Statistical Inference Project

Chris McKelt

Part 2: Basic Inferential Data Analysis

Description: https://www.coursera.org/learn/statistical-inference/peer/3k8j5/statistical-inference-course-pr

https://stat.ethz.ch/R-manual/R-devel/library/datasets/html/ToothGrowth.html

ToothGrowth: The response is the length of odontoblasts (cells responsible for tooth growth) in 60 guinea pigs.

Each animal received one of three dose levels of vitamin C (0.5, 1, and 2 mg/day) by one of two delivery methods,

orange juice or ascorbic acid (form of vitamin C and coded as VC).

```
##      dplyr      tidyr    ggplot2      knitr   markdown    moments
##       TRUE       TRUE       TRUE       TRUE       TRUE       TRUE
##    nortest      e1071 data.table      sqldf    pastecs
##       TRUE       TRUE       TRUE       TRUE       TRUE
```

Load & tidy data. Create 2 groups according to supplement type (orange juice vs ascorbic acid)

```
library(datasets)
data(ToothGrowth)
tbl <- data.table(ToothGrowth)
setnames(tbl, c('len', 'supp', 'dose'), c('tooth_length', 'supplement', 'dose'))
team_oj <- sqldf("select tooth_length, supplement, dose from tbl where supplement = 'OJ'")
  team_vc <- sqldf("select tooth_length, supplement, dose from tbl where supplement = 'VC'")
  options(digits = 2)
  overview <- rbind(team_oj, team_vc)
  overview$dose <- as.numeric(overview$dose)
```

explore the data

```
summary(tbl)
```

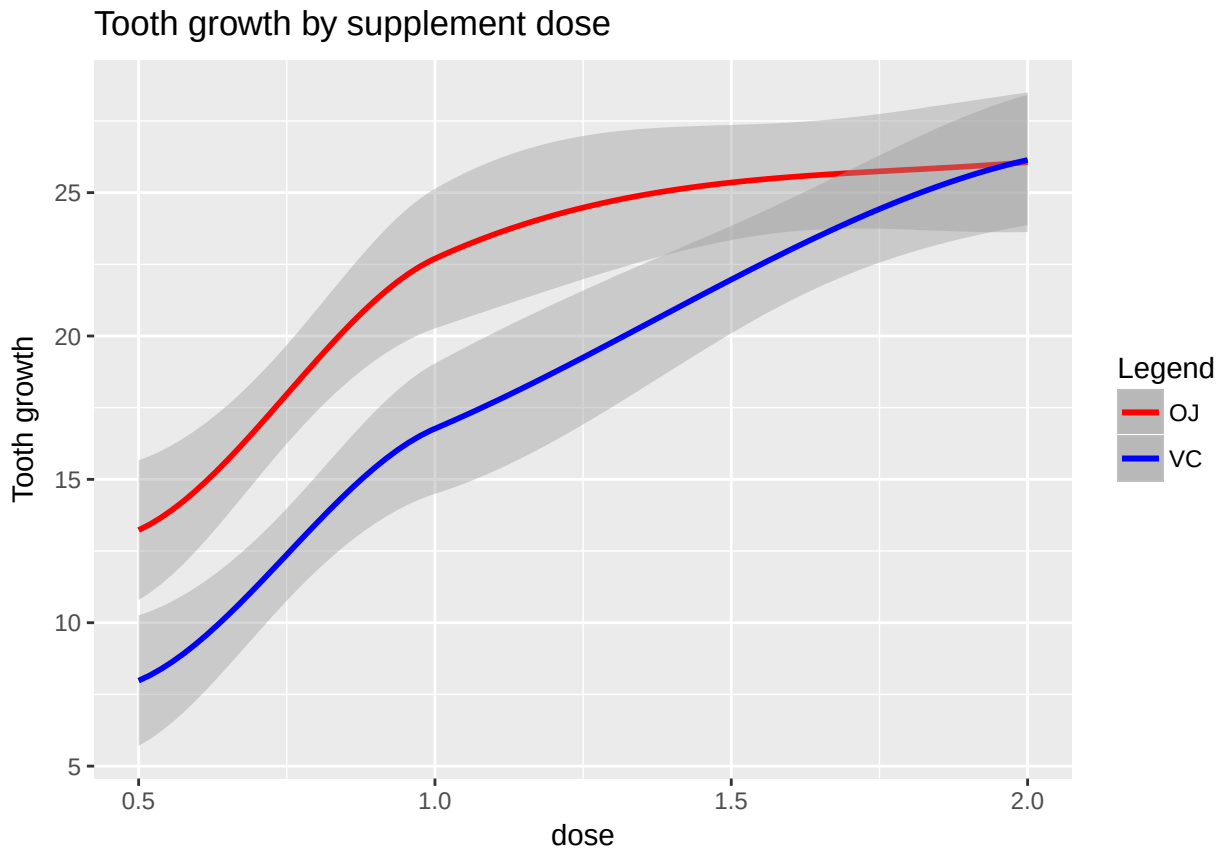
```
##   tooth_length supplement      dose
##  Min.   : 4    OJ:30      Min.   :0.50
##  1st Qu.:13    VC:30      1st Qu.:0.50
##  Median :19               Median :1.00
##  Mean   :19               Mean   :1.17
##  3rd Qu.:25               3rd Qu.:2.00
##  Max.   :34               Max.   :2.00
```

```
  table(tbl$supplement, tbl$dose)
```

```
##
##      0.5  1  2
##   OJ  10 10 10
##   VC  10 10 10
```

