

ST 411/511 Outline 5

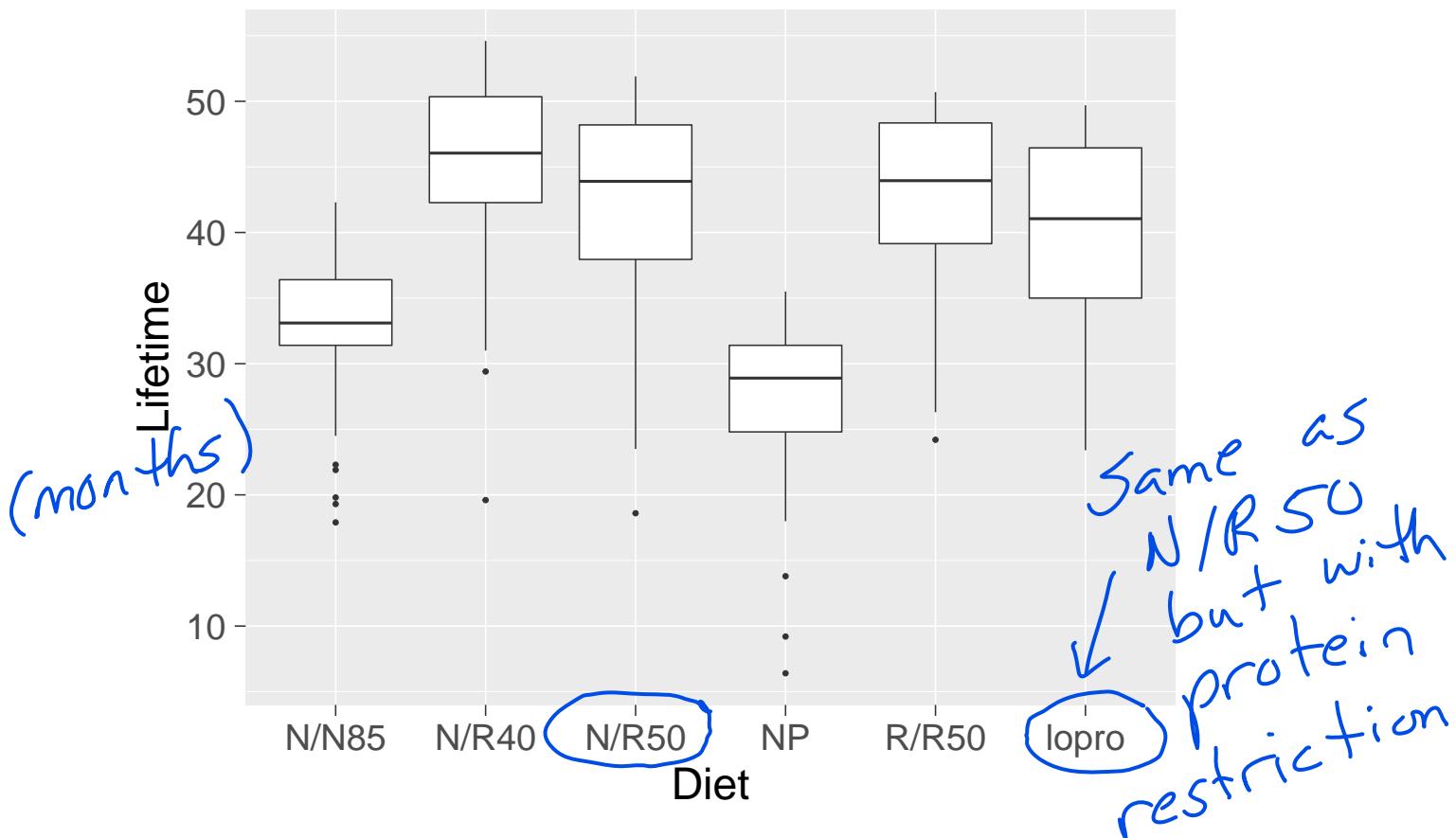
Reading assignment: Chapter 5. This chapter introduced one-way analysis of variance (ANOVA) a generalization the two-sample t-test for more than two groups.

Chapter 5 Comparisons Among Several Samples

Case Study 5.1.1: Diet restriction and longevity in mice

```
> qplot(Diet,Lifetime,data=case0501,geom="boxplot")
```

Want to compare popn means when more than 2 groups.



General research question: Does diet affect lifetime?

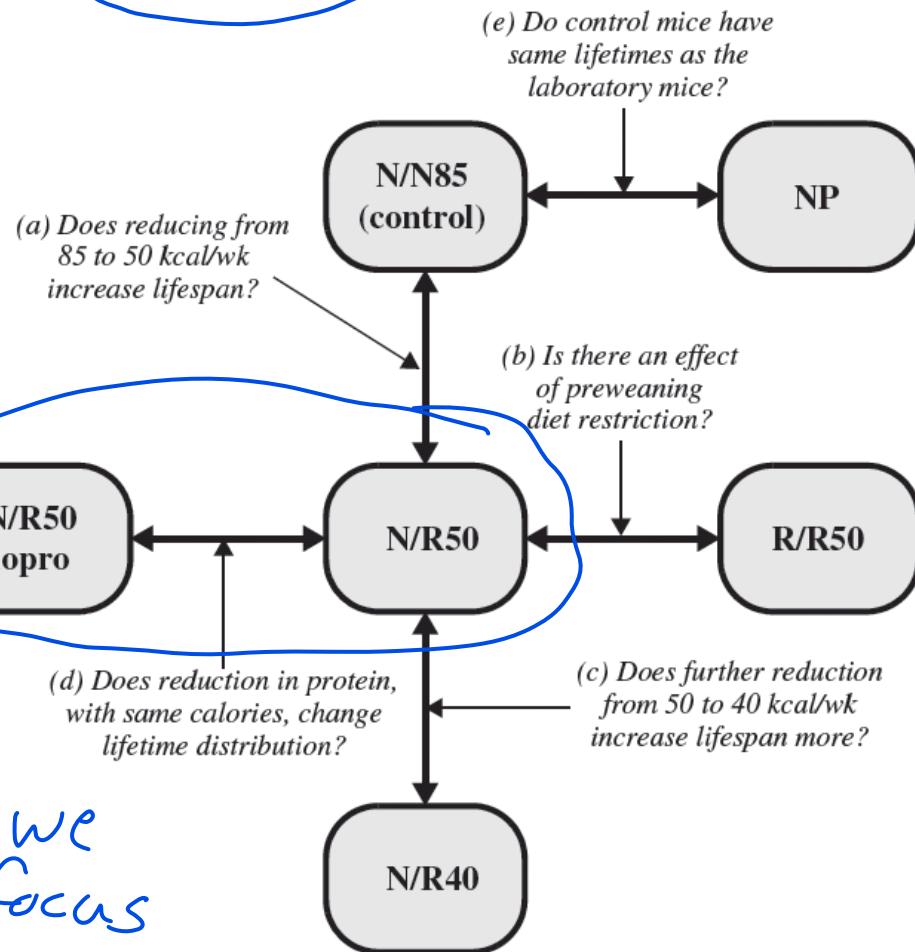
Do popn means differ?

