Lecture 19: Multiple Independent Populations

STAT 324

Ralph Trane
University of Wisconsin–Madison

Spring 2020

WISCONSIN
UNIVERSITY OF WISCONSIN–MADISON

Multiple Independent Populations

Four new formulations of rat poison are being tested, call them 1, 2, 3, and 4. All of the poisons work by thinning the blood, so the response of interest is the time it takes for the blood to coagulate. A longer blood coagulation time indicates a more effective poison. Twenty-four rats were randomly selected, and then randomized to the four poisons. They were fed the poison, and then after a specified length of time, their blood was drawn and the time to blood coagulation was measured.

```
library(tidyverse); theme_set(theme_bw()

rat_poison <- tibble(treatment = rep(1:4
                     coagulation_time =

DT::datatable(rat_poison,
          options = list(paging = FA
                         scrollY = "
                         dom = "t"),
          rownames = FALSE)
```

treatment	coagulation_time
1	62
1	60
1	63
1	59
2	63
2	67
2	71

Multiple Independent Populations

Fundamental research question: are there any differences between the four different kinds of poison?

We can formulate this as a hypothesis test in terms of means:
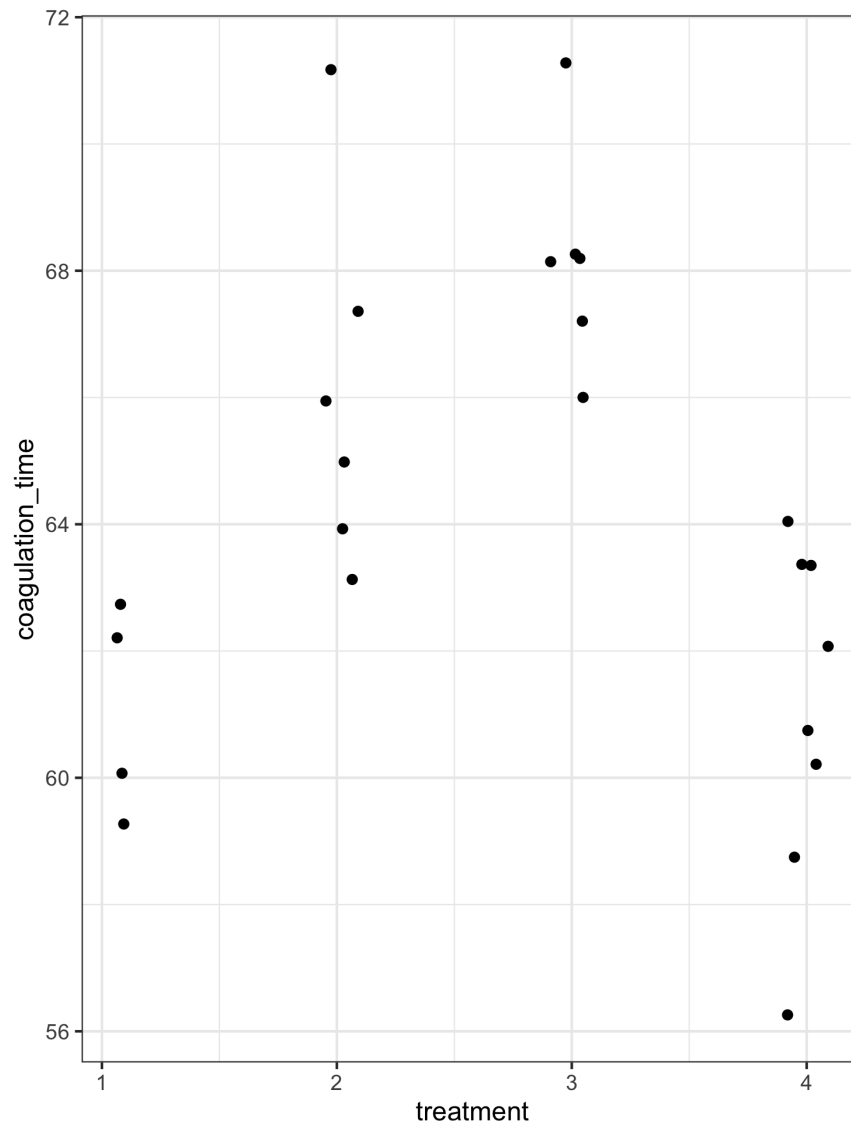
$H_0 : \mu_1 = \mu_2 = \mu_3 = \mu_4$ vs. H_A : at least one mean differs from one other mean.

- alternative is not "all means are different", but rather "not all are the same". Very different statements!
- rejecting the null would tell us nothing about which mean(s) is(are) different, or in which direction any potential difference is
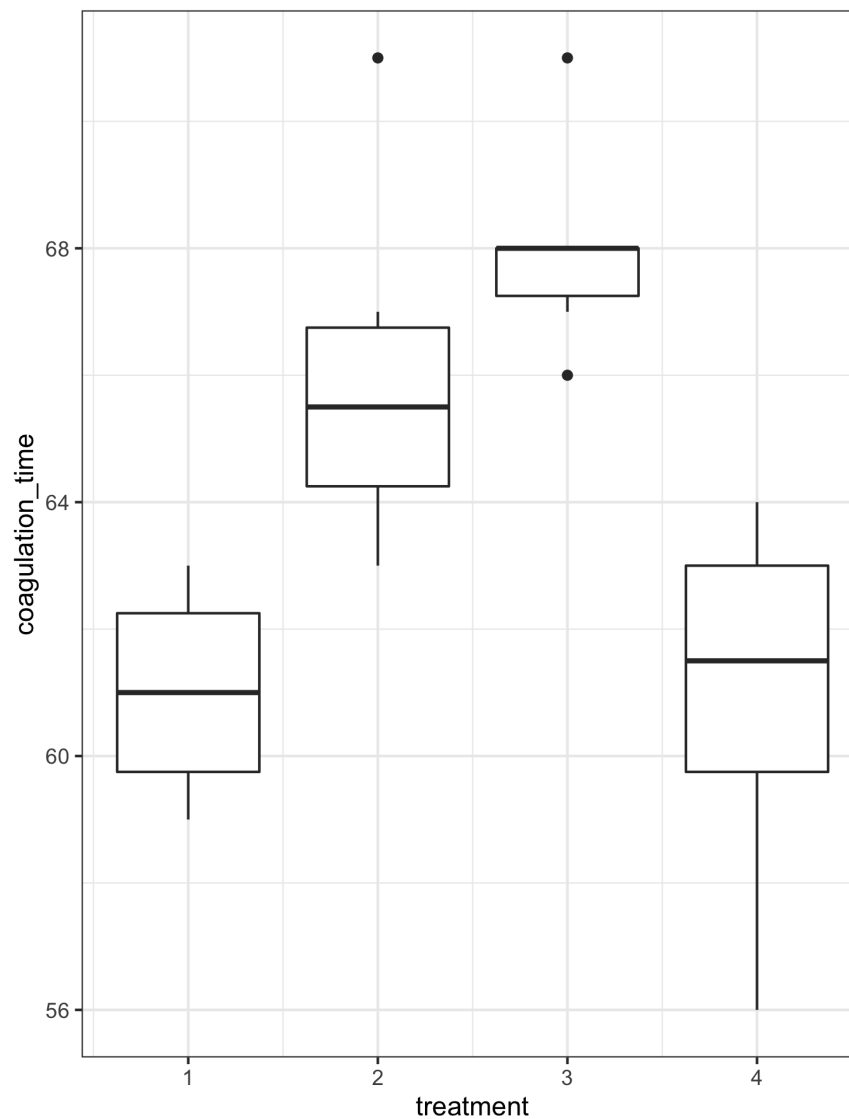
Multiple Independent Populations

```
ggplot(data = rat_poison,
       aes(x = treatment,
           y = coagulation_time)) +
  geom_jitter(width = 0.1)
```

```
ggplot(data = rat_poison,
        aes(x = treatment,
            y = coagulation_time)) +
  geom_boxplot(aes(group = treatment))
```

Most naive approach: use multiple t-tests.