plot to compare tooth growth by supplement type and dose

```
ggplot(overview, aes(x = dose, y = tooth_length)) +
  geom_smooth(data = subset(overview, supplement == "OJ"), aes(colour = "OJ"), linetype = "solid", si
  geom_smooth(data = subset(overview, supplement == "VC"), aes(colour = "VC"), linetype = "solid", si
  scale_color_manual("Legend", values = c("red", "blue")) +
  labs(title = "Tooth growth by supplement dose") +
  labs(x = "dose", y = "Tooth growth")
```



Observation: From the 60 samples spread over 2 supplement types with 3 different dosages, Delivery by Orange Juice results in larger tooth growth. But as dosage increases the growth evens out to where at 2 mg per day the growth between orange juice and ascorbic acid are the same

output statistics of data

```
  stat.desc(team_oj)
```

```
##              tooth_length supplement  dose
## nbr.val             30.00         NA 30.00
## nbr.null             0.00         NA  0.00
## nbr.na               0.00         NA  0.00
## min                  8.20         NA  0.50
## max                 30.90         NA  2.00
## range               22.70         NA  1.50
## sum                619.90         NA 35.00
## median              22.70         NA  1.00
## mean                20.66         NA  1.17
## SE.mean              1.21         NA  0.12
## CI.mean.0.95         2.47         NA  0.24
## var                 43.63         NA  0.40
## std.dev              6.61         NA  0.63
## coef.var             0.32         NA  0.54
```

```
  stat.desc(team_vc)
```

```
##              tooth_length supplement  dose
## nbr.val             30.00         NA 30.00
## nbr.null             0.00         NA  0.00
## nbr.na               0.00         NA  0.00
## min                  4.20         NA  0.50
## max                 33.90         NA  2.00
## range               29.70         NA  1.50
## sum                508.90         NA 35.00
## median              16.50         NA  1.00
## mean                16.96         NA  1.17
## SE.mean              1.51         NA  0.12
## CI.mean.0.95         3.09         NA  0.24
## var                 68.33         NA  0.40
## std.dev              8.27         NA  0.63
## coef.var             0.49         NA  0.54
```

Observation: VC has a higher standard deviation & standard error meaning it is more disbursed.

Compare tooth growth by supplements and dose

Does the amount of supplements affect tooth growth?

```
tresult_supplement <- t.test(tooth_length ~ supplement, data = overview, var.equal = FALSE, paired = F
ho <- 'H0 supplements have not affect on tooth growth'
ha <- 'Ha supplements will affect tooth growth'
test_reject(ho, tresult_supplement)
```

```
## [1] "p.value is  0.0606345078809341  which is greater than alpha (.05) so we reject H0"
## [1] "REJECT NULL HYPOTHESIS:  H0 supplements have not affect on tooth growth"
```

Does the dosage of OJ vs VC affect tooth growth?

need to break up the dose into 3 groups for testing

```
dose_small <- subset(overview, dose == 0.5)
dose_medium <- subset(overview, dose == 1.0)
dose_high <- subset(overview, dose == 2.0)
```

small dose

```
tresult_dose_small <- t.test(tooth_length ~ supplement, data = dose_small, var.equal = FALSE, paired =
ho <- 'small dose .05 mg is best delivered by a single supplement type to aid tooth growth'
test_reject(ho, tresult_dose_small)
```

```
## [1] "FAIL TO REJECT NULL HYPOTHESIS:  small dose .05 mg is best delivered by a single supplement typ
```

medium dose

```
tresult_dose_medium <- t.test(tooth_length ~ supplement, data = dose_medium, var.equal = FALSE, paired
ho <- 'medium dose of 1.0 mg is best delivered by a single supplement type to aid tooth growth'
test_reject(ho, tresult_dose_medium)
```

```
## [1] "FAIL TO REJECT NULL HYPOTHESIS:  medium dose of 1.0 mg is best delivered by a single supplement
```

high dose

```
tresult_dose_high <- t.test(tooth_length ~ supplement, data = dose_high, var.equal = FALSE, paired = F
ho <- 'high dose of 2.0 mg is best delivered by a single supplement type to aid tooth growth'
test_reject(ho, tresult_dose_high)
```

```
## [1] "p.value is  0.963851588723373  which is greater than alpha (.05) so we reject H0"
## [1] "REJECT NULL HYPOTHESIS:  high dose of 2.0 mg is best delivered by a single supplement type to a
```

Conclusion: Vitamin C aids in the overall tooth growth of guinea pigs. In smaller doses (.5-1.5) delivering with Orange Juice is more effective Over larger doses the growth results trend towards similar results per dosage $>$ 2 mg.

Data may be invalid –$>$ Does the Orange Juice already have Vitamin C in it???