ANÁLISE DOS DADOS COM R

Analysis of Variance

James R. Hunter, PhD

Retrovirologia, EPM, UNIFESP

ANALYSIS OF VARIANCE

- One-Way ANOVA - Regression's Kissing Cousin

ANALYSIS OF VARIANCE - ANOVA

- Tool developed in 1920's by Sir Ronald Fisher
- Clinical trials
  - Go-to method of analyzing clinical trial data

CONCEPTS

- Focus on differences in group means (not variances)
- And how a dependent variable can depend on them
  - As opposed to regression
- Balanced vs. Unbalanced Models
  - Same number of cases per group (balanced)
  - Different number of cases per group (unbalanced)
- Multiple ANOVA model schemes
  - One-Way ANOVA
  - Two-Way ANOVA
  - Latin Squares and More…

CONCEPTS - 2

- Multiple Comparisons
  - Post-hoc tests to isolate differences
  - Correction for aggressive estimation of p-value
  - Bonferroni, Benjamini-Hochberg(BH), Tukey's HSD
- Use some functions `fromcar` package

IMAGINARY CLINICAL TRIAL OF NEW DRUG FOR COVID-19 TREATMENT

- New drug - “Nontussis”
  - Directly attacks SARS-CoV-2 virus in respiratory system
- Trial will compare Nontussis to an imaginary version of HCQ and a placebo
  - HCQ - Hydroxychloroquine
- Trial will be double-blinded
- Patients will be divided between
  - Those on supplemental O_2 (50%)
  - Those with no supplemental O_2 (50%)
- Outcome measure is a scaled difference (between 1 and 10) of the difference between viral load at enrollment and viral load after 5 days of treatment

LOAD AND VIEW THE DATA

```r
clin_trial <- read_rds(here("covid_clin_trial.rds"))
glimpse(clin_trial)
```

```
Rows: 66
Columns: 4
$ pac_id <chr> "1", "2", "3", "4", "5", "6", "7", "8", "9", "10", "11", "12",
…
$ drug   <chr> "placebo", "placebo", "placebo", "placebo", "placebo",
"placebo…
$ o2     <chr> "supp", "supp", "supp", "supp", "supp", "supp", "supp",
"supp",…
$ score  <dbl> 1.8, 3.1, 7.3, 4.7, 7.7, 6.4, 2.6, 1.4, 4.7, 5.9, 4.0, 8.5,
3.2…
```

TABLE OF THE DATA

```
clin_trial %>%
  group_by(drug, o2) %>%
  summarise(avg_score = mean(score)) %>%
  knitr::kable()
```

drug	o2	avg_score
hcq	no_supp	4.690909
hcq	supp	4.400000
nontussis	no_supp	6.236364
nontussis	supp	5.981818
placebo	no_supp	4.309091
placebo	supp	4.509091

DESCRIPTIVE STATISTICS OF DATA – SUPPLEMENTAL OXYGEN

```
clin_trial %>%
  filter(o2 == "supp") %>%
  group_by(drug, o2) %>%
descr(stats = c("mean", "sd", "min", "med", "max", "iqr", "cv"),
      transpose = TRUE)
```

SUPPLEMENTAL OXYGEN

```
Descriptive Statistics
clin_trial$score
Group: drug = hcq, o2 = supp
N: 11

             Mean   Std.Dev    Min   Median    Max    IQR     CV
----------- ------ --------- ------ -------- ------ ------ ------
      score   4.40      2.53   1.30     4.70   8.50   4.65   0.58

Group: drug = nontussis, o2 = supp
N: 11

             Mean   Std.Dev    Min   Median    Max    IQR     CV
----------- ------ --------- ------ -------- ------ ------ ------
      score   5.98      2.29   2.80     5.90   9.20   3.70   0.38

Group: drug = placebo, o2 = supp
N: 11

             Mean   Std.Dev    Min   Median    Max    IQR     CV
```

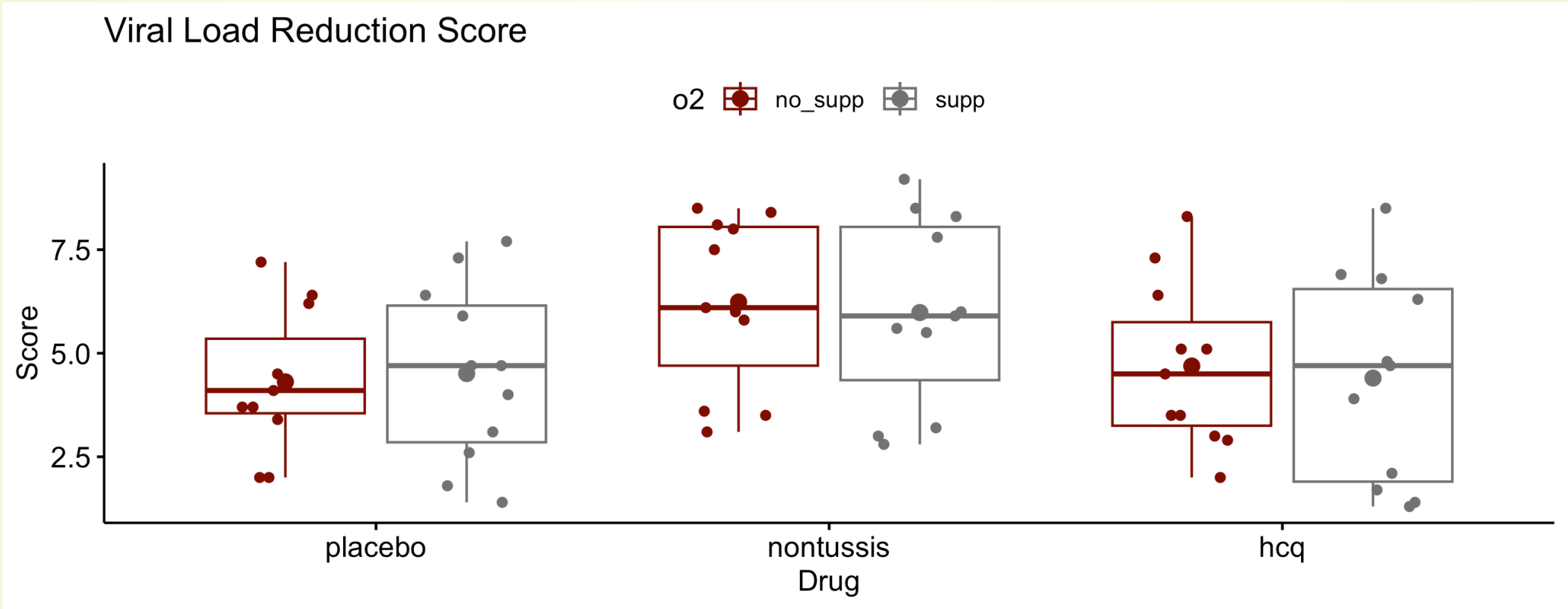
DESCRIPTIVE STATISTICS OF DATA – NO SUPPLEMENTAL OXYGEN

```
clin_trial %>%
  filter(o2 == "no_supp") %>%
  group_by(drug, o2) %>%
descr(stats = c("mean", "sd", "min", "med", "max", "iqr", "cv"),
      transpose = TRUE)
```

