MAD – Data Analysis & Biostatistics in R

Power - More Regression - Logistic Regression

James R. Hunter, Ph.D.

DIPA, EPM, UNIFESP

9 October 2020


MAD

Section 1

Today's Program

Today's Program

1. Statistical Power

Today's Program

1. Statistical Power
2. Polynomial Regression

Today's Program

1. Statistical Power
2. Polynomial Regression
3. Logistic Regression

Today's Program

1. Statistical Power
2. Polynomial Regression
3. Logistic Regression
4. Regression and Analysis of Variance (ANOVA) – Kissing Cousins

Section 2

Statistical Power

Conceptual Review

- Null Hypothesis (H_0): a parameter = some theoretical value
- Alternative Hypothesis (H_1): $\geq$, $\neq$, $\leq$ theoretical value (not H_0)
- Test statistic calculated from the data *assuming* H_0 is true
- p-value: probability of observing the theoretical value being tested or a more extreme value
- Small p-vales provide evidence against the H_0
- Level of significance defines the limit of "how small is small enough" for a p-value
  - Notation: α

Power of a Statistical Test

- α provides an idea of the performance of the test across **multiple samples**
- If H_0 is true and $\alpha = 0.01$
  - In 1% of the **samples**, we would reject H_0 erroneously
  - In 99% of the **samples**, we would reject H_0 (correct decision)
- We want to reject H_0 when it is really false
- If α is too small, we can reject H_0 even if the correct value of the parameter is close to the value in H_0
- **Power** of a test is the probability of making a correct decision (i.e., reject H_0 when it is really false)
- Higher levels of power in a test mean that the test is more sensitive

Power – Definition

Power is the probability of correctly rejecting a false null hypothesis when the alternative hypothesis is correct

Power Analysis

- Allows us to determine the probability that a statistical significance test will reject a null hypothesis
- Permits the calculation of the number of cases (n) that would be necessary in a sample to achieve a given level of power
- Put more simply: allows us to determine *a priori* the probability that we make a correct decision

2 Types of Statistical Errors

- **Type I** – Reject H_0 when it is **true**
  - Occurs with probability α
  - Condemning a person falsely of a crime
- **Type II** – Not reject H_0 when it is **false**
  - Occurs with probability β
  - $\beta = 1$ - power
  - A criminal freed in error

Possible Results of a Hypothesis Test

Conclusão do Teste	Estado de Natureza		
		Nula Verdadeira	Nula Falsa
	Não rejeitar H_o	Correta $p = 1 - \alpha$	Erro Tipo II $p = \beta$
	Rejeitar H_o	Erro Tipo I $p = \alpha$	Correta $p = 1 - \beta$

Factors in Calculating Power - BEAN

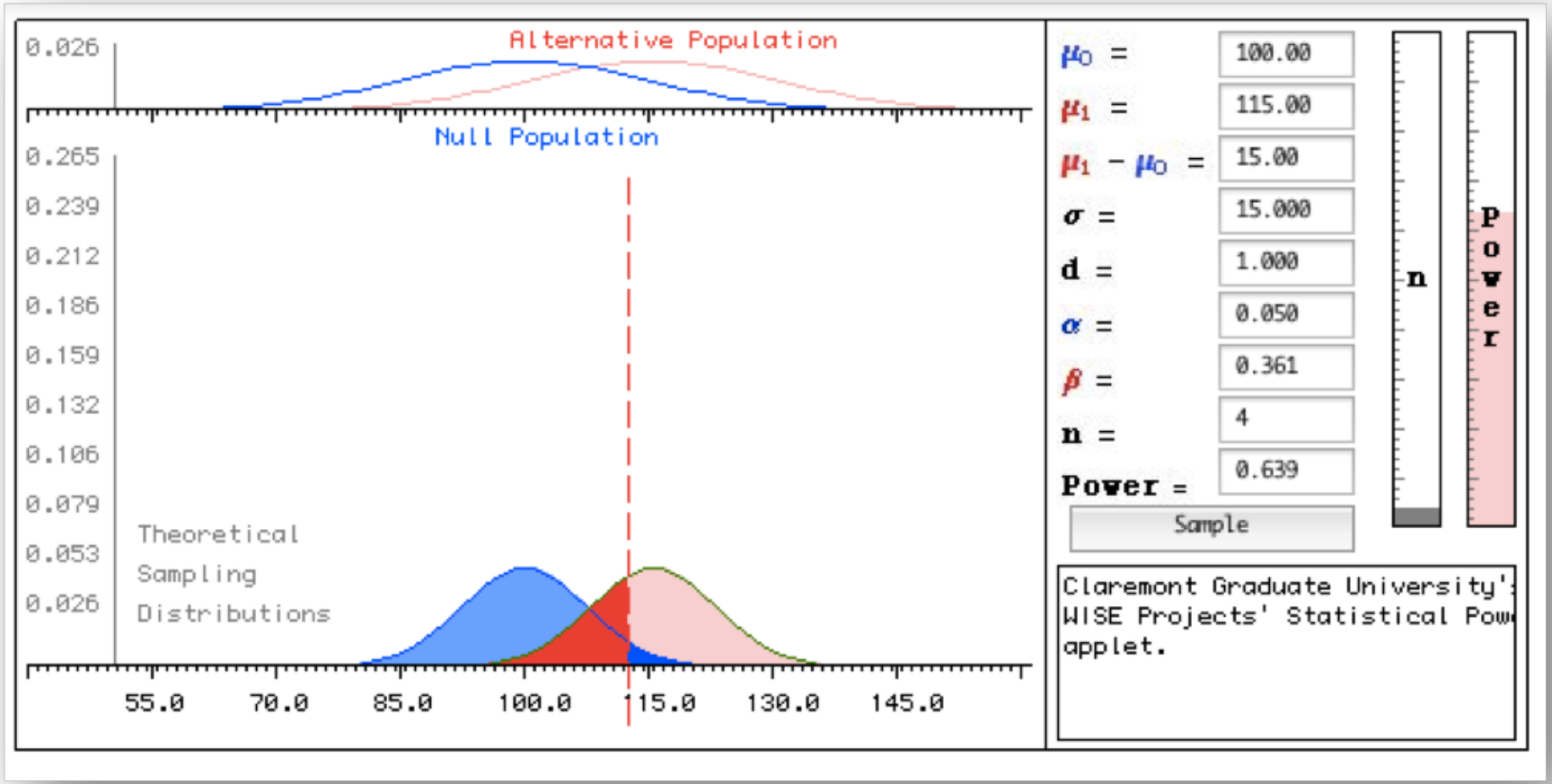
- **B**: β Error (Type II)
  - Probability that significance test will fail to reject H_0 when it is **false**
  - As β error increases, power decreases
- **E**: Effect Size
  - Magnitude of difference between real population mean and null hypothesized mean relative to standard deviation
  - $\frac{(\mu - \mu_{H_0})}{\sigma}$
  - As effect size increases, power increases
- **A**: α Error (Type I)
  - Probability that statistical test will produce a significant finding when H_0 is **true**
  - If $\alpha = 0.05$ *and* H_0 true, false positive would be found in 5 samples out of 100
- **N**: Sample size
  - With larger sample, variability of sample means decreases

Calculating Power

- If you know three, you can calculate the fourth factor
- n and α are under your direct control

Power – Maximization

- Goal in determining power: maximize the power while maintaining the significance level and minimizing the sample size
- Maximize the probability of finding a true effect while minimizing the chance of finding an effect that does not exist
- Power calculations and sample sized calculalted at beginning of project – during planning
  - Not after – already executed experiment. Cannot now try to add new cases


Alternative Population
Null Population
0.026
0.265
0.239
0.212
0.186
0.159
0.132
0.106
0.079
0.053
0.026
Theoretical
Sampling
Distributions
55.0
70.0
85.0
100.0
115.0
130.0
145.0
μ0 = 100.00
μ1 = 115.00
μ1 − μ0 = 15.00
σ = 15.000
d = 1.000
α = 0.050
β = 0.361
n = 4
Power = 0.639
Sample
n
Power
Claremont Graduate University'
WISE Projects' Statistical Pow
applet.

