
Lab in Human Cognition

Todd M. Gureckis
Department of Psychology
New York University

Agenda for Today

- Quick background on ANOVA
- How to report an ANOVA in your paper
- What to put in Lab 2 write up
- Time to work on Lab 2

First

Background on ANOVA

Variables

- **Dependent variables** - typically the one we are observing (can be continuous, multivariate, etc...)
- **Independent variable** - the ones we manipulated in the experiment (can be qualitative, continuous, or based on things like individual differences)
- **Confounding variable** - not held constant or manipulated but could also impact our dependent variable
- **Control variable** - potential nuisance, held constant

t-test revisited

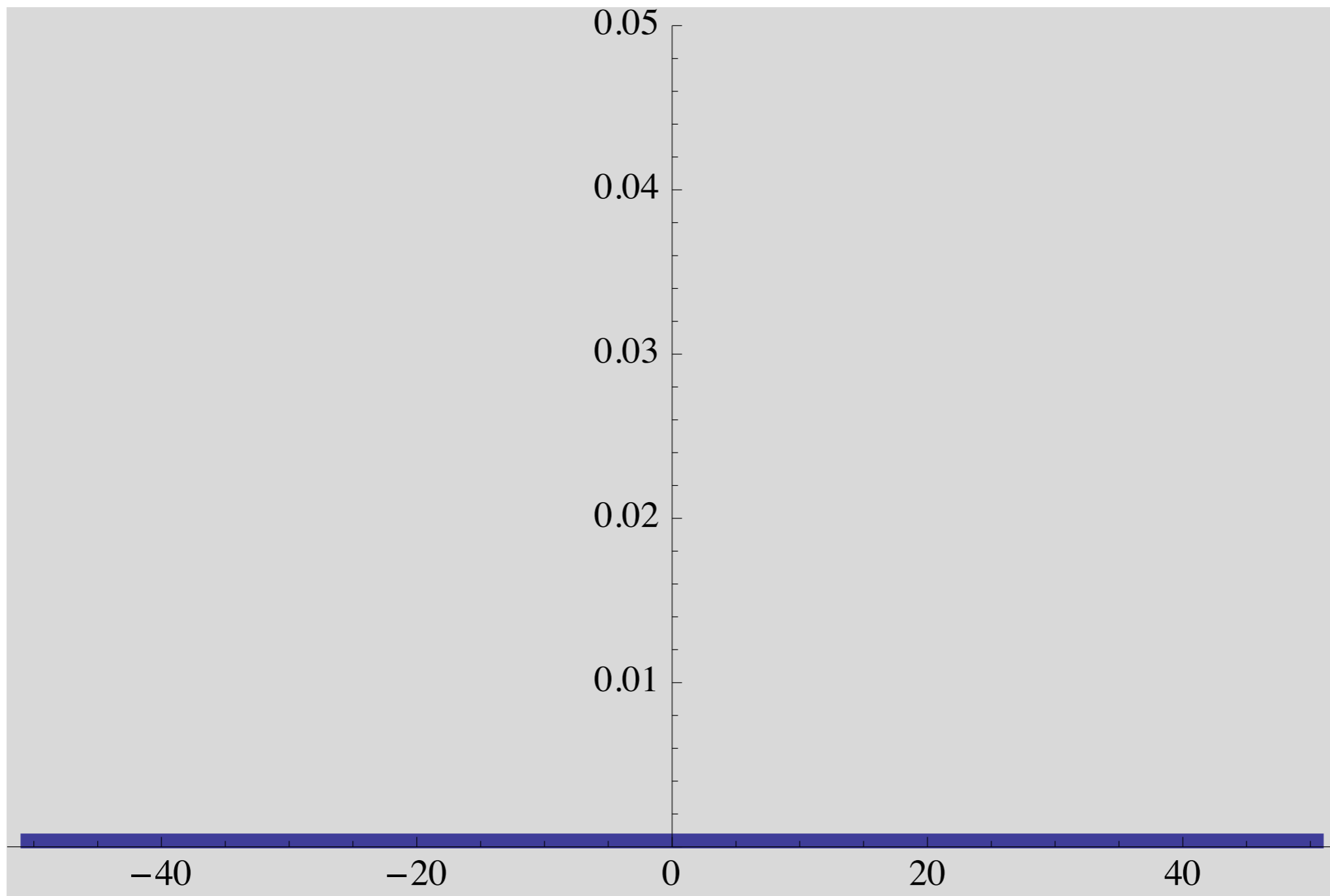
- So we discussed the t-test extensively so far
- There are three varieties of t-test:
 - One-sample
 - Two sample (paired) - really the same as one sample but on the differences
 - Two sample (independent populations)

