Inference concepts

DAAG Chapter 4

Learning objectives

- Point estimation
- Confidence intervals and hypothesis tests
- Contingency tables
- One-way and two-way comparisons, ANOVA
- Response curves
- Nested structures, pseudoreplication
- Maximum likelihood estimation
- Bayesian estimation

Inference

- Interested in population quantities
  - Parameters $\boldsymbol{\theta}$ (e.g. μ, σ^2)
- Collect sample **X**
- Use a sample statistic $\hat{\theta}(\boldsymbol{X})$ to estimate a population quantity θ
- The sampling distribution f_X implies $f_{\hat{\theta}|\theta}$
- We use $f_{\hat{\theta}|\theta}$ for inference about θ, or
- We use $f_{\theta|\hat{\theta}}$ for inference about θ (Bayesian)

Point estimation

- What is the population mean μ?
  - A point estimate of μ is the sample mean $\bar{x}$
- Look to the sampling distribution $f_{\bar{X}|\mu}$
  - According to CLT, $f_{\bar{X}|\mu} \sim N(\mu, \sigma^2/n)$
  - The standard error of the mean is thus $\sigma/\sqrt{n}$
  - Can approximate SEM ≈ $s/\sqrt{n}$
- The sampling distribution of $t = \frac{\bar{x}-\mu}{SEM}$ is $f_{t|\mu}$
  - Includes variability from $\bar{x}$ and s ≈ σ
  - t is the number of SEM units between $\bar{x}$ and μ

Hypothesis tests

- Use $f_{\hat{\theta}|\theta}$ for inference about θ
- In hypothesis testing,
  - Begin by assuming $\theta = \theta_0$ (null hypothesis)
  - What is the sampling distribution $f_{\hat{\theta}|\theta_0}$?
  - Imagine we sample from $f_{\hat{\theta}|\theta_0}$. What values are likely? What values are unlikely?
    - Our answer determines the rejection region of the test
  - Now, collect a sample and compute $\hat{\theta}_{obs}$
    - Is $\hat{\theta}_{obs}$ in the rejection region? Reject our initial hypothesis that $\theta = \theta_0$

Hypothesis tests

- How to decide what is an unlikely value?
  - Formulate an alternative hypothesis
    - $\theta > \theta_0$ or $\theta < \theta_0$ or $\theta \neq \theta_0$
  - Decide on a Type 1 error rate α (false rejection)
  - α, together with alternative hypothesis, implies a rejection region ("unlikely value")
- If we don't want to decide α, compute p-value
  - Smallest α that would result in rejection of null hypothesis

Confidence intervals

- Consider $f_{\hat{\theta}-\theta}$, the sampling distribution of $\hat{\theta}-\theta$
- Given a probability, (e.g. 95% or 99%) we can compute an interval for $\hat{\theta}-\theta$ from $f_{\hat{\theta}-\theta}$
- For μ, use $f_{\bar{X}-\mu} \sim N(0, \sigma^2/\mathrm{n})$ or

$$f_{(\bar{X}-\mu)/SEM} \sim t_{n-1}$$

- Results in confidence intervals for μ

$\bar{x} \pm z_{\alpha/2}\sigma/\sqrt{n}$ or $\bar{x} \pm t_{\frac{\alpha}{2},n-1}s/\sqrt{n}$

A short comment…

- Use hypothesis tests sparingly, and for good reason.
  - Multiple comparisons can result in false alarms
  - Ask directed questions
- Consider alternatives to hypothesis tests
  - They provide little or no information about θ
    - What is the probability of the null hypothesis?
  - Confidence intervals (or Bayesian posterior distributions) provide much more information
- Always report means (point estimates) and standard errors when reporting hypothesis tests

