Jordan Black

CGS2020 202501

Professor Lakshmi Prayaga

2/28/2025

**Assignment Regression Module 4**

1. **What is the purpose of performing a regression analysis?**

**Regression Analysis is used to determine the relationship between a dependent variable**

**And one or more independent variables. It helps in predicting outcomes, identifying trends, and making data-driven decisions.**

1. **Wha are the two requirements we studied to perform regression analysis?**

**What is the Adjusted R Squared: (Adjusted R-squared:( 0.5944**

**What is the value, what is the intercept, and what is the slope? Additionally, going back to the lecture video we have the hypothesis and**

1. **In the equation y = bx + a, which variable is the slope and which is the intercept?**

**b (Slope)**: Represents the rate of change of y concerning x. It shows how the dependent variable changes for each unit increase in the independent variable.

**a (Intercept)**: The value of y when x=0, meaning the predicted value of the dependent variable when the independent variable is absent.

1. **Given speed and stopping distance, which is the dependent variable, which is the independent variable?**

**Dependent Variable: Stopping Distance (feet)**

**Independent Variable: Speed (mph)**

**Since Stopping Distance is affected by speed. It is the dependent variable.**

1. **State the null hypothesis and alternate hypothesis as we did in the income vs. happiness example. Remember null hypothesis is always that there will be no effect of one variable on the other, and the alternate hypothesis is that there will be an influence of one variable on another.**

**Null hypothesis: There is no relationship between speed and stopping distance (b =! 0)**

**Speed does not significantly influence stopping distance.**

**Alternative Hypothesis: There is a significant relationship between speed and stopping distance. Speed affects stopping distance.**

**6. use the dataset speedvsdistance.data and perform a regression analysis to determine the regression equation and produce the regression plot. Deliverables include**