Specific research questions:

How do pop'n means differ

DISPLAY 5.3

Structure of planned comparisons among groups in the diet restriction study



Later we will focus on this one.

These were decided at exp. design stage.

Looking at data to decide on comparisons is data snooping.

Terminology and notation

	2-sample <i>t</i> -test	More than two samples
Number of groups	2	I (index?)
Population means	μ_1, μ_2	$\mu_1, \mu_2, \dots, \mu_I$
Population standard deviation	σ	σ
Sample means	\bar{Y}_1, \bar{Y}_2	$\bar{Y}_1, \bar{Y}_2, \dots, \bar{Y}_I$
Sample standard deviations	s_1, s_2	s_1, s_2, \dots, s_I
Sample sizes	n_1, n_2	n_1, n_2, \dots, n_I
$s_p = \text{pooled estimate of } \sigma$	$\sqrt{\frac{(n_1 - 1)s_1^2 + (n_2 - 1)s_2^2}{n_1 + n_2 - 2}}$	$\frac{(n_1 - 1)s_1^2 + (n_2 - 1)s_2^2 + \dots + (n_I - 1)s_I^2}{n_1 + n_2 + \dots + n_I - I}$ <p style="margin-left: 100px;">\uparrow Sample size - # mean parameters</p>
$df =$ sample size - # mean parameters		\uparrow Sample size - # mean params. $= df$

	2-sample <i>t</i> -test	More than two samples
Usual H_0	$H_0 : \mu_1 - \mu_2 = 0$ $H_0 : \mu_1 = \mu_2$	$H_0 : \mu_1 = \mu_2 = \dots = \mu_I$
Usual H_A	$H_A : \mu_1 - \mu_2 \neq 0$	$H_A : \text{at least one } \mu_i \text{ is different.}$
Test statistic	$t\text{-stat} = \frac{\bar{Y}_1 - \bar{Y}_2}{s_p \sqrt{\frac{1}{n_1} + \frac{1}{n_2}}}$	F-stat (details below) If $I = 2$, $F\text{-stat} = t\text{-stat}^2$

Structure of the F-test

$$H_0 : \mu_1 = \mu_2 = \dots = \mu_I$$

H_0 specifies the null model
"equal means model"

Simple - only one parameter

$$H_A : \text{not all the } \mu_i \text{ are equal}$$

H_A allows each μ_i to be

different "separate means model"

more complex - I parameters

$$F\text{-statistic} = \frac{(\text{extra sum of squares}) / (\text{extra df})}{\hat{\sigma}_{\text{full}}^2}$$

Test stat needs to detect evidence against H_0 in favor of H_A .

Extra sum of squares (SS)

Measures how much better the sep. means model fits the data compared to equal means model.
sep. = separate

Extra df (degrees of freedom)

Measures how much more complex sep. means model is than equal means model.

(extra sum of squares)/(extra df) \approx improvement in model fit per unit of added complexity.

$\hat{\sigma}_{\text{full}}^2 = s_p^2$ = pooled variance using sep. means model.

It's in the denominator of F-stat as a scaling factor so F-stat has an F-dist'n.

Calculating F-statistic and p-value in R

response

```
> # Check structure of data frame  
> head(case0501)
```

Lifetime Diet ← grouping var.

1	35.5	NP
2	35.4	NP
3	34.9	NP
4	34.8	NP
5	33.8	NP
6	33.5	NP

usual formula

```
> # Capture the aov object  
> case0501_aov <- aov(Lifetime ~ Diet, data=case0501)
```

```
> # and get the ANOVA table.  
> anova(case0501_aov)
```

Analysis of Variance Table

Response: Lifetime

	Df	Sum Sq	Mean Sq	F value	Pr(>F)
Diet	5	12734	2546.8	57.104	< 2.2e-16 ***
Residuals	343	15297	44.6	$S_p^2 = \hat{S}_F^2 / I$	

Signif. codes: 0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

$2.2e-16 = 2.2 \cdot 10^{-16} = \text{smallest pos. } \#$
my computer knows.

Statistical conclusion:

Overwhelming evidence that

diff diets caused * differences in mean
mouse lifetimes ($p \approx 0$, ANOVA F-test)

* mice randomly assigned to diets.

ANOVA F-test and two-sample t-test are equivalent for two groups.

Revisit finch data of cast study 2.1.1.

Notation: μ_1, μ_2 are population means in 1976 and 1978, respectively.

Chapter 2 analysis:

$$H_0: \mu_1 - \mu_2 = 0 \quad H_A: \mu_1 - \mu_2 \neq 0$$

< ch. 2

> # Two-sample, two-sided t-test with finch data.

> t.test(Depth ~ Year, data=case0201, var.equal=TRUE)

Two Sample t-test

data: Depth by Year

t = -4.5833, df = 176, p-value = 8.65e-06

alternative hypothesis: true difference in means is not equal to 0

95 percent confidence interval:

-0.9564088 -0.3806698

sample estimates:

mean in group 1976 mean in group 1978

9.469663

10.138202

Chapter 5 analysis:

$$H_0: \mu_1 = \mu_2$$

$$H_A: \mu_1 \neq \mu_2$$

p-values
same

> # ANOVA F-test with finch data.