Pooled variance two-sample t-test

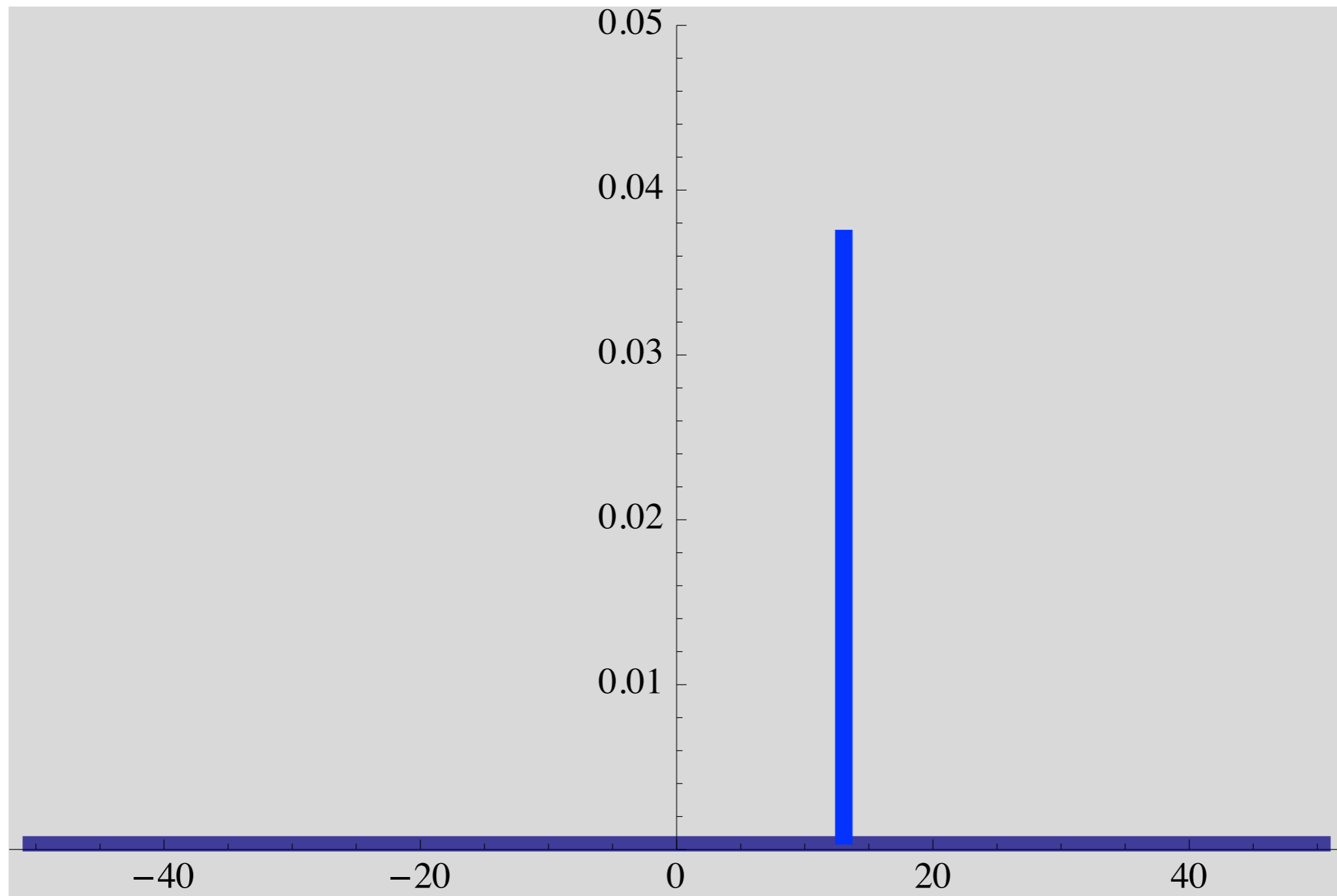
- Examples:
 - Is there a difference in the stroop effect for liberals versus conservatives?
 - Do people with Parkinson's disease learn at a lower rate in a cognitive skill learning test than do normal college undergraduates?
 - In these cases, the groups are not paired (i.e., they are not the same individual tested twice)
 - Thus, we can't just take the difference of the scores and do a one-sample t-test

Pooled variance two-sample t-test

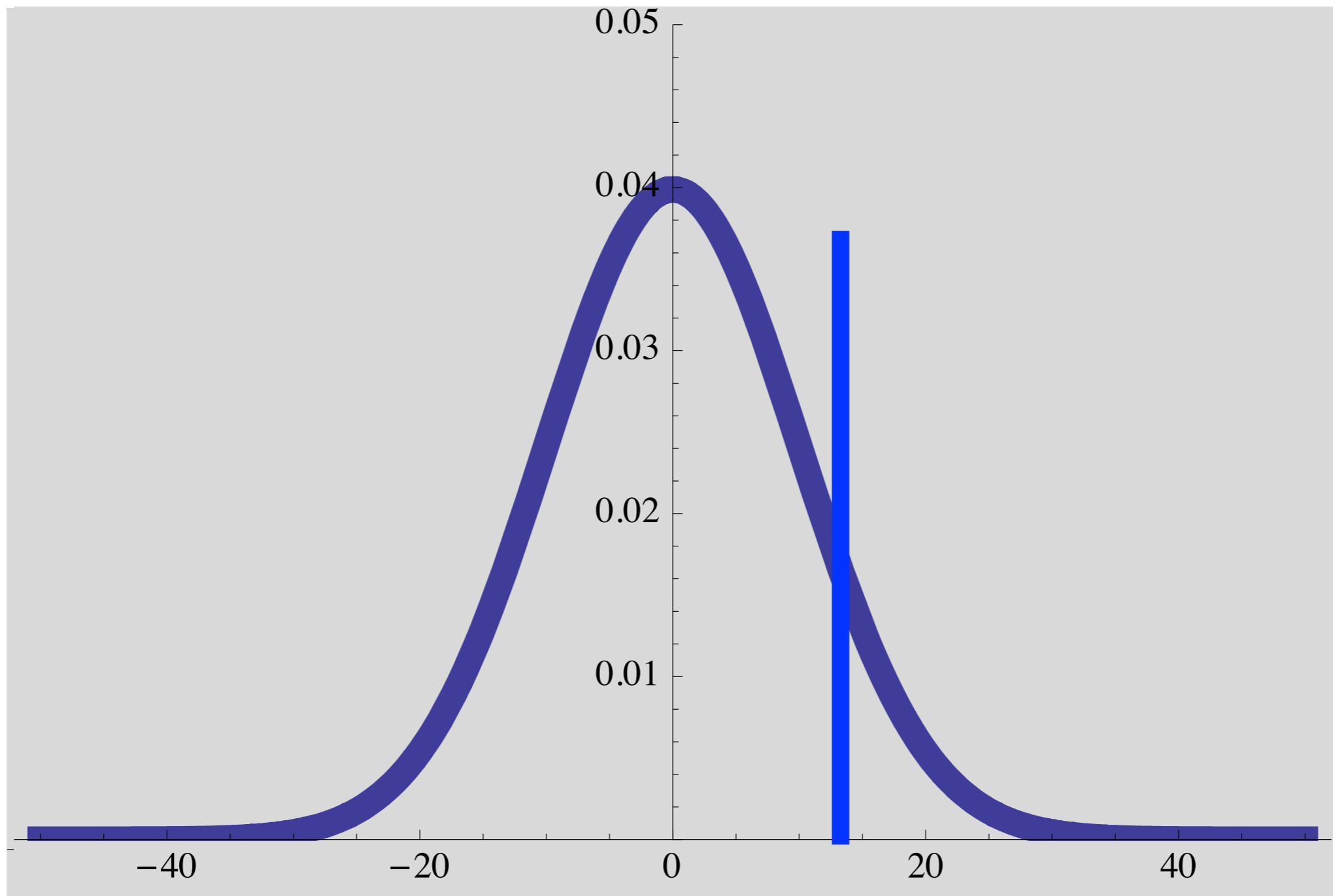
- What do we do?
- Use the logic for the original t-test
 - Compute the difference in the means of the two groups
 - Convert that to a t-value by dividing by the standard error
 - However, in this case, we use what is called a “pooled” estimate of the variance from both groups



