Experimental Design:

Much of the material we will be covering for a while has to do with designing an experimental study that concerns some phenomenon of interest.

We wish to use our subjects in the best way possible.

Our emphasis is on *random assignment* strategies, particularly as to blocking/not blocking on certain factors to make the assessment of some set of treatments more precise.

This is in contrast to *observational studies*, where random assignment is not possible.

We have touched already on many of the analysis methods we will now explore in greater depth (e.g., what comes from multiple regression and the general linear model)

The general argument for randomization is that it will tend to average out between the treatments whatever systematic effects may be present, apparent or hidden, so that comparisons between treatments measure only the pure treatment effects.

We try to eliminate extraneous factors not under the experimenter's control.

In a "completely randomized design", treatments are assigned to subjects at random; the analysis methods will be the various ANOVA models we have and will discuss.

If the subjects are heterogeneous, we may wish to *block* on various factors and randomize within blocks to treatments.

The latter is a form of "restricted randomization"

As emphasized by R. A. Fisher in the first half of the twentieth century, the physical act of randomization can be used to justify the analysis methods we have and will discuss.

For example, consider the completely randomized design:

		Treatment	
1	2	$\cdots$	r
Y_{11}	Y_{21}	$\cdots$	Y_{r1}
Y_{12}	Y_{22}	$\cdots$	Y_{r2}
$\vdots$			$\vdots$
Y_{1n_1}	Y_{2n_2}	$\cdots$	Y_{rn_r}

Under the usual null hypothesis that all the means are equal, we compare

$\frac{MSTR}{MSE}$ to F_{r-1,n_T-r},

where $n_T = n_1 + \cdots + n_r$, the total sample size.

If the treatments do not influence the scores, then the observations we see would be the same before or after the imposition of the treatments.

Consider the pool of n_T observations:

Under the (null) hypothesis of no treatment influence, what you see is a random allocation of the n_T observations to the r groups.

Is this a reasonable conjecture?

Suppose I evaluate $\frac{MSTR}{MSE}$ for my observed data, and evaluate it against all possible $(\frac{n_T!}{n_1!...n_r!})$ ways the data could be subdivided.

If my observed statistic is extreme with respect to this distribution, reject the null hypothesis

As Fisher and others have argued, F_{r-1,n_T-r} is a good approximation to this distribution (with no assumption of normality or random sampling)

Besides relying on an approximation, we could carry out "complete enumeration" or some sampling of the complete distribution.

Also, once we get into this, we can consider:

1) equivalent statistics

2) other measures or test statistics

3) use of ranks instead of the original observations (e.g., the Kruskal-Wallis analysis-of-variance by ranks)

Also, if the ranks are untied, one might be able to do complete enumeration in table form

Note that the two-independent sample instance is a special case (when $r = 2$)

Or, suppose we have a b dependent-sample problem:

		Treatments			
		1	2	$\cdots$	b
Blocks	1				
	2				
	$\vdots$				
	a				

The usual test as we will see under Model III ANOVA if to compare $\frac{MSTR}{MSAB}$ to $F_{b-1,(a-1)(b-1)}$

If we allocate b within blocks at random, there are $(b!)^a$ equally-likely ways the data could have arisen if there are no treatment effects.

If we use ranks (within blocks), we have Friedman's Test

This specializes to 2 treatments (the paired t-test situation)

Single-factor or One-way Analysis of Variance (ANOVA)

The usual form of the independent sample t-test involved only two groups.

The one-way ANOVA extends the same comparison ideas beyond two groups.

We present one-way ANOVA in a very classical way first; then we relate it to the regression approach

Example: Suppose I have three toothpastes to compare – Colgate, Aim, and Crest

[Goggle: Look Ma, no cavities]

Due to conflicting claims, you decide to test if these all have the same decay prevention power

Some distinctions:

Model I: fixed effects. These are the only three toothpastes I care about. If I did the study again, I would use these same toothpastes.