Effect Size

- Of these 4 factors, the most difficult to understand
- Requires some experience with research to apply well
- Table to help beginners use effect size
  - These are general indications
  - With experience you will get a better idea

Table of Effect Size

	Pequeno	Médio	Grande
teste - t	0,2	0,5	0,8
modelo linear	0,02	0,15	0,35
proporções	0,2	0,5	0,8
Qui-quadrado (χ^2)	0,1	0,3	0,5

Section 3

How to Calculate Power

Data on Normal Human Temperature

```
temps <- read_table(here::here("TempData.txt"), col_names = FALSE)
colnames(temps) <- "tempC"
descr(temps, stats = c("mean", "sd", "min", "q1", "med", "q3", "max", "iqr", "cv"),
      transpose = TRUE)
```

```
## Descriptive Statistics
## temps$tempC
## N: 130
##
##                Mean   Std.Dev      Min       Q1   Median       Q3      Max     IQR      CV
## ----------- ------- --------- ------- ------- -------- ------- ------- ------ ------
##       tempC   36.81      0.41   35.72   36.56    36.83   37.06   38.22   0.50   0.01
```

Power Calculations with the pwr Package

- You can specify 3 of the 4 factors; function calculates the 4th

```
pwr.t.test(n = NULL, d = NULL, sig.level = 0.05, power = NULL,
    type = c("two.sample", "one.sample", "paired"),
    alternative = c("two.sided", "less", "greater"))
```

- d = effect size
- sig.level = α
- You need to specify type and alternative

Execute pwr.t.test() Function

```
library(pwr)
pwr.t.test(n = nrow(temps), d = 0.8, sig.level = 0.05,
           type = "one.sample", alternative = "two.sided")
```

```
##
##      One-sample t test power calculation
##
##               n = 130
##               d = 0.8
##       sig.level = 0.05
##           power = 1
##     alternative = two.sided
```

Conclusion about 130 Person Sample

- Power = 1; Sample larger than necessary
- Test a sample of 10

```
pwr.t.test(n = 10, d = 0.8, sig.level = 0.05,
           type = "one.sample", alternative = "two.sided")
```

```
##
##      One-sample t test power calculation
##
##               n = 10
##               d = 0.8
##       sig.level = 0.05
##           power = 0.6162328
##     alternative = two.sided
```

How Many for a Power = 0.95?

```
pwr.t.test(d = 0.80, sig.level = 0.05, power = 0.95,
           type = "one.sample", alternative = "two.sided")
```

```
##
##      One-sample t test power calculation
##
##               n = 22.32455
##               d = 0.8
##       sig.level = 0.05
##           power = 0.95
##     alternative = two.sided
```

- A sample of 23 would be sufficient to have a very high statistical power
- Tell your colleague that 23 cases would have been enough

Types of Tests Covered in `pwr`

- `pwr.p.test`: one-sample proportion test
- `pwr.2p.test`: two-sample proportion test
- `pwr.2p2n.test`: two-sample proportion test (unequal sample sizes)
- `pwr.t.test`: two-sample, one-sample and paired t-tests
- `pwr.t2n.test`: two-sample t-tests (unequal sample sizes)
- `pwr.anova.test`: one-way balanced ANOVA
- `pwr.r.test`: correlation test
- `pwr.chisq.test`: chi-squared test (goodness of fit and association)
- `pwr.f2.test`: test for the general linear model (logistic regression, etc.)

Section 4

Final Thoughts on Hypothesis Tests

1. Be careful when you interpret a p-value
   - p-value does not tell you if H_0 is true
   - What it says: how probable would be the data observed **if** H_0 were true
2. How you collect the data is the key step in arriving at a correct conclusion
3. Always use two-sided tests if you are not *absolutely* certain that one of the sides is totally uninteresting
4. Statistical significance $\neq$ practical significance $\neq$ importance

Effect of Sample Size on p-value

- Three samples: flip a fair coin (p = 0.5); always have 60% heads

heads	trials	statistic	p_value
6	10	6	0.7539063
60	100	60	0.0568879
600	1000	600	0.0000000

5. Lack of significance does not mean that H_0 is true
   - Large p-values can occur because of
     - ★ Luck/chance
     - ★ Problems in data collection
     - ★ H_0 really false
6. Specify your hypotheses before collecting data
7. Proportion tests and tests of means require
   - Independence of observations
   - Distribution of the estimators approximately normal
8. t-test very robust to outliers
   - Only the most extreme outliers can invalidate t-statistic and p-value

Section 5

Polynomial Regression

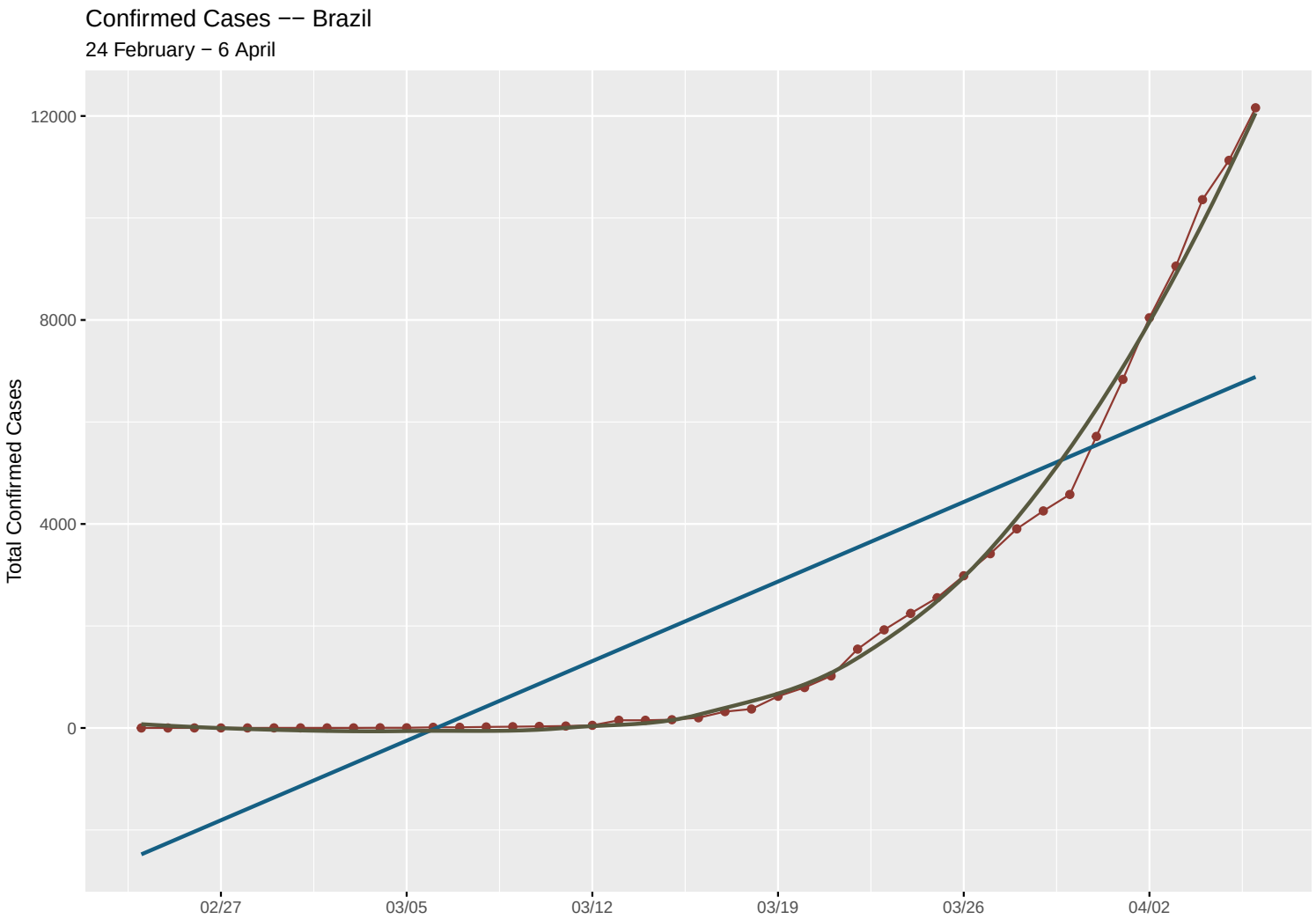
What Is Polynomial Regression

- A way to capture non-linear properties of a variable and retain the linearity of the overall function

