

ToothGrow Inferential Analysis

Jayesh Gokhale

4/18/2021

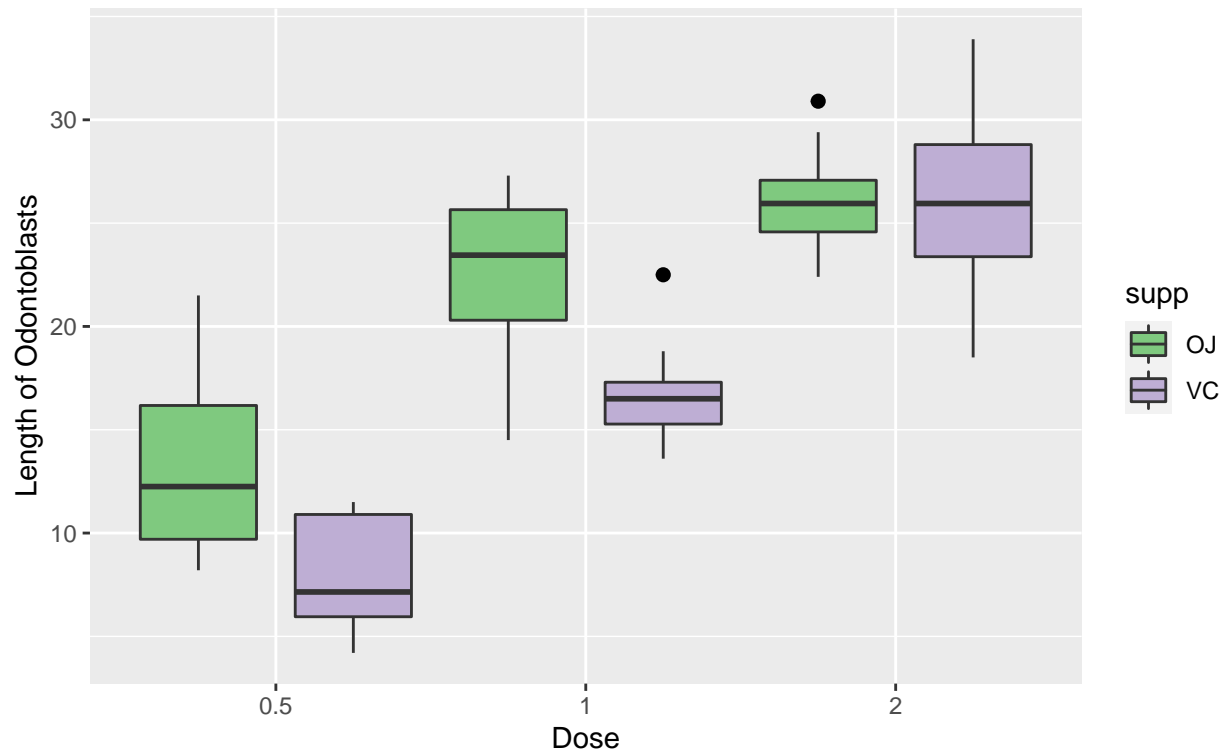
ToothGrow Dataset Inferential Analysis

This is an analysis on the ToothGrow dataset in library datasets. In short this dataset is a study on length of Odontoblasts in 60 guinea pigs where each animal received a certain dose (0.5,1 or 2 mg/day) of Vitamin C through one of the two delivery methods viz. Orange Juice (OJ) or Ascorbic Acid Vitamin C (VC)

Let us first load the dataset and perform some basic EDA.

```
tg <- ToothGrowth
tg$dose <- as.factor(tg$dose)
ggplot(data=tg, aes(x=dose,y=len,fill=supp)) +
  geom_boxplot(outlier.colour="black",
               outlier.size=2,position=position_dodge(1)) +
  labs(title = "Tooth Grow Data Set",
        subtitle = "",
        y = "Length of Odontoblasts", x = "Dose") +
  scale_fill_brewer(palette="Accent") +
  theme(plot.title = element_text(hjust = 0.5))
```

Tooth Grow Data Set



So we can see that visually the Length increases with increase in dosage and generally the length is higher for OJ (Orange Juice) for Dosage = 0.5 and 1 as compared to VC (Ascorbic Acid Vitamin C). For Dosage = 2, we can see that the median length for both OJ and VC seem to be the same but there is a high amount of variance with VC as compared to OJ.

