Physics 440 (ILab) – Exercise in Statistics
Due Thurs., Nov. 29, 2007

Use the data on life expectancy and other parameters that is contained in the file http://www.courses.umass.edu/phys440/stats/TVs_Life_expectancy_data.OPJ (an Origin file). For each of the forty largest countries in the world (according to 1990 population figures), data are given for the country's life expectancy at birth, number of people per television set, and number of people per physician. SOURCE: The World Almanac and Book of Facts (1993), New York (Pharos Books). (This datafile was found online at http://www.amstat.org/publications/jse/datasets/televisions.dat and http://www.amstat.org/publications/jse/datasets/televisions.txt)

1. Look at the life expectancy data (first data column):
   (a) what is the mean? (*i.e.* the average over the values for all countries listed here)
   (b) what is the standard deviation? (Note that Origin might use the N-1 normalization)
   (c) what is the standard error of the mean?

2. (a) Compute the correlation of {life expectancy} with {# of people per TV}.
   (b) Make a plot of one of these parameters vs. the other. (Write the correlation value on the plot).
   (c) State whether these are correlated or not (*i.e.* statistically independent parameters? Strongly correlated? Anti?) Does the correlation value make sense from the plot?

3. Open the datafile http://www.courses.umass.edu/phys440/stats/sum_statistics.OPJ . This data was obtained using a random number generator to make a random sequence of 1 and -1. We sorted these random #s into groups of size n=25 and found their sum. These are listed in the datafile for a large number of runs (N=10000). (You can think of this is a list of displacements of a random walker who has moved 25 steps.)

   (a) Create and plot a histogram. (In Origin, you can do this by right-clicking the data column and selecting 'frequency count;' if you choose step size = 2, you will avoid a lot of misleading zero entries.)
   (b) Fit this plot to a Gaussian function. (Note, if you choose 'Non-linear Curve Fit…" and select GaussAmp from the list of 'Origin Basic Functions,' you will get the variance with the usual definition. Other Origin "Gaussian" functions use non-standard definitions.)
   Record the fit parameters. Does the data follow a Gaussian distribution?
   (c) Compute the mean and standard deviation directly from the list of x-values. Do these agree with your fit parameters?
   (d) What is the standard error of the mean (provide a value and explain how you determined it)? Does $\langle x \rangle$ equal zero within the std. error?

(end)