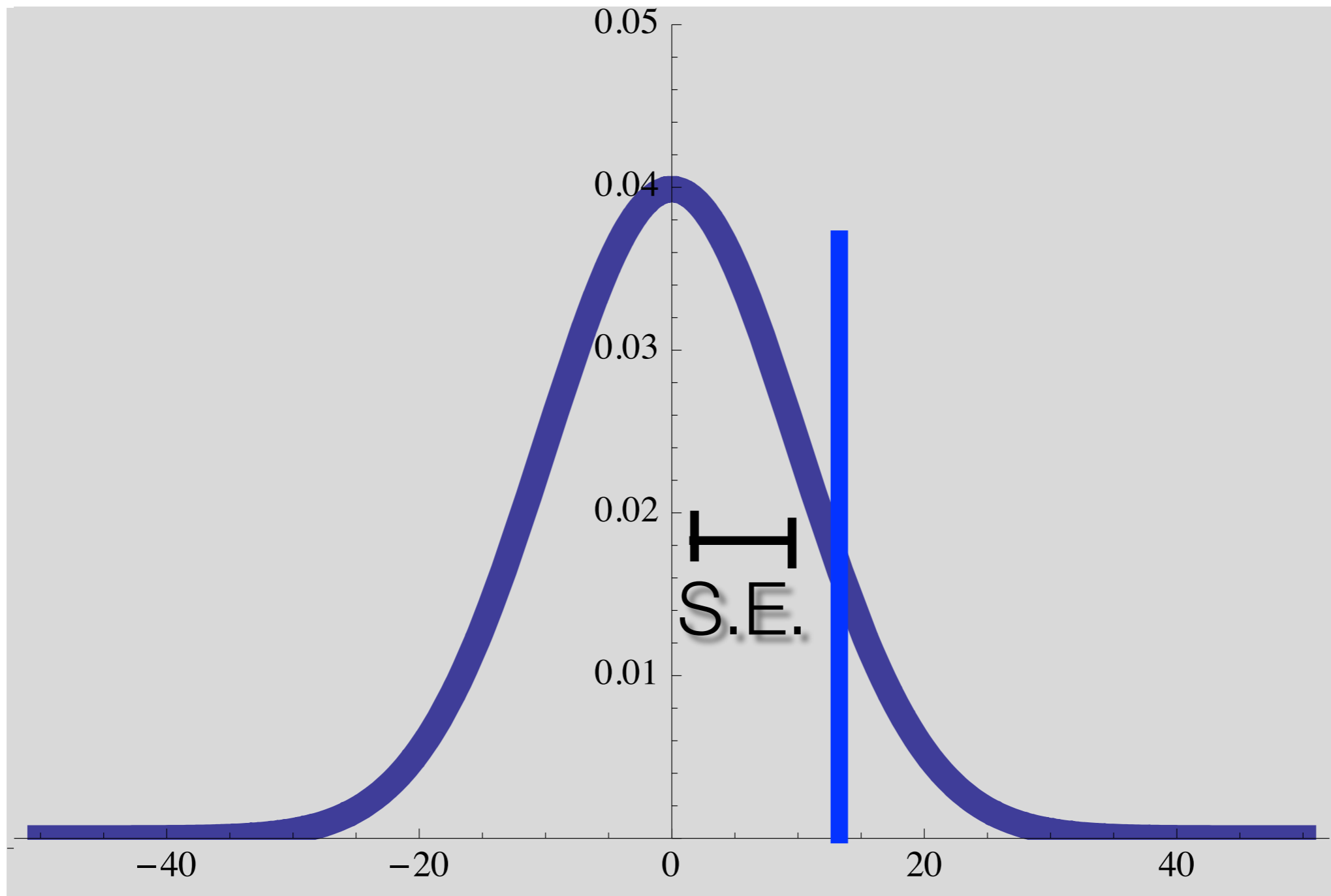
Difference in Means



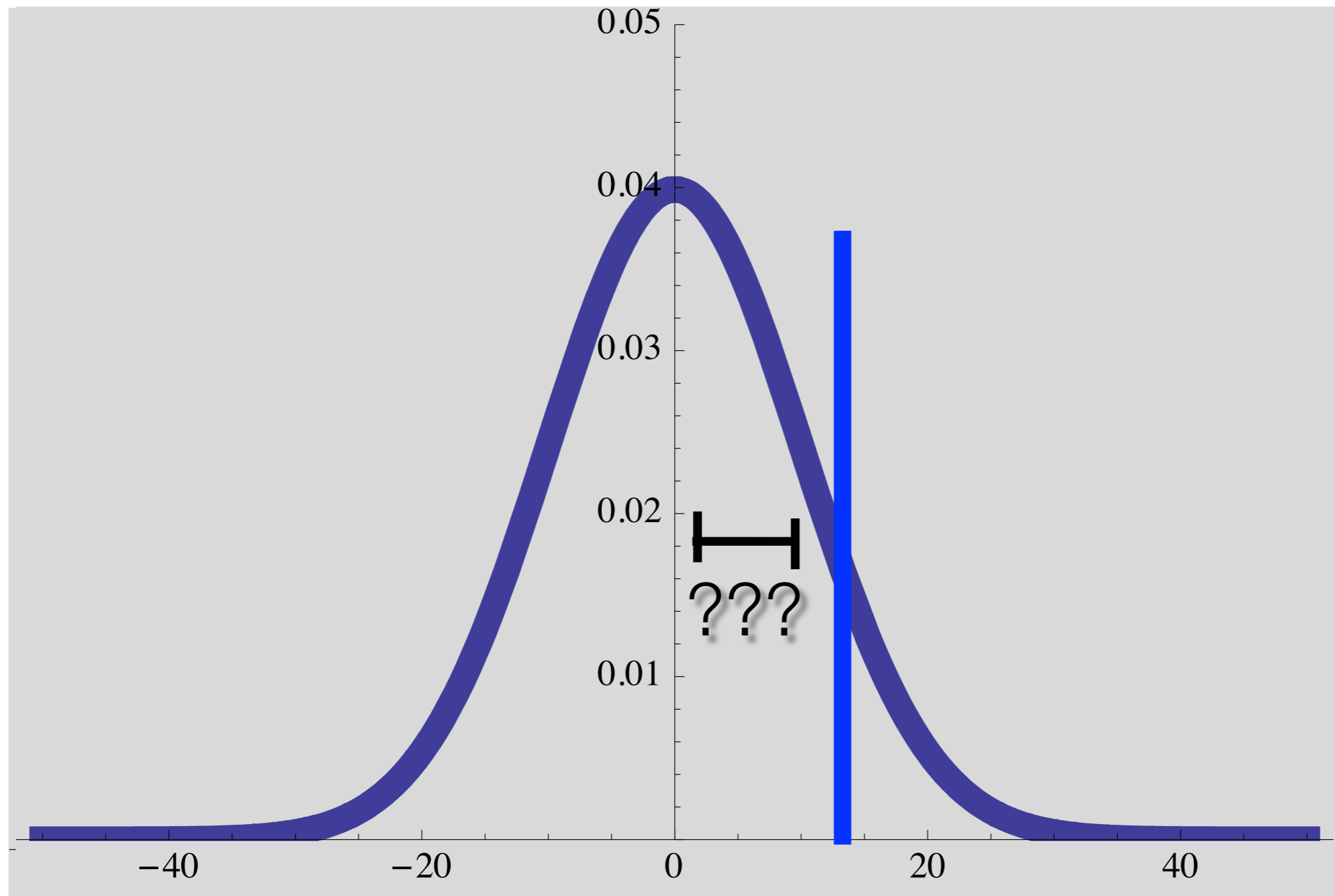
Difference in Means



Difference in Means



Difference in Means



Difference in Means

Pooling Means

Pooled Means

$$\bar{x}_{12} = \frac{n_1 \bar{x}_1 + n_2 \bar{x}_2}{n_1 + n_2}$$

- A “weighted” average of the mean... weight by n of the sample

Pooled Variance

$$s_{12}^2 = \frac{(n_1 - 1) s_1^2 + (n_2 - 1) s_2^2}{(n_1 - 1) + (n_2 - 1)}$$

- A “weighted” average of the individual variances... weighted by $n-1$ (the degrees of freedom that go into each individual variance).

our “old” t-value calculation

$$t = \frac{\bar{x}_1 - \bar{x}_2}{s / \sqrt{n}}$$

- degrees of freedom = $n - 1$

t-value formula in terms of variance (instead of s.d.)

$$t = \frac{\bar{x}_1 - \bar{x}_2}{s_{12} \sqrt{\frac{1}{n_1} + \frac{1}{n_2}}}$$

- degrees of freedom = $n_1 + n_2 - 2$
- (assumes equal variances)

wait!

**what about more than two
groups?**

ANOVA - Analysis of Variance

- Situation: We control one or more independent variables
 - Also known as “factors” (see R)
 - Must contain two or more *levels* (sub categories)
- Observe an effect on the dependent variable

