

# Analysis of variance

# Packages

```
library(tidyverse)
library(smmr)
library(PMCMRplus)
```

# Jumping rats

- Link between exercise and healthy bones (many studies).
- Exercise stresses bones and causes them to get stronger.
- Study (Purdue): effect of jumping on bone density of growing rats.
- 30 rats, randomly assigned to 1 of 3 treatments:
  - ▶ No jumping (control)
  - ▶ Low-jump treatment (30 cm)
  - ▶ High-jump treatment (60 cm)
- 8 weeks, 10 jumps/day, 5 days/week.
- Bone density of rats ( $\text{mg}/\text{cm}^3$ ) measured at end.

## Jumping rats 2/2

- See whether larger amount of exercise (jumping) went with higher bone density.
- Random assignment: rats in each group similar in all important ways.
- So entitled to draw conclusions about cause and effect.

# Reading the data

Values separated by spaces:

```
my_url <- "http://ritsokiguess.site/datafiles/jumping.txt"  
rats <- read_delim(my_url," ")
```

## The data (some random rows)

```
# rats %>% slice_sample(n=10)
```

```
rats
```

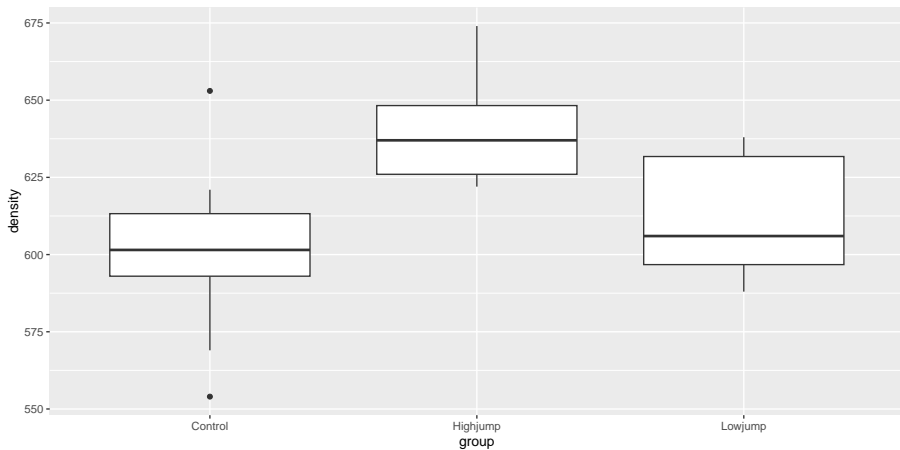
```
# A tibble: 30 x 2
```

	group	density
	<chr>	<dbl>
1	Control	611
2	Control	621
3	Control	614
4	Control	593
5	Control	593
6	Control	653
7	Control	600
8	Control	554
9	Control	603
10	Control	569

```
# i 20 more rows
```

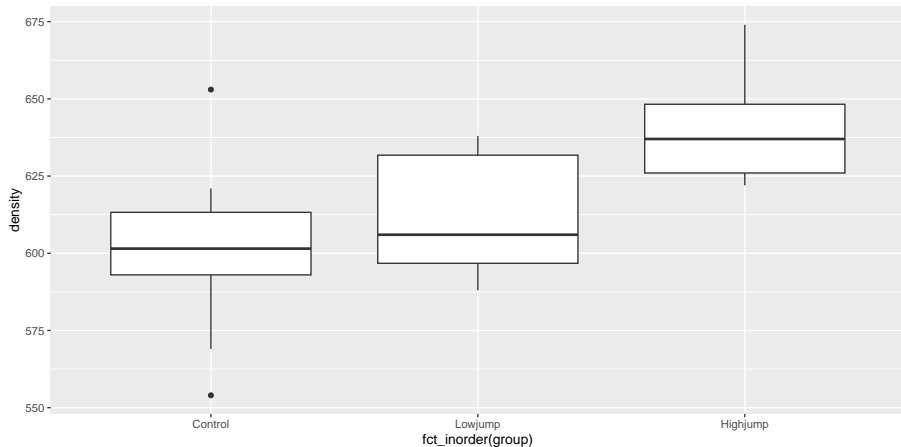
# Boxplots

```
ggplot(rats, aes(y=density, x=group)) + geom_boxplot()
```



## Or, arranging groups in data (logical) order

```
ggplot(rats, aes(y=density, x=fct_inorder(group))) +  
  geom_boxplot()
```





# Analysis of Variance

- Comparing  $> 2$  groups of independent observations (each rat only does one amount of jumping).
- Standard procedure: analysis of variance (ANOVA).
- Null hypothesis: all groups have same mean.
- Alternative: “not all means the same”, at least one is different from others.

