EPSE 592: Design & Analysis of Experiments

Ed Kroc

University of British Columbia

ed.kroc@ubc.ca

January 30, 2020

Last Time

- Hypothesis tests, test statistics, and p-values
- Z-test
- t-tests (independent samples and paired)
- F-tests (testing equality of variances)

Today

- Type I and type II errors
- Multiple testing and adjustments for inflated type I errors
- P-value interpretations (orders of magnitude rule)
- One-way ANOVA (testing mean differences for more than 2 groups)

Example: three experimental groups of interest

Suppose we are interested in studying how amount of higher education correlates with self-reported anxiety levels. We have a survey designed to measure anxiety and give it to 18 people at UBC: 6 who have obtained Bachelor's degrees, 6 who have obtained Master's degrees, and 6 who have obtained PhDs (chosen how?).

Bachelor's	Master's	PhD
6.2	6.2	6.9
5.8	6.9	9.0
6.0	6.2	7.7
5.9	7.7	9.1
6.6	6.8	8.3
6.2	7.9	8.0

Table: Self-reported anxiety levels, 10 point scale. 18 respondents.

Example: three experimental groups of interest

- Could perform 3 independent-samples t-tests to test the 3 null hypotheses:
  - $H_{0,1} : \mu_B = \mu_M$

Independent Samples T-Test

		statistic	df	p
A	Student's t	−2.63	10.0	0.025

  - $H_{0,2} : \mu_M = \mu_P$

Independent Samples T-Test

		statistic	df	p
C	Student's t	2.71	10.0	0.022

  - $H_{0,3} : \mu_B = \mu_P$

Independent Samples T-Test

		statistic	df	p
E	Student's t	−5.73	10.0	<.001

Example: three experimental groups of interest

- Could perform 3 independent-samples t-tests to test the 3 null hypotheses:
  - $H_{0,1} : \mu_B = \mu_M \implies p\text{-value} < 0.05$
  - $H_{0,2} : \mu_M = \mu_P \implies p\text{-value} < 0.05$
  - $H_{0,3} : \mu_B = \mu_P \implies p\text{-value} << 0.05$
- But what about inflated Type I error?

Type I and Type II Errors

- Recall: when p-value small, conclude data inconsistent with H_0.
- Recall: when p-value large, conclude data consistent with H_0.
- Whenever we make a decision about a hypothesis based on a p-value, we have a chance of making an error.

	Given H_0 true	Given H_0 false
data inconsistent with H_0	Type I error *false positive*	Correct decision *true positive*
data consistent with H_0	Correct decision *true negative*	Type II error *false negative*

Type I and Type II Errors

- Traditionally, we set a predetermined *significance level*, α, such that

$$\Pr(\text{Type I error}) = \Pr(p - \mathit{value} < \alpha \mid H_0 \text{ true}) = \alpha.$$

- Then α, sample size, variability, and choice of test determine

$$\Pr(\text{Type II error}) = \Pr(p - \mathit{value} > \alpha \mid H_0 \text{ false}) = \beta.$$

- The *confidence level*, or *specificity*, of a test is defined as

$$\Pr(p - \mathit{value} > \alpha \mid H_0 \text{ true}) = 1 - \alpha.$$

- The *power*, or *sensitivity*, of a test is defined as

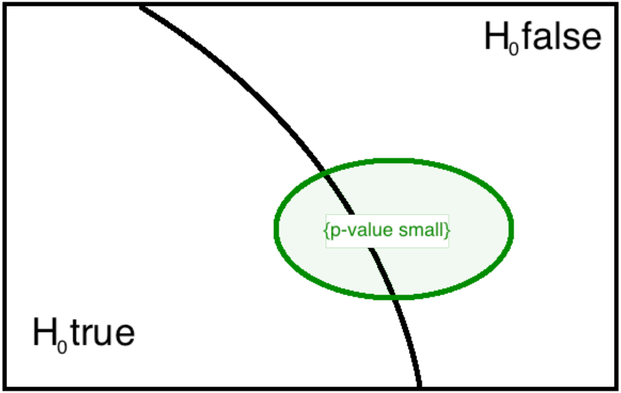
$$\Pr(p - \mathit{value} < \alpha \mid H_0 \text{ false}) = 1 - \beta.$$