Randomized Design

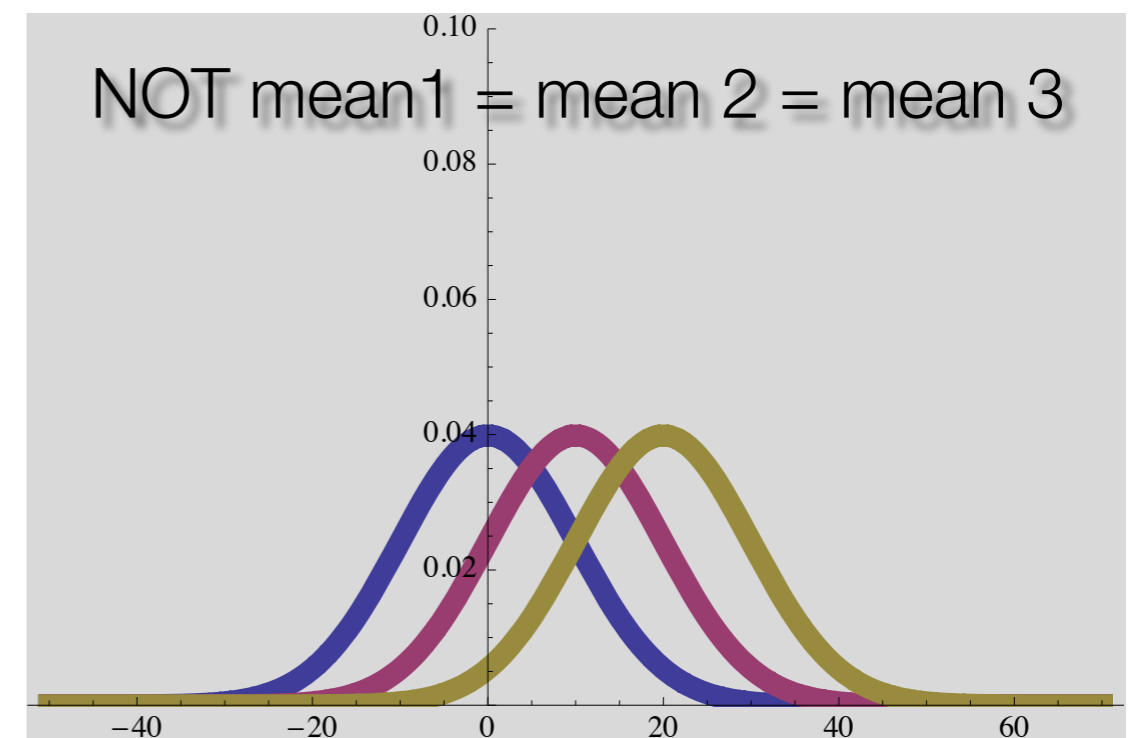
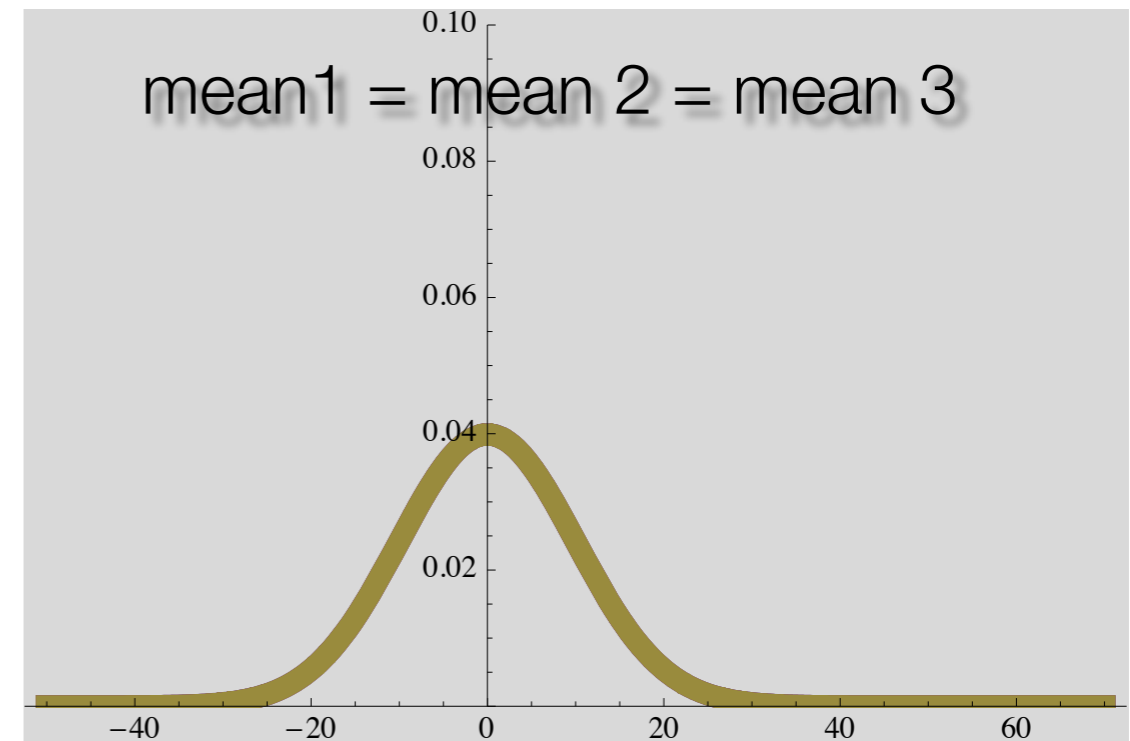
- Assign people (or trials) randomly to the treatment groups
- For example, in our experiment, trials were randomly assigned to be COLOR-UPRIGHT, COLOR-UPSIDEDOWN, etc...
- The manipulations are our factors, and the different values (e.g., “upright”, “upside down”) are the *levels*

ANOVA assumptions

- **Randomized** - Samples in each treatment group are random and independent