Example: Start of COVID-19 Pandemic in Brazil

- Typical of the initial phase of a new infectious disease outbreak
- Deal with period from start of confirmed cases in February through start of April
- Data from Johns Hopkins COVID repository and dashboard

Data


Confirmed Cases −− Brazil
24 February − 6 April
Total Confirmed Cases
12000
8000
4000
0
02/27
03/05
03/12
03/19
03/26
04/02

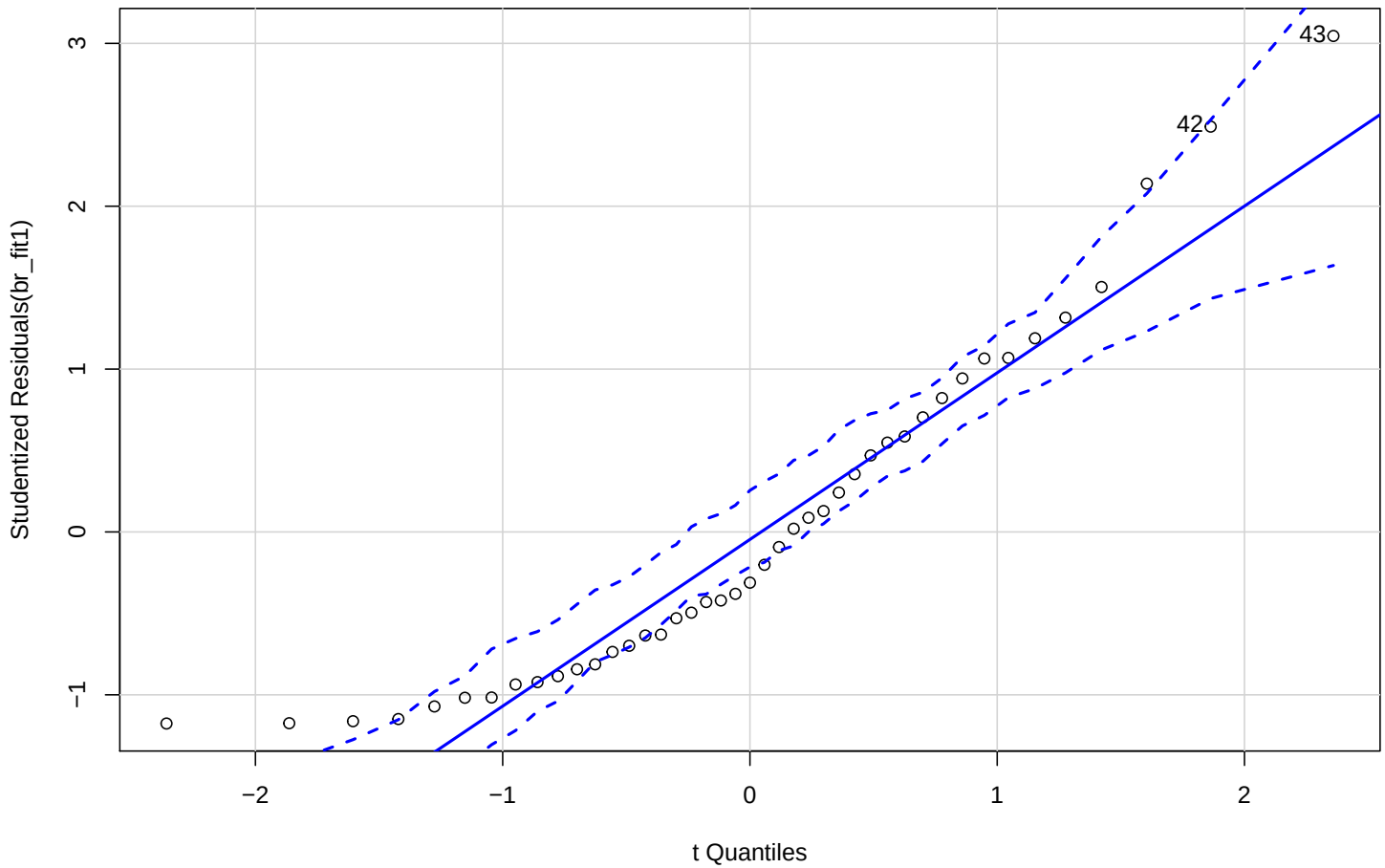
Linear Regression of Cases on Dates

```
br_fit1 <- lm(cases ~ elapsed_date, data = br_cases)
summary(br_fit1)
```

```
##
## Call:
## lm(formula = cases ~ elapsed_date, data = br_cases)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -2301.3 -1630.4  -617.5  1254.1  5277.5
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)  -2700.86     618.11  -4.370 8.29e-05 ***
## elapsed_date   222.89      24.47   9.108 2.11e-11 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 1991 on 41 degrees of freedom
## Multiple R-squared:  0.6693, Adjusted R-squared:  0.6612
## F-statistic: 82.96 on 1 and 41 DF,  p-value: 2.113e-11
```

