Hypothesis Tests
Analysis of Variance (ANOVA)

DASC 512

Overview

- Introduction to Design of Experiments (DoE)
- One-Factor Experiments (One-Way ANOVA)
- Two-Factor Experiments (Two-Way ANOVA)
- Many-Factor Experiments

Introduction to DoE

Design of Experiments

Observational study vs. Designed experiment

Comparisons of two population means often result from observational studies, when outcomes are observed but factors aren’t controlled.

In a designed experiment, researchers attempt to control the levels of one or many factors of interest to determine their effect on an outcome of interest.

If factors are sufficiently controlled, this allows us to infer causal linkage

Definitions

Response Variable (Dependent Variable): Variable of interest to be measured in the experiment

Factors (Independent Variables): Variables whose effect on the response is of interest to the experimenter (may be qualitative or quantitative)

- Factors are typically controlled but are measured even if control is impossible

Levels (Factor Levels): Values of the factor(s) used in the experiment

Treatments: The factor level combinations used in the experiment

Observational/Experimental Unit: Object on which the response and factors are observed or measured.

Designed Experiment example

Pediatric researchers at PSU carried out a designed experiment to test whether a teaspoon of honey before bed calms a child's cough and published their results in Archives of Pediatrics and Adolescent Medicine (Dec 2007).

A sample of 105 children who were ill with an upper respiratory tract infection and their parents participated in the study. On the first night, parents rated their child's cough symptoms of five scales with total scores ranging 0 to 30.

On the second night, parents were instructed to give their child a dosage of liquid "medicine" prior to bedtime. Some were given a dosage of dextromethorphan (DM), some honey, and some a placebo.

Again the parents rated their child's cough through the night and the difference in score was computed. Researchers want to compare the means of each group to determine if there are differences in cough between the treatments.

Designed Experiment example

Response Variable: Improvement in cough score

Factor(s): Medicine given

Levels: 3 (DM, Honey, Placebo)

Treatments: 3 (DM, Honey, Placebo)

Experimental Unit: Sick child

Design of Experiments

The field of Design of Experiments details experimental designs that fully disentangle the effects of multiple factors for analysis.

In a <u>full factorial</u> experiment, all factors are varied across their levels in combination. This allows measurement of all effects and <u>interactions</u>.

Example: Safety researchers are interested in braking reaction times while driving under impairments by sleeping pills and alcoholic drinks.

- Response variable: Braking reaction time
- Factors: 2 – Pills and drinks
- Levels: 2 x 3 – Pills (Yes, No), Drinks (0, 2, 4)
- Treatments: 6 (Y0, Y2, Y4, N0, N2, N4)

Design of Experiments

Full factorial designs grow exponentially with factors/levels.

DoE provides techniques for reducing the size of factorial designs by assuming high-level interaction effects are negligible.

We won't get into those details in this class, but these experiments are designed to take best advantage of the strength of the Analysis of Variance (ANOVA) technique.

One-Factor Experiments

One-Way Analysis of Variance (ANOVA)

Assumptions

- Type of data – Numerical
- Randomization – Data gathered randomly (iid)
- Population distribution – Assumes population data distributed normally, with equal variance
  - ANOVA is robust to deviations from normality
- Sample size – Must be one greater than the number of treatments at bare minimum.

Hypotheses

Null hypothesis: The mean is equal for all treatments.

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

Alternative hypothesis: At least one mean differs from the others.

$$H_a: \mu_i \neq \mu_j, \text{for at least one } i, j \in \{1,2, \dots, k\}$$

Test Statistic: Another F

The ANOVA uses the F distribution following a similar logic to how we developed the F-test for variance.

In the last F-test, $F = \frac{s_1^2}{s_2^2}$. In the ANOVA, we'll take the ratio $\frac{MST}{MSE}$, where

- MST is the Mean Square for Treatment, and
- MSE is the Mean Square for Error.

Total Sum of Squares

Let k be the number of treatments. Let n_j be the number of observations for treatment j. Then we can calculate the <u>total sum of squares</u> (TSS) to be

$$TSS = \sum_{j=1}^{k} \sum_{i=1}^{n_j} \left(y_{ij} - \bar{y}\right)^2$$

This is the numerator for the total sample variance calculation.

