AP Test Review – Describing Relationships

I. Scatterplots

A. Explanatory and Response Variables

B. Analyzing scatterplots

a) Direction – Positive or Negative

b) Form – Linearity or curved

c) Strength

d) Outliers

When analyzing a scatterplot, always do so in context of the problem.

C. Correlation, r – Describes a linear relationship between 2 quantitative variables.

a) $-1 \le r \le 1$

b) not a resistant measure

c) correlation **DOES NOT** mean causation!!

Remember, r only measures strength of linear relationships. Do not use if the relationship is not linear.

II. Least-Squares Regression Line (LSRL)

A. $\hat{y} = a + bx$ or $\hat{y} = b_0 + b_1x$

a) used for prediction of a scatterplot

b) calculate on your calculator or using formula on sheet

c) always goes through $(\bar{x}, \bar{y})$

d) the coefficient of determination, r^2, gives the fraction of variation in the values of y that is accounted by the LSRL of y on x (i.e. overall measure of success the LSRL is in relating x and y)

B. Residuals

a) difference of observed and predicted y values

b) observed y – predicted y or residual = $y - \hat{y}$

c) LSRL makes the residuals as small as possible

C. Residual plot*

a) plot of residuals (y axis) against the explanatory variable (x axis_

b) assess whether a linear model is appropriate (if well scattered about y = 0)

Tips and Common Mistakes:

- Remember to label and title all graphical displays
- Be aware of Minitab output on p. 181
- An observation is influential if removing is drastically changes the result of a calculation
- Beware of extrapolation when using LSRL's
- You should be able to interpret slopes, y-intercepts, correlation, etc. in context
- LSRL's often replace meaningful names instead of x and $\hat{y}$, for instance:
  $\widehat{salary}$ = 40,000 + .1 (commission)
- Transforming for linearity, i.e. plotting logs of response variable against explanatory variable if data is exponential.