**
- Calculate mean yields for all yield data points in each strip.**
- Analyze as if each strip were a plot in a small-plot experiment (RCB, 2 treatments, 4 replicates)**
- Write a model**

$$y_{ij} = \mu + \rho_j + \tau_i + e_{ij}$$

$\tau_0 = \text{unsprayed}$, $\tau_1 = \text{sprayed}$, $\rho_1 \dots \rho_4 = \text{pairs}$, $e_{ij} \sim \mathcal{N}(0, \sigma^2)$

- State a null hypothesis**

$$H_0 : \tau_1 = \tau_2 = 0$$

Naive Analysis

(Naive in the sense of ignoring aspects of experimental design)

- Each treated strip covers 4 harvest passes.**
- Pair each treated strip with an adjacent strip of untreated harvest passes.**
- Calculate mean yields for all yield data points in each strip.**
- Analyze as if each strip were a plot in a small-plot experiment (RCB, 2 treatments, 4 replicates)**
- Write a model**

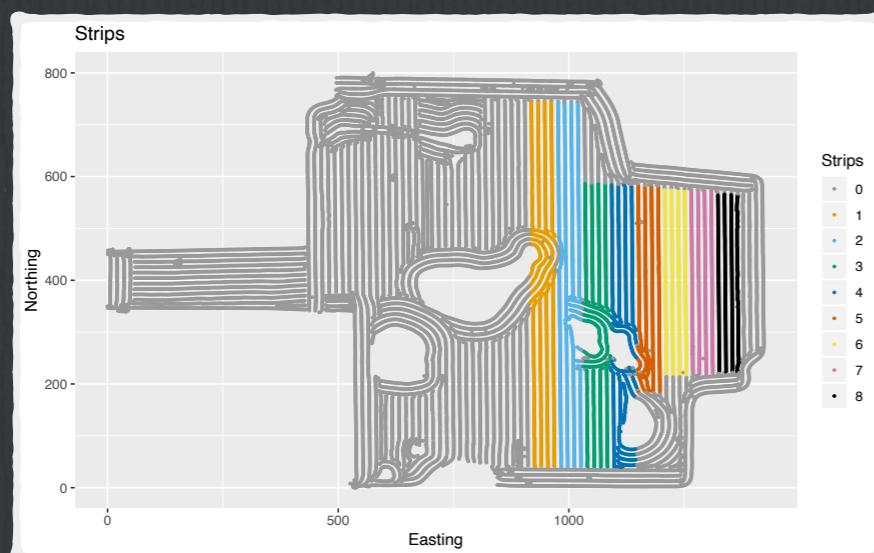
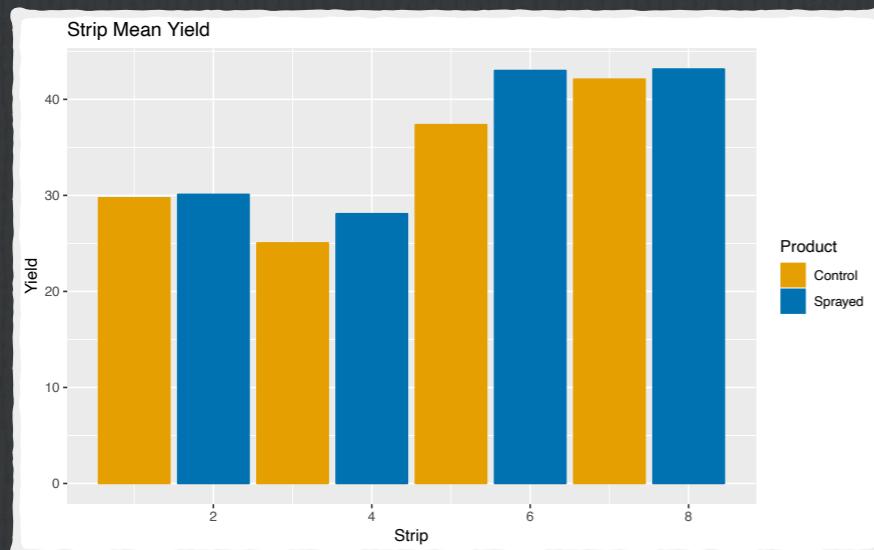
$$y_{ij} = \mu + \rho_j + \tau_i + e_{ij}$$

τ_0 = **unsprayed**, τ_1 = **sprayed**, $\rho_1 \dots \rho_4$ = **pairs**, $e_{ij} \sim \mathcal{N}(0, \sigma^2)$

- State a null hypothesis**

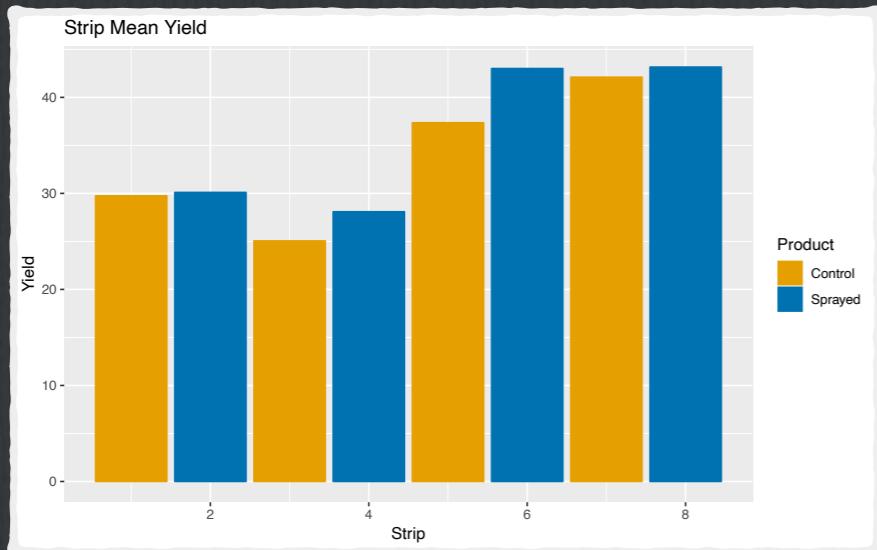
$$H_0 : \tau_1 = \tau_2 = 0$$

Results



- > wilcox.test(Yield ~ Product, paired=TRUE, ...)
Wilcoxon signed rank test
data: Yield by Product
V = 0, p-value = 0.125
 - > t.test(Yield ~ Product, paired=TRUE, ...)
Paired t-test
data: Yield by Product
t = -2.1319, df = 3, p-value = 0.1228
sample estimates:
mean in group Control mean in group Sprayed
33.56637 36.09020
 - > friedman.test(Yield ~ Block | Product, ...)
Friedman rank sum test
data: Yield and Block and Product
Friedman chi-squared = 6, df = 3, **p-value = 0.1116**
 - > anova(Yield ~ Block + Product, ...)
Analysis of Variance Table
Response: Yield
- | | Df | Sum Sq | Mean Sq | F value | Pr(>F) |
|-----------|----|--------|---------|---------|-----------------|
| Block | 3 | 363.52 | 121.173 | 43.230 | 0.005732 ** |
| Product | 1 | 12.74 | 12.739 | 4.545 | 0.122791 |
| Residuals | 3 | 8.41 | 2.803 | | |

We fail to reject the null hypothesis

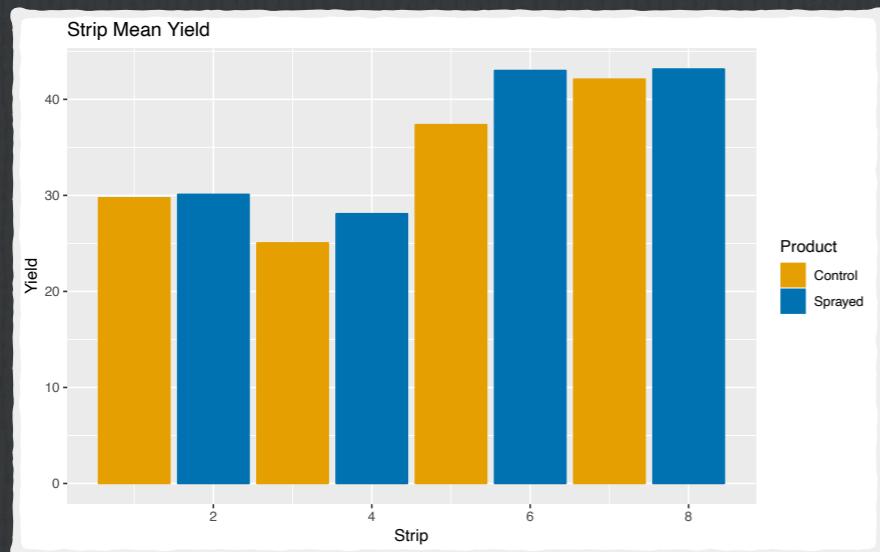


- Even though :
 - For each pair of strips, the treated strip had greater yield
 - Average yield in the treated strips was greater (36.1 vs 33.6 bu/acre - ~ 7.5% increase)

- > wilcox.test(Yield ~ Product, paired=TRUE, ...)
Wilcoxon signed rank test
data: Yield by Product
V = 0, p-value = 0.125
- > t.test(Yield ~ Product, paired=TRUE, ...)
Paired t-test
data: Yield by Product
t = -2.1319, df = 3, p-value = 0.1228
sample estimates:
mean in group Control mean in group Sprayed
33.56637 36.09020
- > friedman.test(Yield ~ Block | Product, ...)
Friedman rank sum test
data: Yield and Block and Product
Friedman chi-squared = 6, df = 3, **p-value = 0.1116**
- > anova(Yield ~ Block + Product, ...)
Analysis of Variance Table
Response: Yield

	Df	Sum Sq	Mean Sq	F value	Pr(>F)
Block	3	363.52	121.173	43.230	0.005732 **
Product	1	12.74	12.739	4.545	0.122791
Residuals	3	8.41	2.803		

We fail to reject the null hypothesis



- This result is disappointing, but may be the best outcome for a designed based analysis of these data.

- > wilcox.test(Yield ~ Product, paired=TRUE, ...)
Wilcoxon signed rank test
data: Yield by Product
V = 0, p-value = 0.125
- > t.test(Yield ~ Product, paired=TRUE, ...)
Paired t-test
data: Yield by Product
t = -2.1319, df = 3, p-value = 0.1228
sample estimates:
mean in group Control mean in group Sprayed
33.56637 36.09020
- > friedman.test(Yield ~ Block | Product, ...)
Friedman rank sum test
data: Yield and Block and Product
Friedman chi-squared = 6, df = 3, **p-value = 0.1116**
- > anova(Yield ~ Block + Product, ...)
Analysis of Variance Table
Response: Yield

	Df	Sum Sq	Mean Sq	F value	Pr(>F)
Block	3	363.52	121.173	43.230	0.005732 **
Product	1	12.74	12.739	4.545	0.122791
Residuals	3	8.41	2.803		

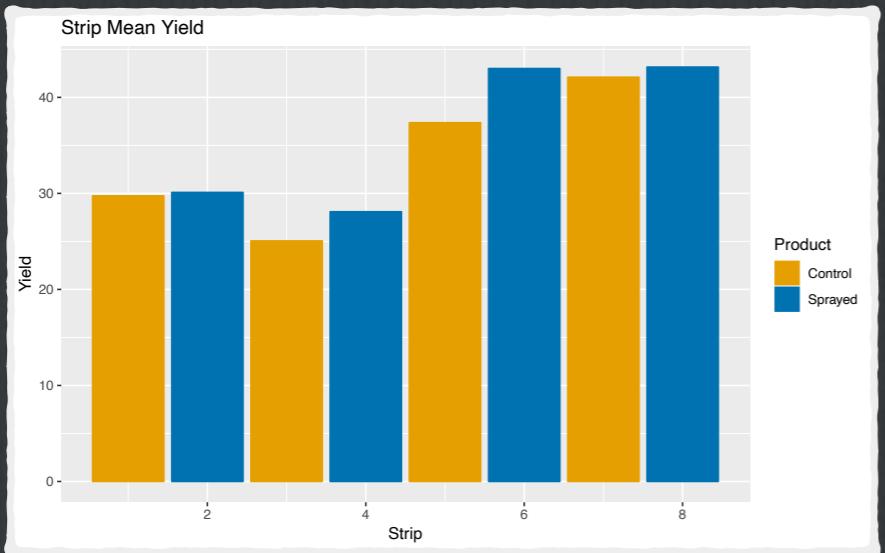
We fail to reject the null hypothesis



- Do I tell him he wasted his time?

- > wilcox.test(Yield ~ Product, paired=TRUE, ...)
Wilcoxon signed rank test
data: Yield by Product
V = 0, p-value = 0.125
 - > t.test(Yield ~ Product, paired=TRUE, ...)
Paired t-test
data: Yield by Product
t = -2.1319, df = 3, p-value = 0.1228
 - > friedman.test(Yield ~ Block | Product, ...)
Friedman rank sum test
data: Yield and Block and Product
Friedman chi-squared = 6, df = 3, **p-value = 0.1116**
 - > anova(Yield ~ Block + Product, ...)
Analysis of Variance Table
Response: Yield
- | | Df | Sum Sq | Mean Sq | F value | Pr (>F) |
|-----------|----|--------|---------|---------|-----------------|
| Block | 3 | 363.52 | 121.173 | 43.230 | 0.005732 ** |
| Product | 1 | 12.74 | 12.739 | 4.545 | 0.122791 |
| Residuals | 3 | 8.41 | 2.803 | | |

Clean the data?



Perhaps our problem is that we not comparing similar 'experimental units'?



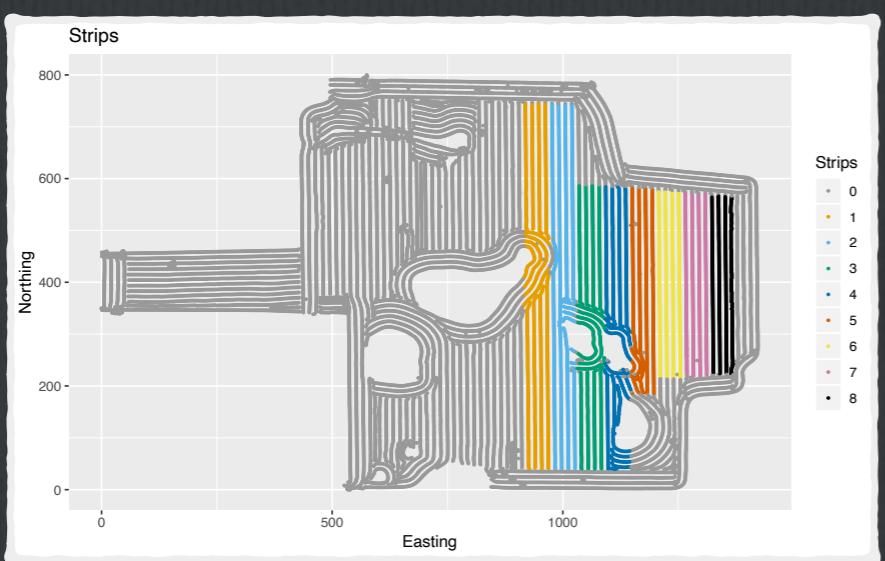
Only 2 strips run the entire length of the field.



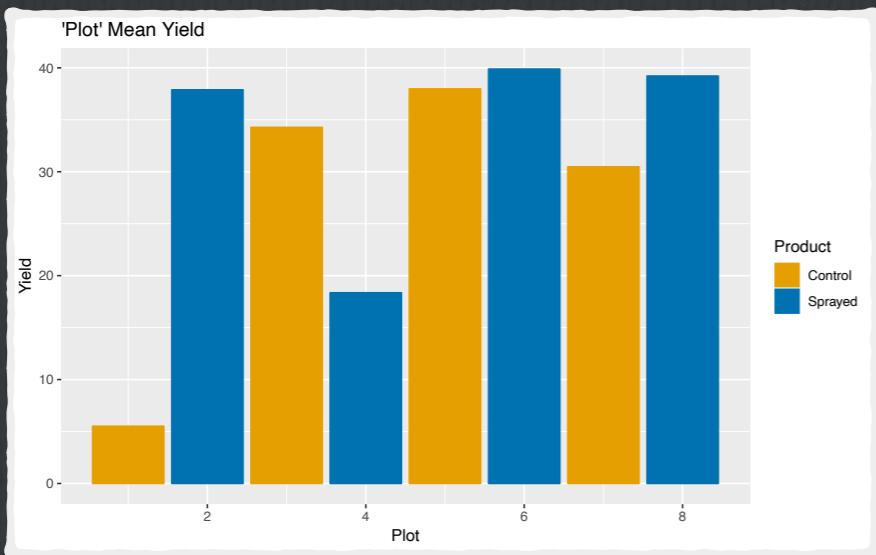
Only 3 of the strips were planted and harvested in continuous passes for the entire length of the strip - the others avoid wet areas



Should we analyze just a portion of the field with uniformly-sized treated and untreated areas?

