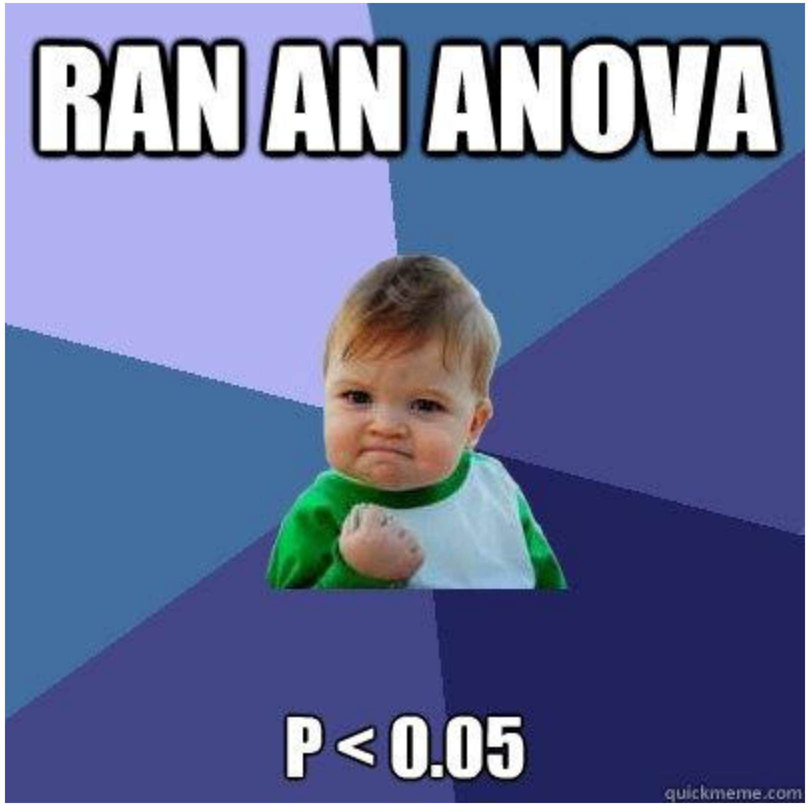
Analysis of Variance



Overview

One-way analysis of variance (ANOVA) concepts and R

Connections between ANOVA and linear regression

Paired tests after running an ANOVA

Quick discussion of your maps...

Let's take a few minutes and show your neighbor the map you created

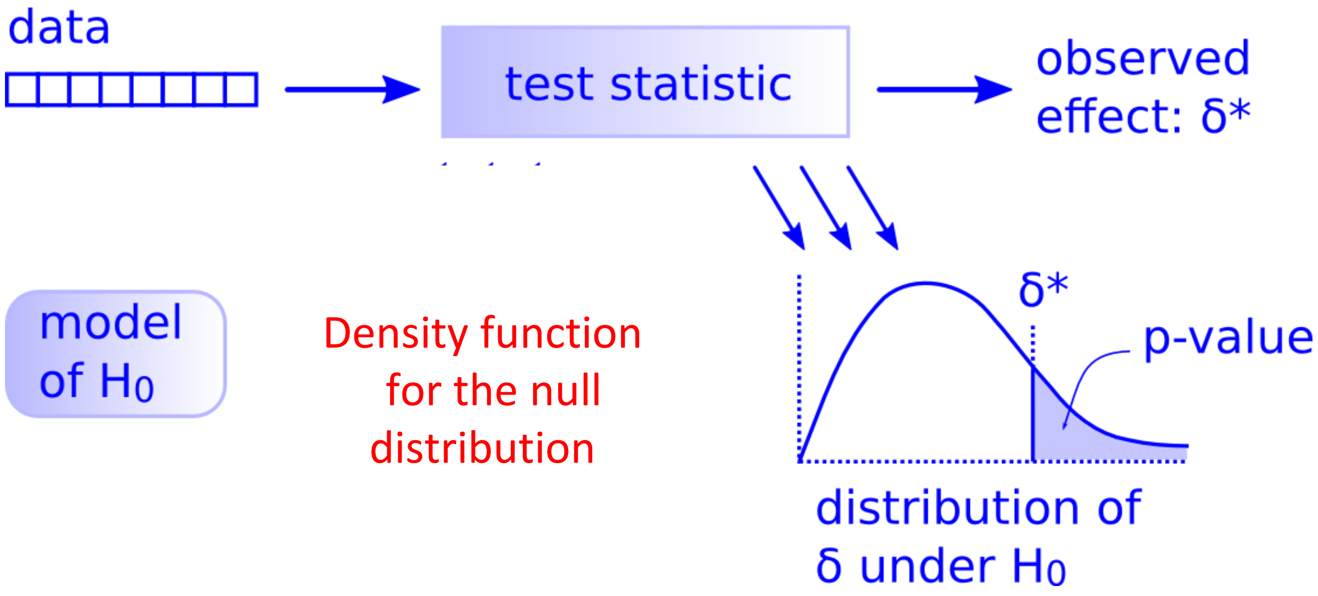
Did anyone find anything particularly interesting?

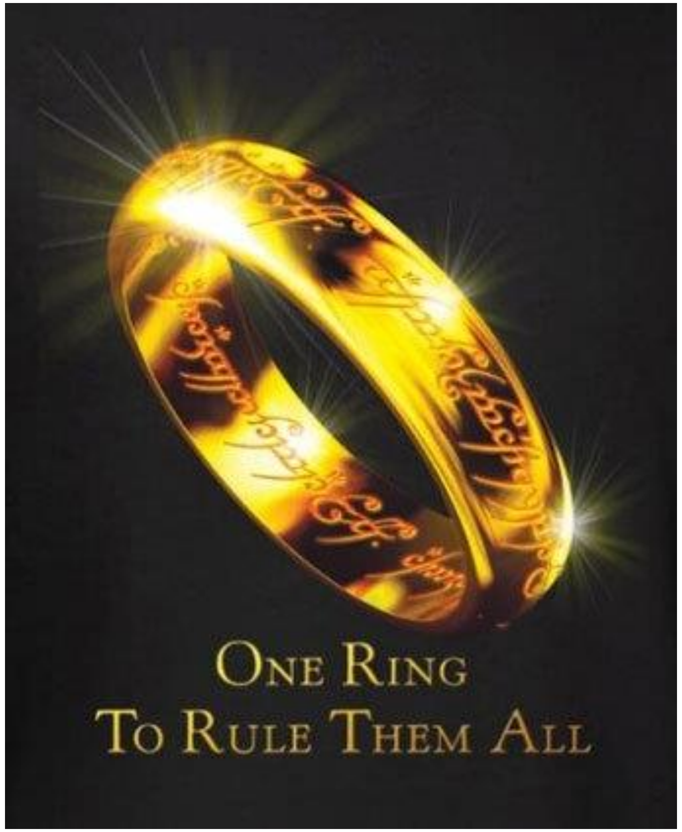
One-way analysis of variance (ANOVA)

One-way ANOVA

A **one-way analysis of variance (ANOVA)** is a parametric hypothesis test that can be used to examine if a set of means are all the same.

There is only one hypothesis test!





Just follow the 5 hypothesis tests steps!

One-way ANOVA

A **one-way analysis of variance (ANOVA)** is a parametric hypothesis test that can be used to examine if a set of means are all the same.

H_0: $\mu_1 = \mu_2 = \ldots = \mu_k$

H_A: $\mu_i \neq \mu_j$ for some i, j

Q: Have we run a test comparing multiple means yet?

Faculty salaries again…

Silly question: Do Assistant, Associate and Fully Professors get paid the same on average?

One-way ANOVA

A **one-way analysis of variance (ANOVA)** is a parametric hypothesis test that can be used to examine if a set of means are all the same.

