

Statistics 191:
Introduction
to Applied
Statistics

Jonathan
Taylor
Department of
Statistics
Stanford
University

Statistics 191: Introduction to Applied Statistics

Qualitative Variables

Jonathan Taylor
Department of Statistics
Stanford University

February 17, 2009

Qualitative variables + interactions

Statistics 191:
Introduction
to Applied
Statistics

Jonathan
Taylor
Department of
Statistics
Stanford
University

Outline

- Qualitative / categorical variables.
- Regression equations differing by group.
- Interactions.
- Analysis of Variance Models

Categorical variables

Statistics 191:
Introduction
to Applied
Statistics

Jonathan
Taylor
Department of
Statistics
Stanford
University

Categorical variables

- Most variables we have looked at so far were continuous: height, rating, etc.
- In many situations, we record a categorical variable: sex, state, country, etc.
- How do we include this in our model?

Categorical variables

Statistics 191:
Introduction
to Applied
Statistics

Jonathan
Taylor
Department of
Statistics
Stanford
University

A simple example

- One example that we have looked at *does* have categorical variables.
- Two sample problem with equal variances: suppose

$$Y = (Z_1, \dots, Z_m, W_1, \dots, W_n)$$

with $Z_j \sim N(\mu_1, \sigma^2), 1 \leq j \leq m$ and
 $W_j \sim N(\mu_2, \sigma^2), 1 \leq j \leq n + m.$

- For $1 \leq i \leq n$, let

$$X_i = \begin{cases} 1 & 1 \leq i \leq m \\ 0 & \text{otherwise.} \end{cases}$$

Categorical variables

Statistics 191:
Introduction
to Applied
Statistics

Jonathan
Taylor
Department of
Statistics
Stanford
University

A simple example

- Design matrix

$$X_{(n+m) \times 2} = \begin{pmatrix} 1 & 1 \\ \vdots & \vdots \\ 1 & 1 \\ 1 & 0 \\ \vdots & \vdots \\ 1 & 0 \end{pmatrix}$$

Example

Statistics 191:
Introduction
to Applied
Statistics

Jonathan
Taylor
Department of
Statistics
Stanford
University

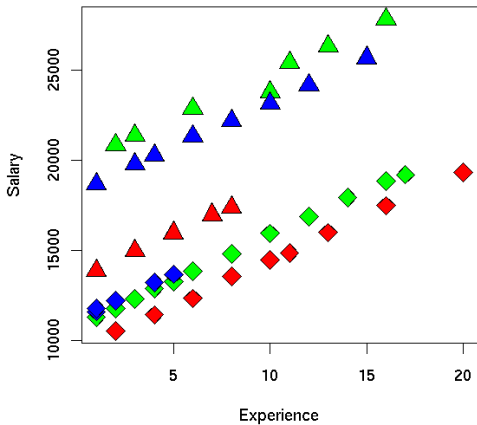
IT salary data

- Outcome: S , salaries for IT staff in a corporation.
- Predictors: X , experience (years); E , education (3 levels): 1=Bachelor's, 2=Master's, 3=Ph.D; M , management (2 levels): 1=management, 0=not management.

IT salary

Statistics 191:
Introduction
to Applied
Statistics

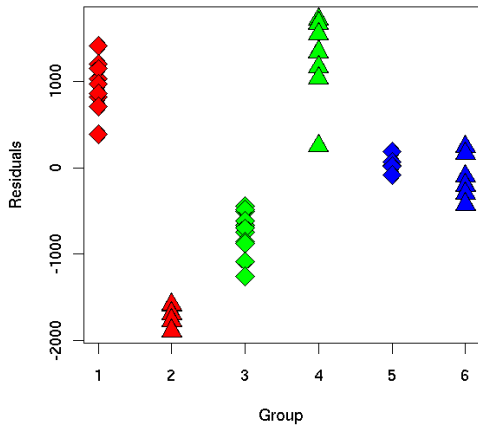
Jonathan
Taylor
Department of
Statistics
Stanford
University



IT salary: residuals

Statistics 191:
Introduction
to Applied
Statistics

Jonathan
Taylor
Department of
Statistics
Stanford
University



Two solutions

Statistics 191:
Introduction
to Applied
Statistics

Jonathan
Taylor
Department of
Statistics
Stanford
University

Solution #1: stratification

- One solution is to “stratify” data set by this categorical variable.
- We could break data set up into groups by education and management, and fit fit model

$$S_i = \beta_0 + \beta_1 X_i + \varepsilon_i$$

in each group.

- Problem: this results in smaller samples in each group: lose degrees of freedom for estimating σ^2 within each group.

