Analysis of Variance

Vahid Partovi Nia

Chapitre 16

Links 2

History **3**

Reference Book 5

Preparation **7**

What is the question? 8

T-Test 13

T-Test Output 14

T-Test and ANOVA 15

What is SS? Variation! **16**

What is df ? Chi-square! **20**

What is MS? A Division **23**

What is Fisher's F? Another Division **25**

Links

- http://probstat.ca/slide.pdf
- http://probstat.ca/note.pdf

2 / 28

History

3 / 28

Fisher

Ronald Fisher: British biologist and statistician.
He also invented Fisher's distribution, maximum likelihood, linear discriminant, and many other data analysis techniques. He is the father of modern statistics (along with Karl Pearson, Egon Pearson, and Jersey Neyman).

4 / 28

Reference Book

Table 16.3

Northeast	Midwest	South	West
15	17	11	10
10	12	7	12
13	18	9	8
14	13	13	7
13	15		9
	12		
13.0	14.5	10.0	9.2

Table 16.4

Source	df	SS	MS = SS/df	F-statistic
Treatment	$k-1$	$SSTR$	$MSTR = \frac{SSTR}{k-1}$	$F = \frac{MSTR}{MSE}$
Error	$n-k$	SSE	$MSE = \frac{SSE}{n-k}$	
Total	$n-1$	SST		

$\Rightarrow$

Table 16.5

Source	df	SS	MS = SS/df	F-statistic
Treatment	3	97.5	32.500	6.32
Error	16	82.3	5.144	
Total	19	179.8		

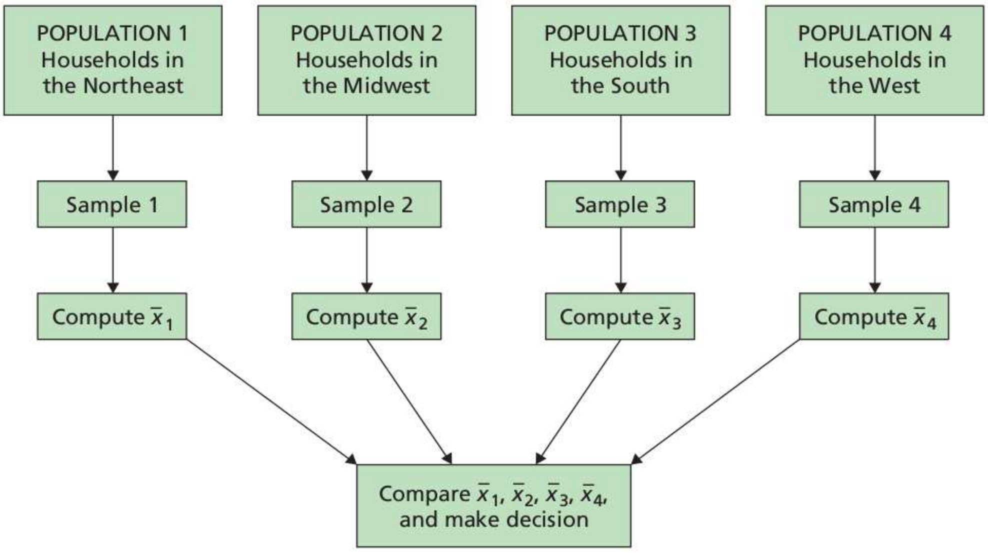
Roadmap

- ☐ Preparation
- ☐ What is SS (SSE, SSTR, SST)?
- ☐ What is df ?
- ☐ What is MS ?
- ☐ What is Fisher's F?
- ☐ When ANOVA does work?
- ☐ When ANOVA does not work?
- ☐ Wrong interpretations of ANOVA
- ☐ Related topics