H_0: $\mu_1 = \mu_2 = \ldots = \mu_k$

H_A: $\mu_i \neq \mu_j$ for some i, j

The statistic we use for a one-way ANOVA is the F-statistic

$$F = \frac{\text{between-group variability}}{\text{within-group variability}} = \frac{\frac{1}{K-1}\sum_{i=1}^{K} n_i(\bar{x}_i - \bar{x}_{tot})^2}{\frac{1}{N-K}\sum_{i=1}^{K}\sum_{j=1}^{n_i}(x_{ij} - \bar{x}_i)^2}$$

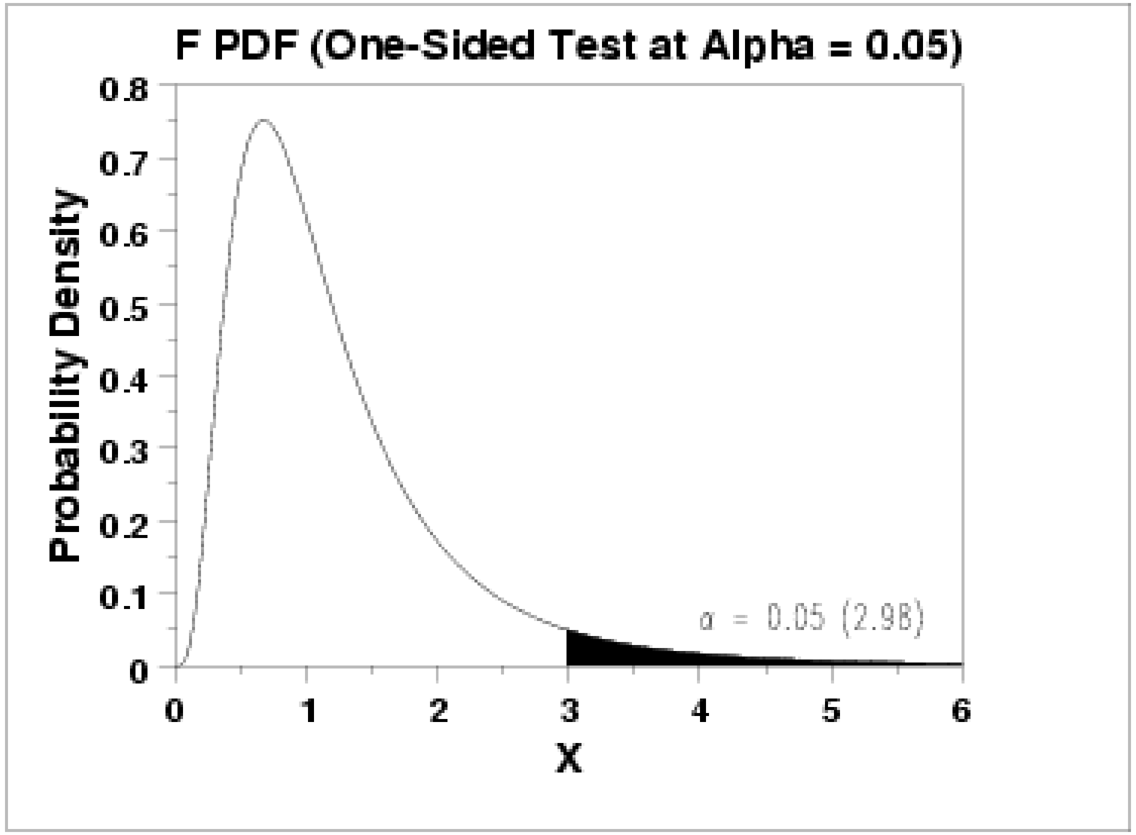
One-way ANOVA – the central idea

If H_0 is true, the F-statistic will come from an F distribution with parameters

- df_1 = K - 1
- df_2 = N - K

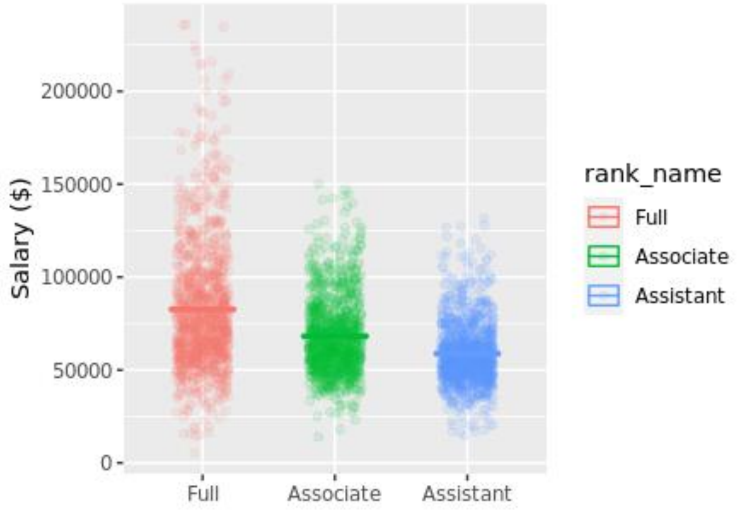
The F-distribution is valid if these conditions are met:

- The data in each group should follow a normal distribution
- The variances in each group should be approximately equal

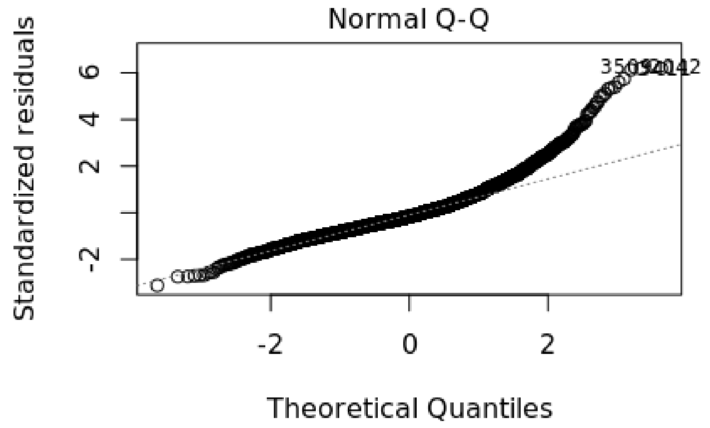


Checking ANOVA conditions (‘assumptions’)

1. We can check if the data in each group is relatively normal by visually examining the residuals between each point and its group mean:
    - Residuals as a function of the group mean
    - Q-Q plots
    - Histograms of residuals



2. We can check the equal variance condition by seeing if the ratio of the largest to smallest standard deviation is greater than 2
    - $s_{max}/s_{min} < 2$



Note: the one-way ANOVA is fairly robust to violations of these conditions.

Calculating the observed F-statistic

$$F = \frac{\text{between-group variability}}{\text{within-group variability}} = \frac{\frac{1}{K-1}\sum_{i=1}^{K} n_i(\bar{x}_i - \bar{x}_{tot})^2}{\frac{1}{N-K}\sum_{i=1}^{K}\sum_{j=1}^{n_i}(x_{ij} - \bar{x}_i)^2}$$

K: the number of groups

N: total number of points

$\bar{x}_{tot}$: the mean across all the data

$\bar{x}_i$: the mean of group i

n_i: the number of points in group i

x_{ij} : the j^{th} data point from group i

K = 3 different times of day

N = 48 total downloads (16 * 3)

