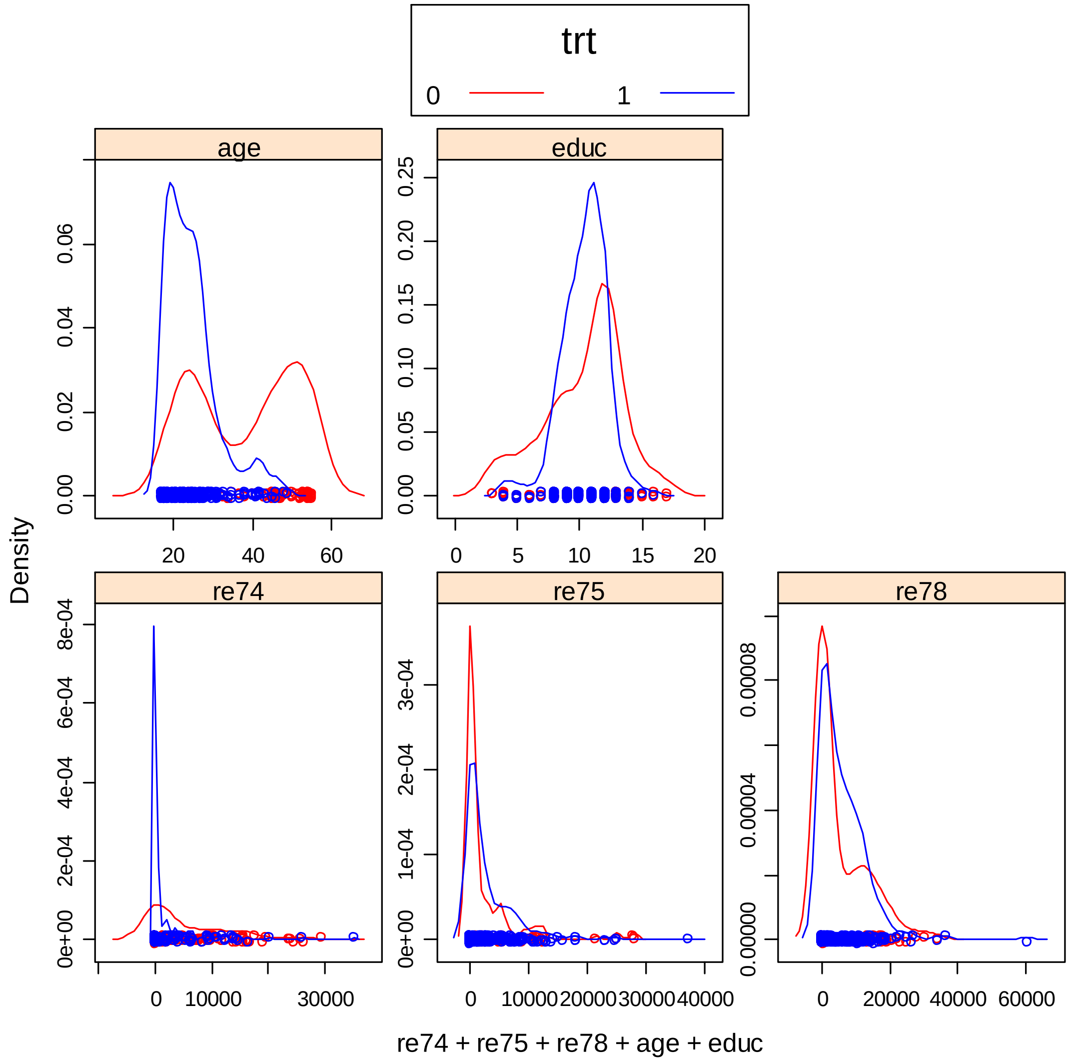
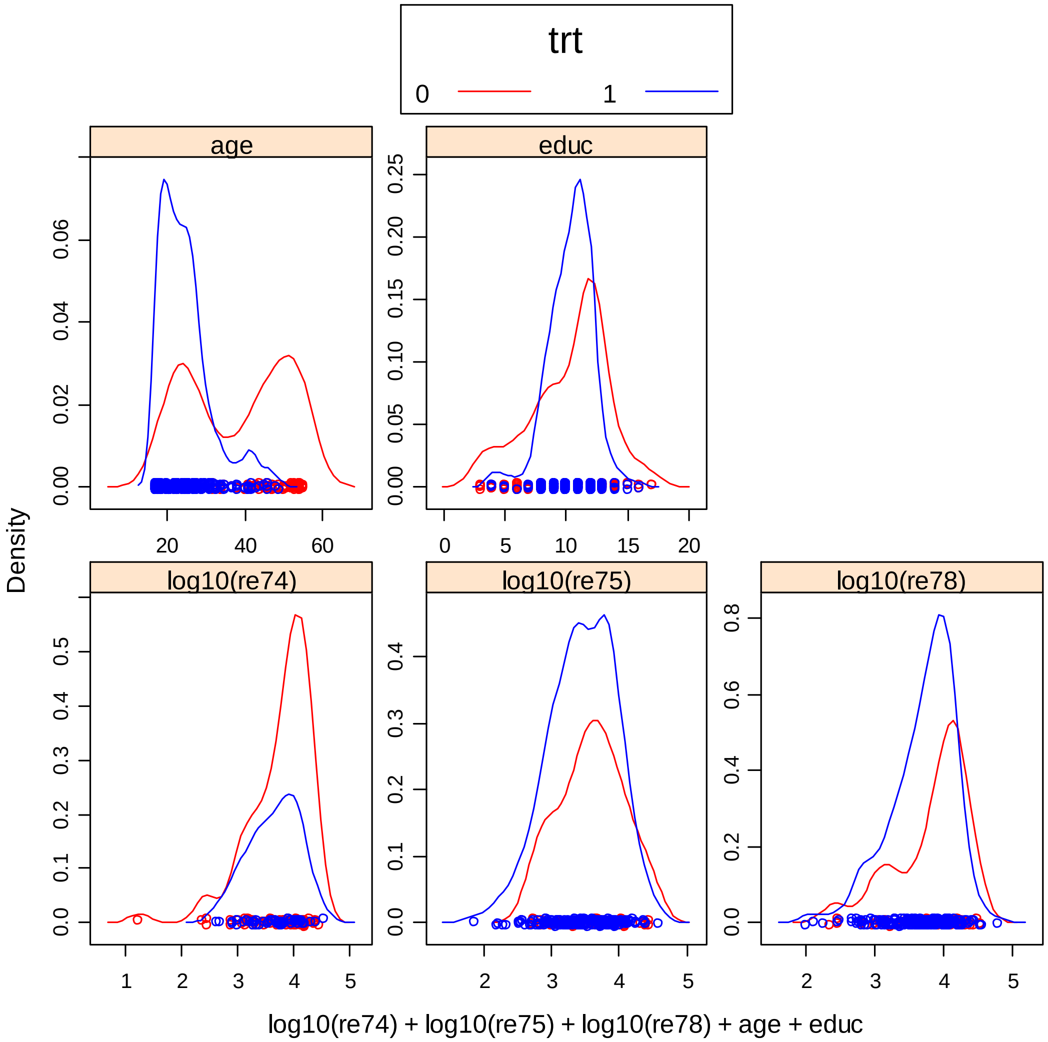
Stat 462: Lab 5 solutions

```
# 4.1
# a)
# Define function to compute confidence intervals
  t.ci <- function(x, cl){ t.test(x, conf.level = cl)$conf.int }
# Apply the confidence interval function to each group using
aggregate()
  nswdemo.ci <- aggregate( re75 ~ trt, data = nswdemo, FUN = t.ci, cl
= 0.95 )
# Pretty up the names for printing (not strictly necessary)
  colnames( nswdemo.ci$re75 ) <- c("lower.95","upper.95")
  print( nswdemo.ci )

  trt re75.lower.95 re75.upper.95
1   0      2530.773      3522.593
2   1      2509.407      3622.789

# b)
  nswdemo.ci <- aggregate( re78 ~ trt, data = nswdemo, FUN = t.ci, cl
= 0.95 )
  colnames( nswdemo.ci$re78 ) <- c( "lower.95","upper.95")
  print( nswdemo.ci )

  trt re78.lower.95 re78.upper.95
1   0      4544.861      5635.236
2   1      5185.685      6767.019

# c)
  t.test( re78 ~ trt, data = nswdemo )$conf.int

[1] -1845.25051    72.64305
attr(,"conf.level")
[1] 0.95
```

```
# 4.8
nswpsid3 <- rbind( psid3, subset( nswdemo, trt==1 ) )