Type I and Type II Errors

- In practice, $\alpha = 0.05$ is a common choice.
- Note: all of $1-\alpha$, β, and $1-\beta$ are determined once α has been fixed, the data have been collected, and the choice of analysis made.
- Good studies will strive to have $1-\beta \geqslant 0.80$. Most studies will have much lower power.

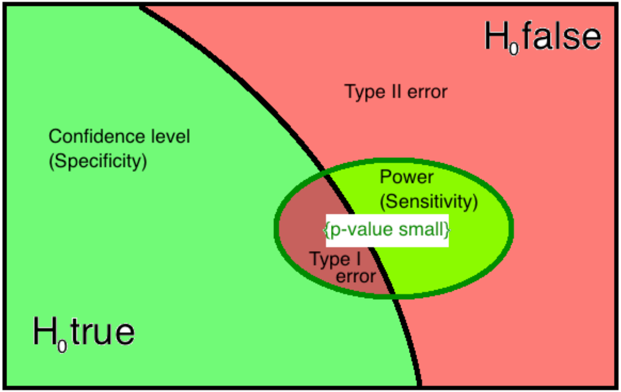
	Given H_0 true	Given H_0 false
Pr(data inconsistent with $H_0 \mid \cdots$)	α	$(1-\beta)$
Pr(data consistent with $H_0 \mid \cdots$)	$(1-\alpha)$	β

Type I and Type II Errors



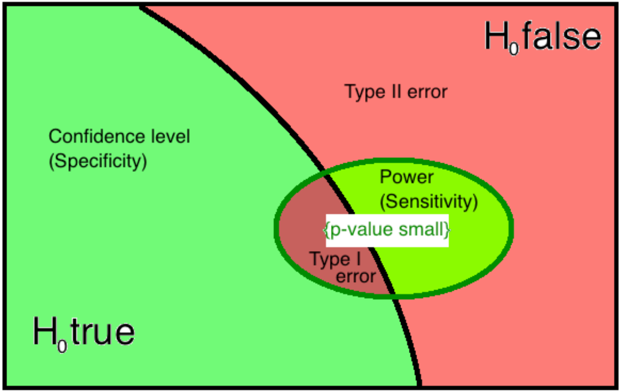
- Can split the universe of possibilities up into two disjoint pieces: H_0 true or H_0 false.
- Event of interest (when the p-value is "small") lives somewhere on the two pieces; its complement (p-value is "large") occupies the remainder of the universe.

Type I and Type II Errors



- Keeping all else the same (e.g. sample size, choice of statistical test), if we force α to be smaller, then this has to shrink the size of the event of interest, {p-value small}; thus, we *necessarily* increase β.

Type I and Type II Errors



- The only way to *simultaneously* decrease α and β (i.e. both kinds of errors) is to increase our sample size or choose a better (i.e. more powerful) statistical test.

Multiple Testing

- Each time we conduct a statistical test of hypothesis, we have a chance of committing a Type I or Type II error.
- The choice of α controls our chance of Type I error for a single test.
- Thus, if our study requires more than one test, each one has a chance of error.
- Thus, if our study requires more than one test, we should be concerned with the *family-wise* error rate: the probability of committing *at least one* Type I error.