Total Sum of Squares

Let $\bar{y_j}$ be the mean of the observations of treatment j. We can then split SST into the sum of two parts.

$$TSS = \sum_{j=1}^{k}\sum_{i=1}^{n_j}\left(y_{ij} - \bar{y}\right)^2 = \boxed{\sum_{j=1}^{k} n_j\left(\bar{y}_j - \bar{y}\right)^2} + \boxed{\sum_{j=1}^{k}\sum_{i=1}^{n_j}\left(y_{ij} - \bar{y}_j\right)^2}$$

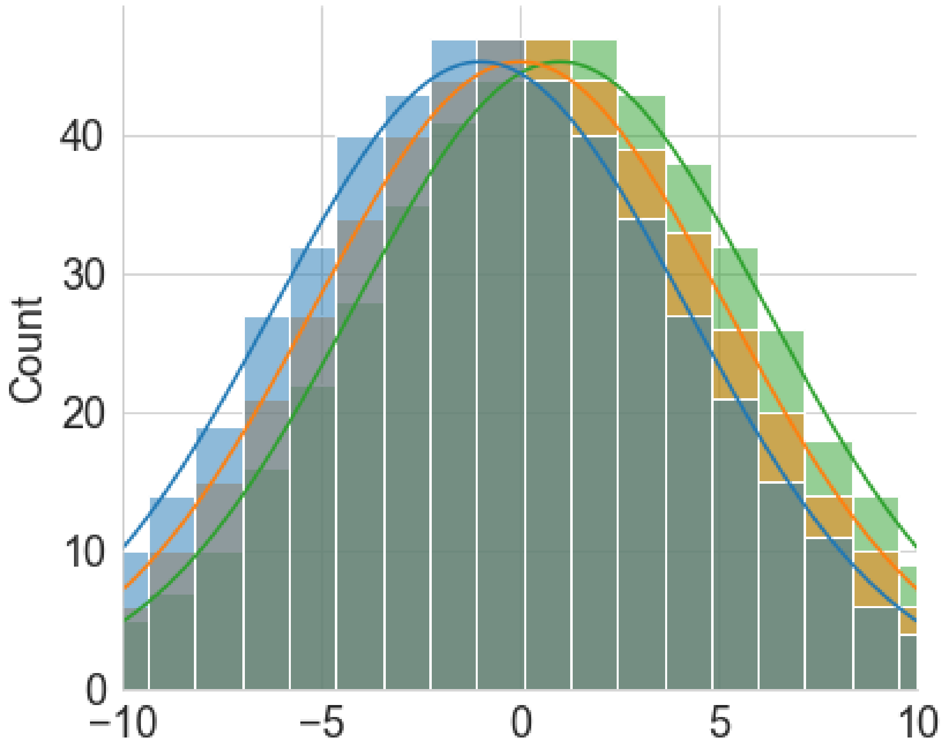
SS Treatments SS Error

If the mean variability between treatments is greater than the mean variability within treatments (i.e., error), we can say the treatments differ.