Model II: random effects. I choose at random three toothpastes from the population of pastes and would like to generalize my results to the whole population of pastes. If I did the experiment over again, I would not necessarily choose these same pastes.

We will talk about fixed effects ANOVA for now.

Suppose I have a group of people and put them on pastes for a year. I randomly assign subject to the pastes and measure the number of cavities after one year.

Terms:

Factor – the independent variable studied. In this case it is "paste"

Level of a factor – 3 levels here

Single factor (one-way) – only one factor is under study and that is of paste

Experimental (manipulated) factor – under the control of the experimenter

Classification (status) factor – not under the control of the experimenter

Qualitative factor – levels are not ordered

Quantitative factor – levels are ordered in some way

ANOVA considers the factors to be qualitative even though they may not be

Notation:

	Theoretical means	Sample size
Group 1 (Colgate)	μ_1	n_1
Group 2 (Aim)	μ_2	n_2
Group 3 (Crest)	μ_3	n_3

Thus, μ_1 is the theoretical mean of cavities for all people I could have observed under group 1, and so on.

In general, we have r groups with means $\mu_1, \ldots, \mu_r$ and sample sizes of $n_1, \ldots, n_r$; the total sample size n_T is equal to $\sum_{i=1}^{r} n_i$

We let Y_{ij} be the response of subject j in group i, so $1 \leq i \leq r$ and $1 \leq j \leq n_i$

Model: (The full-rank linear model)

$Y_{ij} = \mu_i + \epsilon_{ij}$, where $\epsilon_{ij} \sim N(0, \sigma^2)$ and independent (also, constant variance), and μ_i is fixed

Thus, $Y_{ij} \sim N(\mu_i, \sigma^2)$, and all are independent

My task (if I should decide to take it), is to develop a procedure for testing

$H_o : \mu_1 = \cdots = \mu_r$

Some Notation:

In estimating the model we will use the sample mean in group i:

$$\widehat{\mu}_i = \frac{1}{n_i}\sum_{j=1}^{n_i} Y_{ij} = \frac{1}{n_i}(Y_{i.}) = \bar{Y}_{i.}$$

Grand mean:

$$\frac{1}{n_T}\sum_{i=1}^{r}\sum_{j=1}^{n_i} Y_{ij} = \frac{1}{n_T}Y_{..} = \bar{Y}_{..}$$

Partition of the Total Sum of Squares:

Ideally all people in one group would have the same number of cavities. We would not have any variability *within* groups.

The only differences would be dependent on what particular group a person is placed into, i.e., we would have variation *between* groups

Unfortunately, it is never that easy.

Consequently, I would like to have some mechanism of comparing the variation between groups to that within groups and if between groups variation is "large", then I would feel justified in rejecting $H_o : \mu_1 = \cdots = \mu_r$

So, look at the Sum of Squares Total (SSTO):

$$\sum_{i=1}^{r} \sum_{j=1}^{n_i} (Y_{ij} - \bar{Y}_{\cdot\cdot})^2$$

and rewrite as

$$\sum_{i=1}^{r} \sum_{j=1}^{n_i} (Y_{ij} - \bar{Y}_{i\cdot} + \bar{Y}_{i\cdot} - \bar{Y}_{\cdot\cdot})^2$$

This expands:

$$\sum_{i=1}^{r}\sum_{j=1}^{n_i}(Y_{ij}-\bar{Y}_{i\cdot})^2+$$

$$\sum_{i=1}^{r}\sum_{j=1}^{n_i}(\bar{Y}_{i\cdot}-\bar{Y}_{\cdot\cdot})^2+$$

$$\sum_{i=1}^{r}\sum_{j=1}^{n_i}2(Y_{ij}-\bar{Y}_{i\cdot})(\bar{Y}_{i\cdot}-\bar{Y}_{\cdot\cdot})$$

