

Statistical Analysis part2

Deepthi Murthy

July 25, 2015

“

Overview

In this short report we are going to analyze the `ToothGrowth` data of the R data sets package which describes the effect of vitamin C on tooth growth in guinea pigs

The `ToothGrowth` data set consists of 60 observations of 3 variables:

- `len`: Tooth length in millimeters (numeric variable)
- `supp`: Supplement type (factor variable with levels VC and OJ)
- `dose`: Dose in milligrams (numeric variable)

Even though `dose` is a numeric variable we will treat it as a factor in the subsequent analysis because only three dose levels of Vitamin C (0.5, 1, and 2 mg) were used.

In total, we deal with 60 unique guinea pigs. Each pig was assigned to a group which received a specific dose level of vitamin C. In addition, in each of those groups two different delivery methods (orange juice or ascorbic acid) were applied leaving 10 pigs per subgroup:

```
“{r echo=FALSE}
with(ToothGrowth, table(dose, supp))
“
```

Exploratory data analysis

The average guinea pig tooth length is `r round(mean(ToothGrowth$len), 3)` with a standard deviation of `r round(sd(ToothGrowth$len), 3)`.

However, we are more interested in finding differences in tooth length across different groups of supplement types, dose levels and their respective combinations.

```
“{r echo=FALSE, fig.width=7, fig.height=2.5, fig.align='center'}

g_len_by_supp <- ggplot(ToothGrowth, aes(supp, len)) + geom_boxplot() + xlab('Supplement type') +
ylab('Tooth length (mm)') + ggtitle('Tooth length by supplement type')

g_len_by_dose <- ggplot(ToothGrowth, aes(as.factor(dose), len)) + geom_boxplot() + xlab('Dose (mg)') +
ylab('Tooth length (mm)') + ggtitle('Tooth length by dose level')

grid.newpage() # Push viewport to graphic device pushViewport(viewport(layout = grid.layout(1, 2,
widths = c(0.5, 0.5)))) print(g_len_by_supp, vp = viewport(layout.pos.row = 1, layout.pos.col = 1))
print(g_len_by_dose, vp = viewport(layout.pos.row = 1, layout.pos.col = 2))
“
```

The left box plot above reveals that guinea pigs which received their dose of vitamin C using orange juice have on average a greater tooth length than those pigs which got their dose using ascorbic acid. The orange juice group distribution is skewed to the left whereas the ascorbic acid group seems to be fairly symmetric.

Grouping by dose level (right hand side box plot) reveals even greater differences in tooth length and their respective means. The differences between the dose level groups are so big that not even the interquartile range boundaries overlap:

```
“{r echo = FALSE}
```

```
ToothGrowth %>% group_by(dose) %>% summarize( q25 = quantile(len, 0.25), q75 = quantile(len, 0.75) )
%>% as.data.frame
```

```
“
```

Both trends can be further observed when examining tooth length differences across both groups at the same time.

```
“{r echo=FALSE, fig.width=5, fig.height=3.5}
```

```
ggplot(ToothGrowth, aes(as.factor(dose), len)) + geom_boxplot(aes(fill = as.factor(dose))) + facet_grid(. ~
supp) + xlab('Dose (mg)') + ylab('Tooth length (mm)') + ggtitle('Tooth length by dose level and supplement
type') + scale_fill_manual(values = wes_palette("FantasticFox")) + theme(legend.position = "none")
```

```
“
```

Hypothesis tests

Conditions

Since we dealing with subsets of $n \leq 30$ our standard error estimate will generally not be accurate. Therefore, we will use the t distribution for our hypothesis tests and when constructing confidence intervals.

Based on the current information it is hard to check if the 2 necessary prerequisites for this setup are met:

1. Independence of the observations
2. Observations come from a nearly normal distribution.

However, we assume that the guinea pigs were randomly assigned to one of the groups (guarantees independence) and that they were sampled from a nearly normal population.

Differences in supplement types

First, we will check if the observed difference in tooth length means between pigs who received their dose using orange juice and those who received their dose via ascorbic acid is statistically different from 0.

Our conservative choice will be a t -test with unequal variances of the two samples. The table below shows the results:

```
“{r echo=FALSE}
```

```
t_diff_supp <- t.test(len ~ supp, ToothGrowth, var.equal = FALSE) tidy(t_diff_supp)
```

```
“
```

With a p -value of `r round(t_diff_supp$p.value, 3)` which is larger than the significance value, 0.05, we fail to reject the null hypothesis. The data does not provide strong evidence that the average difference in tooth length across the two supplement types is different from 0.

The hypothesis test result corresponds to the respective confidence interval `r round(t_diff_supp$conf.int[1:2], 3)` because it includes 0.

Differences in dose levels

Since we are dealing with 3 different dose levels we need to apply 3 t-tests to cover all factor level combinations. Again, our null hypothesis in all cases is that the average difference in tooth length is 0.

