Statistical Thinking for Forensic Practitioners

Quiz on Part 5: Probability Models and Uncertainty

1 Identifying probability models

Probability models allow us to treat seemingly disparate situations using the same mathematical tools. A handful of probability models are used so often that they are given names. The probability models discussed in the Part 5 lecture slides are examples of such probability models. Dr. Rice explained, in general terms, situations in which these models are most often applied. We will consider specific examples here. Statisticians often consider a series of simple questions to determine which model adequately describes a situation. In the following exercises, you will practice identifying which probability model discussed in the lecture notes best describes each situation.

1. A footwear impression analyst from Irvine, California is interested in determining the local prevalence of Carhartt boots. They randomly survey 10,000 people living in Irvine and record whether they own Carhartt boots.

    (a) Identify the quantity of interest from the sample that we can model using a probability distribution. (Hint: it may be helpful to start your answer with "the number of...").

    (b) Is the quantity of interest you identified above discrete or continuous?

    (c) What are the possible outcomes of one survey response?

(d) What probability distribution should be used to model the quantity of interest?

(e) Interpret the unknown parameter of the probability distribution identified above within the context of the problem.

(f) Assume the quantity of interest you identified in part (a) is a random variable that follows the probability distribution you identified in part (d). Interpret the expectation of this random variable in the context of the problem.

2. A researcher is interested in the prevalence of deltas in the latent prints from a population of interest. The researcher considers a sample of 122 latent prints from this population.

(a) Identify the quantity of interest from the sample that we can model using a probability distribution. (Hint: it may be helpful to start your answer with "the number of...").

(b) Is the quantity of interest you identified above discrete or continuous?

(c) What types of values can the quantity of interest you identified above take on?

(d) What probability distribution should be used to model the quantity of interest?

(e) Assume the quantity of interest you identified in part (a) is a random variable that follows the probability distribution you identified in part (d). Interpret the expectation of this random variable in the context of the problem.

3. Forensic examiners consider the frequency of an allele in a sample of 381 genotypes.

   (a) Identify the quantity of interest from the sample that we can model using a probability distribution. (Hint: it may be helpful to start your answer with "the number of...").

   (b) Is the quantity of interest you identified above discrete or continuous?

   (c) What are the possible outcomes for a single genotype?

   (d) What probability distribution should be used to model the quantity of interest?

(e) Interpret the unknown parameter of the probability distribution identified above within the context of the problem.

(f) Assume the quantity of interest you identified in part (a) is a random variable that follows the probability distribution you identified in part (d). Interpret the expectation of this random variable in the context of the problem.

4. Investigators receive a tip of an unknown number of contraband shipping containers on a barge of 2,000 containers. Due to the large number of containers, investigators decide to randomly select 20 containers without replacement to search. They intend to use the resulting number of contraband containers as an estimate for the number of contraband containers on the whole ship.

(a) Identify the quantity of interest from the sample that we can model using a probability distribution. (Hint: it may be helpful to start your answer with "the number of...").

(b) Is the quantity of interest you identified above discrete or continuous?

(c) What types of values can the quantity of interest you identified above take on?

(d) What probability distribution should be used to model the quantity of interest?

(e) Only one of the parameters of the distribution you identified in part (d) is unknown. Interpret this unknown parameter within the context of the problem.

5. Electron microscopy is used to digitally scan the surface of a cartridge case. The result is a "surface matrix" representation of the cartridge case containing surface height values. Practitioners visually compare variations in the height values (e.g., breech face impressions) to determine whether two cartridge cases were fired from the same firearm. Researchers are interested in modeling the height values of a set of 40 cartridge cases.

(a) Identify the quantity of interest from the sample that we can model using a probability distribution.

(b) Is the quantity of interest you identified above discrete or continuous?

(c) The researchers are unsure whether to use a normal model or a log-normal model. Consider the following histogram of height values for a cartridge case scan. Based on what you know about the normal and log-normal probability models, which would make the most sense to model the height values? Explain your reasoning.
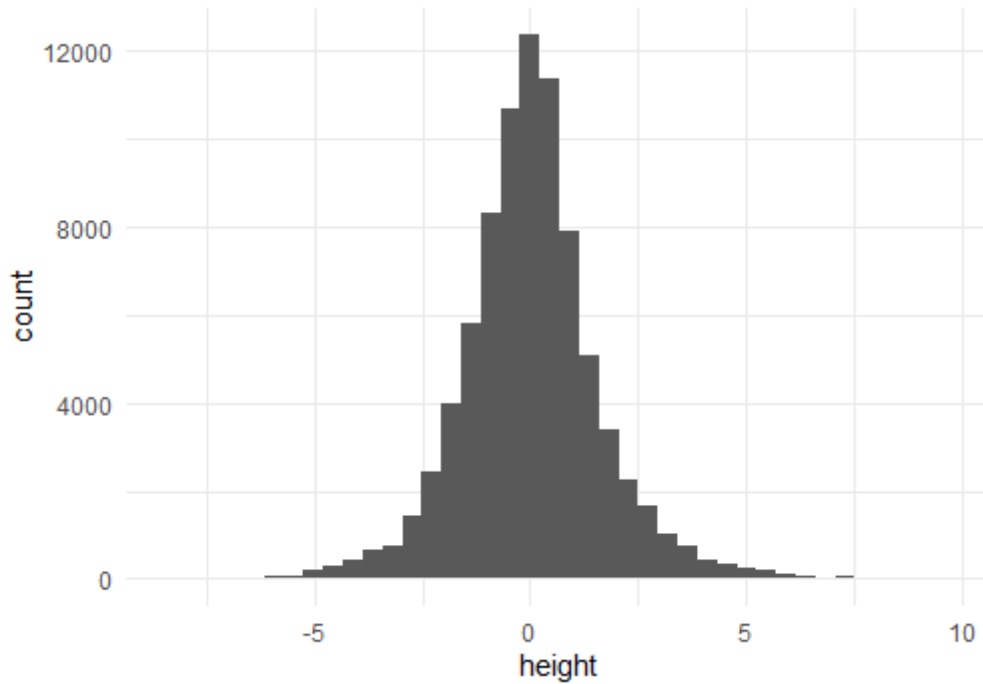
Figure 1: Histogram of height values of a cartridge case scan centered on the average height.

(d) There are two unknown parameters for the distribution you chose in part (c). Interpret the parameters of the probability distribution identified above within the context of the problem.

6. A longitudinal study (a study performed over time) was performed over 6 months to examine how shoes develop Randomly Acquired Characteristics (RACS). The researchers were interested in the number of RACs developed on 200 pairs of sneakers over the 6 month period.

(a) Identify the quantity of interest from the sample that we can model using a probability distribution. (Hint: it may be helpful to start your answer with "the number of...").

(b) Is the quantity of interest you identified above discrete or continuous?

(c) What types of values can the quantity of interest you identified above take on?

(d) What probability distribution should be used to model the quantity of interest?

(e) Assume the quantity of interest you identified in part (a) is a random variable that follows the probability distribution you identified in part (d). Interpret the expectation of this random variable in the context of the problem.