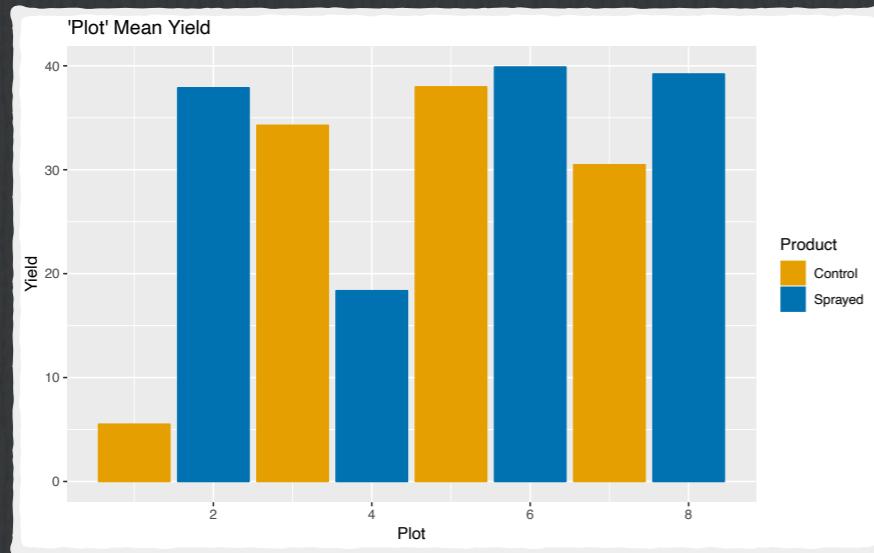


Equal sized ‘plots’



- > wilcox.test(Yield ~ Product, paired=TRUE, ...)
Wilcoxon signed rank test
data: Yield by Product
V = 3, **p-value = 0.625**
 - > t.test(Yield ~ Product, paired=TRUE, ...)
Paired t-test
data: Yield by Product
t = -0.67812, df = 3, **p-value = 0.5463**
mean in group Control mean in group Sprayed
27.05887 33.83388
 - > friedman.test(Yield ~ Block | Product, ...)
Friedman rank sum test
data: Yield and Block and Product
Friedman chi-squared = 4.2, df = 3, **p-value = 0.2407**
 - > anova(Yield ~ Block + Product, ...)
Analysis of Variance Table
Response: Yield
- | | Df | Sum Sq | Mean Sq | F value | Pr (> F) |
|-----------|----|--------|---------|---------|---------------|
| Block | 3 | 369.77 | 123.258 | 0.6174 | 0.6492 |
| Product | 1 | 91.80 | 91.802 | 0.4599 | 0.5463 |
| Residuals | 3 | 598.90 | 199.633 | | |

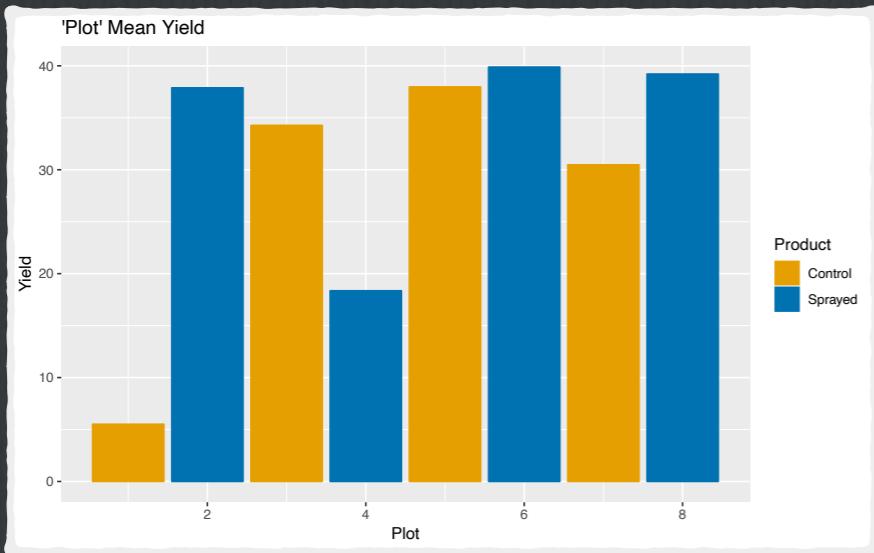
Equal sized ‘plots’



- Even though :
 - Difference in average yield was more extreme (33.8 vs 27.0 bu/acre)

- > wilcox.test(Yield ~ Product, paired=TRUE, ...)
Wilcoxon signed rank test
data: Yield by Product
V = 3, **p-value = 0.625**
 - > t.test(Yield ~ Product, paired=TRUE, ...)
Paired t-test
data: Yield by Product
t = -0.67812, df = 3, **p-value = 0.5463**
mean in group Control mean in group Sprayed
27.05887 33.83388
 - > friedman.test(Yield ~ Block | Product, ...)
Friedman rank sum test
data: Yield and Block and Product
Friedman chi-squared = 4.2, df = 3, **p-value = 0.2407**
 - > anova(Yield ~ Block + Product, ...)
Analysis of Variance Table
Response: Yield
- | | Df | Sum Sq | Mean Sq | F value | Pr (> F) |
|-----------|----|--------|---------|---------|---------------|
| Block | 3 | 369.77 | 123.258 | 0.6174 | 0.6492 |
| Product | 1 | 91.80 | 91.802 | 0.4599 | 0.5463 |
| Residuals | 3 | 598.90 | 199.633 | | |

Equal sized ‘plots’



- The results of AOV suggest our assumptions of the statistical model are not supported
- Specifically,
 $e_{ij} \sim \mathcal{N}(0, \sigma^2)$

- > wilcox.test(Yield ~ Product, paired=TRUE, ...)
Wilcoxon signed rank test
data: Yield by Product
V = 3, **p-value = 0.625**
 - > t.test(Yield ~ Product, paired=TRUE, ...)
Paired t-test
data: Yield by Product
t = -0.67812, df = 3, **p-value = 0.5463**
mean in group Control mean in group Sprayed
27.05887 33.83388
 - > friedman.test(Yield ~ Block | Product, ...)
Friedman rank sum test
data: Yield and Block and Product
Friedman chi-squared = 4.2, df = 3, **p-value = 0.2407**
 - > anova(Yield ~ Block + Product, ...)
Analysis of Variance Table
Response: Yield
- | | Df | Sum Sq | Mean Sq | F value | Pr (>F) |
|-----------|----|--------|---------|---------|---------------|
| Block | 3 | 369.77 | 123.258 | 0.6174 | 0.6492 |
| Product | 1 | 91.80 | 91.802 | 0.4599 | 0.5463 |
| Residuals | 3 | 598.90 | 199.633 | | |

For our present purpose, however, it is only necessary to recognize that, whatever degree of care and experimental skill is expended in equalizing the conditions, other than the one under test, which are liable to affect the result, *this equalization must always be to a greater or less extent incomplete, and in many important cases will certainly be grossly defective.*

-R. A Fisher "The Design of Experiments, 9. Randomization; the Physical Basis of the Validity of the Test".

Theoretical Randomizations



If we assume a standard RCB with 4 blocks and 2 treatments, we have 2^4 possible arrangements

from

101 1	102 2	201 1	202 2	301 1	302 2	401 1	402 2
----------	----------	----------	----------	----------	----------	----------	----------

to

101 2	102 1	201 2	202 1	301 2	302 1	401 2	402 1
----------	----------	----------	----------	----------	----------	----------	----------

Theoretical Randomizations



□ ... and all possible permutations in between

101 1	102 2	201 2	202 1	301 1	302 2	401 1	402 2
----------	----------	----------	----------	----------	----------	----------	----------

101 1	102 2	201 1	202 2	301 2	302 1	401 1	402 2
----------	----------	----------	----------	----------	----------	----------	----------

101 1	102 2	201 2	202 1	301 1	302 2	401 2	402 1
----------	----------	----------	----------	----------	----------	----------	----------

...

101 2	102 1	201 2	202 1	301 1	302 2	401 2	402 1
----------	----------	----------	----------	----------	----------	----------	----------

...

Theoretical Randomizations



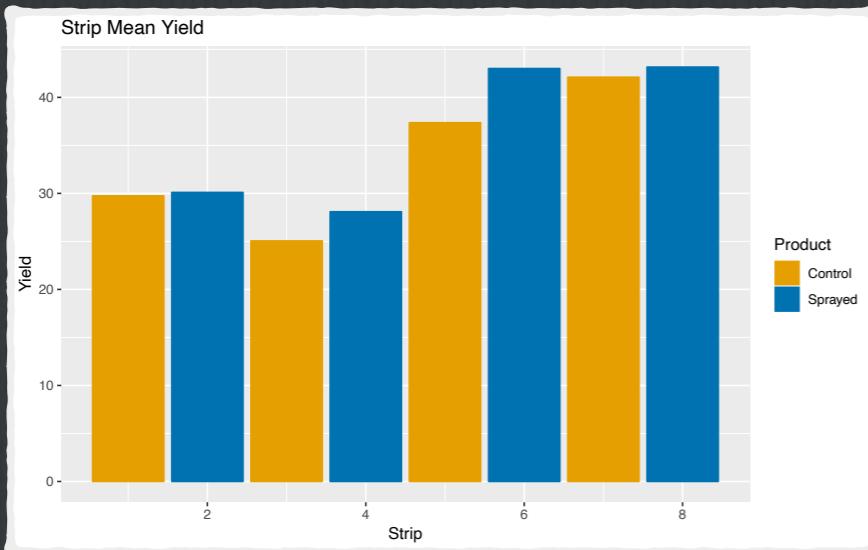
- Under the null hypothesis**
- 1 arrangement would have assigned treatment to the 4 most fertile strips (per pair) (1/16)**
- 4 arrangements would have assigned treatment to 3 of the most fertile strips (per pair) (3/16)**
- 6 arrangements would have assigned treatment to 2 of the most fertile strips (per pair) (6/16)**
- 4 arrangements would have assigned treatment to 1 of the most fertile strips (per pair) (3/16)**
- 1 arrangement would have assigned treatment to none most fertile strips (per pair) (1/16)**

Theoretical Randomizations



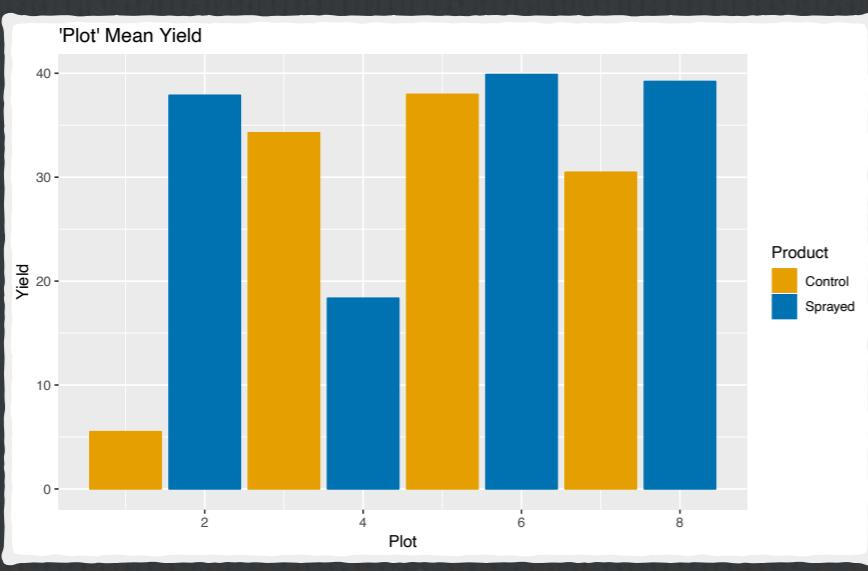
- Under the null hypothesis**
- The most extreme result (treated yield higher on 4 of 4 pairs, or untreated yield higher on 4 of 4 pairs)** could have happened in 2 out of 16 = 0.125 arrangements.
- The next most extreme result (treated yield higher on 3 of 4 pairs, or untreated yield higher on 3 of 4 pairs)** could have happened in 8 out of 16 arrangements; only 6 of 16 would have been less extreme ($p = 1 - 6/16 = 0.625$)

Results



- > wilcox.test(Yield ~ Product, paired=TRUE,...)
Wilcoxon signed rank test
data: Yield by Product
V = 0, p-value = 0.125
- > t.test(Yield ~ Product, paired=TRUE, ...)
Paired t-test
data: Yield by Product
t = -2.1319, df = 3, p-value = 0.1228
- > friedman.test(Yield ~ Block | Product, ...)
Friedman rank sum test
data: Yield and Block and Product
Friedman chi-squared = 6, df = 3, **p-value = 0.1116**
- > anova(Yield ~ Block + Product, ...)
Analysis of Variance Table
Response: Yield

	Df	Sum Sq	Mean Sq	F value	Pr(>F)
Block	3	363.52	121.173	43.230	0.005732 **
Product	1	12.74	12.739	4.545	0.122791
Residuals	3	8.41	2.803		



- > wilcox.test(Yield ~ Product, paired=TRUE,...)
Wilcoxon signed rank test
data: Yield by Product
V = 3, p-value = 0.625
- > t.test(Yield ~ Product, paired=TRUE, ...)
Paired t-test
data: Yield by Product
t = -0.67812, df = 3, p-value = 0.5463
- > friedman.test(Yield ~ Block | Product, ...)
Friedman rank sum test
data: Yield and Block and Product
Friedman chi-squared = 4.2, df = 3, **p-value = 0.2407**
- > anova(Yield ~ Block + Product, ...)
Analysis of Variance Table
Response: Yield

	Df	Sum Sq	Mean Sq	F value	Pr(>F)
Block	3	369.77	123.258	0.6174	0.6492
Product	1	91.80	91.802	0.4599	0.5463
Residuals	3	598.90	199.633		

Experiment as Executed

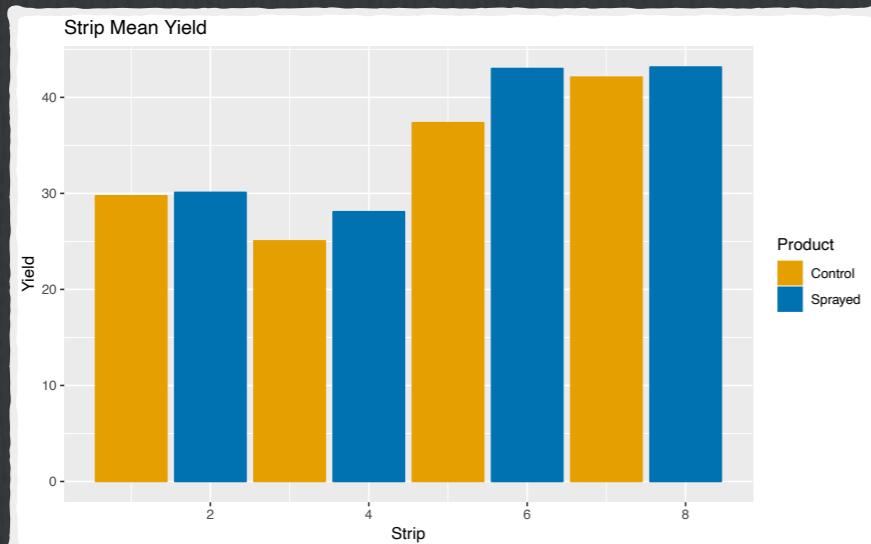


- In the event, there was only one possible arrangement
- Enter the field at the north gate, start spraying.
- Requirements of the null hypothesis test have not been met!

Apart, therefore, from the avoidable error of the experimenter himself introducing with his test treatments, or subsequently, other differences in treatment, the effects of which the experiment is not intended to study, *it may be said that the simple precaution of randomization will suffice to guarantee the validity of the test of significance*, by which the result of the experiment is to be judged.

-R. A Fisher "The Design of Experiments, 10. The Effectiveness of Randomization".

Modeling a Trend by Position



- Because we did not randomize over blocks, we instead model a yield trend by position from the East edge of the field

$$H_1 : y_{ij} = \beta_0 + \beta_1 E_{ij} + \tau_i + e_{ij}$$

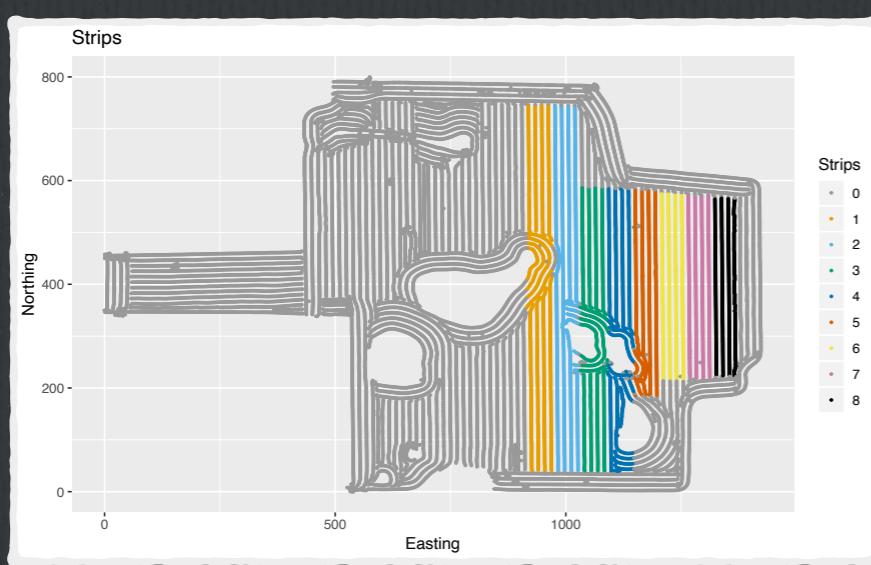
- Increasing the order of the polynomial gives us a new hypotheses.

$$H_2 : y_{ij} = \beta_0 + \beta_1 E_{ij} + \beta_2 E_{ij}^2 + \tau_i + e_{ij}$$

$$H_3 : y_{ij} = \beta_0 + \beta_1 E_{ij} + \dots + \beta_3 E_{ij}^3 + \tau_i + e_{ij}$$

$$H_4 : y_{ij} = \beta_0 + \beta_1 E_{ij} + \dots + \beta_4 E_{ij}^4 + \tau_i + e_{ij}$$

- How do we select among competing hypothesis?



In a problem of estimation, we start with a knowledge of the mathematical form of the population sampled, but without knowledge of the values of one or more of the parameters which enter into this form ...

The *probability* of occurrence of our entire sample is therefore expressible as a function of these unknown parameters, and the *likelihood* is defined merely as a function of these parameters proportional to this probability.

-Johnny Appleseed

Probability vs Likelihood

- Probability and likelihood are mathematically similar, but differ in the choice of free parameters:

- (Normal) Probability Density Function

$$f(y | \mu, \sigma^2) = \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{(y-\mu)^2}{2\sigma^2}}$$

- (Normal) Likelihood Function

$$\mathcal{L}(\mu, \sigma^2 | y) = \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{(y-\mu)^2}{2\sigma^2}}$$

Probability vs Likelihood

- We we talk about p-values, we use the integral of the probability density function; with likelihood, we refer to the value of the probability function for a select value.

- (Normal) Cumulative Density Function

$$F_Y(y | \mu, \sigma^2) = \frac{1}{2} \left[1 + \operatorname{erf}\left(\frac{y - \mu}{\sqrt{2}\sigma}\right) \right] = P(Y \leq y)$$

$$\operatorname{erf}(y) = \frac{2}{\sqrt{\pi}} \int_0^y e^{-t^2} dt$$

- (Normal) Likelihood Function

$$\mathcal{L}(\mu, \sigma^2 | y) = \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{(y-\mu)^2}{2\sigma^2}} = P_{\mu, \sigma^2}(Y = y)$$

Probability vs Likelihood

- We can extend the likelihood concept from single to multiple observations.