The "middle term" (i.e., the one directly above) is zero, giving the final reduction as

$$\sum_{i=1}^{r}\sum_{j=1}^{n_i}(Y_{ij}-\bar{Y}_{i\cdot})^2+$$

$$\sum_{i=1}^{r}n_i(\bar{Y}_{i\cdot}-\bar{Y}_{\cdot\cdot})^2$$

This final reduction is the Sum of Squares Treatment (SSTR) and Sum of Squares Error (SSE)

The degrees of freedom also partitions:

$n_T - 1$ for SSTO is $= \sum_{i=1}^{r}(n_i - 1)$ for SSTR plus $r - 1$ for SSE: or

$$n_T - 1 = (n_T - r) + (r - 1)$$

We also have Mean Square for Treatment:

$$MSTR = \frac{SSTR}{(r-1)}$$

and Mean Square Error: $MSE = \frac{SSE}{(n_T - r)}$

Can show:

$E(MSE) = \sigma^2$ (Always)

$E(MSTR) = \sigma^2 + \frac{\sum_{i=1}^{r} n_i(\mu_i - \mu_\cdot)^2}{r-1}$

where $\mu_\cdot = \frac{\sum_{i=1}^{r} n_i \mu_i}{n_T}$

So, again

$\frac{MSE}{\sigma^2} \sim \chi^2_{n_T - r}/(n_T - r)$

$\frac{MSTR}{\sigma^2} \sim \chi^2_{r-1}/(r-1)$, when H_o is true that $\mu_1 = \cdots = \mu_r$

The two ratios are independent, so

$\frac{MSTR}{MSE} \sim F_{r-1, n_T - r}$ when H_o is true; otherwise, it tends to be larger

Also, $\frac{MSTR}{MSE}$ reduces to a t^2 for two groups – but we cannot do a one-tailed test as we could using just the t (probably just as well)

The ANOVA Table has the following form:

Source	df	SS	MS	F
Between (Treatments)	$r-1$	SSTR	MSTR	MSTR/MSE $\sim F_{r-1,n_T-r}$
Within (Error)	n_T-r	SSE	MSE	
Total	n_T-1	SSTO		

Computational Formulas:

$$\text{SSTO} = (\sum_{i=1}^{r}\sum_{j=1}^{n_i} Y_{ij}^2) - \frac{Y_{\cdot\cdot}^2}{n_T}$$

$$\text{SSTR} = (\sum_{i=1}^{r}\frac{Y_{i\cdot}^2}{n_i}) - \frac{Y_{\cdot\cdot}^2}{n_T}$$

$$\text{SSE} = (\sum_{i=1}^{r}\sum_{j=1}^{n_i} Y_{ij}^2) - \sum_{i=1}^{r}\frac{Y_{i\cdot}^2}{n_i}$$

The Regression approach to one-way ANOVA:

$$\boldsymbol{Y} = \begin{bmatrix} Y_{11} \\ \vdots \\ Y_{1n_1} \\ Y_{21} \\ \vdots \\ Y_{2n_2} \\ \vdots \\ Y_{r1} \\ \vdots \\ Y_{rn_r} \end{bmatrix} = \begin{bmatrix} 1 & 0 & \cdots & 0 \\ \vdots & \vdots & & \vdots \\ 1 & 0 & \cdots & 0 \\ 0 & 1 & \cdots & 0 \\ \vdots & \vdots & & \vdots \\ 0 & 1 & \cdots & 0 \\ \vdots & \vdots & & \vdots \\ 0 & 0 & \cdots & 1 \\ \vdots & \vdots & & \vdots \\ 0 & 0 & \cdots & 1 \end{bmatrix} \begin{bmatrix} \mu_1 \\ \mu_2 \\ \vdots \\ \mu_r \end{bmatrix} + \begin{bmatrix} \epsilon_{11} \\ \vdots \\ \epsilon_{1n_1} \\ \epsilon_{21} \\ \vdots \\ \epsilon_{2n_2} \\ \vdots \\ \epsilon_{r1} \\ \vdots \\ \epsilon_{rn_r} \end{bmatrix}$$

