Jacob Igel Red – Code Input

ISA 291 – A Blue – Output

Dr. Waldyn Highlight – Important Information

1. Use R to generate a scatter plot of y = accuracy and x = distance. Briefly (in one or two sentences) describe the relationship.
   * plot(PGADRIVER$DISTANCE, PGADRIVER$ACCURACY)

![Chart, scatter chart

Description automatically generated]()

* + Between distance and accuracy, the plot appears to have a negative relationship. This means as the distance increases, the accuracy of the golfers decrease.

1. Compute the correlation coefficient between accuracy and distance. Interpret the correlation coefficient.
   * > cor(PGADRIVER$DISTANCE,PGADRIVER$ACCURACY)
   * [1] -0.906395
   * Since this correlation coefficient is so close to -1, this means that the relationship between distance and accuracy has a very strong negative relationship.
2. Compute the least squares regression equation between y = accuracy and x = distance. Include the R table of coefficients.
   * State the simple linear regression equation.
     + Coefficients:

Estimate Std. Error t value Pr(>|t|)

(Intercept) 250.14203 14.23101 17.58 < 0.0000000000000002 \*\*\*

DISTANCE -0.62944 0.04759 -13.23 0.000000000000000848 \*\*\*

* Y = 250.14203 – 0.62944x
  + Interpret the slope coefficient in the language of the problem.
    - Slope = – 0.62944x
    - For every one yard increase of distance, the accuracy of a PGA golfer landing in the fairway decreases by 62%.
  + Interpret the intercept coefficient in the language of the problem. Does the intercept term have a practical meaning?  Why or why not?
    - Intercept = 250.14203
    - If the distance is zero then PGA golfers accuracy landing in the fair way is over 200%.
    - Since the value of x = 0 is not in the range observed, it has n other interpretation other than being the y intercept.

1. State the simple regression assumptions. Evaluate whether the model assumptions are reasonable for this analysis by creating and referring to specific plots.  Include and clearly label the relevant plots from the output.
   * We can assume that the mean of the probability distribution is equal to zero.
     + > mean(reg1$residuals)
     + [1] 0.000000000000000006938894
   * We can say the variance is constant throughout all variables/settings of x.
     + > var(reg1$residuals)
     + [1] 87356657
   * We can say that the probability is normally distributed.
     + hist(reg1$residuals, breaks = 30)

![Chart, histogram

Description automatically generated]()

* + We can assume that the errors associated with both observations are independent because the distance data is not equal to the accuracy data.

1. Conduct a test of significance for the slope.
   * Clearly, the slope of the least squares regression equation is not zero. Explain why you should conduct this test.
     + We still want to conduct this test to make sure that if we enter any prediction values, that they will be at least 95% accurate with the current equation we have.
   * State Ho and Ha
     + Ho: B1 = 0
     + Ha: B1 0
   * Show how the test statistic is computed.
     + > summary(reg1)
     + Coefficients:

Estimate. Std. Error t value Pr(>|t|)

(Intercept) 250.14203 14.23101 17.58 < 0.0000000000000002 \*\*\*

DISTANCE -0.62944 0.04759 -13.23 0.000000000000000848 \*\*\*

Residual standard error: 2.236 on 38 degrees of freedom

* + What are the degrees of freedom of this test statistic?
    - 38 degrees of freedom
  + State the p-value and your decision in the language of the problem.
    - p-value: 0.000000000000000848
    - With a p-value that is zero and a slope greater than zero, we can confidently reject the null hypothesis in favor of the alternative. This brings us to the conclusion that the distance is a significant linear predictor of accuracy on the fairway among PGA golfers.
  + Compute a confidence interval for the slope and interpret it.
    - > confint(reg1, level = .95)

2.5 % 97.5 %

(Intercept) 221.3328452 278.951209

DISTANCE -0.7257813 -0.533105

* + - We are 95% confident that as the distance increases by 1 yard, average accuracy decreases between -72% and -53%

1. Compute and interpret a confidence interval for the accuracy of golfers with distance of 300 yards.
   * > predict(reg1, newdata = data.frame(DISTANCE=300), interval = "confidence")

fit lwr upr

1 61.30908 60.58614 62.03203

* + We can say with 95% confidence that the average accuracy for a golfer with a distance of 300 is between 60% and 62%

1. Compute and interpret a prediction interval for the accuracy of golfers with distance of 300 yards.
   * > predict(reg1, newdata = data.frame(DISTANCE=300), interval = "prediction")

fit lwr upr

1 61.30908 56.72439 65.89378

* + We can predict that 95% of the PGA golfers accuracy with a distance of 300 is between 56% and 65%.

1. Briefly (in a few sentences) summarize your findings as they relate to the professional golfer’s concern regarding the accuracy of his drive if he increases the distance.
   * Overall, it is safe to say with this data that as distance increases, golfers lose quite a bit of accuracy. If a golfer increases his distance in is drive by even a yard, their accuracy can fall dramatically – by upwards of 62%. The golfers need to be aware of this so they can better prepare for actual tournaments where the accuracy of your swings matter