A	B	Mean A	Mean B	Estimated Difference	95% LL	95% UL	p_value
1	2	61	66	-5	-8.395	-1.605	0.009
1	3	61	68	-7	-9.774	-4.226	0.001
1	4	61	61	0	-2.968	2.968	1
2	3	66	68	-2	-5.086	1.086	0.174
2	4	66	61	5	1.72	8.28	0.007
3	4	68	61	7	4.488	9.512	0

Here, we would reject that $\mu_A = \mu_B$, which would lead us to reject H_0.

Is this a good test? We used $\alpha = 0.05$, so we would like
$P(\text{Type I}) = P(\text{reject } H_0 | H_0 \text{ true}) = 0.05$. Let's investigate through simulations!

```
library(distributions3)
X <- Normal()

set.seed(120953)
simulations <- tibble(i = 1:2000) %>%
  mutate(A = map(i, ~random(X, 20)),
         B = map(i, ~random(X, 20)),
         C = map(i, ~random(X, 20)),
         D = map(i, ~random(X, 20)),
         test_AB = map2_dbl(A, B, ~t.test(x = .x, y = .y)$p.value),
         test_AC = map2_dbl(A, C, ~t.test(x = .x, y = .y)$p.value),
         test_AD = map2_dbl(A, D, ~t.test(x = .x, y = .y)$p.value),
         test_BC = map2_dbl(B, C, ~t.test(x = .x, y = .y)$p.value),
         test_BD = map2_dbl(B, D, ~t.test(x = .x, y = .y)$p.value),
         test_CD = map2_dbl(C, D, ~t.test(x = .x, y = .y)$p.value),
         reject = pmap_lgl(.l = list(test_AB, test_AC, test_AD, test_BC, test_BD, tes
                           .f = ~any(c(..1, ..2, ..3, ..4, ..5, ..6) < 0.05)))
```
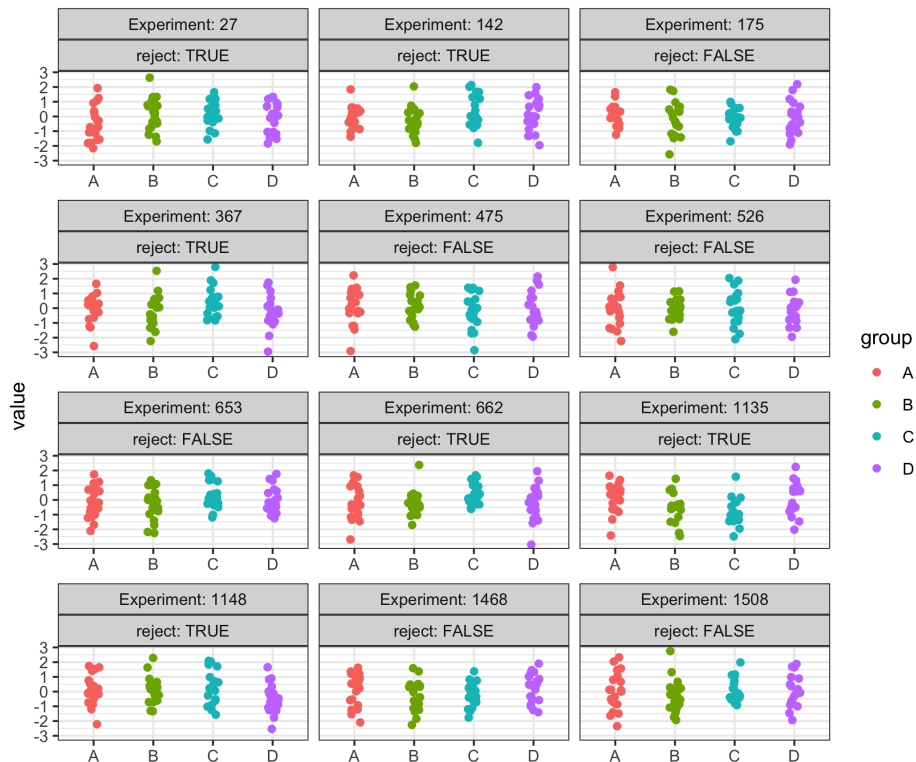
12 of the simulated data sets.

Question: what proportion of experiments were rejected? If this is a good test, it should be close to 5.

```
simulations %>%
  count(reject) %>%
  mutate(prop = n/sum(n)) %>%
  DT::datatable(options = list(dom = "t"
                rownames = FALSE)
```

reject	n	prop
false	1582	0.791
true	418	0.209

This is no surprise. In the following, I'll use this notation:

$$H_0^1 : \mu_A = \mu_B$$
$$H_0^2 : \mu_A = \mu_C$$
$$H_0^3 : \mu_A = \mu_D$$
$$H_0^4 : \mu_B = \mu_C$$
$$H_0^5 : \mu_B = \mu_C$$
$$H_0^6 : \mu_C = \mu_D$$

We are really interested in testing $H_0 : \mu_A = \mu_B = \mu_C = \mu_D$. Remember, the data is simulated such that $\mu_A = \mu_B = \mu_C = \mu_D = 0$. So, what is the probability of rejecting H_0, when we know it is true, if we use the multiple t-tests approach? Easier to calculate the probability of NOT rejecting H_0.

Let's pretend all the tests are independent:

$$P(\text{no type I error} \mid H_0 \text{ true}) = P(\text{not rejecting } H_0^1 \text{ AND not rejecting } H_0^2$$
$$\ldots \text{AND not rejecting } H_0^6 \mid H_0 \text{ true})$$
$$(\text{because independent}) = P(\text{not rejecting } H_0^1 | H_0 \text{ true}) \cdots P(\text{not rejecting } H_0^6 | H_0 \text{ true})$$
$$= 0.95 \cdot 0.95 \cdots 0.95 = 0.95^6 = 0.73509$$

So, what's the probability of making *at least one* type I error?

$$P(\text{at least one type I error} \mid H_0 \text{ true}) = 1 - P(\text{no type I error} \mid H_0 \text{ true})$$
$$= 1 - 0.73509 = 0.26491.$$

Multiple Independent Populations

I.e., pairwise tests is not a good idea. But we still want to test for a difference between means. How would we do that?

Since we are considering means, it is natural to look at group averages. (Same argument as for t-tests.)