> case0201_aov <- aov(Depth ~ Year, data=case0201)

> anova(case0201_aov)

Analysis of Variance Table

Response: Depth

	Df	Sum Sq	Mean Sq	F value	Pr(>F)	
Year	1	19.889	19.8890	21.006	8.65e-06	***
Residuals	176	166.638	0.9468			
				$F_{full}^2 = S_p^2$		
Signif. codes:	0	'***'	0.001	'**'	0.01	'*'
				0.05	'. '	0.1
						1

Pooled estimate of standard deviation

Two-sample t-test:

$$s_p = \sqrt{\frac{(n_1 - 1)s_1^2 + (n_2 - 1)s_2^2}{n_1 + n_2 - 2}}$$

$$df = n_1 + n_2 - 2$$

ANOVA F-test:

$$\hat{s}_{\text{full}} = s_p = \sqrt{\frac{(n_1 - 1)s_1^2 + \dots + (n_I - 1)s_I^2}{n_1 + \dots + n_I - I}}$$

$$\text{res } df = n_1 + \dots + n_I - I$$

total sample size — # mean parameters

DISPLAY 5.6 Pooled estimate of standard deviation; diet restriction data		
Group	n	Sample SD
NP	49	6.1
N/N 85	57	5.1
N/R50	71	7.8
R/R50	56	6.7
N/R50 lopro	56	7.0
N/R40	60	6.7

Calculate the pooled estimate of variance, s_p^2 . 1

$$s_p^2 = \frac{(49 - 1)(6.1)^2 + (57 - 1)(5.1)^2 + (71 - 1)(7.8)^2 + (56 - 1)(6.7)^2 + (56 - 1)(7.0)^2 + (60 - 1)(6.7)^2}{(49 - 1) + (57 - 1) + (71 - 1) + (56 - 1) + (56 - 1) + (60 - 1)}$$

$$= \frac{15,313.90}{343} = 44.647; \quad s_p = \sqrt{44.647} = 6.68$$

s_p is the square root. 2

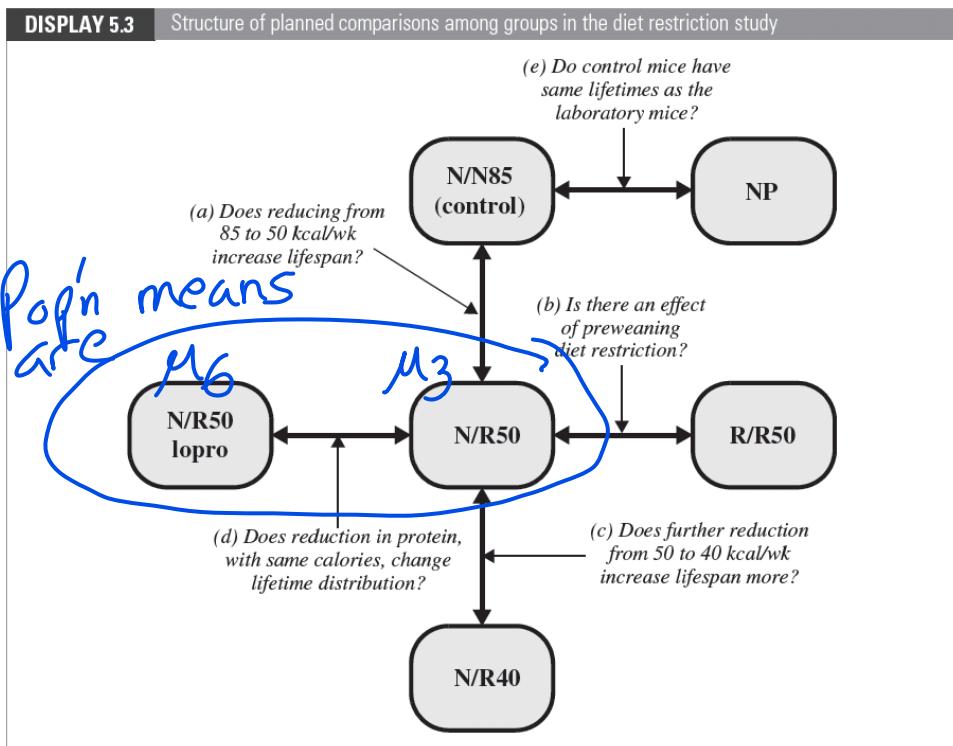
$d.f.$ is the denominator. 3

```
> # Getting the group standard deviations.
> # Split into individual groups, then apply sd()
>
> with(case0501, unlist(lapply(split(Lifetime, Diet), sd)))
  N/N85      N/R40      N/R50       NP      R/R50      lopro
  5.125297  6.703406  7.768195  6.133701  6.683152  6.991695 =  $s_1, \dots, s_I$ 

> # Can find sample sizes as well.
> with(case0501, unlist(lapply(split(Lifetime, Diet), length)))
N/N85 N/R40 N/R50      NP R/R50 lopro
      57      60      71      49      56      56 =  $n_1, \dots, n_I$ 
We won't calculate  $s_p$  by hand. Find  $s_p^2$  on anova() output.

> # Total sample size
> nrow(case0501)
[1] 349 =  $n_1 + \dots + n_I$ 
```

Post ANOVA F-test Analysis



Focus on one of these preplanned comparisons

see p. 32 of outline 2

95% Confidence interval for $\mu_3 - \mu_6$

$$\alpha = 0.05$$

$$pt\ est \pm t_{df}(1 - \alpha/2)SE(pt\ est)$$

$$\bar{Y}_3 - \bar{Y}_6$$

exactly as
before

Same as

before except

$$df = n_1 + \dots + n_6 - 6$$

df associated with s_p

$$s_p \sqrt{\frac{1}{n_3} + \frac{1}{n_6}}$$

Same as before

except use s_p
calculated using
all the groups.

Get s_p^2 & df from residual mean square & df on ANOVA table.

```

> # Need group sample means.
> with(case0501, unlist(lapply(split(Lifetime, Diet), mean)))
  N/N85    N/R40    N/R50      NP    R/R50    lopro
  32.69123 45.11667 42.29718 27.40204 42.88571 39.68571

```

$$1 - \frac{\alpha}{2}$$

$$\bar{Y}_3$$

$$\bar{Y}_6$$

> # Appropriate t quantile for 95% CI.