Two solutions

Statistics 191:
Introduction
to Applied
Statistics

Jonathan
Taylor
Department of
Statistics
Stanford
University

Solution #2: qualitative predictors

- IF it is reasonable to assume that σ^2 is constant for each observation.
- THEN, we can incorporate all observations into 1 model.

$$S_i = \beta_0 + \beta_1 X_i + \beta_2 E_{i2} + \beta_3 E_{i3} + \beta_4 M_i + \varepsilon_i$$

where

$$E_{i2} = \begin{cases} 1 & \text{if } E_i = 2, \\ 0 & \text{otherwise.} \end{cases}, E_{i3} = \begin{cases} 1 & \text{if } E_i = 3, \\ 0 & \text{otherwise,} \end{cases}$$

Categorical variables: details

Statistics 191:
Introduction
to Applied
Statistics

Jonathan
Taylor
Department of
Statistics
Stanford
University

Things to notice

- Although E has 3 levels, we only added 2 variables to the model. In a sense, this is because “intercept” absorbs one level.
- If we added three variables then the columns of design matrix would be linearly dependent.
- Assumes β_1 – effect of experience is the same in all groups, unlike when we fit the model separately. This may or may not be reasonable.

Interactions

Statistics 191:
Introduction
to Applied
Statistics

Jonathan
Taylor
Department of
Statistics
Stanford
University

Effect of experience

- Our model has enforced the constraint the β_1 is the same within each group.
- Graphically, this seems OK, but how can we “test” this?
- We could fit a model with different slopes in each group, but keeping as many d.f. as we can.
- This model has “interactions” in it: the effect of experience depends on what level of education you have.

Interactions

Statistics 191:
Introduction
to Applied
Statistics

Jonathan
Taylor
Department of
Statistics
Stanford
University

Interaction between experience and education

- Model:

$$S_i = \beta_0 + \beta_1 X_i + \beta_2 E_{i2} + \beta_3 E_{i3} + \beta_4 M_i \\ + \beta_5 E_{i2} X_i + \beta_6 E_{i3} X_i + \varepsilon_i.$$

- Note that we took each column corresponding to education and multiplied it by the column for experience to get two new predictors.
- To test whether the slope is the same in each group we would just test $H_0 : \beta_5 = \beta_6 = 0$.
- Based on figure, we expect not to reject H_0 .

Interactions

Statistics 191:
Introduction
to Applied
Statistics

Jonathan
Taylor
Department of
Statistics
Stanford
University

Interaction between management and education

- Based on figure, we expect an interaction effect.
- Fit model

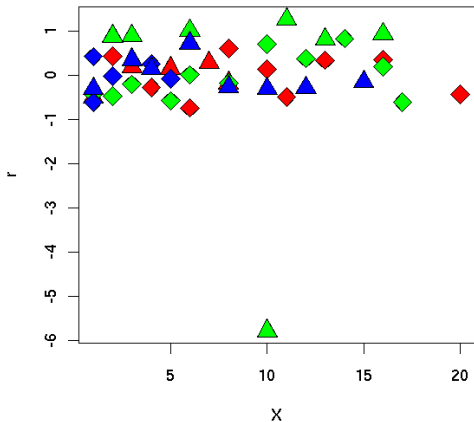
$$S_i = \beta_0 + \beta_1 X_i + \beta_2 E_{i2} + \beta_3 E_{i3} + \beta_4 M_i \\ + \beta_5 E_{i2} M_i + \beta_6 E_{i3} M_i + \varepsilon_i.$$

- Again, testing for interaction is testing $H_0 : \beta_5 = \beta_6 = 0$.

IT salary: outlier

Statistics 191:
Introduction
to Applied
Statistics

Jonathan
Taylor
Department of
Statistics
Stanford
University



Example

Statistics 191:
Introduction
to Applied
Statistics

Jonathan
Taylor
Department of
Statistics
Stanford
University

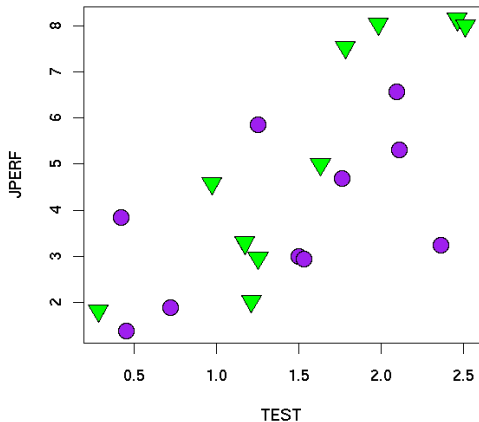
Minority employment data

<i>TEST</i>	job aptitude test score
<i>RACE</i>	1 if minority, 0 otherwise
<i>JPERF</i>	job performance evaluation