Recall when we first discussed hypothesis tests: we argued that it wasn't enough to compare the average to the hypothesized value. We had to take the variation of the data into account. Similar situation here: not enough simply to compare the averages.
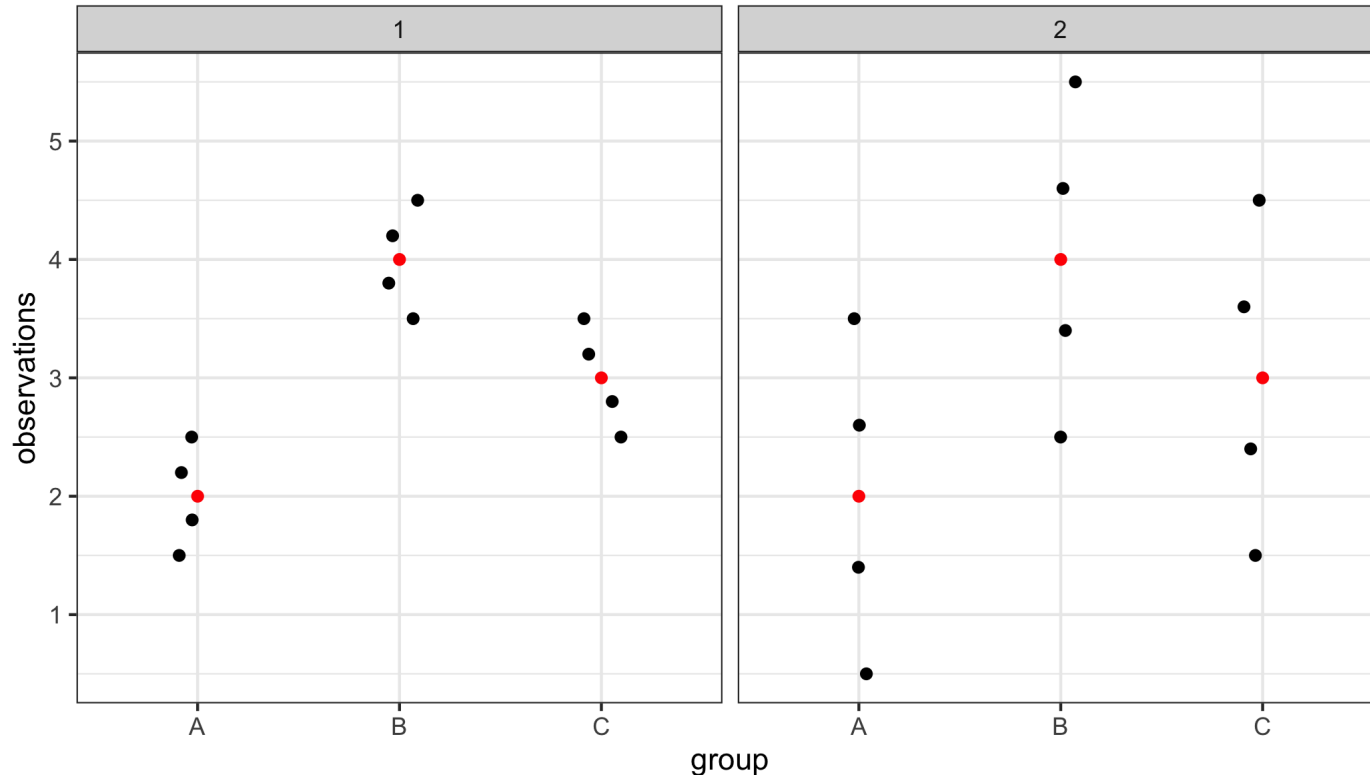
When moving to two sample tests, we sort of compared the *variation* between groups ($\bar{X} - \bar{Y}$) to a measure of variation within groups ($\mathrm{Var}(\bar{X} - \bar{Y}) = \mathrm{Var}(\bar{X}) + \mathrm{Var}(\bar{Y})$).

This idea of variation *between* groups compared to variation *within* groups is the foundation of Analysis of Variance (ANOVA).

Which of the following figures seems to be most indicative of a differences between means?

Although the groups have the same averages, data set 1 seems to provide more evidence suggesting differing means than data set 2.

Basically, we compare the variation *between* groups to the variation *within* groups. The between groups variation can be thought of as the SD of group means, while the within groups variation pools together the variation in each of the groups - it is actually an exact generalization of the pooled variance we already know from the two-sample t-test with equal variance.

