

Coursera Statistical Inference Course

Final Project - Part 2

Darrell Gerber

4/12/2021

Part 2: Basic Inferential Data Analysis

1.1 Introduction

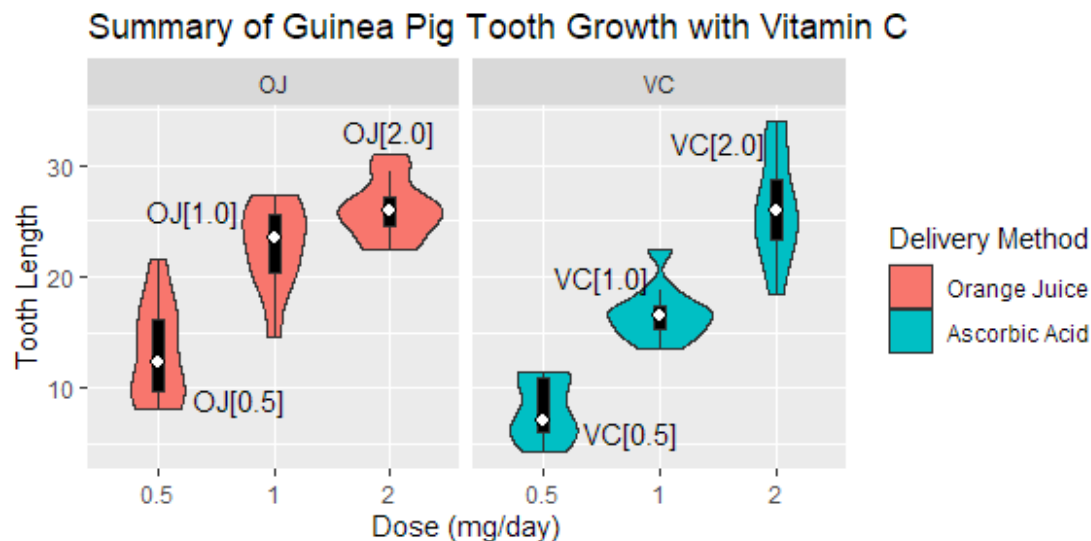
The second part of the project analyzes Vitamin C's influence on tooth growth in Guinea Pigs under two factors: three dose levels (0.5, 1, and 2 mg/day) and two delivery methods (orange juice or ascorbic acid).

Reviewer NOTE: Many of the code chunks used to produce this report are long. Not only does this cause the report to exceed the page limit, but it also makes the report more difficult to read. For those reasons, I do not show the code in the report's main body (i.e., `echo=FALSE`). The Appendix contains all code chunks, however.

1.2 Exploratory Analysis

An initial investigation reveals that the data contains 60 observations split evenly between each delivery type and dosing level. [code 1.2-1]

Var1	Var2	Freq
OJ	0.5	10
VC	0.5	10
OJ	1	10
VC	1	10
OJ	2	10
VC	2	10



A combined box plot and violin plot show that, no matter the delivery method, tooth length increases with increasing Vitamin C dosage. For doses of 0.5 and 1.0 mg/day, orange juice appears to be a more effective delivery method than ascorbic acid. However, orange juice shows a response drop-off at the 2.0 mg/day dosage ($OJ_{2.0}$). All observations are pretty compact, although $OJ_{0.5}$, $OJ_{2.0}$, $VC_{0.5}$, and $VC_{1.0}$ show non-normal distributions. [code 1.2-2]

The table below summarizes the key statistics for each observation group [code 1.2-3]

dose	meanOJ	meanVC	varOJ	varVC
0.5	13.23	7.98	19.89	7.54
1.0	22.70	16.77	15.30	6.33
2.0	26.06	26.14	7.05	23.02

1.3 Hypothesis Testing

Two effects we would like to investigate are:

1. the effect of dose on tooth length, and
2. the effect of delivery method on tooth length.

To determine the effect of dose on tooth length, test four hypotheses comparing two dosage responses for each delivery method.

$$H_{01.1}: \mu_{OJ_{0.5}} - \mu_{OJ_{1.0}} = 0 \text{ and } H_{a1.1}: \mu_{OJ_{0.5}} - \mu_{OJ_{1.0}} <> 0$$

$$H_{01.2}: \mu_{OJ_{1.0}} - \mu_{OJ_{2.0}} = 0 \text{ and } H_{a1.2}: \mu_{OJ_{1.0}} - \mu_{OJ_{2.0}} <> 0$$

$$H_{01.3}: \mu_{VC_{0.5}} - \mu_{VC_{1.0}} = 0 \text{ and } H_{a1.3}: \mu_{VC_{0.5}} - \mu_{VC_{1.0}} <> 0$$

$$H_{01.4}: \mu_{VC_{1.0}} - \mu_{VC_{2.0}} = 0 \text{ and } H_{a1.4}: \mu_{VC_{1.0}} - \mu_{VC_{2.0}} <> 0$$