# Note that nswpsid3 compares treatment to a non-experimental
# "control" group whereas nswdemo compares treatment to an
# experimental control group
# a)
library( lattice )
# Change colors to something that a colour blind guy could see
par.settings <- list( superpose.symbol =
                      list( col = c("red", "blue")
                          , fill = c("red", "blue") )
                    , superpose.line =
                          list( col = c("red", "blue") ) )
densityplot( ~ re74 + re75 + re78 + age + educ , data = nswpsid3
           , groups = trt, scales = "free"
           , par.settings = par.settings
           , auto.key = list(columns=2, title="trt"
                  , border = TRUE, padding.text = 4 )
           )
```

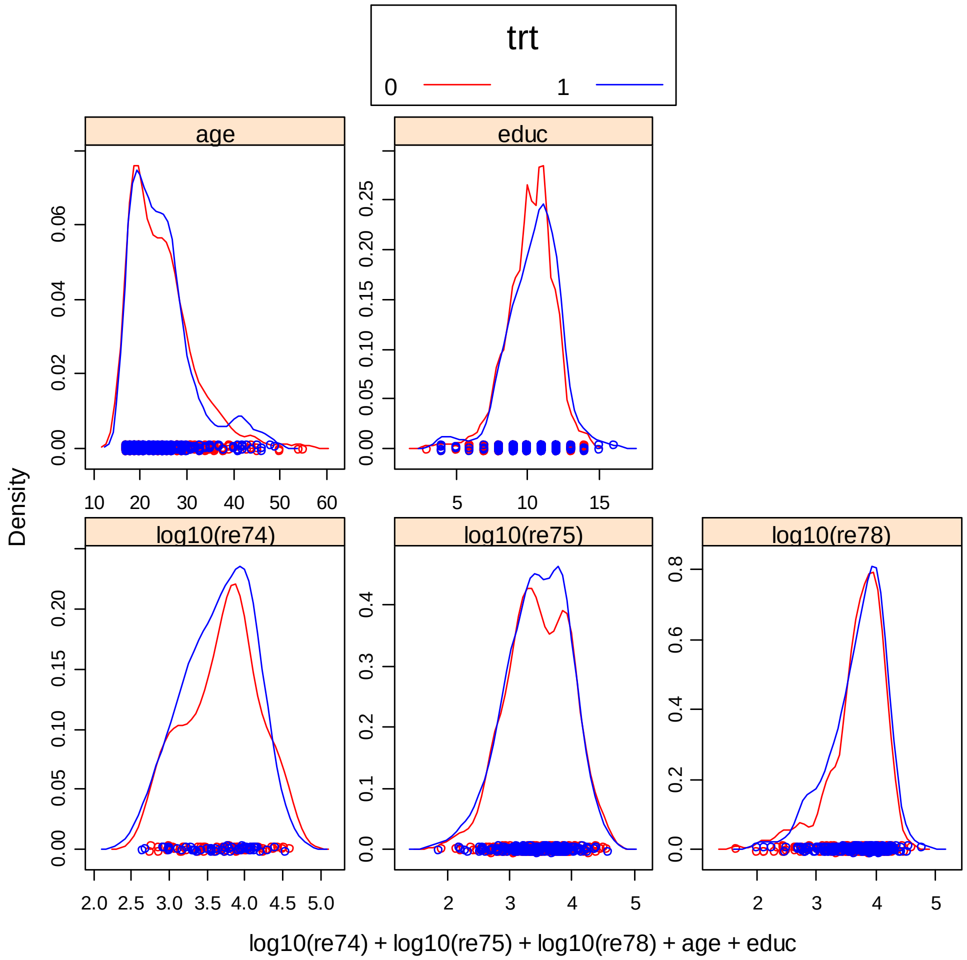


```
# Try log earnings to see differences more clearly
densityplot( ~ log10(re74) + log10(re75) + log10(re78) + age + educ
           , data = nswpsid3, groups = trt, scales = "free"
           , par.settings = par.settings
           , auto.key = list(columns=2, title="trt"
           , border = TRUE, padding.text = 4) )
```



```
# Compute two-way tables for the binary variables
with( nswpsid3, table( black, trt ) )