Q-Q Plot of Linear Fit

```
car::qqPlot(br_fit1)
```



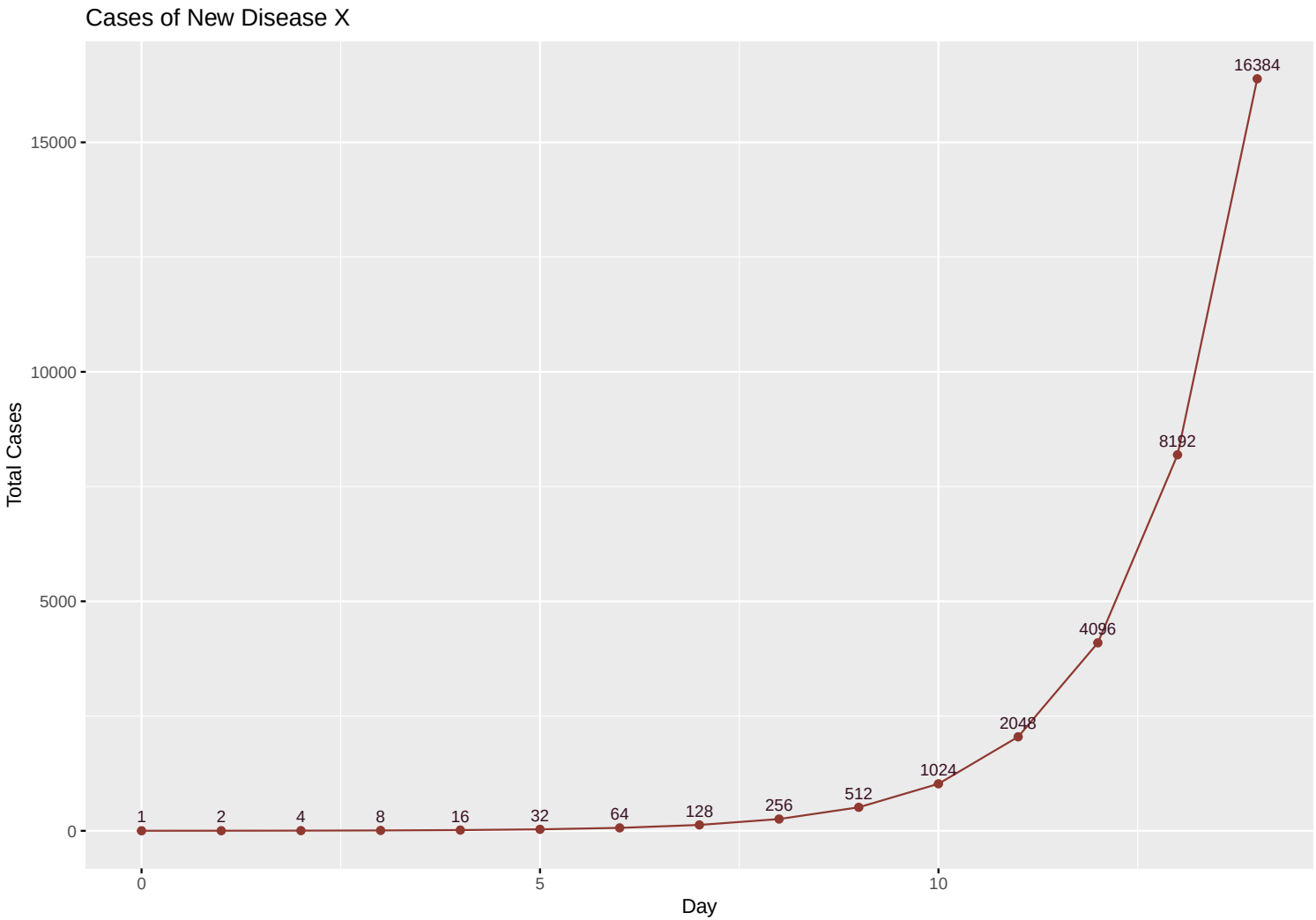

```
## [1] 42 43
```

Create a Polynomial Model of the Data

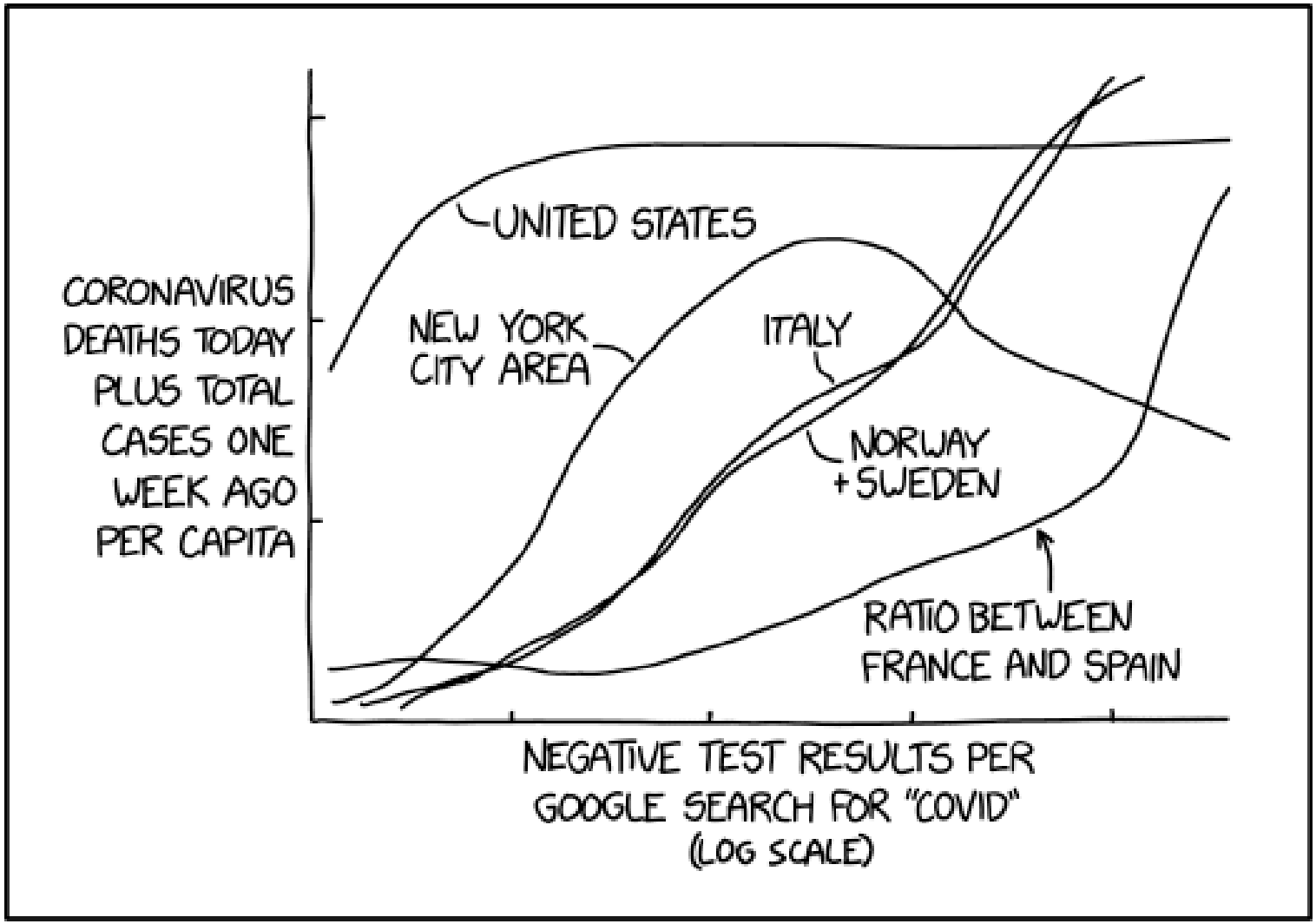
- We know that the initial phases of a new infection follow an exponential curve

```
options(scipen = 1)
r <- 2
times <- 14  #14 days = 2 weeks
x <- r^(0:times)
x
```

```
##  [1]     1     2     4     8    16    32    64   128   256   512  1024  2048
## [13]  4096  8192 16384
```


Cases of New Disease X
Total Cases
15000
10000
5000
0
1
2
4
8
16
32
64
128
256
512
1024
2048
4096
8192
16384
0
5
10
Day

Sort of Like This



I'M A HUGE FAN OF WEIRD GRAPHS, BUT EVEN I ADMIT SOME OF THESE CORONAVIRUS CHARTS ARE LESS THAN HELPFUL.