Multiple Testing: example

- Suppose we test two hypotheses that are independent of each other:
  - $H_{0,1}$: mean iron concentration in blood equal between 2 groups
  - $H_{0,2}$: mean anxiety levels equal between same 2 groups
- Suppose we set $\alpha =$

$$\Pr(\text{test 1 significant} \mid H_{0,1} \text{ true}) = \Pr(\text{test 2 significant} \mid H_{0,2} \text{ true}).$$

- Rules of probability then tell us:
  $\Pr(\text{test 1 or 2 significant} \mid H_{0,1} \text{ and } H_{0,2} \text{ true}) =$

$$\begin{aligned} \Pr(T_1 \text{ sig.}|H_{0,1}) + \Pr(T_2 \text{ sig.}|H_{0,2}) - \Pr(T_1 \text{ and } T_2 \text{ sig.}|H_{0,1}, H_{0,2}) &= \alpha + \alpha - \alpha \cdot \alpha \\ &= 2\alpha - \alpha^2 \\ &> \alpha, \text{ since } 0 < \alpha < 1. \end{aligned}$$

- Therefore, family-wise error rate > individual error rate.

Adjustments for Multiple Tests

- Practically, this means the more hypotheses we test, the less confident we can be that our "significant" results are actually significant.
- However, there are many ways to *correct* for this inflation of Type I error due to multiple testing:
  - Bonferroni adjustment (most common, most conservative)
  - Šidák and Holm adjustments
  - Tukey adjustment
  - Scheffé adjustment
  - Benjamini-Hochberg adjustment
  - ...and many others

Adjustments for Multiple Tests

- Bonferroni adjustement says:
  - Set an original α rate of Type I error.
  - Take this α and divide by the total number of tests, n, you will perform: $\alpha' := \alpha/n$.
  - This new α' level is what you should use in each test to determine if the p-value is "significant" or not.
- The Bonferroni procedure guarantees that the chance of making *any* Type I errors in any tests is no bigger than the original α level.
- That is, Bonferroni ensures family-wise Type I error rate is no bigger than α.
- Bonferroni is *very conservative*: always works, but if tests are not independent, can be a massive overcorrection.

Adjustments for Multiple Tests: example

Recall our data on self-reported anxiety levels:

Bachelor's	Master's	PhD
6.2	6.2	6.9
5.8	6.9	9.0
6.0	6.2	7.7
5.9	7.7	9.1
6.6	6.8	8.3
6.2	7.9	8.0

- We performed three t-tests of hypotheses to compare if the pairwise means of these three groups were different.

Adjustments for Multiple Tests: example

- $H_{0,1} : \mu_B = \mu_M$

Independent Samples T-Test

		statistic	df	p
A	Student's t	−2.63	10.0	0.025

- $H_{0,2} : \mu_M = \mu_P$

Independent Samples T-Test

		statistic	df	p
C	Student's t	2.71	10.0	0.022

- $H_{0,3} : \mu_B = \mu_P$

Independent Samples T-Test

		statistic	df	p
E	Student's t	−5.73	10.0	< .001

Adjustments for Multiple Tests: example

Using the Bonferroni correction, we would find

$$\alpha' = 0.05/3 = 0.017.$$

- Comparing our p-values to the adjusted significane level yields:
  - $H_{0,1} : \mu_B = \mu_M \Longrightarrow p$-value > 0.017 (not significant)
  - $H_{0,2} : \mu_M = \mu_P \Longrightarrow p$-value > 0.017 (not significant)
  - $H_{0,3} : \mu_B = \mu_P \Longrightarrow p$-value < 0.017
- Two issues here:
  (1) Bonferroni too conservative (hypotheses not independent); means we *lose power* to detect effects.
  (2) There is no meaningful difference between a p-value of, say, 0.022 and 0.012. Yet here, the former is not "significant" while the latter is "significant".

Adjustments for Multiple Tests

- Two issues here:
  (1) Bonferroni too conservative (hypotheses not independent); means we *lose power* to detect effects.
  (2) There is no meaningful difference between a p-value of, say, 0.022 and 0.012. Yet here, the former is not "significant" while the latter is "significant".

- How to fix these issues?
  (1) Choose a better test of hypotheses: ANOVA
  (2) Discourage the enforcement of arbitrary thresholds; apply the **orders of magnitude rule**: *p-values that differ by less than one order of magnitude are practically indistinguishable as measures of evidence.*

The Analysis of Variance (ANOVA) Paradigm

The general ANOVA methodology can be described as follows:

- Rather than testing if each pair of m groups exhibit an average difference, test only the null hypothesis

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

- Then, if the data are inconsistent with H_0, we can start to test individual pairs (or contrasts) for average differences, making proper adjustments for inflated Type I errors along the way.