Minority employment data

Statistics 191:
Introduction
to Applied
Statistics

Jonathan
Taylor
Department of
Statistics
Stanford
University



Interactions

Statistics 191:
Introduction
to Applied
Statistics

Jonathan
Taylor
Department of
Statistics
Stanford
University

General model

- In theory, there may be a linear relationship between $JPERF$ and $TEST$ but it could be different by group.
- Model:

$$JPERF_i = \beta_0 + \beta_1 TEST_i + \beta_2 RACE_i + \beta_3 RACE_i * TEST_i + \varepsilon_i.$$

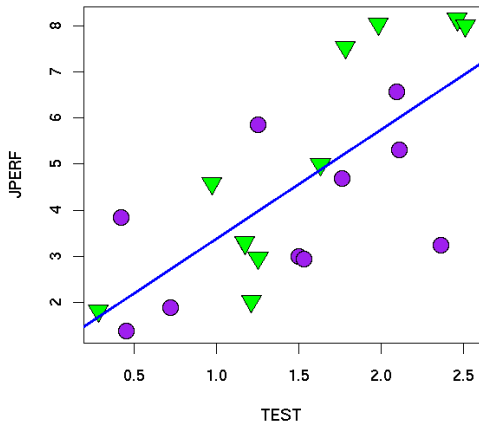
- Regression functions:

$$Y_i = \begin{cases} \beta_0 + \beta_1 TEST_i + \varepsilon_i & \text{if } i\text{-th ind. is white} \\ (\beta_0 + \beta_2) + (\beta_1 + \beta_3) TEST_i + \varepsilon_i & \text{if } i\text{-th ind. is black.} \end{cases}$$

No difference

Statistics 191:
Introduction
to Applied
Statistics

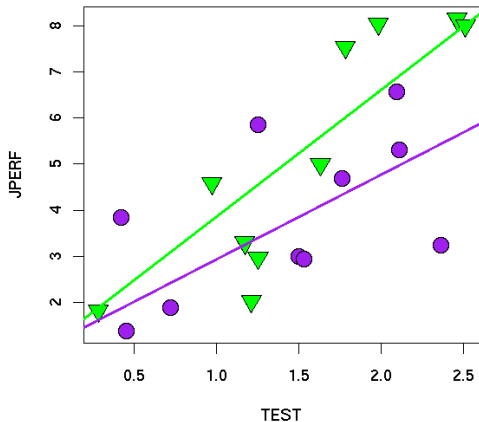
Jonathan
Taylor
Department of
Statistics
Stanford
University



Different slopes, same intercept

Statistics 191:
Introduction
to Applied
Statistics

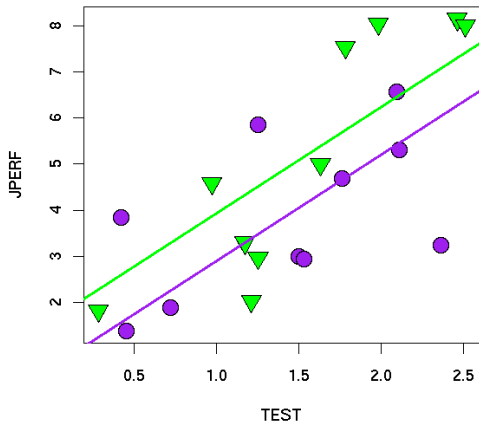
Jonathan
Taylor
Department of
Statistics
Stanford
University



Different intercepts, same slope

Statistics 191:
Introduction
to Applied
Statistics

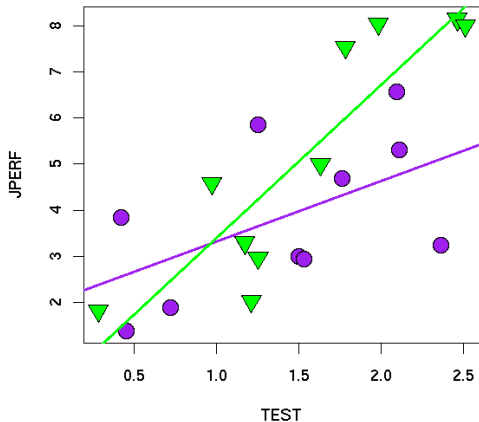
Jonathan
Taylor
Department of
Statistics
Stanford
University



