

# HW9

## Chapter 13

*Your Name Here*

The code below just loads some packages and makes it so that enough digits are printed that you won't get confused by rounding errors.

```
library(dplyr) # functions like summarize
library(ggplot2) # for making plots
library(readr)

options("pillar.sigfig" = 10) # print 10 significant digits in summarize output
```

### Problem 1: Sleuth3 Exercise 13.19

This is a problem about IQ scores. I'm assigning it because I think it's an interesting context and illustrates ideas we've discussed recently. However, whenever we discuss IQ scores it's worth keeping in mind some broader context about IQ scores and what they measure – which is not just intelligence. For example, here's an article from science magazine with some optional reading about some of the complexities of IQ scores: <https://www.sciencemag.org/news/2011/04/what-does-iq-really-measure>

Here's a quote from the book describing the problem:

A 1989 study investigated the effect of heredity and environment on intelligence. From adoption registers in France, researchers selected samples of adopted children whose biological parents and adoptive parents came from either the very highest or the very lowest socio-economic status (SES) categories (based on years of education and occupation). They attempted to obtain samples of size 10 from each combination: (1) high adoptive SES and high biological SES, (2) high adoptive SES and low biological SES, (3) low adoptive SES and high biological SES, and (4) low SES for both parents. It turned out, however, only eight children belonged to combination three. The 38 selected children were given intelligence quotient (IQ) tests. The scores are [loaded in the R code below]. (Data from C. Capron and M. Duyme, "Children's IQs and SES of Biological and Adoptive Parents in a Balanced Cross-fostering Study," *European Bulletin of Cognitive Psychology* 11 (2) (1991): 323-48.) Does the difference in mean scores for those with high and low SES biological parents depend on whether the adoptive parents were high or low SES? If not, how much is the mean IQ score affected by the SES of adoptive parents, and how much is it affected by the SES of the biological parents? Is one of these effects larger than the other? Analyze the data and write a report of the findings.

Here are some things I'll be looking for:

- Create an appropriate plot of the data
- Conduct the analyses suggested by the book above. Each of the three questions posed should be answered with a specific confidence interval and/or hypothesis test for one or more model parameters, with a corresponding interpretation in context.
- Address the scope of inference for this data analysis. Although the book used the word "effect" in framing the questions to answer, is it possible to draw conclusions about cause and effect in this example?

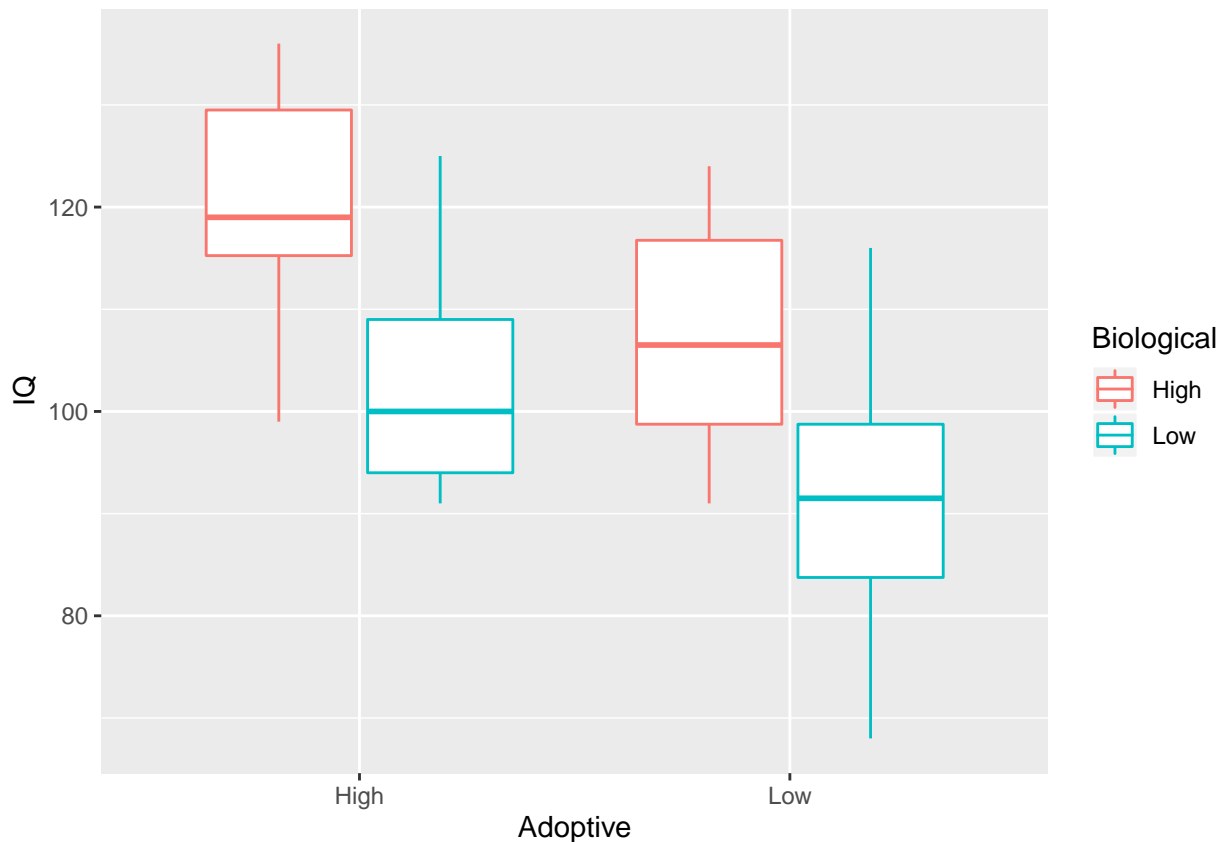
```
nature_nurture <- read_csv("http://www.evanlray.com/data/sleuth3/ex1319_nature_nurture.csv")
```

```
## Parsed with column specification:
## cols(
```

```
## IQ = col_double(),
## Adoptive = col_character(),
## Biological = col_character()
## )
```