- **Normality** - the populations follow a normal distribution (this isn't always true, but the central limit theorem helps us)
- **Homogeneity of variance** - simply means the populations or conditions have equal variance. Harder to control.

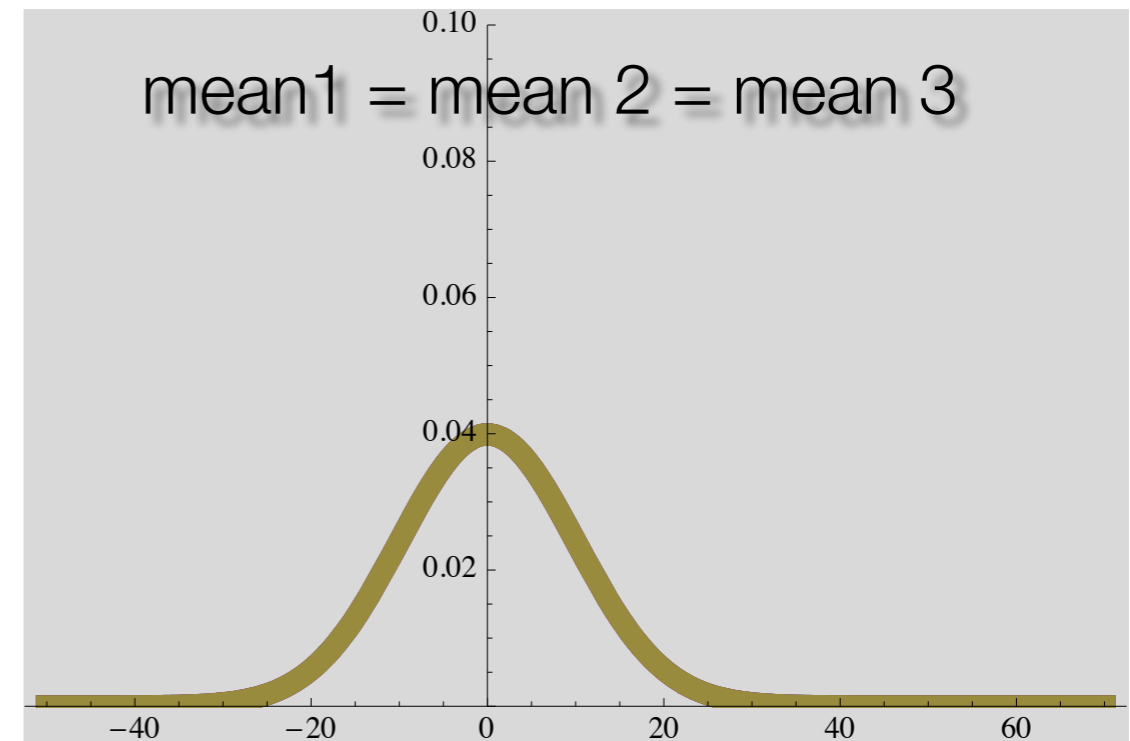
The null hypothesis in the ANOVA

- **Null hypothesis** - the means of each group are *all* equal
- **Alternative hypothesis** - the means are not all equal (at least 1 but maybe more are “different”).