Or, $\boldsymbol{Y} = \boldsymbol{X}\boldsymbol{\beta} + \boldsymbol{\epsilon}$

The Sum of Squared Error for this Full Model is denoted by SSE(F); it has $n_T - r$ degrees-of-freedom

The Reduced Model below has Sum of Squared Error denoted by SSE(R); it has $n_T - 1$ degrees-of-freedom:

$$\boldsymbol{Y} = \begin{bmatrix} 1 \\ 1 \\ \vdots \\ 1 \end{bmatrix} [\mu] + \begin{bmatrix} \epsilon_{11} \\ \vdots \\ \epsilon_{rn_r} \end{bmatrix}$$

Thus,

$$\frac{\frac{SSE(R) - SSE(F)}{(n_T - 1) - (n_T - r)}}{\frac{SSE(F)}{(n_T - r)}} =$$

$$\frac{SSTR}{(r-1)} / MSE \sim F_{r-1, n_T - r}$$

Alternative formulation (not a full-rank formulation):

$Y_{ij} = \mu_i + \epsilon_{ij}$ is replaced by

$Y_{ij} = \mu_\cdot + \tau_i + \epsilon_{ij}$

Here, τ_i is the "effect" of being in group i;

$\mu_\cdot$ is the "beginning level" and it is unclear at this point how we might define it

In any case, $H_o : \mu_1 = \cdots = \mu_r$ is equivalent to

$H_o : \tau_1 = \cdots = \tau_r$

Everything else stays the same

The Regression form of this reformulation:

$$\boldsymbol{Y} = \begin{bmatrix} Y_{11} \\ \vdots \\ Y_{1n_1} \\ Y_{21} \\ \vdots \\ Y_{2n_2} \\ \vdots \\ Y_{r1} \\ \vdots \\ Y_{rn_r} \end{bmatrix} = \begin{bmatrix} 1 & 1 & 0 & \cdots & 0 \\ \vdots & \vdots & & \vdots & \\ 1 & 1 & 0 & \cdots & 0 \\ 1 & 0 & 1 & \cdots & 0 \\ \vdots & \vdots & & \vdots & \\ 1 & 0 & 1 & \cdots & 0 \\ \vdots & \vdots & & \vdots & \\ 1 & 0 & 0 & \cdots & 1 \\ \vdots & \vdots & & \vdots & \\ 1 & 0 & 0 & \cdots & 1 \end{bmatrix} \begin{bmatrix} \mu. \\ \tau_1 \\ \vdots \\ \tau_r \end{bmatrix} + \begin{bmatrix} \epsilon_{11} \\ \vdots \\ \epsilon_{1n_1} \\ \epsilon_{21} \\ \vdots \\ \epsilon_{2n_2} \\ \vdots \\ \epsilon_{r1} \\ \vdots \\ \epsilon_{rn_r} \end{bmatrix}$$

The design matrix $\boldsymbol{X}$ in this case is of size $n_T \times (r+1)$ and is not of full rank; the first column is the sum of the remaining columns.

Thus, we can't do the estimation unless we fix things up

A possible fix is in defining $\mu_{\cdot}$ as a weighted or unweighted mean; representing τ_r as a function of the other taus; and replacing the rows in the design matrix for the r^{th} group observations by different values

a) For an unweighted mean: $\mu_{\cdot} = \frac{1}{r}\sum_{i=1}^{r}\mu_i$ and

$\tau_i = \mu_i - \mu_{\cdot}$, which implies $\sum_{i=1}^{r}\tau_i = 0$

Thus if all the taus are equal, the common value must be zero

Also, we can represent τ_r as $-\tau_1 - \cdots - \tau_{r-1}$

Here's the linear model formulation using the unweighted mean:

$$\boldsymbol{Y} = \begin{bmatrix} Y_{11} \\ \vdots \\ Y_{1n_1} \\ Y_{21} \\ \vdots \\ Y_{2n_2} \\ \vdots \\ Y_{r1} \\ \vdots \\ Y_{rn_r} \end{bmatrix} = \begin{bmatrix} 1 & 1 & 0 & \cdots & 0 \\ \vdots & \vdots & & \vdots & \\ 1 & 1 & 0 & \cdots & 0 \\ 1 & 0 & 1 & \cdots & 0 \\ \vdots & \vdots & & \vdots & \\ 1 & 0 & 1 & \cdots & 0 \\ \vdots & \vdots & & \vdots & \\ 1 & -1 & -1 & \cdots & -1 \\ \vdots & \vdots & & \vdots & \\ 1 & -1 & -1 & \cdots & -1 \end{bmatrix} \begin{bmatrix} \mu_{\cdot} \\ \tau_1 \\ \vdots \\ \tau_{r-1} \end{bmatrix} + \boldsymbol{\epsilon}$$

b) For a weighted mean: $\mu_{\cdot} = \sum_{i=1}^{r} (\frac{n_i}{n_T}) \mu_i$ and $\tau_i = \mu_i - \mu_{\cdot}$,

$\sum_{i=1}^{r} n_i \tau_i = 0$

Again, if all the taus are equal, this common value must be zero

Here, $\tau_r = -\frac{n_1}{n_r}\tau_1 - \cdots - \frac{n_{r-1}}{n_r}\tau_{r-1}$

For a weighted mean the linear model would now be formulated as follows:

$$\boldsymbol{Y} = \begin{bmatrix} Y_{11} \\ \vdots \\ Y_{1n_1} \\ Y_{21} \\ \vdots \\ Y_{2n_2} \\ \vdots \\ Y_{r1} \\ \vdots \\ Y_{rn_r} \end{bmatrix} = \begin{bmatrix} 1 & 1 & 0 & \cdots & 0 \\ \vdots & \vdots & & \vdots & \\ 1 & 1 & 0 & \cdots & 0 \\ 1 & 0 & 1 & \cdots & 0 \\ \vdots & \vdots & & \vdots & \\ 1 & 0 & 1 & \cdots & 0 \\ \vdots & \vdots & & \vdots & \\ 1 & -\frac{n_1}{n_r} & -\frac{n_2}{n_r} & \cdots & -\frac{n_{r-1}}{n_r} \\ \vdots & \vdots & & \vdots & \\ 1 & -\frac{n_1}{n_r} & -\frac{n_2}{n_r} & \cdots & -\frac{n_{r-1}}{n_r} \end{bmatrix} \begin{bmatrix} \mu_{\cdot} \\ \tau_1 \\ \vdots \\ \tau_{r-1} \end{bmatrix}$$

Power:

If $H_o : \mu_1 = \cdots = \mu_r$ is not true, then MSTR/MSE has a noncentral F-distribution that depends on the noncentrality parameter:

$$\phi = \frac{1}{\sigma}\sqrt{\sum_{i=1}^{r} n_i(\mu_i - \mu_\cdot)^2}$$

where $\mu_\cdot$ is the weighted mean.

So, to specify power we must specify σ and $\mu_1, \ldots, \mu_r$;

There are a variety of (Pearson-Hartley) charts we then can consult

We use Feldt and Mahmoud, Power Function Charts for Specification of Sample Size in Analysis of Variance (Psychometrika, 1958, pp. 201–210)

If we have two groups and equal n's, the non-centrality parameter reduces to

$$(\frac{\sqrt{n}}{2})(\frac{1}{\sigma})|\mu_1 - \mu_2|,$$

where

$(\frac{1}{\sigma})|\mu_1 - \mu_2|$ is Cohen's effect size

For small, medium, and large effects, Cohen suggests values of .2, .5, and .8, respectively

So, if you ever need to "appeal to authority" ...