This is all ANOVA is - comparing variation *between* groups to variation *within* groups. If variation between groups is large compared to variation within groups, then we reject the null of no difference, and vice versa.
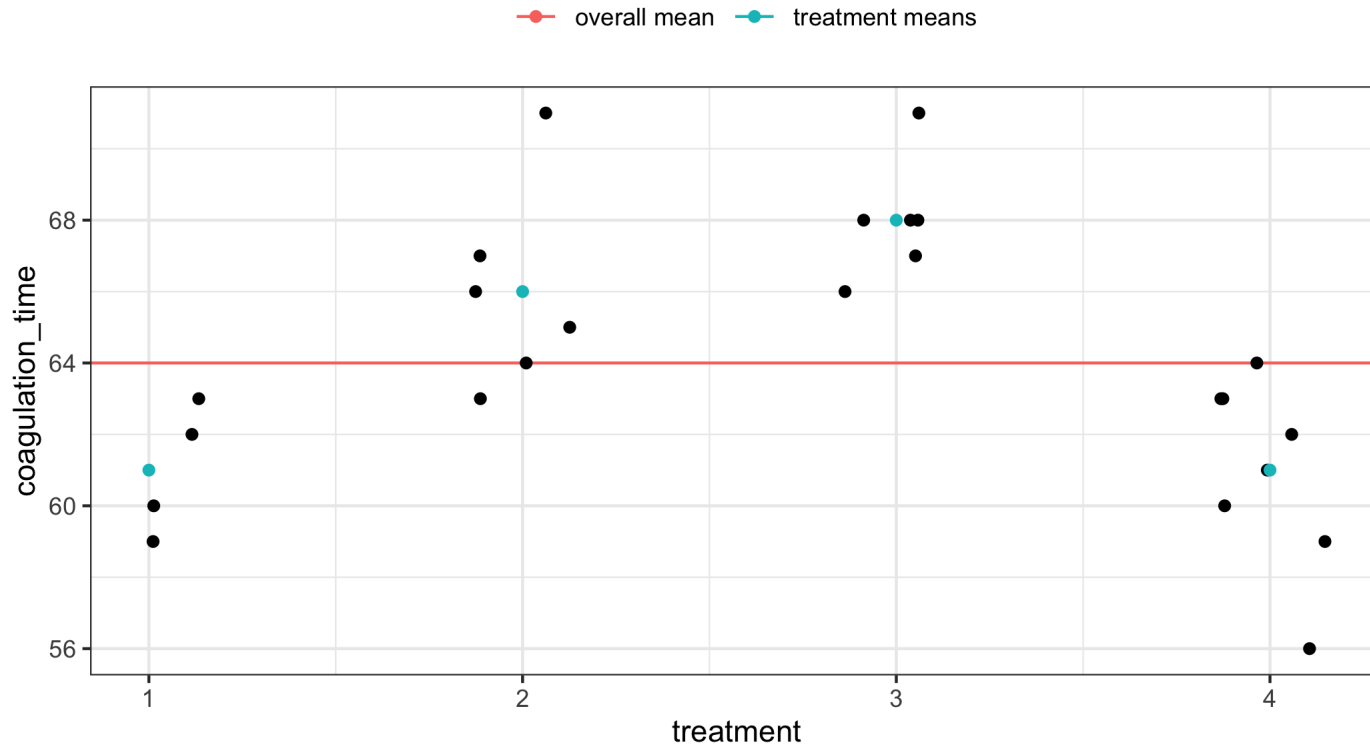
To build the test statistic we will use to formally test the null hypothesis H_0 : all means are equal, we need to do a tiny bit of math...
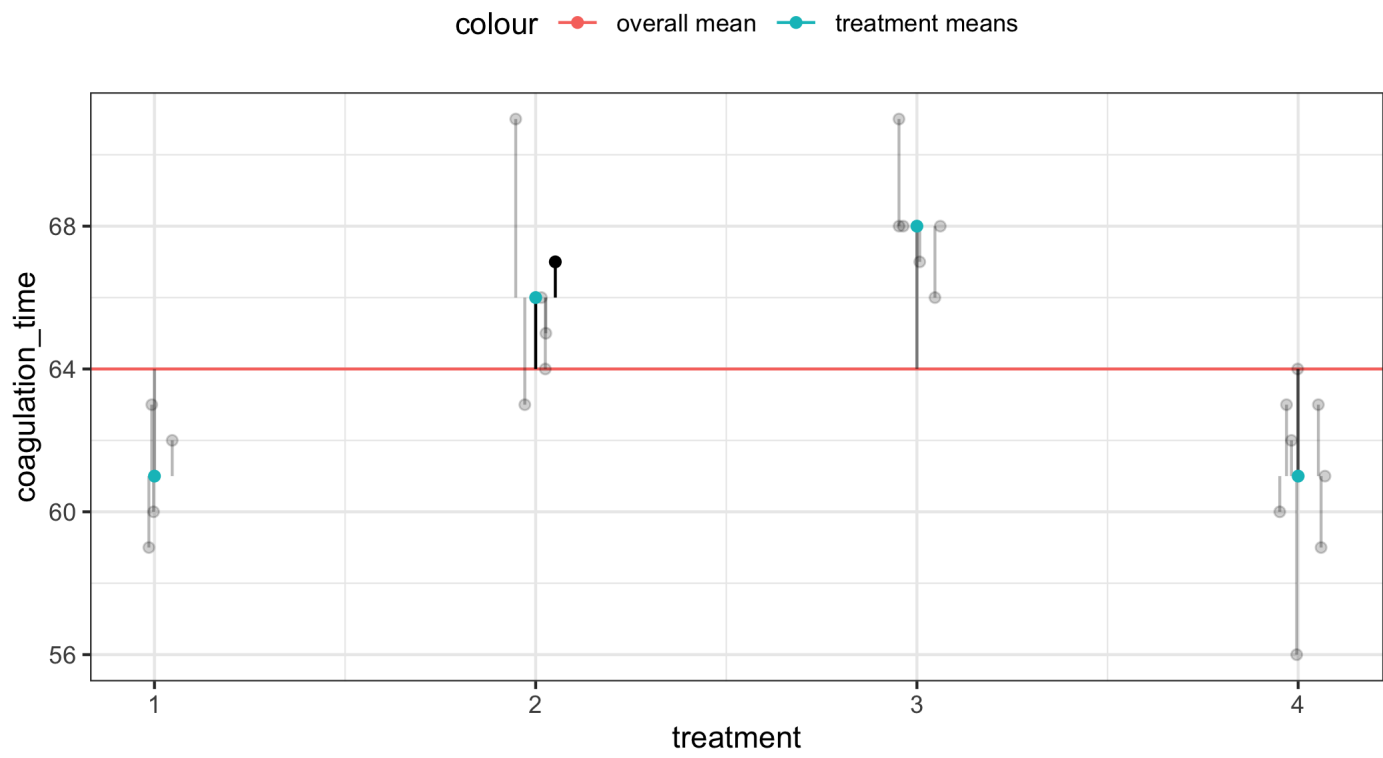
The first building block is the following realization:

$$\text{observation} = \text{overall mean} + \text{group mean - overall mean} + \text{observation - group mean}$$

For example:

Some (very heavy) notation:

- t = number of groups (here, 4).
- i = index of group (here, $i = 1, 2, 3, 4$).
- n_i = number of observations in group i (here, $n_1 = 4, n_2 = 6, n_3 = 6, n_4 = 8$).
- N = total number of observations (i.e. $N = \sum_{i=1}^{t} n_i$. Here, $N = 4 + 6 + 6 + 8 = 24$).
- y_{ij} = observation j from group i (here, $y_{11} = 62, y_{12} = 60$, etc.).
- $\bar{y}_{i.}$ = average of observations in group i (i.e. $\bar{y}_{i.} = \frac{1}{n_i} \sum_{j=1}^{n_i} y_{ij}$. Here, $\bar{y}_{1.} = 61, \bar{y}_{2.} = 66, \bar{y}_{3.} = 68, \bar{y}_{4.} = 61$).
- $\bar{y}_{..}$ = overall (or grand) mean (i.e. $\bar{y}_{..} = \frac{1}{N} \sum_{i=1}^{t} \sum_{j=1}^{n_i} y_{ij}$. Here, $\bar{y}_{..} = 64$).

So, the observation we just made can be written as

$$y_{ij} = \bar{y}_{..} + (\bar{y}_{i.} - \bar{y}_{..}) + (y_{ij} - \bar{y}_{i.})$$

(Side note: nothing magical here. If you actually do the math above, you'll see that everything on the right hand side cancels out except y_{ij}, so it really just says $y_{ij} = y_{ij}$.)

From

$$y_{ij} = \bar{y}_{..} + (\bar{y}_{i.} - \bar{y}_{..}) + (y_{ij} - \bar{y}_{i.})$$

it only takes a small step to get

$$y_{ij} - \bar{y}_{..} = (\bar{y}_{i.} - \bar{y}_{..}) + (y_{ij} - \bar{y}_{i.})$$

(Pause: we are slowly getting somewhere now. Look at the right hand side: the first parenthesis is deviation from group mean to overall mean. If there's big variation between groups, this will be large. The second parenthesis is deviation from observation to group mean. If there's big variation within groups, this will be large.)

Since the above holds for any observation y_{ij}, it also holds if you sum up them all up:

$$\sum_{i=1}^{t}\sum_{j=1}^{n_i}(y_{ij} - \bar{y}_{..}) = \sum_{i=1}^{t}\sum_{j=1}^{n_i}(\bar{y}_{i.} - \bar{y}_{..}) + \sum_{i=1}^{t}\sum_{j=1}^{n_i}(y_{ij} - \bar{y}_{i.})$$

(Pause: again, no magic here. If $a = b$ and $c = d$, then $a + c = b + d$.)