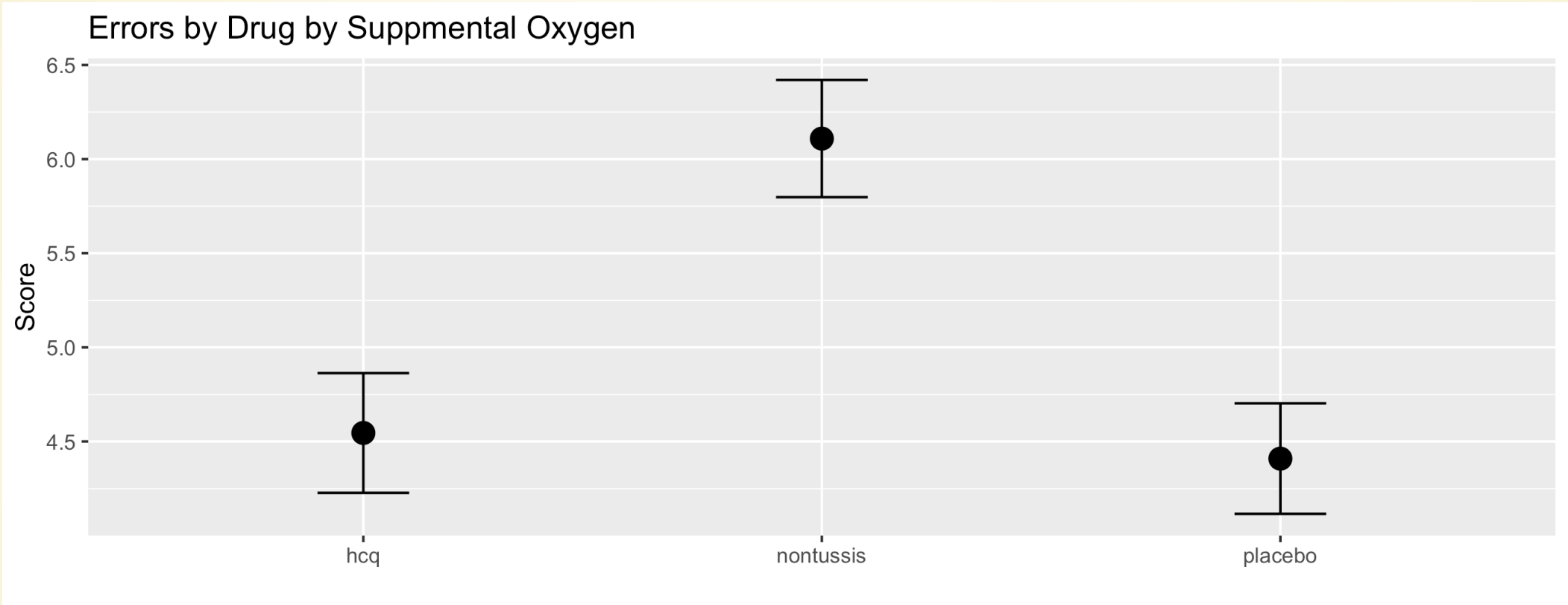
NO SUPPLEMENTAL OXYGEN

```
Descriptive Statistics
clin_trial$score
Group: drug = hcq, o2 = no_supp
N: 11

             Mean   Std.Dev    Min   Median    Max    IQR     CV
----------- ------ --------- ------ -------- ------ ------ ------
      score   4.69      1.98   2.00     4.50   8.30   2.50   0.42

Group: drug = nontussis, o2 = no_supp
N: 11

             Mean   Std.Dev    Min   Median    Max    IQR     CV
----------- ------ --------- ------ -------- ------ ------ ------
      score   6.24      2.06   3.10     6.10   8.50   3.35   0.33

Group: drug = placebo, o2 = no_supp
N: 11

             Mean   Std.Dev    Min   Median    Max    IQR     CV
```

GRAPH OF DATA



GRAPH OF VIRAL LOAD REDUCTION SCORE ERRORS



VARIANCE OF Y (`score`)

