Unit 7 – R for Analysis of One & Two Samples
Homework
SOLUTIONS

```
# Preliminary – load sepsis.Rdata.
setwd("/Users/cbigelow/Desktop")
load(file = "sepsis.Rdata")
```

```
# Q1 – Paired t test
library(dplyr)
library(stargazer)

q1data <- sepsis %>%                          # With dataset sepsis DO
    filter(race==2) %>%                       # Keep only race==2 observations. THEN DO
    select(temp0,temp1) %>%                   # Retain only the variables of interest.  THEN DO
    mutate(chg_temp=temp0-temp1,na.rm=TRUE)   # Create chg_temp = temp0 - temp1

q1data <- as.data.frame(q1data)               # What could go wrong:  stargazer works on dataframes only
stargazer(q1data[c("temp0","temp1", "chg_temp")],type="text",summary.stat=c("n", "mean", "sd", "min",
"p25", "median", "p75", "max"))

t.test(q1data$temp0,q1data$temp1, paired=TRUE)
```

```
===========================================================================
Statistic N   Mean   St. Dev.  Min   Pctl(25) Median  Pctl(75)   Max
---------------------------------------------------------------------------
temp0     32 100.740  2.127   95.900  99.845  100.760 102.000  105.800
temp1     28 99.981   1.799   96.260  98.575  99.900  101.075  104.600
chg_temp  28  0.662   1.182   -2.000  -0.185   0.570   1.295    3.400
---------------------------------------------------------------------------

         Paired t-test

data:  q1data$temp0 and q1data$temp1
t = 2.9648, df = 27, p-value = 0.006263
alternative hypothesis: true difference in means is not equal to 0
95 percent confidence interval:
 0.2038967 1.1203896
sample estimates:
mean of the differences
              0.6621432
```

INTERPRETATION:
In the sub-group of 32 who are classified as race='other', data on temp0 and temp1 were complete for n=28, or 87.5%. Among these n=28, the mean temperatures at baseline and two-hours were 100.7 and 99.9, respectively. The mean and standard deviation of the baseline to two-hour change in temperature were 0.67 and 1.2, respectively. A two-sided paired t-test of the null hypothesis that the mean change $\mu_d = 0$ yielded a paired t-statistic value of 2.96 and associated two sided p-value = 0.006. This is statistically significant. The null hypothesis is rejected. Conclude that these data provide statistically significant evidence that the change in temperature between the baseline and 2-hour occasions of measurement $\mu_d \neq 0$.

 # Q2 – 90% CI for change
 # Note: R calculates change as [var_first] – [var_second] = [temp0] – [temp1]

```
# Two solutions

t.test(q1data$temp0,q1data$temp1, conf.level=0.90, paired=TRUE)$conf.int   # from t.test output

q2ci <- q1data %>%                                                         # by "hand"
    dplyr::summarise (
      n=sum(!is.na(chg_temp)),
      dfreedom=n-1,
      tcrit=qt(.95,df=dfreedom),
      ave=mean(chg_temp,na.rm=TRUE),
      sd=sd(chg_temp,na.rm=TRUE),
      se=sd/sqrt(n),
      lower95 = ave - tcrit*se,
      upper95=ave + tcrit*se,
      varname="Temp0 - Temp1") %>%
      select(varname,n,ave,lower95,upper95)
q2ci
```

```
[1] 0.2817385 1.0425478                                  #  Solution I – from t.test output
attr(,"conf.level")
[1] 0.9


        varname  n       ave   lower95  upper95          #  Solution II – by "hand".  Phew - matches.
1 Temp0 - Temp1 28 0.6621432 0.2817385 1.042548
```

INTERPRETATION:
Based on this sample of n=28, with 90% confidence, the unknown mean change in temperature baseline → two hours among persons with race classification 'other' is estimated to be between 0.28 and 1.04.

 # Q3 Two independent samples t-test: Data in WIDE format

```
library(FSA)

sepsis$treatf <- factor(sepsis$treat,                  # create factor variable treatf from treat
    levels = c(0,1),
    labels = c("Placebo", "Ibuprofen"))


FSA::Summarize(apache~treatf,data=sepsis,na.rm=TRUE)    # get descriptives by group
var.test(data=sepsis, apache~treatf)                    # preliminary test of equal variances
t.test(data=sepsis,apache~treatf, var.equal=TRUE)       # 2 indep samples t test - equal variances
```

```
     treatf   n nvalid     mean       sd min Q1 median Q3 max  percZero        # desriptives by group
1   Placebo 231    230 15.18696 6.922831   0 10   14.5 19  41 0.4347826
2 Ibuprofen 224    224 15.47768 7.261882   3 10   14.0 21  37 0.0000000


       F test to compare two variances                          # Preliminary test of equal variances

data:  apache by treatf
F = 0.9088, num df = 229, denom df = 223, p-value = 0.4724      # null of equal variances NOT rejected
alternative hypothesis: true ratio of variances is not equal to 1
95 percent confidence interval:
 0.6995696 1.1800701
sample estimates:
ratio of variances
         0.9088018

       Two Sample t-test                                    # 2 indep samples t test – equal variances

data:  apache by treatf
t = -0.43668, df = 452, p-value = 0.6626
alternative hypothesis: true difference in means is not equal to 0
95 percent confidence interval:
 -1.599088  1.017644
sample estimates:
  mean in group Placebo mean in group Ibuprofen
               15.18696                15.47768
```

INTERPRETATION:
In this cohort of n_0 = 230 treated with placebo and n_1 = 224 treated with ibuprofen, the mean APACHE scores were 15.19 (s_1 = 6.9) and 15.48 (s_2 = 7.3), respectively. A two sample t-test of the null hypothesis of equality of means yielded a t-statistic value of -0.44 and associated two sided p-value = 0.66. This is not statistically significant. The null hypothesis is not rejected. Conclude that these data do not provide statistically significant evidence of a treatment effect on APACHE score.

```
# Q4 – 95% CI for difference of two independent means
# Note:  R calculates change as [var_first] – [var_second] = [temp0] – [temp1]
```

```
t.test(data=sepsis,apache~treatf, conf.level=0.95,var.equal=TRUE)$conf.int       # from t-test output
```

```
[1] -1.599088  1.017644
attr(,"conf.level")
[1] 0.95
```

INTERPRETATION:
Based on this cohort of n_0 = 230 treated with placebo and n_1 = 224 treated with ibuprofen, respectively, with 95% confidence, the unknown difference in APACHE score (placebo minus ibuprofen) is estimated to be between -1.60 and +1.02.