

Regression Inference

Learning Outcomes

- Given a research question, construct the null and alternative hypotheses in words and using appropriate statistical symbols
- Describe and perform theory-based hypothesis tests for the slope
- Interpret and evaluate a p-value
- Construct and interpret a theory-based confidence interval for slope
- Use a confidence interval to determine the conclusion of a hypothesis test

Terminology

The following terms will be covered in this activity.

- Correlation
- Slope
- Coefficient of determination

For further explanation of these topics review Chapter 3 and 7 in the textbook.

Hand Dexterity

Physical therapists often evaluate manual (hand) dexterity by having patients complete simple tasks, such as moving pegs on a board or threading objects through holes. Researchers want to examine the manual dexterity of children as part of a follow-up study of a test originally designed for adults to see how manual dexterity changes with age. In this test, 174 participants were given a board with 16 pegs, each in their own hole, arranged in a 4x4 grid. Participants were instructed to pick up the peg with one hand, flip it over by rotating their wrist, then reinsert it in the same hole. Using this test, researchers want to know if as people age the speed at which they can flip all 16 pegs increases.

The variables in this dataset consist of the following:

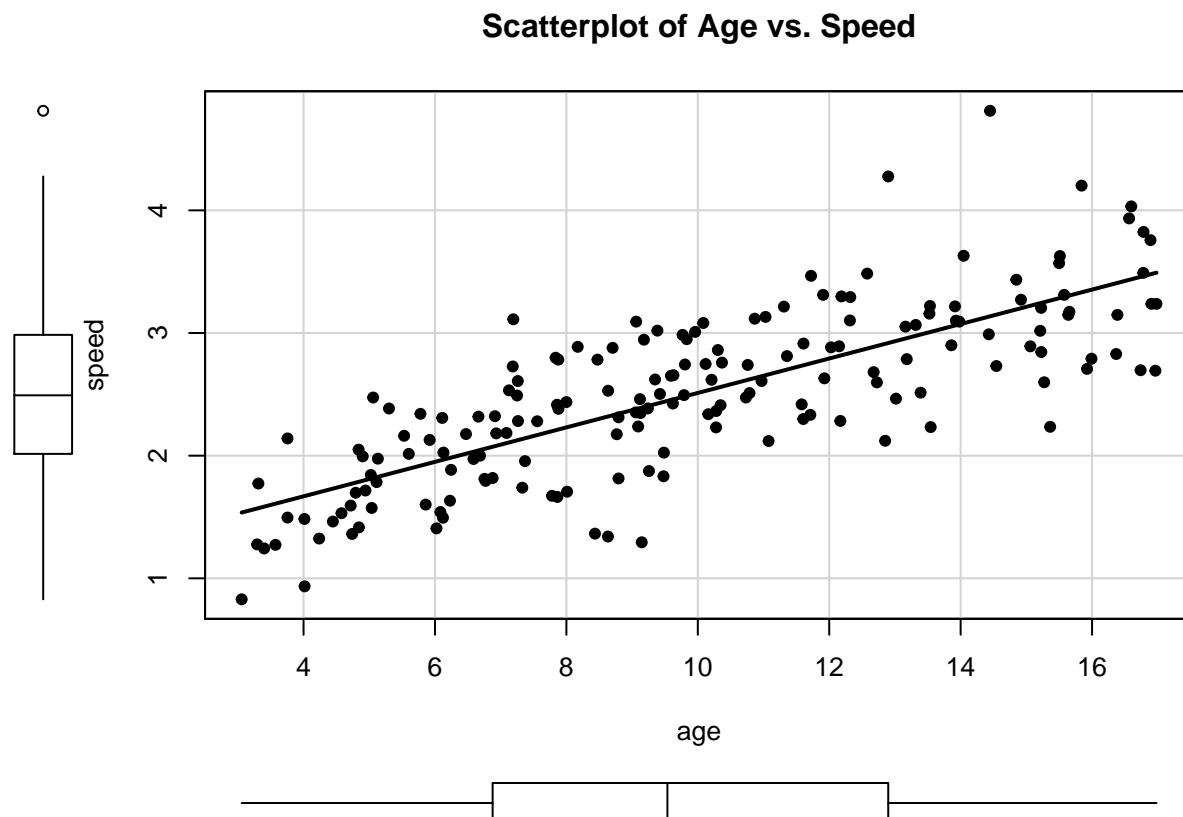
- **time:** Recorded time to flip all 16 pegs, measured in seconds.
- **speed:** The average speed to flip a peg for each participant (seconds per peg).
- **age:** Age of the participants, measured in years.
- **dominant:** Whether the participant's dominant hand was used, coded as 0 for no, 1 for yes.
- **gender:** The participant's gender, recorded as a binary variable, 0 for male, 1 for female.
- **HD:** The dominant hand of the participant, recorded as "R" for right hand, "L" for left hand.
- **handUsed:** Which hand the participant used to complete the test, recorded as "R" for right hand, "L" for left hand.

