

Laboratory Activity 6 for Math 080: Topics from Chapter 6

Jordan C. Hanson

July 25, 2022

1 Formula Area

Normal distribution PDF:

$$p(x) = \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left(-\frac{1}{2} \frac{(x-\mu)^2}{\sigma^2}\right) \quad (1)$$

2 Introduction

Please download the data file “Normal Scores” as an Excel/LibreOffice file.

3 Questions

1. The data in the file Normal Scores represents GPA data from students admitted to Whittier College. There are three data sets: (a) high school GPA, (b) first semester GPA, and (c) second semester GPA.
2. Create three normalized histograms for the three data sets. Do they appear to be normally distributed? List the means and standard deviations of each.
3. Using the normalized histogram data, and knowledge of the means and standard deviations, form a model $N(\mu, \sigma)$ for each of the data sets. That is, write three PDF functions that describe the data sets.
4. For each data sets, for the relative frequencies that are not zero, compute the base-10 logarithm. The function is called `LOG10`. Create a x-y scatter plot of the logarithm of the relative frequencies versus the bin values. The data sets should now appear as quadratic curves. Do you see why? If you take the logarithm of Eq. 1, what do you get?
5. If you add a quadratic trend line in the x-y scatter plots, do the trend lines appear to agree with the data points? Which model, if any, fails to explain the lower GPA numbers in any of the models?¹
6. **Bonus:** try redoing the analysis, but subtract from each GPA value the mean of the data set, and divide the result by the standard deviation. The results will (potentially) be *standard normal* distributions $N(0, 1)$. The deviations from $N(0, 1)$ might be more easily identified.

¹The last time I worked with this data, I noticed the frequency of students scoring GPAs several standard deviations below the means in their first and second semesters at Whittier College was markedly higher than the frequency predicted by $N(\mu, \sigma)$.