Create a Quadratic Regression of Elapsed Date

- "Quadratic" means 2nd degree polynomial
  - Will create an equation

$$\hat{Y}_i = \beta_0 + \beta_1 X_i + \beta_2 X_i^2 + \epsilon_i$$

- `poly()` function (base R) creates regression terms from power 1 up to power in command
  - `poly(var, degree = 2, raw = FALSE)`
  - `degree` = put the degree of the polynomial you want to calculate
  - `raw = FALSE` use orthogonal polynomials - that is polynomials that will have no overlap

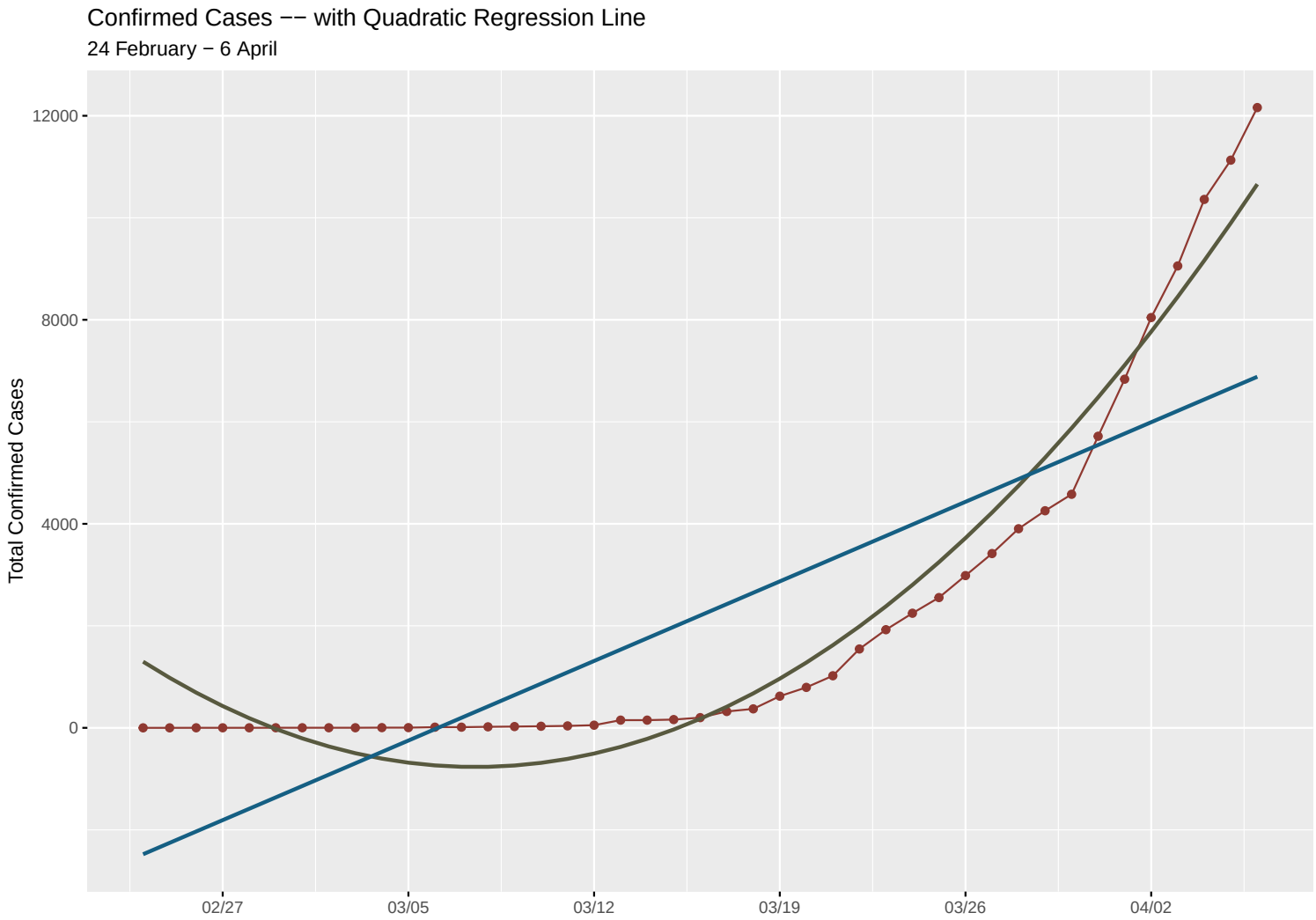
Quadratic Function for elapsed_date

```
br_fit2 <- lm(cases~poly(elapsed_date, 2, raw = TRUE), data = br_cases)
summary(br_fit2)
```

```
##
## Call:
## lm(formula = cases ~ poly(elapsed_date, 2, raw = TRUE), data = br_cases)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -1297.27  -577.19    22.97   606.22  1502.25
##
## Coefficients:
##                                     Estimate Std. Error t value Pr(>|t|)
## (Intercept)                       1640.0117   352.8630   4.648 3.61e-05 ***
## poly(elapsed_date, 2, raw = TRUE)1 -355.8911    36.9896  -9.621 5.81e-12 ***
## poly(elapsed_date, 2, raw = TRUE)2   13.1542     0.8152  16.137  < 2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 735.7 on 40 degrees of freedom
## Multiple R-squared:  0.956,  Adjusted R-squared:  0.9538
## F-statistic: 434.1 on 2 and 40 DF,  p-value: < 2.2e-16
```

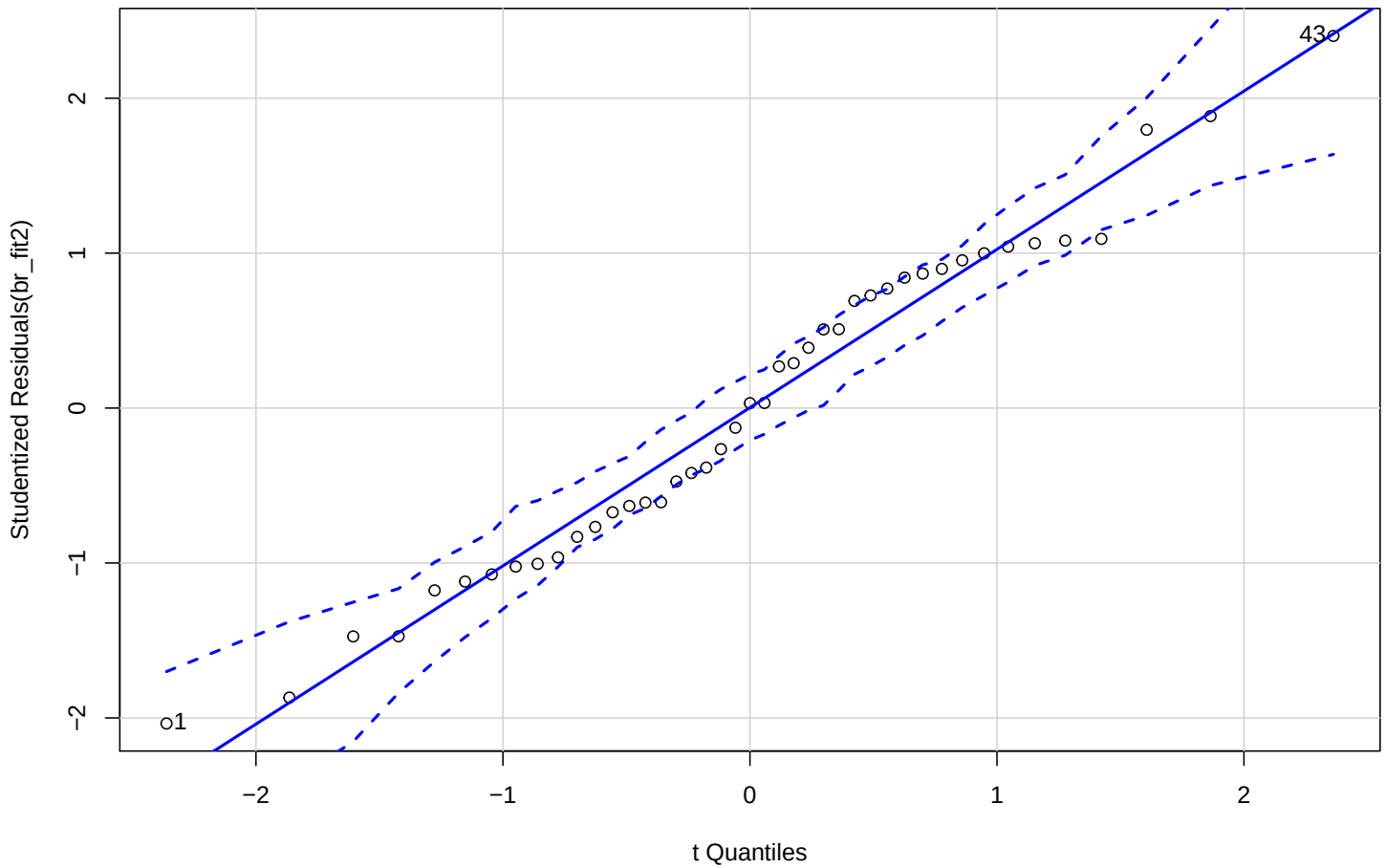
Plot of Regression

```
br_cases %>%
  mutate(fitted_quad = br_fit2$fitted.values) %>%
  ggline(x = "date",
         y = "cases",
         color = "#8F3931FF",
         palette = "uchicago",
         title = "Confirmed Cases -- with Quadratic Regression Line",
         subtitle = "24 February - 6 April",
         xlab = FALSE,
         ylab = "Total Confirmed Cases",
         ggtheme = theme_gray()) +
  geom_line(aes(y = fitted_quad), color = "#58593FFF", size = 1, show.legend = TRUE)
  geom_smooth(color = "#155F83FF", method = "lm", se = FALSE) +
  scale_x_date(date_breaks = "7 days", date_labels = "%m/%d")
```