Comparing Sum of Squares

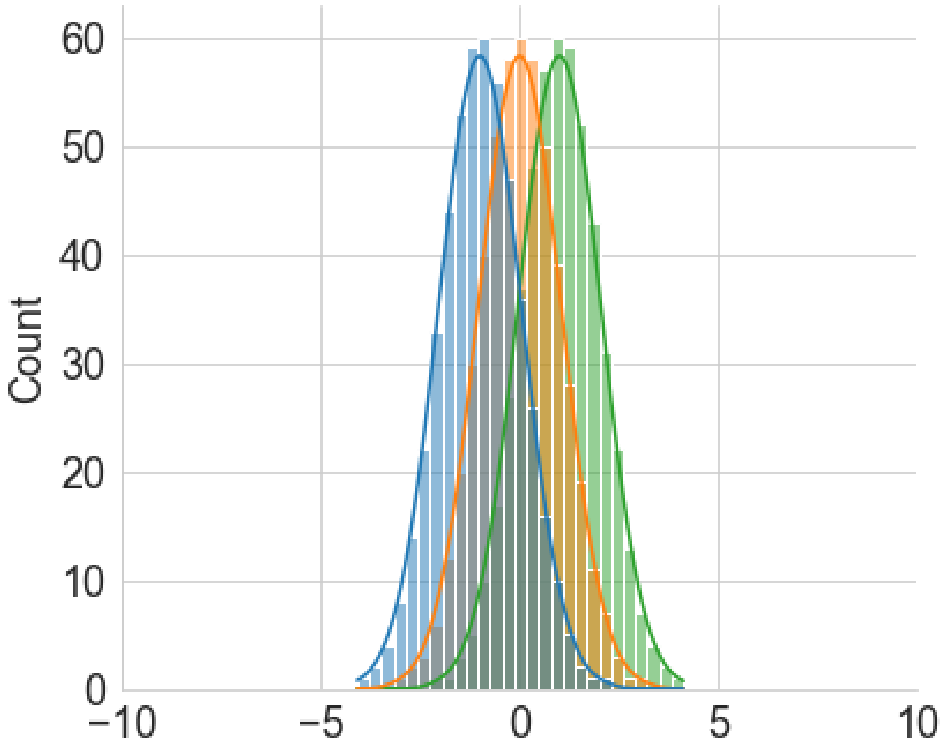
SS Treatment is Small

SS Error is Large



SS Treatment is Small

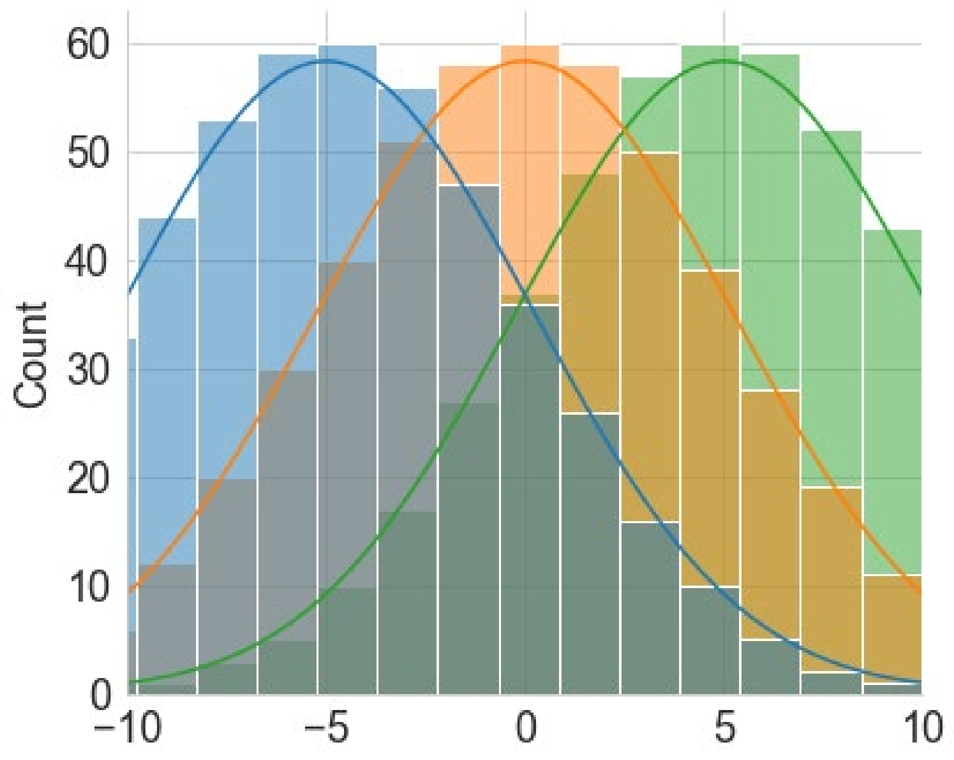
SS Error is Small



Comparing Sum of Squares

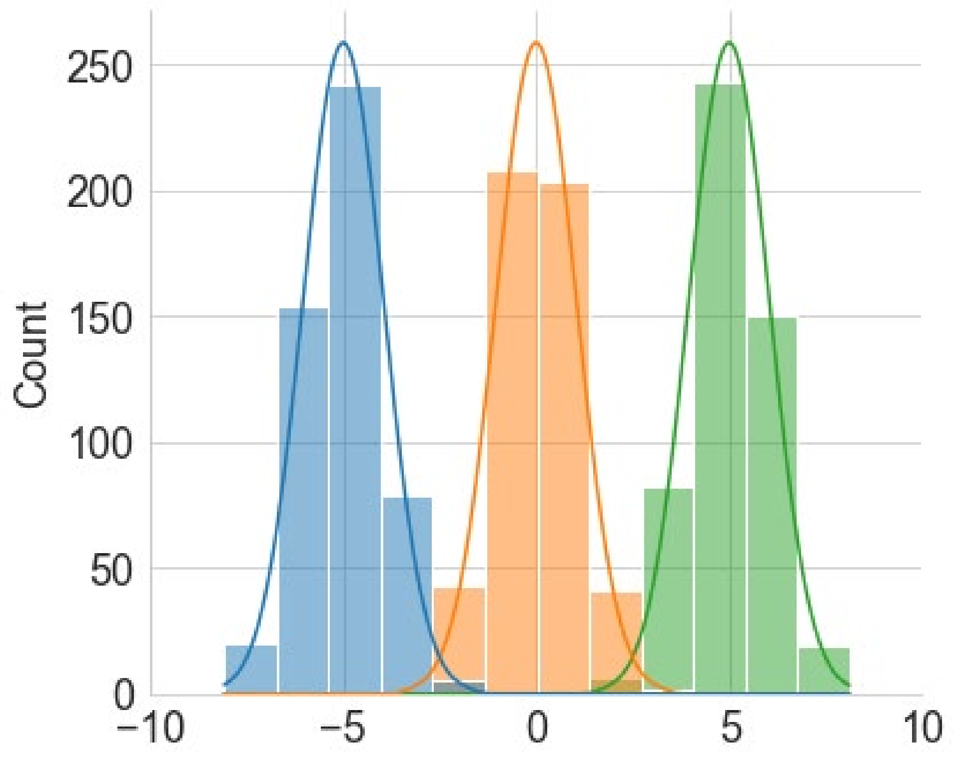
SS Treatment is Large

SS Error is Large



SS Treatment is Large

SS Error is Small



Total Sum of Squares

Blue	Orange
1	2
3	4

Consider two treatments, Blue and Orange.