Different intercepts, different slopes

Statistics 191:
Introduction
to Applied
Statistics

Jonathan
Taylor
Department of
Statistics
Stanford
University



Interactions

Statistics 191:
Introduction
to Applied
Statistics

Jonathan
Taylor
Department of
Statistics
Stanford
University

Interpreting different models

- Both $\beta_2, \beta_3 \neq 0$ – main effect for *RACE* and interaction effect between *TEST* and *RACE*.
- $\beta_2 \neq 0, \beta_3 = 0$ – main effect for *RACE*, no interaction between *TEST* and *RACE*.
- $\beta_2 = 0, \beta_3 \neq 0$ – no main effect for *RACE*, interaction between *TEST* and *RACE*.
- [R code](#)

ANOVA models

Statistics 191:
Introduction
to Applied
Statistics

Jonathan
Taylor
Department of
Statistics
Stanford
University

General definition of ANOVA model

- Models with only qualitative variables.
- Can be thought of as extensions of “two-sample” t -test to more than two groups at once, and more than one grouping variable.
- Example: in a simple experiment studying blood pressure we might start by considering only the overall health (Poor, Moderate, Good).
- Data would then have one categorical variable with three levels.

ANOVA models

Statistics 191:
Introduction
to Applied
Statistics

Jonathan
Taylor
Department of
Statistics
Stanford
University

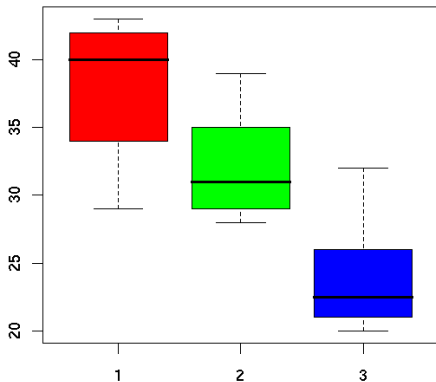
Example: rehab surgery

- How does prior fitness affect recovery from surgery?
Observations: 24 subjects' recovery time.
- Three fitness levels: below average, average, above average.
- If you are in better shape before surgery, does it take less time to recover?

Rehab time

Statistics 191:
Introduction
to Applied
Statistics

Jonathan
Taylor
Department of
Statistics
Stanford
University



ANOVA models

Statistics 191:
Introduction
to Applied
Statistics

Jonathan
Taylor
Department of
Statistics
Stanford
University

One-way ANOVA

- First generalization of two sample t -test: more than one level.
- One-way ANOVA model: observations:
 $Y_{ij}, 1 \leq i \leq r, 1 \leq j \leq n_i$: r groups and n_i samples in i -th group.

$$Y_{ij} = \mu + \alpha_i + \varepsilon_{ij}, \quad \varepsilon_{ij} \sim N(0, \sigma^2).$$

- Constraint: $\sum_{i=1}^r \alpha_i = 0$. This constraint is needed for “identifiability”. This is “equivalent” to only adding $r - 1$ columns to the design matrix for this qualitative variable.

ANOVA models

Statistics 191:
Introduction
to Applied
Statistics

Jonathan
Taylor
Department of
Statistics
Stanford
University

One-way ANOVA

- Model is easy to fit:

$$\hat{Y}_{ij} = \frac{1}{n_i} \sum_{j=1}^{n_i} Y_{ij} = \bar{Y}_{i..}$$

If observation is in i -th group: predicted mean is just the sample mean of observations in i -th group.

- Simplest question: is there any group (main) effect?

$$H_0 : \alpha_1 = \dots = \alpha_r = 0?$$

- Test is based on F -test with full model vs. reduced model. Reduced model just has an intercept.
- Other questions: is the effect the same in groups 1 and 2?

ANOVA models

Statistics 191:
Introduction
to Applied
Statistics

Jonathan
Taylor
Department of
Statistics
Stanford
University

ANOVA table: One-way

Source	SS	df	E(MS)
Treatments	$SSTR = \sum_{i=1}^r n_i (\bar{Y}_{i.} - \bar{Y}_{..})^2$	$r - 1$	$\sigma^2 + \frac{\sum_{i=1}^r n_i \alpha_i^2}{r-1}$
Error	$SSE = \sum_{i=1}^r \sum_{j=1}^{n_i} (Y_{ij} - \bar{Y}_{i.})^2$	$\sum_{i=1}^r n_i - r$	σ^2

- Note that $MSTR$ measures “variability” of the “cell” means. If there is a group effect we expect this to be large relative to MSE .
- We see that under $H_0 : \alpha_1 = \dots = \alpha_r = 0$, the expected value of $MSTR$ and MSE is σ^2 . This tells us how to test H_0 using ratio of mean squares, i.e. an F test.