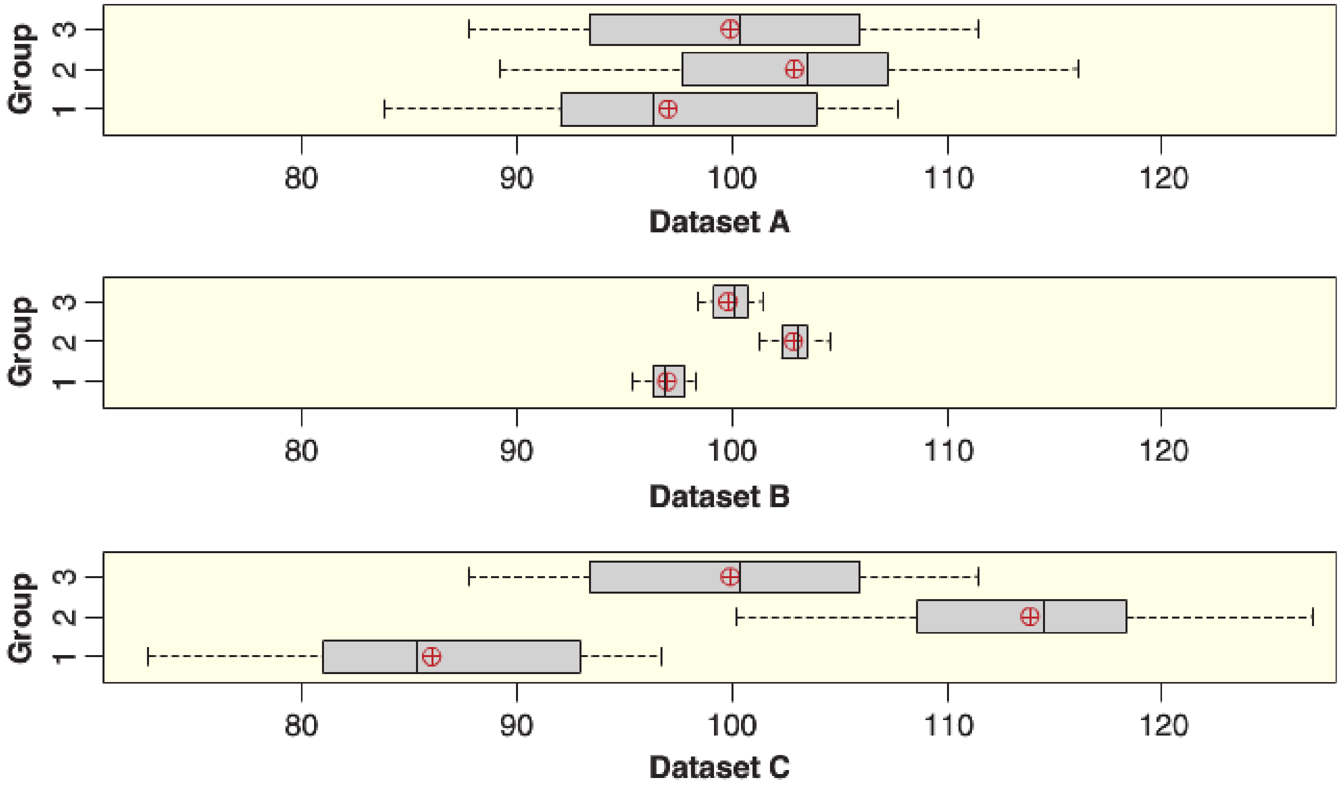
$\bar{x}_{tot}$: the mean speed across all data

$\bar{x}_i$: the means for the i^{th} time of day

n_i = 16 downloads for each time of day

x_{ij} : the j^{th} download at the i^{th} time of day

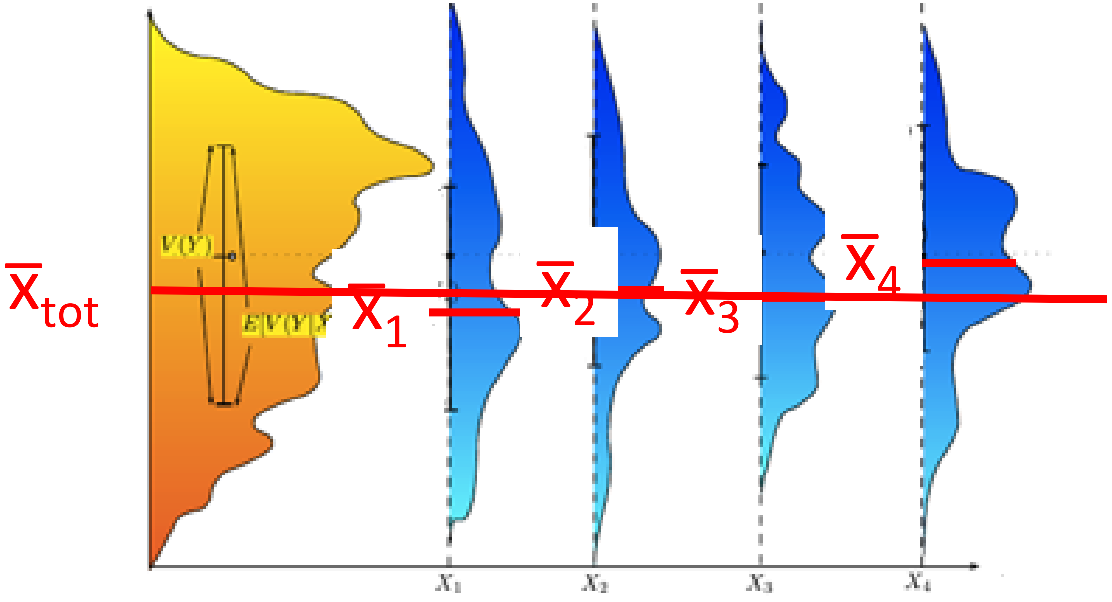
Why use the F-Statistic?



Which dataset gives the strongest evidence that there is a difference in population means?

