Scientific Papers

Math & Stats Tutorial

Day 8

Prabesh Dhakal

03 June 2019

REVIEW: CORRELATION

- Describes the relationship of two variables
- Requires ordinal or continuous data (interval/ratio)
- Value ranges between -1 and 1 where
  - -1 : total negative relationship (one increases when other decreases and vice versa)
  - 0 : no relationship
  - +1 : total positive relationship (both tend to rise/fall together)

REVIEW: REGRESSION I

- **Simple linear regression** model: $y = \beta_0 + \beta_1 x + \varepsilon$
  - Is used to show/predict the relationship between two variables or factors in form of a straight line
  - The straight line approximates the relationship between the two variables
- More than 2 variables: *multiple regression*
- The regression model is judged based on coefficient of determination.

REVIEW: REGRESSION II

- **Coefficient of determination** (r-squared) reflects the proportion of variance (variability) that can be reduced by using independent variable.
  - Higher r-squared means better model
  - https://www.youtube.com/watch?v=IMjrEeeDB-Y&t=81s

Very good resource on simple linear regression, r-squared, and correlation: https://newonlinecourses.science.psu.edu/stat501/node/250/)

ANOVA

Applies to cases with more than 2 groups:

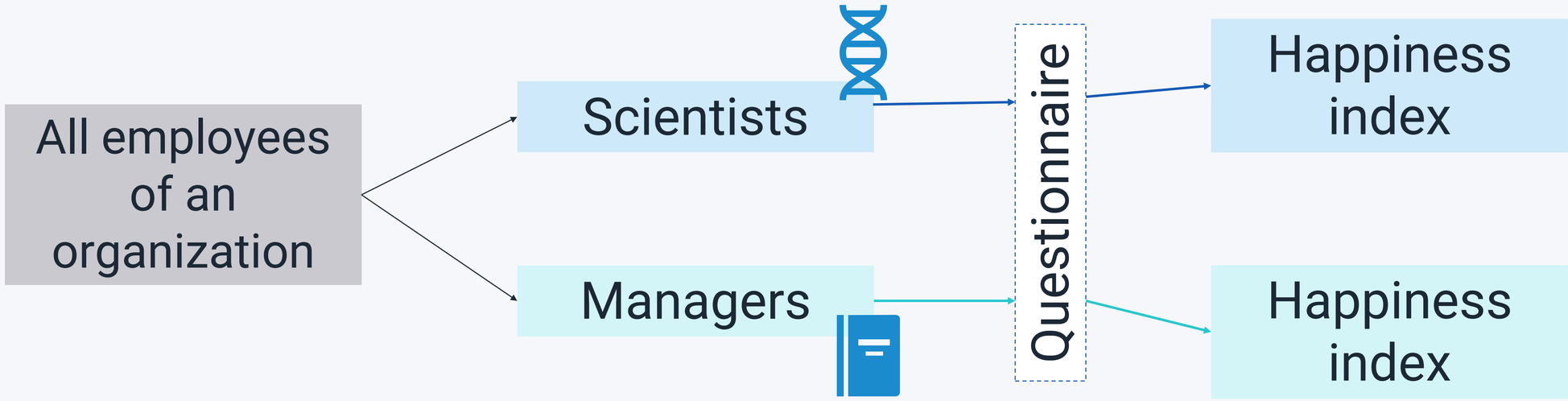
Null Hypothesis: Means of the groups are equal.

Two Research Questions:

1. Are scientists happier than managers?
2. Which of the 4 fertilizers work best?

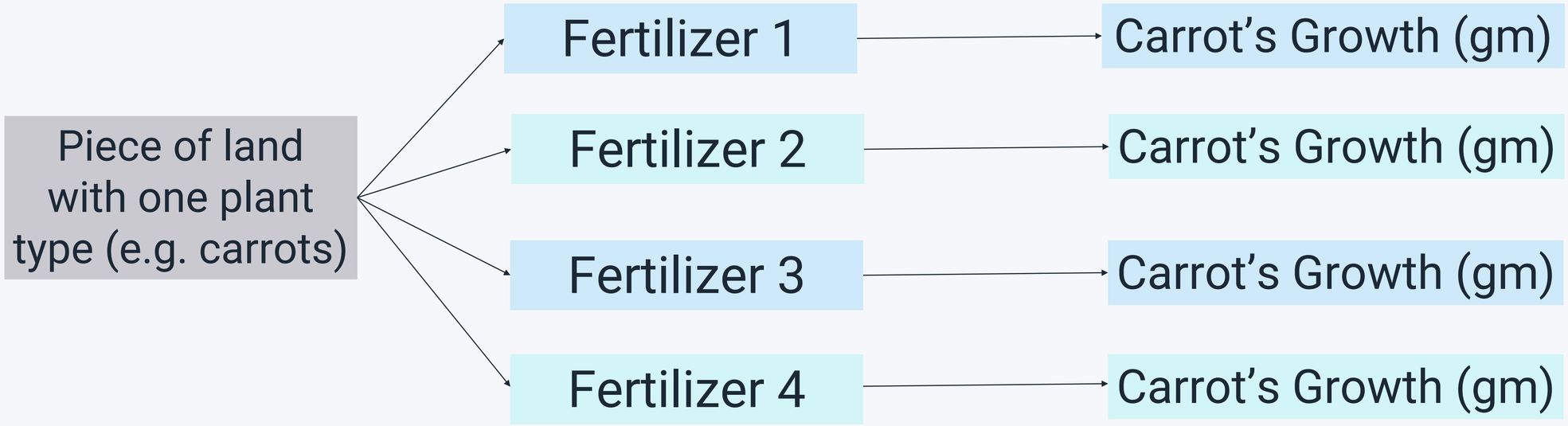
ARE SCIENTISTS HAPPIER THAN MANAGERS?

This is a case of **comparison of two groups → t-Test**



WHICH OF THE 4 FERTILIZERS WORK THE BEST?

