ps9.R

benjamin

2020-12-01

```r
## Problem Set 9
## Ben Reichert

## 1
cat("1.\n ")
```

```
## 1.
##
```

```r
# a)
cat("a) This data should be analyzed using a one-sample t-test because there are two sets of data,\n",
    "we are taking into account the difference between the two, which is in fact one set of data.\n",
    "And for analyzing one sample population or set of data we use the one-sample t-test.\n ")
```

```
## a) This data should be analyzed using a one-sample t-test because there are two sets of data,
##  we are taking into account the difference between the two, which is in fact one set of data.
##  And for analyzing one sample population or set of data we use the one-sample t-test.
##
```

```r
# b)
decafm <- 53
decafsd <- 19
cofm <- 41.5
cofsd <- 17
n <- 10

delta.hat <- decafm - cofm
se <- sqrt(decafsd*decafsd/n + cofsd*cofsd/n)
t.welch = delta.hat/se
nu = ( (decafsd^2/n) + (cofsd^2/n) )^2 / ( ((decafsd^2)/n)^2/(n-1) + ((cofsd^2/n)^2)/(n-1) )
p.value <- 2*(1 - pt(abs(t.welch), df = nu))
cat("b) When people with type-2 diabetes consume dates with cofee...\n",
    "Null hypothesis: Glycemic index is not affected.\n",
    "Alternative hypothesis: Glycemic index is different when coffee is consumed.\n",
    "t-statistic =",t.welch,", P-value =",p.value,"\n ")
```

```
## b) When people with type-2 diabetes consume dates with cofee...
##  Null hypothesis: Glycemic index is not affected.
##  Alternative hypothesis: Glycemic index is different when coffee is consumed.
##  t-statistic = 1.426399 , P-value = 0.171075
##
```

```r
# c)
difm <- 11.5
difsd <- 21
```

```r
lower <- difm - qt(.975, df = n-1)* difsd/sqrt(n)
upper <- difm + qt(.975, df = n-1)* difsd/sqrt(n)
#q <-qt(0.975, df=nu)
#lower <- delta.hat-q*se
#upper <- delta.hat+q*se
cat("c) 95% confidence interval: (",lower,",",upper,")\n ")
```

```
## c) 95% confidence interval: ( -3.522495 , 26.5225 )
##
```

```r
# d)
cat("d) I think it is only safe to say that within this sample set (of people with type-2 diabetes)\n",
    "we can conclude that on average dates will produce a similar glycemic index with or without coffee
    "However, I do NOT think it would be correct to state that on average, no matter what the sample\n"
    "population might be, that dates would produce the same glycemic index with or without coffee.\n\n"]
```

```
## d) I think it is only safe to say that within this sample set (of people with type-2 diabetes)
##   we can conclude that on average dates will produce a similar glycemic index with or without coffee.
##   However, I do NOT think it would be correct to state that on average, no matter what the sample
##   population might be, that dates would produce the same glycemic index with or without coffee.
## 2 Trosset 11.4 Problem Set C, parts 1-3
cat("2.\n ")