- ANOVA procedure is more efficient than Bonferroni and other adjustments.

- ANOVA is a direct generalization of a t-test to a comparison of more than two groups.

The Analysis of Variance (ANOVA) Paradigm

Most importantly:

- The ANOVA procedure can be generalized to account for a variety of secondary effects (confounding variables).
- ANOVA gives us a framework to study *interaction effects*; i.e. how one explanatory variable can *mediate* the effect of another explanatory variable on the response of interest.
- ANOVA procedure is flexible enough to account for a large variety of experimental designs (e.g. repeated measures, nested designs, random effects, etc.)
- We will explore all of these and more in the coming weeks.

Data types

An ANOVA model posits a linear relationship between *categorical* explanatory variables (factors) and a *continuous* response of interest.

- Nominal data: categorical, no ordering
  - E.g. sex, preferred electoral candidate
- Ordinal data: categorical, with ordering
  - E.g. rankings (Likert responses, maybe), severity of disease
- Count data: ordering with equal distances
  - E.g. age*, number of occurrences
- Continuous data: ordered continuum
  - E.g. time, space, height, weight, age*

Choice of model and analysis will depend on data type.

Note: **Always ignore Stevens's levels of measurement: nominal, ordinal, interval, ratio - these are irrelevant in practice and in theory.**

The One-way, Fixed Effects ANOVA Model

The one-way (one-factor), fixed effects ANOVA model:

$$Y = \mu + \tau_X + \varepsilon$$

- Y is the continuous response of interest
- X is the categorical variable, with observations in all categories, used to explain variation in Y
- A *fixed effects* model is one where the explanatory variable(s) X have their values fixed by the experimenter, and/or are exhausted by the experimental design.
- μ is the *grand mean*; i.e. the average of all Y values
- τ_X is the *average treatment effect* of X on Y; i.e. the average of all $Y - \mu$ values for each fixed value of X
- ε is the leftover error; i.e. the variation in Y *unexplained* by μ and τ_X.

The One-way, Fixed Effects ANOVA Model: example

The one-way ANOVA model for our anxiety (Y) vs. education (X) data:

$$Y_{anx} = \mu + \tau_{edu} + \varepsilon$$

- Levels of X were fixed by experimental design; thus, τ_{edu} is a *fixed effect* that, here, can assume three values.
- Y is a random variable, so ε is too.
- Note: $\tau_X \neq \mu_X$
  - $\mu_X =$ average of all Y values for each fixed X value
  - $\tau_X =$ average of all $Y - \mu$ values for each fixed X value
- Thus, testing the hypothesis

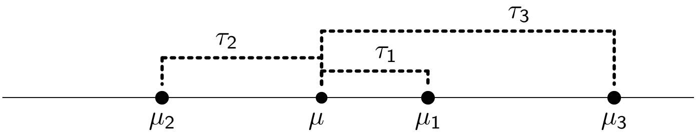
$$H_0 : \mu_B = \mu_M = \mu_P$$

is *equivalent* to testing the hypothesis

$$H_0 : \tau_B = \tau_M = \tau_P = 0$$

The One-way, Fixed Effects ANOVA Model

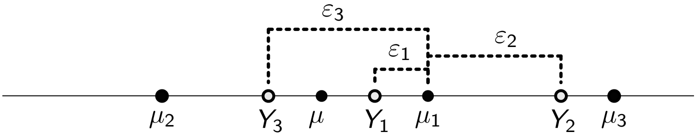
Understanding the treatment effect encoded by τ_X:



- In general, $\tau_X =$ average of all $Y - \mu$ values for each fixed X value
- Expressed another way, $\tau_X = \mu_X - \mu$
- So, if all treatments have the same effect, then they all equal the grand mean μ and $\tau_X = 0$ for all fixed values of X.