- Given the probability of a single observation,

$$f(y | \mu, \sigma^2) = \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{(y-\mu)^2}{2\sigma^2}}$$

- we calculate likelihood from a series of observations

$$\mathcal{L}(\mu, \sigma^2 | y_1, \dots, y_n) = \prod_{i=1}^n \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{(y_i-\mu)^2}{2\sigma^2}} = \left(\sqrt{2\pi\sigma^2}\right)^{-n} \exp\left\{-\frac{1}{2\sigma^2} \sum (y_i - \mu)^2\right\}$$

Maximum Likelihood

- Find parameters such that the likelihood functions achieves a maximum value, e.g. the values

$$\hat{\mu}_i = \frac{\sum_{j=1}^n y_{ij}}{N_i}, \hat{\sigma}^2 = \frac{\sum_{j=1}^n (y_{ij} - \hat{\mu}_i)^2}{n - 1}$$

are maximum likelihood estimates for $\mathcal{L}(\mu, \sigma^2 | y)$

- The value of $\mathcal{L}(\dots | y)$ with respect to the maximum likelihood estimates is a measure of the correctness of a statistical model for data y

Likelihood Ratio

- Given likelihood estimates for two alternative hypothesis:

$$H_1 | \mu = \mu_1 : \mathcal{L}_1(\mu_1, \sigma^2 | y_1, \dots, y_n)$$

$$H_2 | \mu = \mu_2 : \mathcal{L}_1(\mu_2, \sigma^2 | y_1, \dots, y_n)$$

- we can compute a relative likelihood

$$\frac{\mathcal{L}_2(\mu_2, \sigma^2 | y_1, \dots, y_n)}{\mathcal{L}_1(\mu_1, \sigma^2 | y_1, \dots, y_n)}$$

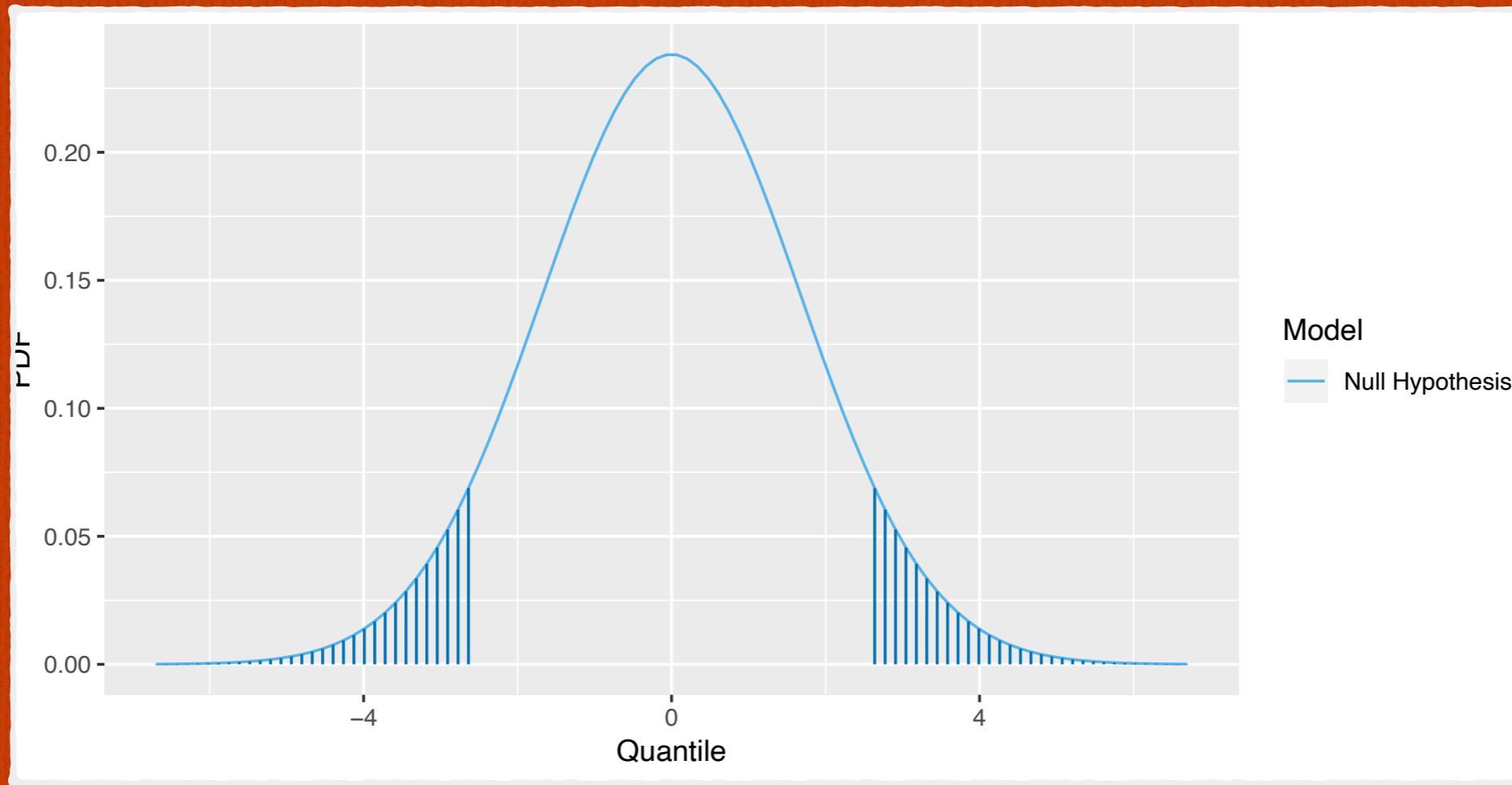
Likelihood Ratio

- We typically use log-likelihoods

$$\ell_1(\mu_1, \sigma^2 | y_1, \dots, y_n) = \log \left\{ \mathcal{L}_1(\mu_1, \sigma^2 | y_1, \dots, y_n) \right\}$$

- and

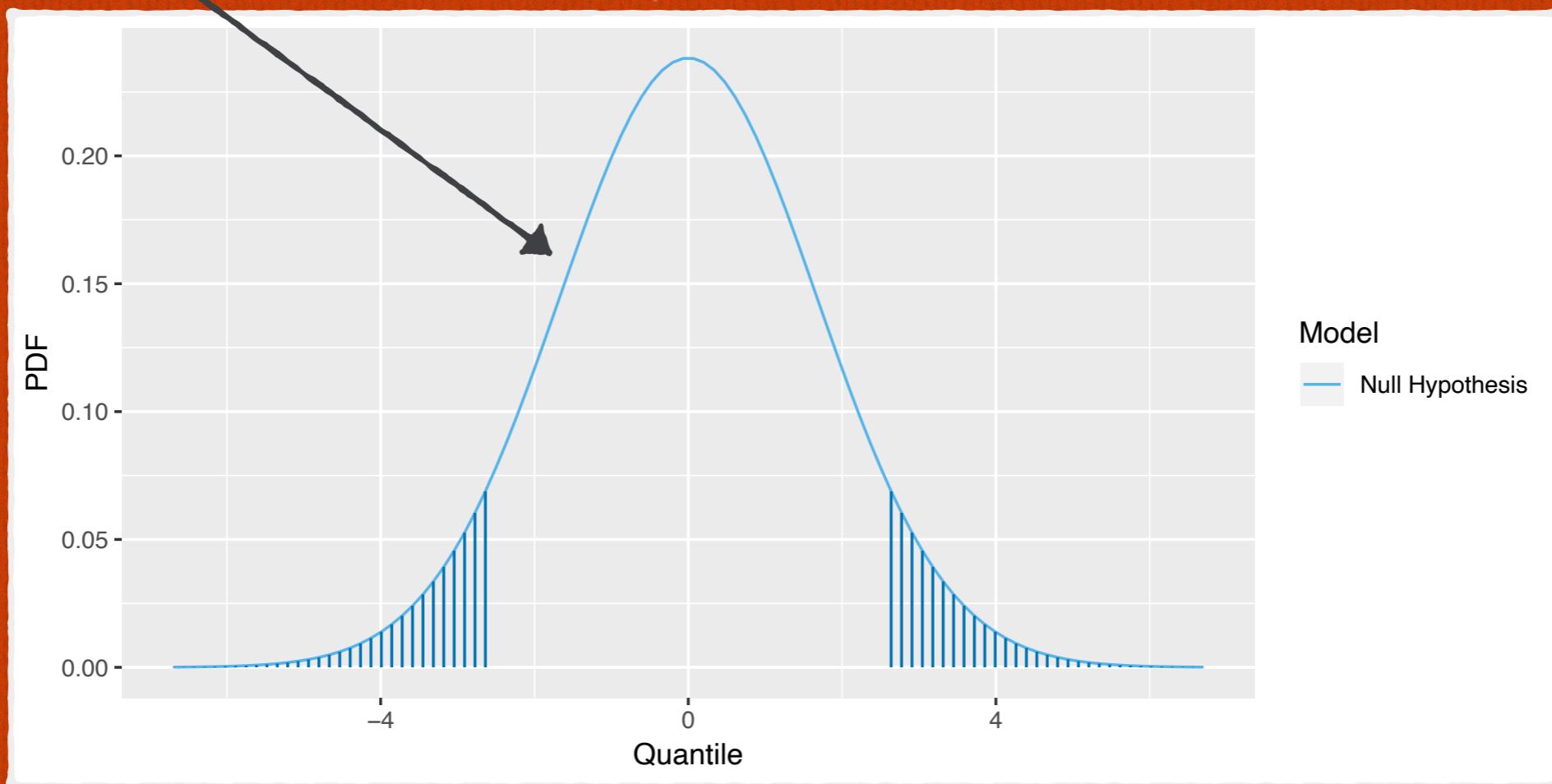
$$\frac{\mathcal{L}_2(\mu_2, \sigma^2 | y_1, \dots, y_n)}{\mathcal{L}_1(\mu_1, \sigma^2 | y_1, \dots, y_n)} = \ell_2(\mu_2, \sigma^2 | y_1, \dots, y_n) - \ell_1(\mu_1, \sigma^2 | y_1, \dots, y_n)$$



Probability vs Likelihood

Using maximum likelihood estimates from our 'naive' analysis,
 $\hat{\mu} = 2.52, \hat{\sigma}^2 = 2.83$

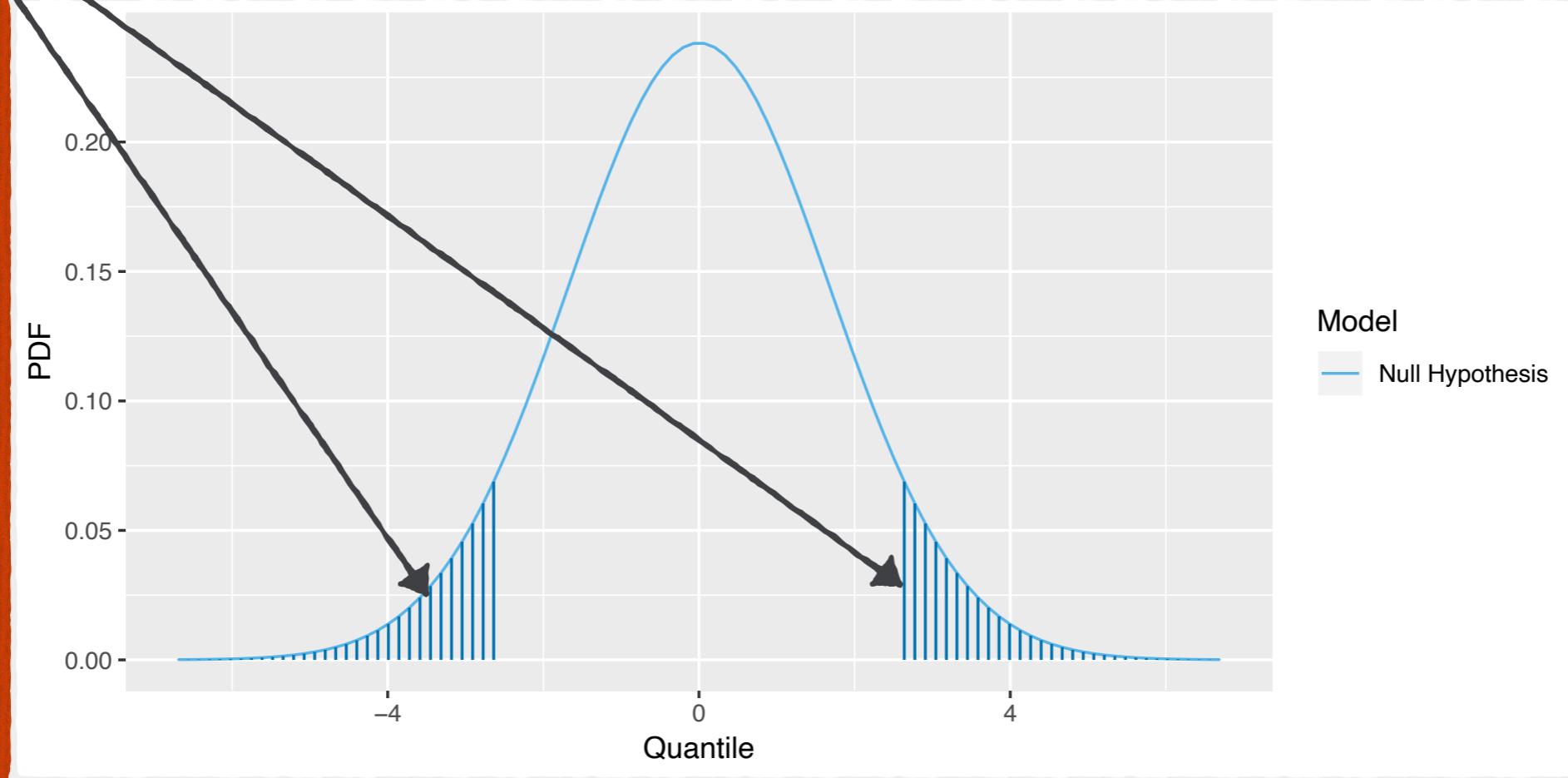
$$f(y | \mu_0 = 0, \sigma^2 = \hat{\sigma}^2)$$



Probability vs Likelihood

The null hypothesis asserts that μ is (axiomatically) 0, but σ^2 can take its maximum likelihood estimate

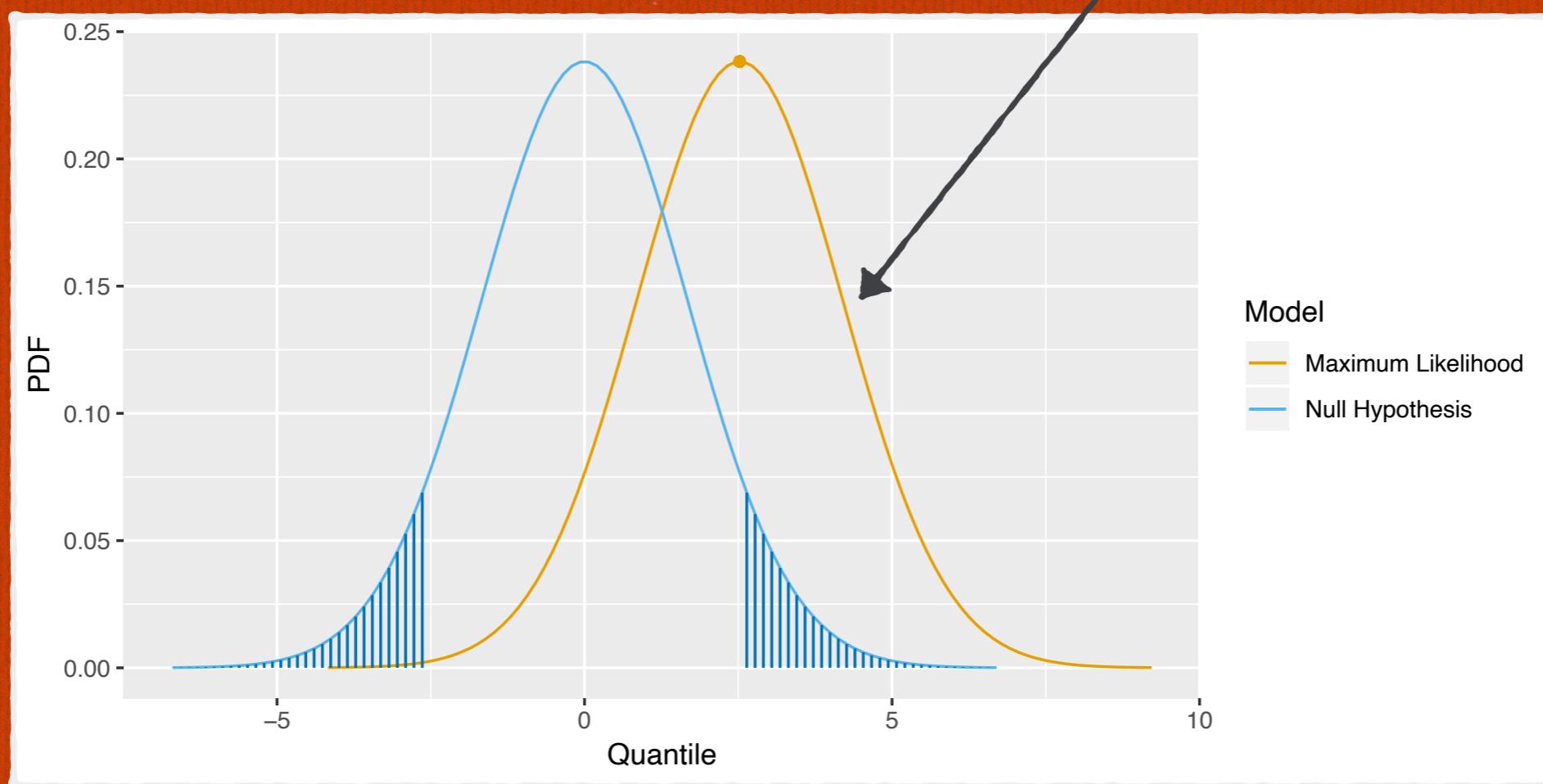
$$p(|y| > \hat{\mu} | \mu_0 = 0, \sigma^2 = \hat{\sigma}^2)$$