- To understand how this works, need to review **variance** and **sum of squares**
- Variance is again based on deviation of individual point from a mean
  - Here, it is Y for each cell compared to the overall mean for Y ($\bar{Y}$)
  - $(Y_{ik} - \bar{Y})$ **Note subscripts**
- Subscripts (and cells) represent
  - Each **group** (k)
  - Each **member** within the group (`i`)
- Formula just assures you pick up all the cells of a cross table

$$Var(Y) = \frac{1}{N} \sum_{k=1}^{K} \sum_{i=1}^{N_k} (Y_{ik} - \bar{Y})^2$$

VARIANCE LOOKS LIKE A MEAN

- **Because it is a mean**
- Squared deviations divided by N
- Mean of the squared deviations

VARIANCES AND SUM OF SQUARES

- ANOVA works with Total Sum of Squares (SST)
- SS_{tot} is **variance** not adjusted by N

$$SS_{tot} = \sum_{k=1}^{K} \sum_{i=1}^{N_k} (Y_{ik} - \bar{Y})^2$$

- Divides SST into 2 components
  - SS_{betw} - Sum of squares between groups
  - $SS_{w/in}$ - Sum of squares within groups

SS_{betw} - BETWEEN GROUPS

- Difference between group means ($\bar{Y}_k$) and overall mean ($\bar{Y}$)

$$SS_{betw} = \sum_{k=1}^{K} \sum_{i=1}^{N_k} (\bar{Y}_k - \bar{Y})^2$$

$SS_{w/in}$ - WITHIN GROUPS

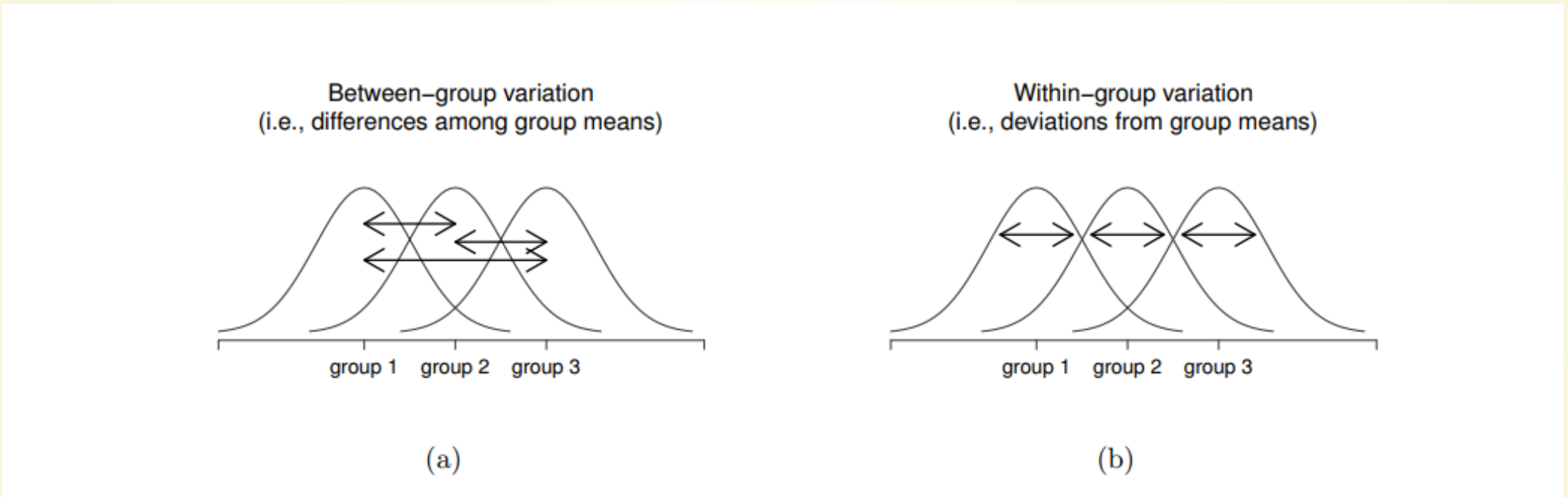
- Difference between the group means ($\bar{Y}_k$) and the individual values (Y_{ik})

$$SS_{w/in} = \sum_{k=1}^{K} \sum_{i=1}^{N_k} (Y_{ik} - \bar{Y}_k)^2$$

AS WITH REGRESSION ...

$$SS_{tot} = SS_{betw} + SS_{w/in}$$

GRAPHIC DEPICTION OF VARIABILITY WITHIN AND BETWEEN GROUPS



source: Navarro, lsr v.6, p.430.

CONSTRUCT NULL HYPOTHESIS FOR THE ANOVA

- Null hypothesis: there will be no difference among the drugs in the reduction of viral load

$$H_0 : \mu_{placebo} = \mu_{hcq} = \mu_{nt}$$

- Alternative hypothesis: at least one of the means of the drugs will be different than the others

$$H_1 : \mu_{placebo} \neq \mu_{hcq} \neq \mu_{nt}$$

TESTING THE HYPOTHESIS: THE F RATIO

- Sums of Squares represent variability: differences of values and group means from other means
- The **F ratio** is a ratio of variances
- Remember that variances have a **degree of freedom**
- Here we are working with 2 variances: $SS_{w/in}$ and SS_{betw}
  - We use the sum of squares and not the variances because the N's cancel out
- ∴ we have two different degrees of freedom
  - df_{betw} = K - 1 (number of groups less 1)
  - $df_{w/in}$ = N - K (number in sample less the number of groups)

CALCULATE MEAN SQUARES FROM SUM OF SQUARES

- Divide each type of sum of squares by its degrees of freedom

$$MS_{betw} = \frac{SS_{betw}}{df_{betw}}$$

$$MS_{w/in} = \frac{SS_{w/in}}{df_{w/in}}$$

F RATIO IS RATIO OF MEAN SQUARES

$$F = \frac{MS_{betw}}{MS_{w/in}}$$

- Bigger the F: relatively bigger the separation between group peaks and narrower the curves
- F test statistic has 2 parameters: its 2 different degrees of freedom
- F ratio must be larger than 1 to have any chance of rejecting H_0
  - Between groups variance needs to be greater than within group variance
  - If not, model contaminated by *noise*