> qt(0.975, 343)

[1] 1.966904

"

df for residuals on p. 6

$$t_{343}(0.975)$$

> # Inappropriate t quantile for 95% CI

> qt(0.975, 125)

[1] 1.979124

$$n_3 + n_6 - 2$$

This would give a wider CI - not so precise.

Confidence Interval: $\text{pt est} \pm t_{\text{df}}(1 - \alpha/2)\text{SE}(\text{pt est})$

$$42.297 - 39.686 \pm 1.9669 \cdot \sqrt{44.6}$$

$$\sqrt{\frac{1}{71} + \frac{1}{56}}$$

$$\approx (0.264, 5).$$

residual
mean S_e

from ANOVA
table on p. 6

Sample
size from
page 8

> 42.29718 - 39.68571 - 1.966904*sqrt(44.6)*sqrt(1/71 + 1/56)

[1] 0.2638421

> 42.29718 - 39.68571 + 1.966904*sqrt(44.6)*sqrt(1/71 + 1/56)

[1] 4.959098

Statistical Conclusion: We estimate that pop'n mean mouse lifetime on N/R50 diet is 0.26 to 5 months more than pop'n mean mouse lifetime on lopro diet (95% CI).

Hypothesis Test to Compare Two Means

Hypotheses:

$$H_0: \mu_3 - \mu_6 = 0$$

$$H_A: \mu_3 - \mu_6 \neq 0$$

Could do one-sided test

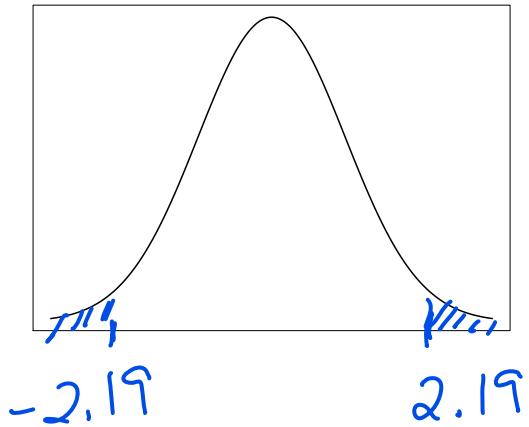
Test Statistic:

$$t\text{-stat} = \frac{\text{pt est} - \text{value under } H_0}{\text{SE(pt est)}}$$

$$= \frac{42.297 - 39.686 - 0}{\sqrt{44.6} \sqrt{\frac{1}{71} + \frac{1}{55}}} \\ \text{SE from CI}$$

$$\approx 2.19$$

Sampling distribution of test statistic under H_0 :



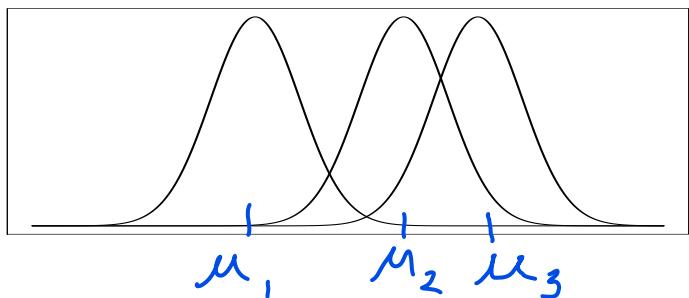
t with $df = 343$
(df must go with Sp)

```
> # Find (two-sided) p-value (cf. item 5(d) from Lab 2).
> 2 * (1-pt(2.19, 343))
[1] 0.02919721
```

Statistical conclusion Moderate evidence that pop'n mean lifetimes differ between N/R50 & lopro diets ($p \approx 0.029$, 2-sided t -test).

Assumptions for ANOVA F-test

- Normality



3 popns
($I = 3$)

- Equal standard deviations

Each normal dist'n has same σ

- Independence No clustering or serial dependence.

e.g. If several mice housed in the same cage, then that would be a cluster.

Checking Assumptions with Residual Plots

Notation and terminology:

Y_{ij} = j^{th} response in i^{th} group

\bar{Y}_i = sample mean in i^{th} group =
est. of M_i

$Y_{ij} - \bar{Y}_i$ = ij^{th} residual

= ij^{th} obs. - estimated mean of its popn

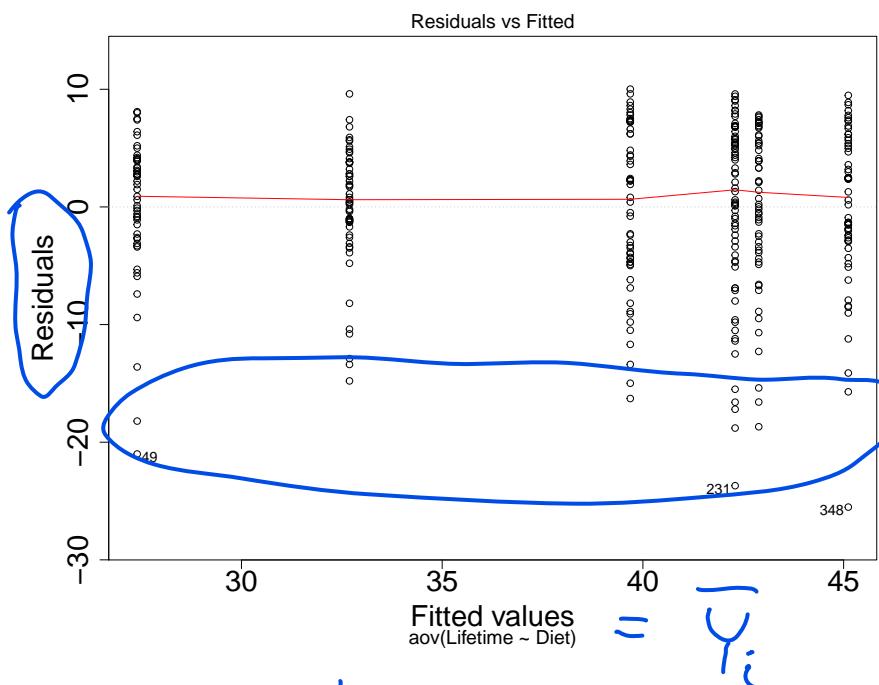
= diff between what we observe
and what model estimates.