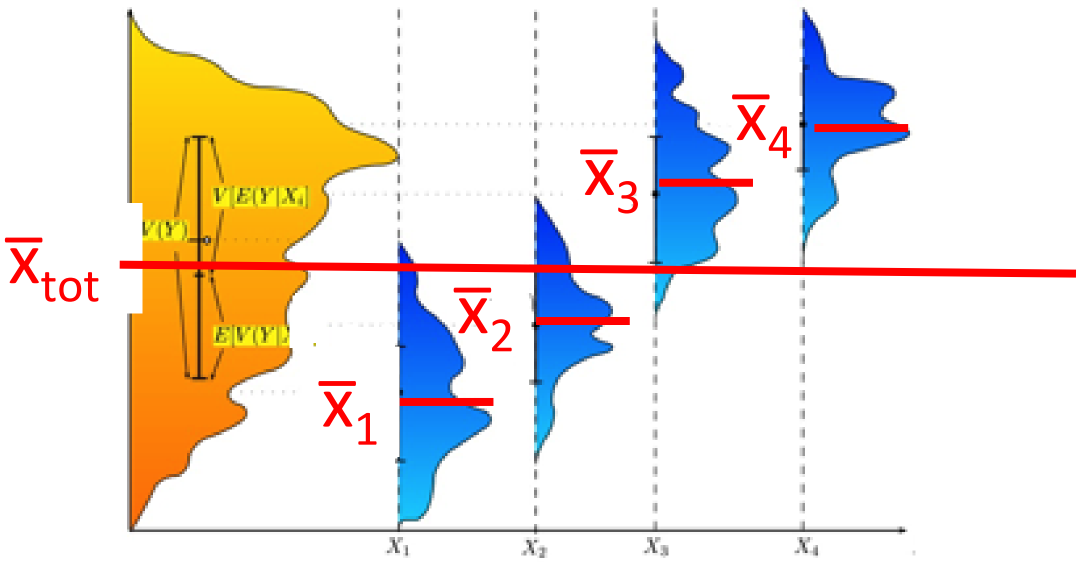
The F-Statistic

If H_0 is true, the data from all groups have the same means



- Similar means $\bar{x}_i$
- Similar spread s_i

If H_0 is not true, the data from all groups have different means



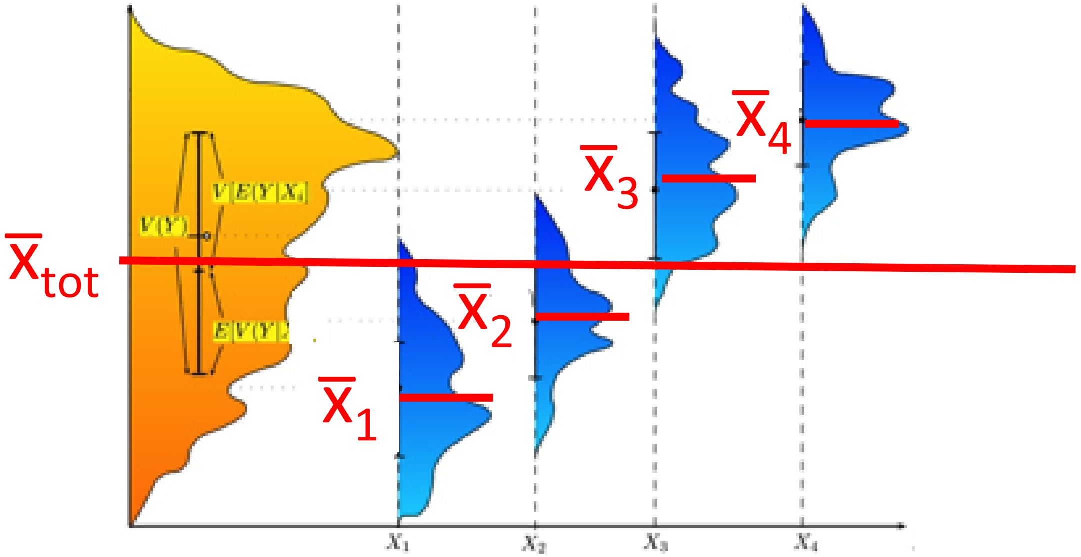
- Different means $\bar{x}_i$
- Smaller spreads s_i

The F-statistic

$$F = \frac{\text{between-group variability}}{\text{within-group variability}} = \frac{\frac{1}{K-1}\sum_{i=1}^{K} n_i(\bar{x}_i - \bar{x}_{tot})^2}{\frac{1}{N-K}\sum_{i=1}^{K}\sum_{j=1}^{n_i}(x_{ij} - \bar{x}_i)^2}$$

The F statistic measures a fraction of:

$$F = \frac{\text{variability between group means}}{\text{variability within each group}}$$



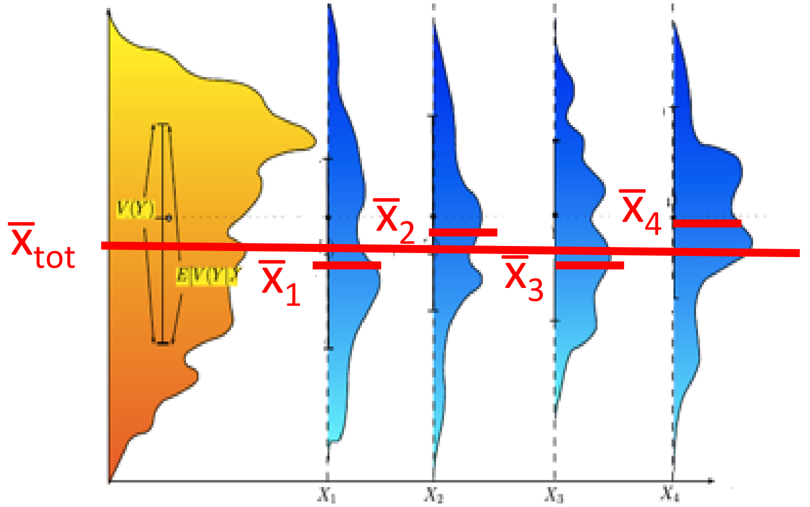
The F-statistic

$$F = \frac{\text{between-group variability}}{\text{within-group variability}} = \frac{\frac{1}{K-1}\sum_{i=1}^{K} n_i(\bar{x}_i - \bar{x}_{tot})^2}{\frac{1}{N-K}\sum_{i=1}^{K}\sum_{j=1}^{n_i}(x_{ij} - \bar{x}_i)^2}$$

$s_i^2 \approx \sigma^2$

The F statistic measures a fraction of:

$$F = \frac{\text{variability between group means}}{\text{variability within each group}}$$



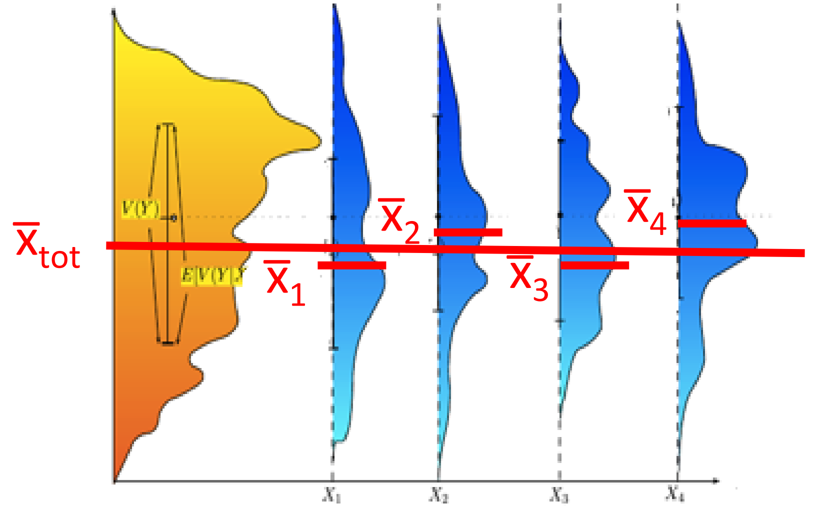
The F-statistic

$$F = \frac{\text{between-group variability}}{\text{within-group variability}} = \frac{\frac{1}{K-1}\sum_{i=1}^{K} n_i(\bar{x}_i - \bar{x}_{tot})^2}{\frac{1}{N-K}\sum_{i=1}^{K}\sum_{j=1}^{n_i}(x_{ij} - \bar{x}_i)^2}$$

$s_i^2 \approx \sigma^2$

The F statistic measures a fraction of:

$$F = \frac{\text{variability between group means}}{\approx \sigma^2}$$



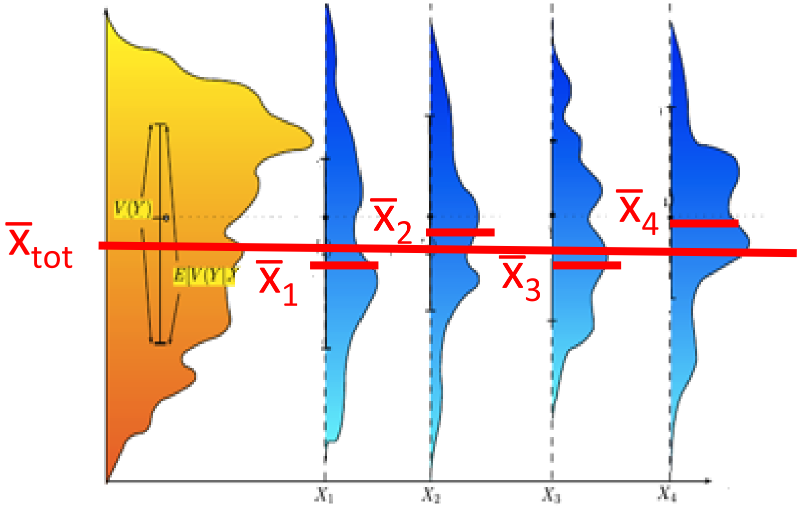
The F-statistic

$$F = \frac{\text{between-group variability}}{\text{within-group variability}} = \frac{\frac{1}{K-1}\sum_{i=1}^{K} n_i(\bar{x}_i - \bar{x}_{tot})^2}{\frac{1}{N-K}\sum_{i=1}^{K}\sum_{j=1}^{n_i}(x_{ij} - \bar{x}_i)^2}$$

