Logistics

- Peer reviews are due tomorrow before midnight. Use UNM learn to submit.

- Today Analysis of Variance

- Next Tuesday Cross Validation and Wrap Up
- Next Thursday – Final Exam Review (May 2nd)
- May 7^{th} – Final Exam (Material from second half of class)

- Assignment 4 will be a regular homework covering ANOVAs.

Analysis of Variance (Part 1)

MATTHEW FRICKE

1.0 – SEND CORRECTIONS TO MFRICKE@UNM.EDU

Introduction

We have seen that most experiments involve studying the effect of one or more factors on a response variable.

The factors can be qualitative or quantitative.

Introduction

We have seen that most experiments involve studying the effect of one or more factors on a response variable.

The factors can be qualitative or quantitative.

If we want to understand the relative impact multiple factors have on the response we can use Analysis of Variation (ANOVAs).

Introduction

The goal of ANOVAs is to identify important independent variables and rank their importance.

The idea is based on measuring the variability and applying various statistical identities to define a test statistic for the affect factors have on the response variable.

The test statistic (like others you have seen) is defined so that we can get a p-value from a lookup table. This allows us to say what factors have a significant affect on the response variable.

Introduction

The goal of ANOVAs is to identify important independent variables and rank their importance.

The idea is based on measuring the variability and applying various statistical identities to define a test statistic for the affect factors have on the response variable.

The test statistic (like others you have seen) is defined so that we can get a p-value from a lookup table. This allows us to say what factors have a significant affect on the response variable.

This test statistic if called F.

The procedure

First recall that the variability of a set of measurements is proportional to the sum of squares of deviations from the mean:

$$\sum_{i=1}^{n} (y_i - \bar{y})^2$$

This is how we calculated the sample variance.

The procedure

The ANOVA partitions the total sum of squares (TSS) into parts. Each of these parts is attributed to a particular independent variable (factor) in the experiment.

$$\mathrm{TSS} = \sum_{i=1}^{n} (y_i - \bar{y})^2 = \mathrm{factor}_1 + \cdots + \mathrm{factor}_k + \mathrm{SSE}$$

Where SSE is the sum of square errors (recall from linear regression lecture).

The procedure

Let's think about the null hypothesis that there is no relationship.

When the factors have no relationship to the response each of the factor contributions to the sum each estimate the variance of the response.

$$\mathrm{TSS} = \sum_{i=1}^{n}(y_i - \bar{y})^2 = \mathrm{factor}_1 + \cdots + \mathrm{factor}_k + \mathrm{SSE}$$

Dividing each term by a constant means TSS is a good estimate of the variance of the response.

The procedure

If one of the factors is related to the response then the contribution of its variance estimate to TSS will be inflated.

$$\text{TSS} = \sum_{i=1}^{n}(y_i - \bar{y})^2 = \text{factor}_1 + \cdots + \text{factor}_k + \text{SSE}$$

The contribution of a factor to the total sum of squares is called the "factor's sum of squares".

The procedure

If one of the factors is related to the response then the contribution of its variance estimate to TSS will be inflated.

$$\text{TSS} = \sum_{i=1}^{n} (y_i - \bar{y})^2 = \text{factor}_1 + \cdots + \text{factor}_k + \text{SSE}$$

The contribution of a factor to the total sum of squares is called the "factor's sum of squares".

The procedure

So now we have the basis for a test statistic.

We will compare the expected sum of squares for each factor to the actual factors sum of squares and define a probability of seeing the combination if the null hypothesis is true.

$$\text{TSS} = \sum_{i=1}^{n} (y_i - \bar{y})^2 = \text{factor}_1 + \cdots + \text{factor}_k + \text{SSE}$$

The procedure – an example

Consider the special case where we have two equal samples of size n_1 and n_2 where:

$$n_1 = n_2$$

$$\mu_1 = \mu_2$$

$$\sigma_1^2 = \sigma_2^2$$

The procedure – an example

Consider the special case where we have two equal samples of size n_1 and n_2 where:

$$n_1 = n_2$$

$$\mu_1 = \mu_2$$

$$\sigma_1^2 = \sigma_2^2$$

$$\text{TSS} = \sum_{i=1}^{2}\sum_{j=1}^{n_1}(y_{ij} - \bar{y})^2$$

Where y_{ij} is the jth observation in the ith sample and $\bar{y}$ is the mean over all $n_1 + n_2$ samples.

The procedure – an example

Now we partition the Total Sum of Squares:

$$n_1 = n_2$$

$$\mu_1 = \mu_2$$

$$\sigma_1^2 = \sigma_2^2$$

$$\text{TSS} = \sum_{i=1}^{2}\sum_{j=1}^{n_1}(y_{ij} - \bar{y})^2$$

$$= n_1 \sum_{i=1}^{2}(\bar{y_i} - \bar{y})^2$$

$$+ \sum_{i=1}^{2}\sum_{j=1}^{2}(y_{ij} - \bar{y_i})^2$$

Where y_{ij} is the jth observation in the ith sample and $\bar{y}$ is the mean over all $n_1 + n_2$ samples.

The procedure – an example

Now we partition the Total Sum of Squares:

$$\text{TSS} = \sum_{i=1}^{2}\sum_{j=1}^{n_1}(y_{ij} - \bar{y})^2$$

$$n_1 = n_2$$

$$\mu_1 = \mu_2$$

$$\sigma_1^2 = \sigma_2^2$$

Sum of Squares for the treatment (SST). i.e. sum of all the factor contributions.

$$= n_1 \sum_{i=1}^{2}(\bar{y}_i - \bar{y})^2$$

Sum of Squares for the treatment (SST). i.e. sum of all the factor contributions.

$$+ \sum_{i=1}^{2}\sum_{j=1}^{2}(y_{ij} - \bar{y}_i)^2$$

Where y_{ij} is the jth observation in the ith sample and $\bar{y}$ is the mean over all $n_1 + n_2$ samples.

The procedure – an example

Now we partition the Total Sum of Squares:

$$\text{TSS} = \sum_{i=1}^{2}\sum_{j=1}^{n_1}(y_{ij} - \bar{y})^2$$

$$n_1 = n_2$$

$$\mu_1 = \mu_2$$

$$\sigma_1^2 = \sigma_2^2$$

Sum of Squares for the treatment (SST). i.e. sum of all the factor contributions.

$$= n_1 \sum_{i=1}^{2}(\bar{y_i} - \bar{y})^2$$

Sum of squares error (SSE)

$$+ \sum_{i=1}^{2}\sum_{j=1}^{2}(y_{ij} - \bar{y_i})^2$$

Where y_{ij} is the jth observation in the ith sample and $\bar{y}$ is the mean over all $n_1 + n_2$ samples.