```

```
## 2.
##
```

```r
# 1
# a)
cat("a) Experimental unit: Cholesterol level\n ")
```

```
## a) Experimental unit: Cholesterol level
##
```

```r
# b)
cat("b) Experimental units drawn from 2 different populations:\n",
    "Type A: urgent, aggressive, and ambitious heavy, middle aged men.\n",
    "Type B: noncompetitive, more relaxed, and less hurried heavy, middle aged men.\n",
    "20 units drawn from each population.\n",
    "2-sample problem.\n ")
```

```
## b) Experimental units drawn from 2 different populations:
##   Type A: urgent, aggressive, and ambitious heavy, middle aged men.
##   Type B: noncompetitive, more relaxed, and less hurried heavy, middle aged men.
##   20 units drawn from each population.
##   2-sample problem.
##
```

```r
# c
cat("c) Only one measurement was taken on each experimental unit, cholesterol level.\n ")
```

```
## c) Only one measurement was taken on each experimental unit, cholesterol level.
##
```

```r
# d
typea <- c(233,291, 312, 250, 246, 197, 268, 224, 239,239,
           254,276, 234, 181, 248, 252, 202, 218, 212,325)
typeb <- c(344,185, 263, 246, 224, 212, 188, 250, 148,169,
```

```r
          226,175, 242, 252, 153, 183, 137, 202, 194,213)
typec <-typea-typeb
print(t.test(typea, typeb))
```

```
##
##  Welch Two Sample t-test
##
## data:  typea and typeb
## t = 2.5621, df = 35.413, p-value = 0.01481
## alternative hypothesis: true difference in means is not equal to 0
## 95 percent confidence interval:
##   7.227071 62.272929
## sample estimates:
## mean of x mean of y
##    245.05    210.30
```

```r
cat("d) Parameters of interest:\n",
    "Type A mean =",mean(typea),"\n",
    "Type B mean =",mean(typeb),"\n",
    "Delta hat =",mean(typea)-mean(typeb),"\n ")
```

```
## d) Parameters of interest:
##  Type A mean = 245.05
##  Type B mean = 210.3
##  Delta hat = 34.75
##
```

```r
# e
cat("e) For heavy middle aged men...\n",
    "Null hypothesis: Cholesterol levels remain the same despite personality differences.\n",
    "Alternative hypothesis: More aggressive, higher intensity individuals have higher cholesterol\n",
    "levels and thus a higher chance coronary heart disease.\n ")
```

```
## e) For heavy middle aged men...
##  Null hypothesis: Cholesterol levels remain the same despite personality differences.
##  Alternative hypothesis: More aggressive, higher intensity individuals have higher cholesterol
##  levels and thus a higher chance coronary heart disease.
##
```

```r
# 2
plot(density(typea))
lines(density(typeb), col="red")
```