$$TSS = \sum_{j=1}^{k}\sum_{i=1}^{n_j}\left(y_{ij} - \bar{y}\right)^2 = (1-2.5)^2 + (3-2.5)^2 + (2-2.5)^2 + (4-2.5)^2 = 5$$

$$SST = \sum_{j=1}^{k} n_j\left(\bar{y}_j - \bar{y}\right)^2 = 2(2-2.5)^2 + 2(3-2.5)^2 = 1$$

$$SSE = \sum_{j=1}^{k}\sum_{i=1}^{n_j}\left(y_{ij} - \bar{y}_j\right)^2 = (1-2)^2 + (3-2)^2 + (2-3)^2 + (4-3)^2 = 4$$

Mean Squares

We assume a model where $Y = \mu_j + \varepsilon$, where μ_j is the mean for the treatment j and $\varepsilon \sim N(0, \sigma^2)$ is the residual error.

Sums of Squares (Treatment, Error, and Total) are then random variables in the scale family of the χ^2 distribution (similar to s^2 in the F-test)

The Mean Squares (Treatment, Error) are the Sum of Squares divided by their degrees of freedom.

$$MST = \frac{SST}{k-1}, \qquad MSE = \frac{SST}{n-k}$$

Test Statistic

$$F = \frac{MST}{MSE} \sim F(k-1, n-k)$$

If $F > F_{ISF}(\alpha, r_1 = k-1, r_2 = n-k)$ then reject the null hypothesis.

P-value: $p = P(F(k-1, n-k) > F) = F_{SF}(F, k-1, n-k)$

One-Way ANOVA Table

Source	df	SS	MS	F
Treatment	$k-1$	$SST = \sum_{j=1}^{k} n_j(\bar{y}_j - \bar{y})^2$	$MST = \frac{SST}{k-1}$	$F = \frac{MST}{MSE}$
Error	$n-k$	$SSE = \sum_{j=1}^{k}\sum_{i=1}^{n_j}(y_{ij} - \bar{y}_j)^2$	$MSE = \frac{SSE}{n-k}$	
Total	$n-1$	$TSS = SST + SSE = \sum_{j=1}^{k}\sum_{i=1}^{n_j}(y_{ij} - \bar{y})^2$		

One-Way ANOVA Table

Blue	Orange
1	2
3	4

Source	df	SS	MS	F
Treatment	1	1	1	1/2
Error	2	4	2	
Total	3	5		

$$F^* = F_{ISF}(\alpha = 0.2, 1, 2) = 3.56$$

$$p = F_{SF}(F, k - 1, n - k) = 0.5528$$

We fail to reject the null hypothesis that the groups are equal.

One-Way ANOVA Table

Blue	Orange
1	2
3	4

Source	df	SS	MS	F
Treatment	1	1	1	1/2
Error	2	4	2	
Total	3	5		

What if we had used a t-test instead?

$$t = \frac{\bar{y}_1 - \bar{y}_2}{s_{\bar{y}_1 - \bar{y}_2}} = \frac{1}{\sqrt{2}}, \qquad \nu = 2 + 2 - 2 = 2$$

$$p = 0.5528$$

One-Way ANOVA Table

Blue	Orange
1	2
3	4

Source	df	SS	MS	F
Treatment	1	1	1	1/2
Error	2	4	2	
Total	3	5		

What if we had used a t-test instead?

Fun fact: if $T \sim t(\nu)$ then $T^2 \sim F(1, \nu)$. For two groups, they are equivalent tests.

Designed Experiment example

Pediatric researchers at PSU carried out a designed experiment to test whether a teaspoon of honey before bed calms a child’s cough and published their results in Archives of Pediatrics and Adolescent Medicine (Dec 2007).

A sample of 105 children who were ill with an upper respiratory tract infection and their parents participated in the study. On the first night, parents rated their child’s cough symptoms of five scales with total scores ranging 0 to 30.

On the second night, parents were instructed to give their child a dosage of liquid “medicine” prior to bedtime. Some were given a dosage of dextromethorphan (DM), some honey, and some a placebo.