The procedure – an example

Now we partition the Total Sum of Squares:

$$n_1 = n_2$$

$$\mu_1 = \mu_2$$

$$\sigma_1^2 = \sigma_2^2$$

$$\text{TSS} = \sum_{i=1}^{2}\sum_{j=1}^{n_1}(y_{ij} - \bar{y})^2$$

Sum of Squares for the treatment (SST). i.e. sum of all the factor contributions.

$$= n_1 \sum_{i=1}^{2}(\bar{y_i} - \bar{y})^2$$

Sum of squares error (SSE)

$$+ \sum_{i=1}^{2}\sum_{j=1}^{2}(y_{ij} - \bar{y_i})^2$$

Notice that the SST increases as the means of the two samples diverge and SSE decreases.

The procedure – an example

Now we partition the Total Sum of Squares:

$$n_1 = n_2$$

$$\mu_1 = \mu_2$$

$$\sigma_1^2 = \sigma_2^2$$

$$\text{TSS} = \sum_{i=1}^{2}\sum_{j=1}^{n_1}(y_{ij} - \bar{y})^2$$

Sum of Squares for the treatment (SST). i.e. sum of all the factor contributions.

$$= n_1 \sum_{i=1}^{2}(\bar{y_i} - \bar{y})^2$$

Sum of squares error (SSE)

$$+ \sum_{i=1}^{2}\sum_{j=1}^{2}(y_{ij} - \bar{y_i})^2$$

Notice that the SST increases as the means of the two samples diverge and SSE decreases.

The procedure – an example

Now we partition the Total Sum of Squares:

$$n_1 = n_2$$

$$\mu_1 = \mu_2$$

$$\sigma_1^2 = \sigma_2^2$$

$$\text{TSS} = \sum_{i=1}^{2}\sum_{j=1}^{n_1}(y_{ij} - \bar{y})^2$$

Sum of Squares for the treatment (SST). i.e. sum of all the factor contributions.

$$= n_1 \sum_{i=1}^{2}(\bar{y_i} - \bar{y})^2$$

Sum of squares error (SSE)

$$+ \sum_{i=1}^{2}\sum_{j=1}^{2}(y_{ij} - \bar{y_i})^2$$

So the larger SST is the greater the evidence for rejecting H_0.

The procedure – an example

So the larger SST is the greater the evidence for rejecting H_0.

How much evidence does this value provide?

The procedure

So the larger SST is the greater the evidence for rejecting H_0.

How much evidence does this value provide?

Let Y_i denote a normally distributed random variable that generated observation y_{ij} then,

The expected value for the variance is:

$$E\left(\frac{\text{SSE}}{2n_1 - 2}\right) = \sigma^2$$

Recall this was an estimator of variance from the first half of class.

The procedure

So the larger SST is the greater the evidence for rejecting H_0.

How much evidence does this value provide?

It is possible to prove that the expected value of the treatment sum of squares is

$$E(\text{SST}) = \sigma^2 + \frac{n_1}{2}(\mu_1 - \mu_2)^2$$

So when there is a difference in means the SST exceeds the variance

The procedure

This allows us to define the Z statistic

$$Z = \frac{\hat{Y}_1 - \hat{Y}_2}{\sqrt{2\sigma^2/n_1}}$$

The procedure

This allows us to define the Z statistic

$$Z = \frac{\hat{Y}_1 - \hat{Y}_2}{\sqrt{2\sigma^2 / n_1}}$$

Since Z is normally distributed under the null hypothesis...

Z^2 follows a χ^2 distribution with 1 degree of freedom.

The procedure

And $$Z^2 = \left(\frac{n_1}{2}\right)\frac{(\hat{Y}_1 - \hat{Y}_2)^2}{\sigma^2} = \frac{\mathrm{SST}}{\sigma^2}$$

The whole point is that we have been able to relate the sum of squares of the treatment to a distribution we can use to generate a p–value.

The procedure

Assuming SST and SSE are independent it follows that

$$\frac{Z^2}{\mathrm{SSE}/(2n_1 - 2)\sigma^2}$$

Defines an F distribution.

The procedure – Mean squares

Sums of squares divided by their degrees of freedom are called mean squares.

So we can define a mean squared for the treatment and the error.

Under $H_0 : \mu_1 = \mu_2$ both MST and MSE estimate the response variance.

The procedure – Mean squares

Sums of squares divided by their degrees of freedom are called mean squares.

So we can define a mean squared for the treatment and the error.

Under $H_0 : \mu_1 = \mu_2$ both MST and MSE estimate the response variance.

When $\mu_1 \neq \mu_2$ MST will be larger than the response variance and tend to be larger than MSE.

The procedure – Mean squares

So finally we have something we can test!

$$F = \frac{\mathrm{MST}}{\mathrm{MSE}}$$

If $F \geq F_\alpha$ then we can reject H_0 at the significance level α
As usual we look up F_α in a table (or more likely we use MATLAB or R.)

The procedure – Mean squares

This explanation was just for a two factor ANOVA. But it is generalizable to any number of factors.

Example

Consider two factors:

A	B
6.1	9.1
7.1	8.2
7.8	8.6
6.9	6.9
7.6	7.5
8.2	7.9

Are the means different at the 0.05 significance level?

Example

$$\text{Total Sum of Squares} = \sum_{i=1}^{2}\sum_{j=1}^{6}(y_{ij} - \bar{y})^2$$

$$\text{Total Sum of Squares} = \sum_{i=1}^{2}\sum_{j=1}^{6} y_{ij} - \frac{\sum_{i=1}^{2}\sum_{j=1}^{6} y_{ij}}{12}$$

