Statistics and Probability, Grades 6 – 8

The work of teaching statistics and probability

Teachers need to design activities that help student see statistical reasoning as a four-step investigative process that involves:

- formulating questions that can be answered with data;
- designing and using a plan to collect relevant data;
- analyzing the data with appropriate methods;
- interpreting results and drawing valid conclusions from the data that relate to the questions posed.

Teachers guide students to use random sampling to collect data and learn to differentiate between the variability in a sample and the variability inherent in a sample statistic when samples are repeatedly selected from the same population.

Teachers help students to understand random sampling by first building their understanding of elementary probability. The probability of a chance event as approximated by a long-run relative frequency is introduced using coins, number cubes, cards, spinners, bead bags. Once the connection between relative frequency and theoretical probability is clear, probability experiments are simulated with technology. It is critical that teachers help students understand how the product rule for counting outcomes for chance events is used in situations such as tossing three coins or rolling two number cubes.

Teachers introduce students to the analysis of bivariate measurement data graphed on a scatterplot and help them describe shape (a cloud of points on a plane), center (a line drawn through the cloud that captures the essence of its shape) and spread (how far the data points stray from this central line). Teachers also introduce students to numerical descriptions of center and spread such as median, quartiles, the interquartile range, and mean absolute deviation.

Key understandings to support this work

- Understand and know how to guide students in the 4-step investigative statistical process.
- Recognize that approximating the probability of a chance event by observing its long-run relative frequency may result in strings and patterns in the short run. Understand that a basic tenet of statistical reasoning is that random sampling allows results from a sample to be generalized to a much larger body of data, namely, the population from which the sample was selected.
- Distinguish between variability in a sample and variability in a sample statistic when samples are repeatedly selected from the same population.
- Know how to find and explain various measures of shape, center, and spread of univariate and bivariate data (including mean, median, quartiles, mean absolute deviation). Identify clusters, gaps, and unusual data points for bivariate data.

- Identify skewness when looking at a graph of data and explain what it means in the context of the problem.
- Choose the best measure of center for a given situation, and explain why it is best.
- Know how to find theoretical probability in uniform and non-uniform probability models, and be able to explain to students why theoretical and empirical probabilities differ for a given situation.
- Know the product rule for finding outcomes of two or three independent events (such as rolling 2 dice or flipping 3 coins) and be able to help students understand why this is a multiplicative (rather than additive) situation.
- Know when and how to use simulations to help students make sense of statistical ideas.
- Interpret the slope of the line fitting a scatter plot as a rate of change.
- Describe the association between two categorical variables.

Illustrative examples

1. In the last ten games of the 2006 season, the Gwinnett Women's Soccer team had an average score of 2 goals per game. Create four different dot plots with data values for the ten games. Remember that the mean number of goals for the ten games is 2. Label your plots a, b, c, and d. The plots should satisfy the following conditions:

Plot a: Only one game had 2 goals.
Plot b: Exactly two games had 4 goals.
Plot c: One game had an amazing 7 goals!
Plot d: The median of the data set is 3.

For the next soccer game played, we might use the mean of 2 to predict how many goals a team will score given that plot a, b, c, or, d describes the previous ten games. For which distribution do you think the prediction of 2 is most likely to be closest to the actual number of goals scored? Give a statistical justification for your answer.

2. In 1972, 48 male bank supervisors were each given the same personnel file and asked to judge whether the person should be promoted to a branch manager job that was described as "routine" or whether the person's file should be held and other applicants interviewed. The files were identical except that half of the supervisors had files showing the person was male while the other half had files showing the person was female. Of the 48 files reviewed, 35 were promoted. (Reference. B. Rosen and T. Jerdee (1974), "Influence of sex role stereotypes on personnel decisions," *J. Applied Psychology*, 59:9-14.)

Create three tables for the data, one showing no discrimination, one showing strong evidence of discrimination, and one where discrimination is not obvious without further investigation.

What percentage of males and females were recommended for promotion? Without exploring the data any further, do you think there was discrimination?

Conduct a simulation of this situation, and record the results in a dot plot. Using these data, estimate the likelihood that 21 or more males out of 35 will be selected if the selection process is random. Look at the dot plot and describe the shape, center, and variability of the distribution. Is the behavior of this distribution what you might expect? Based on the simulation, does there appear to be evidence to support the claim of discrimination?