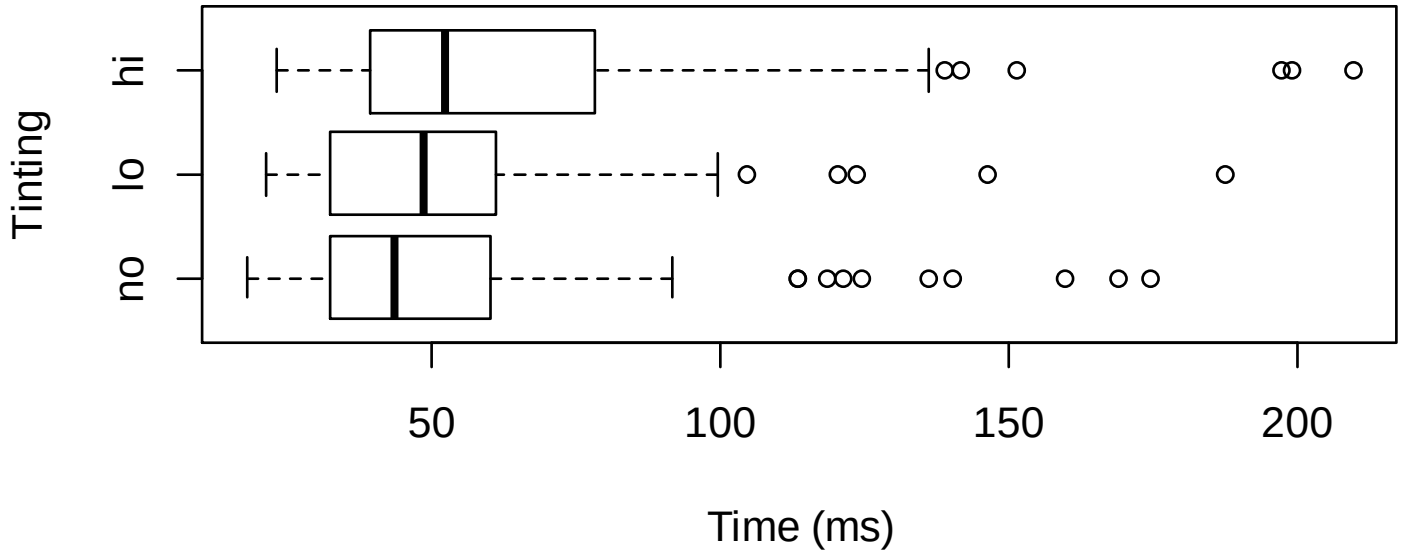
Contingency tables

- Comparing two or more categorical variables
- Common question: are the variables independent? Which categories have more or fewer units than expected?

	Men	Women	Totals
Brown Eyes	42	39	81 (81/174)
Blue Eyes	35	38	73 (73/174)
Other	12	8	20 (20/174)
Totals	89 (89/174)	75 (75/174)	174

One-way comparisons

- Data: tinting
- Experiment: time to discriminate a target for different window tinting levels