Example

$$\text{Total Sum of Squares} = \sum_{i=1}^{2}\sum_{j=1}^{6}(y_{ij} - \bar{y})^2$$

$$\text{Total Sum of Squares} = 711.35 - \frac{(91.9)^2}{12} = 7.5492$$

Example

Now we calculate the Treatment and error sums of squares:

$$\mathrm{SST} = n_1 \sum_{i=1}^{2} (\bar{y_i} - \bar{y})^2$$

$$= \frac{n_1}{2} (\bar{y_1} - \bar{y_2})^2 = 1.6875$$

Example

Now we calculate the Treatment and error sums of squares:

$$\text{SST} = n_1 \sum_{i=1}^{2} (\bar{y_i} - \bar{y})^2 = 1.6875$$

$$\text{SSE} = \sum_{i=1}^{2} \sum_{j=1}^{6} (y_{ij} - \bar{y_i})^2 = 5.8617$$

Example

Finally we take the mean sum of squares

$$\mathrm{MST} = \mathrm{SSE}/1 = 1.6875$$

$$\mathrm{MSE} = \mathrm{SSE}/(2n_1 - 2) = 0.58617$$

Example

Finally we take the mean sum of squares

$$\mathrm{MST} = \mathrm{SSE}/1 = 1.6875$$

$$\mathrm{MSE} = \mathrm{SSE}/(2n_1 - 2) = 0.58617$$

Example

Finally we take the mean sum of squares

$$\mathrm{MST} = \mathrm{SSE}/1 = 1.6875$$

$$\mathrm{MSE} = \mathrm{SSE}/(2n_1 - 2) = 0.58617$$

$$\frac{\mathrm{MST}}{\mathrm{MSE}} = 2.88$$

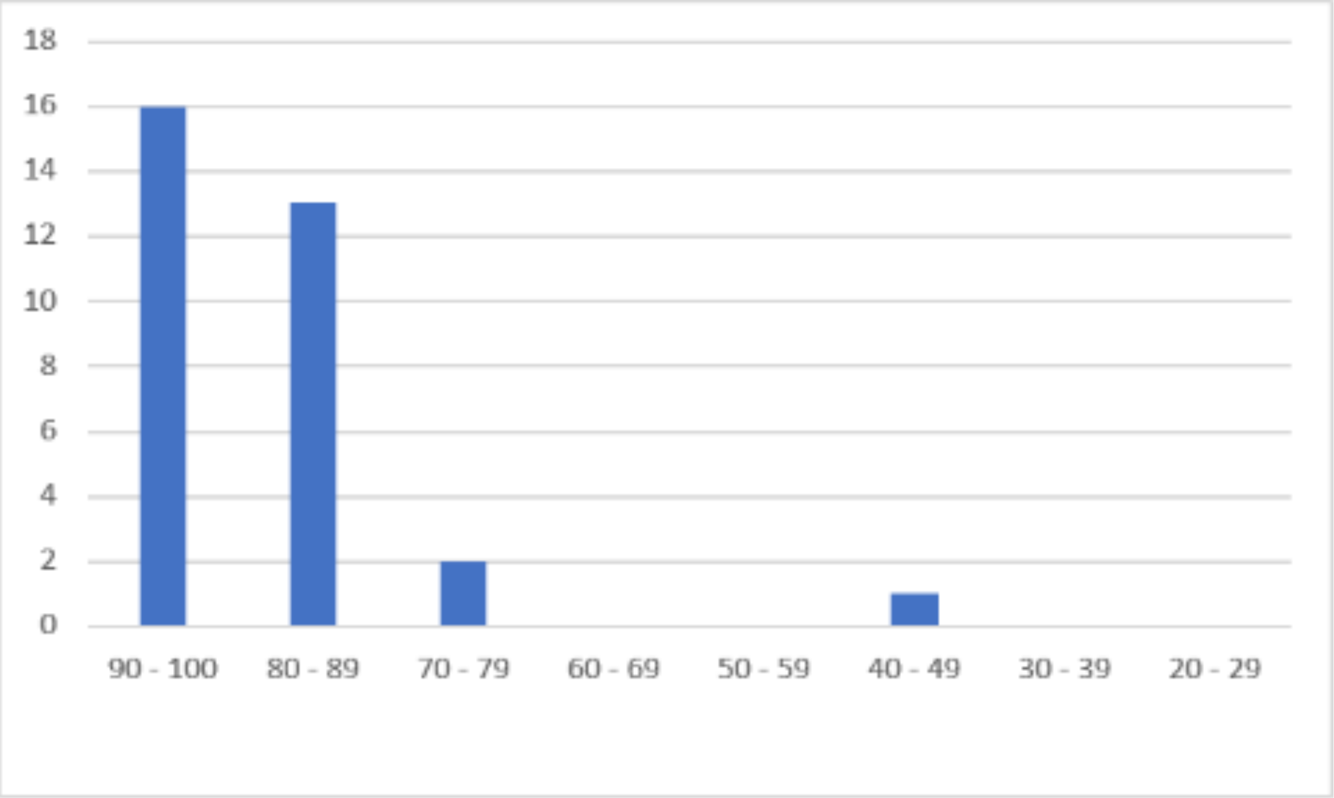
The F statistic for 0.05 is 4.96 so these are not statistically different.

Logistics

- Last Lecture! (You made it!)
- Thursday – final exam review. I will post an example final tonight.
- Assignment 4 is due on Saturday by 11:59.

Logistics

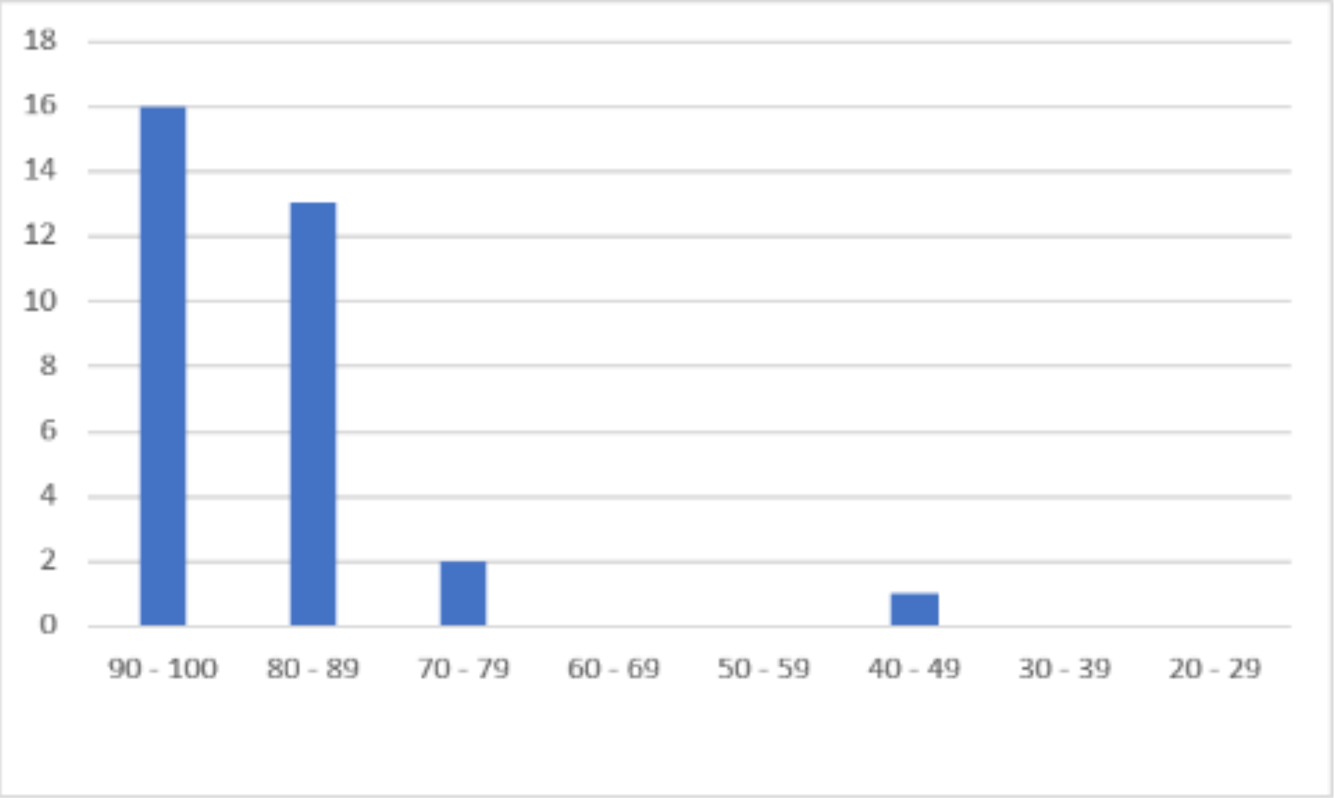
- Assignment 3 has been graded.
- Grades were generally good