ANOVA models

Statistics 191:
Introduction
to Applied
Statistics

Jonathan
Taylor
Department of
Statistics
Stanford
University

Testing for any main effect

- Rows in the ANOVA table are, in general, independent.
- Therefore, under H_0

$$F = \frac{MSTR}{MSE} = \frac{\frac{SSTR}{df_{TR}}}{\frac{SSE}{df_E}} \sim F_{df_{TR}, df_E}$$

the degrees of freedom come from the df column in previous table.

- Reject H_0 at level α if $F > F_{1-\alpha, df_{TR}, df_E}$.

ANOVA models

Statistics 191:
Introduction
to Applied
Statistics

Jonathan
Taylor
Department of
Statistics
Stanford
University

Inference for linear combinations

- Suppose we want to “infer” something about

$$\sum_{i=1}^r a_i \mu_i$$

where $\mu_i = \mu + \alpha_i$ is the mean in the i -th group. For example:

$$H_0 : \mu_1 - \mu_2 = 0 \quad (\text{same as } H_0 : \alpha_1 - \alpha_2 = 0)?$$

Is there a difference between below average and average groups in terms of rehab time?

-

$$\text{Var} \left(\sum_{i=1}^r a_i \bar{Y}_i \right) = \sigma^2 \sum_{i=1}^r \frac{a_i^2}{n_i}.$$

ANOVA models

Statistics 191:
Introduction
to Applied
Statistics

Jonathan
Taylor
Department of
Statistics
Stanford
University

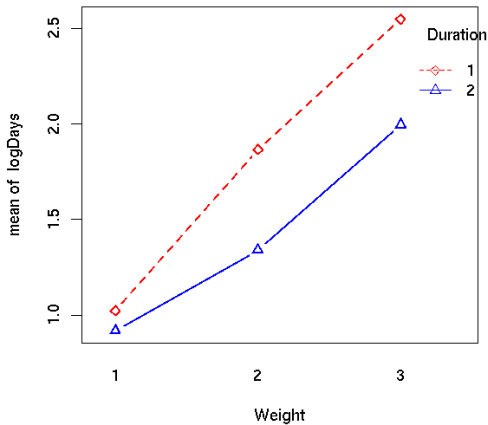
Two categorical variables: kidney failure

- Time of stay in hospital depends on weight gain between treatments and duration of treatment.
- Two levels of duration, three levels of weight gain.
- Is there an interaction? Main effects?

Kidney failure

Statistics 191:
Introduction
to Applied
Statistics

Jonathan
Taylor
Department of
Statistics
Stanford
University



ANOVA models

Statistics 191:
Introduction
to Applied
Statistics

Jonathan
Taylor
Department of
Statistics
Stanford
University

Two-way ANOVA

- Second generalization: more than one grouping variable.
- Two-way ANOVA model: observations:
 $(Y_{ijk}), 1 \leq i \leq r, 1 \leq j \leq m, 1 \leq k \leq n_{ij}$: r groups in first grouping variable, m groups in second and n_{ij} samples in (i, j) - "cell":

$$Y_{ijk} = \mu + \alpha_i + \beta_j + (\alpha\beta)_{ij} + \varepsilon_{ijk}, \quad \varepsilon_{ijk} \sim N(0, \sigma^2).$$

- In kidney example, $r = 3$ (weight gain), $m = 2$ (duration of treatment), $n_{ij} = 10$ for all (i, j) .

ANOVA model

Statistics 191:
Introduction
to Applied
Statistics

Jonathan
Taylor
Department of
Statistics
Stanford
University

Two-way ANOVA: main questions of interest

- Are there main effects for the grouping variables?

$$H_0 : \alpha_1 = \cdots = \alpha_r = 0, \quad H_0 : \beta_1 = \cdots = \beta_m = 0.$$

- Are there interaction effects:

$$H_0 : (\alpha\beta)_{ij} = 0, 1 \leq i \leq r, 1 \leq j \leq m.$$

ANOVA models

Statistics 191:
Introduction
to Applied
Statistics

Jonathan
Taylor
Department of
Statistics
Stanford
University

Constraints on the parameters

- Many constraints are needed, again for identifiability. Let's not worry about the details ...
- Constraints:
 - $\sum_{i=1}^r \alpha_i = 0$
 - $\sum_{j=1}^m \beta_j = 0$
 - $\sum_{j=1}^m (\alpha\beta)_{ij} = 0, 1 \leq i \leq r$
 - $\sum_{i=1}^r (\alpha\beta)_{ij} = 0, 1 \leq j \leq m.$