Again the parents rated their child’s cough through the night and the difference in score was computed. Researchers want to compare the means of each group to determine if there are differences in cough between the treatments.

Designed Experiment example

- Response Variable: Improvement in cough score
- Factor(s): Medicine given
- Levels: 3 (DM, Honey, Placebo)
- Treatments: 3 (DM, Honey, Placebo)
- Experimental Unit: Sick child

- DM: [4,6,9,4,7,7,7,9,12,10,11,6,3,4,9,12,7,6,8,12,12,4,12,13,7,10,13,9,4,4,10,15,9]
- Honey: [12,11,15,11,10,13,10,4,15,16,9,14,10,6,10,8,11,12,12,8,12,9,11,15,10,1,9,13,8,12,10,8,9,5,12]
- Placebo: [5,8,6,1,0,8,12,8,7,7,1,6,7,7,12,7,9,7,9,5,11,9,5,6,8,8,6,7,10,9,4,8,7,3,1,4,3]

One-Way ANOVA Table

Source	df	SS	MS	F	P
Treatment	2	259.8	129.9	13.3	<0.001
Error	102	998.1	9.8		
Total	104	1258.0			

$$F^* = 3.08$$

$$p < 0.001$$

We reject the null hypothesis. At least one of the treatments' means differs from the others. But which one(s)?

Summary

ANOVA uses the power of the F distribution to test the null hypothesis that the mean across treatments is equal.

Check assumptions of normality and equal variance. ANOVA is robust, but severe violations may lead you to non-parametric methods such as the Kruskal-Wallis H-test

ANOVA Tables specify SST, SSE, MST, MSE, df, F, and sometimes p.

If the null hypothesis is rejected, post-hoc analysis may be needed.

If the null hypothesis is not rejected, the means may be equal, the effect may be small relative to noise, or other uncontrolled factors may be masking the effect.

Two-Factor Experiments

Two-Way Analysis of Variance (ANOVA)

Two-Factor Experiments

- Earlier we saw a one-factor (three-level) experiment.
- We could test multiple factors at once instead.
  - Factor A: a levels
  - Factor B: b levels
  - Treatments: $a \times b$ treatments
- Including all possible treatments yields a “full factorial design.”

Experiment

Earlier, we discussed factor interactions using driving reaction time as affected by drinks and sleeping pills. This data is fake, do not try at home.

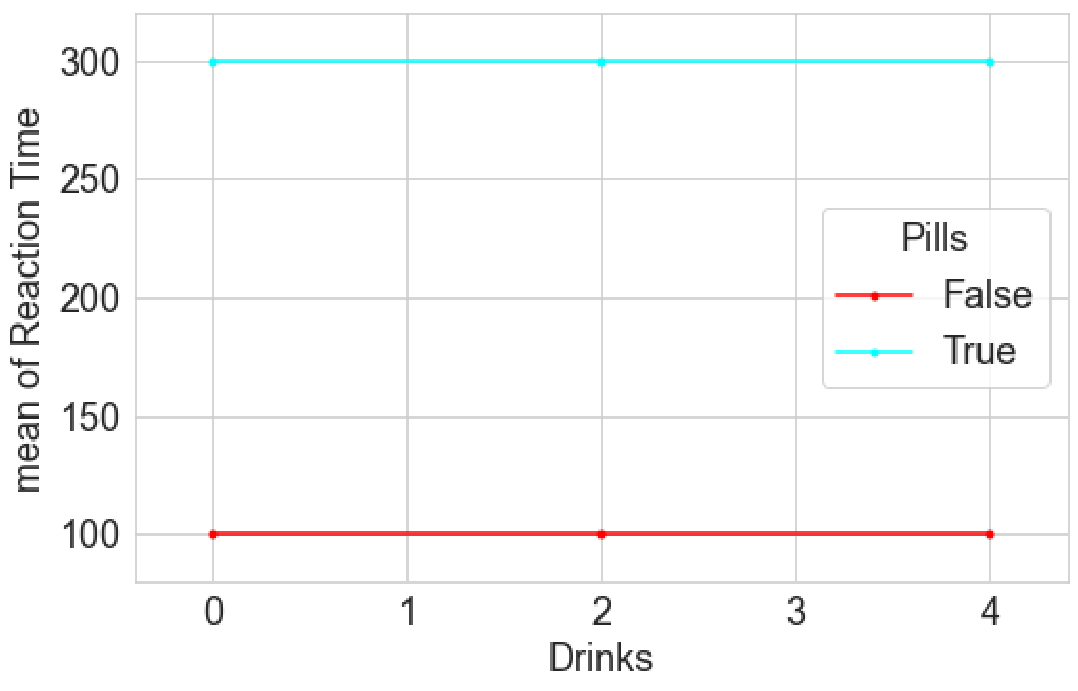
Pretend that we tested this with the following factors/levels:

- Drinks: 0, 2, 4
- Pills: False, True

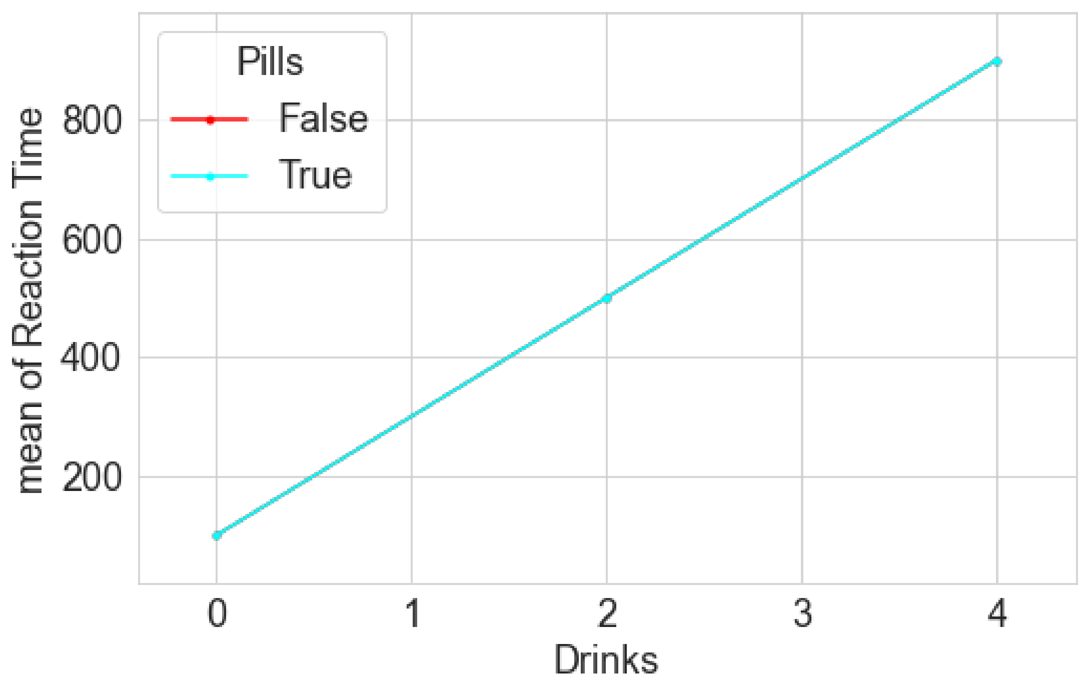
We measure response time in milliseconds.

