One-way Analysis of Variance

Hao Zhang

- One-way ANOVA
  - Model
  - Hypotheses
  - ANOVA
- Power and Sample Sizes

One-way ANOVA: The Big Picture

Goal: to test whether all treatments means are equal.
↳ single factor (eg. driver, diet, drug) with v treatment groups.

Model

$$Y_{it} = \mu_i + \epsilon_{it} \text{ or } Y_{it} = \mu + \tau_i + \epsilon_{it}, \text{ for } i = 1, \cdots, v, t = 1, \cdots, r_i.$$

where μ_i is the treatment mean for treatment i, and ϵ_{it} are i.i.d. $N(0, \sigma^2)$.

Y_{it}: response from treatment i, unit t

μ_i: mean of treatment i

$\tau_i = \mu_i - \mu$: effect of treatment i

$\varepsilon_{it} \sim N(0, \sigma^2)$: random error

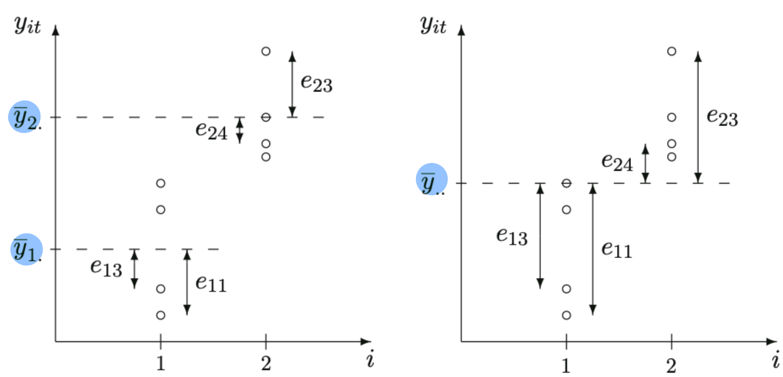
Hypotheses

H_0 : all treatment means are equal versus H_1 : not all treatment means are equal. Equivalently:

$$H_0 : \mu_1 = \mu_2 = \cdots = \mu_v \text{ or } H_1 : \mu_1, \mu_2, \cdots, \mu_v \text{ are not all equal.}$$

If H_0 is true, we have a reduced model

$$Y_{it} = \mu + \epsilon_{it}.$$



Residuals; full model

Residuals; reduced model

- Reduced Model: The sum of squares for error for this reduced model is

$$SSE_0 = \sum_i \sum_t (Y_{it} - \bar{Y}_{..})^2.$$

Then $SSE_0/(n-1)$ is an unbiased estimator of σ^2 under the null hypothesis.

- Under the full model (i.e., without assuming the treatment means are all equal), the sum of squares error is

(Error Sum of Squares) $$SSE = \sum_i \sum_t (Y_{it} - \bar{Y}_i)^2 = \sum_i (r_i - 1) S_i^2$$

$SSE/(n-v)$ is an unbiased estimator of σ^2. In practice, we use this estimator because it is still an unbiased even when the treatment means are not equal.

ANOVA

Consider the difference of sums of squares, which is attributed to the treatments and called the sum squares for treatment (SST)

(Total Sum of Squares) $$SST = SSE_0 - SSE = \sum_i \sum_t (Y_{it} - \bar{Y}_{..})^2 - \sum_i \sum_t (Y_{it} - \bar{Y}_{i.})^2.$$

Simplified into

(Treatment Sum of Squares) $$SST = \sum_i r_i (\bar{Y}_i - \bar{Y}_{..})^2 = \sum_i r_i (\bar{Y}_{i.})^2 - n(\bar{Y}_{..})^2.$$

Note the following

$\boxed{SSTotal = SST + SSE}$

- SST is independent of SSE because the sample means $\bar{Y}_{i\cdot}$ and $\bar{Y}_{\cdot\cdot}$ are independent of SSE.
- $E(SST) = E(SSE_0) - E(SSE) = (n-1)\sigma^2 - (n-v)\sigma^2 = (v-1)\sigma^2$ if H_0 is true.
- $MST = SST/(v-1)$ is an unbiased estimator of σ^2 under H_0.

Under H_0, $F \sim F_{v-1,n-v}$

- The ratio $F = \frac{MST}{MSE}$ has an F-distribution with $v-1$ and $n-v$ degrees of freedom if H_0 is true. Reject H_0 if $F > F_{v-1,n-v,\alpha}$ where α is the significance level of the test and also the probability of Type I error. A Type I error is to reject H_0 when it is true. The upper critical value $F_{v-1,n-v,\alpha}$ is provided in Table A. 6 and can be calculated using the SAS function `finv(1-\alpha, df1, df2)`.
- Intuitively, both MST and MSE are unbiased estimators of σ^2 if the treatment means are all equal. Hence the ratio should be excessively large if H_0 is true.
- SST represents the between-treatment variation and SSE the within-treatment variation.

Source of variation	Degrees of freedom	Sum of squares	Mean square	Ratio	Expected mean square
Treatments	v-1	ssT	$MST=\frac{ssT}{v-1}$	$\frac{msT}{msE}$	$\sigma^2 + Q(\tau_i)$
Error	n-v	ssE	$MSE=\frac{ssE}{n-v}$		σ^2
Total	n-1	$sstot$			

Computational formulae

$ssT = \sum_i r_i \bar{y}_{i.}^2 - n\bar{y}_{..}^2$ $\qquad$ $ssE = \sum_i \sum_t y_{it}^2 - \sum_i r_i \bar{y}_{i.}^2$

$sstot = \sum_i \sum_t y_{it}^2 - n\bar{y}_{..}^2$

$Q(\tau_i) = \sum_i r_i(\tau_i - \sum_h r_h\tau_h/n)^2/(v-1)$

The purpose of ANOVA is to compare the means.