This is a case of **comparison of more than 2 groups → ANOVA**

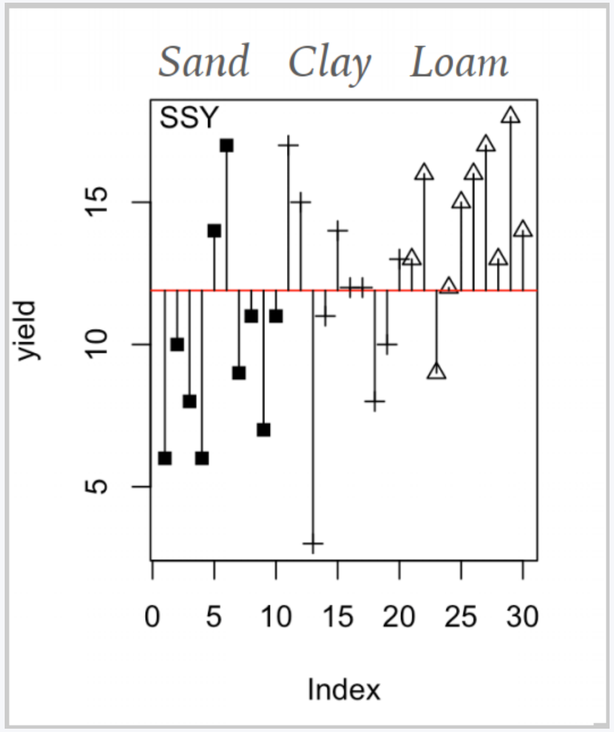


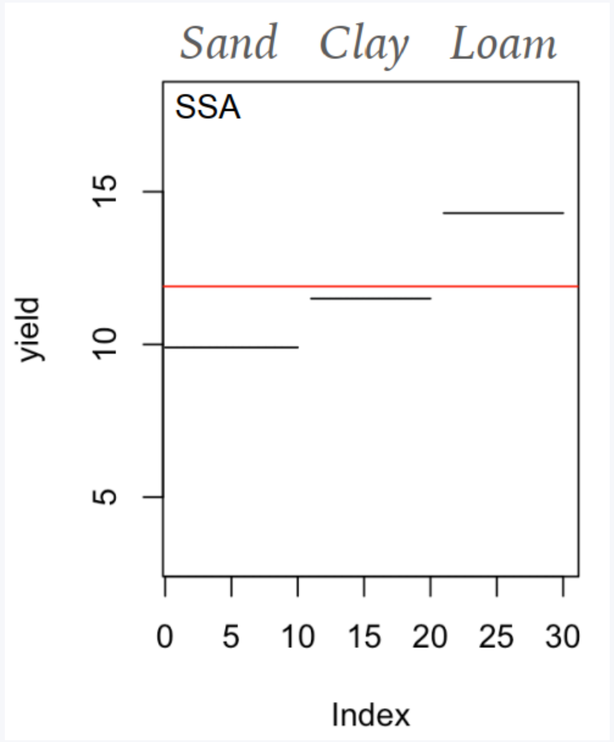
Example Source: Lisa Gotzian

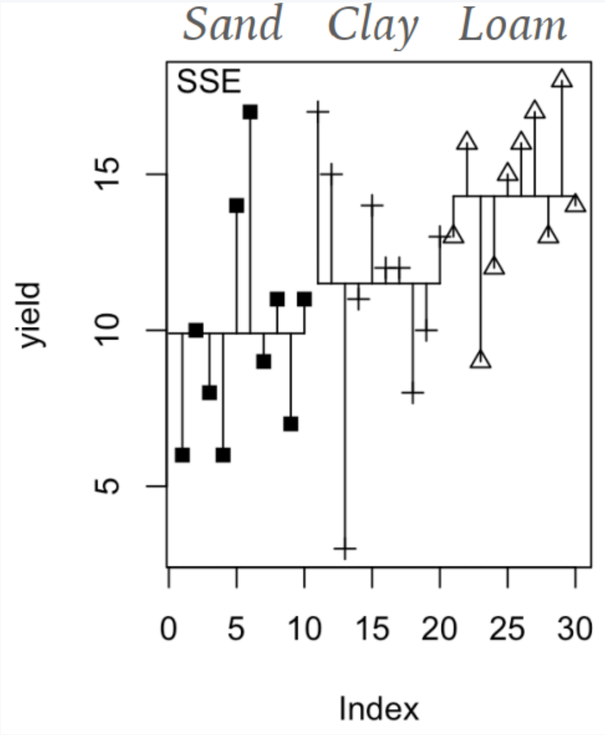
ANOVA Intuition

Total Variation (SSY) = **Explained Variation (SSA)** + **Unexplained Variation (SSE)**

data point – overall mean | treatment mean – overall mean | data points – treatment means







ANOVA Test statistic (F Ratio)

$$F - ratio = SSA/df_{treatment} \;/\; SSE/df_{residuals}$$

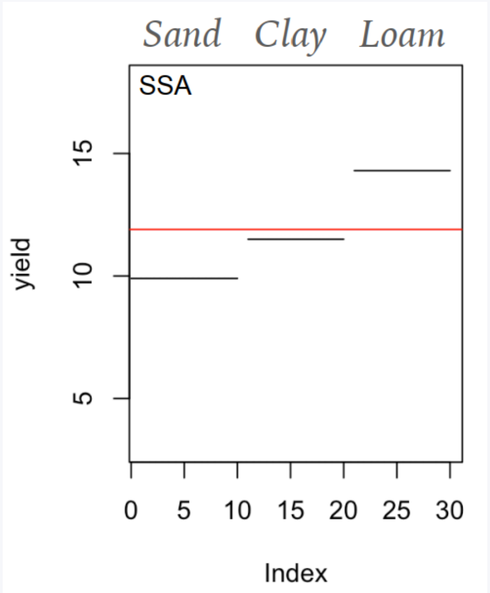
$= k - 1$ (for $SSA/df_{treatment}$)

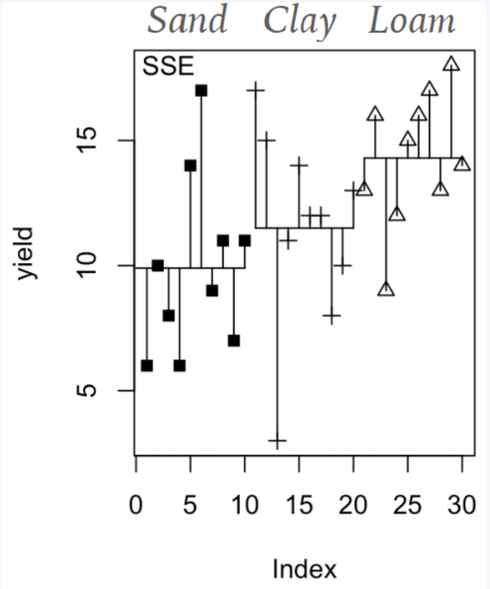
$= k * (n - 1)$ (for $SSE/df_{residuals}$)