What do residuals tell us about assumptions?

If normality and equal standard deviation assumptions are met, then residuals will be normal with population mean approximately 0 and approximately the equal standard deviation.

\bar{Y}_{ij} estimates μ_{ij} , so $Y_{ij} - \bar{Y}_{ij} \approx Y_{ij} - \mu_{ij}$,
so if i^{th} popn is normal, then ij^{th} residual is approx a random draw from a normal dist'n with mean 0.

```
> # Earlier, we captured the aov object.  
> case0501_aov <- aov(Lifetime ~ Diet, data=case0501)  
> plot(case0501_aov, which=1)
```

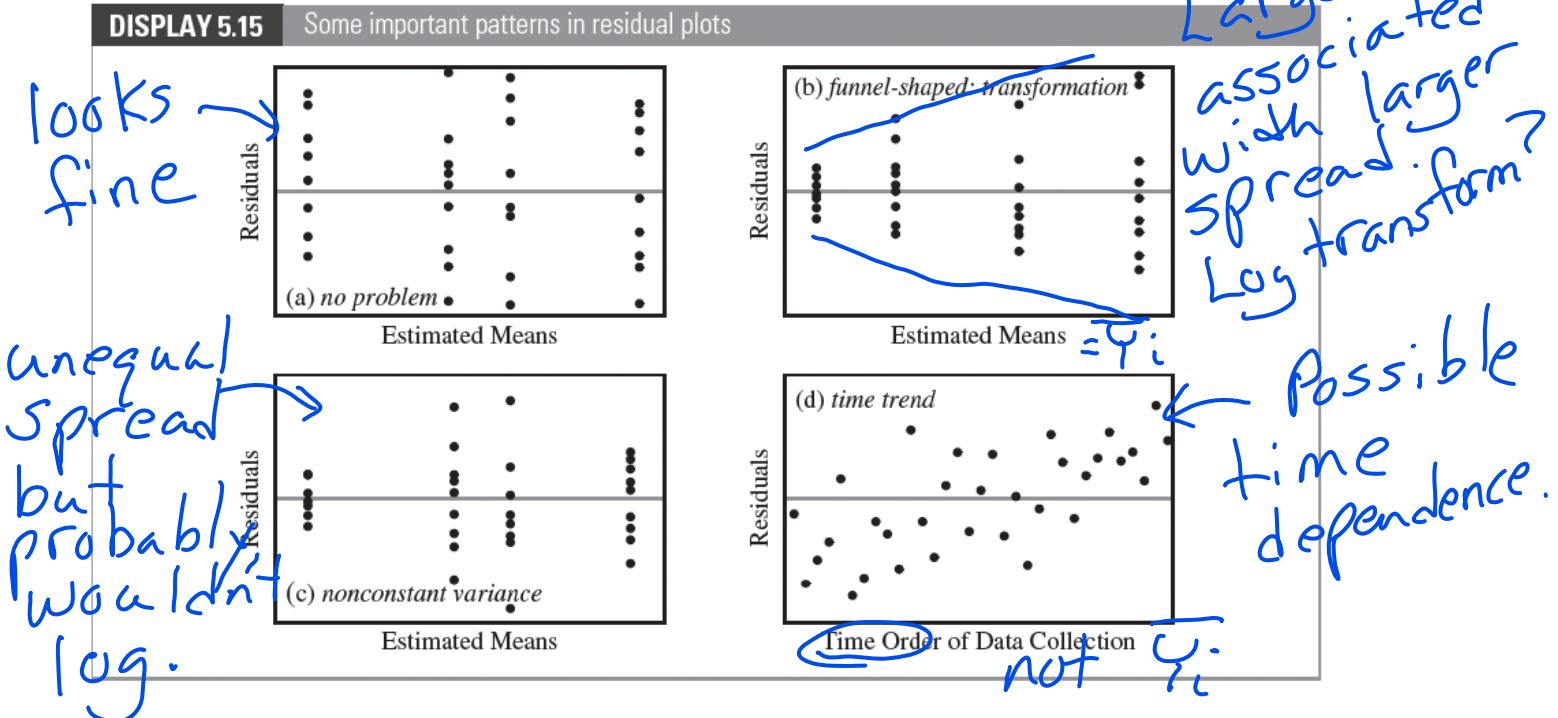


6 groups, so
6 sets of
residuals

maybe a
little skewness
Nearly-equal
spread say
equal var. is
OK

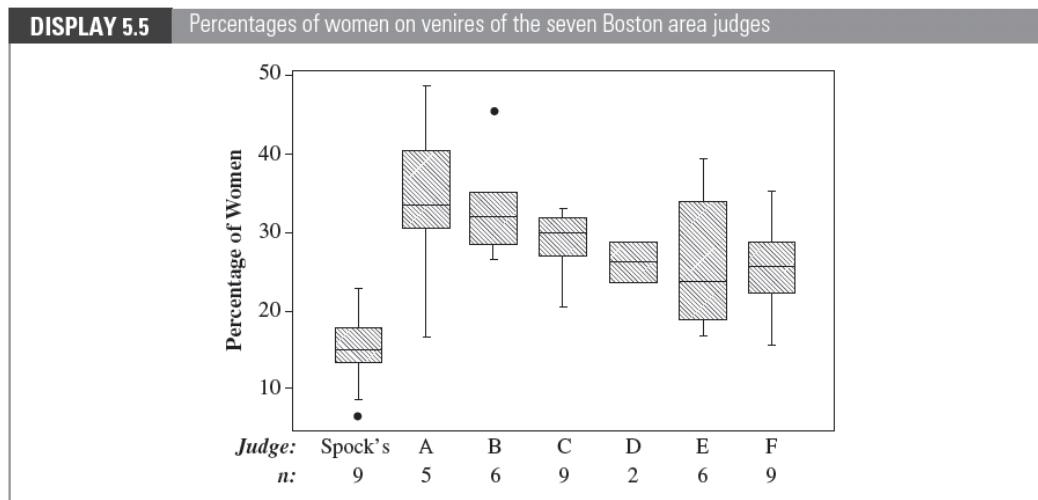
If normality OK, then each group
should look like a sample from
a normal dist'n with mean 0.

