SEIS 631

Dan Ward

Assignment 8

**Q1) What type of plot would you use to display the relationship between runs and one of the other numerical variables? Plot this relationship using the variable at\_bats as the predictor. Does the relationship look linear? If you knew a team’s at\_bats, would you be comfortable using a linear model to predict the number of runs?**

A scatter plot would be the best plot for the relationship between runs and another numerical variable. The relationship looks moderately positive linear. Knowing the teams at bats I would feel comfortable using a linear model to predict the number of runs.

**Q2) What correlation did you get? Describe the relationship between at\_bats and runs? Make sure to discuss the form, direction, and strength of the relationship as well as any unusual observations.**

The correlation is .610627. The relationship is positive, linear, and moderate to weak in strength, with a few outliers.

**Q3) Fit a new model that uses homeruns to predict runs. Using the estimates from the R output, write the equation of the regression line. What does the slope tell us in the context of the relationship between success of a team and its home runs?**

For every homerun the runs increase by 1.8345.

**Q4) If a team manager saw the least squares regression line and not the actual data, how many runs would he or she predict for a team with 5,579 at-bats? Is this an overestimate or an underestimate, and by how much? In other words, what is the residual for this prediction?**

Philadelphia Phillies had 5579 at-bats and 713 runs. So, the least squares regression line overestimated the number of runs by 15 runs.

**Q5) Is there any apparent pattern in the residuals plot? What does this indicate about the linearity of the relationship between runs and at-bats?** The data is fairly consistent both above and below the 0 line. There is more data in the first half of the at bats vs the higher count of at bats. There is also only a couple of outliers. This indicates that the data follows a linear model.

**Q6) Based on the histogram and the normal probability plot, does the nearly normal residuals condition appear to be met?**  Yes, the histogram is nearly normal and the qq plot has some deviation, but if follows the normal line fairly closely.

**Q7) Based on the plot in (1), does the constant variability condition appear to be met?** Yes, the variability is fairly constant.