lab test solution Code

Statistics (University of Melbourne)

MAST20005/MAST90058: Computer Lab Test Solutions Appendix

```r
# Load the data.
HExer <- read.table("HExer.txt", header = TRUE)$HExer

# Q1
summary(HExer)

##    Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##   6.080   7.375  10.310   9.526  11.570  12.560

# Q2
t.test(HExer, conf.level = 0.90)

##
##  One Sample t-test
##
## data:  HExer
## t = 18.408, df = 19, p-value = 1.43e-13
## alternative hypothesis: true mean is not equal to 0
## 90 percent confidence interval:
##   8.631188 10.420812
## sample estimates:
## mean of x
##     9.526

# Q3
# Any of the following are acceptable.
wilcox.test(HExer, mu = 10)

##
##  Wilcoxon signed rank test
##
## data:  HExer
## V = 80, p-value = 0.3683
## alternative hypothesis: true location is not equal to 10

wilcox.test(HExer, mu = 10, exact = FALSE)

##
##  Wilcoxon signed rank test with continuity correction
##
## data:  HExer
## V = 80, p-value = 0.3604
## alternative hypothesis: true location is not equal to 10

wilcox.test(HExer, mu = 10, exact = FALSE, correct = FALSE)

##
##  Wilcoxon signed rank test
##
## data:  HExer
## V = 80, p-value = 0.3507
## alternative hypothesis: true location is not equal to 10
```

```r
# Q4
# Any of the following are acceptable.
prop.test(sum(HExer > 10), length(HExer), alternative = "greater")

##
##  1-sample proportions test without continuity correction
##
## data:  sum(HExer > 10) out of length(HExer), null probability 0.5
## X-squared = 0, df = 1, p-value = 0.5
## alternative hypothesis: true p is greater than 0.5
## 95 percent confidence interval:
##  0.3274038 1.0000000
## sample estimates:
##   p
## 0.5

binom.test(sum(HExer > 10), length(HExer), alternative = "greater")

##
##  Exact binomial test
##
## data:  sum(HExer > 10) and length(HExer)
## number of successes = 10, number of trials = 20, p-value = 0.5881
## alternative hypothesis: true probability of success is greater than 0.5
## 95 percent confidence interval:
##  0.3019539 1.0000000
## sample estimates:
## probability of success
##                    0.5

# Q5
x <- table(cut(HExer, c(-Inf, 7, 10, Inf)))
x

##
##  (-Inf,7]    (7,10] (10, Inf]
##         5         5        10

chisq.test(x)

##
##  Chi-squared test for given probabilities
##
## data:  x
## X-squared = 2.5, df = 2, p-value = 0.2865
```

```r
# Load the data.
salesdata <- read.csv("sales.csv", header = TRUE)
daysofweek <- c("Mon", "Tue", "Wed", "Thu", "Fri", "Sat", "Sun")
salesdata$days <- factor(salesdata$days, daysofweek)
salesdata

##    customers sales days
## 1        230   934  Mon
## 2        179   760  Tue
## 3        134   728  Wed
## 4        237   599  Thu
## 5        149   395  Fri
## 6        315   634  Sat
## 7        335  1063  Sun
## 8        230  1267  Mon
## 9        169   234  Tue
## 10       220   523  Wed
## 11       140   596  Thu
## 12       187   763  Fri
## 13       229   635  Sat
## 14       287  1111  Sun
## 15       190   543  Mon
## 16       265   750  Tue
## 17       155   566  Wed
## 18       209   444  Thu
## 19       213   601  Fri
## 20       209   332  Sat
## 21       311  1328  Sun

# Q6
mfit <- lm(sales ~ customers, salesdata)
summary(mfit)

##
## Call:
## lm(formula = sales ~ customers, data = salesdata)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -351.90 -159.38  -78.29  170.79  529.03
##
## Coefficients:
##             Estimate Std. Error t value Pr(>|t|)
## (Intercept)  67.0893   212.9286   0.315  0.75614
## customers     2.9169     0.9425   3.095  0.00597 **
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 244.3 on 19 degrees of freedom
## Multiple R-squared:  0.3351,Adjusted R-squared:  0.3001
## F-statistic: 9.577 on 1 and 19 DF,  p-value: 0.005965

# Q7
confint(mfit, level = 0.95)
```

```
##                   2.5 %     97.5 %
## (Intercept) -378.5753071 512.753990
## customers      0.9441196   4.889594

# Q8
newdata <- data.frame(customers = 200)
predict(mfit, newdata, interval = "prediction", level = 0.9)

##        fit      lwr      upr
## 1 650.4607 216.9488 1083.973

# Q9
anova(lm(customers ~ factor(days), salesdata))

## Analysis of Variance Table
##
## Response: customers
##              Df Sum Sq Mean Sq F value  Pr(>F)
## factor(days)  6  41994  6998.9  3.8867 0.01705 *
## Residuals    14  25211  1800.8
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

# Q10
on.weekend <- factor(salesdata$days %in% c("Sat", "Sun"))
t.test(customers ~ on.weekend, salesdata)

##
##  Welch Two Sample t-test
##
## data:  customers by on.weekend
## t = -3.776, df = 7.5397, p-value = 0.006041
## alternative hypothesis: true difference in means is not equal to 0
## 95 percent confidence interval:
##  -141.02435  -33.37565
## sample estimates:
## mean in group FALSE  mean in group TRUE
##               193.8               281.0

t.test(customers ~ on.weekend, salesdata, var.equal = TRUE)

##
##  Two Sample t-test
##
## data:  customers by on.weekend
## t = -4.2293, df = 19, p-value = 0.000454
## alternative hypothesis: true difference in means is not equal to 0
## 95 percent confidence interval:
##  -130.35457  -44.04543
## sample estimates:
## mean in group FALSE  mean in group TRUE
##               193.8               281.0

anova(lm(customers ~ on.weekend, salesdata))
```

```
## Analysis of Variance Table
##
## Response: customers
##            Df Sum Sq Mean Sq F value   Pr(>F)
## on.weekend  1  32588   32588  17.887 0.000454 ***
## Residuals  19  34616    1822
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```