SE^2

The F statistic measures a fraction of:

$$F = \frac{\text{variability between group means}}{\approx \sigma^2}$$



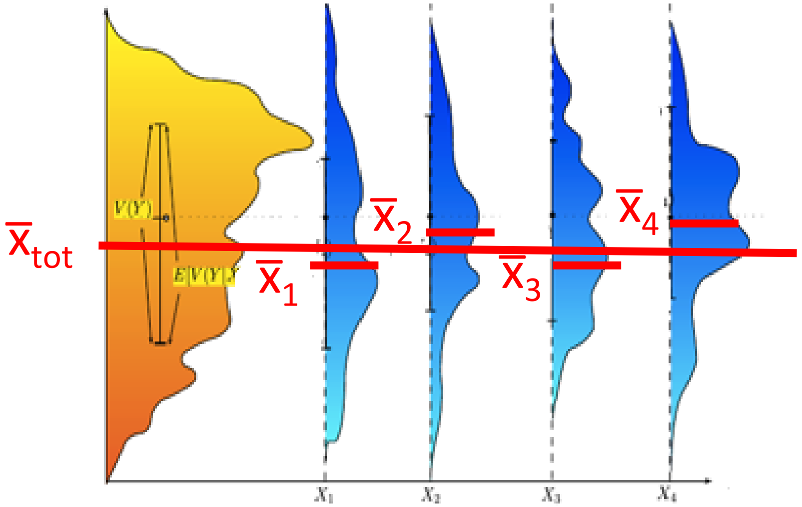
The F-statistic

$$F = \frac{\text{between-group variability}}{\text{within-group variability}} = \frac{\frac{1}{K-1}\sum_{i=1}^{K} n_i(\bar{x}_i - \bar{x}_{tot})^2}{\frac{1}{N-K}\sum_{i=1}^{K}\sum_{j=1}^{n_i}(x_{ij} - \bar{x}_i)^2}$$

$SE^2 \approx \sigma^2/n$

The F statistic measures a fraction of:

$$F = \frac{\approx \sigma^2}{\approx \sigma^2} \approx 1$$



The F-statistic

Sum of Squares Group (SSG)

$$F = \frac{\text{between-group variability}}{\text{within-group variability}} = \frac{\frac{1}{K-1}\sum_{i=1}^{K} n_i(\bar{x}_i - \bar{x}_{tot})^2}{\frac{1}{N-K}\sum_{i=1}^{K}\sum_{j=1}^{n_i}(x_{ij} - \bar{x}_i)^2}$$

The F statistic measures a fraction of:

$$F = \frac{\approx \sigma^2}{\approx \sigma^2} \approx 1$$

The F-statistic

Mean Squares Group (MSG)

$$F = \frac{\text{between-group variability}}{\text{within-group variability}} = \frac{\frac{1}{K-1}\sum_{i=1}^{K} n_i(\bar{x}_i - \bar{x}_{tot})^2}{\frac{1}{N-K}\sum_{i=1}^{K}\sum_{j=1}^{n_i}(x_{ij} - \bar{x}_i)^2}$$

The F statistic measures a fraction of:

$$F = \frac{\text{MSG}}{\approx \sigma^2} \approx 1$$

The F-statistic

Sum of Squares Error (SSE)

$$F = \frac{\text{between-group variability}}{\text{within-group variability}} = \frac{\frac{1}{K-1}\sum_{i=1}^{K} n_i(\bar{x}_i - \bar{x}_{tot})^2}{\frac{1}{N-K}\sum_{i=1}^{K}\sum_{j=1}^{n_i}(x_{ij} - \bar{x}_i)^2}$$

The F statistic measures a fraction of:

$$F = \frac{\text{MSG}}{\approx \sigma^2} \approx 1$$

The F-statistic

Mean of Squares Error (MSE)

$$F = \frac{\text{between-group variability}}{\text{within-group variability}} = \frac{\frac{1}{K-1}\sum_{i=1}^{K} n_i(\bar{x}_i - \bar{x}_{tot})^2}{\frac{1}{N-K}\sum_{i=1}^{K}\sum_{j=1}^{n_i}(x_{ij} - \bar{x}_i)^2}$$

The F statistic measures a fraction of:

$$F = \frac{\text{MSG}}{\text{MSE}} \approx 1$$

Same as what we saw in regression: SSTotal = SSG + SSE

ANOVA table

Source	df	Sum of Sq.	Mean Square	F-statistic	p-value
Groups	$k-1$	SSG	$MSG = \frac{SSG}{k-1}$	$F = \frac{MSG}{MSE}$	Upper tail $F_{k-1,n-k}$
Error	$n-k$	SSE	$MSE = \frac{SSE}{n-k}$		
Total	$n-1$	$SSTotal$			

Where:

$$SSG = \sum_{i=1}^{k} n_i(\bar{x}_i - \bar{x}_{tot})^2$$

$$SST = \sum_{i=1}^{k}\sum_{j=1}^{n_i}(x_{ij} - \bar{x}_{tot})^2$$

$$SSE = \sum_{i=1}^{k}\sum_{j=1}^{n_i}(x_{ij} - \bar{x}_i)^2$$

ANOVA table

Just as we saw for linear regression, we have the relationship:

SST = SSG + SSE

Where:

$$SSG = \sum_{i=1}^{k} n_i(\bar{x}_i - \bar{x}_{tot})^2$$

$$SST = \sum_{i=1}^{k}\sum_{j=1}^{n_i}(x_{ij} - \bar{x}_{tot})^2$$

$$SSE = \sum_{i=1}^{k}\sum_{j=1}^{n_i}(x_{ij} - \bar{x}_i)^2$$

Running a one-way ANOVA

Step 1: State the null and alternative hypothesis

Step 2: Calculate the F-statistic on using actual data

Step 3: Create the appropriate F-distribution

Step 4: Calculate the p-value

Step 5: Make a decision

Check our underlying assumptions were met

Let’s try it out in R!

Connections between regression and ANOVA

ANOVA as regression with only categorical predictors

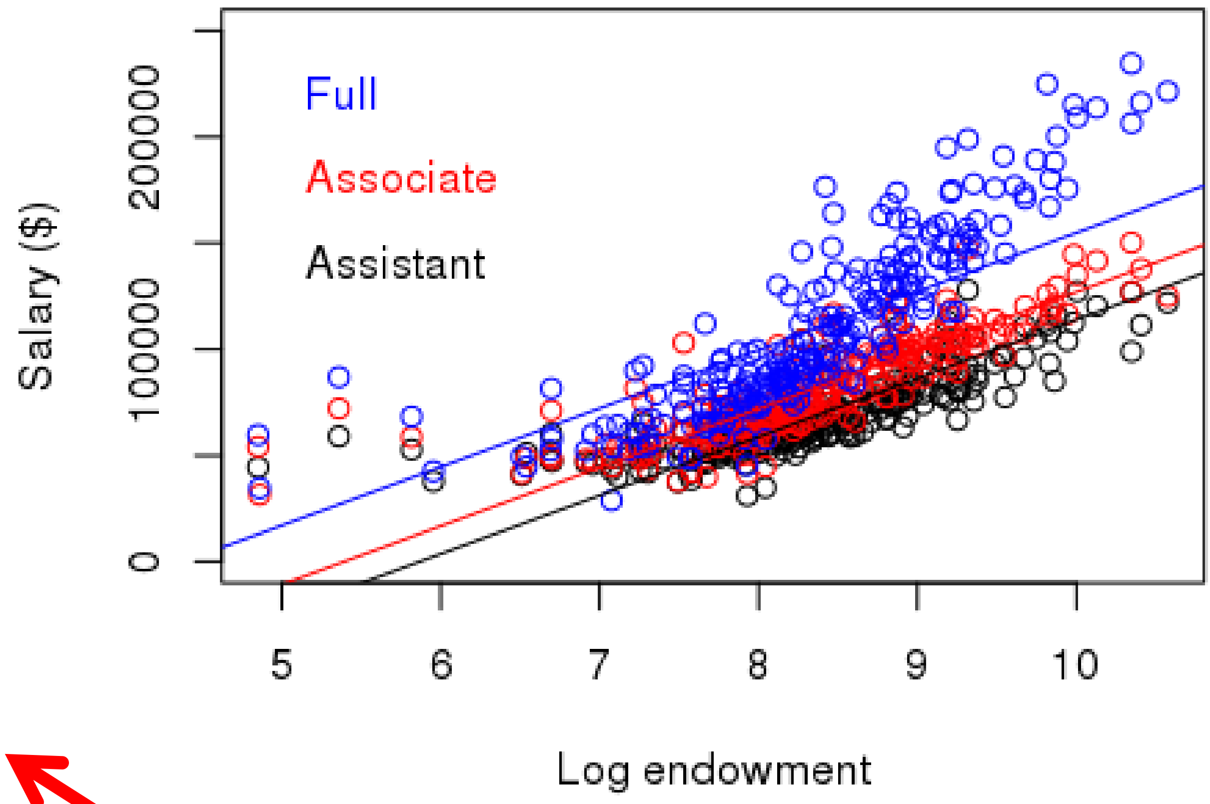
Recall we can have categorical predictors with k levels in a regression model by using k -1 dummy variables:

- e.g., we would need two dummy variables to have different intercepts for Assistant, Associate and Full Professors

$$x_{i1} = \begin{cases} 1 & \text{if Assistant Professor} \\ 0 & \text{if Full Professor} \end{cases} \qquad x_{i2} = \begin{cases} 1 & \text{if Associate Professor} \\ 0 & \text{if Full Professor} \end{cases}$$

$$y_i \;=\; \beta_0 + \beta_1 x_{i1} + \beta_2 x_{i2} + \epsilon_i \;=\; \begin{cases} \beta_0 + \beta_1 + \epsilon_i & \text{if Assistant Professor} \\ \beta_0 + \beta_2 + \epsilon_i & \text{if Associate Professor} \\ \beta_0 + \epsilon_i & \text{if Full Professor} \end{cases}$$

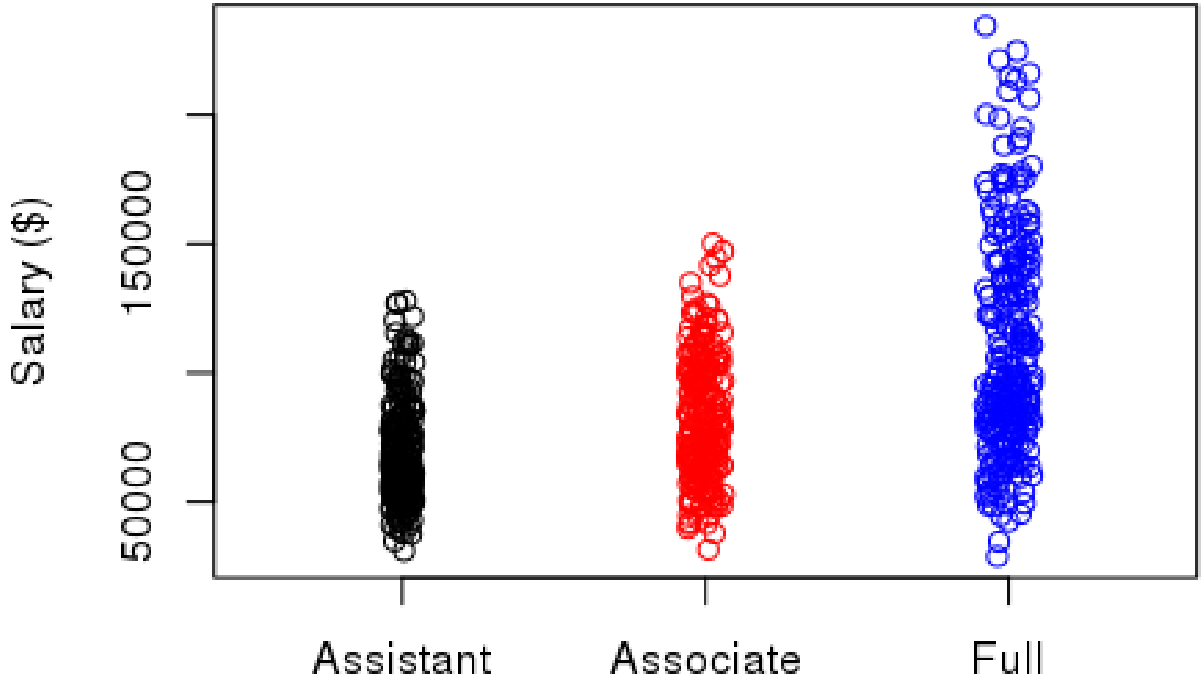
ANOVA as regression with only categorical predictors



Common slope for log endowment

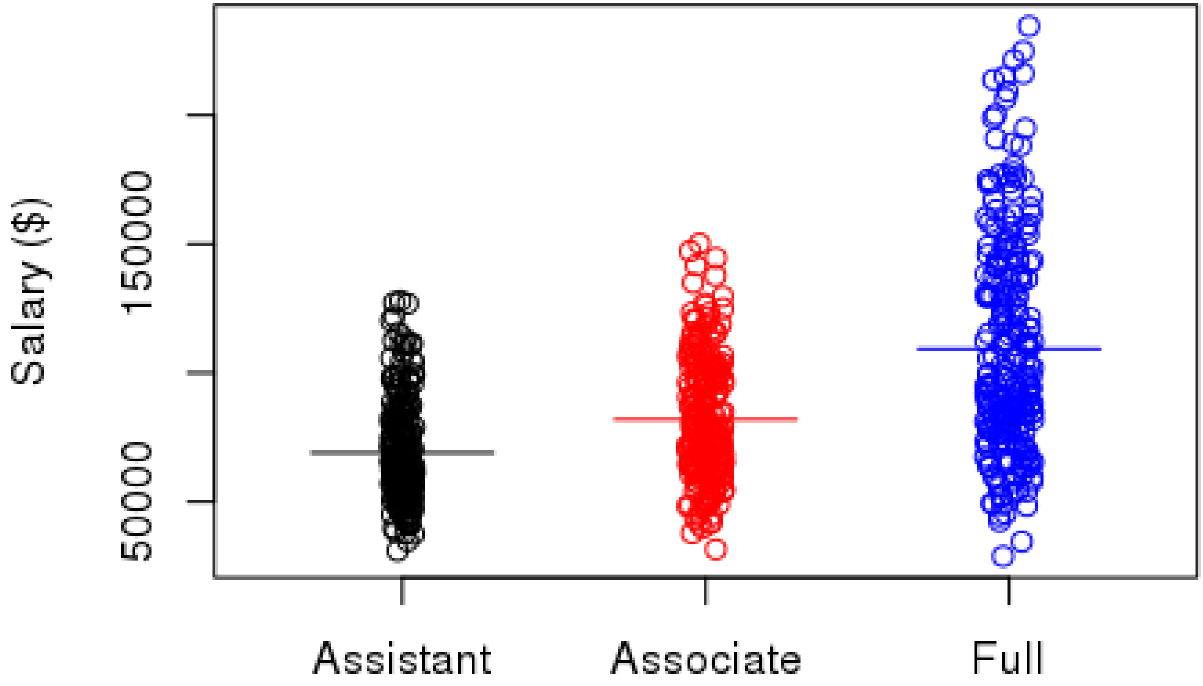
$$y_i = \beta_0 + \beta_1 x_{i1} + \beta_2 x_{i2} + \epsilon_i = \begin{cases} \beta_0 + \beta_1 + \beta_3 x_i + \epsilon_i & \text{if Assistant Professor} \\ \beta_0 + \beta_2 + \beta_3 x_i + \epsilon_i & \text{if Associate Professor} \\ \beta_0 + \beta_3 x_i + \epsilon_i & \text{if Full Professor} \end{cases}$$