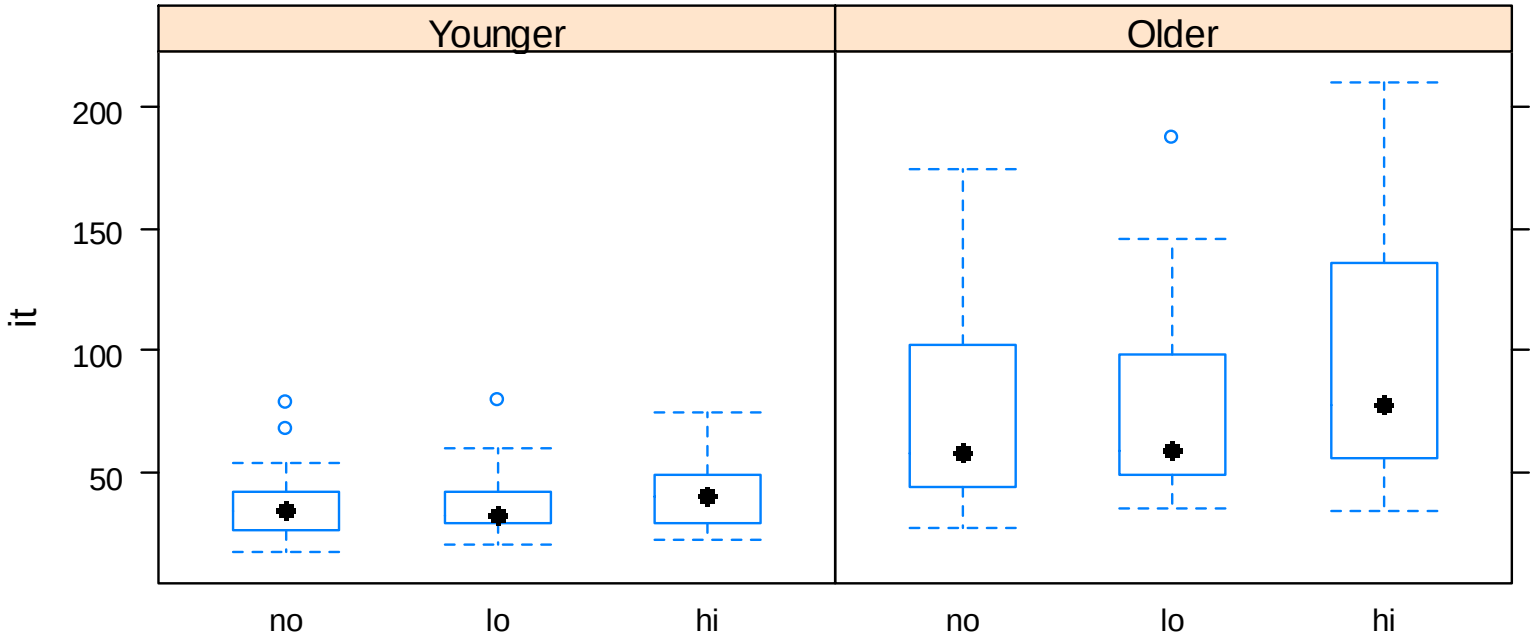
One way ANOVA

```
Analysis of Variance Table

Response: it
           Df Sum Sq Mean Sq F value Pr(>F)
tint        2   6597  3298.4  2.1769 0.1164
Residuals 179 271220  1515.2
```

Two-way comparisons

- There are other factors that might influence time to discriminate a target, e.g. age