density.default(x = typea)

N = 20   Bandwidth = 12.91
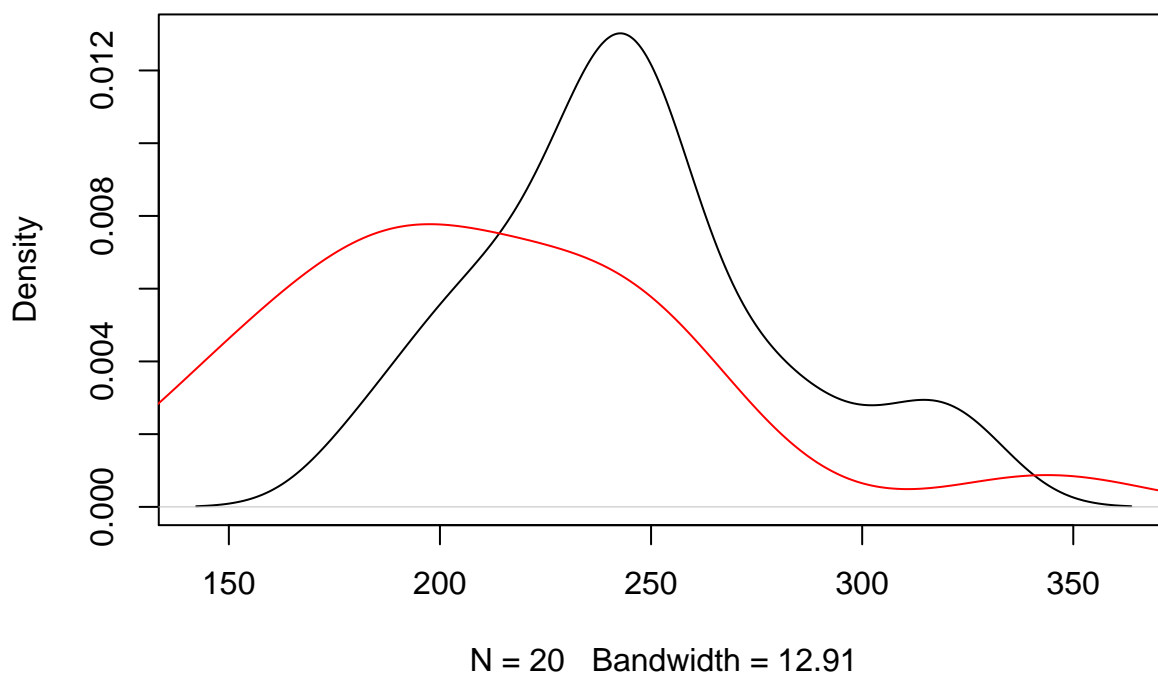
```r
cat("2) After seeing the density plots for the two sets of data it is safe to say that neither of them\n
    "are normally distributed.\n ")
```

```
## 2) After seeing the density plots for the two sets of data it is safe to say that neither of them
##  are normally distributed.
##
```

```r
# 3
typeam <- mean(typea)
typeasd <- sd(typea)
typebm <- mean(typeb)
typebsd <- sd(typeb)
na <- length(typea)
nb <- length(typeb)

delta.hat2 <- typeam - typebm
se2 <- sqrt(typeam*typeasd/na + typebm*typebsd/nb)
t.welch2 = delta.hat2/se2
nu2 = ( (typeasd^2/na) + (typebsd^2/nb) )^2 / ( ((typeasd^2)/na)^2/(na-1) + ((typebsd^2/nb)^2)/(nb-1) )
p.value2 <- 2*(1 - pt(abs(t.welch2), df = nu2))
cat("3)\n ",
    "a) P-value = 0.01481\n ",