What you might see in a plot of residuals vs. fitted values:



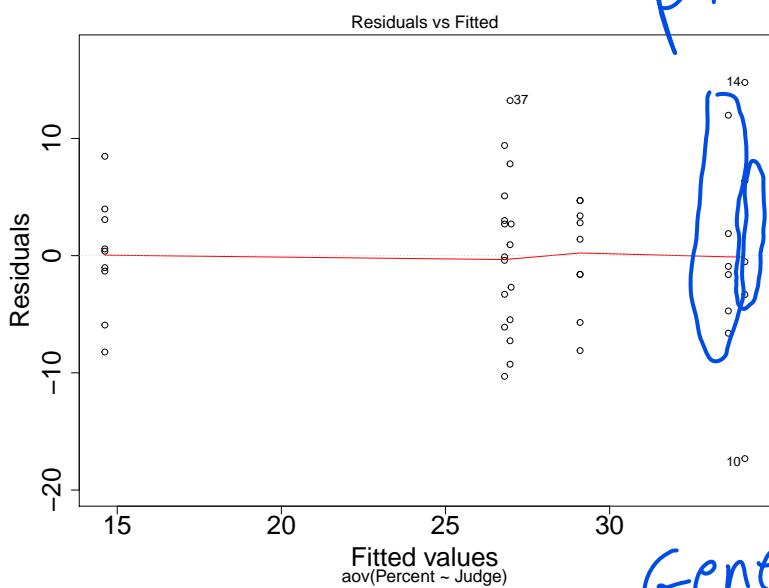
Normality: symmetry across 0
 Equal std. dev.: similar vertical spread for all groups.

Case Study 5.1.2: Dr. Spock conspiracy trial



```
> head(case0502)
Percent Judge
1 6.4 Spock's
2 8.7 Spock's
3 13.3 Spock's
4 13.6 Spock's
5 15.0 Spock's
6 15.2 Spock's
```

```
> case0502_aov <- aov(Percent~Judge, data=case0502)
> plot(case0502_aov, which=1)
```



This command makes
plot of residuals
vs. fits

Hard to see
all 7 groups

```
> with(case0502, kruskal.test(Percent, Judge))
```

Generalizes rank-sum
test for more than 2 groups

Kruskal-Wallis rank sum test

```
data: Percent and Judge
Kruskal-Wallis chi-squared = 21.965, df = 6, p-value = 0.001229
```

Just be aware this test exists.
No HW + exam in ST 411/511.

Statistical Models

Notation and terminology

Says something about random variation in data.

$\mu\{Y_{ij}\}$ = pop'n mean of pop'n where we got y_{ij}
 "pop'n mean of y_{ij} "

Our (full) model in Chapter 5

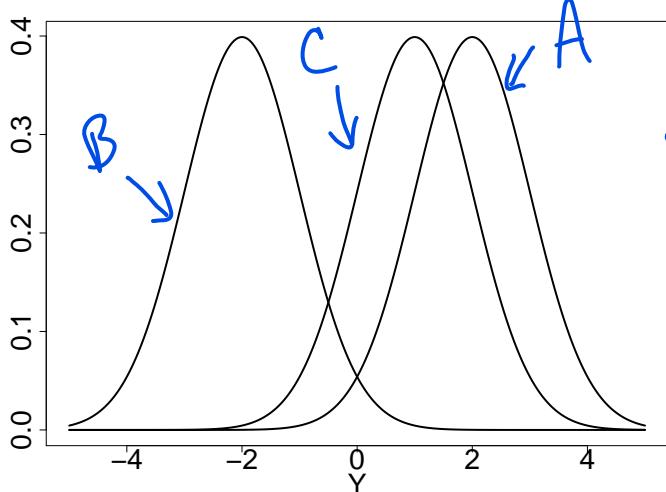
Separate means

$\mu\{Y_{ij}\} = \mu_i$ = i-th pop'n mean

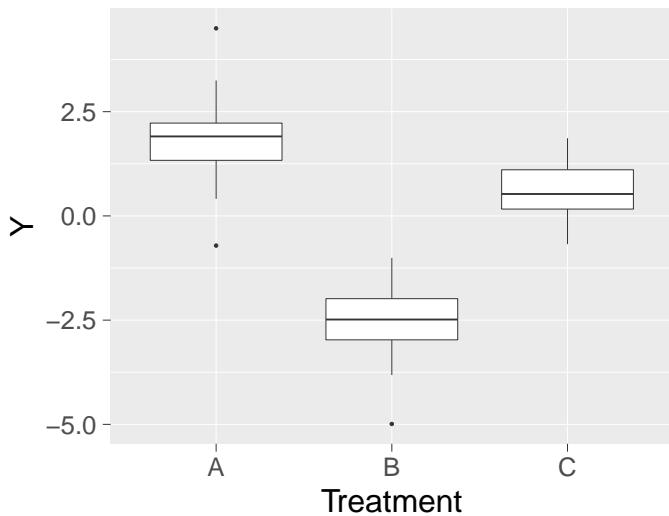
and furthermore Y_{ij} is normally distributed with standard deviation σ .

In pictures:

Hypothetical example



Normal pop'n
for each group.
We don't see this.



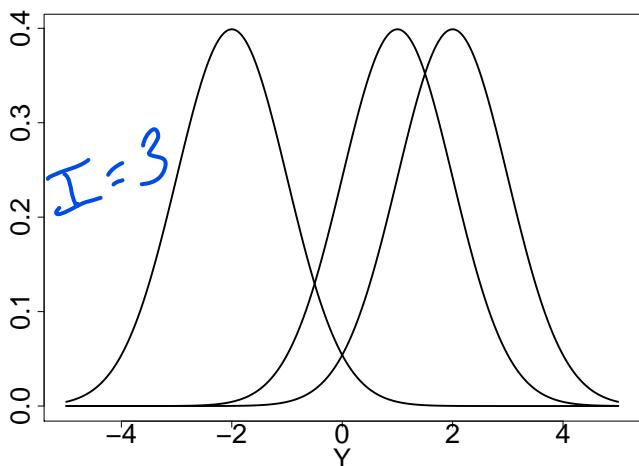
What we see
is just the
data

ANOVA F-test Details—Comparing Two Models

$H_0: \mu_1 = \dots = \mu_I$ all pop's have same mean
 $H_A: \text{at least } \mu_i \text{ is different}$
 one