Logistics

- Assignment 3 has been graded.
- Grades were generally good

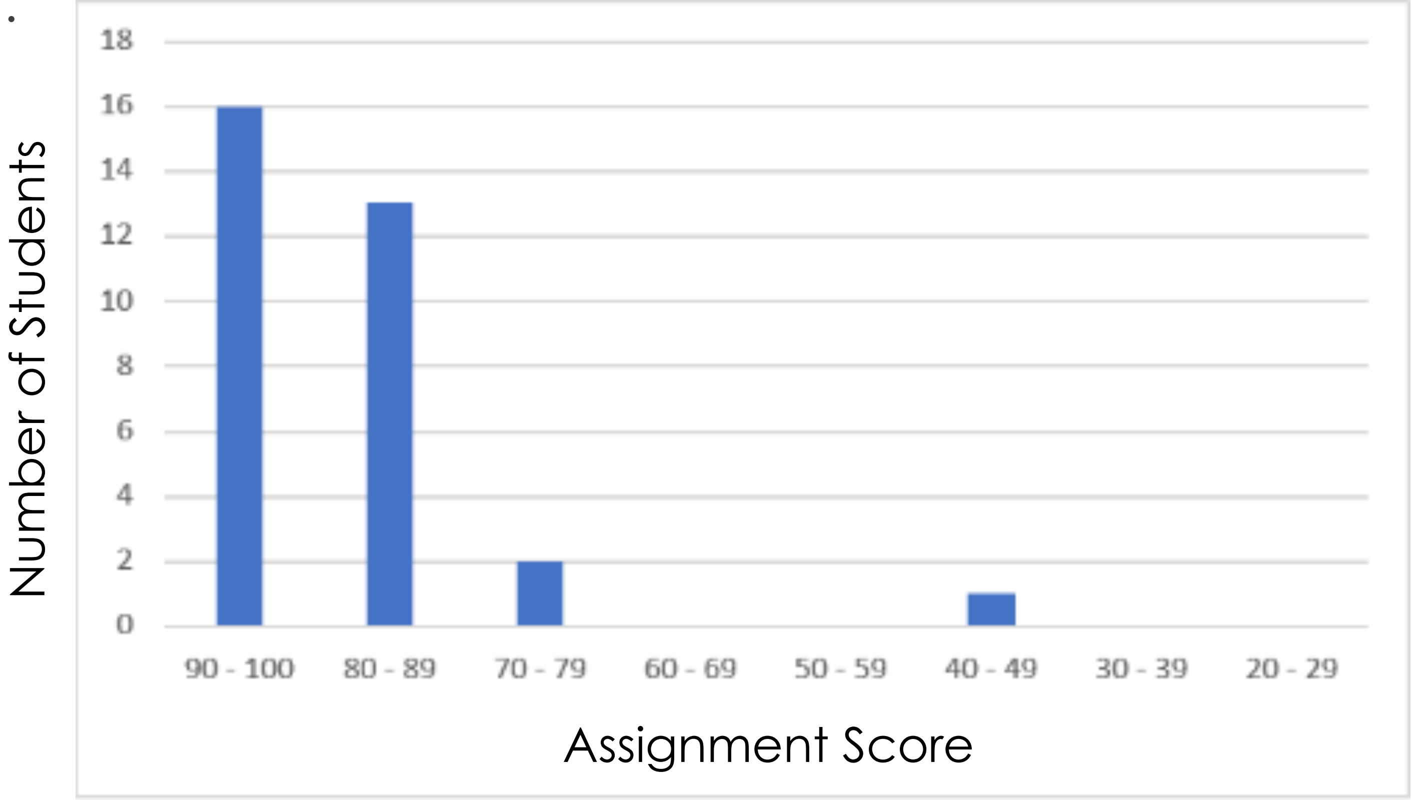
- Some recurring issues:
  1. Label you axes!