## Testing: ANOVA in R

```
rats.aov <- aov(density~group,data=rats)
summary(rats.aov)
```

	Df	Sum Sq	Mean Sq	F value	Pr(>F)
group	2	7434	3717	7.978	0.0019 **
Residuals	27	12579	466		

---

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

- Usual ANOVA table, small P-value: significant result.
- Conclude that the mean bone densities are not all equal.
- Reject null, but not very useful finding.

# Which groups are different from which?

- ANOVA really only answers half our questions: it says “there are differences”, but doesn’t tell us which groups different.
- One possibility (not the best): compare all possible pairs of groups, via two-sample t.
- First pick out each group:

```
rats %>% filter(group=="Control") -> controls  
rats %>% filter(group=="Lowjump") -> lows  
rats %>% filter(group=="Highjump") -> highs
```

## Control vs. low

```
t.test(controls$density, lows$density)
```

Welch Two Sample t-test

data: controls\$density and lows\$density

t = -1.0761, df = 16.191, p-value = 0.2977

alternative hypothesis: true difference in means is not equal

95 percent confidence interval:

-33.83725 11.03725

sample estimates:

mean of x mean of y

601.1 612.5

No sig. difference here.

## Control vs. high

```
t.test(controls$density, highs$density)
```

Welch Two Sample t-test

data: controls\$density and highs\$density

t = -3.7155, df = 14.831, p-value = 0.002109

alternative hypothesis: true difference in means is not equal

95 percent confidence interval:

-59.19139 -16.00861

sample estimates:

mean of x mean of y

601.1 638.7

These are different.

## Low vs. high

```
t.test( lows$density, highs$density)
```

Welch Two Sample t-test

```
data:  lows$density and highs$density
t = -3.2523, df = 17.597, p-value = 0.004525
alternative hypothesis: true difference in means is not equal
95 percent confidence interval:
 -43.15242  -9.24758
sample estimates:
mean of x mean of y
  612.5    638.7
```

These are different too.

# But...

- We just did 3 tests instead of 1.
- So we have given ourselves 3 chances to reject  $H_0$  : all means equal, instead of 1.
- Thus  $\alpha$  for this combined test is not 0.05.

# John W. Tukey



- American statistician, 1915–2000
- Big fan of exploratory data analysis
- Popularized boxplot
- Invented “honestly significant differences”
- Invented jackknife estimation
- Coined computing term “bit”
- Co-inventor of Fast Fourier Transform



# Honestly Significant Differences

- Compare several groups with one test, telling you which groups differ from which.
- Idea: if all population means equal, find distribution of highest sample mean minus lowest sample mean.
- Any means unusually different compared to that declared significantly different.

## Tukey on rat data

```
rats.aov <- aov(density~group, data = rats)
TukeyHSD(rats.aov)
```

Tukey multiple comparisons of means  
95% family-wise confidence level

Fit: aov(formula = density ~ group, data = rats)

\$group

	diff	lwr	upr	p adj
Highjump-Control	37.6	13.66604	61.533957	0.0016388
Lowjump-Control	11.4	-12.53396	35.333957	0.4744032
Lowjump-Highjump	-26.2	-50.13396	-2.266043	0.0297843

- Again conclude that bone density for highjump group significantly higher than for other two groups.

# Why Tukey's procedure better than all t-tests

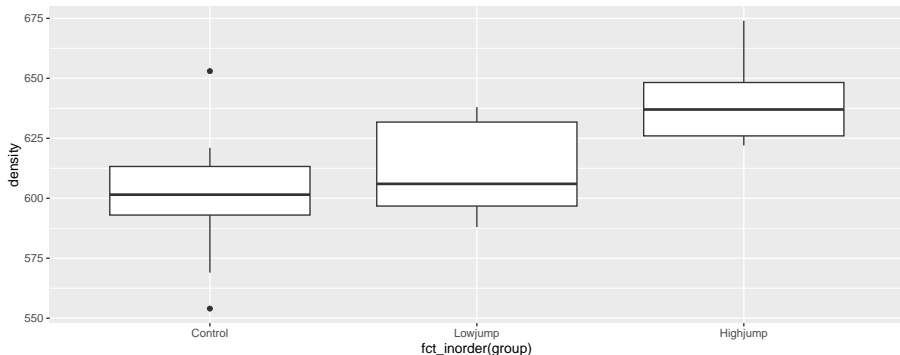
Look at P-values for the two tests:

Comparison	Tukey	t-tests
-----		
Highjump-Control	0.0016	0.0021
Lowjump-Control	0.4744	0.2977
Lowjump-Highjump	0.0298	0.0045

- Tukey P-values (mostly) higher.
- Proper adjustment for doing three t-tests at once, not just one in isolation.