    "If we adopt a significance level of 5% we should not reject the null hypothesis.\n  ")
```

```
## 3)
##   a) P-value = 0.01481
##   If we adopt a significance level of 5% we should not reject the null hypothesis.
##
```

```r
difm <- mean(typec)
difsd <- sd(typec)
```

```r
lower <- difm - qt(.95, df = n-1)* difsd/sqrt(n)
upper <- difm + qt(.95, df = n-1)* difsd/sqrt(n)
cat("b) 90% confidence interval: (",lower,",",upper,")\n")
```

```
## b) 90% confidence interval: ( -0.4387703 , 69.93877 )
```

```r
## 3 Trosset 11.4 Problem Set D, parts 1-4
norm <- c(4.1, 6.3, 7.8, 8.5, 8.9, 10.4, 11.5, 12, 13.8, 17.6, 24.3, 37.2)
diab <- c(11.5,12.1,16.1,17.8,24,28.8,33.9,40.7,51.3,56.2,61.7,69.2)
cat("3.\n ")
```
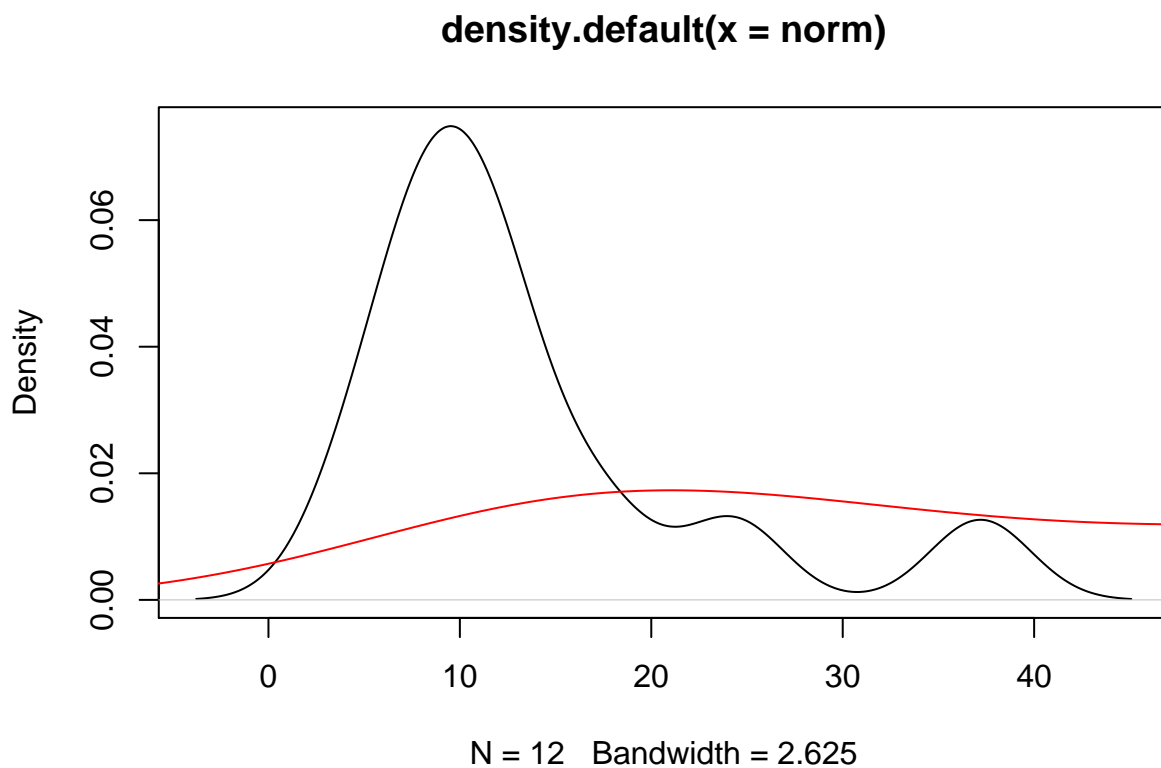
```
## 3.
##
```

```r
# 1
plot(density(norm))
lines(density(diab), col="red")
```