     trt
black   0   1
    0  70  59
    1  58 238
with( nswpsid3, table( hisp, trt ) )
    trt
hisp   0   1
   0 113 269
   1  15  28
```

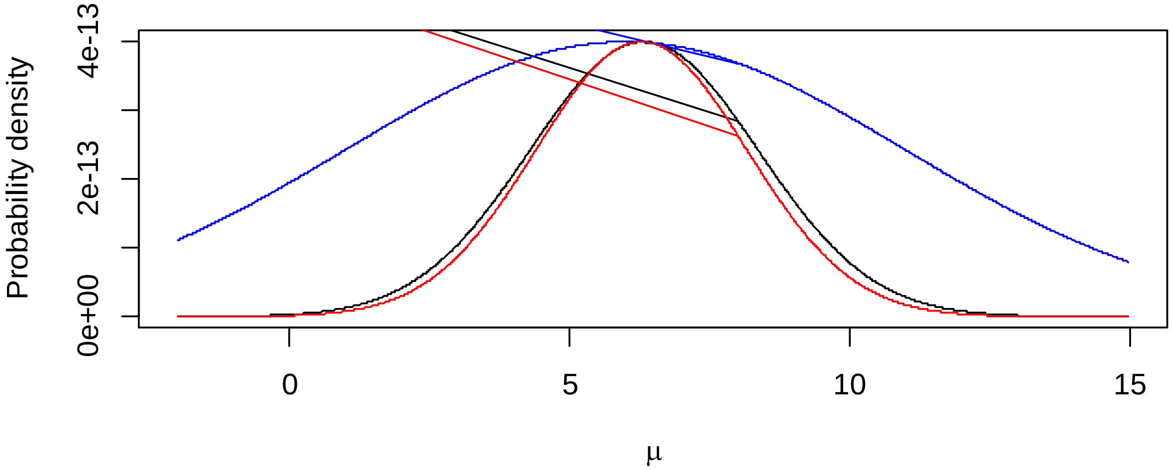
```
with( nswpsid3, table( marr, trt ) )
    trt
marr   0   1
   0  39 247
   1  89  50
with( nswpsid3, table( nodeg, trt ) )
     trt
nodeg   0   1
    0  63  80
    1  65 217

# There are very large differences in the number of blacks,
# the number of married, and the number with no degree.
# There is also a large difference in the distribution of ages.

# b) Now the nswdemo data
# Again use log earnings to see differences more clearly
densityplot( ~ log10(re74) + log10(re75) + log10(re78) + age + educ
           , data = nswdemo, groups = trt, scales = "free"
           , par.settings = par.settings
           , auto.key = list(columns=2, title="trt"
           , border = TRUE, padding.text = 4) )
```

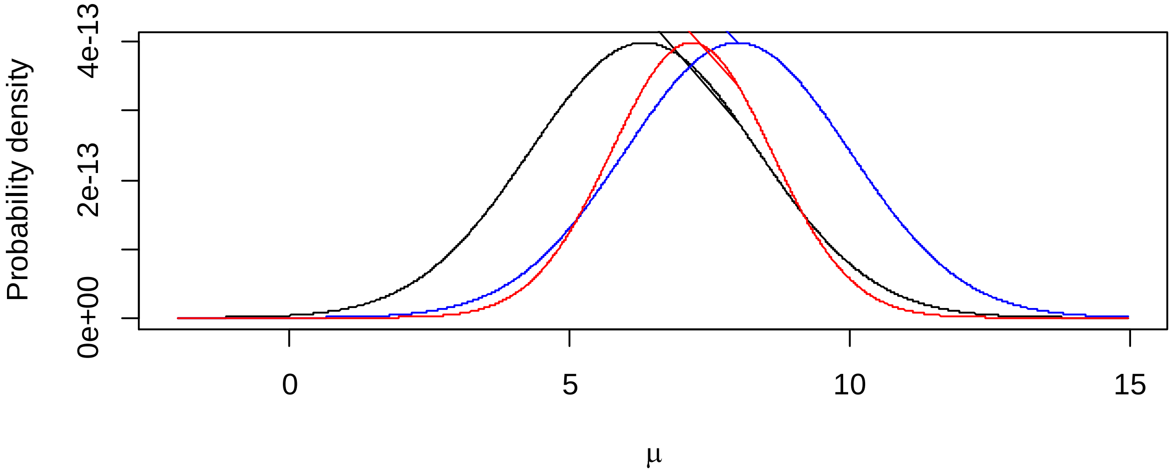


```
# Compute two-way tables for the binary variables
with( nswdemo, table( black, trt ) )
     trt
black   0   1
    0  85  59
    1 340 238
with( nswdemo, table( hisp, trt ) )
    trt