Similarly, test the following three hypotheses to determine the significance of delivery method at each dosage.

$$H_{02.1}: \mu_{OJ_{0.5}} - \mu_{VC_{0.5}} = 0 \text{ and } H_{a2.1}: \mu_{OJ_{0.5}} - \mu_{VC_{0.5}} <> 0$$

$$H_{02.2}: \mu_{OJ_{1.0}} - \mu_{VC_{1.0}} = 0 \text{ and } H_{a2.2}: \mu_{OJ_{1.0}} - \mu_{VC_{1.0}} <> 0$$

$$H_{02.3}: \mu_{OJ_{2.0}} - \mu_{VC_{2.0}} = 0 \text{ and } H_{a2.3}: \mu_{OJ_{2.0}} - \mu_{VC_{2.0}} <> 0$$

Assume a normally distributed population and the observation sets are independent and unpaired. Also, assume that the variances for each data set are not equal.

1.3.1 Influence of Vitamin C dose on tooth length

Perform a Welch Two Sample t-test and determine the 95% Confidence Interval and the p-value for hypotheses $H_{0.1.1}$, $H_{0.1.2}$, $H_{0.1.3}$, and $H_{0.1.4}$: [code 1.3.1-1]

H0	conflow	confhigh	pval
1.1	5.524	13.416	0.0000878491906
1.2	0.189	6.531	0.0391951420462
1.3	6.314	11.266	0.0000006811018
1.4	5.686	13.054	0.0000915560306

The 95% Confidence Intervals for all four tests are positive. We can conclude with 95% confidence that:

1. The Null Hypothesis in all cases can be rejected, and
2. Increasing doses of Vitamin C have a positive influence on tooth length.

The p-values for $H_{0.1.1}$, $H_{0.1.3}$, and $H_{0.1.4}$ indicate an exceptionally low probability that tooth length difference is not due to the Vitamin C dosage. However, the p-value for $H_{0.1.2}$ is significantly higher. It is sufficient to reject the Null Hypothesis with 95% confidence but shows that the dose-response for orange juice is weaker with the higher dosage.

1.3.2 Influence of delivery method on tooth length

Perform a Welch Two Sample t-test and determine the 95% Confidence Interval and the p-value for hypotheses $H_{0.2.1}$, $H_{0.2.2}$, and $H_{0.2.3}$: [code 1.3.2-1]

H0	conflow	confhigh	pval
2.1	-8.781	-1.719	0.006358607
2.2	-9.058	-2.802	0.001038376
2.3	-3.638	3.798	0.963851589

The 95% Confidence Intervals for $H_{0.2.1}$ and $H_{0.2.2}$ are negative. We can conclude with 95% confidence that:

1. The Null Hypothesis (there is no difference in tooth length due to delivery method) can be rejected, and
2. The delivery of Vitamin C via Orange Juice has a more substantial influence on tooth length than the ascorbic acid delivery method.

The p-values for $H_{0.2.1}$ and $H_{0.2.2}$ indicate a low probability that tooth length difference is not due to the Vitamin C delivery method.

However, the 95% Confidence Interval for $H_{0.2.3}$ spans the Null Hypothesis and the p-value is very high. We can not reject $H_{0.2.2}$ and must conclude it is most likely there is no difference in tooth length between the two delivery methods at the 2.0 mg/day dosage.


```

varOJ0.5 <- var(OJ0.5)
meanOJ0.5 <- mean(OJ0.5)
OJ1.0 <- ToothGrowth[which(ToothGrowth$supp=="OJ" &
                           ToothGrowth$dose==1.0),]$len
varOJ1.0 <- var(OJ1.0)
meanOJ1.0 <- mean(OJ1.0)
OJ2.0 <- ToothGrowth[which(ToothGrowth$supp=="OJ" &
                           ToothGrowth$dose==2.0),]$len
varOJ2.0 <- var(OJ2.0)
meanOJ2.0 <- mean(OJ2.0)
VC0.5 <- ToothGrowth[which(ToothGrowth$supp=="VC" &
                           ToothGrowth$dose==0.5),]$len
varVC0.5 <- var(VC0.5)
meanVC0.5 <- mean(VC0.5)
VC1.0 <- ToothGrowth[which(ToothGrowth$supp=="VC" &
                           ToothGrowth$dose==1.0),]$len
varVC1.0 <- var(VC1.0)
meanVC1.0 <- mean(VC1.0)
VC2.0 <- ToothGrowth[which(ToothGrowth$supp=="VC" &
                           ToothGrowth$dose==2.0),]$len
varVC2.0 <- var(VC2.0)
meanVC2.0 <- mean(VC2.0)

stats <- data.frame(dose=c(0.5, 1.0, 2.0))
stats$meanOJ <- c(meanOJ0.5, meanOJ1.0, meanOJ2.0)
stats$meanVC <- c(meanVC0.5, meanVC1.0, meanVC2.0)
stats$varOJ <- round(c(varOJ0.5, varOJ1.0, varOJ2.0), 2)
stats$varVC <- round(c(varVC0.5, varVC1.0, varVC2.0), 2)
ft <- flextable(stats, cwidth = 1)
autofit(ft)