Interaction Plots – Main Effects

Pills Increase Response Time



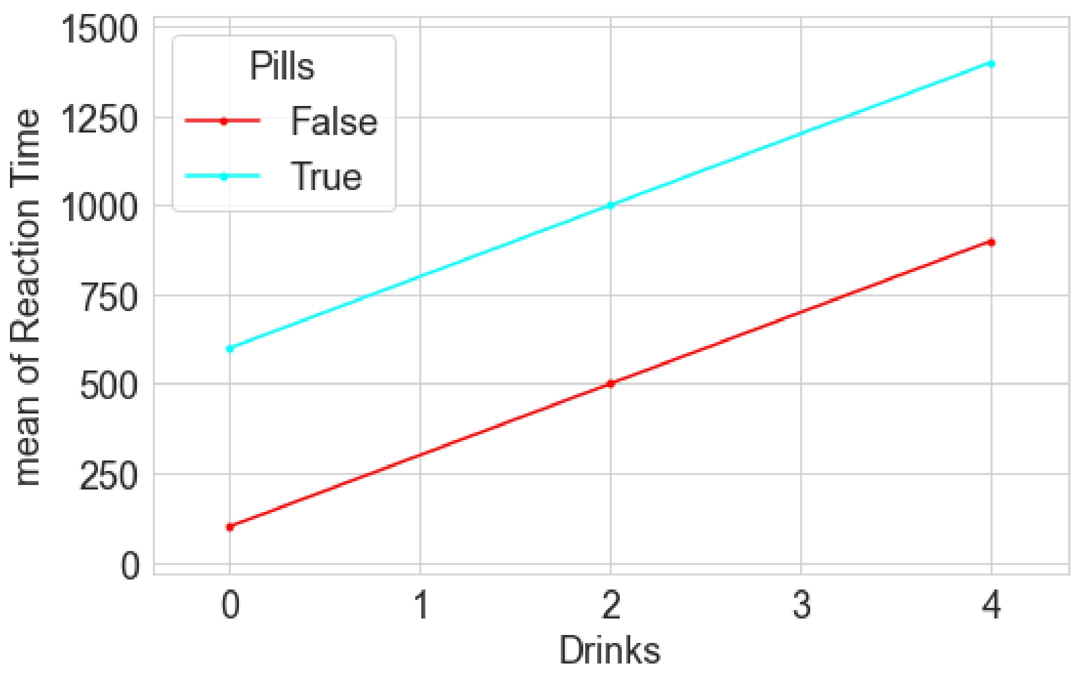
Drinks Increase Response Time



Interaction Plots – Main Effects

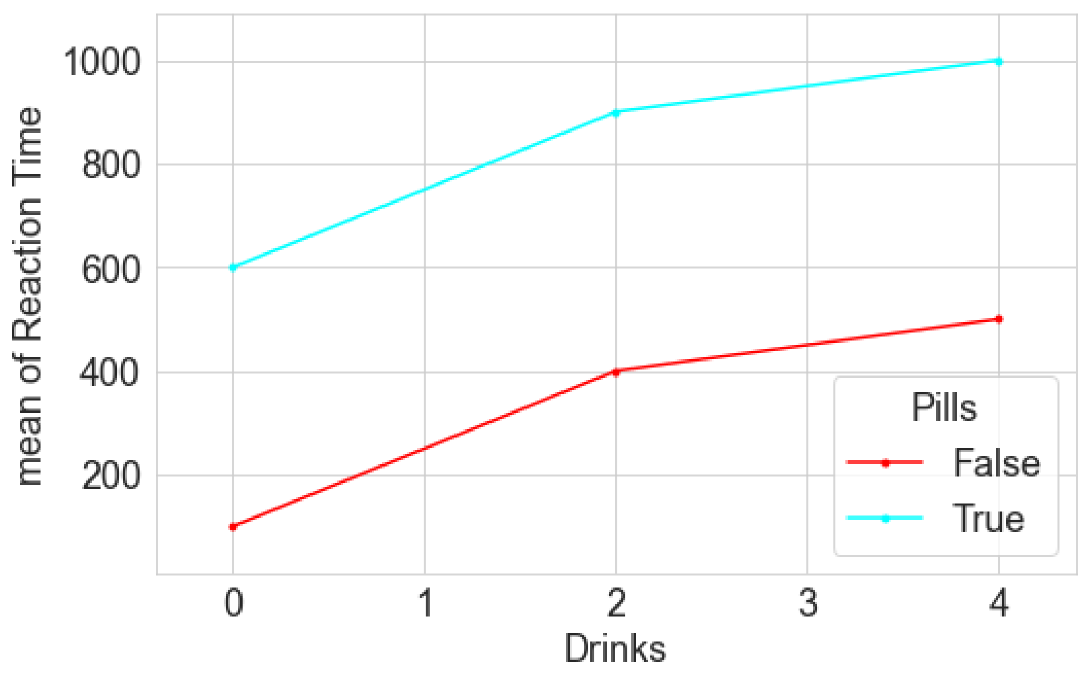
Both Increase Response Time

No Interaction



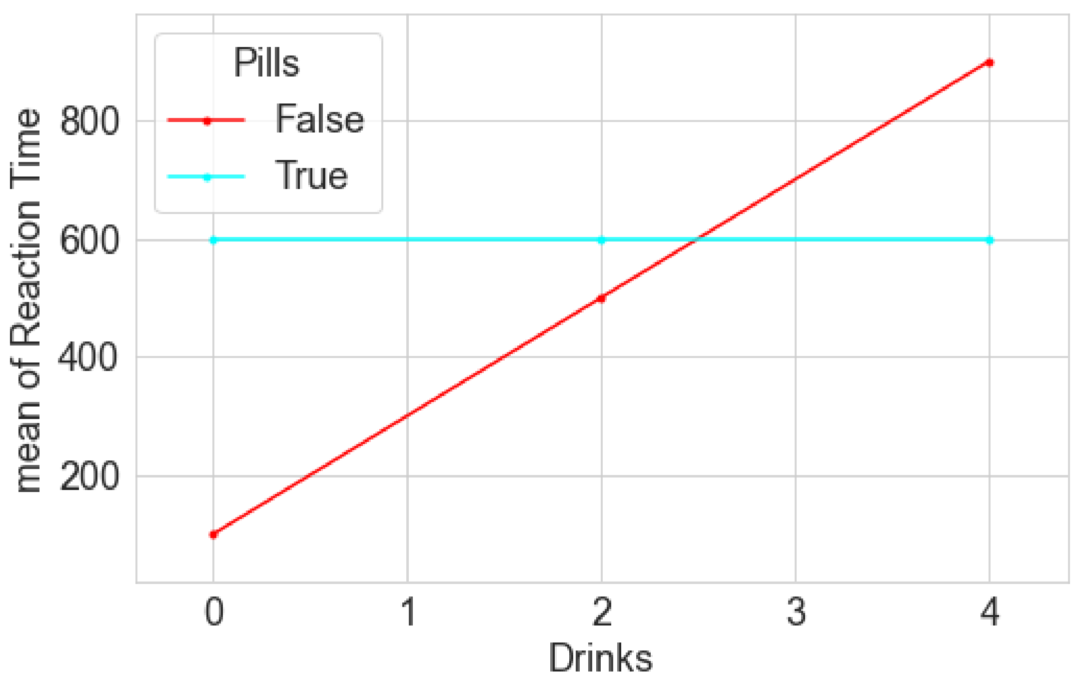
Both Increase Response Time

Drink Effects Nonlinear

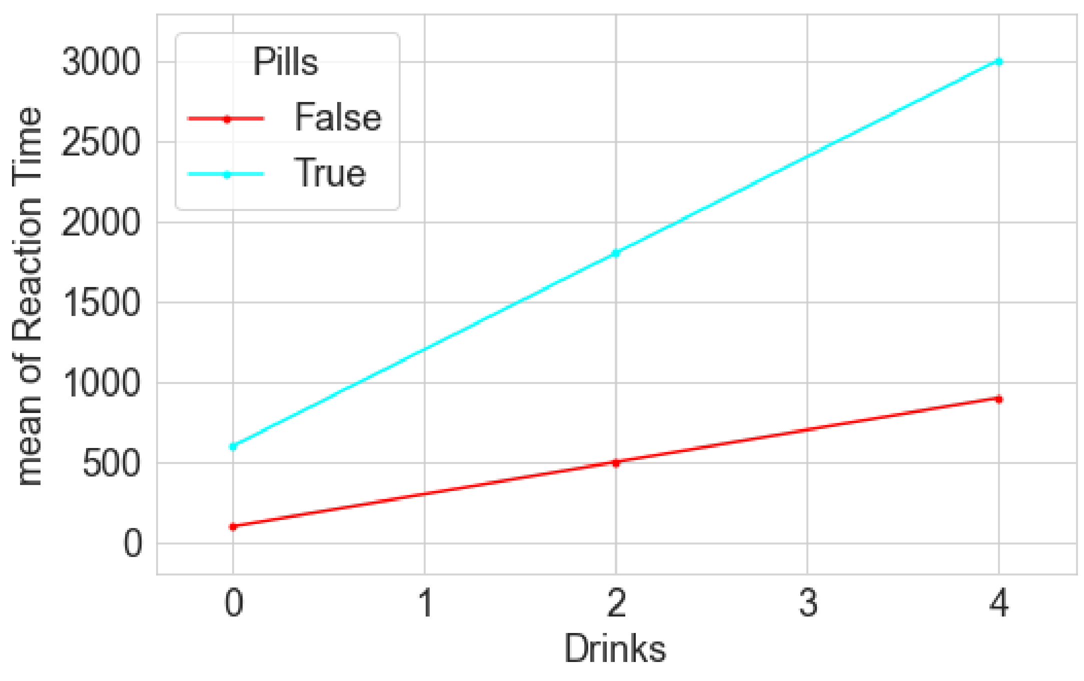


Interaction Plots – Interaction Effects

Both Increase Response Time
Negative Interaction Effect



Both Increase Response Time
Positive Interaction Effect



Total Sum of Squares

We now split TSS into more parts.

$$TSS = SST + SSE$$

$$SST = SSA + SSB + SSAB$$

Two-Way ANOVA Table

Source	df	SS	MS	F
Treatments	$ab-1$	SST=SSA+SSB+SSAB	$MST=\frac{SST}{\mathrm{ab}-1}$	$F=\frac{MST}{MSE}$
A	$a-1$	SSA	$MSA=\frac{SSA}{a-1}$	$F=\frac{MSA}{MSE}$
B	$b-1$	SSB	$MSB=\frac{SSB}{b-1}$	$F=\frac{MSB}{MSE}$
AB	$(a-1)(b-1)$	SSAB	$MSAB=\frac{SSAB}{(a-1)(b-1)}$	$F=\frac{MSAB}{MSE}$
Error	$n-ab$	SSE	$MSE=\frac{SSE}{n-ab}$	
Total	$n-1$	TSS=SST+SSE		

Example time!

More-Way ANOVA?

The same concept can be applied to larger experiments, but then you're typically doing statistical modeling rather than hypothesis tests

That's the focus of the second half of this course!

Recap

- Introduction to Design of Experiments (DoE)
- One-Factor Experiments (One-Way ANOVA)
- Two-Factor Experiments (Two-Way ANOVA)
- Many-Factor Experiments