Logistics

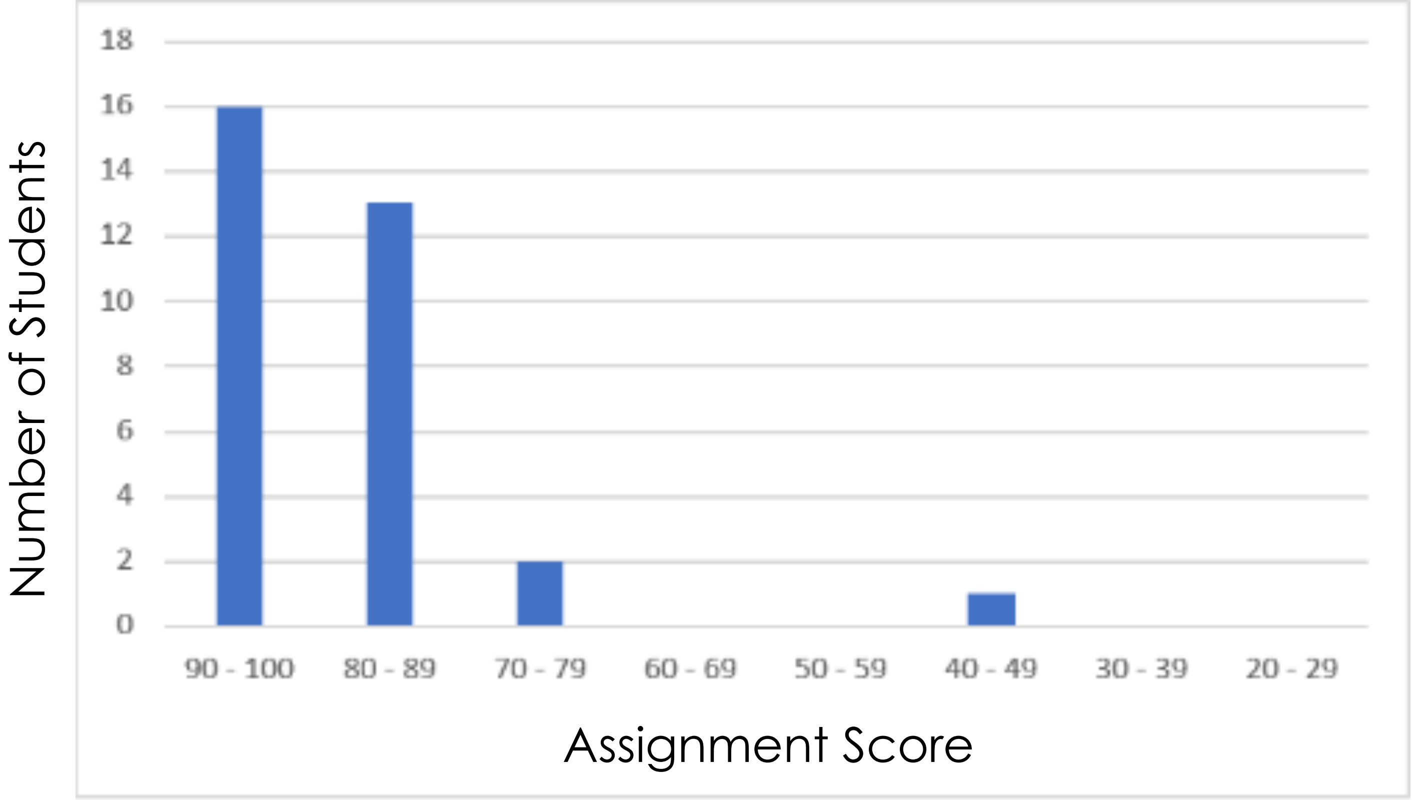
- Assignment 3 has been graded.
- Grades were generally good

- Some recurring issues:
  1. Label you axes!



Logistics

- Assignment 3 has been graded.
- Grades were generally good
- Some recurring issues:
  1. Label you axes!
  2. State **specific** questions.

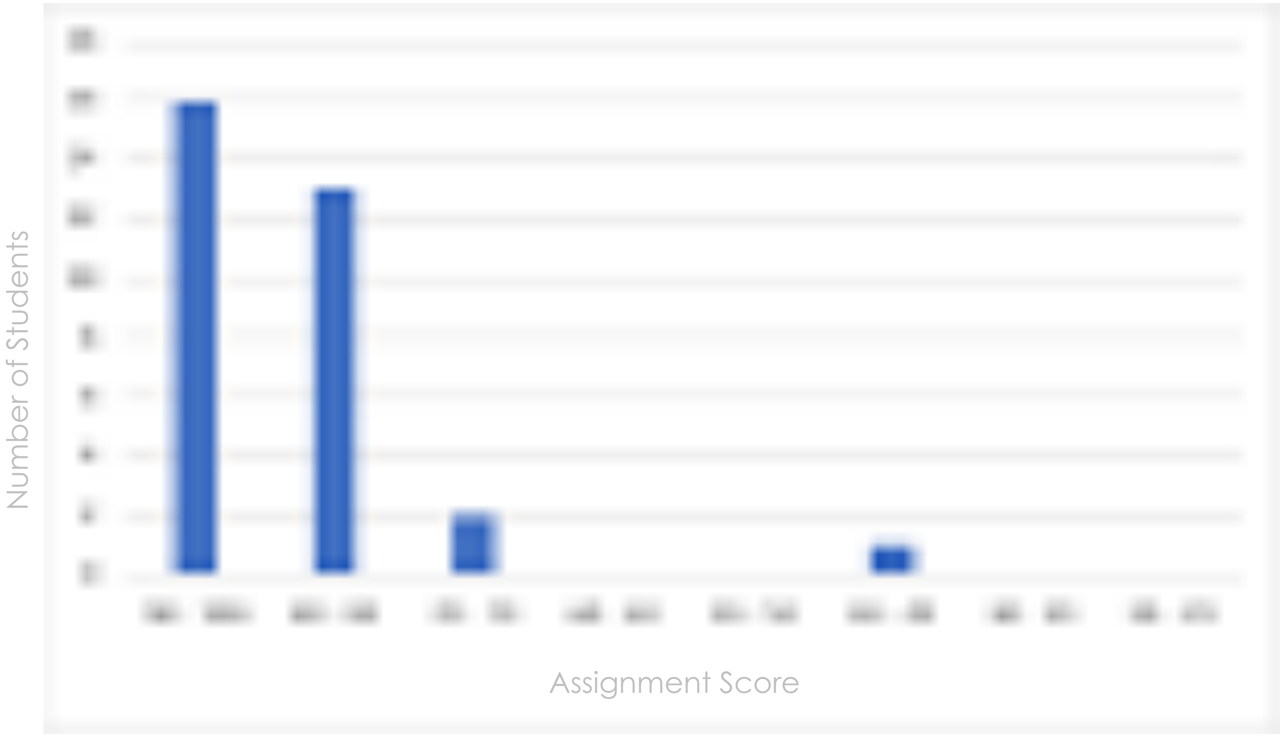


Specific Questions

- "MATLAB is a heavily optimized program for linear algebra so I ask whether it's implementation of matrix multiplication is more efficient than standard C matmul optimized benchmarks."
- Instead of "I examine the relationship between C benchmarks and MATLAB."