The One-way, Fixed Effects ANOVA Model

Understanding the individual error encoded by ε: suppose we have data points on Y (continuous response) and X, a categorical variable with 3 levels. Suppose observations Y_1, Y_2, and Y_3 belong to group X_1.



- In general, ε can be different for every observation/individual; it is the difference between the observed response Y and the group mean μ_X
- Explicitly, $\varepsilon = Y - \mu_X$

The One-way, Fixed Effects ANOVA Model

- The one-way (one-factor), fixed effects ANOVA model:

$$Y = \mu + \tau_X + \varepsilon$$

- Using the previous two slides, this model can be rewritten as:

$$Y - \mu = (\mu_X - \mu) + (Y - \mu_X)$$

- In practice, we do not observe μ or μ_X, but we do observe the *sample* grand mean and *sample* group means.

- Can use these sample statistics to estimate the above equation and then test the hypothesis that $H_0 : \mu_X = \mu$ for all fixed values of X.

Partitioning the ANOVA model into variance components

- We have observations on a response Y and an explanatory factor variable X with K distinct factors.
  - For example, if X is the education level from previous example, then $K = 3$.
- Total sample size $= N$.
  - For example, in the anxiety vs. education example, $N = 18$.
- Sample size within *each factor level* of X is n_j for $1 \leqslant j \leqslant K$. Therefore,

$$\sum_{j=1}^{K} n_j = N.$$

  - For example, in the anxiety vs. education example, $n_j = 6$ for all $1 \leqslant j \leqslant 3$.

Partitioning the ANOVA model into variance components

- NOTATION: Y_{ij} denotes experimental unit i within factor level j.
- NOTATION:
$$\overline{Y_{\cdot j}} = \frac{1}{n_j} \sum_{i=1}^{n_j} Y_{ij}$$
is the sample mean of the responses that all share the same factor level j.
- NOTATION:
$$\overline{Y_{\cdot\cdot}} = \frac{1}{N} \sum_{j=1}^{K} \sum_{i=1}^{n_j} Y_{ij}$$
is the sample mean of *all* responses.

Example: education levels vs. anxiety

Bachelor's ($j = 1$)	Master's ($j = 2$)	PhD ($j = 3$)
$Y_{1,1} = 6.2$	$Y_{1,2} = 6.2$	$Y_{1,3} = 6.9$
$Y_{2,1} = 5.8$	$Y_{2,2} = 6.9$	$Y_{2,3} = 9.0$
$Y_{3,1} = 6.0$	$Y_{3,2} = 6.2$	$Y_{3,3} = 7.7$
$Y_{4,1} = 5.9$	$Y_{4,2} = 7.7$	$Y_{4,3} = 9.1$
$Y_{5,1} = 6.6$	$Y_{5,2} = 6.8$	$Y_{5,3} = 8.3$
$Y_{6,1} = 6.2$	$Y_{6,2} = 7.9$	$Y_{6,3} = 8.0$
$\overline{Y_{\cdot 1}} = 6.12$	$\overline{Y_{\cdot 2}} = 6.95$	$\overline{Y_{\cdot 3}} = 8.12$

$$\overline{Y_{\cdot\cdot}} = 7.08$$

Partitioning the ANOVA model into variance components

Our goal is to partition the observed variation in our response Y into two distinct pieces:

- (1) variation explained by the different factor levels (treatments)
- (2) leftover (residual) variation

- Recall: our ANOVA model can be written as:

$$Y - \mu = (\mu_X - \mu) + (Y - \mu_X) \quad \textit{(theoretical model)}$$

- Since we do *not* observe μ_X or μ, we replace them by their sample estimates $\overline{Y}_{\cdot j}$ and $\overline{Y}_{\cdot\cdot}$
- Also, replace the generic Y by our observed Y_{ij} values:

$$Y_{ij} - \overline{Y}_{\cdot\cdot} = (\overline{Y}_{\cdot j} - \overline{Y}_{\cdot\cdot}) + (Y_{ij} - \overline{Y}_{\cdot j}) \quad \textit{(sample estimate of model)}$$