hisp   0   1
   0 377 269
   1  48  28
with( nswdemo, table( marr, trt ) )
    trt
marr   0   1
   0 358 247
   1  67  50
with( nswdemo, table( nodeg, trt ) )
     trt
nodeg   0   1
    0  79  80
    1 346 217

# We see that the nswdemo control group is much better-matched
# to the treatment group than the psid3 "control" group

# 4.9
# a)
# Create the table as a matrix
acup <- matrix( c(74, 71, 43, 38, 11, 65), ncol = 3
        , dimnames = list( c( "LargeReduction", "SmallReduction" )
                     , c( "Acupuncture", "Sham", "WaitingList" ) )
        )
# Perform chi-square test
chisq.test( acup )

        Pearson's Chi-squared test

data:  acup
X-squared = 32.4862, df = 2, p-value = 8.825e-08

# Looks like Reduction and Treatment are NOT independent

# b)
# Create the table as a matrix
guess <- matrix( c(82, 17, 30, 30, 26, 16), ncol = 2
         , dimnames = list( c( "Chinese", "Other", "DoNotKnow" )
                     , c( "Acupuncture", "Sham" ) )
        )
# Perform chi-square test
chisq.test( guess )
```

```
        Pearson's Chi-squared test

data:  guess
X-squared = 15.3584, df = 2, p-value = 0.0004624