test statistic

treatment mean – overall mean

data points – treatment means





ANOVA RESULT FROM R

```
> model <- lm(yield ~ soil)
> anova(model)
Analysis of Variance Table
Response: yield
          Df  Sum Sq  Mean Sq  F value  Pr(>F)
soil       2   99.2   49.600   4.2447   0.02495 *
Residuals 27  315.5   11.685   ---
Signif. codes: 0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1
```

Note:

SSA is also called treatment SS

SSE is also called error SS

$$F-ratio = SSA/df_{treatment} \;/\; SSE/df_{residuals}$$

test statistic

treatment mean – overall mean **(Soil)**

data points – treatment means **(Residuals)**

ANOVA RESOURCES

(Best Videos):

- One way: *https://www.youtube.com/watch?v=-yQb_ZJnFXw*
- Two way: *https://www.youtube.com/watch?v=cNIIn9bConY*

Literature: *https://newonlinecourses.science.psu.edu/stat502/node/137/*

SCIENTIFIC STUDIES AND PUBLISHING

John Oliver on Scientific Studies:
https://www.youtube.com/watch?v=0Rnq1NpHdmw

- Scientists can submit their research as "**papers**" to journals
- **Journals** are run by some publisher
- Papers are judged by independent **reviewers** and **editors** (other scientists from the same field) → "**peer reviewed**"
  - Journals accept/reject/ask you for a resubmission
  - Journals are ranked

STRUCTURE OF SCIENTIFIC PAPERS

- Different journals require different structures
- The structure loosely conforms to “IMRaD” structure:
  - Introduction – includes research question among other information
  - Methods – study design and methodology
  - Results and – the results of the study
  - Discussion – practical connotations, weaknesses of the study etc.
- Good resource from Prof. Joern Fisher from Leuphana: https://writingajournalarticle.wordpress.com/

PHALAN ET AL., 2011

Reconciling Food Production and Biodiversity Conservation: Land Sharing and Land Sparing Compared

Ben Phalan,[1] Malvika Onial,[1] Andrew Balmford,[1] Rhys E. Green[1,2]

The question of how to meet rising food demand at the least cost to biodiversity requires the evaluation of two contrasting alternatives: land sharing, which integrates both objectives on the same land; and land sparing, in which high-yield farming is combined with protecting natural habitats from conversion to agriculture. To test these alternatives, we compared crop yields and densities of bird and tree species across gradients of agricultural intensity in southwest Ghana and northern India. More species were negatively affected by agriculture than benefited from it, particularly among species with small global ranges. For both taxa in both countries, land sparing is a more promising strategy for minimizing negative impacts of food production, at both current and anticipated future levels of production.

Publication:

Published in ***Science***, one of the best scientific journals.

IMRaD:

- **I**ntroduction
- **M**ethods
- **R**esults **a**nd
- **D**iscussion

CLASS ACTIVITY

Go through the paper (MyStudy) and discuss with your neighbors:

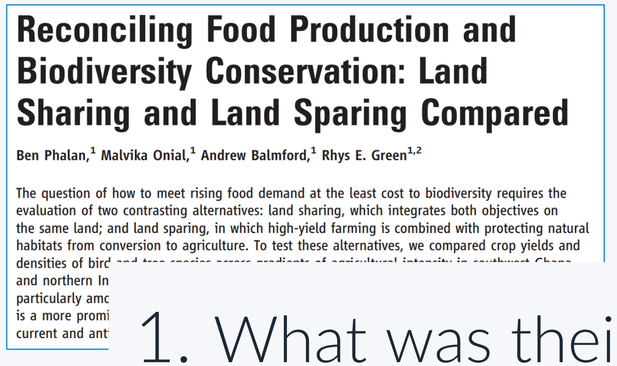
1. What was their research question?
2. What was their study design?
3. What statistical methods were used?
4. What was their result?