Now we are adding all the variations between group means and the overall mean (first parenthesis on the right) to all the variations within groups (second parenthesis). Sadly, it turns out that this is not super interesting:

```
rat_poison %>%
  mutate(overall_mean = mean(coagulation
  DT::datatable(options = list(dom = "t"

rat_poison %>%
  mutate(overall_mean = mean(coagulation
  summarize(sum_of_diffs = sum(coagulati
```

treatment	coagulation_time	overall
1	62	
1	60	
1	63	
1	59	
2	63	

```
# A tibble: 1 x 1
  sum_of_diffs
         <dbl>
1            0
```

Now we are adding all the variations between group means and the overall mean (first parenthesis on the right) to all the variations within groups (second parenthesis). Sadly, it turns out that this is not super interesting:

```r
rat_poison %>%
  mutate(overall_mean = mean(coagulation_time)) %>%
  group_by(treatment, overall_mean) %>%
  summarize(group_mean = mean(coagulation_time)) %>%
  mutate(diffs = group_mean - overall_mean)
```

```
# A tibble: 4 x 4
# Groups:   treatment [4]
  treatment overall_mean group_mean diffs
      <int>        <dbl>      <dbl> <dbl>
1         1           64         61    -3
2         2           64         66     2
3         3           64         68     4
4         4           64         61    -3
```

Now we are adding all the variations between group means and the overall mean (first parenthesis on the right) to all the variations within groups (second parenthesis). Sadly, it turns out that this is not super interesting:

```r
rat_poison %>%
  group_by(treatment) %>%
  mutate(group_mean = mean(coagulation_t
  DT::datatable(options = list(dom = "t"

rat_poison %>%
  group_by(treatment) %>%
  mutate(group_mean = mean(coagulation_t
  summarize(sum_of_diffs = sum(coagulati
```

treatment	coagulation_time	group_
1	62	
1	60	
1	63	
1	59	
2	63	

```
# A tibble: 4 x 2
  treatment sum_of_diffs
      <int>        <dbl>
1         1            0
2         2            0
3         3            0
4         4            0
```

So everything sums to 0! Not very useful.

Fortunately, a tiny bit of mathematical magic happens if we sum the squared deviations. It turns out that the equality still holds:

$$\sum_{i=1}^{t}\sum_{j=1}^{n_i}(y_{ij} - \bar{y}_{..})^2 = \sum_{i=1}^{t}\sum_{j=1}^{n_i}(\bar{y}_{i.} - \bar{y}_{..})^2 + \sum_{i=1}^{t}\sum_{j=1}^{n_i}(y_{ij} - \bar{y}_{i.})^2$$

Now, this is only going to be 0 if all observations are exactly the same (let's just pretend that never happens, cause... it doesn't).

What this really is, is a decomposition of variability:

total variation = variation due to groups (treatment) + variation due to error (random noise)

More accurately, these are *Sum of Squares*, and are refered to as SSTot (Sum of Squares Total), SSTrt (Sum of Squares Treatment), and SSE (Sum of Squares Error).

Remember the pooled variance when we considered two-sample t-tests? Let's revisit this. Let's denote observations from group 1 as $y_{11}, y_{12}, \ldots, y_{n_1 1}$ and observations from group 2 as $y_{21}, y_{22}, \ldots, y_{n_2 2}$. We can then write the pooled variance in the following way:

$$
\begin{aligned}
s^2_{\text{pooled}} &= \frac{(n_1 - 1)s_1^2 + (n_2 - 1)s_2^2}{n_1 + n_2 - 2} \\
&= \frac{(n_1 - 1)\frac{1}{n_1-1}\sum_{j=1}^{n_1}(y_{1j} - \bar{y}_{1.})^2 + (n_2 - 1)\frac{1}{n_2-1}\sum_{j=1}^{n_2}(y_{2j} - \bar{y}_{2.})^2}{n_1 + n_2 - 2} \\
&= \frac{\sum_{i=1}^{2}\sum_{j=1}^{n_1}(y_{ij} - \bar{y}_{i.})^2}{n_1 + n_2 - 2} \\
&= \frac{SSE}{N - t}
\end{aligned}
$$

So, the pooled variance can be written in terms of the SSE! This can naturally be expanded to more than two groups.

For two populations, we said that the pooled variance is a good estimate for an overall variance σ^2 *if the populations in fact have the same variance*. Same is true for a general number of populations: if all populations have same variance, $SSE/(N - t)$ is a good estimate for the overall variance.

Back to ANOVA: we have seen how total variability (SSTot) can be decompositioned into variability due to different treatments (SSTrt), and variability due to random noise (SSE). But these Sum of Squares aren't *really* measures of variability. (What happens to SSTot if we increase the number of observations? What happens to SSTrt if we increase number of treatments? Both increase!)

Already saw that $SSE/(N-t)$ can be thought of as a pooled variance.

What is $\frac{SSTot}{N-1}$?

$$\frac{SSTot}{N-1} = \frac{\sum_{i=1}^{t}\sum_{j=1}^{n_j}(y_{ij} - \bar{y}_{..})^2}{N-1} = s^2$$

is simply the sample variance of *all* the data (each observation subtracted the overall mean, squared, summed, divided by size minus 1).

What about SSTrt? If we think of SSTrt as "variation between treatments" it would be natural to divide by number of treatment groups minus 1.

We call these ratios *Mean Squares*, and will denote them by MSTot, MSTrt, and MSE. The denominators are called *degrees of freedom*.

Pause for a summary:

Source	SS	df	MS
Treatment (between)	SSTrt	t-1	MSTrt=SSTrt/(t-1)
Error (within)	SSE	N-t	MSE=SSE/(N-t)
Total	SSTot	N-1	MSTot=SSTot/(N-1)

Things to remember: for SS and df, columns sum up. I.e. SSTot = SSTrt + SSE, and $df_{\text{Tot}} = df_{\text{Trt}} + df_{\text{E}}$.

We now have some reasonable measures for *between* and *within* treatment variations:

$$\text{MSTrt} = \frac{\sum_{i=1}^{t} \sum_{j=1}^{n_i} (\bar{y}_{i.} - \bar{y}_{..})^2}{t - 1}$$

$$\text{MSE} = \frac{\sum_{i=1}^{t} \sum_{j=1}^{n_i} (\bar{y}_{ij} - \bar{y}_{i.})^2}{N - t}$$

The point of all of this: is *between* large compared to *within*? A reasonable thing to look at: $F = \frac{\text{MSTrt}}{\text{MSE}}$. This will be our test statistic for testing the hypothesis H_0 : all means are equal against the alternative H_A : at least one mean differs from another.

So, in the spirit of general hypothesis tests, we need to determine what "more extreme" means in this scenario. Questions:

1. What values can F possibly take?

2. What does a large value of F tell us about the null?

3. What does a small value of F tell us about the null?

1. What values can F possibly take?
   - F is always positive.

1. What does a large value of F tell us about the null?

   - F large $\Rightarrow$
   - MSTrt >> MSE $\Rightarrow$
   - variations between groups >> variations within groups $\Rightarrow$
   - evidence against the null.

2. What does a small value of F tell us about the null?

   - F small $\Rightarrow$
   - MSTrt << MSE $\Rightarrow$
   - variations between groups << variations within groups $\Rightarrow$
   - evidence "for" the null.

I.e. "more extreme" will here **always** be "greater than".

So, p-value = $P(\text{more extreme than } F_{\text{obs}} \mid H_0 \text{ true}) = P(F > F_{\text{obs}} \mid H_0 \text{ true})$.

All we need now is the distribution of F *assuming* the null hypothesis is true.