```

[code 1.3.1-1]

```

doseH <- data.frame(H=c("1.1", "1.2", "1.3", "1.4"),
                    conflow=c(NA, NA, NA, NA),
                    confhigh=c(NA, NA, NA, NA),
                    pval=c(NA, NA, NA, NA))
doseH[1,]$conflow <- round(t.test(OJ1.0, OJ0.5, paired=FALSE, var.equal =
FALSE)$conf.int[1], 3)
doseH[1,]$confhigh <- round(t.test(OJ1.0, OJ0.5, paired=FALSE, var.equal =
FALSE)$conf.int[2], 3)
doseH[1,]$pval <- t.test(OJ1.0, OJ0.5, paired=FALSE, var.equal = FALSE)$p.value

doseH[2,]$conflow <- round(t.test(OJ2.0, OJ1.0, paired=FALSE, var.equal =
FALSE)$conf.int[1], 3)
doseH[2,]$confhigh <- round(t.test(OJ2.0, OJ1.0, paired=FALSE, var.equal =
FALSE)$conf.int[2], 3)
doseH[2,]$pval <- t.test(OJ2.0, OJ1.0, paired=FALSE, var.equal = FALSE)$p.value

doseH[3,]$conflow <- round(t.test(VC1.0, VC0.5, paired=FALSE, var.equal =
FALSE)$conf.int[1], 3)
doseH[3,]$confhigh <- round(t.test(VC1.0, VC0.5, paired=FALSE, var.equal =
FALSE)$conf.int[2], 3)
doseH[3,]$pval <- t.test(VC1.0, VC0.5, paired=FALSE, var.equal = FALSE)$p.value

doseH[4,]$conflow <- round(t.test(VC2.0, VC1.0, paired=FALSE, var.equal =
FALSE)$conf.int[1], 3)
doseH[4,]$confhigh <- round(t.test(VC2.0, VC1.0, paired=FALSE, var.equal =
FALSE)$conf.int[2], 3)
doseH[4,]$pval <- t.test(VC2.0, VC1.0, paired=FALSE, var.equal = FALSE)$p.value

```

```
flextable(doseH, cwidth = 1.25)
```

[code 1.3.2-1]

```
suppH <- data.frame(H0=c("2.1", "2.2", "2.3"),
                    conflow=c(NA,NA,NA),
                    confhigh=c(NA,NA,NA),
                    pval=c(NA,NA,NA))
suppH[1,]$conflow <- round(t.test(VC0.5, OJ0.5, paired=FALSE, var.equal =
FALSE)$conf.int[1],3)
suppH[1,]$confhigh <- round(t.test(VC0.5, OJ0.5, paired=FALSE, var.equal =
FALSE)$conf.int[2],3)
suppH[1,]$pval <- t.test( VC0.5, OJ0.5,paired=FALSE, var.equal = FALSE)$p.value

suppH[2,]$conflow <- round(t.test( VC1.0, OJ1.0,paired=FALSE, var.equal =
FALSE)$conf.int[1],3)
suppH[2,]$confhigh <- round(t.test(VC1.0, OJ1.0, paired=FALSE, var.equal =
FALSE)$conf.int[2],3)
suppH[2,]$pval <- t.test(VC1.0, OJ1.0, paired=FALSE, var.equal = FALSE)$p.value

suppH[3,]$conflow <- round(t.test(VC2.0, OJ2.0, paired=FALSE, var.equal =
FALSE)$conf.int[1],3)
suppH[3,]$confhigh <- round(t.test(VC2.0, OJ2.0, paired=FALSE, var.equal =
FALSE)$conf.int[2],3)
suppH[3,]$pval <- t.test(VC2.0, OJ2.0, paired=FALSE, var.equal = FALSE)$p.value

flextable(suppH, cwidth = 1.25)
```