Partitioning the ANOVA model into variance components

- Now we square both sides of the equation:

$$
\begin{aligned}
(Y_{ij} - \overline{Y}_{..})^2 &= \left[(\overline{Y}_{.j} - \overline{Y}_{..}) + (Y_{ij} - \overline{Y}_{.j})\right]^2 \\
&= (\overline{Y}_{.j} - \overline{Y}_{..})^2 + (Y_{ij} - \overline{Y}_{.j})^2 + 2(\overline{Y}_{.j} - \overline{Y}_{..})(Y_{ij} - \overline{Y}_{.j})
\end{aligned}
$$

- Now sum over all observations:

$$
\begin{aligned}
\sum_{j=1}^{K}\sum_{i=1}^{n_j}(Y_{ij} - \overline{Y}_{..})^2 =& \sum_{j=1}^{K}\sum_{i=1}^{n_j}(\overline{Y}_{.j} - \overline{Y}_{..})^2 + \sum_{j=1}^{K}\sum_{i=1}^{n_j}(Y_{ij} - \overline{Y}_{.j})^2 \\
&+ 2\sum_{j=1}^{K}\sum_{i=1}^{n_j}(\overline{Y}_{.j} - \overline{Y}_{..})(Y_{ij} - \overline{Y}_{.j})
\end{aligned}
$$

- Examine the last term in the equation:

Partitioning the ANOVA model into variance components

- Examine the last term in the equation:

$$2\sum_{j=1}^{K}\sum_{i=1}^{n_j}(\overline{Y}_{\cdot j} - \overline{Y}_{\cdot\cdot})(Y_{ij} - \overline{Y}_{\cdot j}) = 2\sum_{j=1}^{K}(\overline{Y}_{\cdot j} - \overline{Y}_{\cdot\cdot})\sum_{i=1}^{n_j}(Y_{ij} - \overline{Y}_{\cdot j})$$

- Now, we can simplify the last factor on the RHS as follows:

$$\begin{aligned}
\sum_{i=1}^{n_j}(Y_{ij} - \overline{Y}_{\cdot j}) &= \sum_{i=1}^{n_j} Y_{ij} - \sum_{i=1}^{n_j} \overline{Y}_{\cdot j} \\
&= \frac{n_j}{n_j}\sum_{i=1}^{n_j} Y_{ij} - \overline{Y}_{\cdot j}\sum_{i=1}^{n_j} 1 \\
&= n_j\overline{Y}_{\cdot j} - n_j\overline{Y}_{\cdot j} \\
&= 0
\end{aligned}$$

Partitioning the ANOVA model into variance components

- Therefore, the entire cross-term disappears:

$$\sum_{j=1}^{K}\sum_{i=1}^{n_j}(Y_{ij}-\overline{Y}_{..})^2 = \sum_{j=1}^{K}\sum_{i=1}^{n_j}(\overline{Y}_{.j}-\overline{Y}_{..})^2 + \sum_{j=1}^{K}\sum_{i=1}^{n_j}(Y_{ij}-\overline{Y}_{.j})^2 + 2\sum_{j=1}^{K}\sum_{i=1}^{n_j}(\overline{Y}_{.j}-\overline{Y}_{..})(Y_{ij}-\overline{Y}_{.j})$$

$$\sum_{j=1}^{K}\sum_{i=1}^{n_j}(Y_{ij}-\overline{Y}_{..})^2 = \sum_{j=1}^{K}\sum_{i=1}^{n_j}(\overline{Y}_{.j}-\overline{Y}_{..})^2 + \sum_{j=1}^{K}\sum_{i=1}^{n_j}(Y_{ij}-\overline{Y}_{.j})^2 + 0$$

Partitioning the ANOVA model into variance components

- This final equation is the *fundamental equation of analysis of variance*.

$$\sum_{j=1}^{K}\sum_{i=1}^{n_j}(Y_{ij} - \overline{Y}_{..})^2 = \sum_{j=1}^{K}\sum_{i=1}^{n_j}(\overline{Y}_{.j} - \overline{Y}_{..})^2 + \sum_{j=1}^{K}\sum_{i=1}^{n_j}(Y_{ij} - \overline{Y}_{.j})^2$$

- This equation says that the sample variance in the response variable is equal to the sample variance in the *average response for each treatment* plus the sample variance of the responses *within each treatment*.

