**MA4413 2009 Assignment 2**

# Instructions

The submission deadline is 2pm Monday 9th November 2009. Answers should be typed, 12pt font with 1.5 line spacing, and submitted as a single pdf file with the following naming convention `Assign2-0123456.pdf` where the trailing digits are your UL Student Identification Number. Keep your answers brief and concise. Verbosity will be penalised. There should be no accompanying files. If you need to refer to any of your own R code then you should include it in your report – but only if absolutely necessary. Email your report as a single file attachment (not zipped) to `kevin.hayes@ul.ie`.

## Questions

1. In 1882 Simon Newcomb caried out an experiment to determine the speed of light. His measurements showed the time it took light to travel a fixed distance of 7442 m. You can get these data in the file `newcomb.txt`. The original measurements have been coded by multiplying by $10^9$ and then subtracting 24,800. Why do you think this was done? Use a graphical display (stem-and-leaf plot, histogram, dotplot or boxplot) to inspect the data. Can the measurements be described by a normal distribution? Why or why not? Include your preferred plot in your report.

2. Include a normal probability plot of the variable `Speed`. What does the plot say about the assumption of normality? Make a new series of

speed data, call it `positive.speeds`, by subsetting the original series by excluding negative values. Include a normal probability plot of the variable `positive.speeds`. Comment.

3. In 1798 Henry Cavendish took some experimental measurements of the density of the Earth. You can get these data in the file `cavendish.txt`. Do you think these data can be described by a normal distribution? use a normal probability plot. What was Cavendish's estimate of the density of the Earth? How, after inspecting a graphical summary of the data, might he have obtained a "better" estimate?

4. Data showing the population, *Pop87* and *Pop97*, in thousands of all countries in Africa in 1987 and 1997 are available in the file `africa.txt`. using whichever year's data you prefer, look at the actual values in the `R` console. Which country seems to stand out from the others? Obtain two different graphical displays and a normal probability plot for the populations. What conclusions to you arrive at regarding the validity of the assumption that these data are described by a normal distribution?

5. Find the natural logarithms of the *Pop87* and *Pop97* values. Also find logarithms to base 10 of the *Pop87* and *Pop97* values. Are the logarithms or the original values better described by a normal distribution? What is the difference graphically between using natural logarithms or base 10 logarithms? In what situations would logarithms be an appropriate transformation to get a normal distribution?

6. According to data published by the Pregnancy Outcome unit of the South Australian Health Comision, the weights of new born babies in

outh Australia had a mean of 3400 g and a standard deviation of 550 g, and could approximately described by a normal distribution ('very-low birthweight' babies are excluded). Find the proportion of 'very-low birthweight babies', which are those weighting less than 2500 g. What proportion weighted less than 3500 g? What proportion were 'high birthweight' babies, weighing more than 4000g?

7. Find the weight which had 10% of babies' birthweights above it. Find the lowe 1% point and the upper 5% point of the babies weights. Find the quartiles of the babies' weights.

8. Medical records show that human gestation times (pregnancies) are approximately normally distributed with a mean of 266 days and a standard deviation of 16 days. What proportion of pregnancies last between 250 days and 290 days? What proportion last more than 305 days? The middle 95% of pregnancies last between ____ and ____ days (fill in the blanks). The longest 3% of pregnancies require special care. How long do such pregnancies last?