Logistics

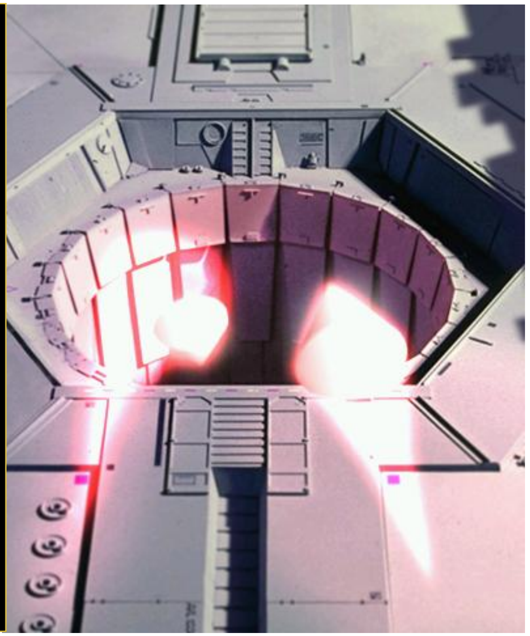
- Assignment 3 has been graded.
- Grades were generally good
- Some recurring issues:
  1. Label you axes!
  2. State **specific** questions.
  3. Make sure your plots are legible.



Logistics

- Assignment 3 has been graded.
- Grades were generally good

- Some recurring issues:
  1. Label you axes!
  2. State **specific** questions.
  3. Make sure your plots are legible.
  4. Avoid digressions from your main questions. Don't use up two pages of tables to make a point that could be summarized in one sentence and that isn't a central question.



Logistics

- Assignment 3 has been graded.
- Grades were generally good

- Some recurring issues:
  1. Label you axes!
  2. State **specific** questions.
  3. Make sure your plots are legible.
  4. Avoid digressions from your main questions. Don't use up two pages of tables to make a point that could be summarized in one sentence and that isn't a central question.
  5. Think of the kittens.

ANOVAs (Part 2)

MATTHEW FRICKE

1.0 SEND CORRECTIONS TO MFRICKE@UNM.EDU

Multiple samples

- We can generalize from testing for differences in population means from part 1 to multiple populations.
- The random selection of independent samples from **p** populations is called "completely randomized experimental design".

Multiple samples

- We can generalize from testing for differences in population means from part 1 to multiple populations.
- The random selection of independent samples from **p** populations is called "completely randomized experimental design".
- Assume the underlying populations are Gaussian with possibly different means but the same variance.

Multiple samples

- We can generalize from testing for differences in population means from part 1 to multiple populations.
- The random selection of independent samples from **p** populations is called "completely randomized experimental design".
- Assume the underlying populations are Gaussian with possibly different means but the same variance.
- We can allow different sample sizes from each population.

$$\mu_i$$
$$\sigma_i^2 = \sigma_{i+1}^2$$

Multiple samples

- Let y_{ij} be the response of the jth experimental unit in the ith sample.
- Let T_i and $\bar{T}_i$ be the total and mean of the ith sample.

Multiple samples

- Let $\mathbf{y}_{ij}$ be the response of the jth experimental unit in the ith sample.
- Let T_i and $\bar{T}_i$ be the total and mean of the ith sample.

$$\mathrm{TSS} = \mathrm{SST} + \mathrm{SSE}$$

Multiple samples

- Let y_{ij} be the response of the jth experimental unit in the ith sample.
- Let T_i and $\bar{T}_i$ be the total and mean of the ith sample.

$$\text{TSS} = \text{SST} + \text{SSE}$$

$$\text{CM (Correction for the mean)} = \frac{(\text{Total for all observations})^2}{n}$$

Multiple samples

$$\text{CM (Correction for the mean)} = \frac{(\text{Total for all observations})^2}{n}$$

Which simplifies straightforwardly to (n times the mean squared)

$$\text{CM} = n\bar{y}^2$$

Multiple samples

$$\text{CM (Correction for the mean)} = \frac{(\text{Total for all observations})^2}{n}$$

Which simplifies straightforwardly to (n times the mean squared)

$$\text{CM} = n\bar{y}^2$$

$$\text{SST} = \sum_{i=1}^{P} \frac{T_i^2}{n_i} - \text{CM}$$

Multiple samples

$$\text{CM (Correction for the mean)} = \frac{(\text{Total for all observations})^2}{n}$$

Which simplifies straightforwardly to (n times the mean squared)

$$\text{CM} = n\bar{y}^2$$

$$\text{SST} = \sum_{i=1}^{P} \frac{T_i^2}{n_i} - \text{CM} \qquad \text{SSE} = \text{TSS} - \text{SST}$$

Multiple samples

$$\text{CM (Correction for the mean)} = \frac{(\text{Total for all observations})^2}{n}$$

Which simplifies straightforwardly to (n times the mean squared)

$$\text{CM} = n\bar{y}^2$$

$$\text{SST} = \sum_{i=1}^{P} \frac{T_i^2}{n_i} - \text{CM} \qquad \text{SSE} = \text{TSS} - \text{SST}$$

Multiple samples

The estimate of variance based on the number of degrees of freedom (numerator below) is (Mean Squares for Error):

$$\mathrm{MSE} = \frac{\mathrm{SSE}}{n_1 + n_2 + \cdots + n_p - p}$$

Multiple samples

$$\mathrm{MSE} = \frac{\mathrm{SSE}}{n_1 + n_2 + \cdots + n_p - p}$$

The mean square for treatments will have p-1 degrees of freedom, i.e. one less than the number of means.

$$\mathrm{MST} = \frac{\mathrm{SST}}{p - 1}$$

Multiple samples

So to test the null hypothesis that all the means are equal we can use the test statistic:

$$\mathrm{F} = \frac{\mathrm{MST}}{\mathrm{MSE}} > \mathrm{F}_\alpha$$

Multiple samples

So to test the null hypothesis that all the means are equal we can use the test statistic:

$$F = \frac{\mathrm{MST}}{\mathrm{MSE}} > F_\alpha$$

All this is to say that the greater that difference between the sample means the greater the probability that the population means are different.

Multiple samples

So to test the null hypothesis that all the means are equal we can use the test statistic:

$$\mathrm{F} = \frac{\mathrm{MST}}{\mathrm{MSE}} > \mathrm{F}_{\alpha}$$

All this is to say that the greater that difference between the sample means the greater the probability that the population means are different.

Multiple samples

Departing from the assumptions do not usually cause problems. But it is important to know the assumptions.

Again: Gaussian distributed populations, random selection of samples from populations, and equal variance.