- This is typically written as a *sum of squares (SS)* equation:

$$SS_{total} = SS_{treatment} + SS_{error}$$

Or:

$$SS_{total} = SS_{between} + SS_{within}$$

Partitioning the ANOVA model into variance components

- This final equation is the *fundamental equation of analysis of variance*.

$$\sum_{j=1}^{K}\sum_{i=1}^{n_j}(Y_{ij}-\overline{Y_{\cdot\cdot}})^2=\sum_{j=1}^{K}\sum_{i=1}^{n_j}(\overline{Y_{\cdot j}}-\overline{Y_{\cdot\cdot}})^2+\sum_{j=1}^{K}\sum_{i=1}^{n_j}(Y_{ij}-\overline{Y_{\cdot j}})^2$$

- Notice how each term is a sum of squared differences from the grand (terms 1 and 2) or treatment (term 3) means. *This is exactly how we always measure variability,* up to a constant multiple.

- Notice: the variance in the response is partitioned into variability *explained by the average treatment effect* (term 2) plus variability *leftover* (term 3).

Examples to clarify the math: Ex. 1

- Suppose we have these sample data on Y over a categorical variable X with 3 factor levels:

$X=1$	$X=2$	$X=3$
$Y_{1,1}=1$	$Y_{1,2}=-1$	$Y_{1,3}=5$
$Y_{2,1}=1$	$Y_{2,2}=-1$	$Y_{2,3}=5$
$Y_{3,1}=1$	$Y_{3,2}=-1$	$Y_{3,3}=5$

  Then:

$$\overline{Y}_{\cdot 1}=1, \qquad \overline{Y}_{\cdot 2}=-1, \qquad \overline{Y}_{\cdot 3}=5$$

  And $\overline{Y}_{\cdot\cdot}=1.67$.

- Now plug into the fundamental equation of ANOVA:

Examples to clarify the math: Ex. 1

Fundamental equation of ANOVA:

$$\sum_{j=1}^{K}\sum_{i=1}^{n_j}(Y_{ij} - \overline{Y_{..}})^2 = \sum_{j=1}^{K}\sum_{i=1}^{n_j}(\overline{Y_{.j}} - \overline{Y_{..}})^2 + \sum_{j=1}^{K}\sum_{i=1}^{n_j}(Y_{ij} - \overline{Y_{.j}})^2$$

- Notice that the last term equals zero!

$$\begin{aligned}\sum_{j=1}^{K}\sum_{i=1}^{n_j}(Y_{ij} - \overline{Y_{.j}})^2 =& (1-1)^2 + (1-1)^2 + (1-1)^2 \\ & + (-1+1)^2 + (-1+1)^2 + (-1+1)^2 \\ & + (5-5)^2 + (5-5)^2 + (5-5)^2 = 0\end{aligned}$$

- So, as expected, all the variability in the response is explained by the different treatment/factor levels of X.

Examples to clarify the math: Ex. 2

- Now, suppose we have these sample data instead on Y over a categorical variable X with 3 factor levels:

$X = 1$	$X = 2$	$X = 3$
$Y_{1,1} = -1.1$	$Y_{1,2} = -4.2$	$Y_{1,3} = 0.5$
$Y_{2,1} = 0.5$	$Y_{2,2} = -0.1$	$Y_{2,3} = 0.6$
$Y_{3,1} = 2.4$	$Y_{3,2} = 6.1$	$Y_{3,3} = 0.7$

Then:

$$\overline{Y_{\cdot 1}} = 0.6, \qquad \overline{Y_{\cdot 2}} = 0.6, \qquad \overline{Y_{\cdot 3}} = 0.6$$

And $\overline{Y_{\cdot\cdot}} = 0.6$.

