Tutorial 3

The questions for this tutorial have been revised directly from the class text "The Statistical Sleuth"

Question 1 in Tutorial 2 (Con'd, revised based on ex 7.27 from Statistical Sleuth)
The file "ex0727.csv" contains measured distances and recession velocities for 10 clusters of nebulae. According to a theory by Hubble the mean of the measured distance, as a function of velocity, should be β_1*velocity (i.e., μ(distance|velocity) = β_1*velocity), and β_1 is the age of the universe.

a) Are the data consistent with the theory that the intercept (β_0) is zero?
c) Produce a 95% confidence interval for the estimate in part (b) of Question 1 in Tutorial 2.

Question 2 in Tutorial 2 (Con'd, revised based on ex 7.29 from Statistical Sleuth)
Black wheatears are small birds of Spain and Morocco. Males of the species demonstrate an exaggerated sexual display by carrying many heavy stones to nesting cavities. Different males carry somewhat different sized stones, prompting a study of whether larger stones may be a signal of higher health status. A study was conducted (M. Soler et al.) which calculated the average stone mass (g) carried by each of 21 male wheatears, along with T-cell response measurements reflecting their immune systems' strengths. The file "ex0729.csv" contains the data.

a) Does the data suggest there is a linear relationship between the mean T-cell response and average stone mass?
b) If a particular wheatear carries stones that have an average mass of 6.5g, what would you predict its T-cell response to be? Provide a 90% prediction interval for this prediction.

Question 1 (revised based on ex 8.23 from Statistical Sleuth)
The data in the file "ex0823.csv" are the average wine consumption rates (in litres per person) and number of ischemic heart disease deaths (per 1000 men aged 55 to 64 years old) for 18 industrialised countries.

a) Do these data suggest that the heart disease rate is associated with average wine consumption? (You will need to transform the data before fitting the SLR)

Question 2 (revised based on ex 8.17 from Statistical Sleuth)
In a study of the effectiveness of biological control of the exotic weed tansy ragwort, researchers manipulated the exposure to the ragwort flea beetle on 15 plots that had been planted with a high density of ragwort. Harvesting the plots the next season, they measured the average dry mass of ragwort remaining (grams/plant) and the flea beetle load (beetles/gram of ragwort mass) to see if the ragwort plants in the plots with high beetle loads were smaller as a result of herbivory by the beetles. The data is contained in the file "ex0817.csv".

a) Use scatterplots of the raw data, along with trail and error, to find a transformation of Y = Ragwort dry mass and X = Flea beetle load that will be suitable for a SLR. Try the following transformations: log, square root, and reciprocal.
To create the 16 plots covering all transformation combinations use the following commands:

```
Ymat = cbind(Y, log(Y), sqrt(Y), 1/Y)
Xmat = cbind(X, log(X), sqrt(X), 1/X)
Ynames=c("Y", "log(Y)", "sqrt(Y)", "1/Y")
```

```
Xnames=c("X", "log(X)", "sqrt(X)", "1/X")
par(mfrow=c(4,4))
for(i in 1:4) {
    for(j in 1:4) {
                plot(Xmat[ , i], Ymat[ , j], xlab=Xnames[i], ylab=Ynames[j])
}
}
```

b) Use a residual plot to check whether the transformations you choose seem appropriate.

Question 3 (revised based on ex 7.30 from "The Statistical Sleuth")
Studies over the last two decades have shown that activity can effect the reorganisation of the human central nervous system. For example, it is known that the part of the brain associated with activity of a finger or limb is taken over for other purposes in individuals whose limb or finger has been lost. In one study, psychologists used magnetic source imaging (MSI) to measure neuronal activity in the brains of nine string players (six violinists, two cellists, and one guitarist) and six controls who had never played a musical instrument, when the thumb and fifth finger of the left hand were exposed to mild stimulation. The researchers felt that stringed instrument players, who use the fingers of their left hand extensively, might show different behaviour in the brain – as a result of this extensive physical activity – than individuals who did not play stringed instruments. The file "violin.csv" contains data on the neuron activity index from the MSI and number of years that the individual had been playing a stringed instrument (zero for the controls).

a). Is the neuron activity different in the stringed musicians and the controls?