# Checking assumptions

```
ggplot(rats,aes(y = density, x = fct_inorder(group)))+  
  geom_boxplot()
```

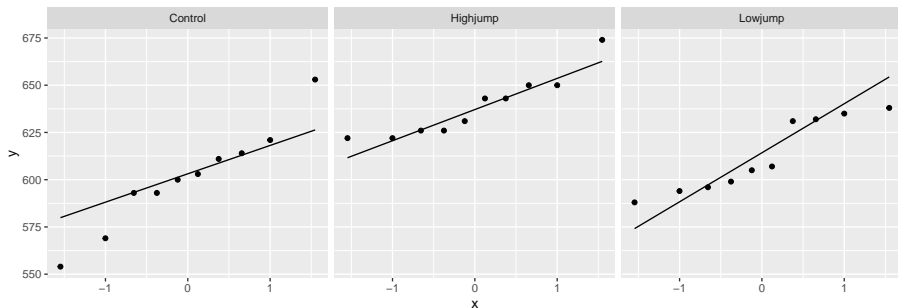


## Assumptions:

- Normally distributed data within each group
- with equal group SDs.

# Normal quantile plots by group

```
ggplot(rats, aes(sample = density)) + stat_qq() +  
  stat_qq_line() + facet_wrap( ~ group)
```



# The assumptions

- Normally-distributed data within each group
- Equal group SDs.
- These are shaky here because:
  - ▶ control group has outliers
  - ▶ highjump group appears to have less spread than others.
- Possible remedies (in general):
  - ▶ Transformation of response (usually works best when SD increases with mean)
  - ▶ If normality OK but equal spreads not, can use Welch ANOVA. (Regular ANOVA like pooled t-test; Welch ANOVA like Welch-Satterthwaite t-test.)
  - ▶ Can also use Mood's Median Test (see over). This works for any number of groups.

## Mood's median for multiple groups

- Find median of all bone densities, regardless of group
- Count up how many observations in each group above or below overall median
- Test association between group and above/below
- Mood's `median_test` (over).

## Mood's median test here

```
median_test(rats, density, group)
```

```
$grand_median
```

```
[1] 621.5
```

```
$table
```

	above	
group	above	below
Control	1	9
Highjump	10	0
Lowjump	4	6

```
$test
```

	what	value
1	statistic	1.680000e+01
2	df	2.000000e+00
3	P-value	2.248673e-04



## Comments

- No doubt that medians differ between groups (not all same).
- This test is equivalent of  $F$ -test, not of Tukey.
- To determine which groups differ from which, can compare all possible pairs of groups via (2-sample) Mood's median tests, then adjust P-values by multiplying by number of 2-sample Mood tests done (Bonferroni):

```
pairwise_median_test(rats, density, group)
```

```
# A tibble: 3 x 4
```

	g1	g2	p_value	adj_p_value
	<chr>	<chr>	<dbl>	<dbl>
1	Control	Highjump	0.000148	0.000443
2	Control	Lowjump	0.371	1
3	Highjump	Lowjump	0.371	1

- Now, lowjump-highjump difference no longer significant.

## Welch ANOVA

- For these data, Mood's median test probably best because we doubt both normality and equal spreads.
- When normality OK but spreads differ, Welch ANOVA way to go.
- Welch ANOVA done by `oneway.test` as shown (for illustration):

```
oneway.test(density~group, data=rats)
```

One-way analysis of means (not assuming equal variances)

data: density and group

F = 8.8164, num df = 2.000, denom df = 17.405, p-value = 0.002

- P-value very similar, as expected.
- Appropriate Tukey-equivalent here called Games-Howell.

# Games-Howell

- Lives in package PMCMRplus. Install first.

```
# gamesHowellTest(density ~ group, data = rats)
gamesHowellTest(density ~ factor(group), data = rats)
```

	Control	Highjump
Highjump	0.0056	-
Lowjump	0.5417	0.0120

Careful: explanatory must be factor (so commented-out line does not work).

# Deciding which test to do

For two or more samples:

