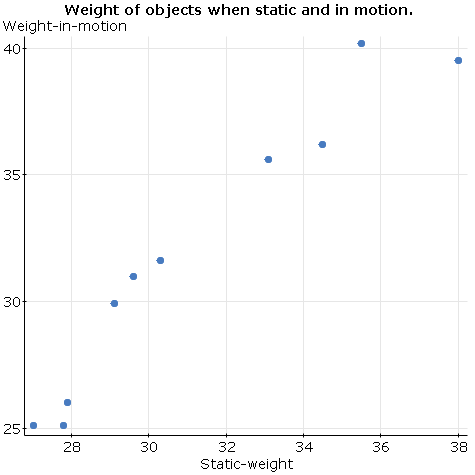
Statistics for computer Science

Philip Dumaresq

Lab 04 - September 28th 2017

Guy Beaulieu

1. The Minnesota Department of Transportation hoped that they could measure the weights of big trucks without actually stopping the vehicles by using a newly developed “weight-in-motion” scale. To see if the new device was accurate, they conducted a calibration test. The weighed several trucks when stopped (Static Weight), assuming that this weight was correct. Then they weighed the trucks again while they were moving to see how well the new scale could estimate the actual weight. Their data is given in the file Vehicle\_weights (on Moodle).
   1. **Make a scatterplot for these data.**



* 1. **Describe the direction, form, and strength of the plot.**

The scatterplot is positive, linear, with a medium strength

* 1. **Write a few sentences telling what the plot says about the data. (Note: The sentences should be about weighing trucks, not about the scatterplot.)**

The plot says that while trucks are in motion, there appears to be a slight increase in the weight of the trucks.

* 1. **Find the correlation.**

The correlation between the static weight of the trucks and of the trucks in motion is 0.965

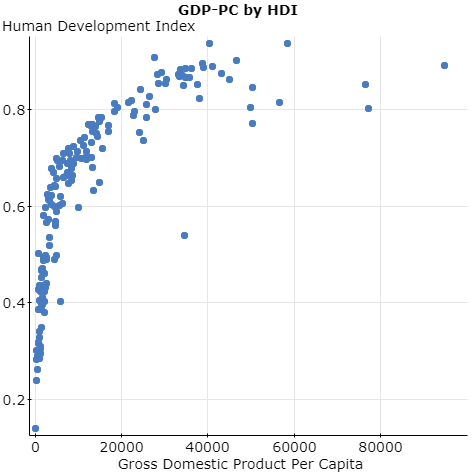
* 1. **If the trucks were weighed in kilograms, how would this change the correlation? (1 kilogram = 2.2 pounds)**

If the data were to be converted all to kilograms, then the data would scale accordingly. The correlation would not change, and the shape of the plot wouldn’t, but each of these points would individually be lower.

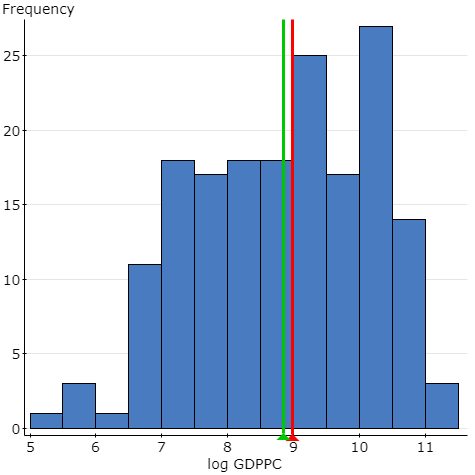
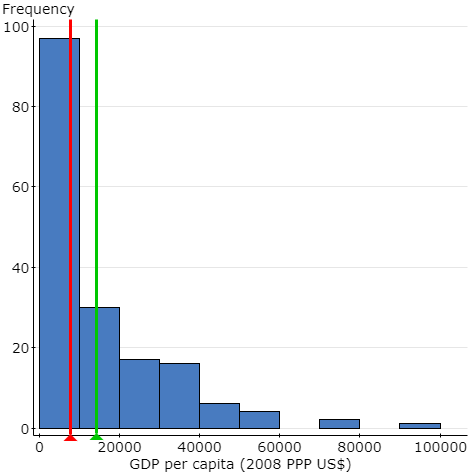
* 1. **Do any points deviate from the overall pattern? What does the plot say about a possible recalibration of the weight-in-motion scale?**

As the mass increases, there are a couple points that seem to start moving away from the pattern. However, there are only two points here. To really draw any conclusion from that data there would need to be more data collected using heavier loads.

1. The United Nations Development Programme (UNDP) uses the Human Development Index (HDI) in an attempt to summarize in one number the progress in health, education, and economics of a country. In 2010, the HDI was as high as 0.938 for Norway and as low as 0.14 for Zimbabwe. The gross domestic product per capita (GDPPC) is used to summarize the overall economic strength of a country. Consider the Human\_Development\_Index\_2010 data (on Moodle).
   1. **Make a scatterplot of GDPPC against HDI for all the countries.**

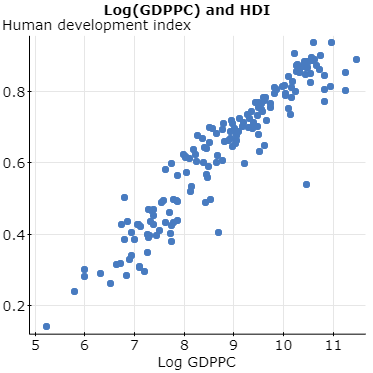


* 1. **GDPPC is measured in dollars. Income and other economic measures tend to be highly right-skewed. Taking logs often makes the distribution more unimodal and symmetric. Compare the histogram of GDPPC to the histogram of log(GDPPC).**



The histogram on the left is unimodal and heavily skewed to the right with a few outlier. The one on the right is the log of the same data and the histogram is bi-modal and relatively symmetric.

* 1. **Use the log re-expression to make a scatterplot of log(GDPPC) against HDI. Comment on the effects of the re-expression.**



The re-expression of the data makes the scatterplot much more linear. The correlation appears to be about the same before and after the log re-expression. The correlation is still positive.

* 1. **The HDR report classifies countries into high, medium, and low development based in part by their HDI. Look at the scatterplot of log(GDPPC) against HDI for only the medium- and high-development countries (currently viewed as the highest 75%, so HDI¿0.48 here). Does this relationship appear stronger or weaker than when we include all the countries?**

With the higher 75% of countries, the relationship between GDPPC and HDI appears to be quite a bit stronger than with the whole dataset.

* 1. **Find the correlation of log(GDPPC) with HDI for all countries. Find the correlation for only the medium- and high-development countries. Which is higher?**

The correlation for all countries is: 0.94347217

The correlation for the medium and high development countries is: 0.88455456

The correlation for all countries is stronger than the correlation for countries with a medium of high GDPPC.