

HW2 Yijin Wang

1.Ignore 'dose' and determine whether there is a significant difference in the mean "len" between the two groups (i.e., OJ vs VC): a) A parametric procedure b) A non-parametric procedure c) Discuss the assumption underlying each of the analyses, their validity, and any remedial measures to be taken.

```
data(ToothGrowth)
OJ<-ToothGrowth$len[ToothGrowth$supp == 'OJ']
VC<-ToothGrowth$len[ToothGrowth$supp == 'VC']
#a) parametric procedure
t.test(OJ, VC)

##
##  Welch Two Sample t-test
##
## data:  OJ and VC
## t = 1.9153, df = 55.309, p-value = 0.06063
## alternative hypothesis: true difference in means is not equal to 0
## 95 percent confidence interval:
##  -0.1710156  7.5710156
## sample estimates:
## mean of x mean of y
##  20.66333  16.96333

#As the p-value is greater than 0.05, we fail to reject the null hypothesis.
#We can conclude that there is no significant difference in
#mean "len" between 2 groups.

#b) nonparametric procedure
wilcox.test(OJ, VC)

## Warning in wilcox.test.default(OJ, VC): cannot compute exact p-value with ties
##
##  Wilcoxon rank sum test with continuity correction
##
## data:  OJ and VC
## W = 575.5, p-value = 0.06449
## alternative hypothesis: true location shift is not equal to 0

#As the p-value is greater than 0.05,
#we fail to reject the null hypothesis.
#We can conclude that supplement OJ has the same effect
#on tooth growth as supplement VC.

#c)
#Parametric test
#Assumptions:
#1. Normal distribution 2. X and Y independent random samples

#Validity:
```

```

#shapiro.test

#Remedial Methods:
#1. Trimmed means - remove outliers
#2. Winsorized variances -
#outliers are "pulled" towards the center of the distribution

#Nonparametric test
#Assumptions:
#1. Random independent samples
#2. Mann-Whitney Test - two samples have same shape
#3. Wilcoxon Signed-Rank Test - symmetric distribution
#4. ANOVA- same shape and equal variances

#Validity:
#F test to compare 2 variances

#Remedial Methods:
#Welch's modified two-sample t test

```

2. Assume that if "len" is above 20, it is classified as "HIGH"; and "LOW", otherwise.
 - a) Ignore 'dose', and determine whether there is a significant difference in the proportions of the two groups classified as "HIGH" using a suitable test and a 95% confidence interval.
 - b) Repeat the above, taking into account "dose", and comment on the results.

```

#a
OJ_HIGH<-OJ[OJ > 20]
OJ_LOW<-OJ[OJ <= 20]
VC_HIGH<-VC[VC > 20]
VC_LOW<-VC[VC <= 20]
temp1<-c(length(OJ_HIGH),length(VC_HIGH))
temp2<-c(length(OJ),length(VC))
prop.test(temp1, temp2)

##
## 2-sample test for equality of proportions with continuity correction
##
## data: temp1 out of temp2
## X-squared = 3.2812, df = 1, p-value = 0.07008
## alternative hypothesis: two.sided
## 95 percent confidence interval:
## -0.009950289 0.543283623
## sample estimates:
## prop 1 prop 2
## 0.6000000 0.3333333

#Since p-value is greater than 0.05, we fail to reject null hypothesis.
#Thus, there is no significant difference in the proportions of the two groups classified as "HIGH"
#95% confidence interval: (-0.009950289 0.543283623)

#b
#dose=0.5
OJ_0.5<-ToothGrowth$len[ToothGrowth$supp == "OJ" & ToothGrowth$dose == 0.5]
VC_0.5<-ToothGrowth$len[ToothGrowth$supp == "VC" & ToothGrowth$dose == 0.5]

```

```

OJ_1<-ToothGrowth$len[ToothGrowth$supp == "OJ" & ToothGrowth$dose == 1]
VC_1<-ToothGrowth$len[ToothGrowth$supp == "VC" & ToothGrowth$dose == 1]

OJ_2<-ToothGrowth$len[ToothGrowth$supp == "OJ" & ToothGrowth$dose == 2]
VC_2<-ToothGrowth$len[ToothGrowth$supp == "VC" & ToothGrowth$dose == 2]

HIGH_OJ_0.5<-OJ_0.5[OJ_0.5>20]
LOW_OJ_0.5<-OJ_0.5[OJ_0.5<=20]
HIGH_VC_0.5<-VC_0.5[VC_0.5>20]
LOW_VC_0.5<-VC_0.5[VC_0.5<=20]

HIGH_OJ_1<-OJ_1[OJ_1>20]
LOW_OJ_1<-OJ_1[OJ_1<=20]
HIGH_VC_1<-VC_1[VC_1>20]
LOW_VC_1<-VC_1[VC_1<=20]

HIGH_OJ_2<-OJ_2[OJ_2>20]
LOW_OJ_2<-OJ_2[OJ_2<=20]
HIGH_VC_2<-VC_2[VC_2>20]
LOW_VC_2<-VC_2[VC_2<=20]

data <- array(c(length(HIGH_OJ_0.5),length(LOW_OJ_0.5),
                 length(HIGH_VC_0.5),length(LOW_VC_0.5),
                 length(HIGH_OJ_1),length(LOW_OJ_1),
                 length(HIGH_VC_1),length(LOW_VC_1),
                 length(HIGH_OJ_2),length(LOW_OJ_2),
                 length(HIGH_VC_2),length(LOW_VC_2)), c(2,2,3))
mantelhaen.test(data)

```

```

##
## Mantel-Haenszel chi-squared test with continuity correction
##
## data: data
## Mantel-Haenszel X-squared = 6.9478, df = 1, p-value = 0.008392
## alternative hypothesis: true common odds ratio is not equal to 1
## 95 percent confidence interval:
## 2.117045 361.562605
## sample estimates:
## common odds ratio
## 27.66667

```

```

#The p-value 0.008392 is lower than 0.05, so we reject the null hypothesis.
#We conclude that there is a significant difference in the proportions
#of the two groups classified as "HIGH" when taking "dose" into account.

```