**Exercise 7: Bootstrapping**

1. Assume we sample 3 numbers: {1,2,3}.
	1. Write all possible bootstrap sampling from this group.
	2. Show that the number of all possible bootstrap samples is nchoosek(2\*n-1,n) where n is the sample size. (hint: look on the web for a formula to compute combination with repetition. For example - <https://www.mathsisfun.com/combinatorics/combinations-permutations.html>). But make sure you understand it
2. x below is a sample of shopping spending values. Let's practice bootstrapping with this sample.
	1. Compute and plot the bootstrap distribution of the mean of x.
	2. Compute the mean of the distribution, the bias and the std.
	3. Compare the std of the bootstrap distribution with the expected SE value of the sampling distribution.
	4. Compute the 95% CI using the t distribution technique and the percentile method

x = [3.11 8.88 9.26 10.81 12.69 13.78 15.23 15.62 17.00 17.39 18.36 18.43 19.27 19.50 19.54 20.16 20.59 22.22 23.04 24.47 24.58 25.13 26.24 26.26 27.65 28.06 28.08 28.38 32.03 34.98 36.37 38.64 39.16 41.02 42.97 44.08 44.67 45.40 46.69 48.65 50.39 2.75 54.80 59.07 61.22 70.32 82.70 85.76 86.37 93.34]

1. The survival times of cancer patients after treatment are typically strongly right-skewed. Here are the survival times (in days) of 72 guinea pigs in a medical trial:

X = [43 45 53 56 56 57 58 66 67 73 74 79 80 80 81 81 81 82 83 83 84 88 89 91 91 92 92 97 99 99 100 100 101 102 102 102 103 104 107 108 109 113 114 118 121 123 126 128 137 138 139 144 145 147 156 162 174 178 179 184 191 198 211 214 243 249 329 380 403 511 522 598]

(a) Make a histogram of the survival times.

(b) Use bootstrap to inspect the sampling distribution of the mean and the median, and compute the bias for each statistic measure.

(c) Compute a 95% CI of the mean and median

1. Let:

X = [-1.2813 -0.1841 0.0482 -0.6145 0.6551 0.7311 0.5106 -2.4743 -0.8332 -0.4330 -0.4136 -0.3690 -0.7189 -0.5005 -0.4866 0.0742 1.3049 0.0640 1.3466 -1.0089]

Y = [-1.4464 -0.6979 -0.7789 0.3631 -0.1148 0.4467 1.5122 0.9879 0.4127 1.6380 -0.6286 -0.7610 -0.9177 -0.5770 0.1648 1.1977 1.5438 1.0619 -0.1745 -0.6828];

1. Compute the correlation between x and y. Use matlab *[r,p]=corr(X’,Y’)* function to find the r and p values of this correlation
2. Use permutation to estimate the p value of this correlation (use 2-sided estimation, and explain whether you assume symmetrical or unsymmetric distribution)
3. Use bootstrap to estimate the confidence interval of this correlation. Is the zero inside this CI?
4. Assume we have 20 points and that we fitted a line using a linear regression model (see the code below). The fitted line has two parameters: slope and intersection with the y axis (b1(2) and b1(1) respectively as in the code below).

We want to estimate the std of these parameters. Hint: you canot resample the points as pairs here. So think what can you resample.

x=(1:20)';

y=x+(randn(1,20))';

b1 = regress(y, [ones(size(x)), x])';

figure;

scatter(x,y)

% add the line:

hold on;

yfit = b1(2)\*x+b1(1);

plot(x',yfit)