ANOVA Table

Source	DoF	SS	MS	F
Treatments	p-1	SST	MST=SST/(p-1)	MST/MSE
Error	n-p	SSE	MSE=SSE/(n-p)	
Total	n-1	TSS		

ANOVA Table Example

```
>> y = [52.7 57.5 45.9 44.5 53.0 57.0 45.9 44.0]';
g1 = [1 2 1 2 1 2 1 2];
g2 = {'hi';'hi';'lo';'lo';'hi';'hi';'lo';'lo'};
g3 = {'may';'may';'may';'may';'june';'june';'june';'june'};
>>
```

ANOVA Table Example

```
>> y = [52.7 57.5 45.9 44.5 53.0 57.0 45.9 44.0]';
g1 = [1 2 1 2 1 2 1 2];
g2 = {'hi';'hi';'lo';'lo';'hi';'hi';'lo';'lo'};
g3 = {'may';'may';'may';'may';'june';'june';'june';'june'};
>>

>> p = anovan(y,{g1,g2,g3})

p =

    0.4174
    0.0028
    0.9140

>>
```

MATLAB ANOVA Table

Figure 1: N-Way ANOVA

File Edit View Insert Tools Desktop Window Help

Analysis of Variance

Source	Sum Sq.	d.f.	Mean Sq.	F	Prob>F
X1	3.781	1	3.781	0.82	0.4174
X2	199.001	1	199.001	42.95	0.0028
X3	0.061	1	0.061	0.01	0.914
Error	18.535	4	4.634		
Total	221.379	7			

ANOVAs for Block Design

The randomized block design implies we have two qualitatively different independent variables “blocks” and ”treatments”.

So now we have the sum of squares for blocks, treatments, and error.

$$\text{TSS} = \text{SSB} + \text{SST} + \text{SSE}$$

With b blocks and p treatments.

ANOVAs for Block Design

$$\text{TSS} = \text{SSB} + \text{SST} + \text{SSE}$$

With b blocks and p treatments.

The TSS, SST, and SSE are calculated as before. Except:

$$\text{SSB} = \frac{\sum_{i=1}^{b} B_i^2}{p} - \text{CM}$$

ANOVAs for Block Design

$$\text{TSS} = \text{SSB} + \text{SST} + \text{SSE}$$

The TSS, SST, and SSE are calculated as before. Except:

$$\text{SSB} = \frac{\sum_{i=1}^{b} B_i^2}{p} - \text{CM} \quad \text{and SST} = \frac{\sum_{j=1}^{p} T_i^2}{b} - \text{CM}$$

ANOVA Table for Block Design

Source	DoF	SS	MS	F
Blocks	b-1	SSB	MSB=SSB/(b-1)	MSB/MSE
Treatments	p-1	SST	MST=SST/(p-1)	MST/MSE
Error	n-b-p+1	SSE	MSE=SSE/(n-p)	
Total	n-1	TSS		

Notice we can ask about block and treatment effects

Tukey’s Honestly Significantly Different test

- ANOVA tells you there is a significant effect but not which pairs of treatments are producing the effect.
- This is a post-hoc test. So you use it if you have p-values that indicate the null hypothesis can be rejected in your ANOVA table.
- The HSD test looks at the pairwise difference in means divided by a proxy for the variance within the treatments.
- The test uses the q statistic which you can find values for in a table – or use MATLAB.

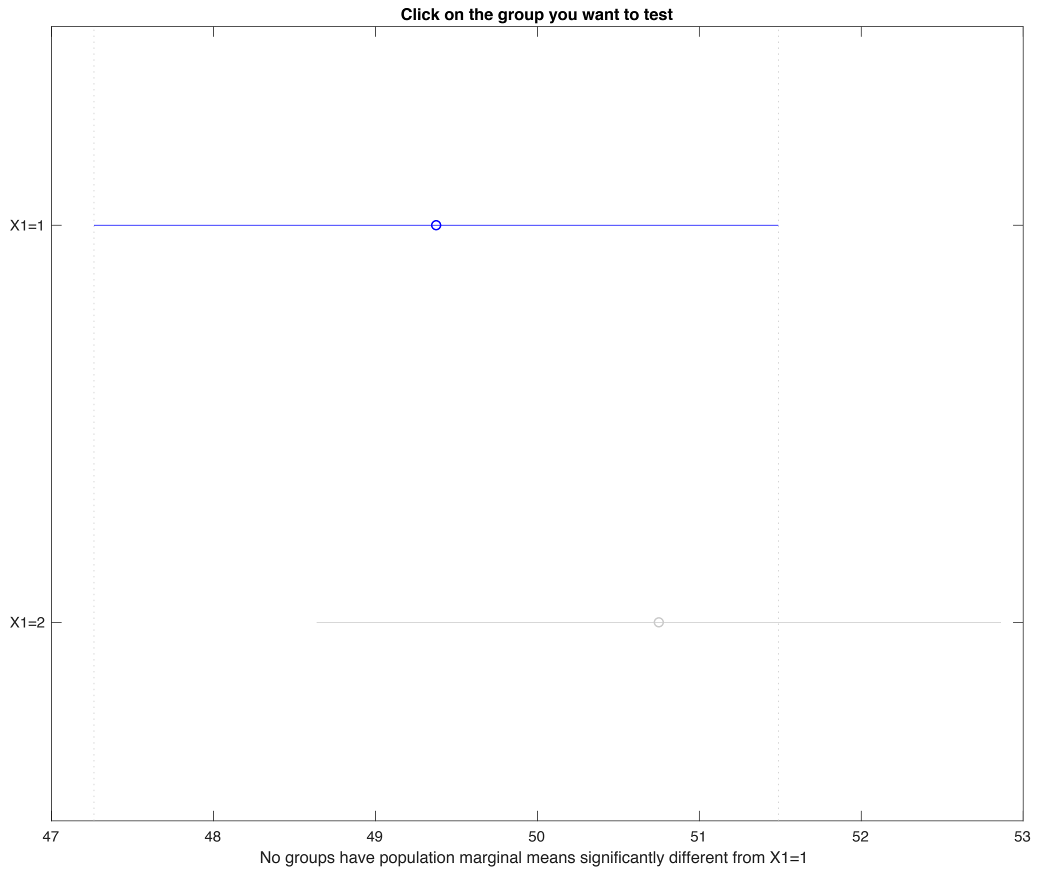
For more see Herv´e Abdi, Lynne J. Williams https://www.utdallas.edu/~herve/abdi-HSD2010-pretty.pdf in In Neil Salkind (Ed.), Encyclopedia of Research Design. Thousand Oaks, CA: Sage. 2010

Tukey's Honestly Significantly Different test

```
>> [p,t,stats] = anovan(y,{g1,g2,g3})

...

[c,m,h,nms] = multcompare(stats);
```



Cohen’s Effect Size

- A study of 14,000 children ages 6-17 showed a “highly significant” ($p < .001$) correlation of r = .11) between height and IQ
- What does this p indicate?
- What’s the magnitude of this correlation?
  - Accounts for 1% of the variance