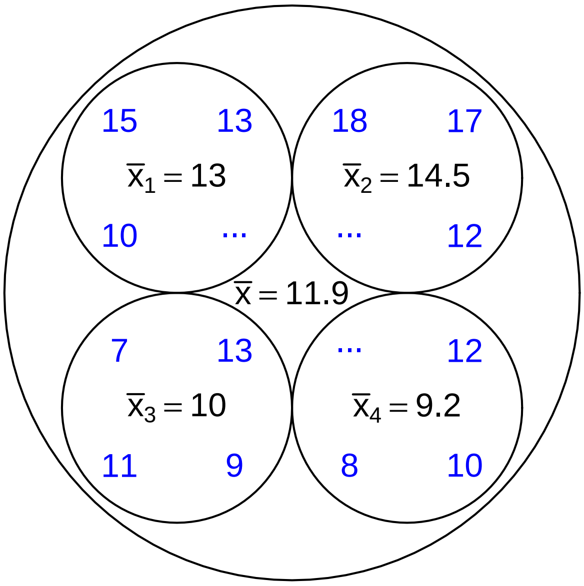
What is the question?

	$k=1$	$k=2$	$k=3$	$k=4$
	Northwest	Midwest	South	West
	15	17	11	10
	10	12	7	12
	13	18	9	8
	14	13	13	7
	13	15		9
		12		
$\overline{x}_k =$	13.0	14.5	10.0	9.2

Figure 16.5



Data Visualization



Testing Hypothesis

Question

$$\mu_{\text{Northwest}} = \mu_{\text{Midwest}} = \mu_{\text{South}} = \mu_{\text{West}}?$$

$$H_0 \;:\; \mu_k = \mu_{k'}$$
$$H_1 \;:\; \exists k \neq k', \text{such that } \mu_k \neq \mu_{k'}$$

When do you reject H_0?

Simplify

Remember the independent T-Test.

Northwest	Midwest
15	17
10	12
13	18
14	13
13	15
	12

T-Test

Remember the independent T-Test with equal variances.

T-Test

```
x <- c(15,10,13,14,13,
       17,12,18,13,15,12)

treat <- c(rep("Northeast",5),
           rep("Midwest",6))

treat <- as.factor(treat)

t.test(x~treat,var.equal=TRUE)
```

T-Test Output

```
        Two Sample t-test

data:  x by treat
t = 1.0783, df = 9, p-value = 0.309
alternative hypothesis: true difference in means is not equal to 0
95 percent confidence interval:
 -1.646908  4.646908
sample estimates:
  mean in group Midwest mean in group Northeast
                   14.5                    13.0
```

