Homework #1:
Data Analysis and Geostatistics, Fall 2020

August 27, 2020

Overview and dataset

Download the two snow depth datasets from the Google Drive, in the folder 'HW1'. The datasets are in ascii files called *depths1.txt* and *depths2.txt*. They are snow depths, in centimeters, measured at two different locations, on the same date. This homework will review basic statistical descriptions of these two datasets.

Load the data into MATLAB using the *load* command (e.g. `D1=load('depths1.txt')`).

1. For each dataset, calculate and report the mean and standard deviation (use *mean.m, std.m*)
2. For each dataset, calculate and report the median, mode, IQR, skewness, and kurtosis.
3. For each dataset, use the *boxplot.m* function to create a notched boxplot.
4. Using the *hist.m* function, create a histogram of each dataset, using 30 bins.

```
hist(D1,nbins) % first input is data, second is either # of bins, or vector of bin centers
```

5. Using the *subplot.m* function, create one figure showing all 4 plots from above.

```
subplot(row,col,current) % subplot with # rows, # columns, and choose which is current
```

6. Use the *axis.m* function to set the axis limits to the same range of values for all 4 plots.

```
axis([minX maxX minY maxY]) % set axis limits. Use square brackets to define 1x4 vector.
```

7. A more useful histogram is a *relative density histogram*, which is normalized such that the area of the histogram is equal to 1. Using the *hist* function, calculate the histogram, normalize it, and plot the *relative density histogram* for each dataset using the *bar* function.
8. The Gaussian (normal) distribution has the form

$$f(x;\mu,\sigma^2) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{(x-\mu).^2}{2\sigma^2}} \tag{1}$$

for mean μ and standard deviation σ. Using your calculated mean and standard deviation from above, calculate the normal distribution curve and plot it with a thick red line on your relative density histograms above.

9. What is the probability of a new measurement at each site being within 20cm of the average value?
10. What is the probability of a new measurement at each site being at least 20cm larger than the average value?
11. What is the probability of a new measurement at each site being at least 20cm smaller than the average value?

1 Appendix: Custom Functions

```
function f=mynormpdf(z,mu,sig)
% f=mynormpdf(z,mu,sig)
% INPUT: z = values to estimate PDF at
%          mu = mean value
%          sig = standard deviation
% OUTPUT: f = probability density function for a normal/Gaussian distribution
A=1/(sig*sqrt(2*pi)); % constant to ensure PDF integrates to 1 over +/- inf
B=(z-mu).^2; % squared deviations from mean; numerator in exponential function
C=2*sig.^2; % constant; denominator in exponential function
f=A*exp(-B./C); % normal PDF
```