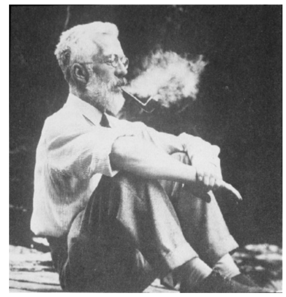
The Analysis of Variance

Ronald Fisher



- Developed by R. A. Fisher in the 1920's
- Fisher was a mathematician then employed by Rothamsted to analyze data from field experiments
- Developed ANOVA as a way to separate treatment and plot effects
- Published a series of reports under the title *Studies in Crop Variation*
- Published the book *Statistical Methods for Research Workers* in 1925
- First citation of his book in an *Agronomy Journal* article was in 1926 – by "Student"

One-factor ANOVA

Design

		Reps			
Treatments	**1**	**2**	**.**	**.**	***J***
1					
2					
.					
.					
.					
I	**.**	**.**	**.**	**.**	***IJ* = *N***

One-factor ANOVA
Example

<u>Corn Yield Trial</u>
- 12 Hybrids
- 4 Reps

One-factor ANOVA
Conceptual Layout

Hybrid i

	1	2	3	4	5	6	7	8	9	10	11	12
Plot (i)j	1	5	9	13	17	21	25	29	33	37	41	45
	2	6	10	14	18	22	26	30	34	38	42	46
	3	7	11	15	19	23	27	31	35	39	43	47
	4	8	12	16	20	24	28	32	36	40	44	48

- Shows relationships among factors in an experiment
- In this one there are 12 treatments (Hybrids)
- Each one is grown (replicated) in four independent plots
- Note that there are a total (4x12) of 48 plots
- Plots are considered to be nested because each one is an individual and can only be assigned to one of the Hybrids

One-factor ANOVA

Example – Field Map

Plot	1	2	3	4	5	6	7	8	9	10	11	12
100	6	12	5	3	5	2	11	6	5	2	9	8
200	6	12	10	10	1	7	1	12	8	3	10	4
300	8	4	7	2	3	6	8	12	7	9	11	4
400	1	11	5	1	9	9	2	3	7	11	4	10

Variance of a Population

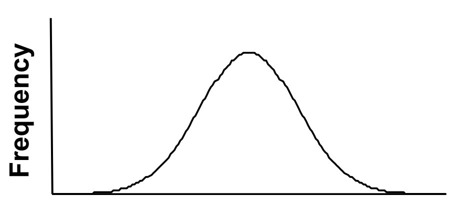
$$\sigma^2 = \frac{\sum_{i=1}^{n}(Y_i - \mu)^2}{n}$$

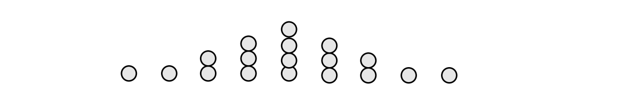
Variance of a Sample Population

$$s^2 = \frac{\sum_{i=1}^{n}(Y_i - \bar{Y})^2}{n-1}$$

The Normal Distribution

$$f = \frac{N}{\sigma\sqrt{2\pi}} e^{-(X-\mu)^2/2\sigma^2}$$



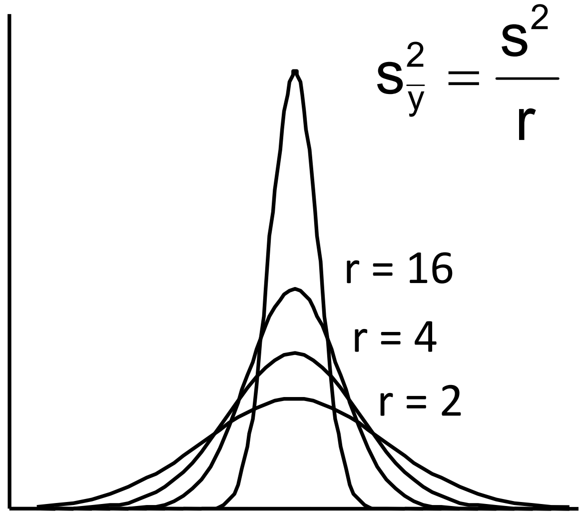


Variance of a Mean

$$s_{\bar{y}}^2 = \frac{s^2}{r}$$

$$s^2 = r s_{\bar{y}}^2$$

Distribution of Sample Means



$$s_{\bar{y}}^2 = \frac{s^2}{r}$$

Switchgrass

Variety Trial (Yield, Mg/ha)

	Cultivar	
Replication	A	B
1	10.8	8.9
2	12.7	8.9
3	11.4	8.6
4	13.6	8.8
5	13.9	8.5
6	13.0	6.1
Mean	12.6	8.3
Variance	1.46	1.24

Switchgrass

Variety Trial (Yield, Mg/ha)

Between Treatment Variance

$$s_{\bar{y}}^2 = \frac{\sum_{i=1}^{n}(\bar{Y}_i - \bar{Y})^2}{n-1} = \frac{(12.6-10.4)^2 + (8.3-10.4)^2}{2-1} = 9.12$$

$$s^2 = rs_{\bar{y}}^2 = 6(9.12) = 54.74$$

Switchgrass

Variety Trial (Yield, Mg/ha)