The basic idea

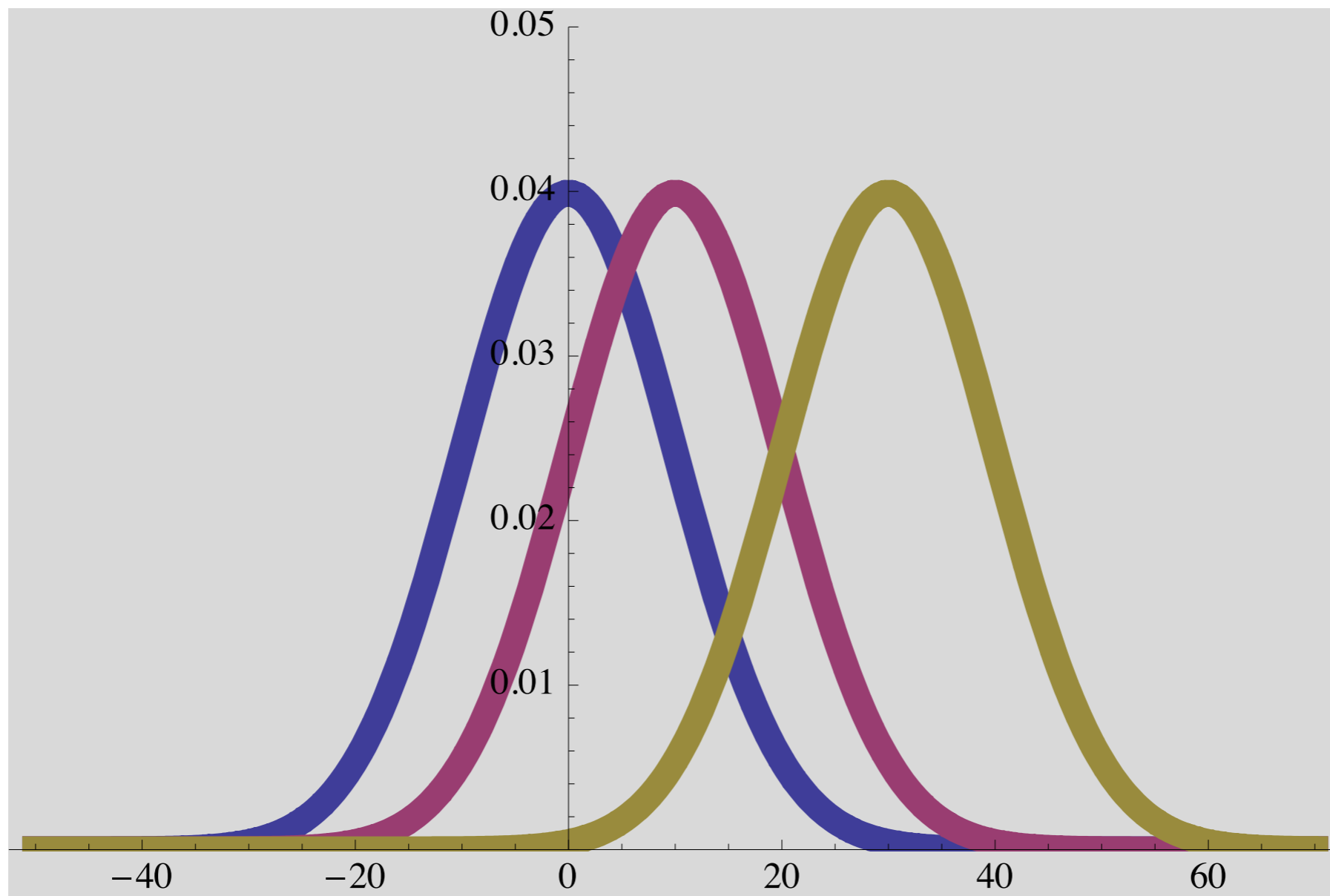
- Compare 2 types of variation to test the equality of means
- **Ratio of variances**
- If the variance between the different treatment groups is greater than the (assumed) random variation within each group, then we conclude the means are not equal
- The variance that go into this are computed by “partitioning” total variation into subsets