RUNNING ANOVA IN R

- Basic command `aov()` with model specified with formula notation
  - As with regression
  - Assign to a model name
  - Use `summary()` to see full result

```
anova_mod1 <- aov(score~drug, data = clin_trial)
summary(anova_mod1)
```

```
            Df Sum Sq Mean Sq F value Pr(>F)
drug         2  39.26  19.630   4.504 0.0148 *
Residuals   63 274.55   4.358
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

SAME INFORMATION AS SHOWN ON REGRESSION OUTPUT

- `lm()` is the same operation as `aov()` behind the curtain

```
lm_mod1 <- lm(score~drug, data = clin_trial)
summary(lm_mod1)
```

```
Call:
lm(formula = score ~ drug, data = clin_trial)

Residuals:
    Min      1Q  Median      3Q     Max 
-3.3091 -1.6205 -0.1091  1.8818  3.9545 

Coefficients:
             Estimate Std. Error t value Pr(>|t|)    
(Intercept)    4.5455     0.4451  10.213 5.47e-15 ***
drugnontussis  1.5636     0.6294   2.484   0.0157 *  
drugplacebo   -0.1364     0.6294  -0.217   0.8292    
---
```

DIFFERENCES IN PRESENTATION OF RESULTS

- Drugs in `lm()` appear as dummy variables
- F statistic and df's the same
- p-value of model the same
- `lm()` estimates of line not useful to ANOVA
  - Interest in ANOVA is in differences in means

EFFECT SIZE (η^2) AND COEFFICIENT OF DETERMINATION (R^2)

- Effect size: ratio of between-group sum of squares and total sum of squares

$$\eta^2 = \frac{SS_{betw}}{SS_{tot}}$$

- They measure equivalent ratios
  - How much of the variance does the model explain
  - Leaving the rest to the residuals
  - `lsr` package has a function that automatically calculates η^2
- η^2 = `lsr::etaSquared(anova_mod1)$eta.sq` = 0.1251055, 0.1251055
- R^2 from `lm()` function = `summary(lm_mod1)$r.squared` = 0.1251055

MULTIPLE COMPARISON *POST HOC* TESTS

- If you have an ANOVA with more than two groups and a significant result …
  - Where did the significance come from?
  - Which comparison of means?
- Null hypothesis really has a 3-way claim:
  - Nontussis is no better than a placebo
  - HCQ is no better than a placebo
  - Nontussis is no better than HCQ
- Running multiple t-tests to find where the significance comes from is **NOT** the answer
- Likely to find significant results simply by chance
  - The whole hypothesis test system based on controlling Type I error rates

ADJUSTMENTS (CORRECTIONS) WHEN MAKING MULTIPLE COMPARISON TESTS

- Bonferroni Correction
  - If I am making `m` number of tests
  - Multiply raw p-values times `m` and then compare to α
  - Very conservative – will accept more H_Os than other methods
- Benjamini-Hochberg Method (False Discovery Rate)
  - Controls the expected proportion of false discoveries among the rejected H_0s
- Many more

MULTIPLE COMPARISON IN R

- Function `pairwise.t.test()` conducts pairwise tests and applies correction of your choice
- Feed it the dependent variable (`score`) and the grouping variable (`drug`)
- Choose an adjustment method (`p.adjust.method`)

WITH BENJAMINI-HOCHBERG METHOD

```r
pairwise.t.test(x = clin_trial$score, g = clin_trial$drug,
                p.adjust.method = "BH")
```

```
    Pairwise comparisons using t tests with pooled SD

data:  clin_trial$score and clin_trial$drug

          hcq   nontussis
nontussis 0.023 -
placebo   0.829 0.023

P value adjustment method: BH
```

WITH BONFERRONI CORRECTION

```r
pairwise.t.test(x = clin_trial$score, g = clin_trial$drug,
                p.adjust.method = "bonferroni")
```

```
    Pairwise comparisons using t tests with pooled SD

data:  clin_trial$score and clin_trial$drug

          hcq   nontussis
nontussis 0.047 -
placebo   1.000 0.027

P value adjustment method: bonferroni
```

MORE ON ANOVA

- `clin_trial` was a *balanced* design
  - All the groups were of the same size
- Unbalanced designs are possible
  - Calculation of SS_x, df_x more complicated
- Very sophisticated designs possible for multiple dimensional studies
- Non-parametric version using ranks
  - Kruskal-Wallis rank sum test
  - `kruskal.test()`
  - Use when normality of data and residuals or equality of variances is a problem
- Wide variety of test designs

EXAMPLE DATA

- New version of COVID clinical trial data
  - This time **unbalanced**
  - 72 cases:
    - 22 placebo, 26 nontussis, 30 hcq
    - 33 supplementary O_2, 39 no supplementary O_2
- Outcome measure is a scaled difference (between 1 and 10) of the difference between viral load at enrollment and viral load after 5 days of treatment

LOAD DATA

```r
clin_trial <- read_rds(here("covid_clin_trial_unbal.rds")) %>%
  mutate(drug = factor(drug),
         o2 = factor(o2))
str(clin_trial)
```

```
tibble [72 × 4] (S3: tbl_df/tbl/data.frame)
 $ pac_id: chr [1:72] "1" "2" "3" "4" ...
 $ drug  : Factor w/ 3 levels "hcq","nontussis",..: 3 3 3 3 3 3 3 3 3 3 ...
 $ o2    : Factor w/ 2 levels "no_supp","supp": 2 2 2 2 2 2 2 2 2 2 ...
 $ score : num [1:72] 1.8 3.1 7.3 4.7 7.7 6.4 2.6 1.4 4.7 5.9 ...
```

TABLE OF THE DATA