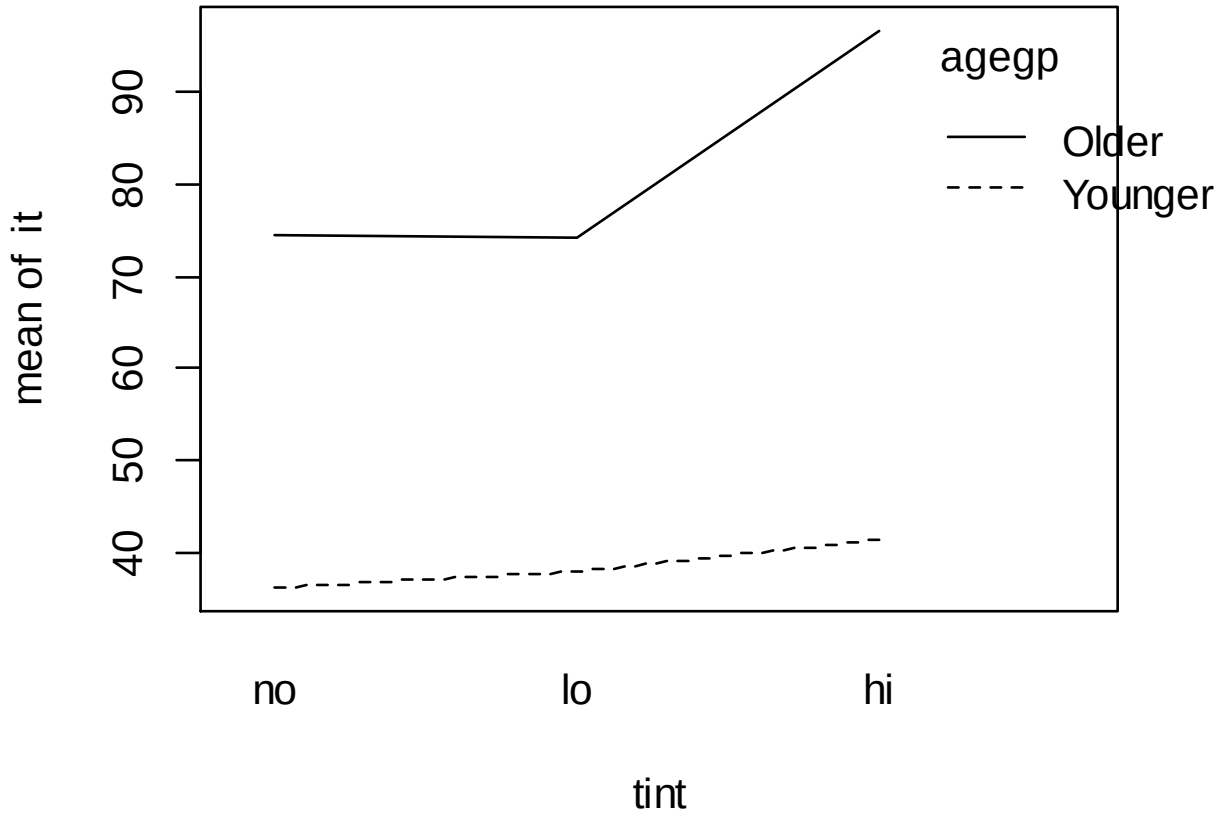


Two way ANOVA

```
Analysis of Variance Table

Response: it
           Df Sum Sq Mean Sq F value    Pr(>F)
tint        2   6597    3298  3.0965   0.04765 *
agegp       1  81612   81612 76.6164 1.567e-15 ***
Residuals 178 189607    1065
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’
0.1 ‘ ’ 1
```

Interaction plots



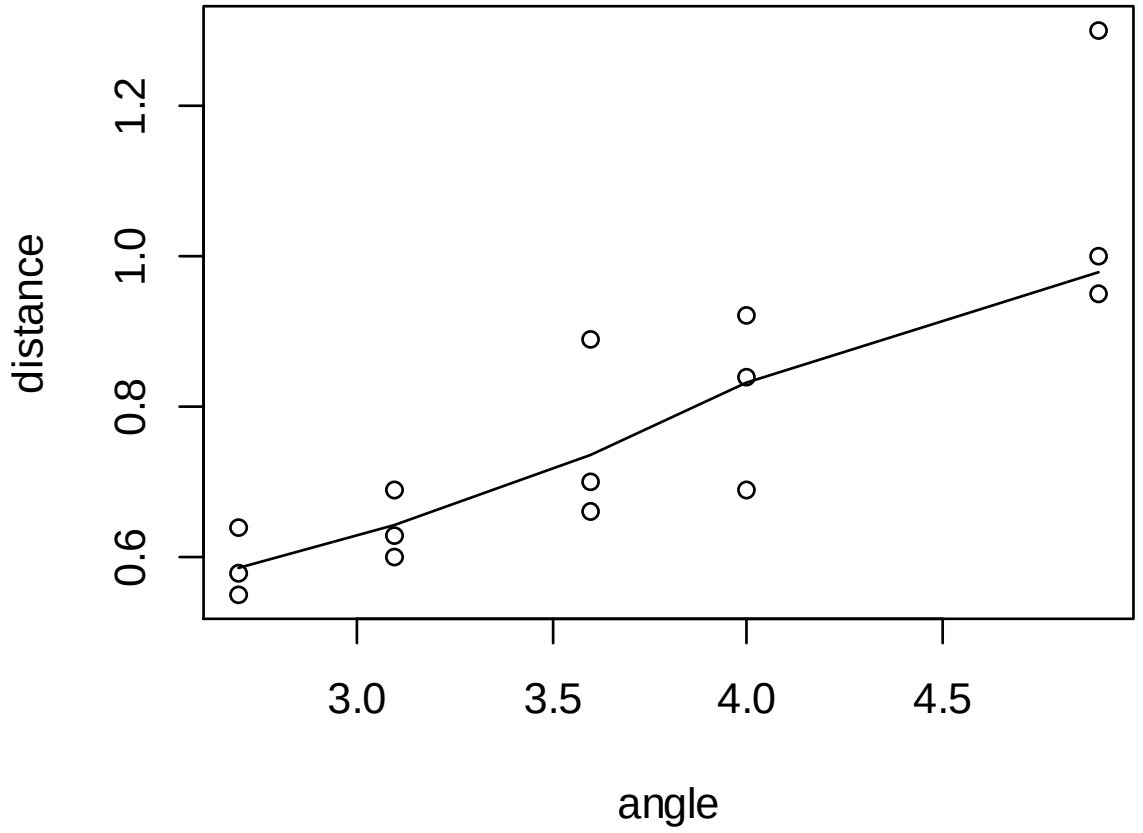
Two-way ANOVA: interaction

```
Analysis of Variance Table

Response: it
             Df Sum Sq Mean Sq F value    Pr(>F)
tint          2   6597    3298  3.1109   0.04702 *
agegp         1  81612   81612 76.9729 1.466e-15 ***
tint:agegp    2   2999    1499  1.4141   0.24590
Residuals   176 186609    1060
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’
0.1 ‘ ’ 1
```