Full model: allowed under H_A

sep. means model

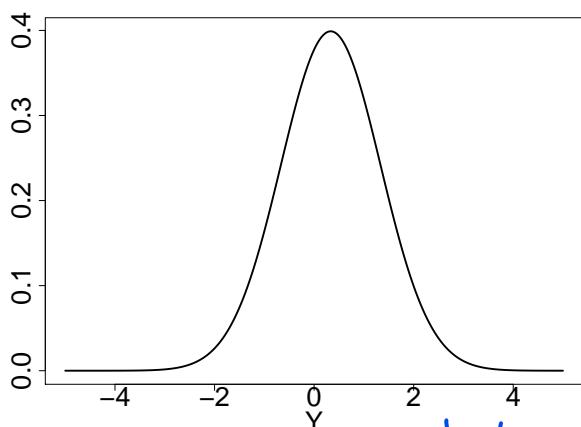


$$\mu\{Y_{ij}\} = \mu_i$$

and all pop's are normally-distributed with std. dev. σ

mean parameters
 $= I$

Reduced model: specified by H_0
 equal means model



$$\mu\{Y_{ij}\} = \mu$$

and pop'n is normally-distributed with

std. dev. σ
 # mean parameters = 1

Equal means model is a special case of
 sep. means model when $\underbrace{\mu_1 = \dots = \mu_I}_{H_0}$
 Full model is bigger.

Test statistic compares residual variation between full and reduced models

$$F\text{-statistic} = \frac{(\text{extra sum of squares}) / (\text{extra df})}{\hat{\sigma}_{\text{full}}^2}$$

ANOVA F-test

Extra sum of squares =

improvement in model fit when
means allowed to differ vs. held constant

full model

reduced model

= residual SS from reduced model - residual SS from full model

Residual SS measures how closely
model matches data.

Recall: ij th residual = ij th observation - model's estimate of population mean of ij th observation

$$Y_{ij} - \bar{Y}_i$$

- \bar{Y}_i ↗ sep. means
model's est.
of $\mu\{Y_{ij}\}$

Residual (separate means model):

$$Y_{ij} - \bar{Y}_i$$

Residual (equal means model):

$$Y_{ij} - \bar{Y}$$

$$\mu\{Y_{ij}\} = \mu$$

\bar{Y} = average of
all data

Residual sum of squares (residual SS)

res SS

residuals for a model squared and summed

Residual SS(separate means model) = $\sum_{i=1}^I \sum_{j=1}^{n_i} (Y_{ij} - \bar{Y}_i)^2$

for all combinations of i & j add up res. for sep. means

Residual SS(equal means model) = $\sum_{i=1}^I \sum_{j=1}^{n_i} (Y_{ij} - \bar{Y})^2$

res. for equal means

extra SS = $\underbrace{\text{resSS(red. model)}}_{\text{equal means}} - \underbrace{\text{resSS(full)}}_{\text{sep means}}$

res. df(equal means model) - res. df(separate means model)

$(n_1 + \dots + n_I - I) - (n_1 + \dots + n_I - I) = I - 1$

Sample size minus 1 (one mean)

number additional parameters in full model = extra df

Mean squares

SS/\bar{df}

Extra mean square =

extra SS / extra df

Residual mean square =

for full model

$\frac{\text{resSS(fall model)}}{n_1 + \dots + n_I - I} = S_p^2$

$MSE = \hat{S}_{full}^2 = \text{"mean squared error"}$

How much better full model fits data

Extra SS

additional parameters
Extra df

$$F\text{-statistic} = \frac{\{\text{residual SS(reduced)} - \text{residual SS(full)}\}}{\{\text{df(reduced)} - \text{df(full)}\}} \cdot \frac{\sigma^2_{\text{full}}}{\text{MSE}}$$

≈ improvement in model fit by full model over reduced model per additional parameter, scaled by MSE.

Large F-stat says full model is worth the extra complexity.

$$MS = \frac{SS}{df}$$

DISPLAY 5.10

Analysis of variance table: a test for equal mean percentages of women in venires of seven judges (Spock data)

Source of Variation	Sum of Squares	d.f.	Mean Square	F-Statistic	p-Value
Between Groups	1,927.08	6	321.18	6.72	0.000061
Within Groups	1,864.45	39	47.81		
Total	3,791.53	45			

extra SS adf

residual for full model

Res. SS (equal means)

res. df equal means

Annotations:

- (1) Sum of squared residuals from fitting the full (separate-means) model.
- (2) Sum of squared residuals from fitting the reduced (equal-means) model.
- (3) degrees of freedom.
- (4) Extra sum of squares is Total minus within (d.f. also obtained by subtraction).
- (5) A mean square is the ratio of a sum of squares to its degrees of freedom.
- (6) The F-statistic is the ratio of the Between MS to the Within MS.
- (7) The p-value comes from an F-distribution with 6 and 39 d.f.

Note: This is s_p^2 .

F-distn has two parameters

numerator df
extra df
6

& denominator df
residual df
39

```
> case0502_aov <- aov(Percent~Judge, data=case0502)
> anova(case0502_aov)
```

Analysis of Variance Table

Response: Percent

	Df	Sum Sq	Mean Sq	F value	Pr(>F)
Judge	6	1927.1	321.18	6.7184	6.096e-05 ***
Residuals	39	1864.5	47.81	= S_p^2	= $\hat{\sigma}_{full}^2$ = MSE *
					resSS (full model)

Signif. codes:	0	***	0.001	**	0.01 *
					0.05 .
					0.1 ' '
					1

$$df = n - I = \text{total sample size} - \# \text{ mean parameters}$$

$$\begin{aligned} \text{extra df} &= \text{diff. in } \# \text{ parameters} \\ &\quad \text{between sep means model} \\ &\quad \text{and equal means model} \\ &= I - 1 \end{aligned}$$

* "Error" and "residual" are synonyms.