```r
clin_trial %>%
  group_by(drug, o2) %>%
  summarise(avg_score = mean(score),
            n = n())
```

```
# A tibble: 6 × 4
# Groups:   drug [3]
  drug      o2      avg_score     n
  <fct>     <fct>       <dbl> <int>
1 hcq       no_supp      4.61    15
2 hcq       supp         4.4     11
3 nontussis no_supp      6.19    13
4 nontussis supp         5.98    11
5 placebo   no_supp      4.31    11
6 placebo   supp         4.51    11
```

DESCRIPTIVE STAT – SUPPLEMENTAL OXYGEN

```
d1 <- clin_trial %>%
  filter(o2 == "supp") %>%
  group_by(drug, o2) %>%
  descr(stats = c("mean", "sd", "min", "med", "max", "iqr", "cv"),
      transpose = TRUE)
```

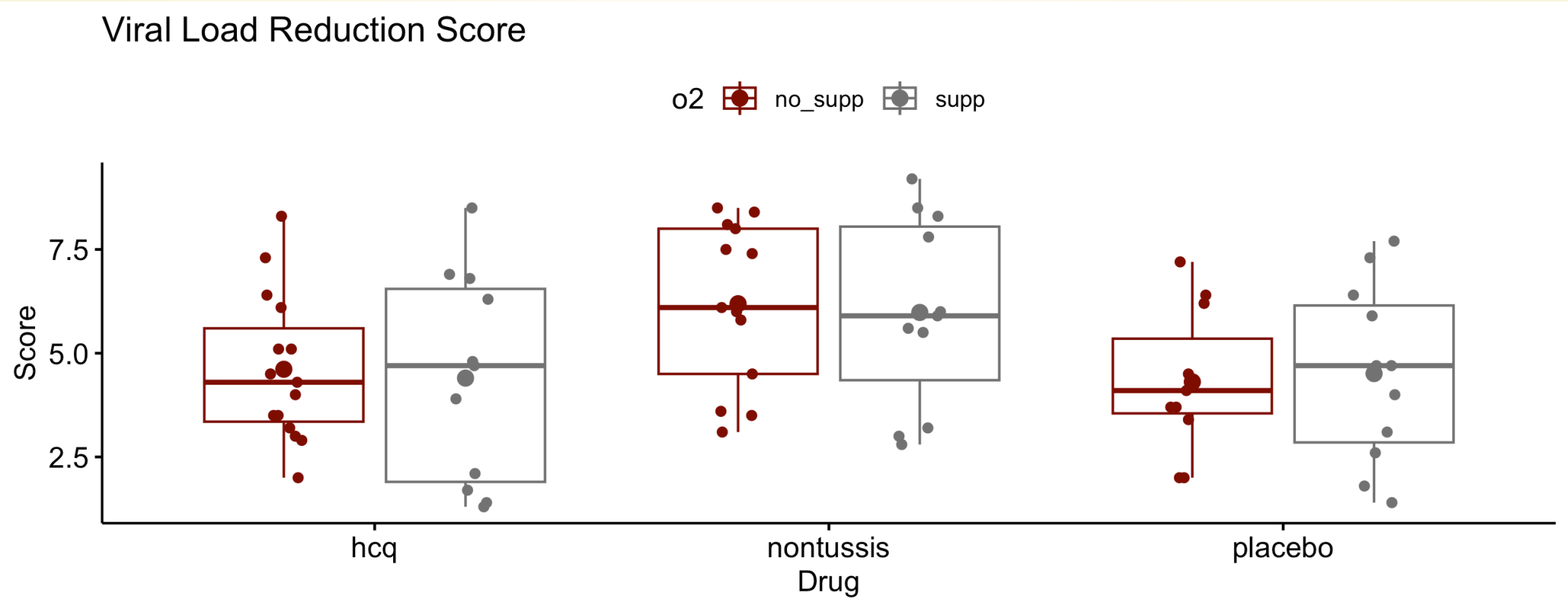
```
Descriptive Statistics
clin_trial$score
Group: drug = hcq, o2 = supp
N: 11

             Mean   Std.Dev    Min   Median    Max    IQR     CV
---------- ------ --------- ------ -------- ------ ------ ------
      score   4.40      2.53   1.30     4.70   8.50   4.65   0.58

Group: drug = nontussis, o2 = supp
N: 11

             Mean   Std.Dev    Min   Median    Max    IQR     CV
---------- ------ --------- ------ -------- ------ ------ ------
```

DESCRIPTIVE STAT – NO SUPPLEMENTAL OXYGEN

```
d2 <- clin_trial %>%
  filter(o2 == "no_supp") %>%
  group_by(drug, o2) %>%
descr(stats = c("mean", "sd", "min", "med", "max", "iqr", "cv"),
      transpose = TRUE)
```

```
Descriptive Statistics
clin_trial$score
Group: drug = hcq, o2 = no_supp
N: 15

             Mean   Std.Dev    Min   Median    Max    IQR     CV
---------- ------ --------- ------ -------- ------ ------ ------
      score   4.61      1.77   2.00     4.30   8.30   2.25   0.38

