

Unit 2 Day 3 CW - Coefficient of Determination

$$\hat{y} = a + bx$$

$$b = r \frac{sy}{sx}$$

$$a = \bar{y} - b\bar{x}$$

For the following:

- 1) Compute appropriate statistics needed to find the least-squares regression equation.
- 2) Use formulas to find the regression equation, and write it in a descriptive manner.
- 3) Find r^2 , and write a sentence which explains its meaning in the context of the problem.

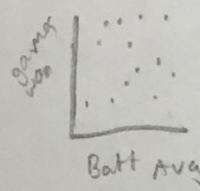
Problem 1: Is the number of games won by a major league baseball team related to its batting average? Before you begin, be sure you understand which variable is the explanatory and which is the response.

Based on the RESID plot between BA and GW, a linear model would be best fit. However, the

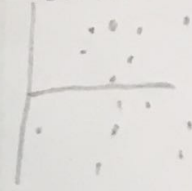
TEAM	GAMES WON	BATTING AVERAGE
NY	87	.277
TOR	83	.275
BALT	74	.272
BOS	85	.267
TB	69	.257
CLE	90	.288
DET	79	.275
CHI	95	.286
KC	77	.288
MIN	69	.270
ANA	82	.280
TEX	71	.283
SEA	91	.269
OAK	91	.270

calc by hand

correlation of .23 shows a positive but weak relationship between the



BA + GW resid



$$\begin{aligned} \bar{x} &= .2747 \\ s_x &= .00817 \\ \bar{y} &= 81.64 \\ s_y &= 8.705 \\ r &= .2319 = .3 \end{aligned}$$

$$\begin{aligned} \hat{y} &= 13.8 + 246.96x \\ &= 246.96x + 13.8 \end{aligned}$$

$$r^2 = .0529 = .09$$

$$\text{games won} = 13.8 + 246.96(\text{batting avg})$$

By the coefficient of determination (r^2), there is only .0529 ($\approx 5\%$) of the games won accounted for the LSRL relating the Batting avg and the Games won.

Problem 2: How is a baby's age related to the number of hours the baby sleeps each day?

AGE (Years)	HOURS SLEPT
0.1	14.9
0.2	14.5
0.3	13.4
0.4	14.1
0.5	13.4
0.6	13.7

form, strength, direction

$$\hat{y} = 14.8 - 2.457(x)$$

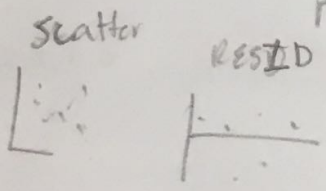
\hat{y} - predicted sleep

x - age (years)

$$r = -.7496$$

$$r^2 = .562$$

Due to the findings (analysis), as a baby's age increases the number of hours slept decreases.



The scatterplot of the hours slept based on age shows a negative, strong ~~relationship~~ ^{association}. This is also strengthened by the $-.7496$. A linear would be satisfactory based on no patterns in the residual plot.