

Author: Chris Franklin

Situation: z-scores

Prompt

In a university statistics course designed for secondary mathematics majors, students were asked to consider the following scenario. [Source: Agresti, A. and Franklin, C. (2009). *Statistics: The Art and Science of learning from Data* (p.88). Upper Saddle River, NJ: Pearson/Prentice Hall.]

Roger Maris, who spent most of his professional baseball career with the New York Yankees, held the record for the most home runs in one season (61) from 1961 until 1998, when the record was broken by Mark McGwire. Maris played in the major leagues from 1947 to 1968. The number of home runs he hit in each year that he played is summarized in MINITAB output as shown.

Variable	N	Mean	Median	StDev	Min	Max	Q1	Q3
RMHR	12	22.92	19.50	15.98	5.00	61.00	10.00	31.75

A sports writer commented that Roger Maris hit only 13 home runs in 1966. Was this unusual for Maris? Comment, using statistical justification.

A student (an in-service secondary teacher) approached answering this question by finding the z-score for 13 as follows:

$$z = (13 - 22.92) / 15.98 = -0.62.$$

The student then commented that since 13 is less than one standard deviation below the mean of 22.92, hitting only 13 homeruns in one season would not be considered unusual. Thirteen would be considered plausible for what Maris might hit in a season. In fact, using the normal distribution, the probability that Maris hit 13 or less homeruns in one season is about .27 or 27%.

Commentary

The student selected a reasonable approach by evaluating the z-score, then commenting in context that 13 homeruns was less than one standard deviation from the mean; therefore, not an unusual value. Unfortunately, the student went one step further and incorrectly assumed the distribution of homeruns for Maris was normally distributed. The distribution is not normally distributed, but instead skewed to the right.

Focus 1

This focus will describe what is a z- score and what it measure. A z-score is a way to standardize data with respect to it's mean and standard deviation.

Focus 2

This focus will describe how the z -score can be used with any distribution of numerical data. Although z -scores are commonly used with normal distributions in order to find probabilities from a normal table, it is only appropriate to find normal probabilities using z -scores when we can assume the distribution is normal versus being a non-normal distribution. It is important for students to learn that using z -scores does not imply normal distribution.

Focus 3

This focus will discuss how the data for Maris' homerun seasons is discrete versus being continuous. Thus, the distribution can't be normally distributed since the normal distribution is continuous. The question is whether it's appropriate to approximate this discrete distribution using the normal distribution.

Focus 4

This focus will describe how one can tell the distribution of home runs for Roger Maris is not approximately bell-shaped but rather right skewed by using the given mean and standard deviation. It will be noted that the minimum value of 5 is only 1.12 standard deviations below the mean while the maximum value of 13 is 2.39 standard deviations above the mean indicating a right skewed distribution (skewed to higher values) rather than a bell-shaped, symmetric distribution.

Focus 5

This focus will describe how one can use the 5 number summary of positions (min, max, Q1, med, and Q3) to construct a boxplot and visually see the distribution is skewed toward higher values, not symmetric.

Focus 6

This focus will describe an alternative approach to answering the original question by using the quartiles of the distribution. Since the first quartile is 10 which indicates that in 25% of the seasons Maris hit 10 or less homeruns, 13 would not be considered unusual.