Group: drug = nontussis, o2 = no_supp
N: 13

             Mean   Std.Dev    Min   Median    Max    IQR     CV
---------- ------ --------- ------ -------- ------ ------ ------
```

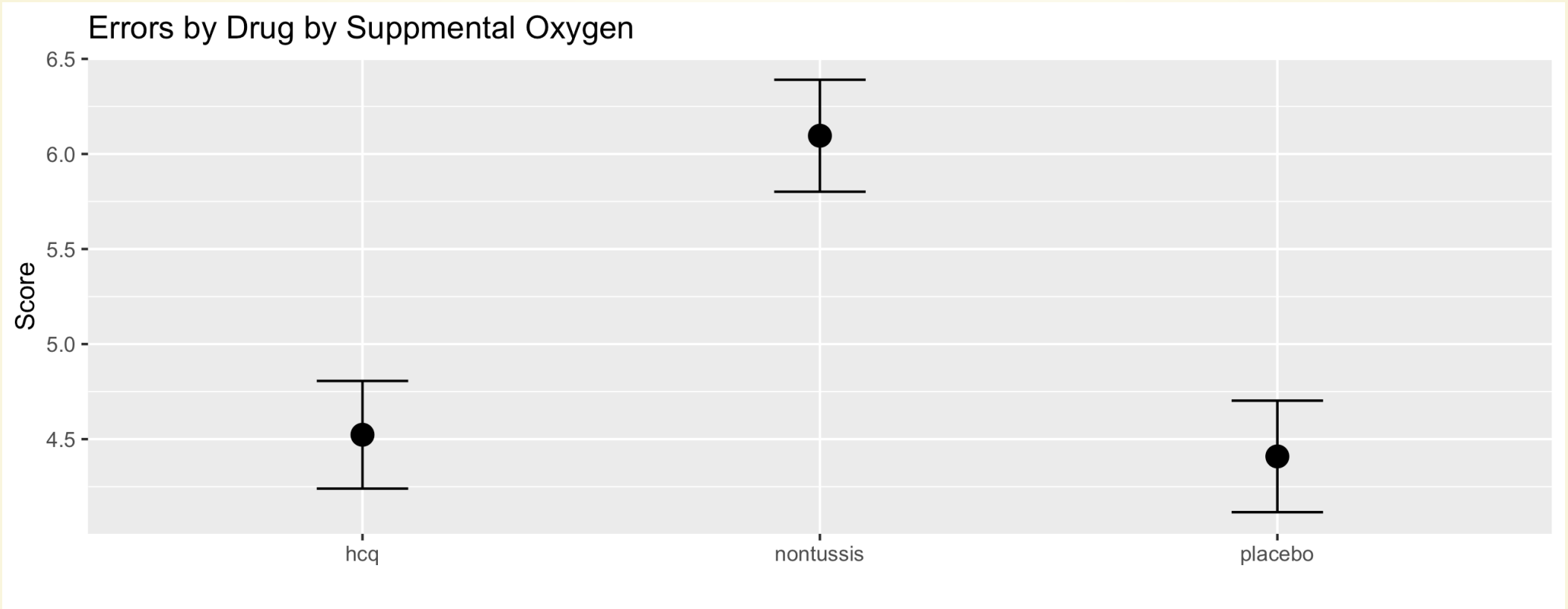
GRAPH OF DATA



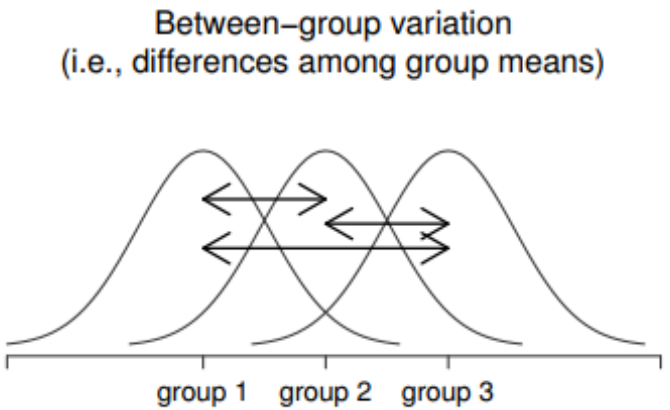
GRAPH OF VIRAL LOAD REDUCTION SCORE ERRORS - CODE

```
gg_error <- clin_trial %>%
  group_by(drug) %>%
  summarise(mean = mean(score),
            sd = sd(score),
            se = sqrt(sd(score)/n())) %>%
  ggplot(aes(x = drug, y = mean)) +
  geom_line() +
  geom_point(size = 4) +
  geom_errorbar(aes(ymin = mean - se, ymax = mean + se), width = .2) +
  labs(title = "Errors by Drug by Suppmental Oxygen", y = "Score", x = "")
```

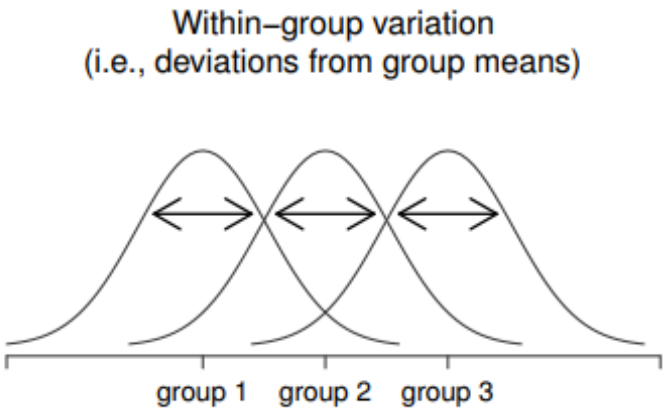
GRAPH


Errors by Drug by Suppmental Oxygen
Score
6.5
6.0
5.5
5.0
4.5
hcq
nontussis
placebo

GRAPHIC DEPICTION OF VARIABILITY WITHIN AND BETWEEN GROUPS



(a)



(b)

ONE-WAY ANOVA

- Focus on drug as factor variable
- Two ways to execute in R
  - `aov()` + `summary()` of model
  - `oneway.test()` - only for one-way ANOVA
- Check assumptions 1st
  - Equality of variances among groups:
    - `car::leveneTest()`
  - Normality of residuals
    - `shapiro.test()`
    - QQ-plot of residuals
    - Needs a calculated ANOVA model to have residuals

CHECK EQUALITY OF VARIANCE ASSUMPTION