density.default(x = norm)

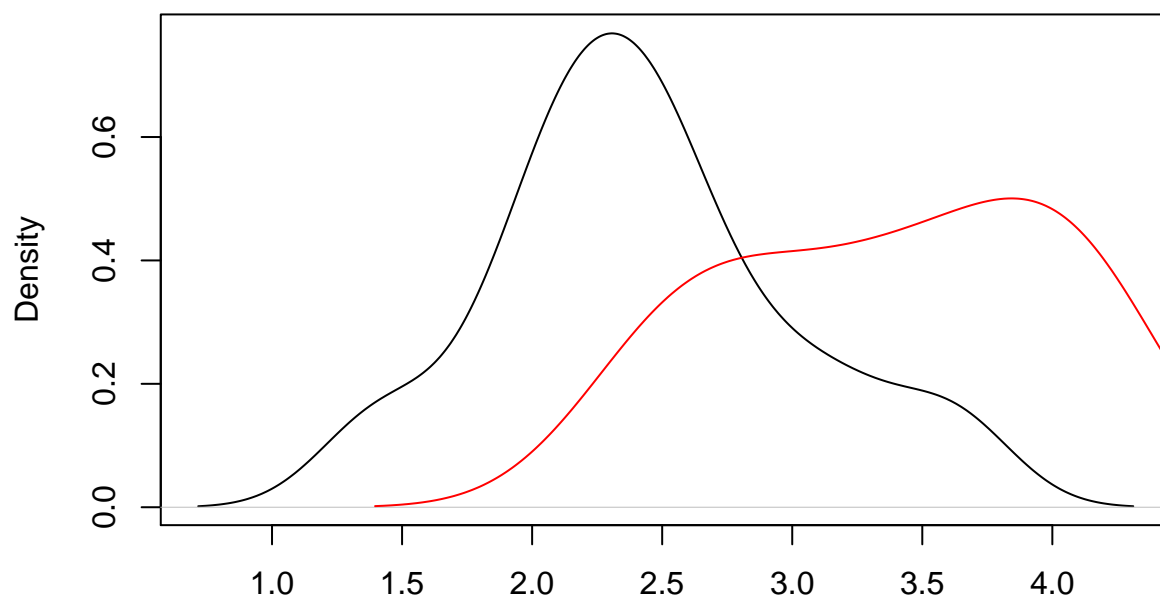N = 12   Bandwidth = 2.625

```r
cat("1) No, according to their graphs, these data sets do not look like symmetric distributions.\n",
    "They are lop sided.")
```

```
## 1) No, according to their graphs, these data sets do not look like symmetric distributions.
##  They are lop sided.
```

```r
# 2
normlog<-log(norm)
diablog<-log(diab)
plot(density(normlog))
lines(density(diablog), col="red")
```
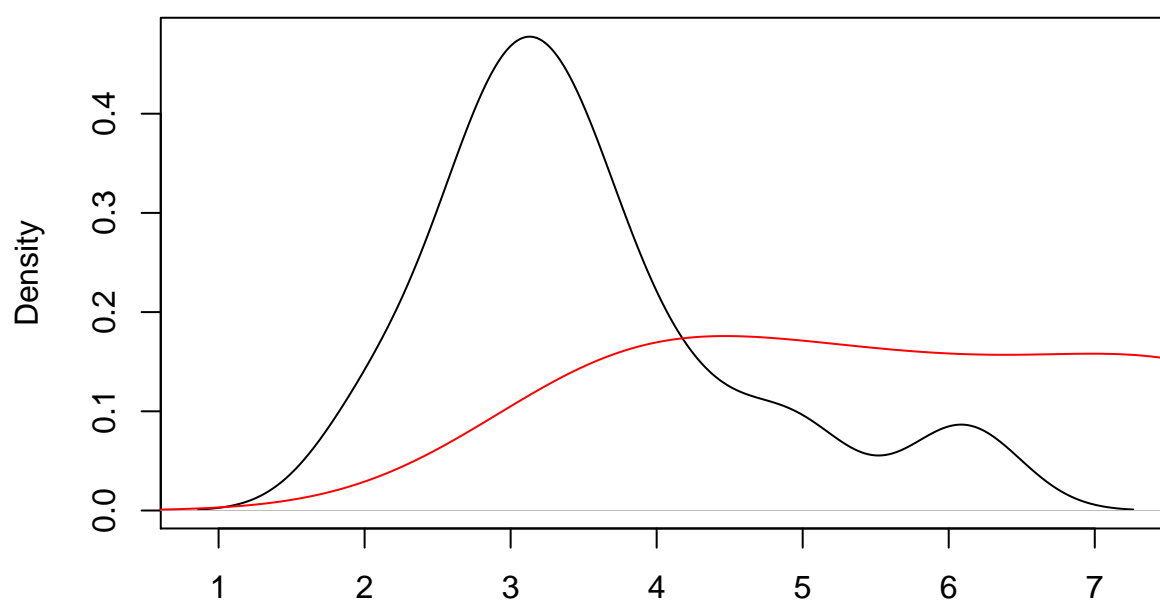
density.default(x = normlog)

N = 12   Bandwidth = 0.2316

```
normsr<-sqrt(norm)
diabsr<-sqrt(diab)
plot(density(normsr))
lines(density(diabsr), col="red")
```

density.default(x = normsr)

N = 12   Bandwidth = 0.3882

```
cat("2) The transformed measurements look slightly more symmetric. Especially the natural logarithm\n",
    "transformation of the normal control group.\n ")
```

```
## 2) The transformed measurements look slightly more symmetric. Especially the natural logarithm
##  transformation of the normal control group.
##
```

```
# 3
normlog<-log(norm)
diablog<-log(diab)
plot(density(normlog))
lines(density(diablog), col="red")
```

density.default(x = normlog)