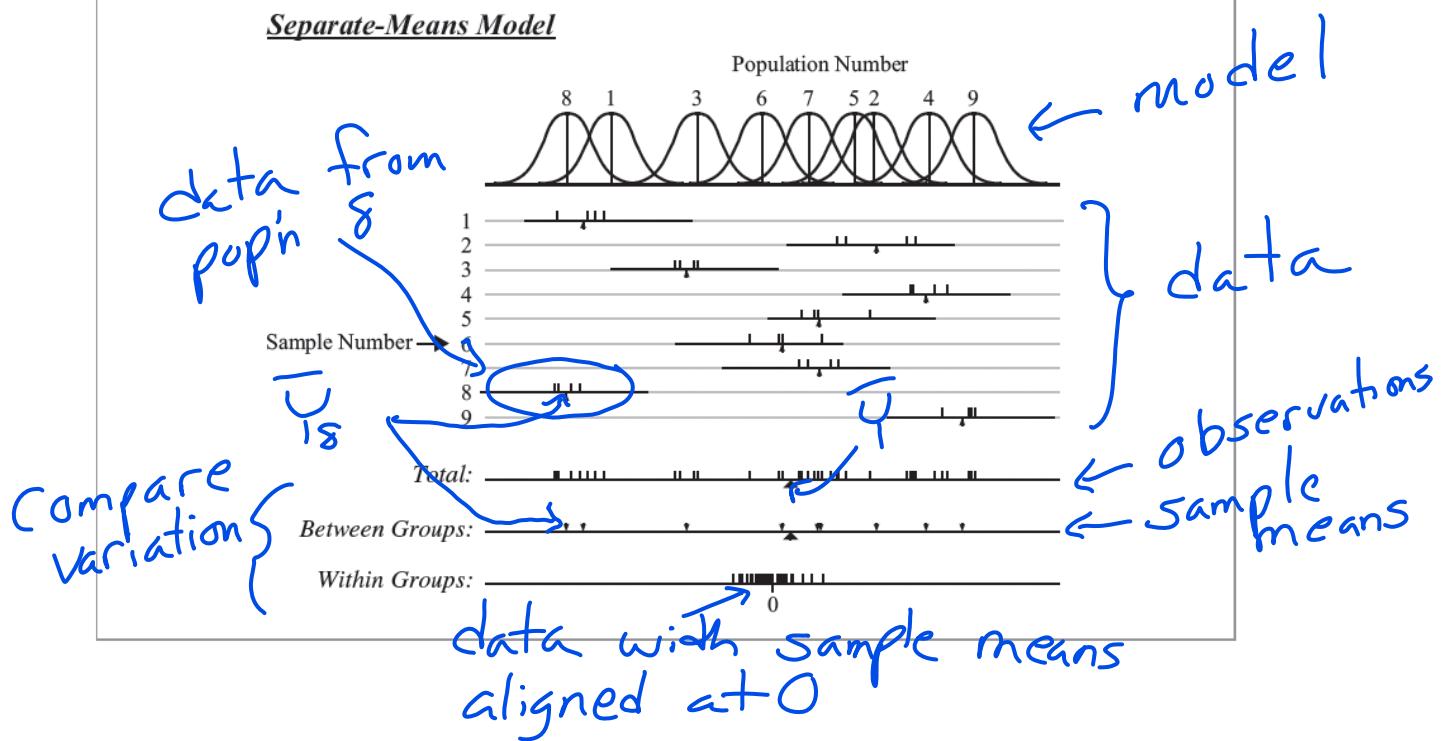
Another perspective on F-test

ANOVA = ANalysis Of VAriance

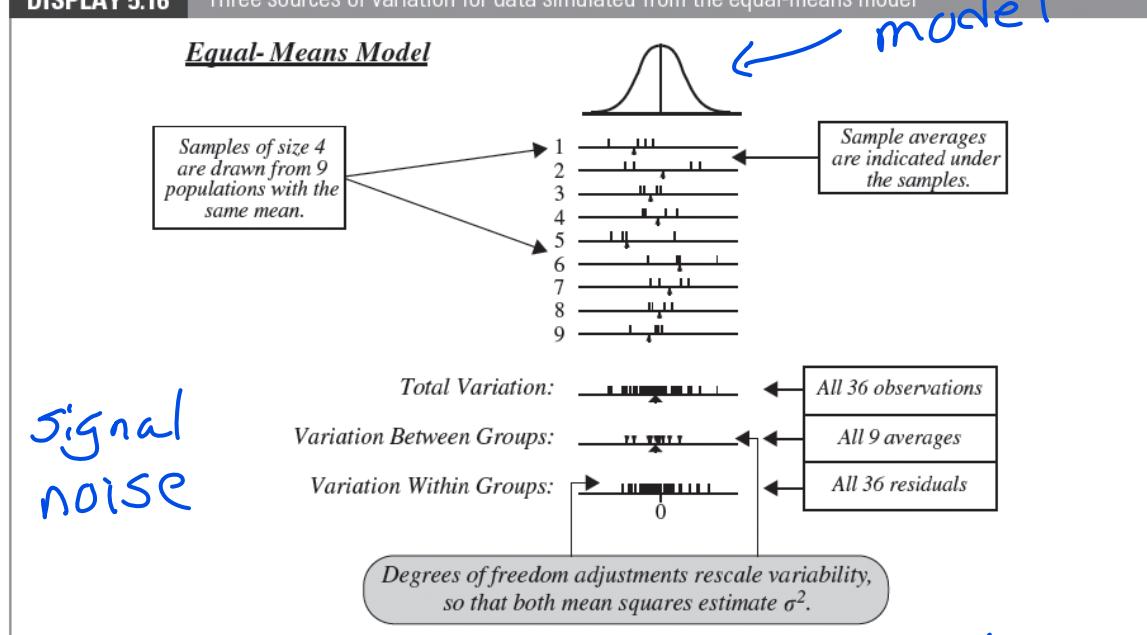
Two sources of variation in data

- Between groups: variation among sample means "signal"
- Within groups: variation of observations around their sample means "noise"

DISPLAY 5.17 Variations in the several-group problem for data simulated from the separate-means model



DISPLAY 5.16 Three sources of variation for data simulated from the equal-means model



F-test asks, "is Signal much larger than noise?"