

Hand Dexterity

10.1 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

10.2 Terminology review

In today's activity, we will use theory-based hypothesis tests and confidence intervals for a linear regression slope. Some terms covered in this activity are:

- Correlation
- Slope
- Regression line

To review these concepts, see Chapters 3 and 7 in the textbook.

10.3 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 data set¹ consist of the following:

Variable	Description
time	Recorded time to flip all 16 pegs (seconds)
speed	Average speed to flip a peg for each participant (seconds per peg)
age	Age of the participants (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 non-male
HD	The dominant hand of the participant, recorded as R for right hand, L for left hand

¹Data source: Hand Dexterity in Children: Administration and Normative Values of the Functional Dexterity Test (FDT), Gogola, G., et al., 2013

Variable	Description
handUsed	Which hand the participant used to complete the test, recorded as R for right hand, L for left hand

```
# Read in data set
hands <- read.csv("data/hands.csv")

# Rename variables (odd original coding)
colnames(hands) <- c("time", "speed", "age", "dominant", "gender",
                    "HD", "handUsed")

# Code categorical variables as factors
hands <- # Write over original data with the following
hands %>% # Pipe data set into
mutate(dominant = factor(dominant), # Recode categorical variables as factors
       gender = factor(gender),
       HD = factor(HD),
       handUsed = factor(handUsed))
```

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. Use the provided R markdown file to create a scatterplot to examine the relationship between the speed at which a participant can flip a peg and the age of the participant by filling in the variable names ('speed' and 'age') for xx and yy. 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.

```
hands %>% # Pipe data set into...
ggplot(aes(x = xx, y = yy)) + #Specify variables
  geom_point() + #Add scatterplot of points
  labs(x = "Age (yrs)", #Label x-axis
       y = "Speed (sec/peg)", #Label y-axis
       title = "Scatterplot of Age vs. Speed") + #Be sure to tile your plots
  geom_smooth(method = "lm", se = FALSE) #Add regression line
```

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

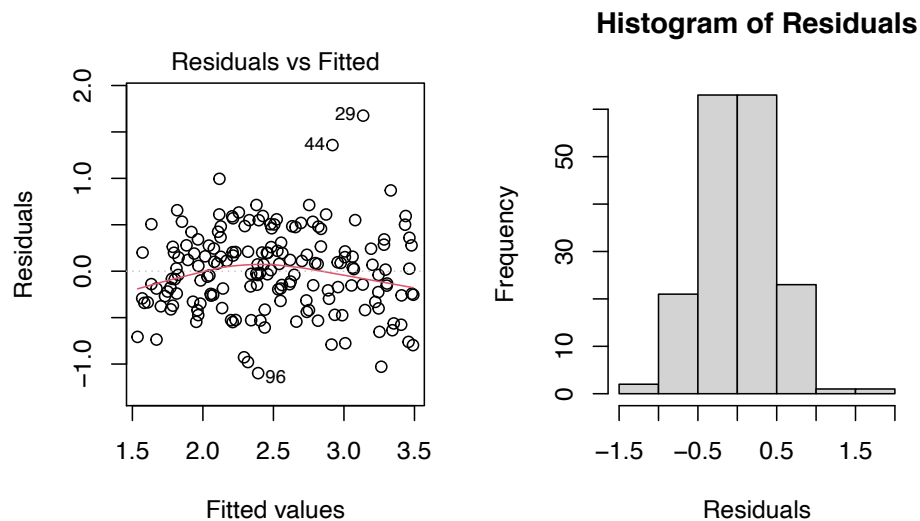
If you indicated there are potential outliers, which points are they?

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 plots will be used to assess the conditions for approximating the data with the t -distribution.



5. Are the conditions met to approximate to use the t -distribution to approximate the sampling distribution of our test statistic?

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

Using the provided R markdown file, enter the response variable into the linear model function for xx and the explanatory variable for yy to get the linear model output.

```
lm.hand <- lm(xx~yy, data=hands) #lm(response~explanatory)
summary(lm.hand)$coefficients
```

8. Using the output from the evaluated R code 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,

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

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

12. Calculate the test statistic for slope. Identify where this calculated value is in the linear model output.

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

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

15. 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.

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

$$b_1 \pm t^* SE(b_1)$$

The t^* multiplier comes from the t -distribution with $n - 2$ df. Recall for a 95% confidence interval, use the 97.5% percentile (95% of the distribution is in the middle, leaving 2.5% in each tail).

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

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

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

Communicate the results and answer the research question

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

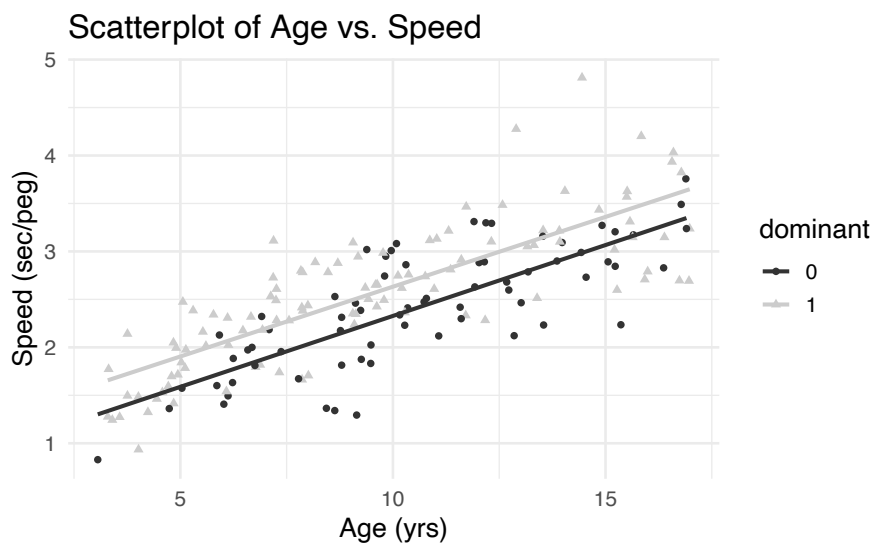
18. Does the p-value agree with the 95% confidence interval? What does each tell you about the null hypothesis?

19. Summarize the results of the study in a written paragraph. Be sure to describe:
 - Summary statistic
 - Test statistic and interpretation
 - P-value and interpretation
 - Confidence interval and interpretation
 - Conclusion (written to answer the research question)
 - Scope of inference

Revisit and look forward

20. Is there an effect due to gender on the linear relationship between age and speed? Explain your answer using the scatterplot below.

```
hands %>% # Pipe data set into...
ggplot(aes(x = age, y = speed, color = dominant))+ #Specify variables
  geom_point(aes(pch = dominant)) + #Add scatterplot of points
  labs(x = "Age (yrs)", #Label x-axis
       y = "Speed (sec/peg)", #Label y-axis
       legend = "Dominant hand", #Label your legend
       title = "Scatterplot of Age vs. Speed") + #Be sure to tile your plots
  geom_smooth(method = "lm", se = FALSE) + #Add regression line
  scale_color_grey() #Make greyscale for printing
```



10.4 Additional notes

Use this space to summarize your thoughts and take additional notes on today's activity.