Many displays of the data are possible. Here is one option:

```
ggplot(data = nature_nurture, mapping = aes(x = Adoptive, y = IQ, color = Biological)) +
  geom_boxplot()
```



The plot suggests that interactions are likely not necessary in this case.

### Question 1

Our first question is: does the difference in mean scores for those with high and low SES biological parents depend on whether the adoptive parents were high or low SES?

This can be answered with an interactions model:

```
fit_interactions <- lm(IQ ~ Adoptive * Biological, data = nature_nurture)
summary(fit_interactions)
```

```
##
## Call:
## lm(formula = IQ ~ Adoptive * Biological, data = nature_nurture)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -24.400  -9.475  -2.000   8.225  23.600
```

```
##
## Coefficients:
##               Estimate Std. Error t value Pr(>|t|)
## (Intercept)      119.600      4.180  28.611  <2e-16 ***
## AdoptiveLow       -12.100      6.270  -1.930   0.0620 .
## BiologicalLow     -16.000      5.912  -2.706   0.0106 *
## AdoptiveLow:BiologicalLow  0.900      8.618   0.104   0.9174
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 13.22 on 34 degrees of freedom
## Multiple R-squared:  0.3883, Adjusted R-squared:  0.3343
## F-statistic: 7.194 on 3 and 34 DF,  p-value: 0.0007261
```

In this model, if  $\beta_3 = 0$  then the difference in mean scores for those with high and low SES biological parents is the same whether the adoptive parents had high or low SES. To see this, we can work through the model step by step. The starting point is to write down the full model equation:

$$\hat{\mu}(IQ|Adoptive, Biological) = \hat{\beta}_0 + \hat{\beta}_1 AdoptiveLow + \hat{\beta}_2 BiologicalLow + \hat{\beta}_3 AdoptiveLow \times BiologicalLow$$

We also need to remember that in this equation, *AdoptiveLow* and *BiologicalLow* are indicator variables. For each individual,

$$AdoptiveLow = \begin{cases} 1 & \text{if that subject's adoptive parents had low SES} \\ 0 & \text{if that subject's adoptive parents had high SES} \end{cases}$$

and

$$BiologicalLow = \begin{cases} 1 & \text{if that subject's biological parents had low SES} \\ 0 & \text{if that subject's biological parents had high SES} \end{cases}$$