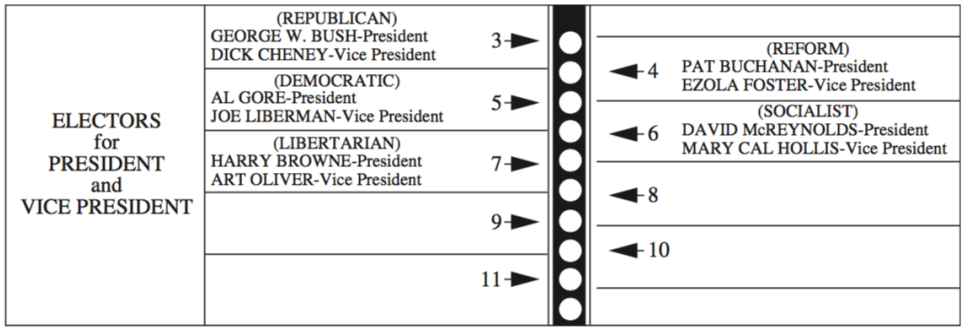
b). Is the amount of neuron activity associated with the number of years the individual had been playing the instrument?

Question 4 (revised based on ex 8.25 from "The Statistical Sleuth")
The U.S. presidential election of November 7, 2000 was one of the closest in history. As returns were counted on election night it became clear that the outcome in the state of Florida would determine the next president. At one point in the evening, television networks projected that the state was carried by the democrat nominee, Al Gore, but a retraction of the projection followed a few hours later. Then, early in the morning of November 8, the networks projected that the Republican nominee George W. Bush, had carried Florida and won the presidency. Gore called Bush to concede. While on route to his concession speech, though, the Florida count changed rapidly in his favour. The networks once again reversed their projection, and Gore called Bush to retract his concession. When the roughly 6 million Florida votes had been counted, Bush was shown to be leading by only 1738 and the narrow margin triggered an automatic recount. The recount, completed in the evening of November 9, showed Bush's lead to be less than 400.

Meanwhile, angry Democrat voters in Palm Beach County complained that a confusing "butterfly" lay-out ballet caused them to accidentally vote for the Reform Party candidate Pat Buchannan instead of Gore. The ballot, shown below, listed presidential candidates on both a left-hand and right-hand page. Voters were to register their vote by punching the circle corresponding to their choice, from the column of circles between the pages. It was suspected that since Bush's name was listed first on the left-hand page, Bush voters likely selected the first circle. Since Gore's name was listed second on the left-hand page, many voters – who already new who they wished to vote for – did not bother examining the right-hand side of the

ballet and consequently selected the second circle in the column; the one actually corresponding to Buchannan. Two pieces of evidence supported this claim: Buchannan had an unusually high percentage of the vote in that county, and an unusually large number of ballots (19000) were discarded because voters had marked two circles (possibly inadvertently voting for Buchannan and then trying to correct the mistake by then voting for Gore.)

DISPLAY 8.22 Confusing ballot in Palm Beach County, Florida



[Above display (Display 8.22) is taken from "The Statistical Sleuth"]

The file "vote.csv" contains data on the number of votes for Buchannan and Bush in all 67 counties in Florida.

a). Produce plots of the number of Buchannan votes versus the number of Bush votes and another plot for the log of these two variables. Does the log-log transformation appear better for performing a simple linear regression?

b). Analyse the data without the Palm Beach County results to obtain an equation for predicting Buchannan votes from Bush votes.

c). Use the residual and normal probability plots from the regression in (b) to check the adequacy of the model.

d). Obtain a 95% prediction interval for the number of Buchannan votes in Palm Beach – assuming the relationship is the same in this county as the others.

e). Comment on the result in (d) given that Buchannan actually received 3407 votes in Palm Beach County.