Example. In an experiment to study the effects of drivers on the mpg of Toyota Prius, 12 new Prius were randomly assigned to three drivers so that each driver drove four cars and obtained the mpgs. This is a completely randomized design. The data are given below.

$d1$	$d2$	$d3$
50.33	48.11	49.08
46.83	50.14	48.89
51.57	43.22	49.96
45.33	47.26	49.70

The SAS program to get the sample means and sample standard deviations:

```
data prius;
input driver mpg;
lines;
1 50.33
1 46.83
1 51.57
1 45.33
2 48.11
2 50.14
2 43.22
2 47.26
3 49.08
3 48.89
3 49.96
3 49.70
;
run;
proc print data=prius;
run;

proc glm data=prius;
class driver;
 model mpg=driver;
run;
```

Partial output:

The SAS System

The GLM Procedure

Dependent Variable: mpg

Source	DF	Sum of Squares	Mean Square	F Value	Pr > F
Model	2	10.03031667	5.01515833	0.87	0.4502
Error	9	51.69105000	5.74345000		
Corrected Total	11	61.72136667			

R-Square	Coeff Var	Root MSE	mpg Mean
0.162510	4.954791	2.396550	48.36833

Source	DF	Type I SS	Mean Square	F Value	Pr > F
driver	2	10.03031667	5.01515833	0.87	0.4502

Source	DF	Type III SS	Mean Square	F Value	Pr > F
driver	2	10.03031667	5.01515833	0.87	0.4502

The p-value is 0.4502. Hence we fail to reject H_0 and there is no significant evidence to support that the three drivers have different MPG.

Try the following SAS program with optional argument

```
proc glm data=prius;
class driver;
 model mpg=driver/solution;
 lsmeans driver;
 means driver;
 run;
```

Power and Sample Sizes

Goal: Find the # of replicates r needed to detect a real difference among treatments with desired power (eg. 95%, 90%)...

Suppose $\mu_1, \cdots, \mu_\nu$ are a given set of treatment means that not all equal (i.e., H_0 is not true). The experimenter would like be able to claim a significant difference with certain probability (this probability is called the power). What sample size is sufficient for that?

In this case, the test statistic F has a noncentral F-distribution with a noncentrality parameter δ^2. To determine the noncentrality parameter, note

$$SST = r \sum_{i=1}^{\nu} \left(\bar{Y}_{i.} - \bar{Y}_{..}\right)^2.$$

The noncentrality parameter

$$\delta^2 = E(SST) = r \sum_{i=1}^{\nu} \left(E\left(\bar{Y}_{i.}\right) - E\left(\bar{Y}_{..}\right)\right)^2 / \sigma^2$$
$$= r \sum_{i=1}^{\nu} (\mu_i - \bar{\mu}.)^2 / \sigma^2$$

where $\bar{\mu}.$ is the average of $\mu_i, i = 1, \cdots, \nu$. The power is $P\left(F > F_{\nu-1, \nu(r-1), \alpha}\right)$ which can be calculated for any given r where F with $(\nu - 1)$ and $(n - \nu)$ degrees of freedom has the F-distribution with the noncentrality parameter δ^2.. Then one can choose the r that yields satisfactory power, say, 0.90.

SAS Code - An Example

> In a study to compare the effects of 4 diets on the weights of mice of 20 days old, the experimenter wishes to detect a difference of 10 grams. The experimenter estimates that the standard deviation σ is no larger than 5 grams. How many replicates is necessary to have a probability of 90% to detect a difference of 10 grams using the F-test for $\alpha = 0.05$?

A difference of 10 grams means that the maximum and the minimum treatment means differ by 10 grams. The scenario that the noncentrality parameter is the largest is when the other two treatment means are all at the middle. Hence the four treatment means are $a, a+5, a+5, a+10$ for some constant a. Then

Δ^2: Variance among treatment means.
Larger ↑r: more power to detect true differences.

$$\delta^2 = r \sum_{i=1}^{\nu} (\mu_i - \bar{\mu}.)^2 / \sigma^2 = \frac{r\Delta^2}{2\sigma^2}$$

SAS code:

```
data power;
input r @@;
v=4;
diff=10;
sigma=5;
alpha=0.05;
df1=v-1;
df2=v*(r-1);
ncp=r*diff**2/(2*sigma**2);
Falpha=finv(1-alpha, df1, df2);
power=1-probf(Falpha, df1, df2, ncp);
lines;
3 4 5 6 7 8 9
;
run;
proc print;
var r power;
run;
```

You will get the output as

Obs	r	power
1	3	0.33906
2	4	0.50370
3	5	0.64423
4	6	0.75459
5	7	0.83613
6	8	0.89360
7	9	0.93258

Note SAS has a procedure proc power that can calculate the necessary sample size for a desired power for a one-way ANOVA model.

```
proc power directly gives required r (sample size per group)
onewayanova test=overall
  alpha=.05
  groupmeans=(0 5 5 10) /* or you can use (5 10 10 15) and get the same result*/
  stddev= 5 /* standard deviation sigma */
  power=.90  /*power for the overall F test*/
  npergroup=.; /*sample size per treatment level*/
run;
```

However, proc power only applies to one-way ANOVA and is not recommended to use for two-way or multiple-way ANOVA.

Summary

Concept	Description
ANOVA model	One continuous outcome, one categorical factor
Goal	Test if treatment means differ
Key idea	Partition total variation into between-group (SST) and within-group (SSE)
F-test	Compares MST and MSE
Power analysis	Helps plan how large your experiment should be