Response curves

- Sometimes a response should be handled as a regression problem rather than ANOVA



Ecological Monographs 54(2), 1984, pp. 187-211

PSEUDOREPLICATION AND THE DESIGN OF ECOLOGICAL FIELD EXPERIMENTS

STUART H. HURLBERT
Department of Biology, San Diego State University, San Diego, California 92182 USA

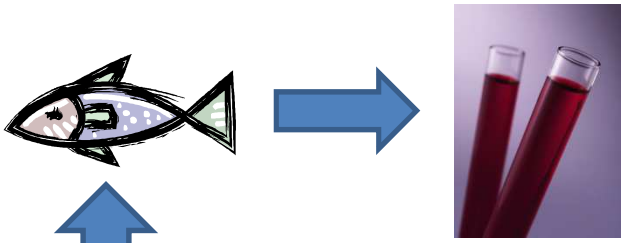
Abstract. Pseudoreplication is defined as the use of inferential statistics to test for treatment effects with data from experiments where either treatments are not replicated (though samples may be) or replicates are not statistically independent. In ANOVA terminology, it is the testing for treatment effects with an error term inappropriate to the hypothesis being considered. Scrutiny of *176* experimental studies published between 1960 and the present revealed that pseudoreplication occurred in *27%* of them, or 48% of all such studies that applied inferential statistics. The incidence of pseudoreplication is especially high in studies of marine benthos and small mammals. The critical features of controlled experimentation are reviewed. Nondemonic intrusion is defined as the impingement of chance events on an experiment in progress. As a safeguard against both it and preexisting gradients, interspersion of treatments is argued to be an obligatory feature of good design. Especially in small experiments, adequate interspersion can sometimes be assured only by dispensing with strict randomization procedures. Comprehension of this conflict between interspersion and randomization is aided by distinguishing pre-layout (or conventional) and layout-specific alpha (probability of type I error). Suggestions are offered to statisticians and editors of ecological journals as to how ecologists' understanding of experimental design and statistics might be improved.

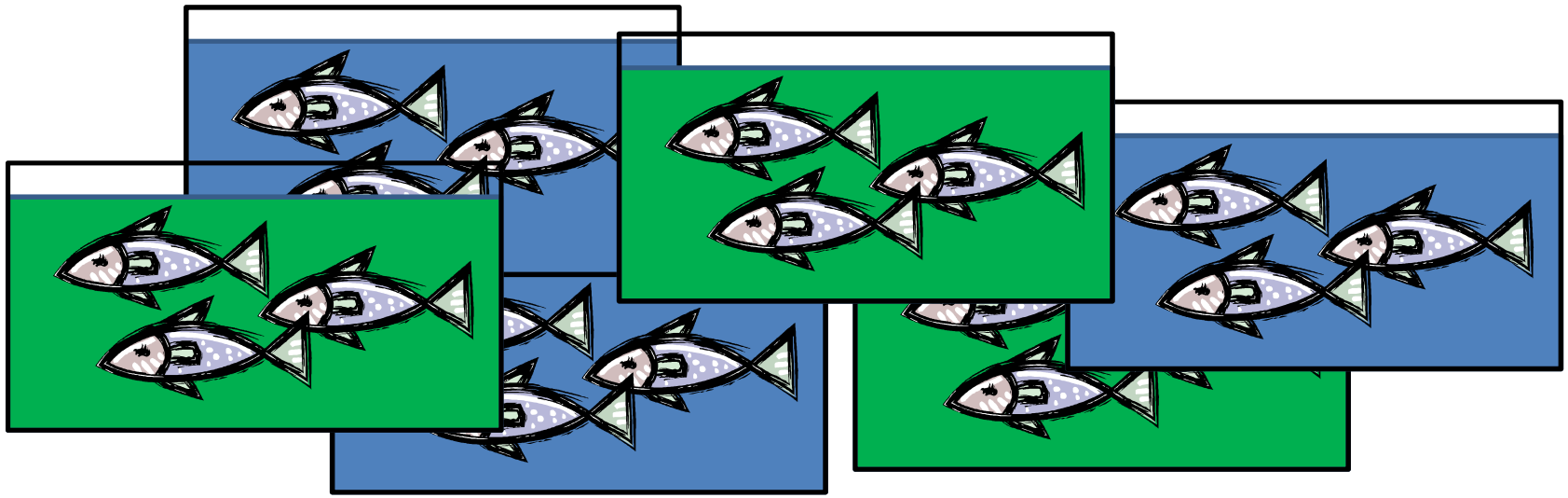
Key words: experimental design; chi-square; R. A. Fisher; W. S. Gossett; interspersion of treatments; nondemonic intrusion; randomization; replicability; type I error.

Nested structures

- If the scale of your effect doesn't match the scale of your experimental unit, don't pretend that it does.

Q: How many experimental units do we have for comparing treatment to control?





Maximum likelihood estimation

- Likelihood is the probability of data $\boldsymbol{X}$ given a population, parameterized by $\boldsymbol{\theta}$
- The value of $\boldsymbol{\theta}$ that maximizes the likelihood is the maximum likelihood estimate $\widehat{\boldsymbol{\theta}_{ML}}$.

$$y_i = \mu + \varepsilon_i,\ \varepsilon_i \sim N(0, \sigma^2),\ \ i = 1,2, \dots, n$$

$$L(\boldsymbol{y}; \mu, \sigma^2) = \prod_{i=1}^{n} \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{(y_i-\mu)^2}{2\sigma^2}}$$

$$l(\boldsymbol{y}; \mu, \sigma^2) = -\frac{1}{2}\log(2\pi\sigma^2) - \sum_{i=1}^{n} \frac{(y_i - \mu)^2}{2\sigma^2}$$

Bayesian estimation

$$P(\boldsymbol{\theta}|\boldsymbol{X}) = \frac{P(\boldsymbol{X}|\boldsymbol{\theta})P(\boldsymbol{\theta})}{P(\boldsymbol{X})}$$

It is often difficult to get $P(\boldsymbol{X})$ directly, but $P(\boldsymbol{X})$ is just a normalizing constant

$$P(\boldsymbol{\theta}|\boldsymbol{X}) \propto P(\boldsymbol{X}|\boldsymbol{\theta})P(\boldsymbol{\theta})$$

so use various tricks to generate samples from

$$P(\boldsymbol{X}|\boldsymbol{\theta})P(\boldsymbol{\theta})$$

The most popular trick is MCMC