ANOVA as regression with only categorical predictors



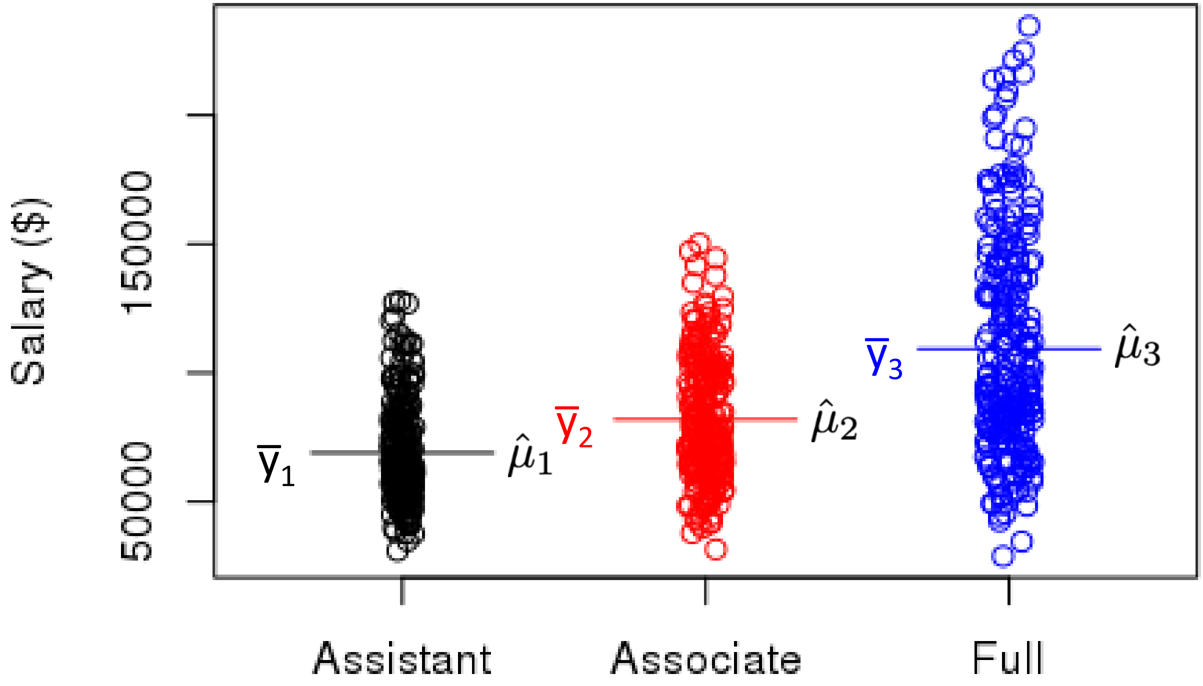
$$y_i = \beta_0 + \beta_1 x_{i1} + \beta_2 x_{i2} + \epsilon_i = \begin{cases} \beta_0 + \beta_1 + \epsilon_i & \text{if Assistant Professor} \\ \beta_0 + \beta_2 + \epsilon_i & \text{if Associate Professor} \\ \beta_0 + \epsilon_i & \text{if Full Professor} \end{cases}$$

Least squares prediction for $\hat{y}_i$ is $\overline{y}_k$



$$y_i = \mu_k + \epsilon_i = \begin{cases} \mu_1 + \epsilon_i & \text{if Assistant Professor} \\ \mu_2 + \epsilon_i & \text{if Associate Professor} \\ \mu_3 + \epsilon_i & \text{if Full Professor} \end{cases}$$

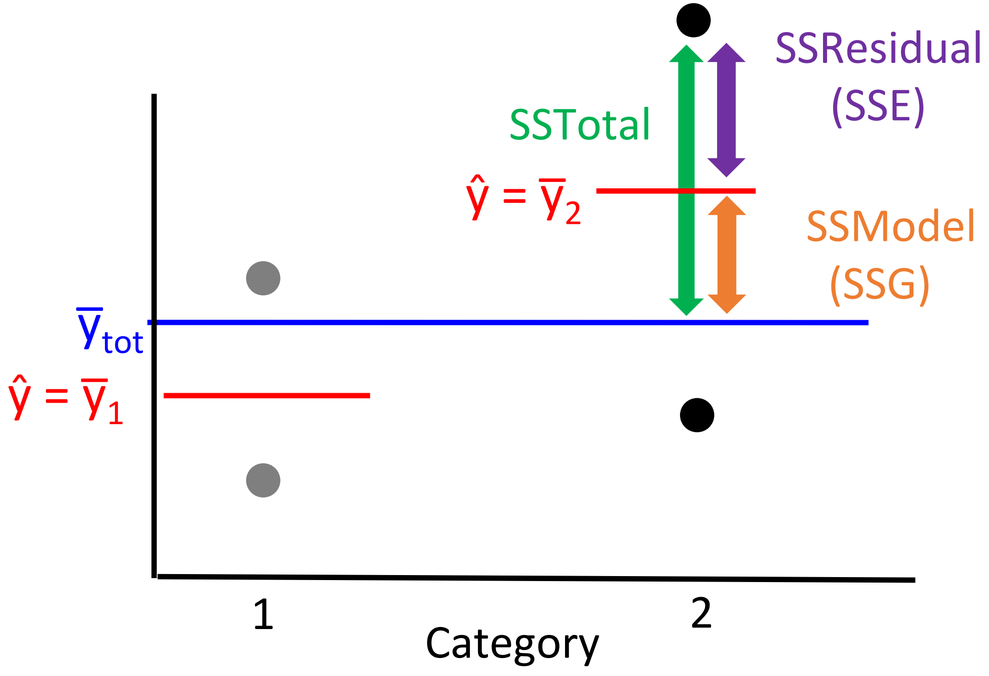
Least squares prediction for $\hat{y}_i$ is $\overline{y}_k$



$$\hat{y}_i = \hat{\mu}_k = \begin{cases} \hat{\mu}_1 & \text{if Assistant Professor} \\ \hat{\mu}_2 & \text{if Associate Professor} \\ \hat{\mu}_3 & \text{if Full Professor} \end{cases}$$

ANOVA decomposition

$$F = \frac{\frac{1}{K-1}\sum_{i=1}^{K} n_i(\bar{y}_i - \bar{y}_{tot})^2}{\frac{1}{N-K}\sum_{i=1}^{K}\sum_{j=1}^{n_i}(y_{ij} - \bar{y}_i)^2}$$



The ANOVA decomposes the variance as:

- SSTotal = SSModel (SSG) + SSResidual (SSE)

$$y_{ij} - \bar{y}_{tot} = (\hat{y}_{ij} - \bar{y}_{tot}) + (y_{ij} - \hat{y}_{ij})$$

$$(y_{ij} - \bar{y}_{tot})^2 = (\hat{y}_{ij} - \bar{y}_{tot})^2 + (y_{ij} - \hat{y}_{ij})^2$$

$$(y_{ij} - \bar{y}_{tot})^2 = (\bar{y}_i - \bar{y}_{tot})^2 + (y_{ij} - \bar{y}_i)^2$$

$\hat{y}_{ji} = \bar{y}_i$

(the prediction for each class is the group mean)

Let’s examine these relationships in R…

Planned comparisons/posthoc tests

Suppose we run a one-way ANOVA and we are able to reject the null hypothesis.

H_0: $\mu_1 = \mu_2 = \ldots = \mu_k$

H_A: $\mu_i \neq \mu_j$ for some i, j

Q: What else would we like to know?

Pairwise comparisons

There are several tests that can be used to examine which pairs of means differed; i.e., to test:

- H_0: $\mu_i = \mu_j$
- H_A: $\mu_i \neq \mu_j$

These tests include:

- Fisher's Least Significant Difference
- Bonferroni procedure/correction
- Tukeys Honest significantly different

Fisher's Least Significant Difference (LSD)

1. Perform the ANOVA

2. If the ANOVA F-test is not significant, stop

3. If the ANOVA F-test is significant, then you can test H_0 for a pairwise comparisons using:

$$t = \frac{\bar{x}_i - \bar{x}_j}{\sqrt{MSE \cdot (\frac{1}{n_i} + \frac{1}{n_j})}}$$

Estimate of the SE
Uses the MSE as a pooled estimate of the σ^2

Use a t-distribution with n-k degrees of freedom

Very 'liberal' tests

- Likely to make Type I errors (lots of false rejections of H_0)
- Less likely to make Type II errors (highest chance of detecting effects)

Bonferroni correction

Controls for the ***family-wise error rate***

- i.e., $\alpha = 0.05$ for making ***any*** Type I error ***over all pairs of comparisons***

