SAVANNAH math labs

Lab Module 1: Determining Outliers

Question 1: Height is a quantitative variable that tends to be nearly normally distributed. Suppose the distribution of heights among males is normally distributed with mean parameter (μ_X) equal to 68 inches and standard deviation parameter (σ_X) equal to 3.5 inches.

Who is the Outlier?

Shaq and Kevin Hart could not be more different with respect to their height! Shaq's height is seven foot one inch, or 85 inches. Kevin Hart's height is five foot four inches, or 64 inches.

(a) Compute a Z-score for Shaq's height with respect to the population distribution of male heights.

(b) Compute a Z-score for Kevin Hart's height with respect to the population distribution of male heights.

(c) For normally distributed data, a Z-score of 0.6745 corresponds to the upper quartile (Q_U) and a Z-score of -0.6745 corresponds to the lower quartile (Q_L). Determine the minimum Z-score magnitude that would correspond with an outlier based on the 1.5 $\times$ IQR criteria.

(d) Based on your Z-score calculations, determine whether Kevin Hart, Shaq, both or neither are outliers with respect to height. Base your answer on the 1.5$\times$ IQR criteria for outliers.

(e) Why can't you simply look at the picture and determine who is an outlier? Why must you know information about the population (i.e. the parameters μ_X and σ_X) to classify outliers? Explain.

Question 2: Suppose you are the manager of a trendy new boutique in Miami Beach named Millie 305. You collect sales data for 100 customers in the first month following the boutique's grand opening. You compute the quartiles of monthly spending at Millie 305 for these 100 customers and obtain the following: Q_L = \$55, Median = \$150, Q_U = \$375. The five customers who spent the least spent the following amounts: (\$5.75, \$6.25, \$8.75, \$9.00 and \$10.25). The five customers who spent the most spent the following amounts: (\$675, \$1,025, \$1,500, \$2,750 and \$3,125).

(a) Sketch a box plot by hand for the customer sales data.

(b) Based on your box plot, determine which customer sales amounts are outlier data values.

(c) Why are you not able to classify outliers based on a Z-score calculation for this data set?