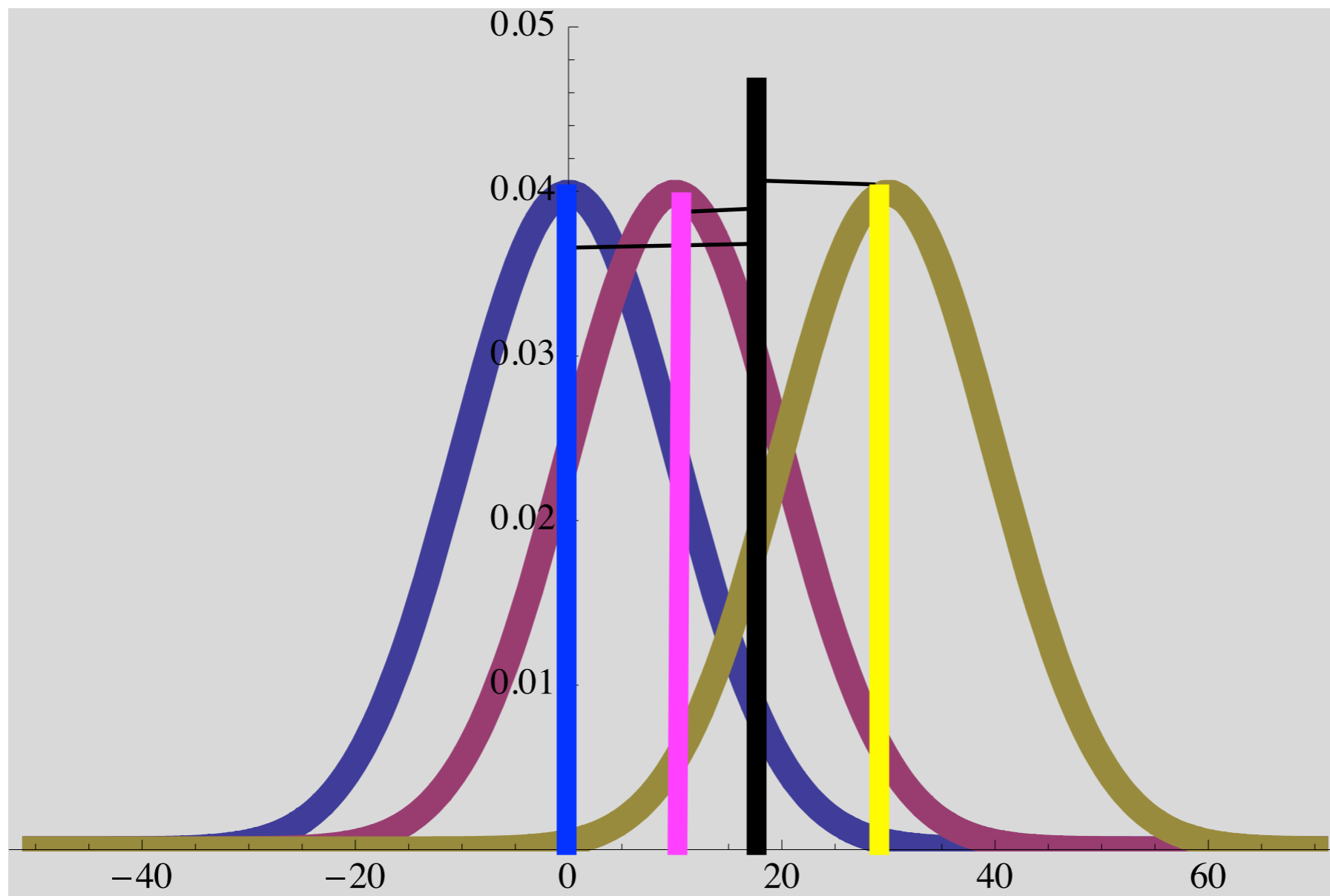
Partitioned Variance Idea



Partitioned Variance Idea

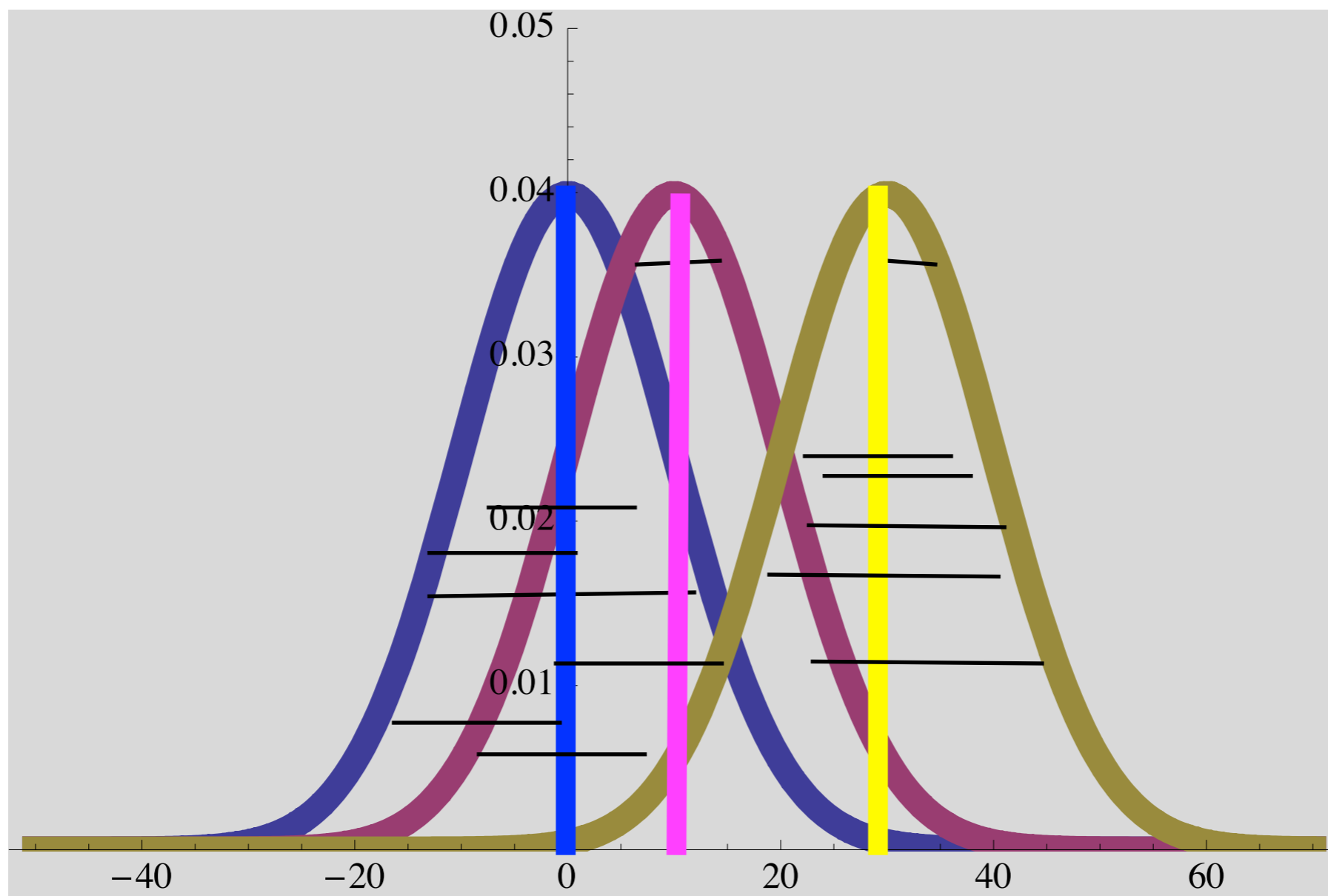


Partitioned Variance Idea



Between-group variance

Partitioned Variance Idea



Between-group variance

ANOVA

$$F = \text{MSB} / \text{MSW}$$

MSB = mean squared error between

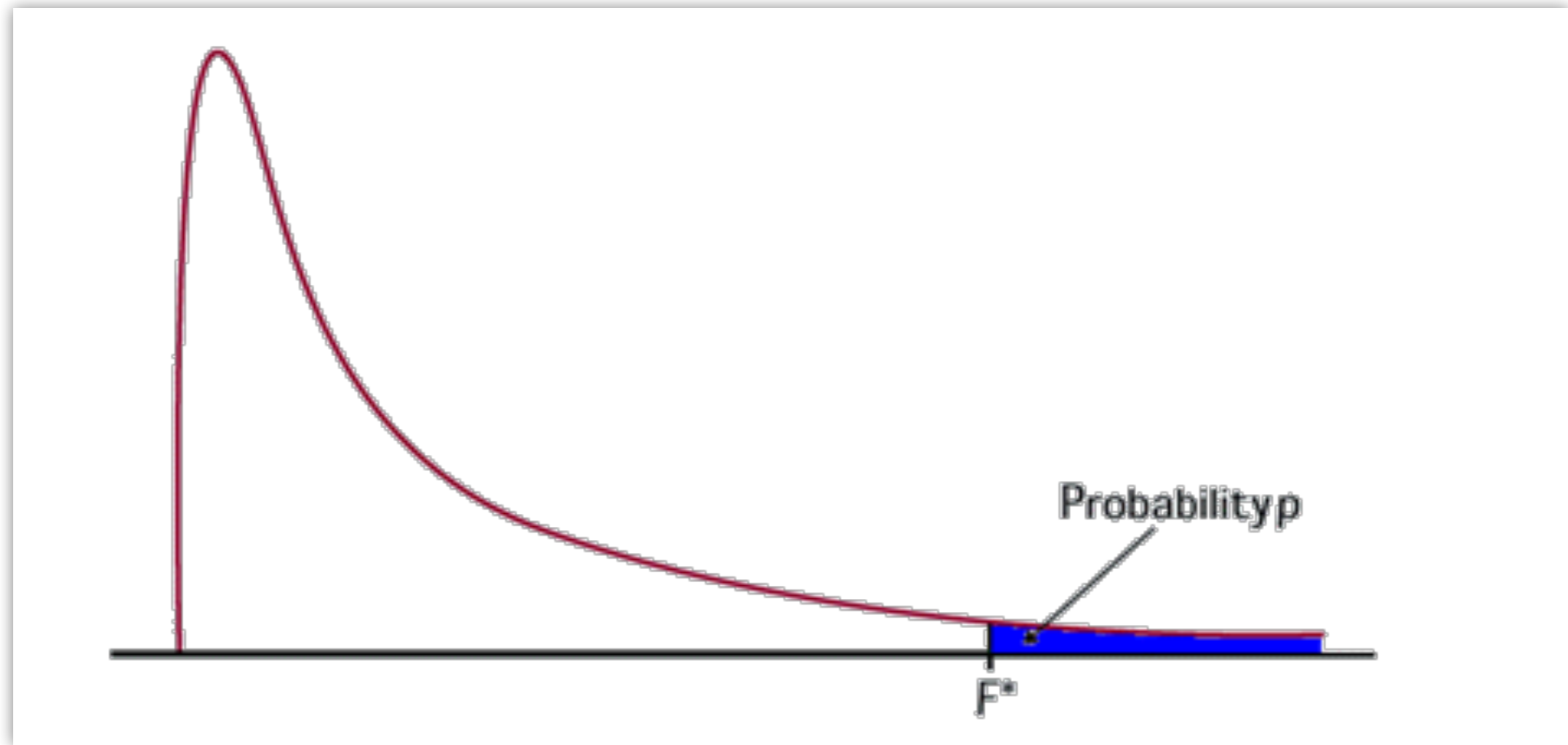
MSW = mean squared error within

df of MSB is # of groups - 1

df of MSW is # of samples - # of groups

ANOVA

$$F = \text{MSB} / \text{MSW}$$



ANOVA

| <u>Source</u> | <u>SS</u> | <u>df</u> | <u>MS (s^2)</u> | <u>F</u> | <u>p</u> | <u>Decision</u> |
|---------------|----------------------|----------------------|------------------------------|----------------------|-----------------------|----------------------|
| Between | <input type="text"/> | <input type="text"/> | <input type="text"/> | <input type="text"/> | <input type="text"/> | <input type="text"/> |
| Within | <input type="text"/> | <input type="text"/> | <input type="text"/> | | | |
| Total | <input type="text"/> | <input type="text"/> | | | | |

How to report

A one-way (repeated measures) ANOVA on trial type revealed a significant effect the presented stimulus on participant's reaction time, $F(1, 64)=14.07, p<.001$.

How to report

| <u>Source</u> | <u>SS</u> | <u>df</u> | <u>MS (s^2)</u> | <u>F</u> | <u>p</u> | <u>Decision</u> |
