Discussion Questions

Conor Mayo-Wilson

Stylized Example: A coin-factory produces two types of coin: one type has bias $\frac{1}{4}$ and the other has bias $\frac{3}{4}$. You have a coin that was produced by this factory, but you don't know its bias. Suppose you flip the coin 52 times, and fourteen flips have landed heads. You are interested in testing the hypothesis that the coin has bias $\frac{1}{4}$. How would Fisher approach this problem? How would Neyman and Pearson? To answer these questions, you should do the following:

1. Define the following terms: hypothesis test, rejection region, null hypothesis, alternative hypothesis, size, power, and p- value. Explain how they might be identified in the above example in question.

2. Which concepts are employed Fisher's analysis of the example? Which are employed in Neyman and Pearson's approach to hypothesis testing?

3. According to Fisher, why is the outcome of a hypothesis test important? For Neyman and Pearson?

1 Goodman

Steven N. Goodman. Toward evidence-based medical statistics. 2: The bayes factor. *Annals of internal medicine*, 130(12):1005–1013, June 1999:

1. What is the "P-value fallacy"?

    - Note: Goodman describes the fallacy in at least two ways in the paper. You should consider whether the two descriptions are equivalent.
    - Hint: The fallacy involves interpreting P-values in two different ways.

2. Why is it a fallacy? That is, what argument (and example) does Goodman give that a P-value cannot be interpreted in both ways? How could Howson and Urbach's discussion of "mixed tests" be used to provide a second example for Goodman?

3. Is their interpretation of P-values legitimate, according to Goodman? Why or why not?

2 Howson and Urbach

Colin Howson and Peter Urbach. *Scientific reasoning: the Bayesian approach*. Open Court, Chicago, 2005:

1. Howson and Urbach identify several problems for Fisherian tests. Explain their claim that for Fisher, the choice of rejection region and test statistic is arbitrary.

2. How do (some) Neyman-Pearsonian tests address the two previous worries? What theorem is relevant here, and when is it applicable?

3. Howson and Urbach describe three interpretations of significance tests. Describe the first two and the problems with each, according to Howson and Urbach.