Descriptive Statistics 2 – Solutions

STAT-UB.0103 – Statistics for Business Control and Regression Models

Standard Deviation and The Empirical Rule

1. Fifty-seven respondents to the class survey reported their SAT scores. The mean score was 2160, and the standard deviation was 140. What can you say about the range of scores reported? Assume that the distribution of reported scores is symmetric and mound-shaped.

> **Solution:** We can use the empirical rule to make the following statements:
>
> - For approximately 68% respondents, reported score is between 2020 and 2300.
> - For approximately 95% respondents, reported score is between 1880 and 2440.
> - For approximately 99.7% respondents, reported score is between 1740 and 2580.
>
> In fact the true percentages in those intervals are 73%, 96%, and 98%. When the distribution of the data is symmetric and mound-shaped, the predictions from the empirical rule are usually only accurate for the 68% and 95% intervals.
>
> One other point: it is impossible to get an SAT score above 2400, so we could report the ranges for 95% and for 99.7% as [1880, 2400] and [1740, 2400].

2. Of those students who reported high school GPAs measured on a 4-point scale, the mean value was 3.9 and the standard deviation was 0.2.

   (a) Complete the following statement with appropriate values for X and Y: "For those students whose high school GPAs were measured on a 4-point scale, approximately 95% of the survey respondents reported values between X and Y."

   > **Solution:** $X = 3.9 - 2 \times 0.2 = 3.5$; $Y = 3.9 + 2 \times 0.2 = 4.3$.

   (b) What assumptions do you need to make for the statement in (a) to be correct? Do you think these assumptions are plausible? How could you check this?

   > **Solution:** That the distribution of GPAs is symmetric and mound-shaped.
   >
   > We could check this with a histogram. In fact, there is a slight skew to the right for the GPA data, but even with this skewness, there is reasonable agreement with the empirical rule: 76% of reported GPAs were between 3.7 and 4.1; 92% of reported GPAs were between 3.5 and 4.3; 98% of reported GPAs were between 3.3 and 4.5.

3. Your company has an annual profit of \$60MM with a standard deviation of \$5MM. Assume that the distribution of your annual profits is symmetric and mound-shaped.

   (a) Would it be unusual for your company to have an annual profit of \$52MM?

Solution: No; 95% of the time, profits are between \$50MM and \$70MM.

(b) Would it be unusual for your company to have an annual profit of \$83MM?

Solution: Yes; this would happen less than 99.7% of the time.

z-scores

4. Fifty-one respondents from the class survey reported an expected annual salary below \$150K. The mean and standard deviation of these values (in \$1K) was $\bar{x} = 68$ and $s = 13$. How many standard deviations above or below the mean are the following values?

   (a) An expected starting salary of \$80K per year.

   Solution: Let $x_1 = 80$ and let z_1 be the number of standard deviations above of below the mean. Then,
   $$x_1 = \bar{x} + sz_1,$$
   so
   $$z_1 = \frac{x_1 - \bar{x}}{s} = \frac{80 - 68}{13} = 0.9.$$
   Thus, x_1 is 0.9 standard deviations above the mean.

   (b) An expected starting salary of \$60K per year.

   Solution: Let $x_2 = 60$. Then,
   $$z_2 = \frac{x_2 - \bar{x}}{s} = \frac{60 - 68}{13} = -0.6.$$
   Thus, x_2 is 0.6 standard deviations below the mean.

   (c) An expected starting salary of \$250K per year.

   Solution: Let $x_3 = 250$. Then,
   $$z_3 = \frac{x_3 - \bar{x}}{s} = \frac{250 - 68}{13} = 14.$$
   Thus, x_3 is 14 standard deviations above the mean.

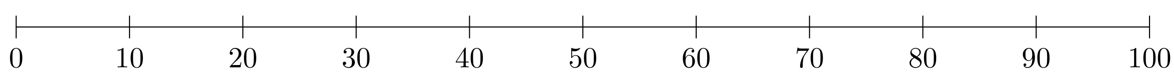
5. In the previous problem, which of the values are unusual?

   Solution: The value $x_3 = 250$ is unusual, since this is 14 standard deviations away from the mean. Typical values are within 2 or 3 standard deviations of the mean (here, "typical" means 95% or 99.7% of the time).

Boxplots

6. Here are the 35 reported expected starting salaries for the male survey respondents (in \$1K per year). Make a boxplot of the data.

50, 50, 50, 50, 60, 60, 60, 60, 60, 60, 60, 60, 60, 62.4 65, 65, 65, 70, 70, 70, 75, 76, 80, 80, 80, 80, 80, 80, 80, 85, 90, 90,100,250,300



7. Here are the 18 reported expected starting salaries for the female survey respondents. Make a boxplot of the data.

40, 45, 54, 60, 60, 60, 60, 60, 65, 67, 70, 70, 70, 70, 80, 80, 85, 100