**a. the code,**

**# Load necessary libraries**

**library(ggplot2)**

**library(dplyr)**

**library(broom)**

**library(ggpubr)**

**# Load dataset (update the file path if needed)**

**data <- read.csv("C:/Users/draft/Downloads/speedvsdistance.csv", header = TRUE)**

**# Check column names**

**colnames(data)**

**# Rename columns if necessary (Ensure that "speed" and "distance" are the actual column names)**

**if (!all(c("speed", "distance") %in% colnames(data))) {**

**colnames(data) <- c("speed", "distance") # Adjust column names to match expected variable names**

**}**

**# Display summary statistics of the dataset**

**summary(data)**

**# Perform linear regression**

**speed.distance.lm <- lm(distance ~ speed, data)**

**# Display regression summary**

**summary(speed.distance.lm)**

**# Scatter plot of speed vs stopping distance**

**ggplot(data, aes(x = speed, y = distance)) +**

**geom\_point() +**

**labs(title = "Stopping Distance vs Speed",**

**x = "Speed (mph)",**

**y = "Stopping Distance (feet)")**

**# Add regression line to the plot**

**speed.graph <- ggplot(data, aes(x = speed, y = distance)) +**

**geom\_point() +**

**geom\_smooth(method = "lm", col = "red") +**

**labs(title = "Regression Analysis: Speed vs Stopping Distance",**

**x = "Speed (mph)",**

**y = "Stopping Distance (feet)")**

**# Display plot**

**print(speed.graph)**

**# Add regression equation to the plot**

**speed.graph <- speed.graph +**

**ggpubr::stat\_regline\_equation(aes(label = paste(..eq.label.., ..rr.label.., sep = "~~~~")),**

**label.x = 10, label.y = 150) # Adjust label position as needed**

**# Display the final plot with equation**

**print(speed.graph)**

**# Apply theme and final formatting**

**speed.graph <- speed.graph +**

**theme\_bw() +**

**labs(title = "Stopping Distance vs Speed",**

**x = "Speed (mph)",**

**y = "Stopping Distance (feet)")**

**# Show the final graph**

**print(seed.graph)A screenshot of a computer

AI-generated content may be incorrect.**

**b. screen shots of plotting data points, regression line, printing the regression equation and titles**

**> # Display summary statistics of the dataset**

**> summary(data)**

**speed distance**

**Min. : 4.0 Min. : 2.00**

**1st Qu.:12.0 1st Qu.: 26.00**

**Median :15.0 Median : 36.00**

**Mean :15.5 Mean : 43.02**

**3rd Qu.:19.0 3rd Qu.: 56.00**

**Max. :29.0 Max. :120.00**

**> # Perform linear regression**

**> speed.distance.lm <- lm(distance ~ speed, data)**

**> # Display regression summary**

**> summary(speed.distance.lm)**

**Call:**

**lm(formula = distance ~ speed, data = data)**

**Residuals:**

**Min 1Q Median 3Q Max**

**-40.146 -10.027 -0.937 8.474 46.049**

**Coefficients:**

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

**(Intercept) -13.3841 7.0061 -1.910 0.0621 .**

**speed 3.6390 0.4265 8.532 3.5e-11 \*\*\***

**---**

**Signif. codes: 0 ‘\*\*\*’ 0.001 ‘\*\*’ 0.01 ‘\*’ 0.05 ‘.’ 0.1 ‘ ’ 1**

**Residual standard error: 16.41 on 48 degrees of freedom**

**Multiple R-squared: 0.6027, Adjusted R-squared: 0.5944**

**F-statistic: 72.8 on 1 and 48 DF, p-value: 3.495e-11**

**> # Scatter plot of speed vs stopping distance**

**> ggplot(data, aes(x = speed, y = distance)) +**

**+ geom\_point() +**

**+ labs(title = "Stopping Distance vs Speed",**

**+ x = "Speed (mph)",**

**+ y = "Stopping Distance (feet)")**

**> # Add regression line to the plot**

**> speed.graph <- ggplot(data, aes(x = speed, y = distance)) +**

**+ geom\_point() +**

**+ geom\_smooth(method = "lm", col = "red") +**

**+ labs(title = "Regression Analysis: Speed vs Stopping Distance",**

**+ x = "Speed (mph)",**

**+ y = "Stopping Distance (feet)")**

**> # Display plot**

**> print(speed.graph)**

**`geom\_smooth()` using formula = 'y ~ x'**

**> # Add regression equation to the plot**

**> speed.graph <- speed.graph +**

**+ ggpubr::stat\_regline\_equation(aes(label = paste(..eq.label.., ..rr.label.., sep = "~~~~")),**

**+ label.x = 10, label.y = 150) # Adjust label position as needed**

**> # Display the final plot with equation**

**> print(speed.graph)**

**`geom\_smooth()` using formula = 'y ~ x'**

**Warning message:**

**The dot-dot notation (`..eq.label..`) was deprecated in ggplot2 3.4.0.**

**ℹ Please use `after\_stat(eq.label)` instead.**

**This warning is displayed once every 8 hours.**

**Call `lifecycle::last\_lifecycle\_warnings()` to see where this warning was generated.**

**> # Apply theme and final formatting**

**> speed.graph <- speed.graph +**

**+ theme\_bw() +**

**+ labs(title = "Stopping Distance vs Speed",**

**+ x = "Speed (mph)",**

**+ y = "Stopping Distance (feet)")**

**> # Show the final graph**

**> print(speed.graph)**

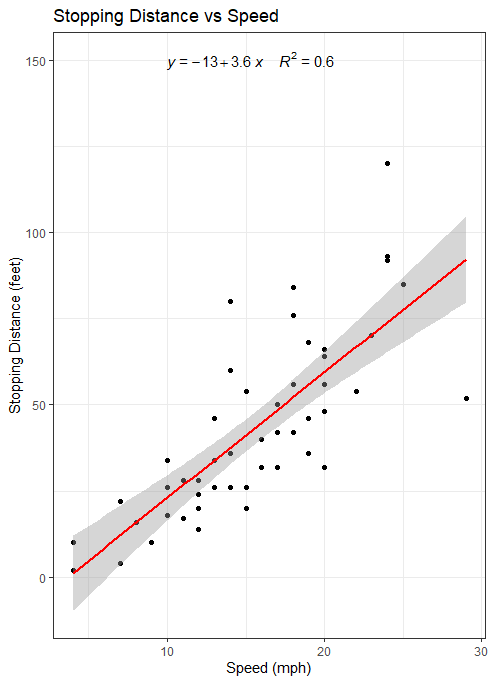
**`geom\_smooth()` using formula = 'y**

**Regression Equation:**

**Y= -13+3.6x**

**R^2=0.6**

**A screenshot of a computer

AI-generated content may be incorrect.**

**A screenshot of a computer

AI-generated content may be incorrect.~ x**

**c. interpret the results referencing the r squared value and p value. What do they tell you about the accuracy of your regression model?**

**R^2 Value:0.6027:**

**This means that 60.27% of the variation in stopping distance can be explained by speed.**

**P-Value: 3.495e -11(<0.05):**

**This means the relationship between speed and stopping distance is statistically significant.**

**Since the p-value is much smaller than 0.05, we reject the null hypothesis and conclude that speed significantly impacts stopping distance.**

**Final summary:**

**Regression Equation:**

**Y=-13+3.6x**

**Suggest that every 1mph increases in speed, stopping distance increases by 3.6 feet.**

**P-Value confirms that a strong relationship, rejecting the null hypothesis.**

**Goodness of fit:**

**The R^2 value (60.27%) shows the speed explains a significant portion of the variation in stopping distance, meaning other factors( road conditions, braking force) may also influence it.**