ANOVA models

Statistics 191:
Introduction
to Applied
Statistics

Jonathan
Taylor
Department of
Statistics
Stanford
University

Fitting model

- Easy to fit:

$$\hat{Y}_{ijk} = \bar{Y}_{ij.} = \frac{1}{n_{ij}} \sum_{k=1}^{n_{ij}} Y_{ijk}.$$

- Inference for combinations

$$\text{Var} \left(\sum_{i=1}^r \sum_{j=1}^m a_{ij} \bar{Y}_{ij.} \right) = \sigma^2 \cdot \sum_{i=1}^r \sum_{j=1}^m \frac{a_{ij}^2}{n_{ij}}.$$

- Usual t -tests, confidence intervals.

ANOVA models

Statistics 191:
Introduction
to Applied
Statistics

Jonathan
Taylor
Department of
Statistics
Stanford
University

ANOVA table: Two-way (assuming $n_{ij} = n$)

Term	SS
A	$SSA = nm \sum_{i=1}^r (\bar{Y}_{i..} - \bar{Y}_{...})^2$
B	$SSB = nr \sum_{j=1}^m (\bar{Y}_{.j.} - \bar{Y}_{...})^2$
AB	$SSAB = n \sum_{i=1}^r \sum_{j=1}^m (\bar{Y}_{ij.} - \bar{Y}_{i..} - \bar{Y}_{.j.} + \bar{Y}_{...})^2$
Error	$SSE = \sum_{i=1}^r \sum_{j=1}^m \sum_{k=1}^n (Y_{ijk} - \bar{Y}_{ij.})^2$

ANOVA models

Statistics 191:
Introduction
to Applied
Statistics

Jonathan
Taylor
Department of
Statistics
Stanford
University

ANOVA table: Two-way (assuming $n_{ij} = n$)

<i>SS</i>	<i>df</i>	<i>E(MS)</i>
<i>SSA</i>	$r - 1$	$\sigma^2 + nm \frac{\sum_{i=1}^r \alpha_i^2}{r-1}$
• <i>SSB</i>	$m - 1$	$\sigma^2 + nr \frac{\sum_{j=1}^m \beta_j^2}{r-1}$
<i>SSAB</i>	$(m - 1)(r - 1)$	$\sigma^2 + n \frac{\sum_{i=1}^r \sum_{j=1}^m (\alpha\beta)_{ij}^2}{(r-1)(m-1)}$
<i>SSE</i>	$(n - 1)mr$	σ^2

- For instance, we see that under $H_0 : (\alpha\beta)_{ij} = 0, \forall i, j$ the expected value of *SSAB* and *SSE* is σ^2 – use these for an *F*-test testing for an interaction.

Fixed and random effects

Statistics 191:
Introduction
to Applied
Statistics

Jonathan
Taylor
Department of
Statistics
Stanford
University

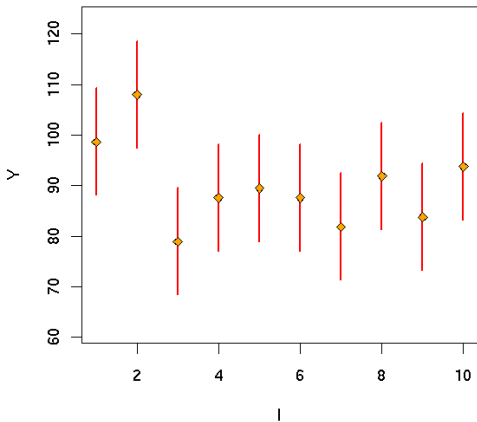
Random effects

- In kidney & rehab examples, the categorical variables are well-defined categories: below average fitness, long duration, etc.
- In some designs, the categorical variable is “subject”.
- Simplest example: repeated measures, where more than one (identical) measurement is taken on the same individual.
- In this case, the “group” effect α_i is best thought of as random because we only sample a subset of the entire population.

Simulated example, $n = 5, r = 10$

Statistics 191:
Introduction
to Applied
Statistics

Jonathan
Taylor
Department of
Statistics
Stanford
University



Fixed and random effects

Statistics 191:
Introduction
to Applied
Statistics

Jonathan
Taylor
Department of
Statistics
Stanford
University

When to use random effects?

- A “group” effect is random if we can think of the levels we observe in that group to be samples from a larger population.
- Example: if collecting data from different medical centers, “center” might be thought of as random.
- Example: if surveying students on different campuses, “campus” may be a random effect.