We can use this to obtain the means for each combination of categories in terms of the coefficients  $\beta_0$ ,  $\beta_1$ ,  $\beta_2$ , and  $\beta_3$ :

$$\mu(IQ|Adoptive = "High", Biological = "High") = \beta_0 + \beta_1 0 + \beta_2 0 + \beta_3 0 \times 0 = \beta_0$$

$$\mu(IQ|Adoptive = "Low", Biological = "High") = \beta_0 + \beta_1 1 + \beta_2 0 + \beta_3 \times 1 \times 0 = \beta_0 + \beta_1$$

$$\mu(IQ|Adoptive = "High", Biological = "Low") = \beta_0 + \beta_1 0 + \beta_2 \times 1 + \beta_3 0 \times 1 = \beta_0 + \beta_2$$

$$\mu(IQ|Adoptive = "Low", Biological = "Low") = \beta_0 + \beta_1 \times 1 + \beta_2 \times 1 + \beta_3 \times 1 \times 1 = \beta_0 + \beta_1 + \beta_2 + \beta_3$$

The difference in population means between those with high and low SES biological parents, among those people with high SES adoptive parents, is

$$\mu(IQ|Adoptive = "High", Biological = "High") - \mu(IQ|Adoptive = "High", Biological = "Low") = \beta_0 - (\beta_0 + \beta_2) = -\beta_2$$

Similarly, the difference in population means between those with high and low SES biological parents, among those people with low SES adoptive parents, is

$$\mu(IQ|Adoptive = "Low", Biological = "High") - \mu(IQ|Adoptive = "Low", Biological = "Low") = \beta_0 + \beta_1 - (\beta_0 + \beta_1 + \beta_2 + \beta_3) = -\beta_2 - \beta_3$$

In order for these differences to be equal, we would need to have  $-\beta_2 = -\beta_2 - \beta_3$ , or  $\beta_3 = 0$ .

Therefore, we can answer this question with either a hypothesis test or a confidence interval for  $\beta_3$ . The test would be of the hypotheses

$$H_0 : \beta_3 = 0 \text{ vs. } H_A : \beta_3 \neq 0$$

The p-value for this test is 0.917. The data offer no evidence against the null hypothesis that  $\beta_3 = 0$ , suggesting that the difference in mean scores for those with high and low SES biological parents does not depend strongly on whether the adoptive parents were high or low SES.

## Questions 2 and 3

The second question is: If not, how much is the mean IQ score affected by the SES of adoptive prents, and how much is it affected by the SES of the biological parents?

We decided above that an interaction term is not needed. To simplify interpretations, we will therefore proceed with a model that does not include the interaction term.

```
fit_additive <- lm(IQ ~ Adoptive + Biological, data = nature_nurture)
summary(fit_additive)
```

```
##
## Call:
## lm(formula = IQ ~ Adoptive + Biological, data = nature_nurture)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -24.188  -9.621  -1.788   7.973  23.812
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)    119.388     3.603   33.132 < 2e-16 ***
## AdoptiveLow    -11.624     4.240   -2.741 0.009574 **
## BiologicalLow  -15.576     4.240   -3.674 0.000793 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 13.03 on 35 degrees of freedom
## Multiple R-squared:  0.3881, Adjusted R-squared:  0.3531
## F-statistic: 11.1 on 2 and 35 DF,  p-value: 0.000185
```

```
confint(fit_additive)
```

```
##              2.5 %      97.5 %
## (Intercept)  112.07282 126.703652
## AdoptiveLow  -20.23156  -3.015502
## BiologicalLow -24.18450  -6.968443
```

We are 95% confident that for a fixed level of SES status for the biological parents, the mean IQ score among children whose adoptive parents have low SES is between 3 and 20 points lower than the mean IQ score among children whose adoptive parents have high SES.

We are 95% confident that for a fixed level of SES status for the adoptive parents, the mean IQ score among children whose biological parents have low SES is between 7 and 24 points lower than the mean IQ score among children whose biological parents have high SES.

The effect of status of the biological parents is slightly larger than the effect of status of the adoptive parents, but they are comparable in magnitude.

## Scope of Inference

The people included in this study were not randomly selected and may not be representative of any specific larger population.

The people in this study were not randomly assigned to biological and adoptive parents with different SES, so it would not be reasonable to use these data to draw any conclusions about cause and effect.