Probability vs Likelihood

We determine a p-value by integrating the area corresponding to values extreme than the maximum likelihood estimate

$$f(y | \mu_1 = \hat{\mu}, \sigma^2 = \hat{\sigma}^2)$$

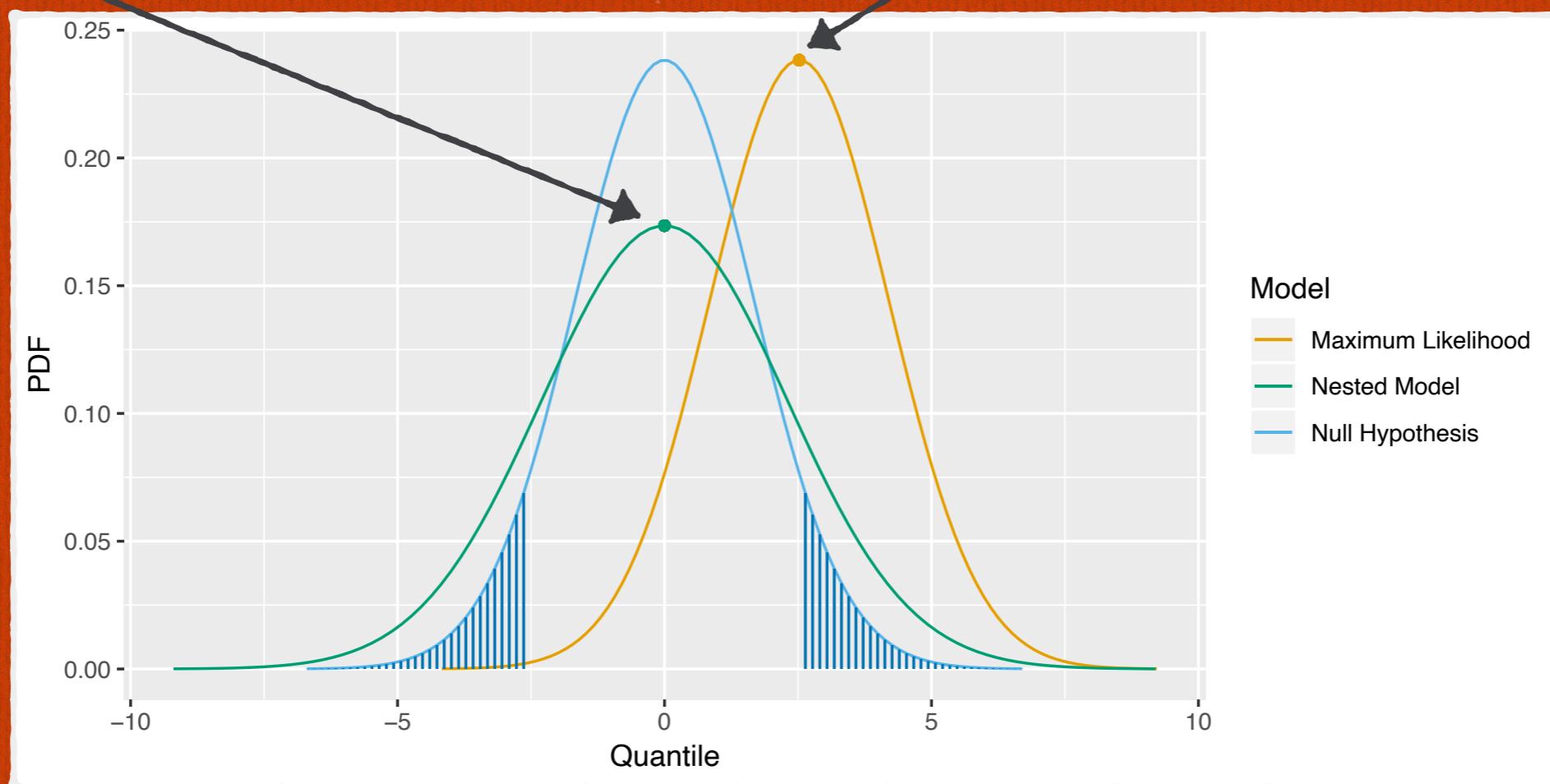


Probability vs Likelihood

We can also calculate PDF values centered on the maximum likelihood estimate of the mean.

$$\mathcal{L}_0(\mu_0 = 0, \hat{\sigma}_0^2 = 5.26)$$

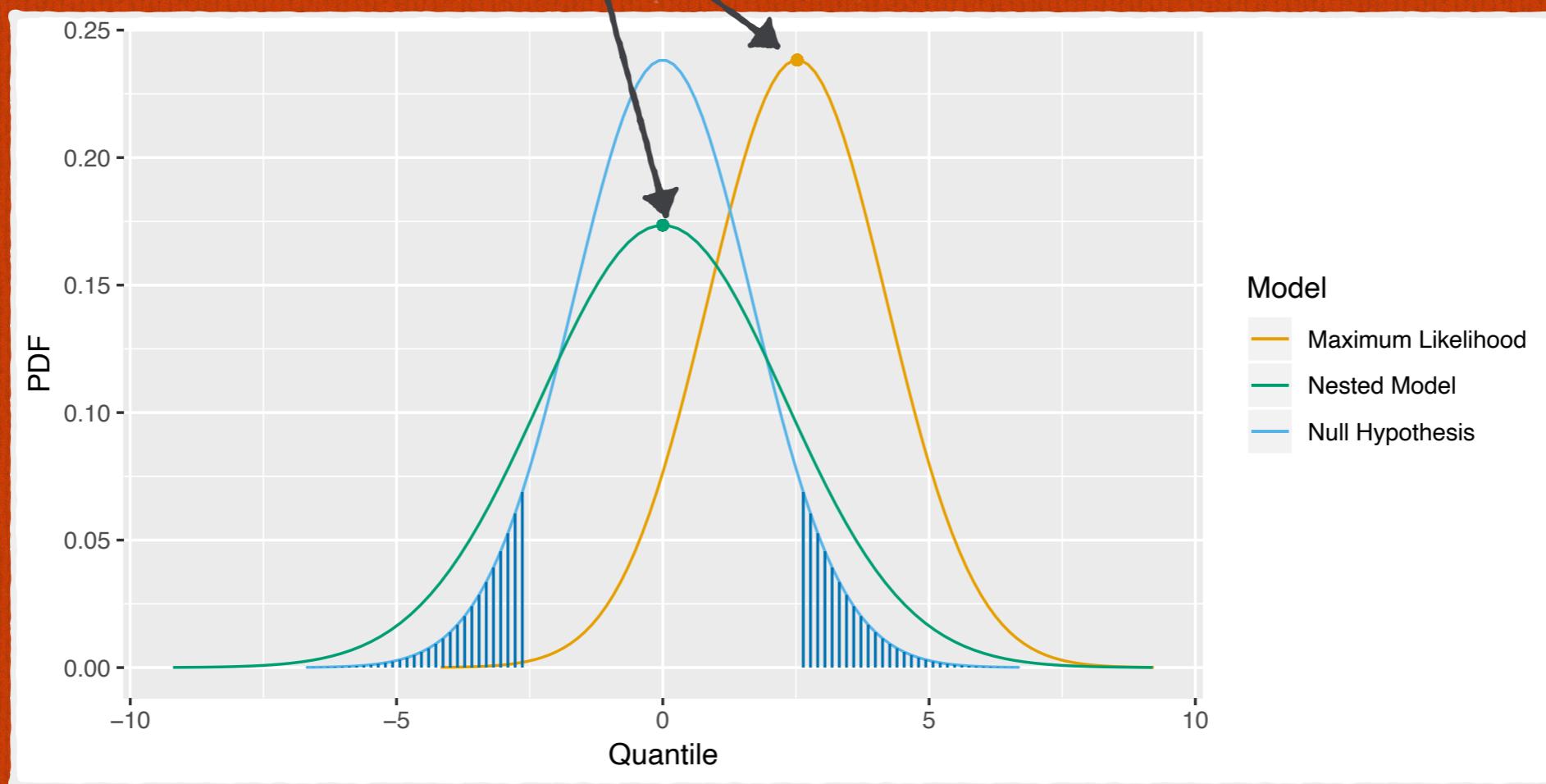
$$\mathcal{L}(\hat{\mu} = 2.52, \hat{\sigma}^2 = 2.83)$$



Probability vs Likelihood

We can also calculate PDF values when one or more parameters is constrained to 0.

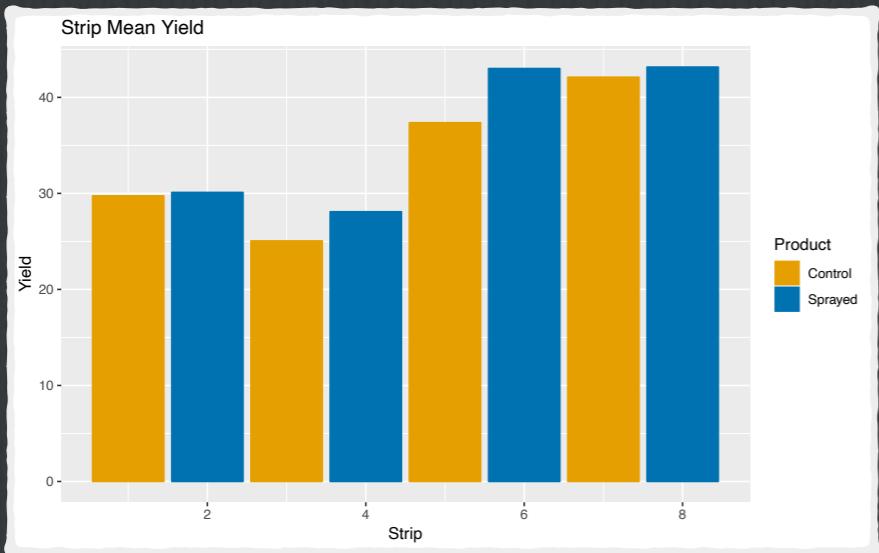
$$\frac{\mathcal{L}_0(\mu_0 = 0, \hat{\sigma}_0^2 = 5.26)}{\mathcal{L}(\hat{\mu} = 2.52, \hat{\sigma}^2 = 2.83)}$$



Probability vs Likelihood

The likelihood ratio is a measure of the relative likelihood of two hypothesis

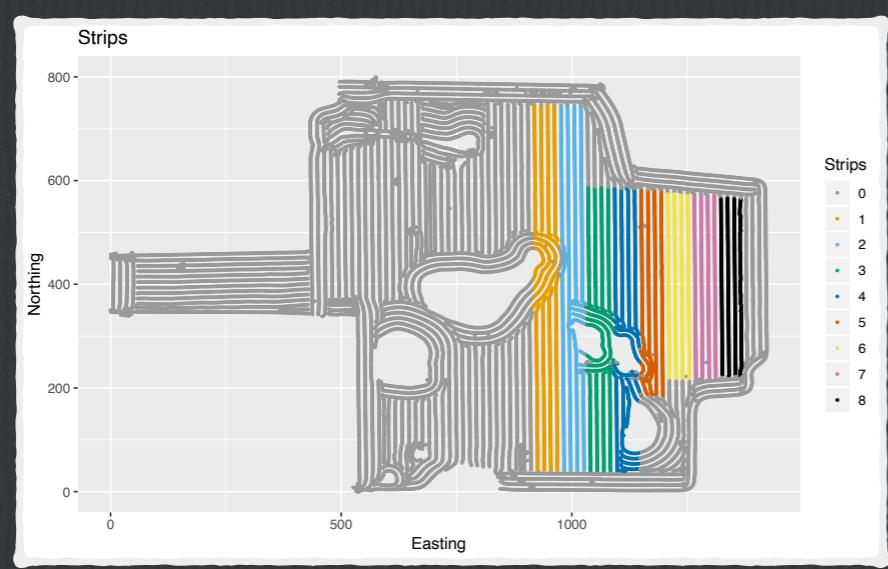
Likelihood Ratio (Naive)



□ We state two hypothesis:

$$H_1 : y_{ij} = \mu + \rho_j + e_{ij}$$

$$H_2 : y_{ij} = \mu + \rho_j + \tau_i + e_{ij}$$



□ The linear models are then

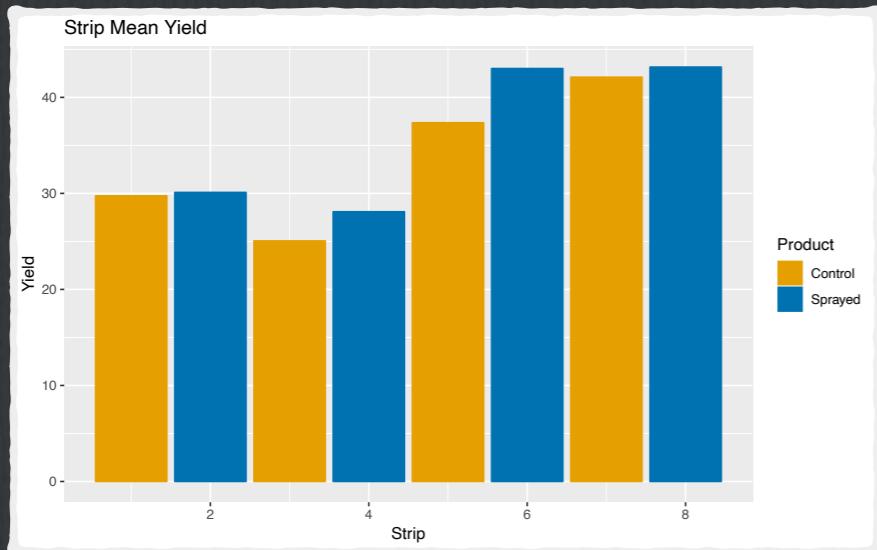
$$H_1 : y = X\beta_1; \beta_1 = \{\mu, \rho_1, \dots, \rho_m\}$$

$$H_2 : y = X\beta_2; \beta_2 = \{\mu, \rho_1, \dots, \rho_m, \tau_i\}$$

□ or

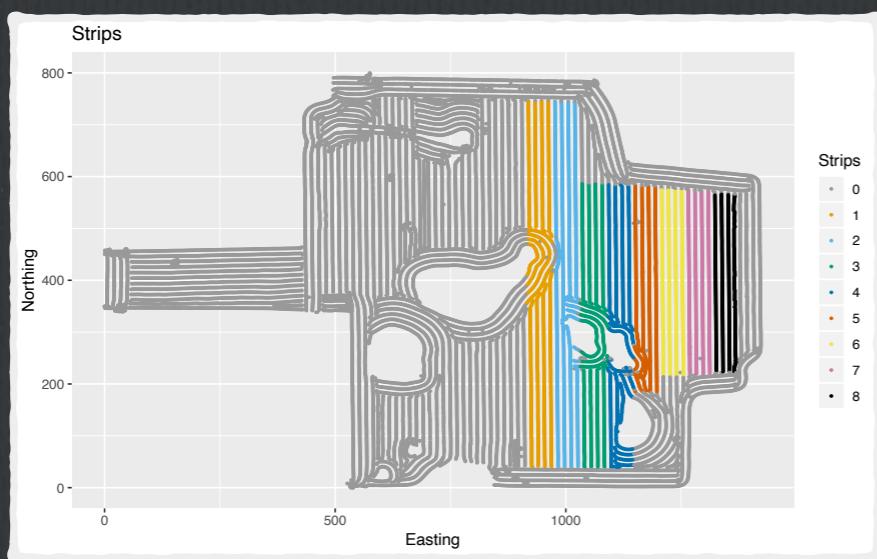
$$\beta_1 = \beta_2 \mid \tau = 0$$

Likelihood Ratio (Naive)

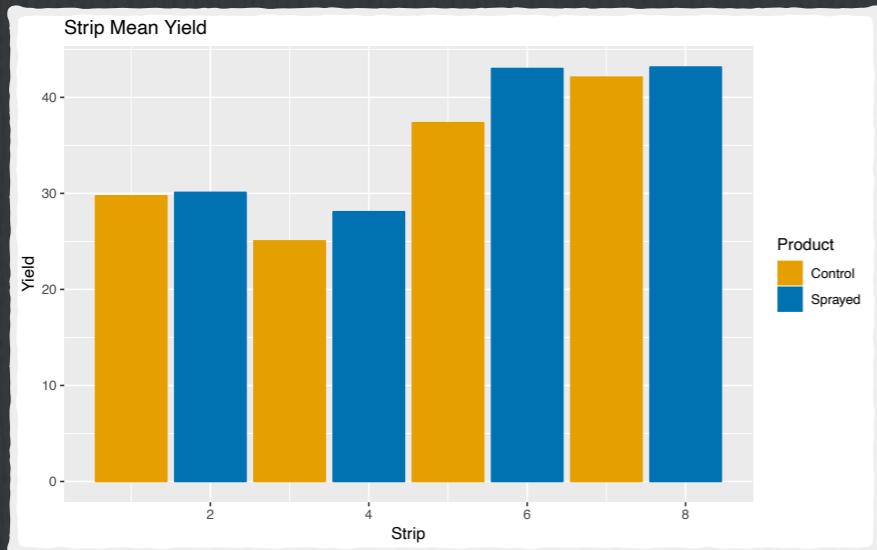


- We can test this by calculating the log-likelihood for two models

```
> H1.lm <- lm(Yield ~ Block,  
  data=meansEqual.dat)  
> H2.lm <- lm(Yield ~ Block + Product,  
  data=meansEqual.dat)  
> logLik(H1.lm)  
'log Lik.' -29.18457 (df=5)  
> logLik(H2.lm)  
'log Lik.' -28.61411 (df=6)  
> logLik(H2.lm)-logLik(H1.lm)  
'log Lik.' 0.5704535 (df=6)  
> exp(logLik(H2.lm)-logLik(H1.lm))  
'log Lik.' 1.769069 (df=6)
```

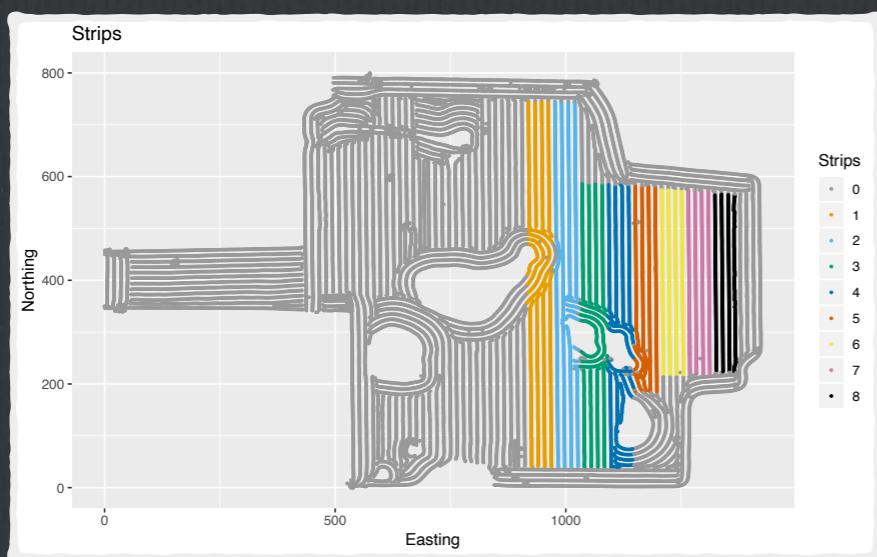


Likelihood Ratio (Naive)



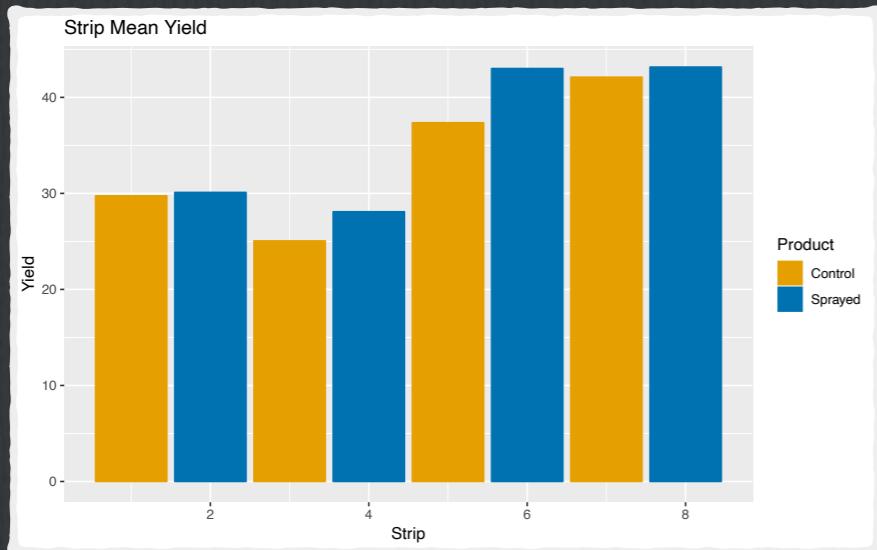
- Subtract the logs, then exponentiate to get a ratio

```
> H1.lm <- lm(Yield ~ Block,  
data=meansEqual.dat)  
> H2.lm <- lm(Yield ~ Block + Product,  
data=meansEqual.dat)  
> logLik(H1.lm)  
'log Lik.' -29.18457 (df=5)  
> logLik(H2.lm)  
'log Lik.' -28.61411 (df=6)  
> logLik(H2.lm)-logLik(H1.lm)  
'log Lik.' 0.5704535 (df=6)  
> exp(logLik(H2.lm)-logLik(H1.lm))  
'log Lik.' 1.769069 (df=6)
```

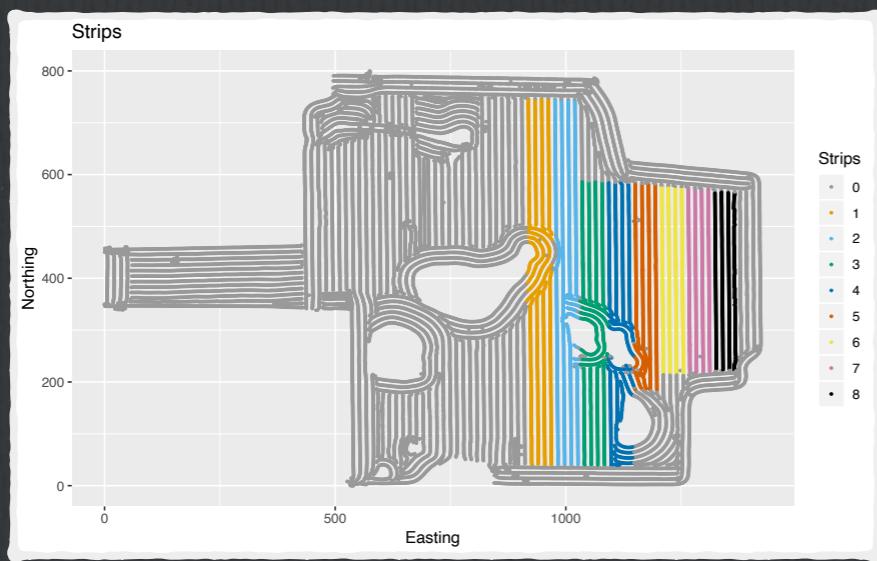


- H2 is more likely, but is it much more likely?