Within Treatment Variance

$$s_p^2 = \frac{(n_1 - 1)s_1^2 + (n_2 - 1)s_2^2}{(n_1 + n_2) - 2} = \frac{5(1.46) + 5(1.24)}{(6 + 6) - 2} = 1.35$$

$$s_p^2 = \frac{s_1^2 + s_2^2}{2} = \frac{1.46 + 1.24}{2} = 1.35$$

Switchgrass

Variety Trial (Yield, Mg/ha)

$$F = \frac{s_b^2}{s_w^2} = \frac{54.74}{1.35} = 40.58$$

The probability that $F_{1,10\ df}$ > 40.58 is 0.000081

Excel Formula: =FDIST(40.58,1,10) =0.000081

Linear Additive Model

$$Y_{ij} = \mu + T_i + \varepsilon_{(i)j}$$

where:

Y_{ij} = Response observed for the ij experimental unit

μ = Overall mean

T_i = Effect of the i^{th} treatment

$\varepsilon_{(i)j}$ = Effect associated with the ij^{th} eu or plot

ANOVA

Assumptions

$$Y_{ij} = \mu + T_i + \varepsilon_{(i)j}$$

1) **Effects** are additive

Errors are:

2) Normally distributed
3) Homogenous
4) Independent

One-Factor ANOVA

Corn Yield Test Example

$$Y_{ij} = \mu + T_i + \varepsilon_{(i)j}$$

Corn yield in plot 101	=	Average of all plots	+	Effect of hybrid 6	+	Effect of plot 101
163.04		154.39		3.08		5.57

Means Model

$$Y_{ij} = \mu_{Ti} + \varepsilon_{(i)j}$$

Y_{ij}	μ_{Ti}	$\varepsilon_{(i)j}$
163.04	157.47	5.57

Deviations Model

$$Y_{ij} - \mu = T_i + \varepsilon_{(i)j}$$

$Y_{ij} - \mu$	T_i	$\varepsilon_{(i)j}$
8.65	3.08	5.57

One-Factor ANOVA

Expected Mean Squares

Term	EMS
T_i	$\sigma^2_e + j\phi T$
$\varepsilon_{(i)j}$	σ^2_e

Row / Column Notation

	Replication (j)					
Treatment (i)	1	2	3	r	Total	Mean
1	Y_{11}	Y_{12}	Y_{13}	Y_{1r}	$T_{1.}$	$\bar{Y}_{1.}$
2	Y_{21}	Y_{22}	Y_{23}	Y_{2r}	$T_{2.}$	$\bar{Y}_{2.}$
3	Y_{31}	Y_{32}	Y_{33}	Y_{3r}	$T_{3.}$	$\bar{Y}_{3.}$
t	Y_{t1}	Y_{t2}	Y_{t3}	Y_{tr}	$T_{t.}$	$\bar{Y}_{t.}$
					$T_{..}$	$\bar{Y}_{..}$

One-Factor ANOVA

Calculations

Source	df	SS	MS	F
T_i	I - 1	$J\sum_{i=1}^{I}(\bar{Y}_{i.}-\bar{Y}_{..})^2$	SS(T)/(I - 1)	MS(T)/MS(E)
$\varepsilon_{(i)j}$	I(J - 1)	$\sum_{i=1}^{I}\sum_{j=1}^{J}(Y_{ij}-\bar{Y}_{i.})^2$	SS(E)/[I(J - 1)]	
Total	IJ - 1	$\sum_{i=1}^{I}\sum_{j=1}^{J}(Y_{ij}-\bar{Y}_{..})^2$		

One-Factor ANOVA

Switchgrass Example

Treatments

- Variety 10
- Reps 6

$$SS_{TOT} = \sum_{i=1}^{10}\sum_{j=1}^{6}(Y_{ij}-\bar{Y}_{..})^2 = 788.10$$

$$SS_{TRT}\, 6\sum_{i=1}^{10}(\bar{Y}_{i.}-\bar{Y}_{..})^2 = 6(59.25) = 355.53$$

$$SS_{ERROR} = SS_{TOT} - SS_{TRT} = 788.10 - 333.53 = 432.57$$

One-Factor ANOVA
Switchgrass Example

Source	DF	Sum of Squares	Mean Square	F Value	Pr > F
Cultivar	9	355.5271	39.50301	4.57	0.0002
Error	50	432.5709	8.651419		
Total	59	788.098			

The probability of the F ratio (4.57) occurring by chance is quite low – only 2 out of 10,000. ∴ we conclude that there was a significant cultivar effect.

One-Factor ANOVA
Hypothesis

F-test in ANOVA tests the null hypothesis:

$$\mu_1 = \mu_2 = \mu_3 = \mu_4 = \mu_5 = \mu_6 = \mu_7 = \mu_8 = \mu_9 = \mu_{10}$$

Based on the ANOVA we conclude that the null hypothesis is false. That is, at least one pair of means among the ten are statistically different from each other.

One-Factor ANOVA

Switchgrass Example

Cultivar	Mean Yield (Mg/ha)
A	12.6
B	8.3
C	14.2
D	12.7
E	9.4
F	6.7
G	13.9
H	9.6
I	9.4
J	9.1

We know that there is at least one significant difference among pairs of means. How do we find out which pairs actually differ?