Confirmed Cases −− with Quadratic Regression Line
24 February − 6 April
Total Confirmed Cases
12000
8000
4000
0
02/27
03/05
03/12
03/19
03/26
04/02

Q-Q Plot of Quadratic Fit

```
car::qqPlot(br_fit2)
```




```
## [1]  1 43
```

Cubic Regression of Date and Cases

- Still can be improved
- Try 3rd Degree - Cubic Regression
- Can we remove the oscillation in the Q-Q graph?
- Can we reduce the regions of over- and under-estimation?

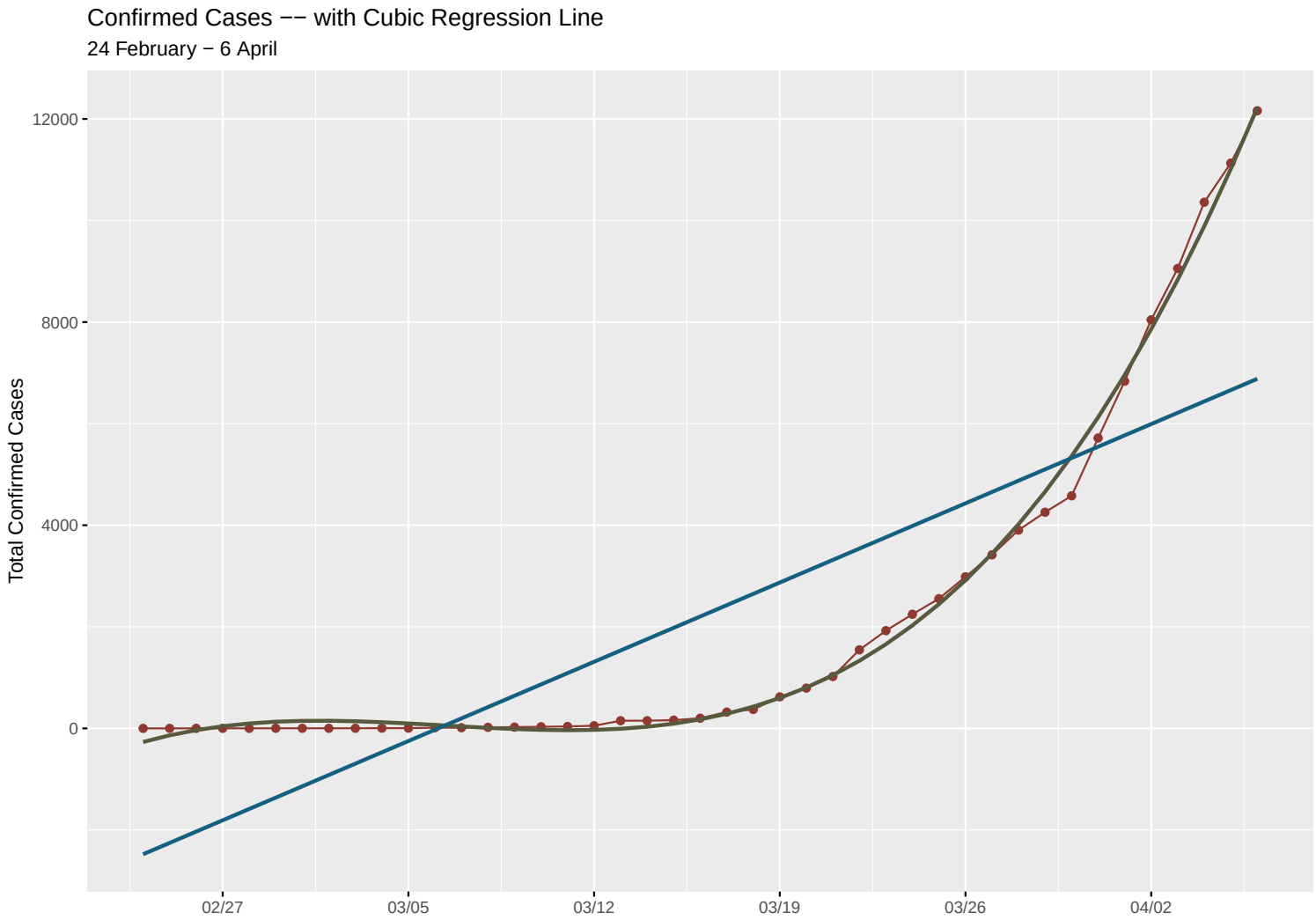
Calculating Cubic Regression

```
br_fit3 <- lm(cases~poly(elapsed_date, 3, raw = TRUE), data = br_cases)
summary(br_fit3)
```

```
##
## Call:
## lm(formula = cases ~ poly(elapsed_date, 3, raw = TRUE), data = br_cases)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -781.37  -94.21   22.30  111.31  474.04
##
## Coefficients:
##                                      Estimate Std. Error t value Pr(>|t|)
## (Intercept)                          -429.60935  141.41184  -3.038  0.00423 **
## poly(elapsed_date, 3, raw = TRUE)1   178.10207   27.51443   6.473 1.14e-07 ***
## poly(elapsed_date, 3, raw = TRUE)2   -16.84034    1.44413 -11.661 2.77e-14 ***
## poly(elapsed_date, 3, raw = TRUE)3    0.45446    0.02159  21.050  < 2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 211.9 on 39 degrees of freedom
## Multiple R-squared:  0.9964, Adjusted R-squared:  0.9962
## F-statistic:  3636 on 3 and 39 DF,  p-value: < 2.2e-16
```