```r
car::leveneTest(score ~ drug, data = clin_trial, center = mean)
```

```
Levene's Test for Homogeneity of Variance (center = mean)
      Df F value Pr(>F)
group  2  0.1928 0.8251
      69
```

- H_0 that variances are more or less equal not rejected - OK!

NORMALITY OF RESIDUALS TEST - SHAPIRO-WILK TEST

```r
aov1 <- aov(score ~ drug, data = clin_trial)
shapiro.test(aov1$residuals)
```

```
    Shapiro-Wilk normality test

data:  aov1$residuals
W = 0.95769, p-value = 0.01639
```

ONE-WAY ANOVA WITH aov()

```r
aov1 <- aov(score ~ drug, data = clin_trial)
aov1
```

```
Call:
   aov(formula = score ~ drug, data = clin_trial)

Terms:
                     drug Residuals
Sum of Squares   42.40483 283.37392
Deg. of Freedom         2        69

Residual standard error: 2.026541
Estimated effects may be unbalanced
```

SUMMARY

```r
summary(aov1)
```

```
            Df Sum Sq Mean Sq F value  Pr(>F)
drug         2   42.4  21.202   5.163 0.00814 **
Residuals   69  283.4   4.107
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

ONE-WAY ANOVA WITH `oneway.test()`

```
1 aov2 <- oneway.test(score ~ drug, data = clin_trial, var.equal = TRUE)
2 aov2
```

```
    One-way analysis of means

data:  score and drug
F = 5.1627, num df = 2, denom df = 69, p-value = 0.008139
```

- If the variances are equal, need to use argument `var.equal = TRUE`
- Calculation `oneway.test()` assumes **un**equal variances

MULTIPLE COMPARISONS

WHY DO WE NEED CORRECTIONS?

- I explained that when we do multiple tests
  - Elevated chance of having a at least p-value < α by chance
- If α = 0.05 and 3 tests (as in this problem), at least 14% prob
  - If number of tests rises to 50 – at least a 92% prob of a significant result by chance

TUKEY'S HSD

- Tukey's Honest Significant Difference
- Uses as input the `aov()` model
- Works for balanced and mildly unbalanced designs
- In base R, so has history and popularity (long time available)

EXECUTING TUKEY'S HSD

```r
TukeyHSD(aov1)
```

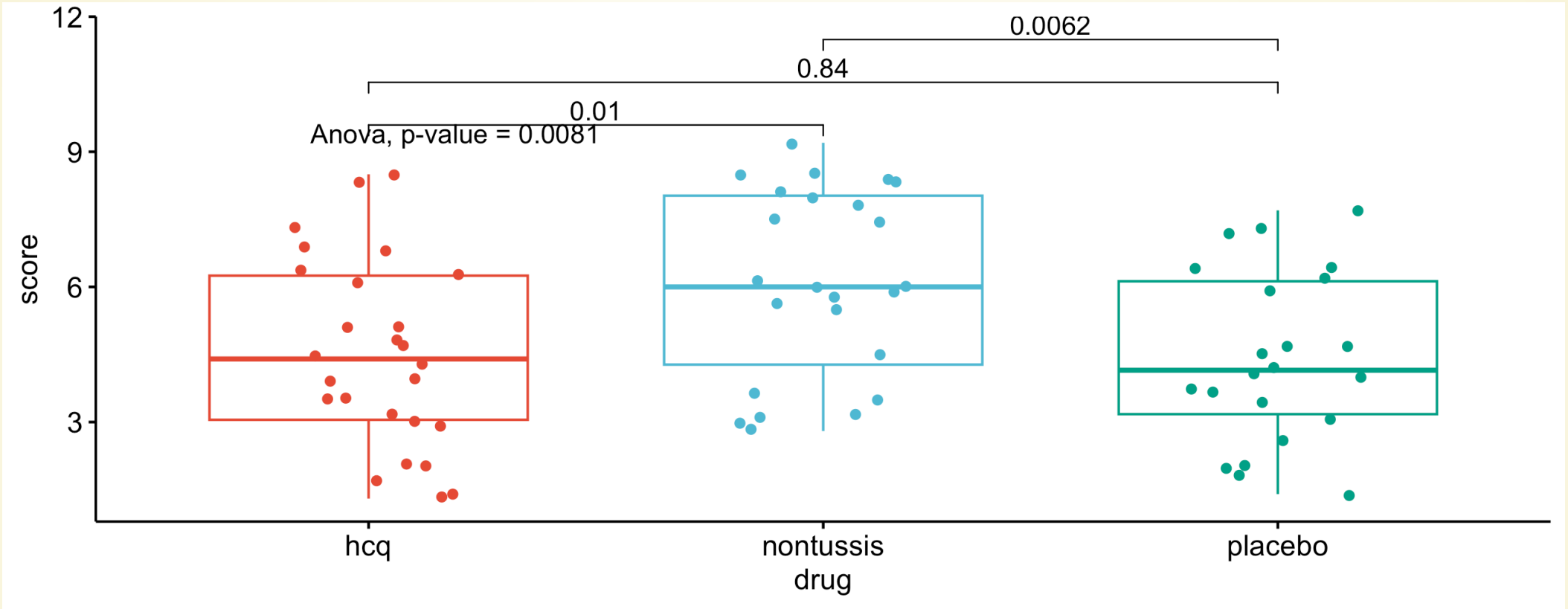
```
  Tukey multiple comparisons of means
    95% family-wise confidence level

Fit: aov(formula = score ~ drug, data = clin_trial)

$drug
                       diff        lwr        upr     p adj
nontussis-hcq      1.572756  0.1986822  2.9468306 0.0209759
placebo-hcq       -0.113986 -1.5201633  1.2921913 0.9794369
placebo-nontussis -1.686742 -3.1195221 -0.2539627 0.0170271
```