PHALAN ET AL., 2011

Reconciling Food Production and Biodiversity Conservation: Land Sharing and Land Sparing Compared

Ben Phalan,[1] Malvika Onial,[1] Andrew Balmford,[1] Rhys E. Green[1,2]

The question of how to meet rising food demand at the least cost to biodiversity requires the evaluation of two contrasting alternatives: land sharing, which integrates both objectives on the same land; and land sparing, in which high-yield farming is combined with protecting natural habitats from conversion to agriculture. To test these alternatives, we compared crop yields and densities of bird and tree species across gradients of agricultural intensity in southwest Ghana and northern India. More species were negatively affected by agriculture than benefited from it, particularly among species with small global ranges. For both taxa in both countries, land sparing is a more promising strategy for minimizing negative impacts of food production, at both current and anticipated future levels of production.



Reconciling Food Production and Biodiversity Conservation: Land Sharing and Land Sparing Compared

Ben Phalan,[1] Malvika Onial,[1] Andrew Balmford,[1] Rhys E. Green[1,2]

The question of how to meet rising food demand at the least cost to biodiversity requires the evaluation of two contrasting alternatives: land sharing, which integrates both objectives on the same land; and land sparing, in which high-yield farming is combined with protecting natural habitats from conversion to agriculture. To test these alternatives, we compared crop yields and densities of bird and tree species across gradients of agricultural intensity in southwest Ghana and northern In
particularly amc
is a more promi
current and ant

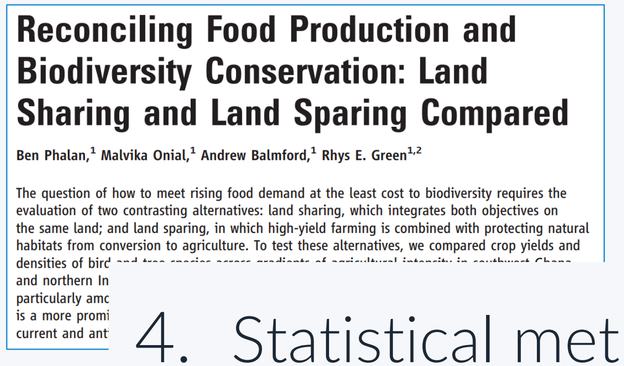
PHALAN ET AL. : IMRaD

1. What was their research question?

- **How does the use of the land impact biodiversity?**

2. What was their study design?

Sampling	**Explanatory variable** *(predictor)*	**Response variable**
No random assignment Land in SW Ghana & Northern India → 3 types of land	3 types of land (yield) 1. Land with “land sharing” 2. Land with “land sparing” 3. Land with conventional agriculture	Biodiversity (density) For each type of land: 1. Density of birds 2. Density of trees


Reconciling Food Production and Biodiversity Conservation: Land Sharing and Land Sparing Compared

Ben Phalan,[1] Malvika Onial,[1] Andrew Balmford,[1] Rhys E. Green[1,2]

The question of how to meet rising food demand at the least cost to biodiversity requires the evaluation of two contrasting alternatives: land sharing, which integrates both objectives on the same land; and land sparing, in which high-yield farming is combined with protecting natural habitats from conversion to agriculture. To test these alternatives, we compared crop yields and densities of bird and tree species across gradients of agricultural intensity in southwest Ghana and northern In particularly amc is a more promi current and ant

PHALAN ET AL. : IMRaD

4. Statistical methods?
    - How does the use of the land impact biodiversity?
5. Results and discussions?
    - What results did they get?
    - What were their interpretations? Did they mention anything else?

EXAM

1. **Question types**: multiple choice and (short) text
2. **Calculator required?** → no.
3. **Calculations required?** → yes.

EXAM: WHAT ABOUT R?

- Know main functions discussed so far:
  - Read data, select column, check data, create simple plots, calculate descriptive values (mean, sd, etc), perform tests, etc.
- **Cheat Sheet** for R
  - Hand-in deadline: 1 week before the exam
  - Typed on a computer (Font: Arial, 11pt, 1.5 spacing, margin default)
  - Cannot copy paste from help window
  - Will be laid out in alphabetical order of name on the day of the exam
    - Be there 30 minutes early to collect your sheets
- **Questions?**

PLAN FOR NEXT WEEK

That’s it for today! :-)

Next week, we are going to discuss:

1. We discuss other papers.
2. We do more exercises.

If you want to reach me, mail me at:

`prabesh.dhakal@stud.leuphana.de`