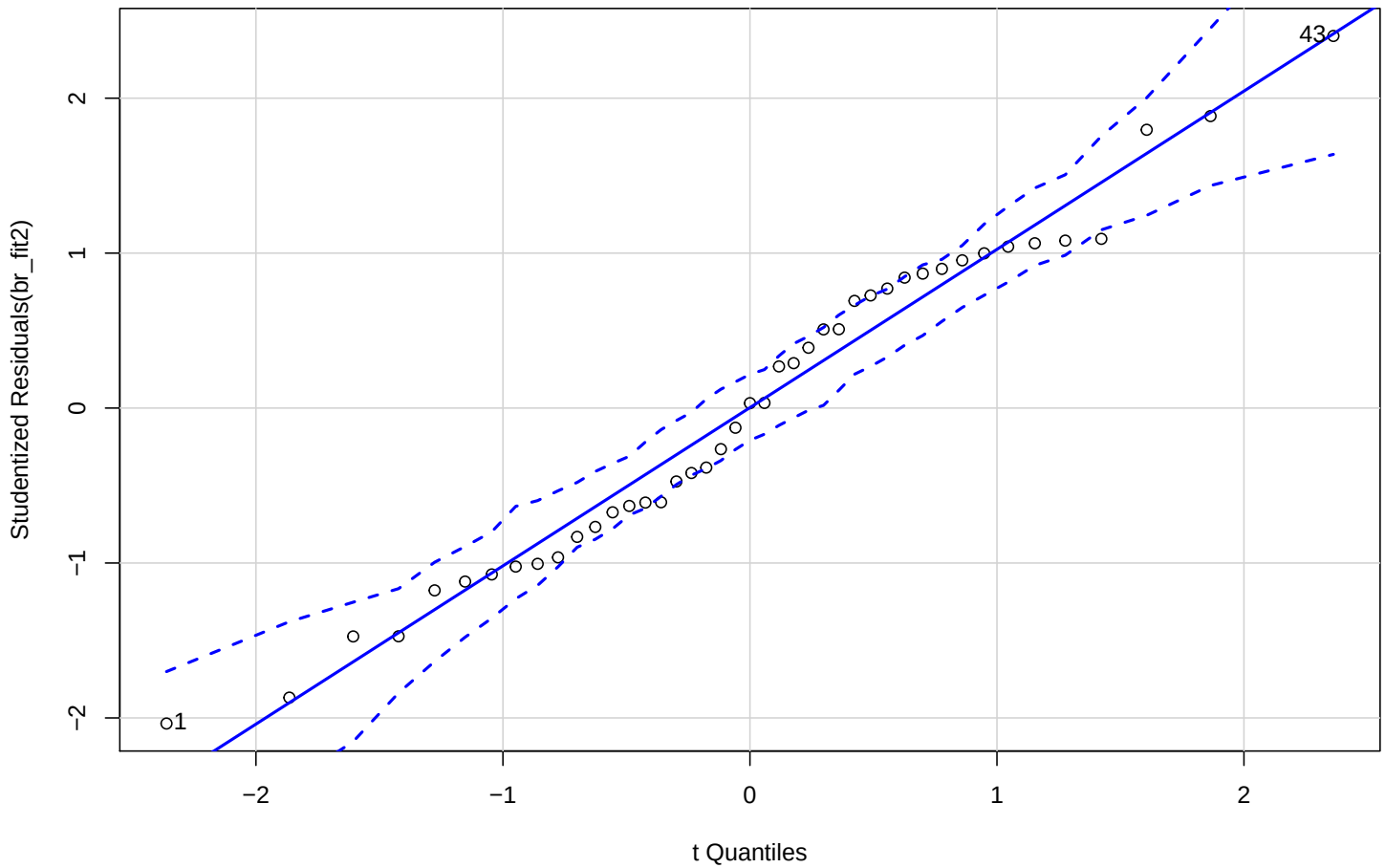
Plot of Cubic Regression

```
br_cases %>%
  mutate(fitted_cubic = br_fit3$fitted.values) %>%
  ggline(x = "date",
         y = "cases",
         color = "#8F3931FF",
         palette = "uchicago",
         title = "Confirmed Cases -- with Cubic Regression Line",
         subtitle = "24 February - 6 April",
         xlab = FALSE,
         ylab = "Total Confirmed Cases",
         ggtheme = theme_gray()) +
  geom_line(aes(y = fitted_cubic), color = "#58593FFF", size = 1, show.legend = TRU
  geom_smooth(color = "#155F83FF", method = "lm", se = FALSE) +
  scale_x_date(date_breaks = "7 days", date_labels = "%m/%d")
```


Confirmed Cases −− with Cubic Regression Line
24 February − 6 April
Total Confirmed Cases
12000
8000
4000
0
02/27
03/05
03/12
03/19
03/26
04/02

Q-Q Plot of Cubic Fit

```
car::qqPlot(br_fit2)
```




```
## [1]  1 43
```

Section 6

One-Way ANOVA - Regression's Kissing Cousin

Analysis of Variance - ANOVA

- Tool developed in 1920's by Sir Ronald Fisher
- Misnomer
  - Technique for comparing differences in means
  - Not variances
- Way to study groups
  - And how some outcome variable can depend on them
- Variety of ANOVA models
- Multiple comparison tests when we want to isolate certain differences
  - Correction for aggressive estimation of p-value
- Clinical trials
  - Go-to method of analyzing clinical trial data

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
- 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

```
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", "plac...
## $ o2     <chr> "supp", "supp", "supp", "supp", "supp", "supp", "supp", "sup...
## $ score  <dbl> 1.8, 8.2, 7.3, 4.7, 7.7, 6.4, 2.6, 1.4, 4.7, 5.9, 4.0, 8.5, ...
```

