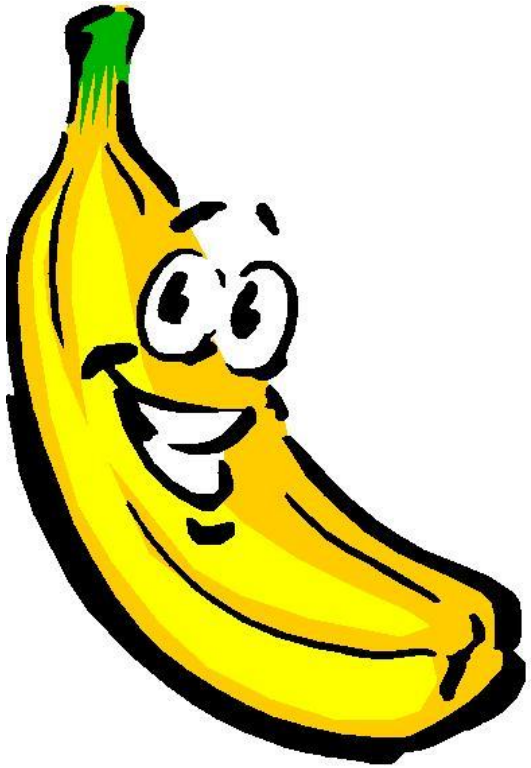

I tried to fit a line to estimate how many people fall per banana peel, but it was a slippery slope.

Larry Lesser and Dennis Pearl
https://www.causeweb.org/cause/resources/library/r12865

VIRTUAL LAB 8

Complete Example

Example

- A professor would like to use attendance to estimate a student's GPA. The professor used his current class of 160 students and found that the average number of days absent was 4 with a standard deviation of 1.39. The average GPA in the class was 1.37 with a standard deviation of 0.95 and a correlation of -0.52. Use this information to answer the following questions…

Questions

- Create the sample regression line for predicting GPA from the number of days missed.

  $\hat{y} = 2.79 - 0.356x$

- What would a predicted GPA at the end of this class if they missed 3 days of class? How about 4 days of class?

  For x=3, $\hat{y}$ = 1.72

  For x=4, $\hat{y}$ = 1.37

- State the hypothesis for a test to determine whether the slope of the true regression line is equal to zero.

  H_0: β_1=0

  H_A: $\beta_1 \neq 0$

Questions, continued…

- Fill in the blanks for the following table:

Parameter	Estimate	Std. Error	t Value	P-value
Intercept	2.79	0.196	14.23	≈0
Slope	-0.356	0.046	-7.74	≈0

- Summarize the results of the hypothesis test of the slope.

Conclusion for the test is that we reject the null hypothesis. There does appear to be a significant LINEAR relationship between the number of absent days from class and class GPA.

- What is the value and meaning of R^2 in this problem?

$R^2 = -0.52^2 = 0.27$…approximately 27% of the variation in class GPA can be explained by the number of absent days from class.

MORE EXAMPLES

More examples

- At the last college football game of the season, a survey of attendants was conducted for a random sample of 95 people to asked them their age and how many games that they have attended this season. The average age was 29 with a standard deviation of 8.4. The average number of games attended was 4.2 with a standard deviation of 2.3. The correlation between these variables is 0.6.

- What is the slope for the simple linear regression line for using age to predict the number of games attended?

0.164

- What is the intercept for the simple linear regression line for using age to predict the number of games attended?

-0.556

Continued…

- The following results were obtained from your regression above:

SSE = 14000

$\Sigma (x - \bar{x})^2 = 12100$

What is the test statistic for the test of the slope of the regression line? $T_{calc}=1.47$

What is the margin of error for a 95% confidence interval for the slope? MOE=1.986*1.11=0.221

What is the proportion of variation in number of games attended that is explained by its linear relationship with age?

$R^2 = 0.6^2 = 0.36$…approximately 36% of the variation in number of games attended can be explained by age.