Turns out, **if the data are normal** and **all groups have same variance**, then $F \sim F_{\text{df}_{\text{Trt}}, \text{df}_{\text{E}}}$ (F-distribution with numerator df of df_{Trt} and denominator df of df_{E}; in `distributions3`, this is `FisherF(df1, df2)`).
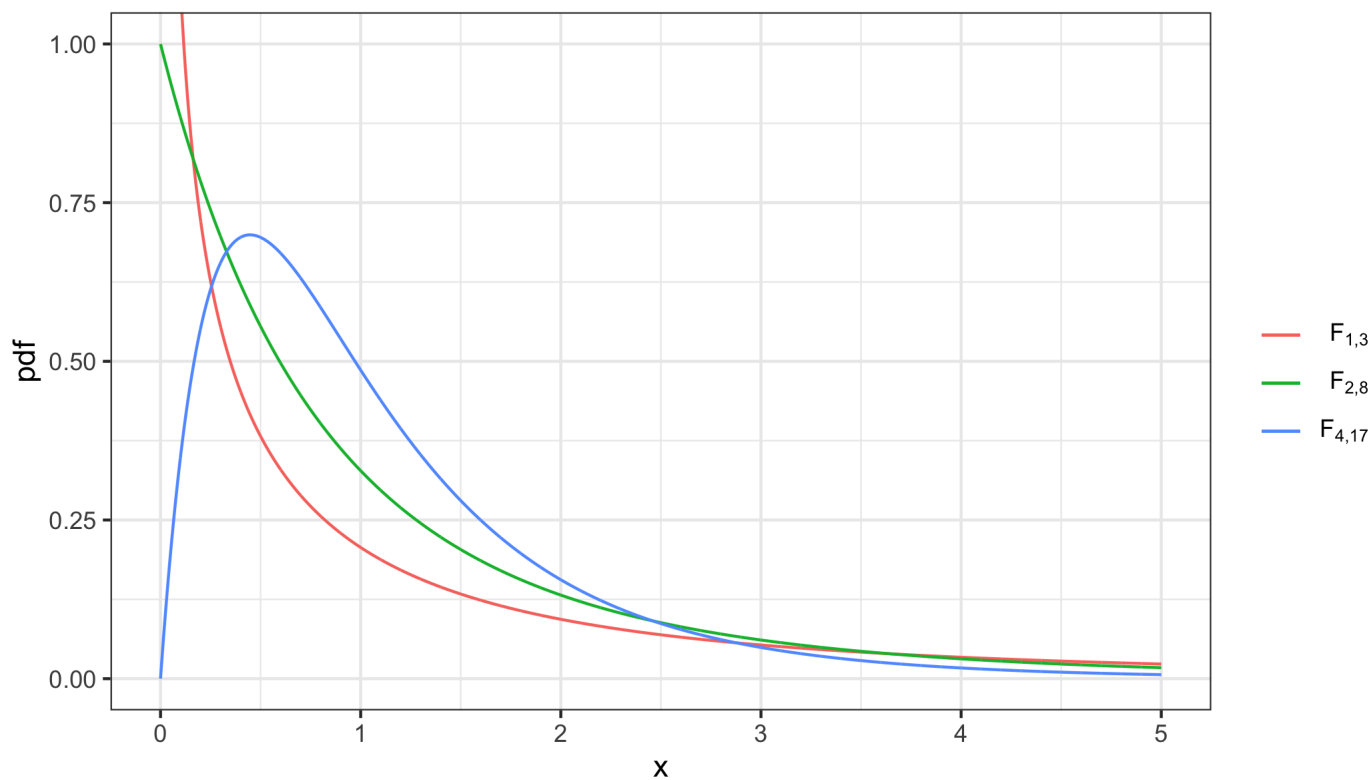
So how do we perform an Analysis of Variance? We fill out the ANOVA table, one step at a time:

Source	SS	df	MS	F_{obs}	p-value
Treatment (between)	SSTrt	t-1	MSTrt=SSTrt/(t-1)	MSTrt/MSE	$P(F > F_{\text{obs}}|H_0 \text{ true})$
Error (within)	SSE	N-t	MSE=SSE/(N-t)		
Total	SSTot	N-1	MSTot=SSTot/(N-1)		

SSTrt $= \sum_{i=1}^{t} \sum_{j=1}^{n_i} (\bar{y}_{i\cdot} - \bar{y}_{\cdot\cdot})^2$:

```
## overall mean
(overall_mean <- rat_poison %>%
  summarize(ybar_dotdot = mean(coagulation_time)))

# A tibble: 1 x 1
  ybar_dotdot
        <dbl>
```

SSTot $= \sum_{i=1}^{t} \sum_{j=1}^{n_i} (\bar{y}_{ij} - \bar{y}_{..})^2$:

```
(SSTot <- rat_poison %>%
  mutate(obs_minus_average = coagulation_time - mean(coagulation_time),
         squares = obs_minus_average^2) %>% print %>%
  summarize(SSTot = sum(squares)))
```

```
# A tibble: 24 x 4
   treatment coagulation_time obs_minus_average squares
       <int>            <dbl>             <dbl>   <dbl>
 1         1               62                -2       4
 2         1               60                -4      16
 3         1               63                -1       1
 4         1               59                -5      25
 5         2               63                -1       1
 6         2               67                 3       9
 7         2               71                 7      49
 8         2               64                 0       0
 9         2               65                 1       1
10         2               66                 2       4
# … with 14 more rows

# A tibble: 1 x 1
  SSTot
```

- Degrees of freedom:

$$df_{Trt} = t - 1 = 4 - 1 = 3$$

$$df_E = N - t = 24 - 4 = 20$$

$$df_{Total} = N - 1 = 23$$

Check: $df_{Total} = df_{Trt} + df_E$!

- Mean squares:

$$MSTrt = \frac{SSTrt}{df_{Trt}} = \frac{228}{3} = 76$$

$$MSE = \frac{SSE}{df_E} = \frac{112}{20} = 5.6$$

- Observed test statistic:

$$F_{obs} = \frac{MSTrt}{MSE} = \frac{76}{5.6} = 13.57$$

- p-value:

```
F_3_20 <- FisherF(3,20)

cdf(F_3_20, 13.57)

[1] 4.66169e-05
```

Filled out ANOVA table:

Source	SS	df	MS	F_{obs}	p-value
Treatment (between)	228	3	76	13.571	0.00004661690082416
Error (within)	112	20	5.6		
Total	340	23	14.783		

As you might have guessed, we can do this easily using R. One important note: you HAVE to make sure that the treatment variable is NOT a numeric variable. You can check if it is by simply printing the data (assuming you used `read_csv` to import it):

```
rat_poison %>% head(n = 6)
```

```
# A tibble: 6 x 2
  treatment coagulation_time
      <int>            <dbl>
1         1               62
2         1               60
3         1               63
4         1               59
5         2               63
6         2               67
```

It's fairly easy to correct:

```
rat_poison <- rat_poison %>%
  mutate(treatment = as.character(treatm
```

We then use `aov` to run the ANOVA, and `summary` to get the ANOVA table:

```
ANOVA <- aov(data = rat_poison,
             coagulation_time ~ treatmen

summary(ANOVA)
```

```
            Df Sum Sq Mean Sq F value    F
treatment    3    228    76.0   13.57 4.6
Residuals   20    112     5.6
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '
```

Notice how the syntax is the same familiar syntax from `t.test` and `wilcox.test` (sweet!), and how the numbers match what we found "by hand".

If we use this technique for all of our simulated experiments, we see that the type I error rate is very close to spot on!

reject	n	prop	test
false	1884	0.942	ANOVA
true	116	0.058	ANOVA
false	1582	0.791	Multiple t-tests
true	418	0.209	Multiple t-tests

Some rather convoluted theory shows that this F-test is actually the most powerful test for testing H_0 : all means equal **IF** all assumptions are met. What are the assumptions?

- Each group should be normal

- Equal variance for all groups

We could evaluate group by group, but there's an easier way to do this: using *residuals*.