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	4.972727

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)
```

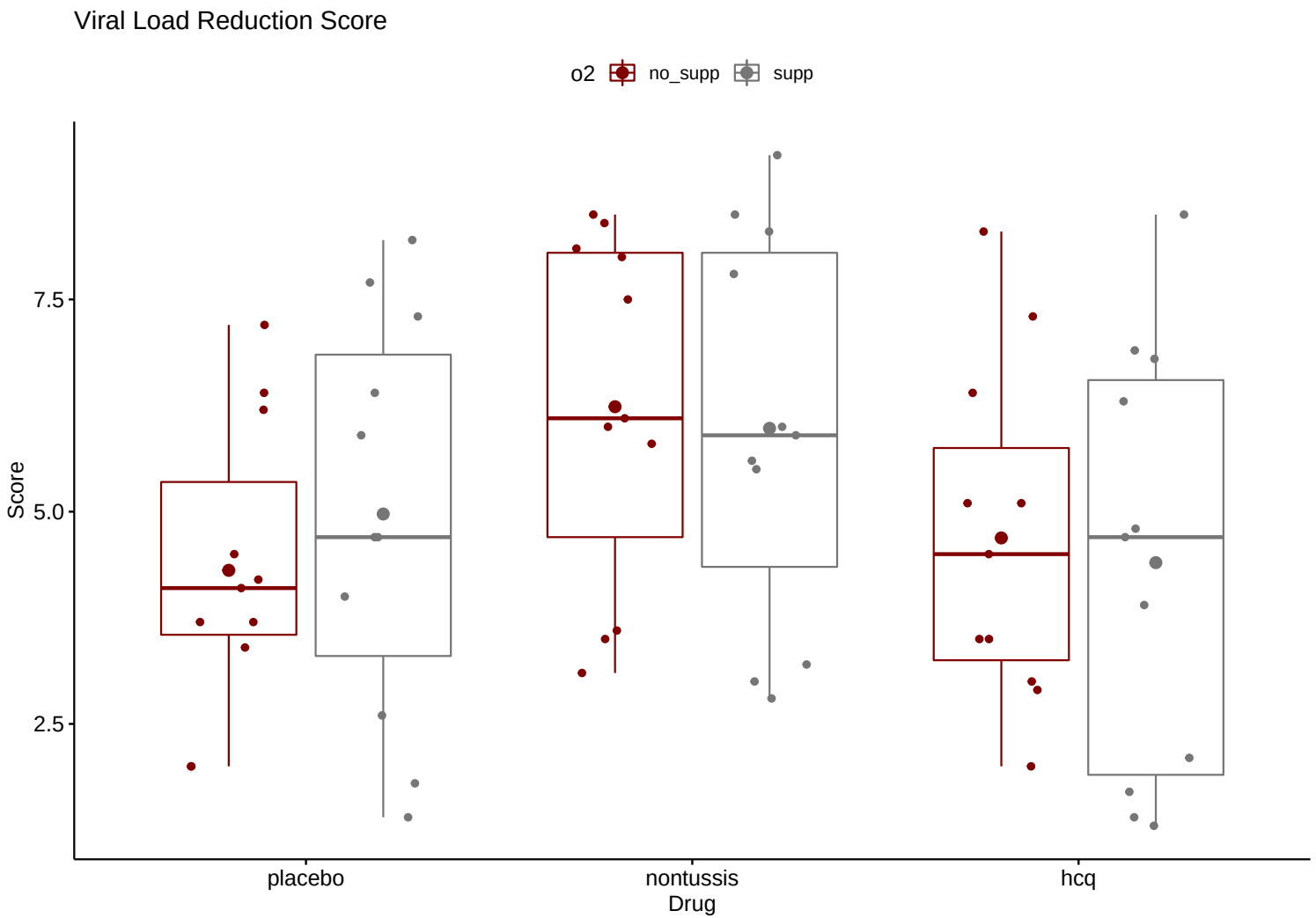
```
## Descriptive Statistics
## clin_trial$score
## Group: drug = hcq, o2 = supp
## N: 11
##
##               Mean   Std.Dev    Min   Median    Max    IQR     CV
## ----------- ------ --------- ------ -------- ------ ------ ------
##       score   4.40      2.53   1.30     4.70   8.50   4.65   0.58
##
## 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
## ----------- ------ --------- ------ -------- ------ ------ ------
##       score   4.97      2.36   1.40     4.70   8.20   3.55   0.47
```

Descriptive Statistics of Data – No Supplemental Oxygen

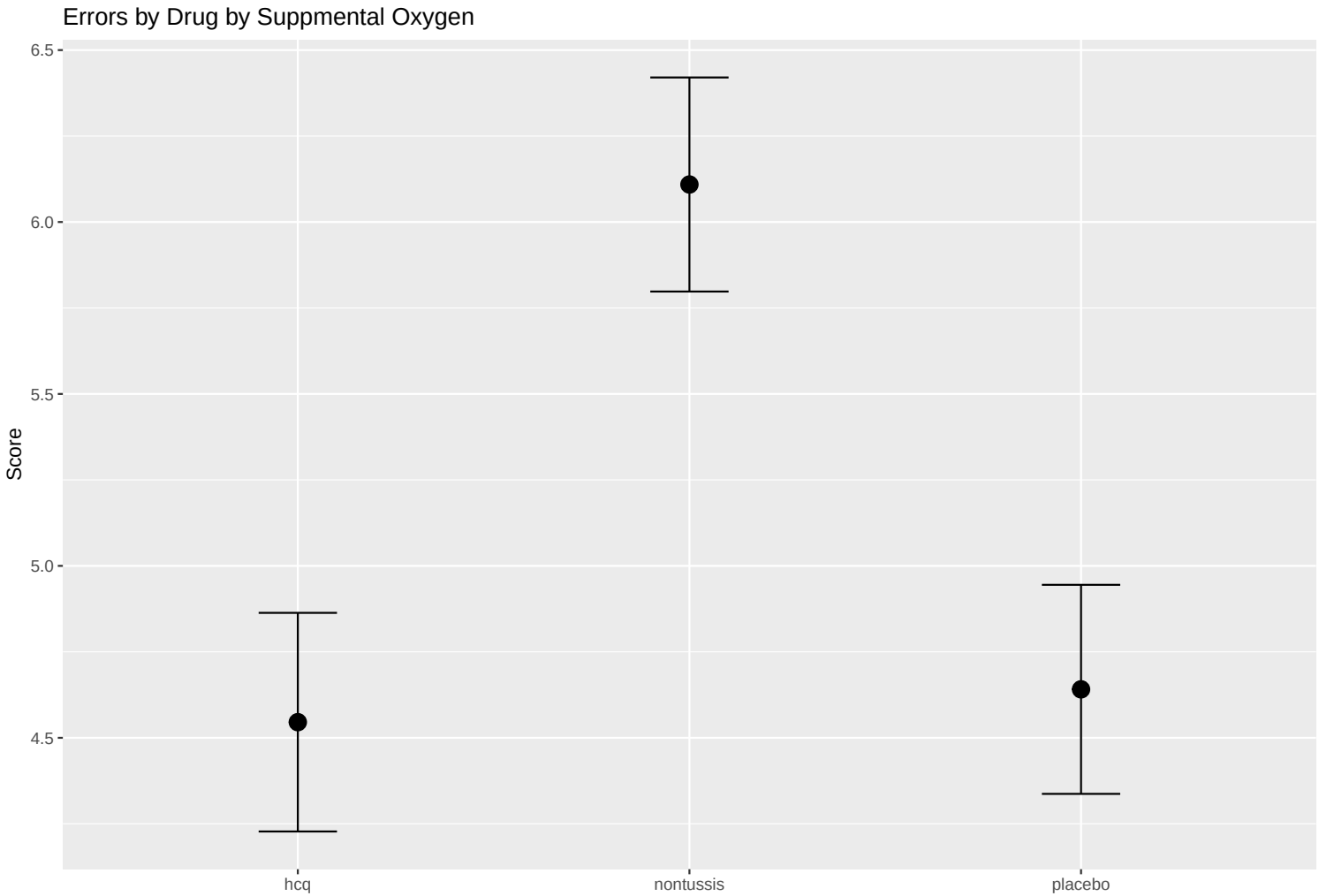
```
clin_trial %>%
  filter(o2 == "no_supp") %>%
  group_by(drug, o2) %>%
descr(stats = c("mean", "sd", "min", "med", "max", "iqr", "cv"),
      transpose = TRUE)
```

```
## 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
## ----------- ------ --------- ------ -------- ------ ------ ------
##       score   4.31      1.69   2.00     4.10   7.20   1.80   0.39
```

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

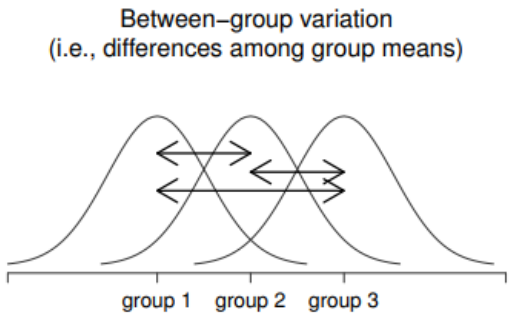
- Difference between the group means ($\bar{Y}_k$) and the individual values (Y_{ik})

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

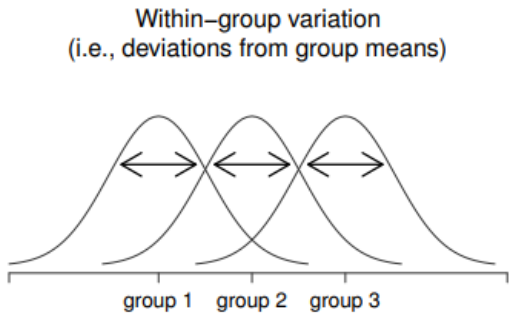
As with regression . . .

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

Graphic Depiction of Variability within and between Groups



(a)



(b)

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
- $\therefore$ 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   33.8   16.90   3.723 0.0296 *
## Residuals    63  286.0    4.54
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

- Review items in table

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  1.8307  3.9545
##
## Coefficients:
##               Estimate Std. Error t value Pr(>|t|)
## (Intercept)    4.54545    0.45428  10.006 1.22e-14 ***
## drugnontussis  1.56364    0.64245   2.434   0.0178 *
## drugplacebo    0.09545    0.64245   0.149   0.8824
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 2.131 on 63 degrees of freedom
## Multiple R-squared:  0.1057, Adjusted R-squared:  0.0773
## F-statistic: 3.723 on 2 and 63 DF,  p-value: 0.02964
```

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)$eta.sq` = 0.1056935, 0.1056935
- R^2 from `lm()` function = `summary(lm_mod1)$r.squared` = 0.1056935

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

```
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 0.039 -
## placebo   0.882 0.039
##
## P value adjustment method: BH
```

With Bonferroni Correction

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

```
##
##  Pairwise comparisons using t tests with pooled SD
##
## data:  clin_trial$score and clin_trial$drug
##
##           hcq   nontussis
## nontussis 0.053 -
## placebo   1.000 0.077
##
## 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