Data source: Hand Dexterity in Children: Administration and Normative Values of the Functional Dexterity Test (FDT), Gogola, G., et al., 2013

Vocabulary Review

1. Explain why regression methods are appropriate to use to address the researchers' question. Make sure you clearly define the variables of interest in your explanation and their roles.
2. What is the scope of inference for this study? Explain your answer.
3. Create a scatterplot to examine the relationship between the speed at which a participant can flip a peg and the age of the participant. Provide this plot. Based on your plot, does it appear that there is a relationship between age and speed? Note: age should be on the x-axis.

```
scatterplot(speed~age, data = hands, smooth=F, col=1, pch=19,  
            main = "Scatterplot of Age vs. Speed")
```



4. Describe the features of the plot you created in Question 3.

If you indicated there are potential outliers, which points are they?

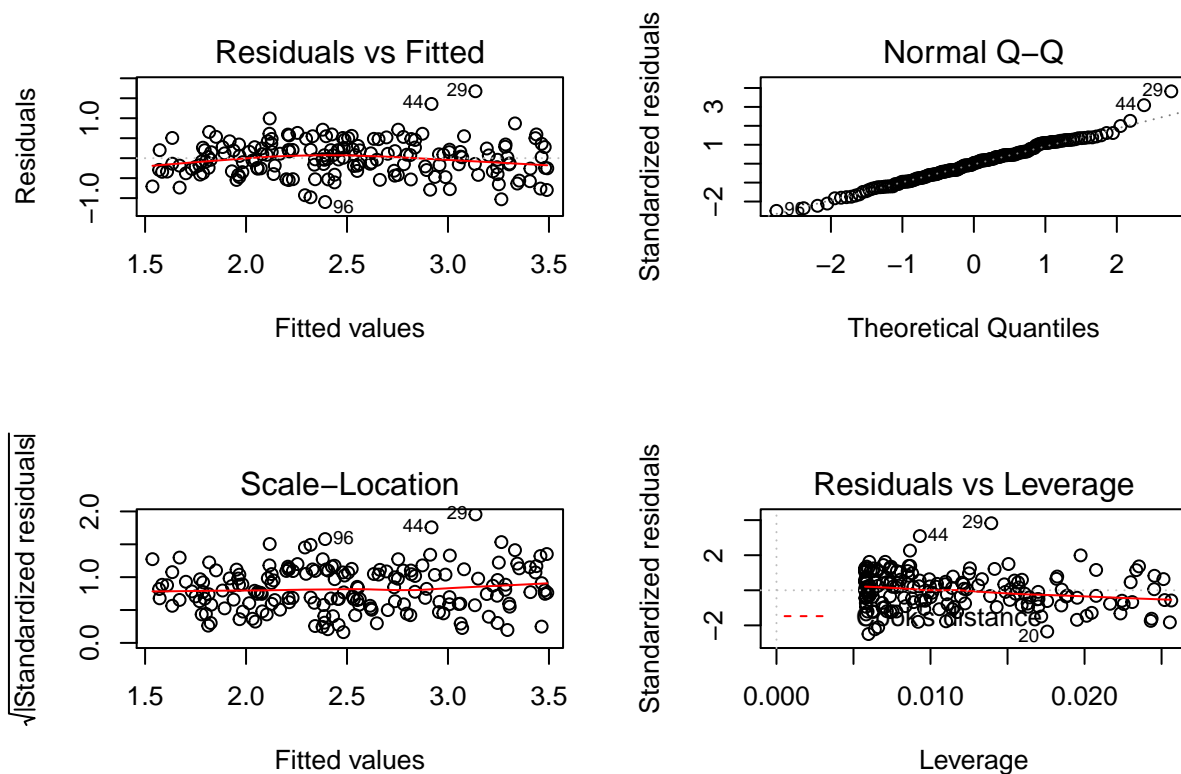
Conditions for the least squares line

When performing inference on a least squares line, the follow conditions are generally required

- Linearity: the data should follow a linear trend
- Nearly normal residuals: residuals must be nearly normal
- Constant variability: the variability of points around the least squares line remains roughly constant
- Independent observations: individual data points must be independent

The scatterplot and the residual plot will be used to assess the conditions for approximating the data with the T-distribution

```
lm.hand <- lm(speed~age, data = hands)
par(mfrow=c(2,2))
plot(lm.hand)
```



5. Are the conditions met

Ask a Research Question

6. Write out the null hypothesis in words.
7. Using the research question, write the alternative hypothesis in notation.

Summarize and Visualize the data

Enter the variable names into the linear model function to get the linear model output.

```
## replace each xxx with the appropriate information  
lm.hand <- lm(speed~age, data=hands)  
summary(lm.hand)$coefficients
```

```
##           Estimate Std. Error t value    Pr(>|t|)  
## (Intercept) 1.1070563 0.093326769 11.86215 4.237698e-24  
## age         0.1404378 0.008796699 15.96483 8.777516e-36
```

8. Using the output above, write the equation of the regression line.