# Looks like the guess and the treatment are NOT independent.

# Although there seemed to be a relationship between the
# treatment and reduction in migraines, the fact that some
# patients knew they might be getting sham treatment is a
# weakness of the design. This is especially troubling since
# there was an apparent relationship between the treatment
# that the patients received and the treatment that they
# thought they were getting.
# The fact that the largest difference was between the
# waiting list and those who received any acupuncture (sham
# or otherwise) suggests that some of the migraine relief
# may be due to a placebo effect.

# 4.14
library(boot)
z.transform <- function(r) 0.5*log((1+r)/(1-r))
z.inverse <- function(z) (exp(2*z)-1)/(exp(2*z)+1)
possum.fun <- function( data, indices ){
  chest <- data$chest[indices]
  belly <- data$belly[indices]
z.transform( cor(belly, chest) )
}
possum.boot <- boot(possum, possum.fun, R=999)
z.inverse(boot.ci(possum.boot, type="perc")$percent[4:5])
[1] 0.4920185 0.7072572

# Now for the confidence interval without transform
possum.fun2 <- function( data, indices ){
  chest <- data$chest[indices]
  belly <- data$belly[indices]
cor(belly, chest)
}
possum.boot2 <- boot(possum, possum.fun2, R=999)
boot.ci(possum.boot2, type="perc")$percent[4:5]
[1] 0.4734588 0.7105285

# The two methods give nearly the same confidence
# interval for the correlation coefficient.
```

```
# 4.22
like <- function( mu, sigma, x=with(pair65, heated-ambient) ){
  prod( dnorm(x, mu, sigma) )
}
# Compute the likelihood for mu given sigma is 6.1
muvals <- seq( -2, 15, 0.01 )
like.mu <- sapply( muvals, FUN = like, sigma = 6.1 )

par( mfrow = c(2,1) )
# Plot the likelihood
plot( muvals, like.mu, xlab = expression(mu)
     , ylab = "Probability density", type="l" )
# Add the prior
# Get re-scaled prior density
prior <- dnorm( muvals, 6, 5 )
prior.scaled <- prior/max(prior)*max(like.mu)
lines( muvals, prior.scaled, col = "blue" )
# Add the posterior
n <- nrow( pair65 )
ybar <- with( pair65, mean( heated - ambient ) )
s.s0 <- ( 6.1^2 )/( 5^2 )
mu.post <- ( n*ybar + 6*s.s0 )/( n + s.s0 )
sig.post <- sqrt( ( 6.1^2 )/( n + s.s0 ) )

post <- dnorm( muvals, mu.post, sig.post )
# Get re-scaled posterior density
post.scaled <- post/max(post)*max(like.mu)
lines( muvals, post.scaled, col = "red" )

# This doesn't show much... consider a different prior:
# Normal prior with mean 8 and standard deviation 2
# Plot the likelihood
plot( muvals, like.mu, xlab = expression(mu)
     , ylab = "Probability density", type="l" )
# Add the prior
# Get re-scaled prior density
prior <- dnorm( muvals, 8, 2 )
prior.scaled <- prior/max(prior)*max(like.mu)
lines( muvals, prior.scaled, col = "blue" )
# Add the posterior
s.s0 <- ( 6.1^2 )/( 2^2 )
mu.post <- ( n*ybar + 8*s.s0 )/( n + s.s0 )
sig.post <- sqrt( ( 6.1^2 )/( n + s.s0 ) )

post <- dnorm( muvals, mu.post, sig.post )
# Get re-scaled posterior density
post.scaled <- post/max(post)*max(like.mu)
lines( muvals, post.scaled, col = "red" )
```


Probability density
0e+00
2e-13
4e-13
0
5
10
15
μ


Probability density
0e+00
2e-13
4e-13
0
5
10
15
μ