|---------------|-----------|-----------|------------------------------|----------|-----------------------|-----------------|
| Between | | | | | | |
| Within | | | | | | |
| Total | | | | | | |

$$F(1, 64) = 14.07, p < .001.$$

df from table

F-stat from table

p value (same rules as before)

How to report

```

> myaov=aov(byslot$rt~byslot$cond)
> summary(myaov)
              Df Sum Sq Mean Sq F value    Pr(>F)
byslot$cond   2 165803   82902   6.9621 0.002779 **
Residuals    36 428672   11908
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

```

$F(2, 36)=6.96, p<.0001.$

How to report

```
> summary(myaov)
```

```
Error: byslot$subj
```

```
      Df Sum Sq Mean Sq F value Pr(>F)
Residuals 12 374819    31235
```

```
Error: Within
```

```
      Df Sum Sq Mean Sq F value    Pr(>F)
byslot$cond  2 165803    82902  36.946 4.716e-08 ***
Residuals   24  53853     2244
```

```
---
```

```
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```
>
```

$F(2, 24)=36.96, p<.0001.$

How to report

```
> summary(myaov)
```

Between-subject SSE

```
Error: byslot$subj
      Df Sum Sq Mean Sq F value Pr(>F)
Residuals 12 374819    31235
```

```
Error: Within
```

```
      Df Sum Sq Mean Sq F value    Pr(>F)
byslot$cond  2 165803    82902  36.946 4.716e-08 ***
Residuals   24  53853     2244
```

```
---
```

```
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```
>
```

$F(2, 24)=36.96, p<.0001.$

Repeated measures ANOVA