Cohen’s Effect Size

- A study of 14,000 children ages 6-17 showed a “highly significant” ($p < .001$) correlation of $r = .11$) between height and IQ
- What does this p indicate?
- What’s the magnitude of this correlation?
  - Accounts for 1% of the variance
  - Based on an r this big, you’d expect that increasing a child’s height by 4 feet would increase IQ by 30 points, and that increasing IQ by 233 points would increase height by 4 inches (as a correlation, the predicted relationship could work in either direction)

The height-IQ correlation: Cohen 1990

Cohen's Effect Size

- A study of 14,000 children ages 6-17 showed a "highly significant" ($p < .001$) correlation of $r = .11$) between height and IQ
- What does this p indicate?
- What's the magnitude of this correlation?
  - Accounts for 1% of the variance
  - Based on an r this big, you'd expect that increasing a child's height by 4 feet would increase IQ by 30 points, and that increasing IQ by 233 points would increase height by 4 inches (as a correlation, the predicted relationship could work in either direction)
  - The Effect Size does NOT tell us whether there is a real effect! P-value does that.

The height-IQ correlation: Cohen 1990

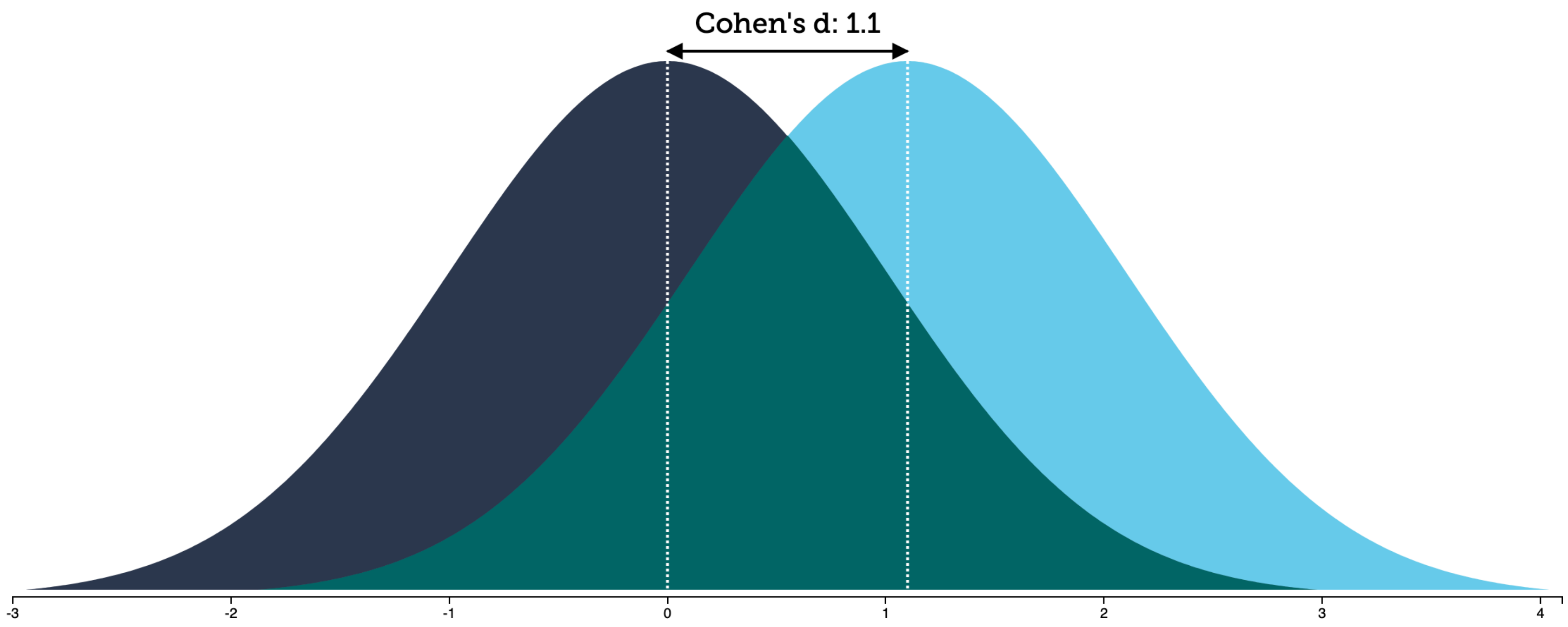
Cohen’s Effect Size

- To calculate Cohen’s effect size:
- **For any two treatment groups, find treatment means, subtract them, and divide by the standard deviation**

The height-IQ correlation: Cohen 1990

Cohen's Effect Size Visual

- https://rpsychologist.com/d3/cohend/



Experimental CS Examples

MATTHEW FRICKE

1.0 SEND CORRECTIONS TO MFRICKE@UNM.EDU

Approximation algorithms for Bin Packing

- Given a list of n weights drawn uniformly at random from the range (0,1), what is the expected asymptotic packing ratio (compared to optimal) of simple algorithms like First Fit, Best Fit, First Fit Decreasing, and Best Fit Decreasing?
- A series of experiments by Jon Bentley and others suggested that all four are asymptotically optimal (which contradicted previous conjectures), and also revealed patterns that inspired the arguments used in the proofs

On Experimental Algorithmics, An Interview with Catherine McGeoch and Bernard Moret
by Richard T. Snodgrass

Dijkstra’s algorithm

- Traditional worst case analysis of Dijkstra’s algorithm for the shortest paths problem shows that the decrease-key operation dominates, and much design work has gone into creating data structures (such as Fibonacci heaps) that reduce the worst case bound.
- Experiments by Andrew Goldberg et al. suggested that, for a large category of input graphs, the decrease-key operation is rare—a property which they went on to prove.
- Those good-worst-case data structures optimized the wrong thing in many cases.

Insufficiency of the RAM Model

- The simple RAM model does not predict computation times on modern machines with sufficient accuracy because it does not take the memory hierarchy into account.
- Anthony LaMarca and Richard Ladner developed experiments to guide their design of a new two-level model of computation that captures the interactions between caches and main memory.
- They reanalyzed classic algorithms (sorting) and data structures (heaps) under the new model; their analyses are much closer to experience, and in some cases flatly contradict conventional design wisdom based on traditional analyses.

Insufficiency of the RAM Model

- The LaMarca and Ladner work was predated by a long and rich history of experimental and theoretical efforts—carried out by both the theory and the systems communities since around 1966—to develop two-level models of computation that describe algorithm performance in virtual memory systems. Peter Denning has a nice article that describes how theory and experiments contributed to develop our understanding of how locality of reference affects computation time [9]. A separate thread of research into cost models for I/O-bound computation has been equally fruitful.