BENJAMINI-HOCHBERG FOR COMPARISON

```r
pairwise.t.test(clin_trial$score, clin_trial$drug, p.adjust.method = "BH")
```

```
    Pairwise comparisons using t tests with pooled SD

data:  clin_trial$score and clin_trial$drug

          hcq   nontussis
nontussis 0.012 -
placebo   0.847 0.012

P value adjustment method: BH
```

BOXPLOT WITH THE P-VALUES SHOWN



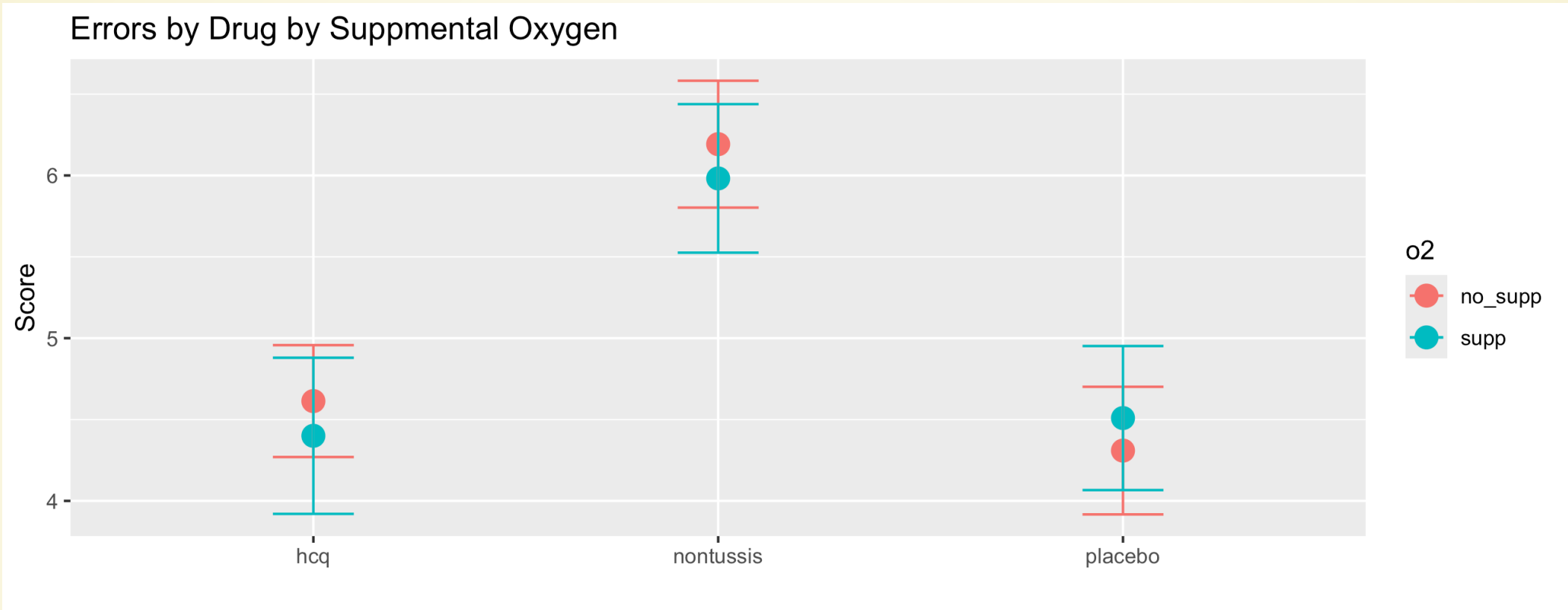
TWO-WAY ANOVA

WHAT IS DIFFERENT FROM ONE-WAY?

- Adding a second factor to be analyzed together
  - Here supplemental oxygen: `o2`
- Equivalent to multiple linear regression with two covariates
- Can also have interactions between the covariates

VISUALIZE DIFFERENCES FOR SUPPLEMENTAL OXYGEN

```
gg_two_way <- clin_trial %>%
  group_by(drug, o2) %>%
  summarise(mean = mean(score),
            sd = sd(score),
            se = sqrt(sd(score)/n())) %>%
  ggplot(aes(x = drug, y = mean, color = o2)) +
  geom_line() +
  geom_point(size = 4) +
  geom_errorbar(aes(ymin = mean - se, ymax = mean + se), width = .2) +
  labs(title = "Errors by Drug by Suppmental Oxygen", y = "Score", x = "")
```


Errors by Drug by Suppmental Oxygen
Score
6
5
4
hcq
nontussis
placebo
o2
no_supp
supp

PREPARE *ADDITIVE* MODEL

- Means no interaction Term

```r
aov3 <- aov(score ~ drug + o2, data = clin_trial)
summary(aov3)
```

```
            Df Sum Sq Mean Sq F value  Pr(>F)
drug         2  42.40  21.202   5.090 0.00871 **
o2           1   0.13   0.128   0.031 0.86151
Residuals   68 283.25   4.165
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

COMMENTS ON ANOVA

- `aov()` internally handles unbalanced models
- If residuals very not normal, use `kruskal.test()` (Kruskal-Wallis test)
- `oneway.test()` does not produce a model object that multiple comparisons can use
- `oneway.test()` assumes model variables **do not** have equal variance
  - Must specify if they do


That's all Folks!
kalilak