Brock Parker
University of Arizona
Department of Astronomy & Steward Observatory
Professor Tim Eifler

Statistical and Computational Methods for Astronomy (ASTR 513)

# Homework 1

August 25, 2023

---

1. **Come up with two definitions what 'Statistics' as a science is all about.**

   (a) **In general.**

   ---

   Statistics is the science of the treatment of data and uncertainty, including collection, processing, analysis, and—most importantly—interpretation. Statistics allows us to rigorously extract and understand the underlying patterns in data and the assumptions and conclusions we can draw therefrom.

   (b) **Specific to Astrophysics.**

   ---

   In the context of astrophysics, statistics provides the ground work to analyze and interpret observations and compare them to existing theories and models in an unbiased and meaningful manner. We utilize statistics to convert collected data into usable interpretations of the physical world.

2. **Have ChatGPT come up with its two definitions.**

   (a) **In general.**

   ---

   Statistics is the scientific discipline that involves collecting, analyzing, and interpreting data to uncover patterns, relationships, and trends in order to make informed decisions and draw meaningful conclusions about the world. It encompasses mathematical techniques for extracting knowledge from data and plays a crucial role in various fields by providing a structured approach to handling uncertainty and making predictions.

   (b) **Specific to Astrophysics.**

   ---

   In astrophysics, statistics involves employing mathematical and computational techniques to process and interpret data gathered from celestial observations. It enables astrophysicists to uncover hidden patterns, measure cosmic parameters, validate theoretical models, and draw meaningful conclusions about the nature and behavior of the universe.

3. **Explain the three main differences between Frequentist and Bayesian approach to statistics.**

   ---

   (a) In frequentist statistics, parameters of interest are considered fixed and the data is random; uncertainties result from sampling error alone, rather than inherent error in the parameters. The results depend purely on the observed data, however the experiment methodology must be known. In Bayesian statistics, the parameters are not assumed fixed, but rather random, requiring an initial probability distribution for the parameter value. Such a distribution—the prior—along with the observed data, is required to extract meaningful results.

   (b) In frequentist statistics, the analysis provides a confidence interval, a range of values that repeated measurements are expected to fall into with a given probability. In contrast, Bayesian statistics result in a more intuitive probability interval wherein the true value is expected to lie with a given probability. This is extracted from the posterior distribution, which describes the overall probability distribution of the parameter.

   (c) Frequentist statistics rely on p-values, which describe the probability of obtaining an equally extreme data set with future repeated experiments assuming the null hypothesis is true. The p-value gives no probability of the alternative hypothesis being true, simply the rejection or acceptance of the null hypothesis. Bayesian statistics directly provides the probability of the hypothesis given the data. This allows for the direct comparison of the probabilistic strengths of different hypotheses, the Bayes factor.