1. Choose an α-level for the family-wise error rate α
2. Decide how many comparisons you will make. Call this m.
3. Reject any hypothesis tests that have p-values less than α/m
   - Pairwise tests typically done using a t-statistic, where the MSE is used in the estimate of the SE

$$t = \frac{\bar{x}_i - \bar{x}_j}{\sqrt{MSE \cdot \left(\frac{1}{n_i} + \frac{1}{n_j}\right)}}$$

Use a t-distribution with n-k degrees of freedom

Very 'conservative' tests

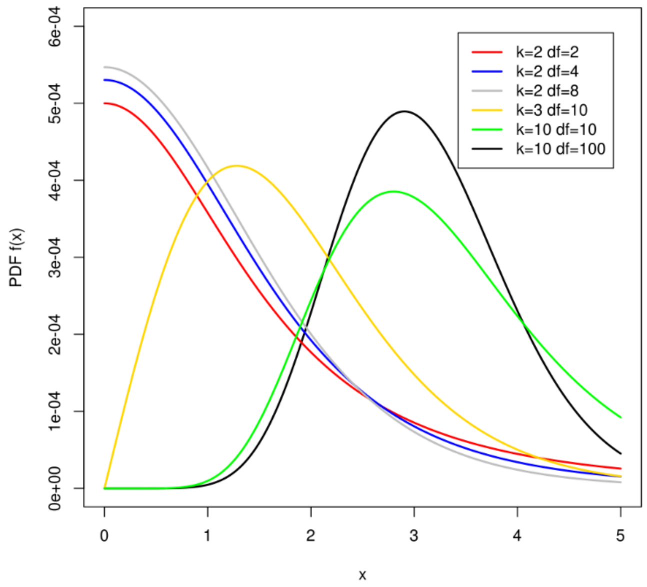
- Unlikely to make Type I errors (few false rejections of H_0)
- Likely to make Type II errors (insensitive at detecting real effects)

Tukey's Honest Significantly Different Test

Controls for the family-wise error rate

$$q = \frac{\sqrt{2}(\bar{x}_{max} - \bar{x}_{min})}{\sqrt{MSE \cdot (\frac{1}{n_{max}} + \frac{1}{n_{min}})}}$$

Where *q* comes from a ***studentized range distribution***



The test is based on the distribution of $|\bar{x}_{max} - \bar{x}_{min}|$ that would be expected under the null hypothesis that none of the pairs of means are different

- Controls for the familywise error rate but less conservative than the Bonferroni correction
- Still based on assumptions that the data in each group is normal with equal variance

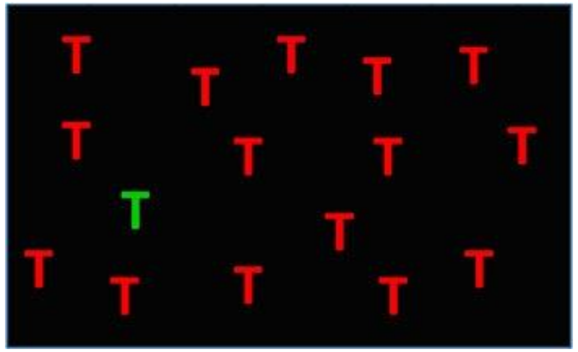
Let’s try the KW test and pairwise comparisons in R…

Homework 10

You will analyze data from a psychophysics experiment that explored popout attention

- Study done at Hampshire College by Jacob Prescott, Tapujit Debnath Tapu, Julian Oks, Kirsten Lydic

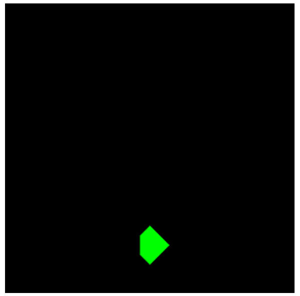
Background:



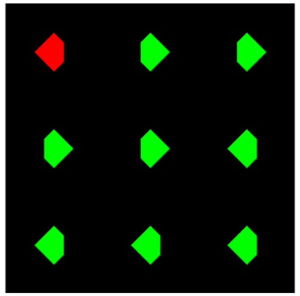
Exogenous attention

Endogenous attention

Single item



Multiple items

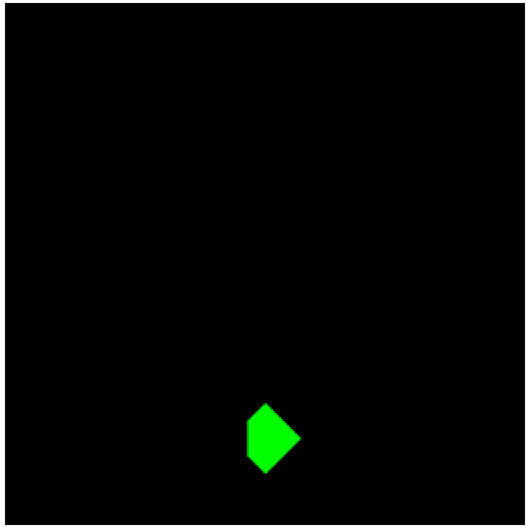


Do reaction times differ for:

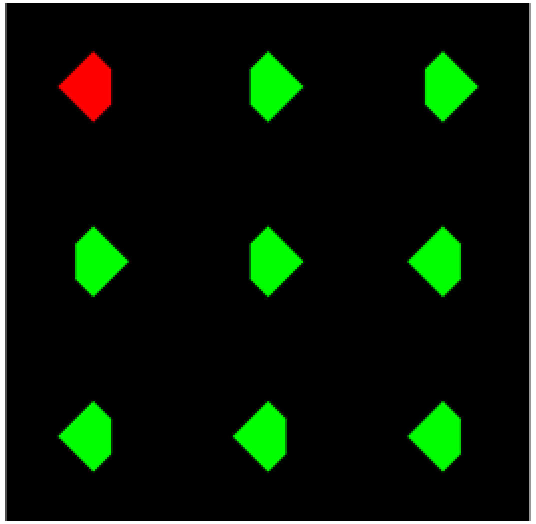
1. Position of target stimulus
2. Single vs. multiple item displays

Homework 10

Participants need to respond as *quickly* and as *accurately* as possible



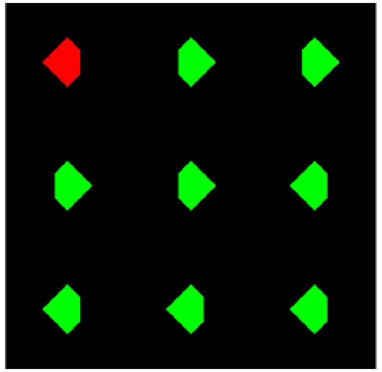
Press "z" because left side is cut off



Press "/" because right side is cut off

You can try it yourself at: https://run.pavlovia.org/xfang/popout_exp1/html/

Homework 10



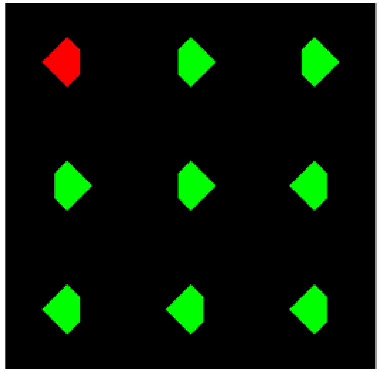
The experiment had a 9 x 2 x 2 x 2 ***factorial*** design:

1. Position (9 levels): 9 locations where the target stimulus could appear
2. Isolated/distractor condition (2 levels): isolated or cluttered display
3. Target color (2 levels): red or green target
    - For cluttered displays, the distractors always had the opposite color of the target
4. Cut direction (2 levels): left or right side of the target diamond was cut off
    - Corresponds to pressing the "z" or "/" key

The experiment had 10 blocks where all 72 (9 x 2 x 2 x 2) stimuli were shown

8 volunteer participants participated in the experiment

Homework 10



On homework 10 you will run:

- A one-way ANOVA to see if the mean reaction time is the same at all target positions
- A two-way ANOVA to look at how both position and isolated/cluttered displays affect mean reaction times.
- Explore another question using this data

Questions?