N = 12   Bandwidth = 0.2316
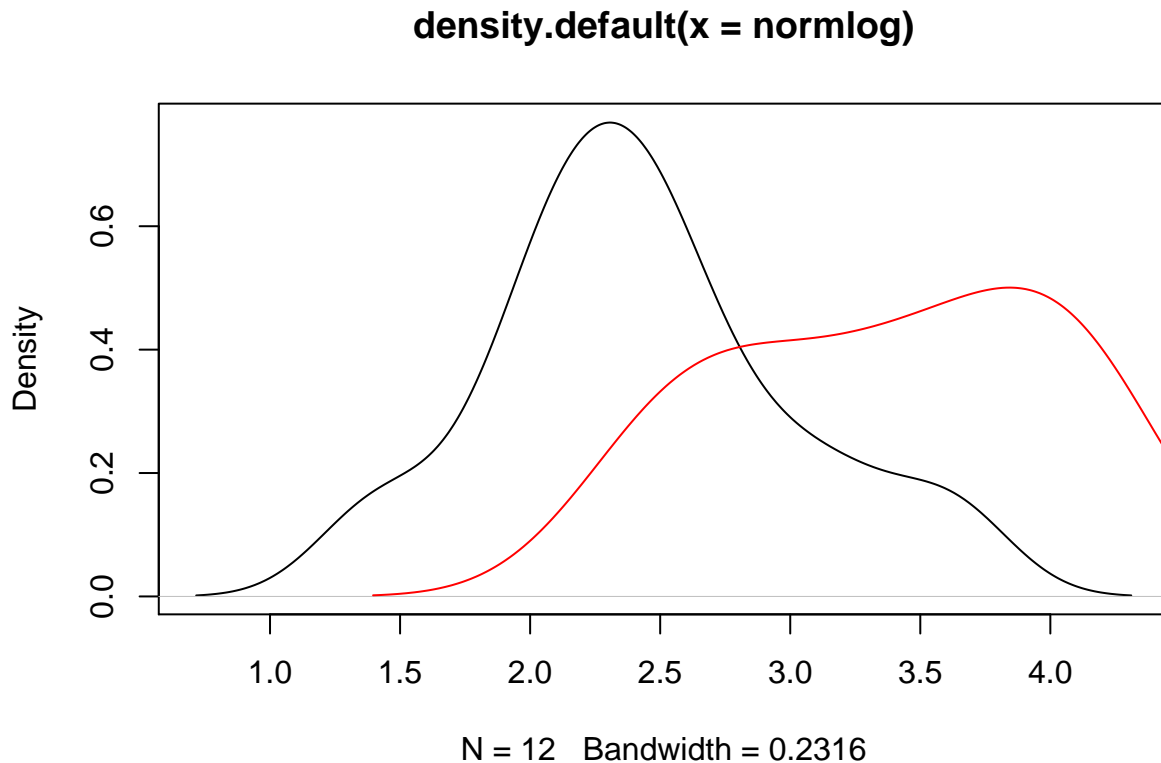
```
cat("3) The transformed measurements do look to be relatively normal, their graphs form near spot on\n"
    "bell curves.\n ")
```

```
## 3) The transformed measurements do look to be relatively normal, their graphs form near spot on
##  bell curves.
##
```

```
# 4
cat("4)\n ")
```

```
## 4)
##
```

```
print(t.test(normlog, diablog))
```

```
##
##  Welch Two Sample t-test
##
## data:  normlog and diablog
```

```
## t = -3.8041, df = 21.9, p-value = 0.0009776
## alternative hypothesis: true difference in means is not equal to 0
## 95 percent confidence interval:
##  -1.4792986 -0.4352589
## sample estimates:
## mean of x mean of y
##  2.433349  3.390628
```

```
cat(" Because the p-value is 0.0009776 which is < 0.05, it is safe to reject the null hypothesis and\n"
    "accept the alternative that the measurements of urinary beta-thromboglobulin excretion is conclusiv
    "higher in people with diabetes-12.")
```

```
##  Because the p-value is 0.0009776 which is < 0.05, it is safe to reject the null hypothesis and
##  accept the alternative that the measurements of urinary beta-thromboglobulin excretion is conclusive
##  higher in people with diabetes-12.
```