Likelihood Ratio (Naive)

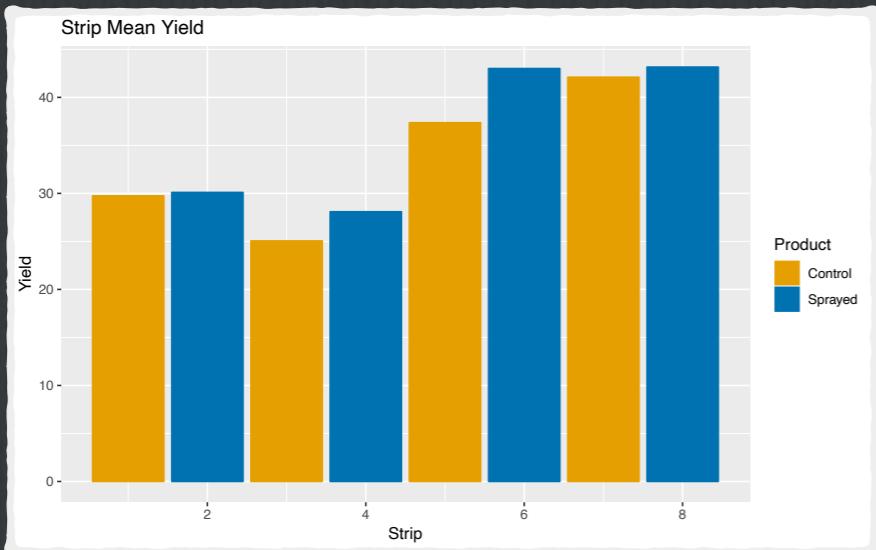


> H1.lm <- lm(Yield ~ Block,
 data=meansEqual.dat)
> H2.lm <- lm(Yield ~ Block + Product,
 data=meansEqual.dat)
> logLik(H1.lm)
'log Lik.' -29.18457 (df=5)
> logLik(H2.lm)
'log Lik.' -28.61411 (df=6)
> logLik(H2.lm)-logLik(H1.lm)
'log Lik.' 0.5704535 (df=6)
> exp(logLik(H2.lm)-logLik(H1.lm))
'log Lik.' **1.769069** (df=6)

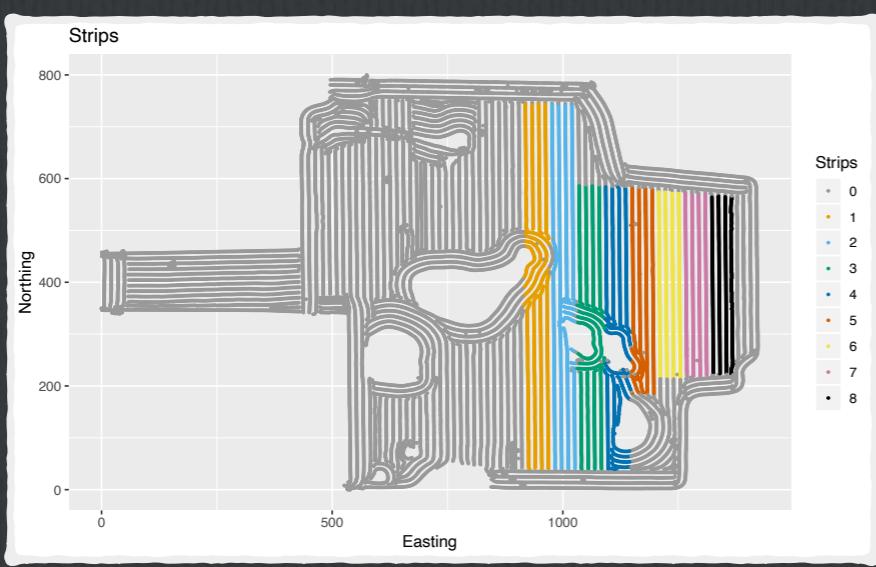


Is it reasonable to say that it's almost twice as likely that the product increased yield by 2.5 b/acre, than that the product did nothing?

Information Criteria



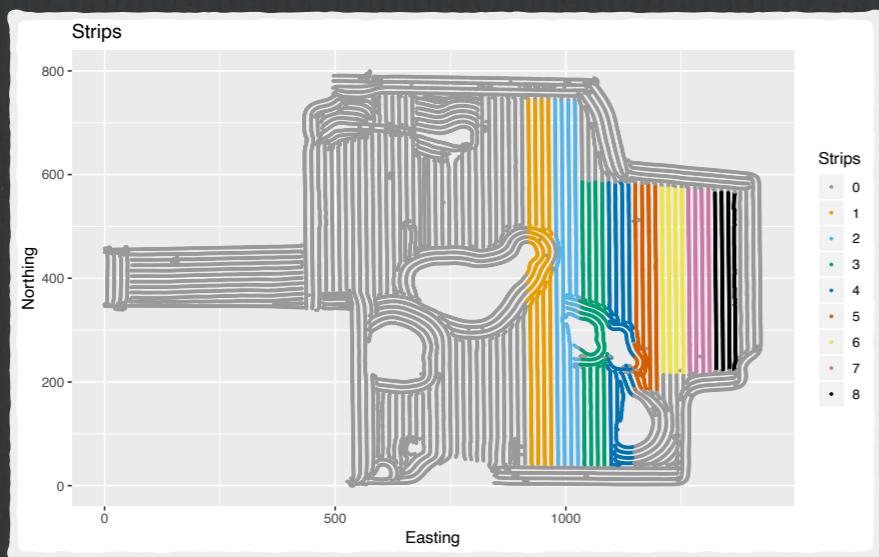
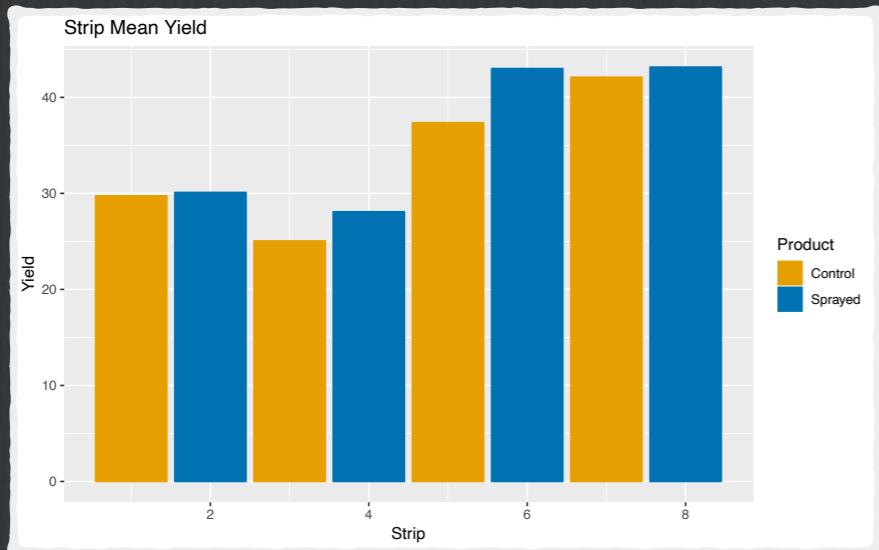
- We might use information criteria,
 $AIC = 2k - 2\ell$
 $BIC = \ln(n)k - 2\ell$
- where k is the number of parameters, and n is the number of observations



```
□ > AIC(H1.lm)
[1] 68.36914
> AIC(H2.lm)
[1] 69.22823
> BIC(H1.lm)
[1] 68.76634
> BIC(H2.lm)
[1] 69.70488
```

- With IC, smaller is better; the model without a parameter for treatment effect is preferred

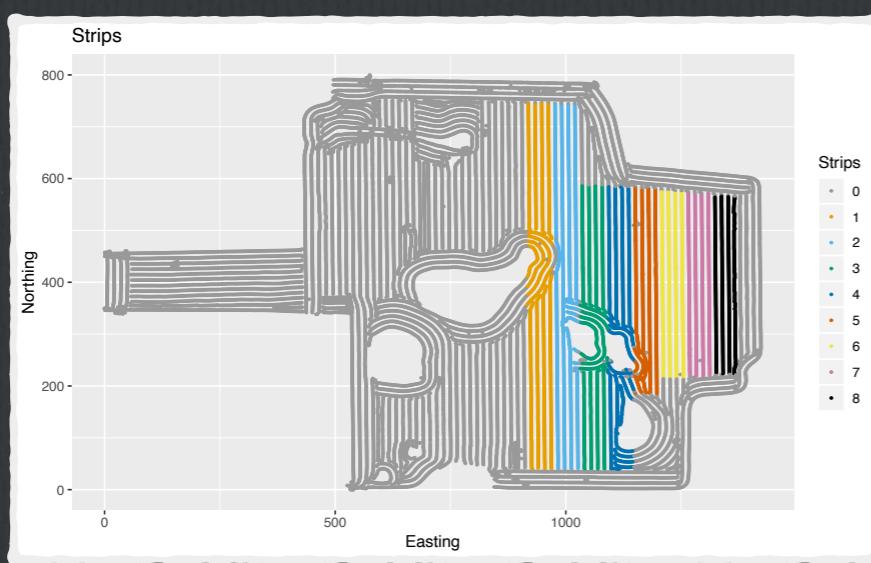
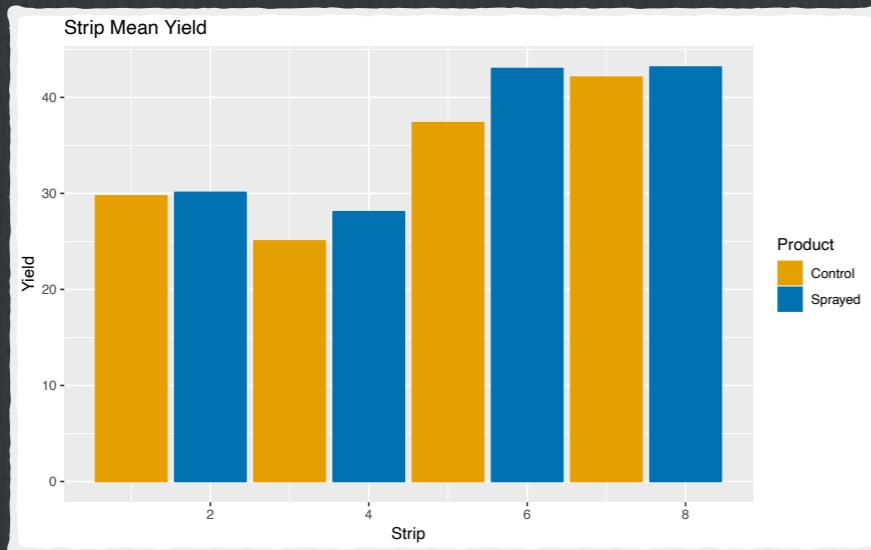
Likelihood Ratio Test



- A likelihood ratio test statistics is written as $LR = -2(\ell_1 - \ell_2)$ approaches a χ^2 distribution, so sometimes this is used as a null hypothesis test:

```
> lrtest(H1.lm, H2.lm)
      Likelihood ratio test
    Model 1: Yield ~ Block
    Model 2: Yield ~ Block + Product
      Df LogLik Df Chisq Pr(>Chisq)
    1     5 -29.185
    2     6 -28.614  1 1.1409   0.2855
```

Modeling a Trend by Position



- Next, consider the trend models

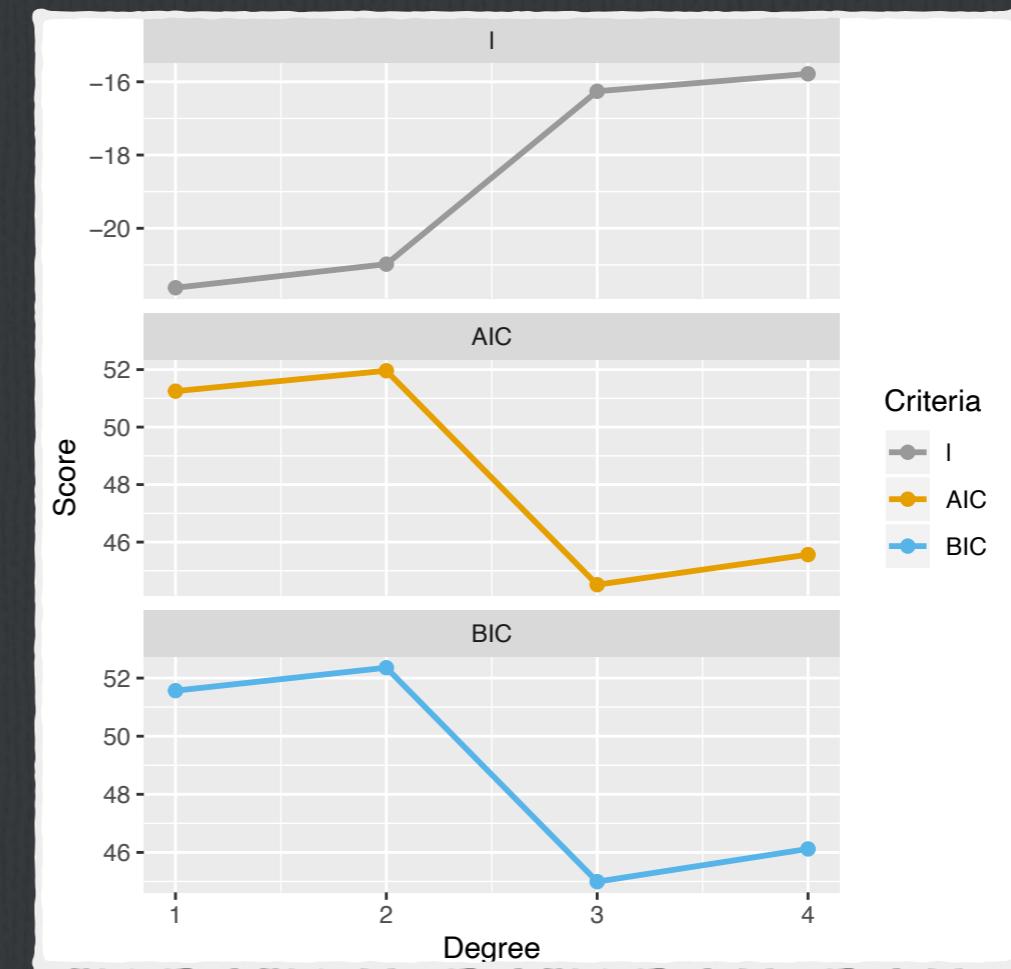
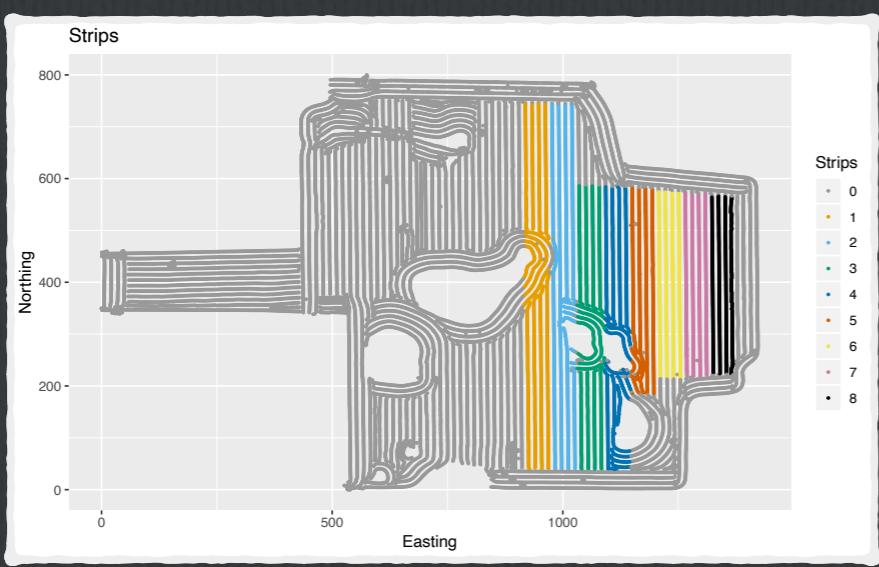
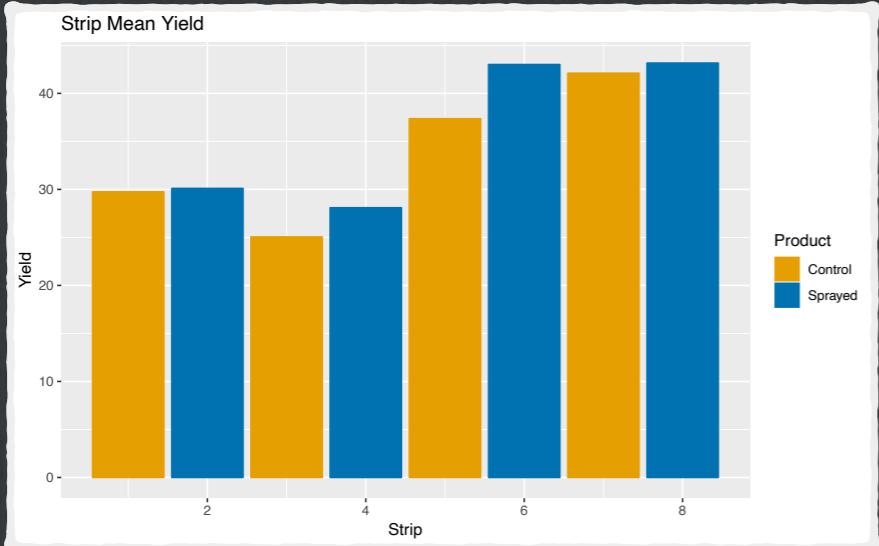
$$H_1 : y_{ij} = \beta_0 + \beta_1 E_{ij} + \tau_i + e_{ij}$$

$$H_2 : y_{ij} = \beta_0 + \beta_1 E_{ij} + \beta_2 E_{ij}^2 + \tau_i + e_{ij}$$

$$H_3 : y_{ij} = \beta_0 + \beta_1 E_{ij} + \dots + \beta_3 E_{ij}^3 + \tau_i + e_{ij}$$

