Guided Thought Questions

MATH 322: Mathematical Statistics II Learning Module 4: Goodness-of-Fit

All textbooks are written at two cognitive levels. The surface level is the literal information provided in the book. The deeper level is the connections between the topics. Textbooks are excellent at the surface level, rarely good at the deeper level. To help with the deeper levels, I am providing several questions for each section of the textbook.

You should take time to answer these questions after reading the section. Answer them in your notes. Make sure that you are able to confidently answer them. In fact, you should feel pressure to ask these questions in class if you cannot answer them.

Readings: Sections 7.1 to 7.4

§1: Introduction

1.

§2: The Chi-Square Tests for Count Data

- 1. Why is the expected count npi?
- 2. Why do the two Chi-square tests require a "large" sample size to be accurate?

§3: GoF to Identify the Probability Distribution

- 1. What are advantages and disadvantages to using the Pearson Chi-square test?
- 2. On what graphic is the Pearson Chi-square test based? Why is knowing this important?
- 3. What is a drawback to the Kolmogorov-Smirnov test?
- 4. On what graphic is the Kolmogorov-Smirnov test based? Why is knowing this important?
- 5. What advantage does the Shapiro-Wilk test have over the other tests discussed in this section?
- 6. What can the P-P plot and the Q-Q plot tell us about the data? How are they different from each other and from the histogram?

§4: Applications: Parametric Analysis

- 1. Why did the author decide upon the Weibull distribution?
- 2. Why is it better to compare empirical and theoretical CDFs than to compare empirical and theoretical pdfs?
- 3. Why is knowing the actual distribution of the data helpful in analysis and prediction?