Let us confirm this with Hypothesis Testing. We have two things to confirm.

1. Increase in Dose causes an Increase in Length of Odontoblasts across both Delivery Methods (OJ and VC)
2. Across each Dose, OJ provides higher increase in length.

```
tg$supp.dose <- paste0(tg$supp,".",tg$dose)
tg$supp.dose <- as.factor(tg$supp.dose)
supp.doses <- levels(tg$supp.dose)
n.supp.dose <- length(supp.doses)

mat.comparison <- matrix(rep(0,n.supp.dose**2),n.supp.dose,n.supp.dose)
rownames(mat.comparison) <- supp.doses
colnames(mat.comparison) <- supp.doses
for (i in 1:n.supp.dose) mat.comparison[i,i] <- "X"

comb.supp.doses <- combn(levels(tg$supp.dose),2)
comb.supp.pvalues <- rbind(comb.supp.doses,
  apply(comb.supp.doses,2,function(x)
  {
    t.test(len~supp.dose,data=tg[tg$supp.dose %in% x,])$p.value
  })
```

```

))
comb.supp.pvalues <- as.data.frame(t(comb.supp.pvalues))
colnames(comb.supp.pvalues) <- c("Method.1", "Method.2", "P.Value")
comb.supp.pvalues$P.Value <- as.numeric(levels(comb.supp.pvalues$P.Value)[comb.supp.pvalues$P.Value])
comb.supp.pvalues$VERDICT <- with(comb.supp.pvalues, ifelse(P.Value < 0.05, "SIGNIFICANT",
  ifelse((P.Value >= 0.05) & (P.Value < 0.1), "BORDERLINE", "INSIGNIFICANT")))
color.significant <- which(comb.supp.pvalues$VERDICT=="SIGNIFICANT")
color.borderline <- which(comb.supp.pvalues$VERDICT=="BORDERLINE")
color.insignificant <- which(comb.supp.pvalues$VERDICT=="INSIGNIFICANT")
comb.supp.pvalues$VERDICT <- as.factor(comb.supp.pvalues$VERDICT)

```

Let us study the Pair-wise T Test Statistics

Table 1: Pair wise Testing

Method.1	Method.2	P.Value	VERDICT
OJ.0.5	OJ.1	0.0000878	SIGNIFICANT
OJ.0.5	OJ.2	0.0000013	SIGNIFICANT
OJ.0.5	VC.0.5	0.0063586	SIGNIFICANT
OJ.0.5	VC.1	0.0460103	SIGNIFICANT
OJ.0.5	VC.2	0.0000072	SIGNIFICANT
OJ.1	OJ.2	0.0391951	SIGNIFICANT
OJ.1	VC.0.5	0.0000000	SIGNIFICANT
OJ.1	VC.1	0.0010384	SIGNIFICANT
OJ.1	VC.2	0.0965261	BORDERLINE
OJ.2	VC.0.5	0.0000000	SIGNIFICANT
OJ.2	VC.1	0.0000002	SIGNIFICANT
OJ.2	VC.2	0.9638516	INSIGNIFICANT
VC.0.5	VC.1	0.0000007	SIGNIFICANT
VC.0.5	VC.2	0.0000000	SIGNIFICANT
VC.1	VC.2	0.0000916	SIGNIFICANT

Conclusion

1. The finding matches with the Box Plot.
2. There is no statistical significance to suggest that for a fixed Dose = 2, the delivery method viz. OJ/VC has any specific impact
3. While the Dose increases from 1 to 2, the Delivery method if changed from OJ to VC, there is no statistical significance to suggest that there is any change in length of Odontoblasts.
4. Apart from the above exceptions, a general trend is that with increase in dosage, length of Odontoblasts increases and Orange Juice (OJ) has greater impact on length as compared with Ascorbic Acid Vitamin C (VC)