The table below summarizes the results of the three pairwise comparisons. Column `null_hypothesis` shows which null hypothesis was used for the respective t-test also indicating which sub group dose levels means were compared to each other.

```
“{r echo = FALSE}

pairwise_results <- t.test(ToothGrowth$len[ToothGrowth$dose == 2], ToothGrowth$len[ToothGrowth$dose == 1]) %>% tidy %>% mutate( null_hypothesis = 'µ2mg - µ1mg = 0' ) %>% select(9, 1:8)

pairwise_results <- t.test(ToothGrowth$len[ToothGrowth$dose == 2], ToothGrowth$len[ToothGrowth$dose == 0.5]) %>% tidy %>% mutate( null_hypothesis = 'µ2mg - µ0.5mg = 0' ) %>% select(9, 1:8) %>% bind_rows(pairwise_results, .)

pairwise_results <- t.test(ToothGrowth$len[ToothGrowth$dose == 1], ToothGrowth$len[ToothGrowth$dose == 0.5]) %>% tidy %>% mutate( null_hypothesis = 'µ1mg - µ0.5mg = 0' ) %>% select(9, 1:8) %>% bind_rows(pairwise_results, .)

print.data.frame(pairwise_results)

“
```

In all three cases the p-value is smaller than the significance level 0.05 which means we reject the null hypothesis each time. The data provides strong evidence that the average tooth length of guinea pigs is different for each dose level.

Interpreting the first row of the table for illustration:

We are 95% confident that the average tooth length of guinea pigs who received a 2mg dose of vitamin C is on average `r round(pairwise_results$conf.low[1], 2)` to `r round(pairwise_results$conf.high[1], 2)` millimeters higher than those who received a 1mg dose.

```
“{r eval=FALSE}

library(ggplot2) library(dplyr) library(broom) library(wesanderson) library(grid)
```

Tables

```
with(ToothGrowth, table(dose, supp))
```

```
ToothGrowth %>% group_by(dose) %>% summarize( q25 = quantile(len, 0.25), q75 = quantile(len, 0.75) ) %>% as.data.frame
```

Generate plots

```
g_len_by_supp <- ggplot(ToothGrowth, aes(supp, len)) + geom_boxplot() + xlab('Supplement type') + ylab('Tooth length (mm)') + ggtitle('Tooth length by supplement type')
```

```
g_len_by_dose <- ggplot(ToothGrowth, aes(as.factor(dose), len)) + geom_boxplot() + xlab('Dose (mg)') + ylab('Tooth length (mm)') + ggtitle('Tooth length by dose level')
```

```
ggplot(ToothGrowth, aes(as.factor(dose), len)) + geom_boxplot(aes(fill = as.factor(dose))) + facet_grid(. ~ supp) + xlab('Dose (mg)') + ylab('Tooth length (mm)') + ggtitle('Tooth length by dose level and supplement type') + scale_fill_manual(values = wes_palette("FantasticFox")) + theme(legend.position = "none")
```

T-test for difference in tooth length means across supplement types

```
t_diff_supp <- t.test(len ~ supp, ToothGrowth, var.equal = FALSE) tidy(t_diff_supp)
```

Pairwise t-tests for differences in tooth lengths means across dose levels

```
pairwise_results <- t.test(ToothGrowthlen[ToothGrowthdose == 2], ToothGrowthlen[ToothGrowthdose == 1]) %>% tidy %>% mutate( null_hypothesis = 'μ2mg - μ1mg = 0' ) %>% select(9, 1:8)
```

```
pairwise_results <- t.test(ToothGrowthlen[ToothGrowthdose == 2], ToothGrowthlen[ToothGrowthdose == 0.5]) %>% tidy %>% mutate( null_hypothesis = 'μ2mg - μ0.5mg = 0' ) %>% select(9, 1:8) %>% bind_rows(pairwise_results, .)
```

```
pairwise_results <- t.test(ToothGrowthlen[ToothGrowthdose == 1], ToothGrowthlen[ToothGrowthdose == 0.5]) %>% tidy %>% mutate( null_hypothesis = 'μ1mg - μ0.5mg = 0' ) %>% select(9, 1:8) %>% bind_rows(pairwise_results, .)
```

```
print.data.frame(pairwise_results)
```

““

Conclusions

There are clear indications that both the supplement as the dosage have clear independent effects on the length of teeth guinea pigs. More those means on average longer teeth. Supplement type has a clear influence too, but OJ has a greater average teethgrowth in combination with dosages 0.5 and 1 then for the VC supplement, while teeth length for the VC supplement vs the OJ in combination with dosage 2 has no significant effect (almost same mean & same confidence interval)

The fact remains however that these assumptions are based on the facts:

- that the guinea pigs are representative for the population of guinea pigs,
- that dosage and supplement were randomly assigned and
- that the distribution of the means is normal.