Fixed and random effects

Statistics 191:
Introduction
to Applied
Statistics

Jonathan
Taylor
Department of
Statistics
Stanford
University

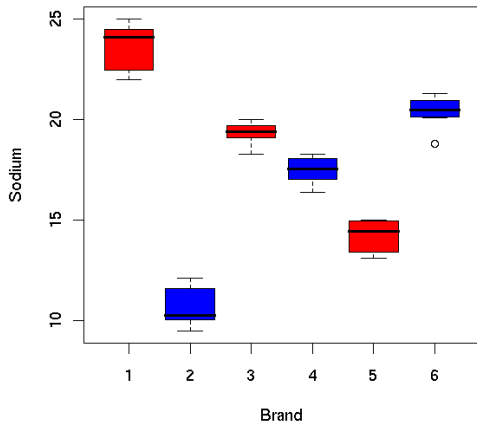
Example: sodium content in beer

- How much sodium is there in North American beer? How much does this vary by brand?
- Observations: for 6 brands of beer, we recorded the sodium content of 8 12 ounce bottles.
- Questions of interest: what is the “grand mean” sodium content? How much variability is there from brand to brand?
- “Individuals” in this case are brands, repeated measures are the 8 bottles.

Sodium content

Statistics 191:
Introduction
to Applied
Statistics

Jonathan
Taylor
Department of
Statistics
Stanford
University



One-way ANOVA (random)

Statistics 191:
Introduction
to Applied
Statistics

Jonathan
Taylor
Department of
Statistics
Stanford
University

One-way random effects model

- Assuming that cell-sizes are the same, i.e. equal observations for each “subject” (brand of beer).
- Observations

$$Y_{ij} \sim \mu. + \alpha_i + \varepsilon_{ij}, 1 \leq i \leq r, 1 \leq j \leq n$$

- $\varepsilon_{ij} \sim N(0, \sigma^2), 1 \leq i \leq r, 1 \leq j \leq n$
- $\alpha_i \sim N(0, \sigma_\mu^2), 1 \leq i \leq r.$
- Parameters:
 - μ is the population mean;
 - σ^2 is the measurement variance (i.e. how variable are the readings from the machine that reads the sodium content?);
 - σ_μ^2 is the population variance (i.e. how variable is the sodium content of beer across brands).

One-way ANOVA (random)

Statistics 191:
Introduction
to Applied
Statistics

Jonathan
Taylor
Department of
Statistics
Stanford
University

Implications for model

- In random effects model, the observations are no longer independent (even if ε 's are independent

$$\text{Cov}(Y_{ij}, Y_{i'j'}) = \sigma_{\mu}^2 \delta_{i,i'} + \sigma^2 \delta_{j,j'}.$$

- In more complicated models, this makes “maximum likelihood estimation” more complicated: least squares is no longer the best solution.
- Also changes the degrees of freedom for some t -statistics.

One-way ANOVA (random)

Statistics 191:
Introduction
to Applied
Statistics

Jonathan
Taylor
Department of
Statistics
Stanford
University

Fitting the model

- Only one parameter in the mean function $\mu_{..}$.
- When cell sizes are the same (balanced),

$$\hat{\mu}_{..} = \bar{Y}_{..} = \frac{1}{nr} \sum_{i,j} Y_{ij}.$$

- Unbalanced models: slightly more tricky.
- This also changes estimates of σ^2 – see ANOVA table below. We might guess that $df = nr - 1$ and

$$\hat{\sigma}^2 = \frac{1}{nr - 1} \sum_{i,j} (Y_{ij} - \bar{Y}_{..})^2.$$

This is *not* the case.

One-way ANOVA (random)

Statistics 191:
Introduction
to Applied
Statistics

Jonathan
Taylor
Department of
Statistics
Stanford
University

ANOVA table

Source	SS	df	E(MS)
Treatments	$SSTR = \sum_{i=1}^r n (\bar{Y}_{i.} - \bar{Y}_{..})^2$	$r - 1$	$\sigma^2 + n\sigma_{\mu}^2$
Error	$SSE = \sum_{i=1}^r \sum_{j=1}^n (Y_{ij} - \bar{Y}_{i.})^2$	$(n - 1)r$	σ^2

- Only change here is the expectation of $SSTR$ which reflects randomness of α_j 's.
- ANOVA table is still useful to setup tests: the same F statistics for fixed or random will work here.
- Test for random effect: $H_0 : \sigma_{\mu}^2 = 0$ based on

$$F = \frac{MSTR}{MSE} \sim F_{r-1, (n-1)r} \quad \text{under } H_0.$$

One-way ANOVA (random)

Statistics 191:
Introduction
to Applied
Statistics

Jonathan
Taylor
Department of
Statistics
Stanford
University

Inference for population mean: μ .