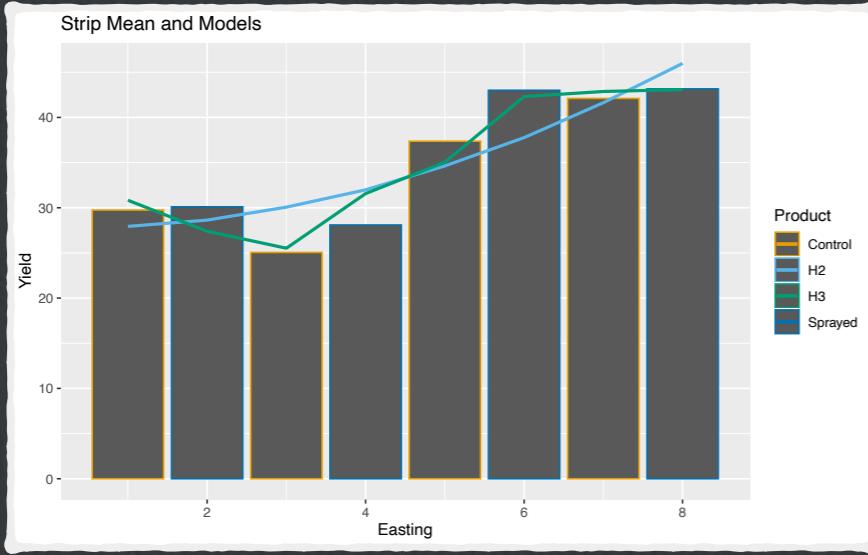
$$H_4 : y_{ij} = \beta_0 + \beta_1 E_{ij} + \dots + \beta_4 E_{ij}^4 + \tau_i + e_{ij}$$

Model Selection



A cubic trend is the likely model

Model Comparison



- Likelihood test to compare H_2 and H_3
- Model 1: Yield ~ poly(Pass, 2) + Product
Model 2: Yield ~ poly(Pass, 3) + Product

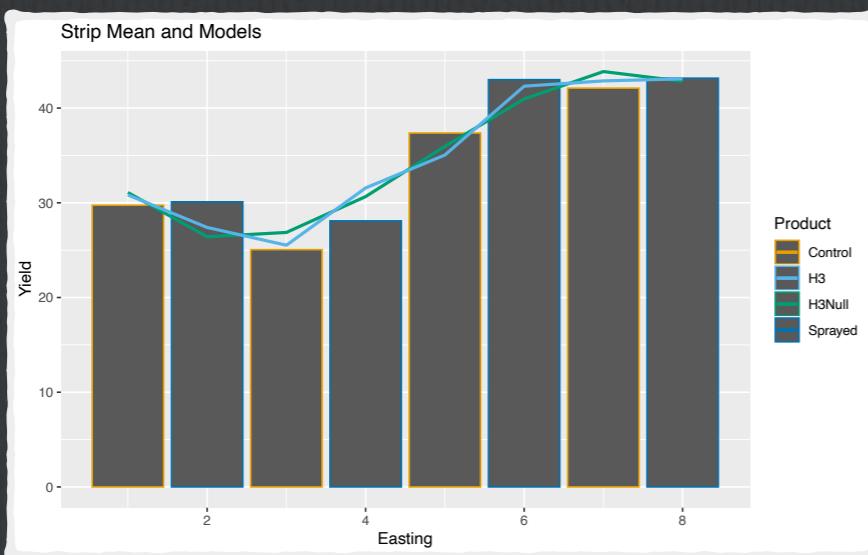
#Df	LogLik	Df	Chisq	Pr(>Chisq)
1	5	-20.980		
2	6	-16.259	1	9.4428 0.00212 **

- Likelihood test to compare H_3 with and without treatment effect

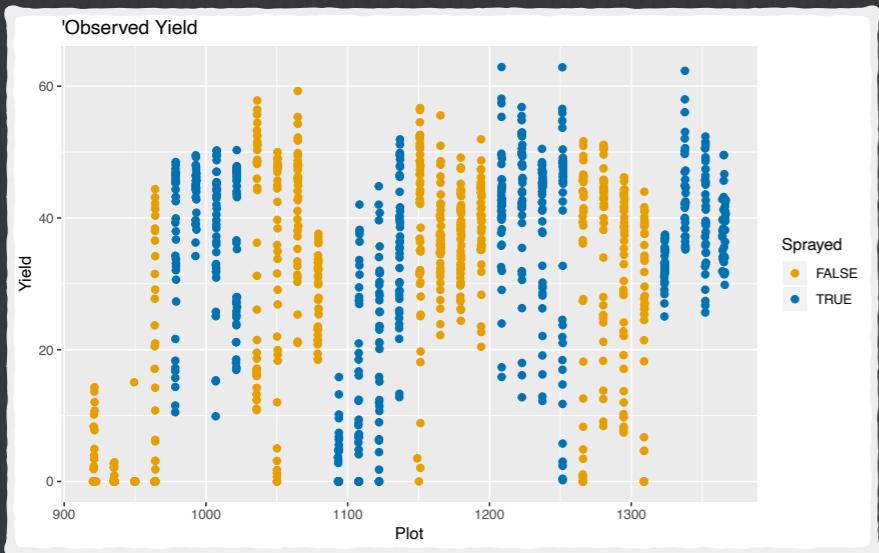
- Likelihood ratio test
- Model 1: Yield ~ poly(Pass, 3)
Model 2: Yield ~ poly(Pass, 3) + Product

#Df	LogLik	Df	Chisq	Pr(>Chisq)
1	5	-17.217		
2	6	-16.259	1	1.9168 0.1662

- and
- $H_3 : \tau = 2.108$
- $LR = 2.61$



Trend Analysis Across ‘Plots’

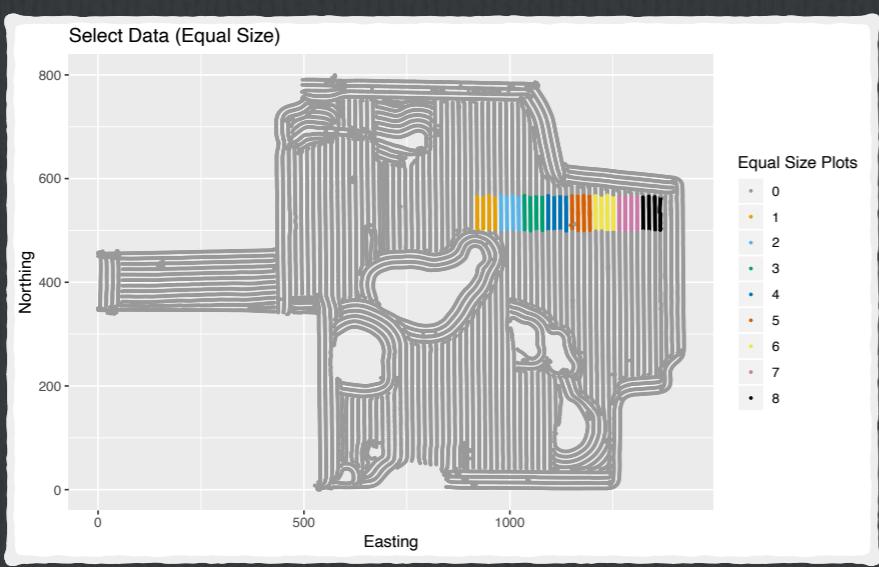
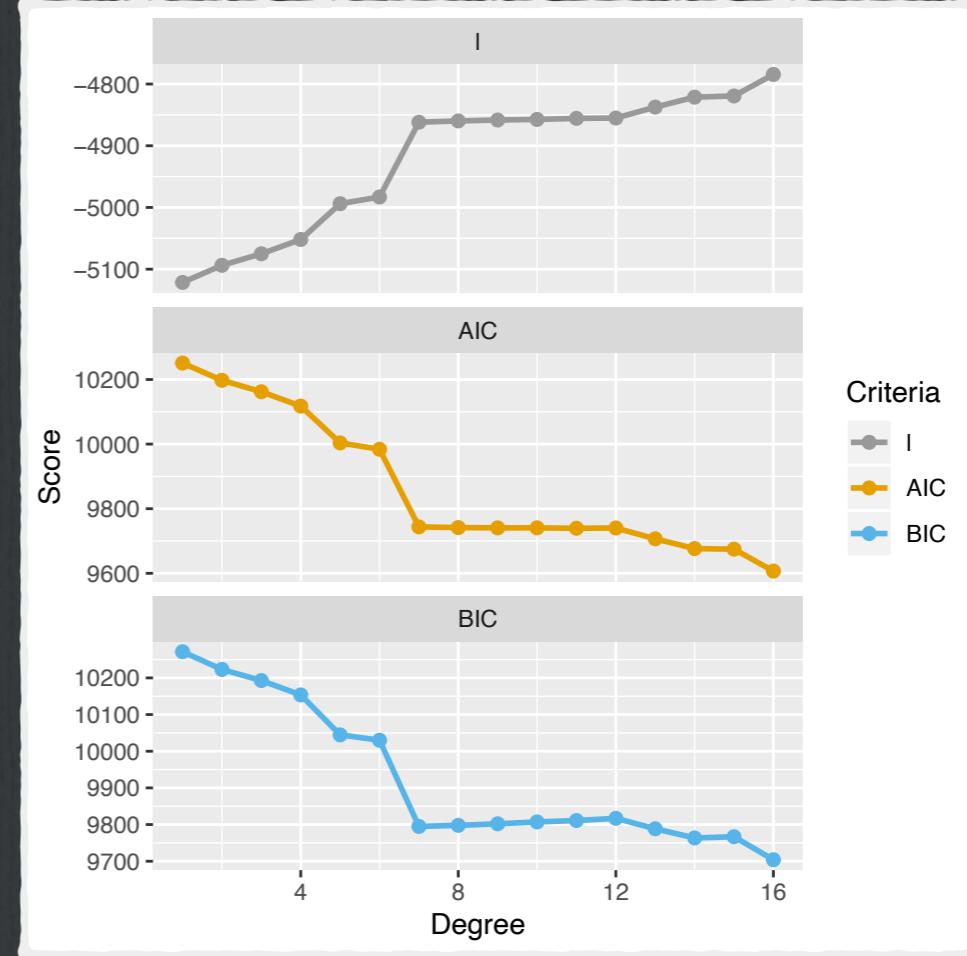
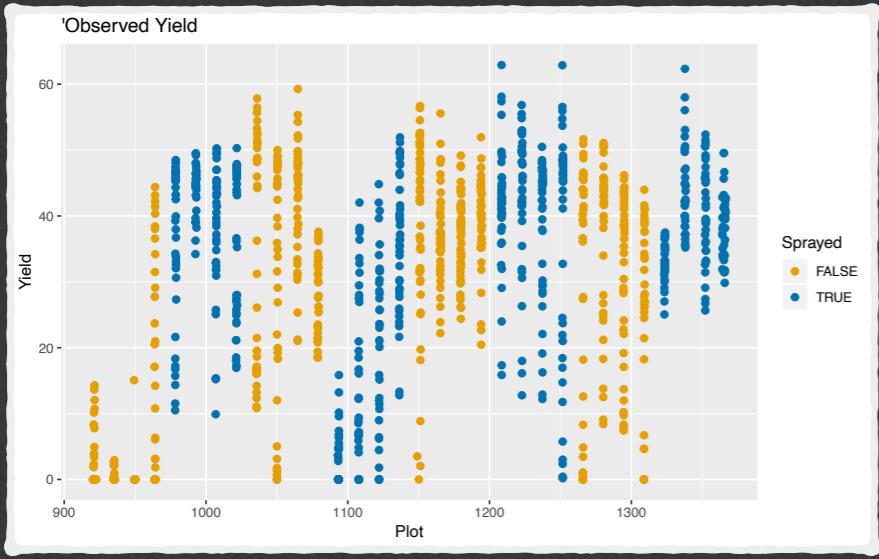


- We revisit the ‘plot’ analysis, but this time using yield observations, and we model by distance from the East edge of the field.



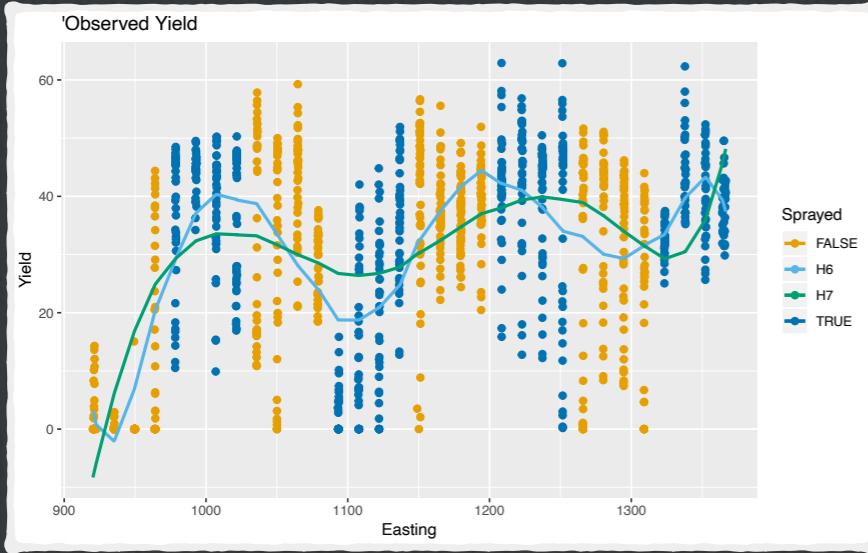
- As before, calculate log-likelihood, AIC and BIC for each hypothesis

Model Selection



- There appears to be little benefit with a polynomial of degree higher than 7

Model Comparison



Likelihood test to compare H_6 and H_7

Likelihood ratio test

	#Df	LogLik	Df	Chisq	Pr(>Chisq)
Model 1:	9	-4983.0			
Model 2:	10	-4861.9	1	242.29	< 2.2e-16 ***

Likelihood test to compare H_7 with and without treatment effect

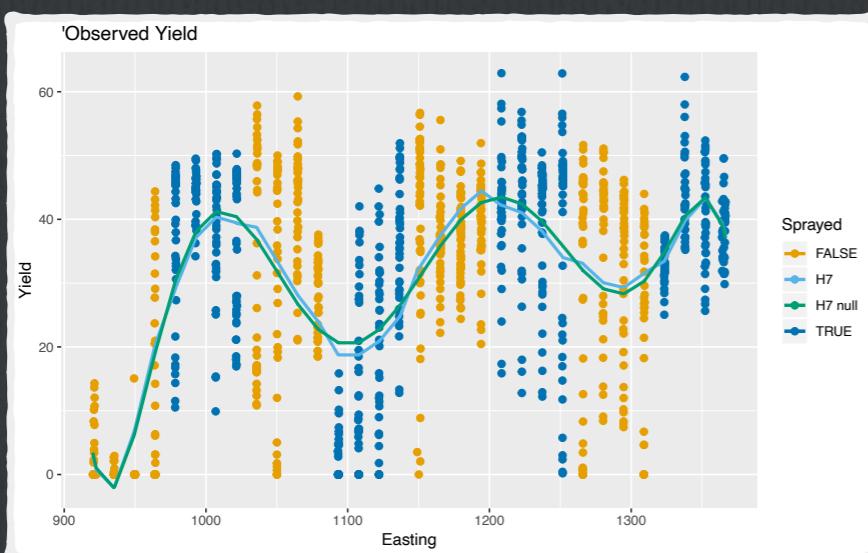
Likelihood ratio test

	#Df	LogLik	Df	Chisq	Pr(>Chisq)
Model 1:	9	-4869.0			
Model 2:	10	-4861.9	1	14.276	0.0001579 ***

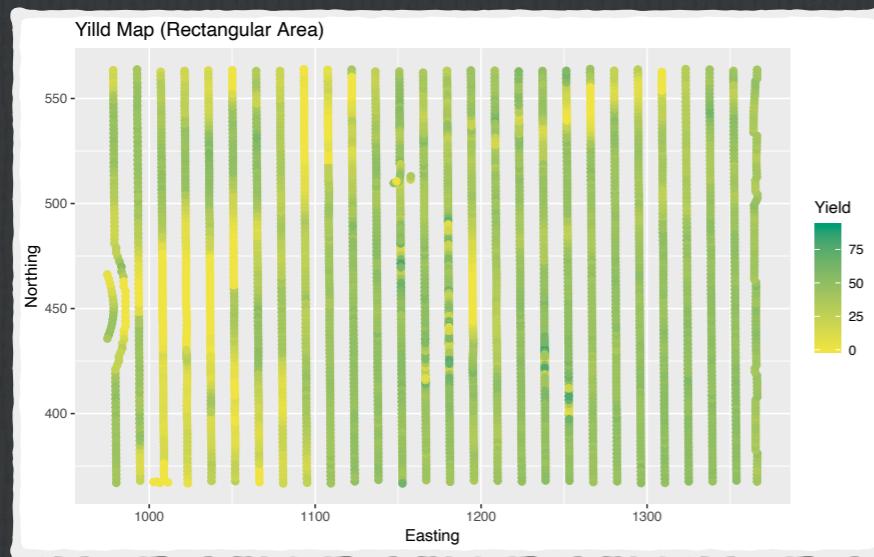
But,

$$H_6 : \tau = -0.86$$

$$H_7 : \tau = -3.02$$



Model in Two Dimensions

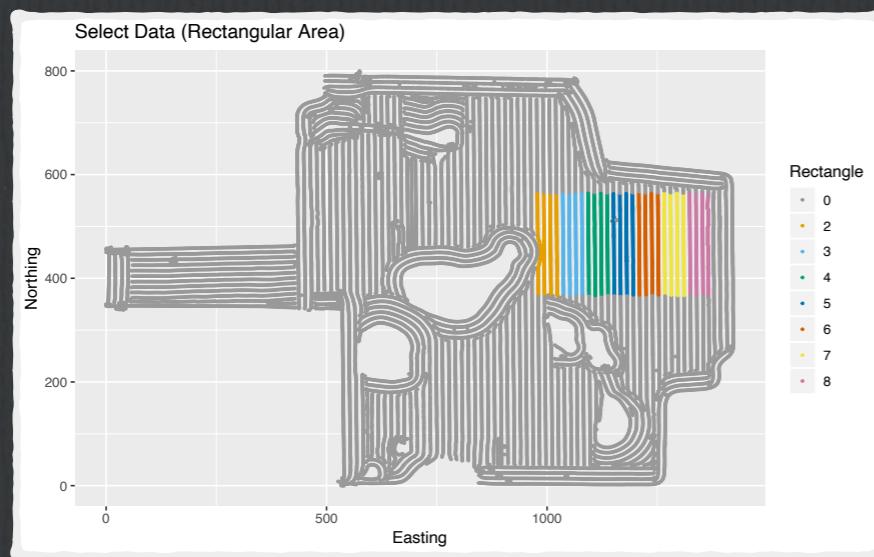


- We extend the model by including position in two-dimensions

$$H_1 : y_{ij} = \beta_0 + \beta_1 E_{ij} + \beta_2 N_{ij} + \tau_i + e_{ij}$$

$$H_2 : y_{ij} = \beta_0 + \beta_1 E_{ij} + \beta_2 N_{ij} + \beta_3 E_{ij}^2 + \beta_4 N_{ij}^2 + \beta_5 E_{ij} \times N_{ij} + \tau_i + e_{ij}$$