9. Interpret the slope in context of the problem.
10. Using your estimated line of best fit, predict the per peg speed for a participant who was 9.18 years old. Show all work.
11. Calculate the residual associated with the observation (9.18, 2.95), using your estimated regression line from question 8.

Use statistical inferential methods to draw inferences from the data

To find the value of the test statistic to test the slope we will use,

$$T = \frac{\text{slope estimate}}{SE} = T = \frac{b_1}{SE(b_1)}$$

We will use the linear model output above to get the estimate for slope and standard error.

12. Calculate the test statistic for slope.

13. What value does the value calculated in question 12 match in the linear model output?

14. Interpret the test statistic in context of the problem.

15. Using the linear model output, report the p-value for the test of significance.

16. Based on the p-value, how much evidence is there against the null hypothesis?

Recall that a confidence interval is calculated by adding and subtracting the margin of error to the point estimate.

point estimate $\pm t^*SE(\text{estimate})$ $b_1 \pm t^*SE(b_1)$

The t^* multiplier comes from the t-distribution with $n - 2$ df.

```
qt(0.95+0.025, 172) #95% t* multiplier
```

```
## [1] 1.973852
```

17. Calculate the 95% confidence interval for the true slope.

18. Calculate the coefficient of determination for a linear model that describes the relationship between **age** and **speed**. Use proper notation.

19. Interpret this value in the context of the study.

Communicate the results and answer the research question.

20. Based on the p-value, write a conclusion in context of the problem.

21. Interpret the 95% confidence interval in context of the problem.

22. Summarize the results of the study in a written paragraph. Be sure to include.

- Summary statistic
- Test statistic and interpretation
- P-value and interpretation
- Confidence interval and interpretation
- Conclusion in context
- Scope of inference

Revisit and Look Forward

23. Is there an effect due to gender on this linear relationship? Explain your answer using the scatterplot.

```
scatterplot(speed~age+dominant, data = hands, smooth=F, col=1,  
            main = "Scatterplot of Age vs. Speed by Gender")
```