We can think of ANOVA as a *model*. It's a formula that aims at describing the data. (Note: the idea of models is very important in statistics, and will be revisited when talking about linear regression, our next and last topic.)

The ANOVA model is based on the decomposition we say earlier:

$$y_{ij} = \bar{y}_{..} + (\bar{y}_{i.} - \bar{y}_{..}) + (y_{ij} - \bar{y}_{i.})$$

We can view this as "observation = overall mean + (treatment deviation from overall mean) + (observation deviation from treatment mean)".

If we were to *predict* (i.e. provide a best guess for) a new observation in a group, we would use the group mean. We adopt this idea that each observation is the group mean plus an error (random noise). This error is what we call the *residual*, i.e. it is what is left over in the observation, once we have accounted for what our model tells us.

If we use our model to "predict" what the observations we have are, we call the results *fitted values*. So, the residuals are

$$\text{residual} = \text{observed} - \text{fitted}$$

We will use these to check our assumptions.

```
rat_poison %>%
  group_by(treatment) %>%
  mutate(fitted = mean(coagulation_time)) %>%
  ungroup() %>%
  mutate(residuals = coagulation_time - fitted) %>%
  DT::datatable(options = list(dom = "t", scrollY = "30vh"),
                rownames = FALSE)
```
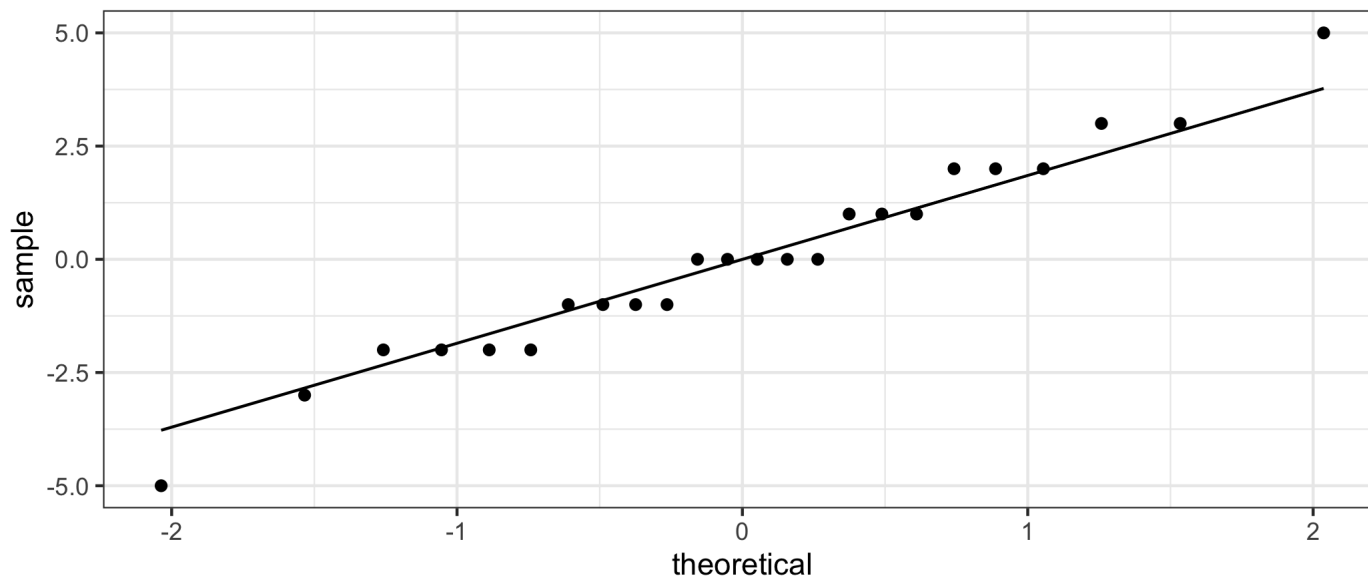
treatment	coagulation_time	fitted	residuals
1	62	61	1
1	60	61	-1
1	63	61	2
1	59	61	-2
2	63	66	-3

Looking at a QQ-plot, the residuals look pretty normal.

```
rat_poison %>%
  group_by(treatment) %>%
  mutate(fitted = mean(coagulation_time)) %>%
  ungroup() %>%
  mutate(residuals = coagulation_time - fitted) %>%
  ggplot(aes(sample = residuals)) +
    geom_qq() +
    geom_qq_line()
```
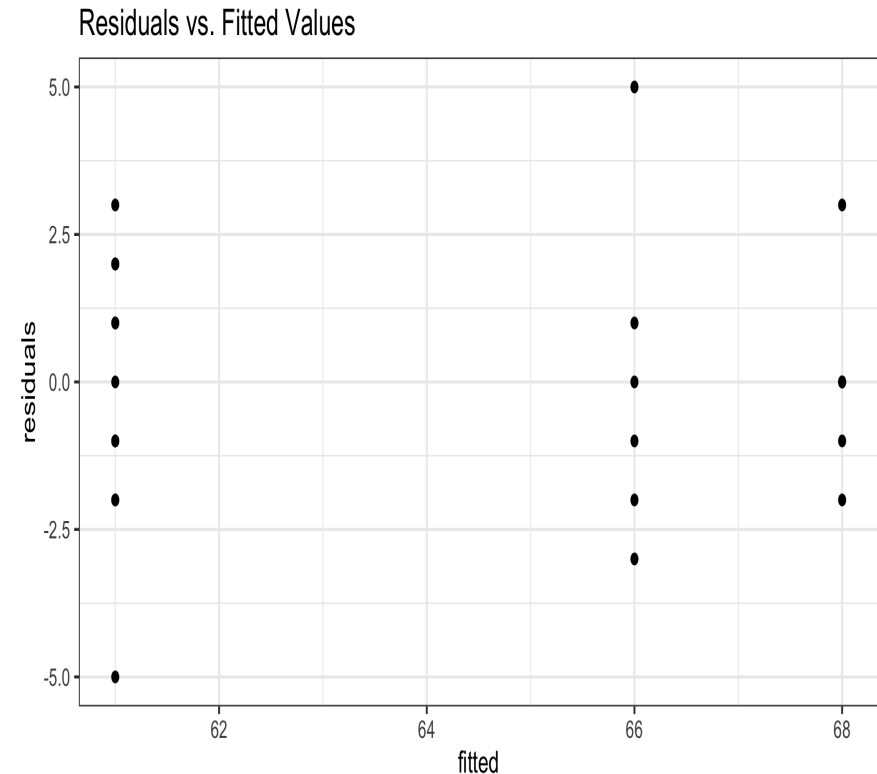
To check for "equal variance" graphically, we usually look at "residuals vs. fitted":

```
rat_poison %>%
  group_by(treatment) %>%
  mutate(fitted = mean(coagulation_time)
  ungroup() %>%
  mutate(residuals = coagulation_time -
  ggplot(aes(x = fitted, y = residuals))
    geom_point() +
    labs(title = "Residuals vs. Fitted V
```
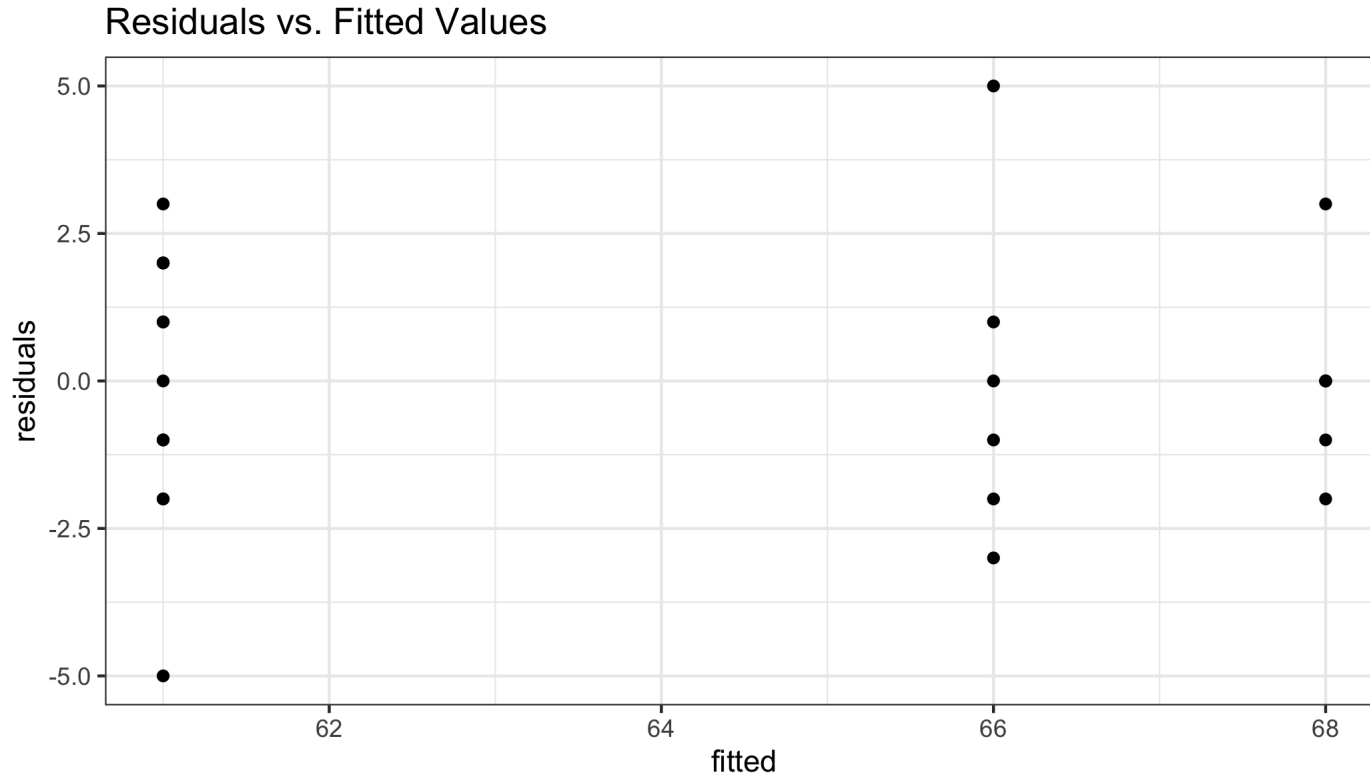