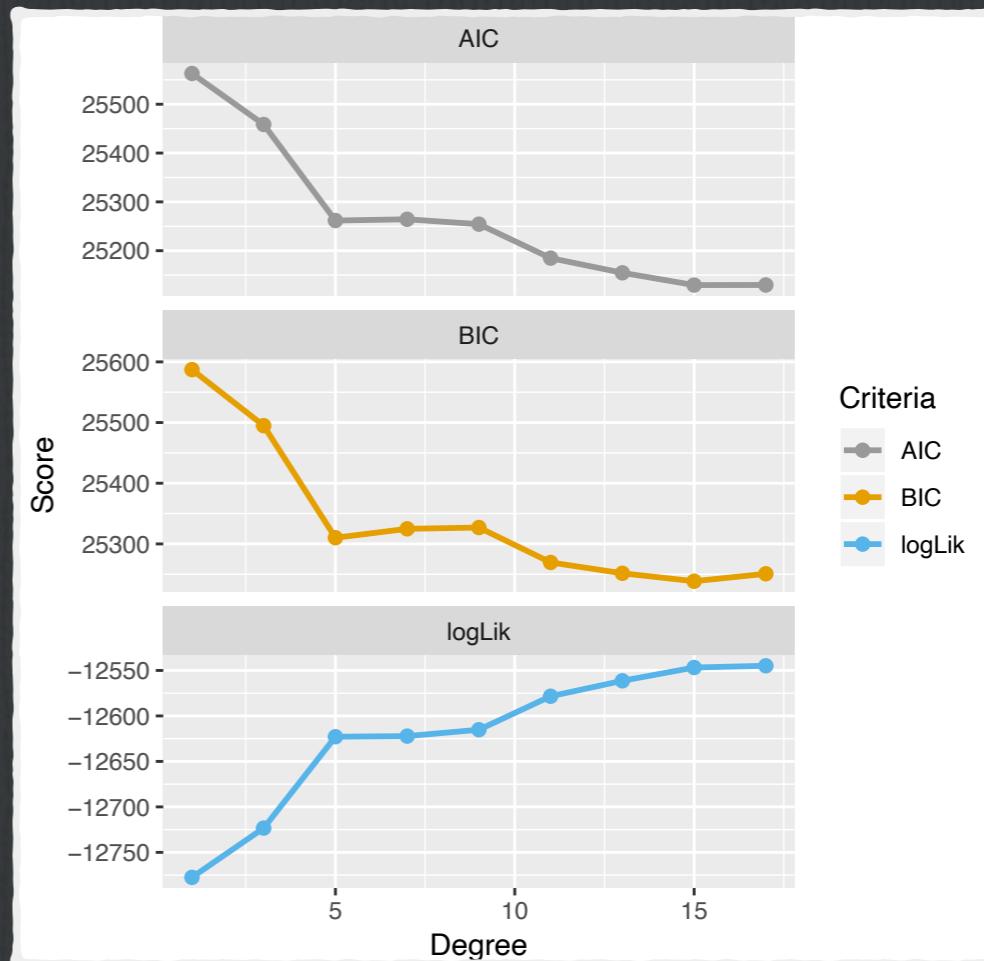
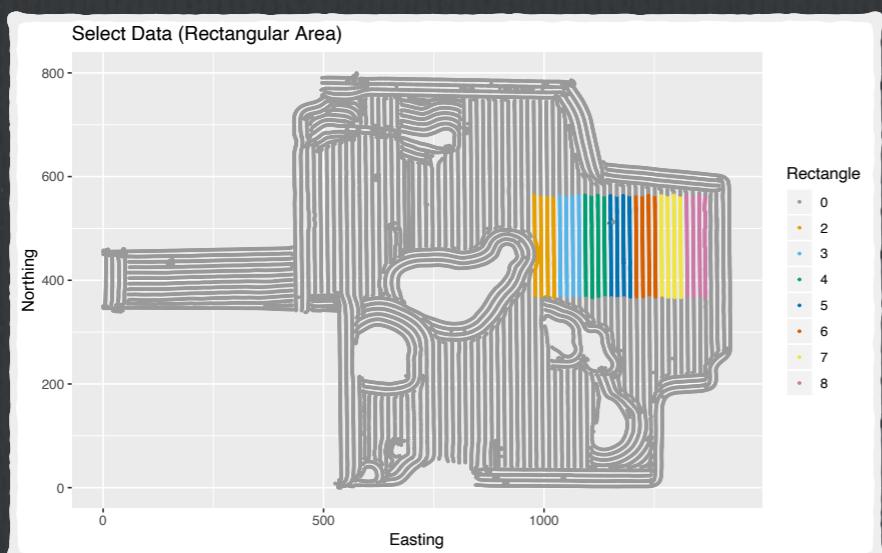
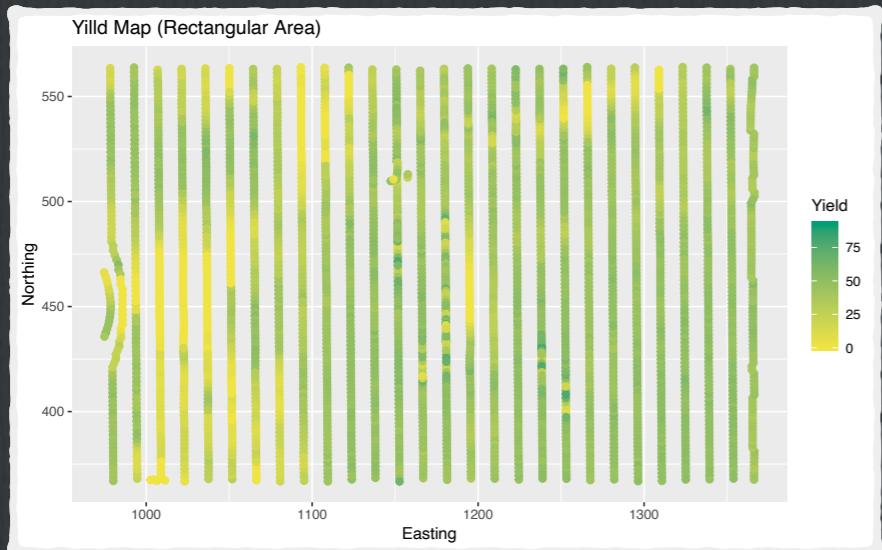
...



- We can expand our area to be analyzed to include more of the treated strips.

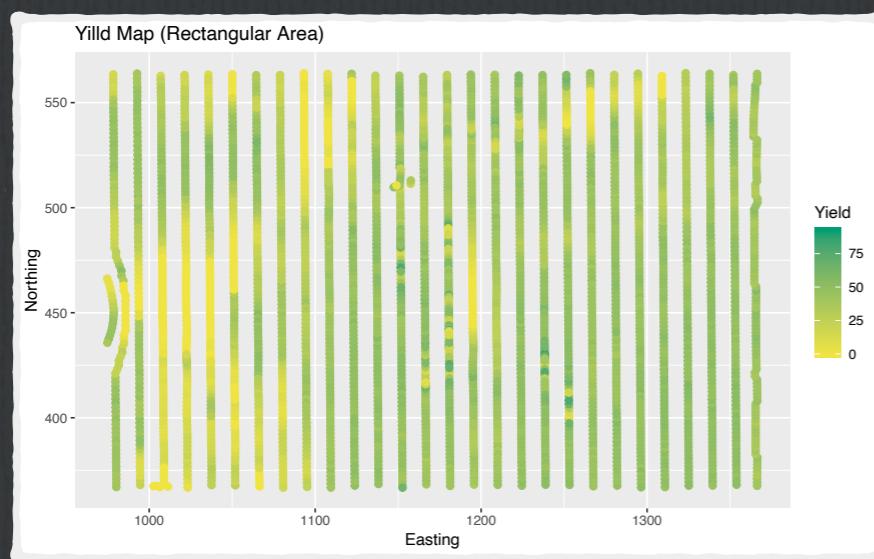
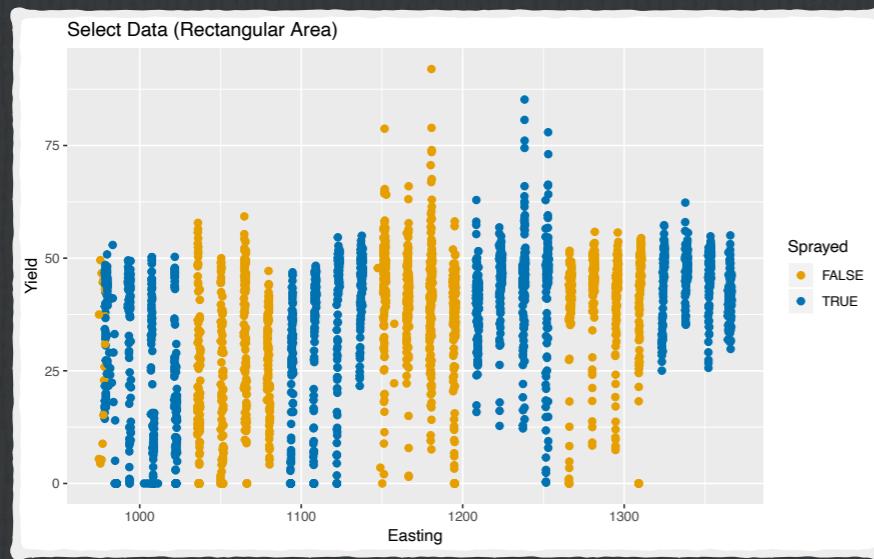
- Since I'm no longer analyzing as pairs, I can exclude part of the untreated area that avoids a wetland.

Model Selection



In 2 dimensions, 5th degree polynomials are sufficient.

Model Comparison



Likelihood test to compare 1D and 2D models

Likelihood ratio test

```
Model 1: Yield ~ poly(Easting, 5) + Sprayed
Model 2: Yield ~ poly(Easting, 7) * poly(Northing, 7) +
Sprayed
#Df LogLik Df Chisq Pr(>Chisq)
1 8 -12623
2 66 -11700 58 1846.4 < 2.2e-16 ***
```

Likelihood test to compare 2D models with and without treatment effect

Likelihood ratio test

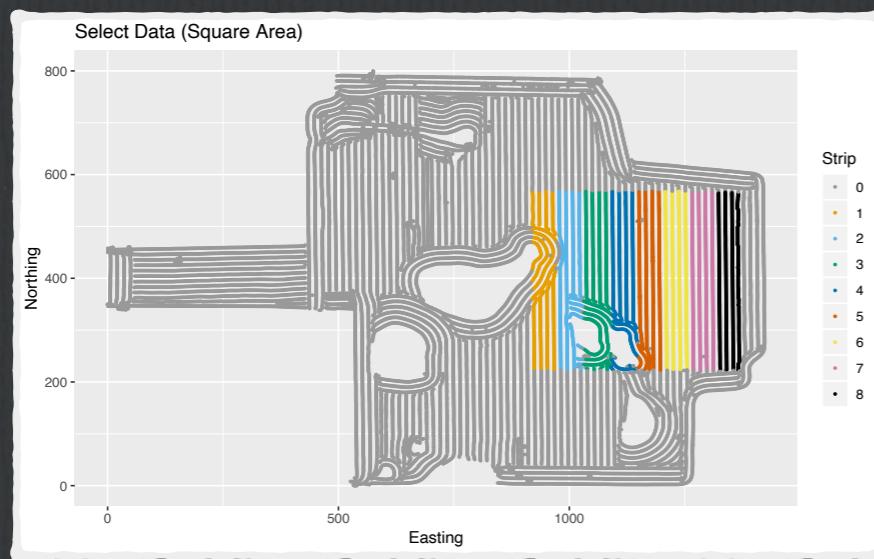
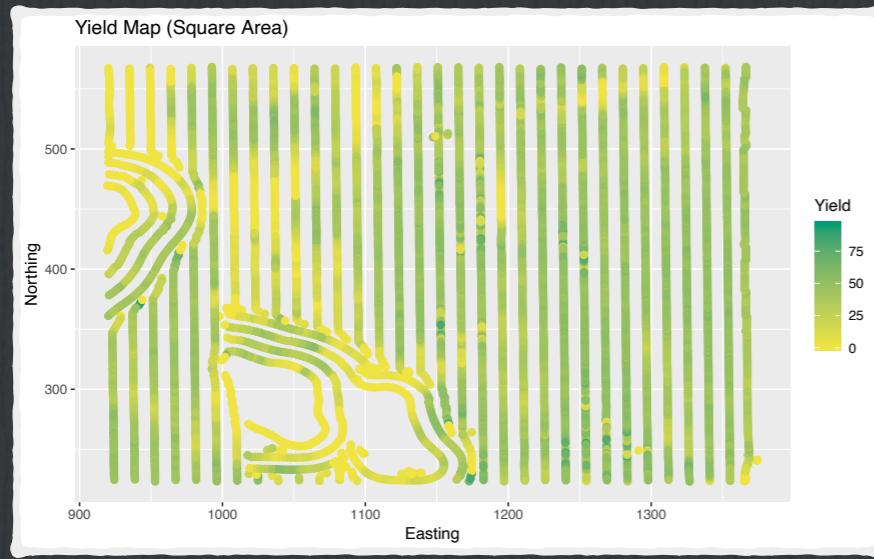
```
Model 1: Yield ~ poly(Easting, 7) * poly(Northing, 7)
Model 2: Yield ~ poly(Easting, 7) * poly(Northing, 7) +
Sprayed
#Df LogLik Df Chisq Pr(>Chisq)
1 65 -11702
2 66 -11700 1 3.8295 0.05036 .
```

and

$$H_6 : \tau = 0.8762, p(t) = 0.101$$

$$H_{2D} : \tau = 0.787, p(t) = 0.052825$$

Expanded Square



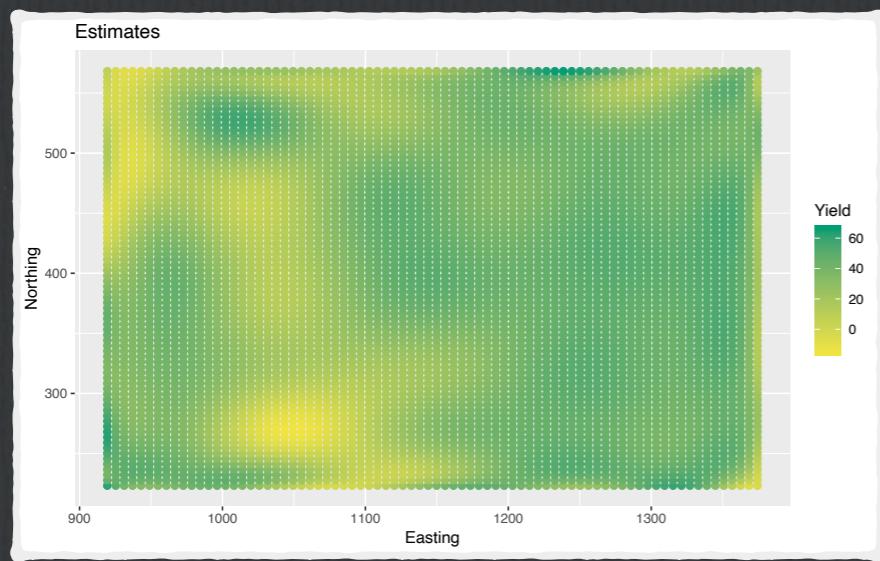
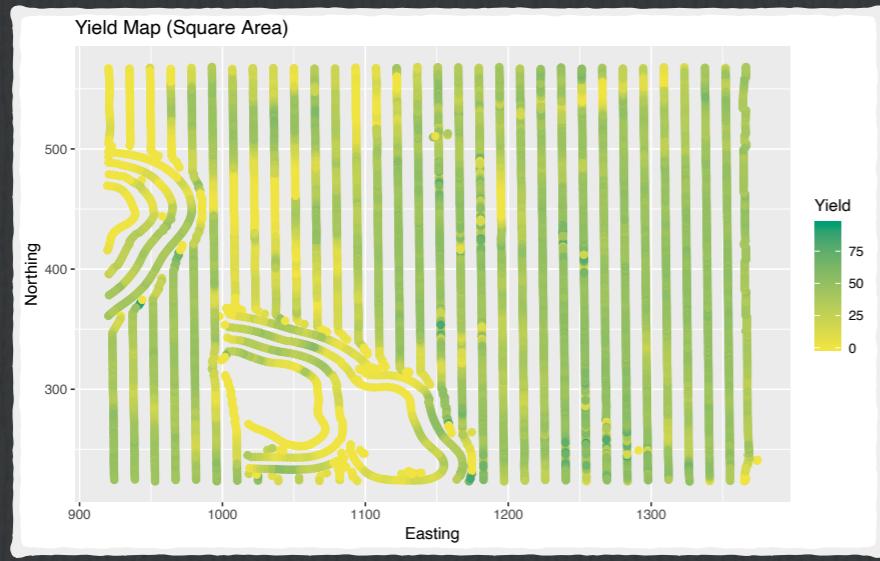
- Likelihood test to compare 2D models with and without treatment effect

- Likelihood ratio test
- Model 1: Yield ~ poly(Easting, 17) * poly(Northing, 17)
- Model 2: Yield ~ poly(Easting, 17) * poly(Northing, 17) + Sprayed
- | | #Df | LogLik | Df | Chisq | Pr (> Chisq) |
|---|-----|--------|----|--------|--------------|
| 1 | 325 | -21602 | | | |
| 2 | 326 | -21601 | 1 | 1.0578 | 0.3037 |

□ and

$$H_{2D} : \tau = 0.648, p(t) = 0.317$$

Expanded Square

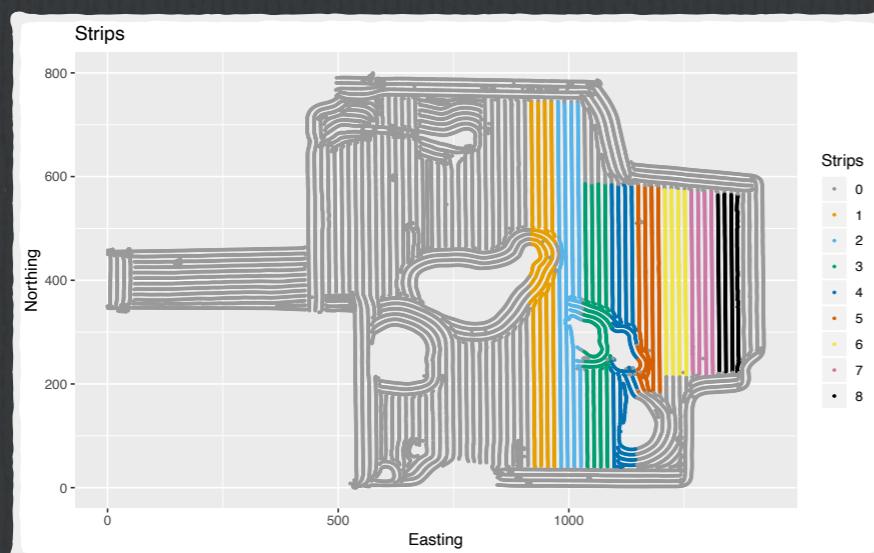
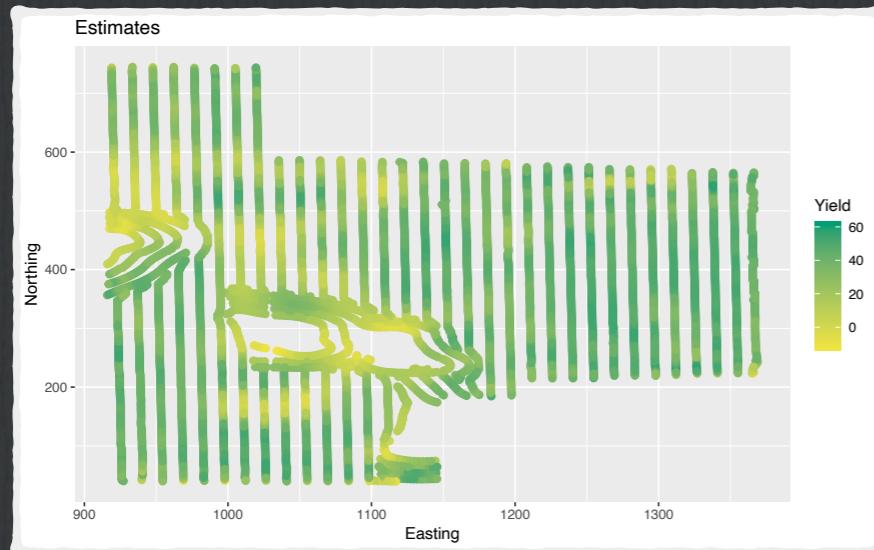


- Likelihood test to compare 2D models with and without treatment effect**
- Likelihood ratio test
 - Model 1: Yield ~ poly(Easting, 17) * poly(Northing, 17)
 - Model 2: Yield ~ poly(Easting, 17) * poly(Northing, 17) + Sprayed

	Df	LogLik	Df	Chisq	Pr (>Chisq)
1	325	-21602			
2	326	-21601	1	1.0578	0.3037

- and
 - $H_{2D} : \tau = 0.648, p(t) = 0.317$
- But we require some degree of interpolation to fill in missing areas.**

All Strips



- Likelihood test to compare 2D models with and without treatment effect

- Likelihood ratio test

Model 1: Yield ~ poly(Easting, 21) * poly(Northing, 21)

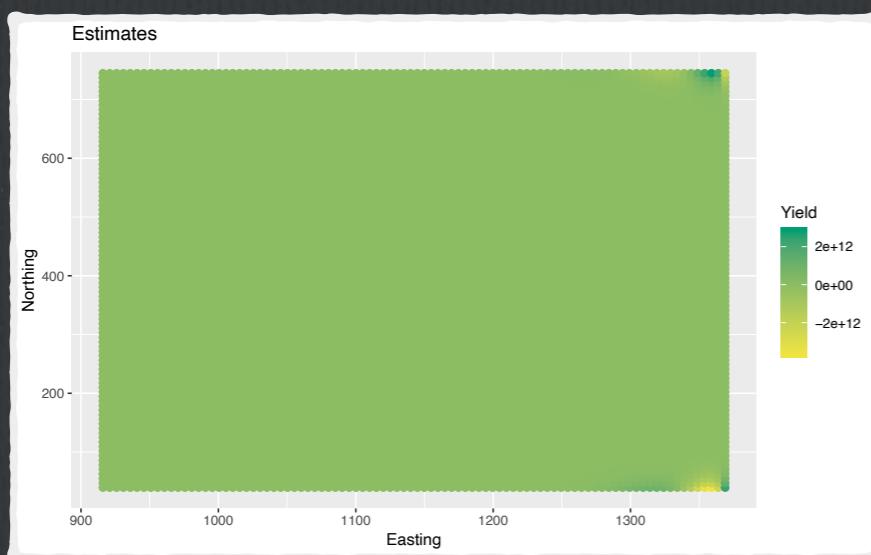
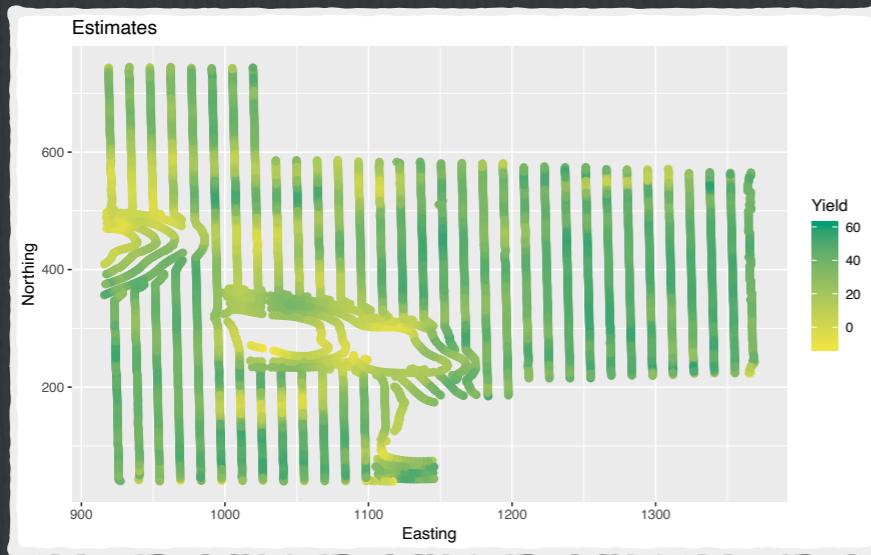
Model 2: Yield ~ poly(Easting, 21) * poly(Northing, 21) + Sprayed

#Df	LogLik	Df	Chisq	Pr (> Chisq)
1	485	-29622		
2	486	-29622	1	1.8792 0.1704

- and

$$H_{2D} : \tau = 0.833, p(t) = 0.183$$

All Strips



- Likelihood test to compare 2D models with and without treatment effect**
 - Likelihood ratio test
 - Model 1: Yield ~ poly(Easting, 21) * poly(Northing, 21)
 - Model 2: Yield ~ poly(Easting, 21) * poly(Northing, 21) + Sprayed
 - #Df LogLik Df Chisq Pr(>Chisq)

#Df	LogLik	Df	Chisq	Pr(>Chisq)
1	485	-29622		
2	486	-29622	1	1.8792 0.1704
- and**
$$H_{2D} : \tau = 0.833, p(t) = 0.183$$
- This requires interpolation and extrapolation.**

Correlated Errors Likelihood

- Up to this point, we've assumed a simple likelihood function, with only one random effect

$$\mathcal{L}_1(\beta_1, \sigma^2 | y_1, \dots, y_n) = \left(\sqrt{2\pi\sigma^2} \right)^{-n} \exp \left\{ -\frac{1}{2\sigma^2} \sum (y_i - X\beta)^2 \right\}$$

- A more appropriate model includes structured random effects

$$\mathcal{L}(\beta, V | y_1, \dots, y_n) = (2\pi)^{-nK/2} |V|^{-n/2} \exp \left\{ -\frac{1}{2} \sum (y_n - X\beta) V^{-1} (y_n - X\beta) \right\}$$

where V is a matrix describing the spatial correlation model, e.g.

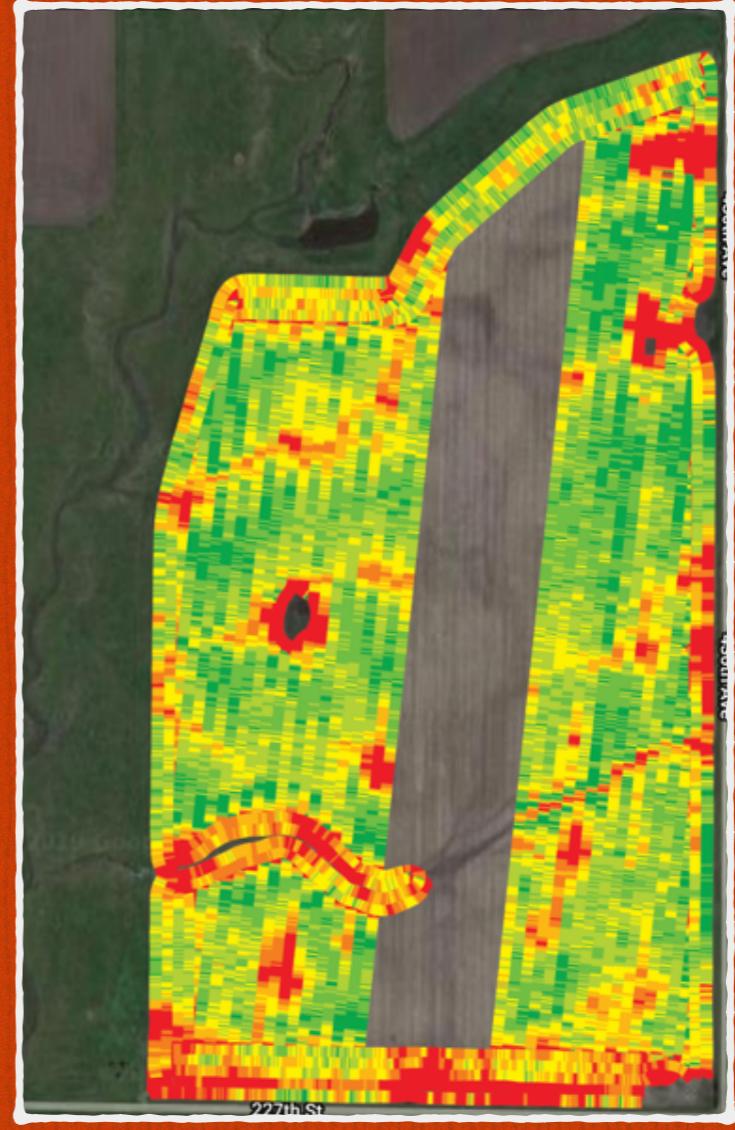
$$V[s_i, s_j] = \text{Cov}(s_i, s_j) = c_0 + \sigma^2 \exp \left(-\|s_i - s_j\| / \alpha \right)$$

Correlated Errors Likelihood



But then I've got to explain that,
to this guy.

Additional Thoughts



Spraying and Yield Maps

The example I've used was the good field

reductio ad absurdum

The two classes of results which are distinguished by our test of significance are, on the one hand, those which show a significant discrepancy from a certain hypothesis; ... ; and on the other hand, results which show no significant discrepancy from this hypothesis.

-R. A Fisher "The Design of Experiments, 8. The Null Hypothesis"

reductio ad absurdum

Proof by Contradiction

- To prove that an assertion is true, we first assume it is false, then show that this assumption leads to a logical contradiction

Proof by Contradiction

Example

- Proposition :
 $\sqrt{2}$ is irrational

- Proof :
Proof is by contraction. Assume $\sqrt{2}$ is rational. Then
 $\sqrt{2} = \frac{p}{q}$,
such that p and q are integers that have no common factors.

Proof by Contradiction

Example

- Given $\sqrt{2} = \frac{p}{q}$,
 $p^2 = 2q^2$, and p is even. Then q is odd.
- Since p is even, $p = 2k$.
So $p^2 = 4k^2 = 2q^2$ and $q^2 = 2k^2$. Then q is even.
- q cannot be both odd and even, therefore $\sqrt{2}$ cannot be rational

Null Hypothesis Test

Proof by Contradiction

Null Hypothesis Test

□ Proposition :

$$\tau_2 \neq \tau_1$$

□ Proof :

Proof is by contraction. Assume $\tau_2 \neq \tau_1$ is not true, that is, $\tau_2 = \tau_1 = 0$.

Then ... ?

Proof by Contradiction

Null Hypothesis Test

Calculate a test statistic

```
 > wilcox.test(Yield ~ Product, paired=TRUE,...)
      Wilcoxon signed rank test
data: Yield by Product
v = 0, p-value = 0.125

 > t.test(Yield ~ Product, paired=TRUE, ...)
      Paired t-test
data: Yield by Product
t = -2.1319, df = 3, p-value = 0.1228

 > friedman.test(Yield ~ Block | Product, ...)
      Friedman rank sum test
data: Yield and Block and Product
Friedman chi-squared = 6, df = 3, p-value = 0.1116

 > anova(Yield ~ Block + Product, ...)
      Analysis of Variance Table
Response: Yield
          Df Sum Sq Mean Sq F value    Pr(>F)
Block      3 363.52 121.173  43.230 0.005732 ***
Product    1  12.74  12.739   4.545 0.122791
Residuals  3    8.41    2.803
```

Is that statistic ‘absurd’ given that that the null hypothesis is taken a axiomatically true?

$p(\text{false}) = 0 < p(t) < p(\text{true}) = 1$

“Significance testing as perverse probabilistic reasoning”

- Consider a typical medical research study, for example designed to test the efficacy of a drug, in which a null hypothesis H_0 ('no effect') is tested against an alternative hypothesis H_1 ('some effect'). Suppose that the study results pass a test of statistical significance (that is P-value <0.05) in favor of H_1 . What has been shown?
- 1. H_0 is false.
- 2. H_1 is true.
- 3. H_0 is probably false.
- 4. H_1 is probably true.
- 5. Both (1) and (2).
- 6. Both (3) and (4).
- 7. None of the above.

Significance testing as perverse probabilistic reasoning

- Only 12 of 246 physicians surveyed in [1] chose #7.
- 1. H_0 is false.
- 2. H_1 is true.
- 3. H_0 is probably false.
- 4. H_1 is probably true.
- 5. Both (1) and (2).
- 6. Both (3) and (4).
- 7. None of the above.

Significance testing as perverse probabilistic reasoning

How should I communicate the results of a significance test to this guy?



- 1. H_0 is false.
- 2. H_1 is true.
- 3. H_0 is probably false.
- 4. H_1 is probably true.
- 5. Both (1) and (2).
- 6. Both (3) and (4).
- 7. None of the above.

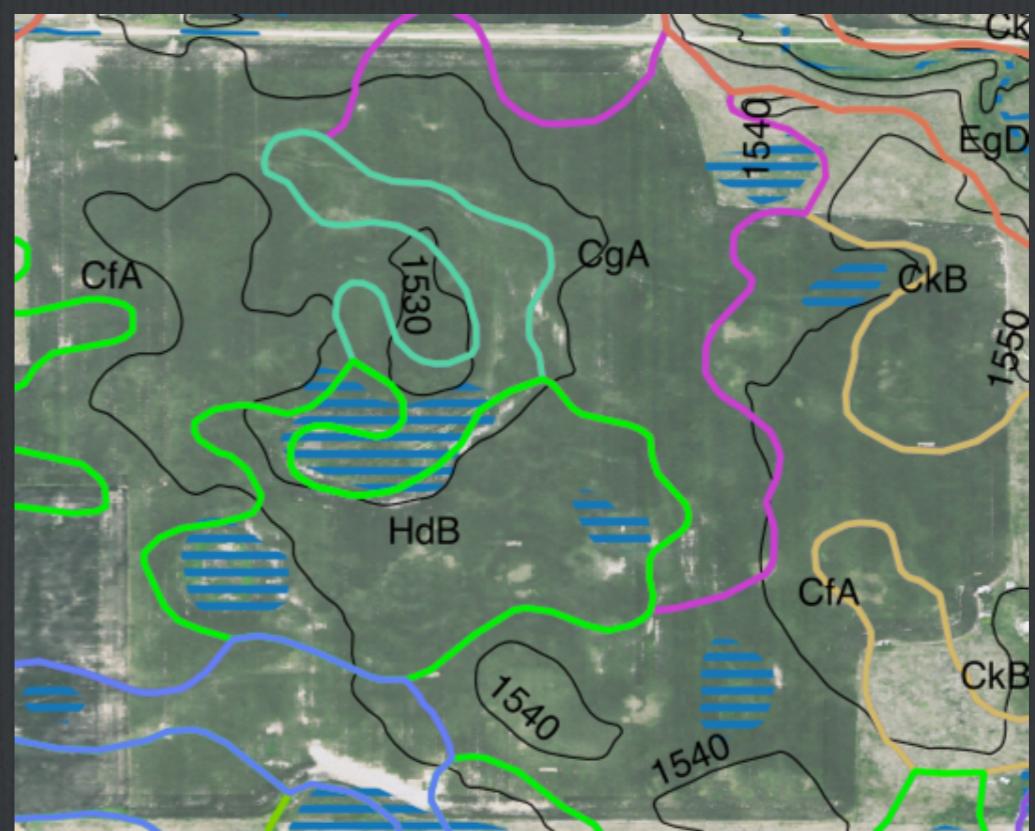
Randomization ; the Physical Basis of the Validity of the Test

-Johnny Appleseed

Exchangeability

- When we analyze a model of the form
$$y_{ij} = \mu + \tau_i + e_{ij}$$
- We typically assume a constraint
$$e_{ij} \sim \mathcal{N}(0, \sigma^2)$$
- This assumption is typically valid if each y_{ij} represent exchangeable units.
 - In a small-plot experiment, we design plots to be exchangeable.

Exchangeability



- Soil map units
- Topography
- Wetlands