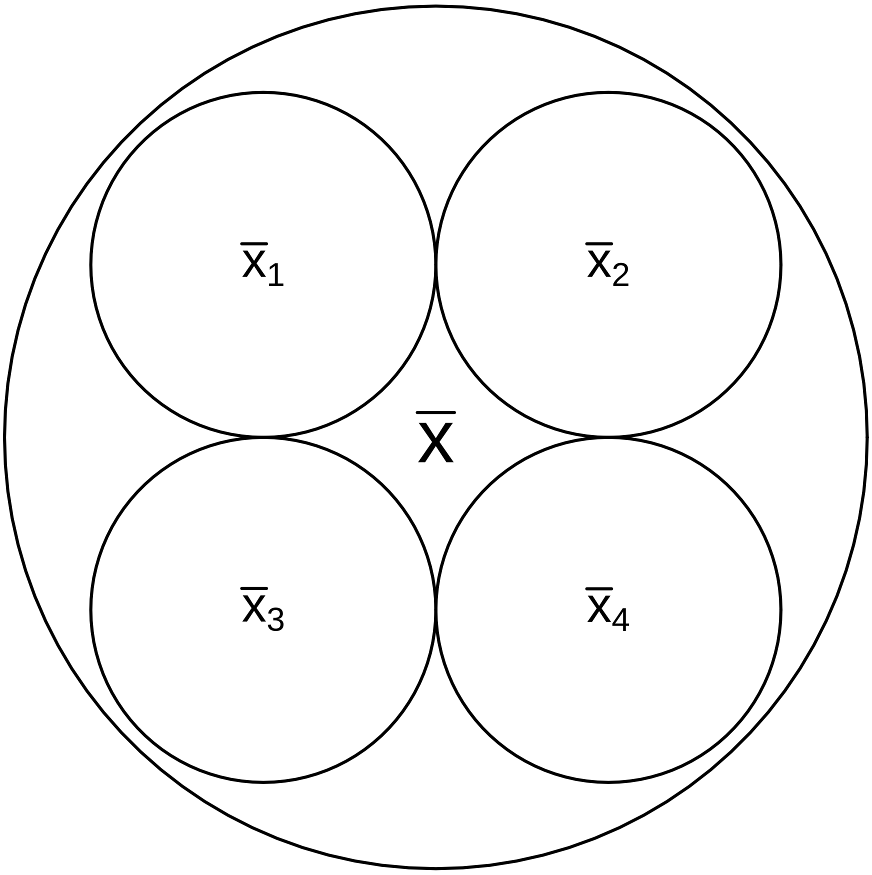

T-Test and ANOVA

```
summary(aov(x~treat))
```

```
            Df Sum Sq Mean Sq F value Pr(>F)
treat        1   6.14   6.136   1.163  0.309
Residuals    9  47.50   5.278
```

What is SS? Variation!

Intuition



Simple Math

Variation of data around $\overline{x}_1$: $\sum_i (x_{i1} - \overline{x}_1)^2$
Compute $\sum_{i=1}^{4} x_{i1}^2 - \frac{(\sum_{i=1}^{4} x_i)^2}{4}$ =14.0

Exercise: Variation of data around $\overline{x}_2$: $\sum_i (x_{i2} - \overline{x}_2)^2$

Variation of data around $\overline{x}_j$: $\sum_i (x_{ij} - \overline{x}_j)^2$

1) SSE: Sum of Error $\sum_j \sum_i (x_{ij} - \overline{x}_j)^2$
Compute

2) SST: Sum of total variation around $\overline{x}$: $\sum_j \sum_i (x_{ij} - \overline{x})^2$

Fisher's Decomposition:

$$SST = SSE + ?$$

18 / 28

Proof

$$\sum_j \sum_i (x_{ij} - \overline{x})^2 =$$

What is df ? Chi-square!

Remember if

$$\begin{aligned} x_1 &\sim \chi^2_{\mathrm{df}_1} \\ x_2 &\sim \chi^2_{\mathrm{df}_2} \end{aligned}$$

independently, then

$$x_1 + x_2 \sim \chi^2_{\mathrm{df}_1 + \mathrm{df}_2}$$

More details

Suppose $x_{ij} \sim N(\mu_j, \sigma^2)$

$$\frac{1}{\sigma^2}\sum_{i=1}^{n_1}(x_{i1} - \overline{x}_1)^2 \sim ?$$

$$\sum_{j=1}^{k}\frac{1}{\sigma^2}\sum_{i=1}^{n_1}(x_{i1} - \overline{x}_1)^2 \sim ?$$

Suppose $x_{ij} \sim N(\mu, \sigma^2)$

$$\frac{1}{\sigma^2}\sum_{j=1}^{k}\sum_{i=1}^{n_j}(x_{ij} - \overline{x})^2 \sim ?$$

22 / 28

What is MS? A Division

23 / 28

MS is simple

$$MS = SS/df$$

24 / 28

What is Fisher's F? Another Division

in ANOVA

$$F = \frac{MSTR}{MSE}$$

Fisher's Distribution

If X_1 is Chi-square with df_1 degrees of freedom, X_2 is another Chi-square with df_2 degrees of freedom independently. Then

$$F = \frac{X_1/\mathrm{df}_1}{X_2/\mathrm{df}_2}$$

is Fisher with numerator df_1 and denominator df_2 degrees of freedom, written as $F(\mathrm{df}_1, \mathrm{df}_2)$.

Exercise

Page 723

- ☐ Exercise 16.24
- ☐ Exercise 16.25

Challenge

- ☐ If T is student-t with n degrees of freedom, what is the distribution of T^2?
- ☐ $F(1, \infty)$ resembles which distribution?

- ☐ https://probstat.shinyapps.io/devoir2/
- ☐ 1681547
- ☐ 1677982