- Now plug into the fundamental equation of ANOVA:

Examples to clarify the math: Ex. 2

Fundamental equation of ANOVA:

$$\sum_{j=1}^{K}\sum_{i=1}^{n_j}(Y_{ij} - \overline{Y_{\cdot\cdot}})^2 = \sum_{j=1}^{K}\sum_{i=1}^{n_j}(\overline{Y_{\cdot j}} - \overline{Y_{\cdot\cdot}})^2 + \sum_{j=1}^{K}\sum_{i=1}^{n_j}(Y_{ij} - \overline{Y_{\cdot j}})^2$$

- Notice that the second term equals zero now!

$$\begin{aligned}\sum_{j=1}^{K}\sum_{i=1}^{n_j}(\overline{Y_{\cdot j}} - \overline{Y_{\cdot\cdot}})^2 =& (0.6-0.6)^2 + (0.6-0.6)^2 + (0.6-0.6)^2 \\ &+ (0.6-0.6)^2 + (0.6-0.6)^2 + (0.6-0.6)^2 \\ &+ (0.6-0.6)^2 + (0.6-0.6)^2 + (0.6-0.6)^2 = 0\end{aligned}$$

- So, as expected, all the variability in the response is explained by the variation *within* each treatment/factor level of X.

Mean sum of squares

- The *fundamental equation of analysis of variance*:

$$\sum_{j=1}^{K}\sum_{i=1}^{n_j}(Y_{ij} - \overline{Y_{\cdot\cdot}})^2 = \sum_{j=1}^{K}\sum_{i=1}^{n_j}(\overline{Y_{\cdot j}} - \overline{Y_{\cdot\cdot}})^2 + \sum_{j=1}^{K}\sum_{i=1}^{n_j}(Y_{ij} - \overline{Y_{\cdot j}})^2$$

- Notice how each term is a sum of squared differences from the grand (terms 1 and 2) or treatment (term 3) means. *This is exactly how we always measure variability,* up to a constant multiple.

- Recall: to define the sample variance, we had to *divide* by a constant:

$$S^2 = \frac{1}{N-1}\sum_{\ell=1}^{N}(Y_\ell - \overline{Y})^2$$

- The same applies for the ANOVA equation:

Mean sum of squares

- An unbiased estimator of the *total variance* is the *total mean square*:

$$MS_{total} = \frac{1}{N-1}\sum_{j=1}^{K}\sum_{i=1}^{n_j}(Y_{ij} - \overline{Y}_{..})^2$$

- An unbiased estimator of the *between treatment variance* is the *treatment mean square*:

$$MS_{treatment} = \frac{1}{K-1}\sum_{j=1}^{K}\sum_{i=1}^{n_j}(\overline{Y}_{\cdot j} - \overline{Y}_{..})^2$$

- An unbiased estimator of the *within treatment variance* is the *error mean square*:

$$MS_{error} = \frac{1}{N-K}\sum_{j=1}^{K}\sum_{i=1}^{n_j}(Y_{ij} - \overline{Y}_{\cdot j})^2$$

Testing the ANOVA null hypothesis

- Recall that the null hypothesis that a one-way (fixed factor) ANOVA model is designed to test is:

$$H_0 : \mu_1 = \mu_2 = \cdots = \mu_K,$$

where μ_j is the mean response over the jth category of the explanatory factor X, $1 \leqslant j \leqslant K$.

- *Under this null hypothesis*, we have that:

$$\frac{MS_{treatment}}{MS_{error}} \sim F(K-1, N-K)$$

- That is, the ratio of the between and within treatment sample variance estimators derived from the ANOVA model give an F-statistic under the null hypothesis.
- Thus, we can use this ratio as a test statistic and calculate p-values!