Residuals vs. Fitted Values

We look at residuals vs. fitted because often variance increases with means: the standard deviation of the weight of 10 elephants will tend to be larger than SD of the weight of 10 hamsters. A couple of classic problems:

Back to our data:

Residuals vs. Fitted Values
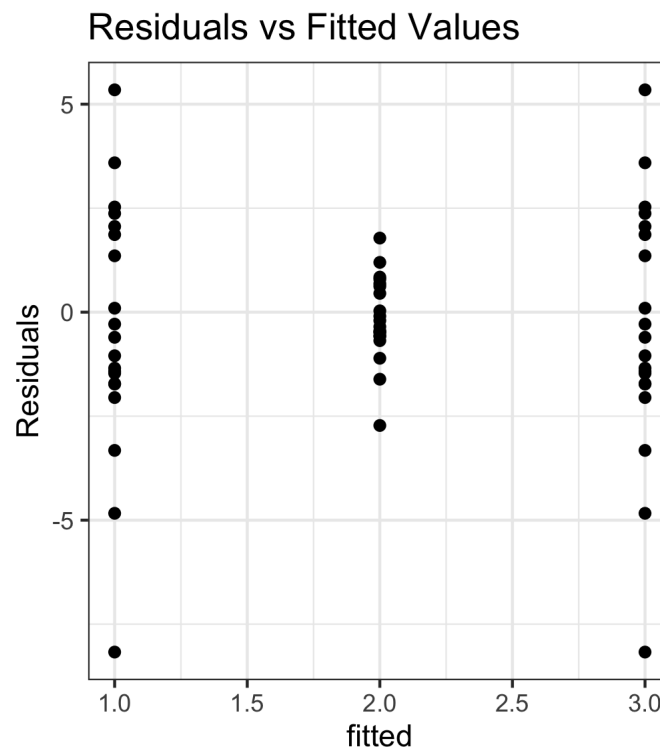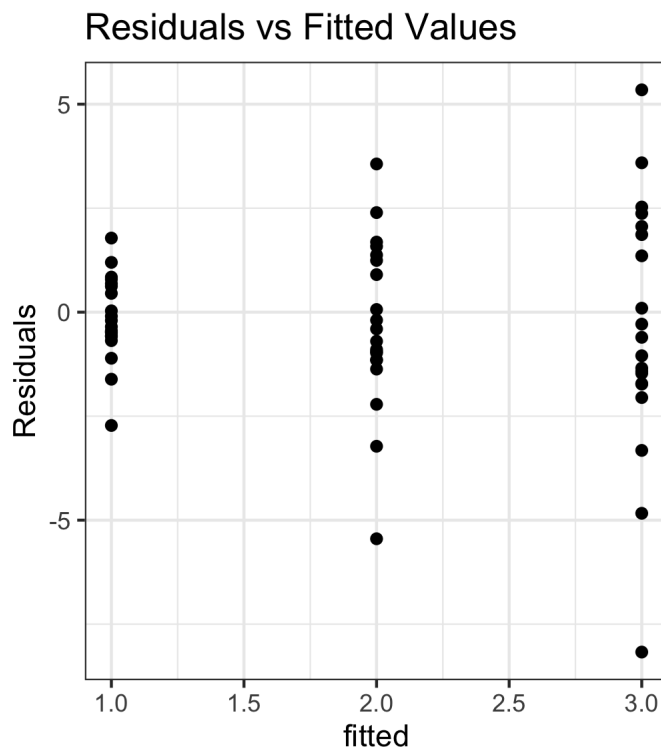
Although the last group might have slightly less spread, it doesn't seem to bad.

Numerically, we use our old rule of thumb, and check that "largest SD"/"smallest SD" < 2:

```
rat_poison %>%
  group_by(treatment) %>%
  summarize(s = sd(coagulation_time)) %>
  summarize(ratio = max(s)/min(s))
```

```
# A tibble: 4 x 2
  treatment      s
  <chr>      <dbl>
1 1           1.83
2 2           2.83
3 3           1.67
4 4           2.62

# A tibble: 1 x 1
  ratio
  <dbl>
1  1.69
```

Looks good!

So using ANOVA to test $H_0 : \mu_1 = \mu_2 = \mu_3 = \mu_4$ against H_A : at least one is different seems valid. Since our p-value was super small (0.000047), we reject the null hypothesis.

The next natural question is then: which treatments are different?!

The ANOVA showed us that there probably is a difference somewhere in the data. More often than not, this is only somewhat interesting. We would much rather know where the differences are.