- Easy to check that

$$E(\bar{Y}_{..}) = \mu.$$

$$\text{Var}(\bar{Y}_{..}) = \frac{n\sigma_{\mu}^2 + \sigma^2}{rn}.$$

- To come up with a t statistic that we can use for test, CIs, we need to find an estimate of $\text{Var}(\bar{Y}_{..})$. ANOVA table says $E(MSTR) = n\sigma_{\mu}^2 + \sigma^2$.
- Therefore,

$$\frac{\bar{Y}_{..} - \mu}{\sqrt{\frac{SSTR}{(r-1)rn}}} \sim t_{r-1}$$

One-way ANOVA (random)

Statistics 191:
Introduction
to Applied
Statistics

Jonathan
Taylor
Department of
Statistics
Stanford
University

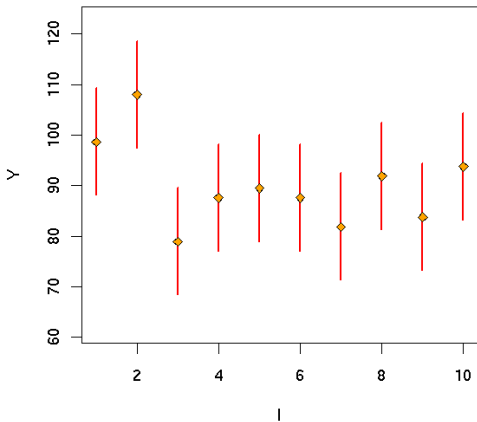
Inference for population mean: μ .

- Why $r - 1$ degrees of freedom? Imagine we could record an infinite number of observations for each individual, so that $\bar{Y}_i \rightarrow \mu_i$, or that $\sigma_{\mu}^2 = 0$.
- To learn anything about μ . we still only have r observations (μ_1, \dots, μ_r) .
- Sampling more within an individual cannot narrow the CI for μ .

Simulated example, $n = 5, r = 10$

Statistics 191:
Introduction
to Applied
Statistics

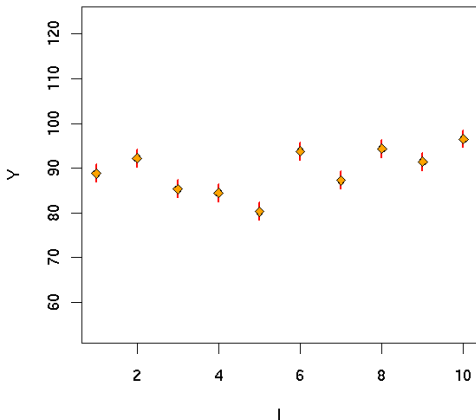
Jonathan
Taylor
Department of
Statistics
Stanford
University



Simulated example, $n = 100, r = 10$

Statistics 191:
Introduction
to Applied
Statistics

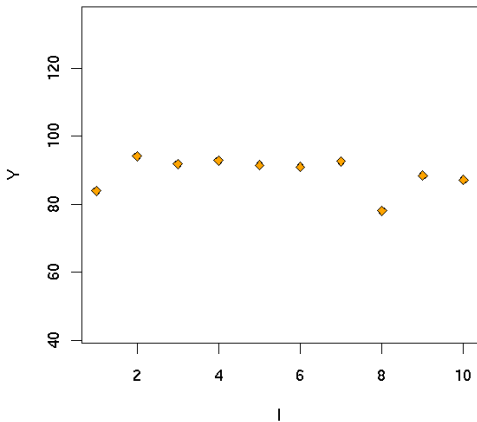
Jonathan
Taylor
Department of
Statistics
Stanford
University



Simulated example, $n = 2000, r = 10$

Statistics 191:
Introduction
to Applied
Statistics

Jonathan
Taylor
Department of
Statistics
Stanford
University



One-way ANOVA (random)

Statistics 191:
Introduction
to Applied
Statistics

Jonathan
Taylor
Department of
Statistics
Stanford
University

Estimating σ_μ^2

- From the ANOVA table

$$\sigma_\mu^2 = \frac{E(SSTR/(r-1)) - E(SSE/((n-1)r))}{n}.$$

- Natural estimate:

$$S_\mu^2 = \frac{SSTR/(r-1) - SSE/((n-1)r)}{n}$$

- Problem: this estimate can be negative! One of the difficulties in random effects model.