

Module 2 Assessment

2/12/2024

Logistics

- Due Feb 20 at 11:59pm
- Please submit both a pdf and .Rmd file to canvas

Salaries data

We will be using the “Salaries” dataset from “An R Companion to Applied Regression” by Fox and Weisberg (2019). It can be loaded into R using the `carData` package as shown below. The dataset has the 2008-2009 salaries from 397 faculty members at a college in the US.

```
# Install the package carData if you don't already have it using the code below  
# install.packages("carData")
```

```
# load the package  
library("carData")
```

```
# load the dataset. Running this should result in a variable named Salaries  
# being created. Note that R is case sensitive  
data(Salaries)
```

```
head(Salaries)
```

```
##      rank discipline yrs.since.phd yrs.service  sex salary  
## 1   Prof          B             19          18 Male 139750  
## 2   Prof          B             20          16 Male 173200  
## 3  AsstProf       B              4           3 Male  79750  
## 4   Prof          B             45          39 Male 115000  
## 5   Prof          B             40          41 Male 141500  
## 6  AssocProf     B              6           6 Male  97000
```

```
# The ? command opens the documentation for the dataset
```

```
?Salaries
```

In the following questions, fit the model which most directly answers the question of interest; i.e., don't take transformations or add additional covariates unless they directly address the question. In many cases, the model you fit will clearly violate some of the assumptions we've discussed in class. For now, just ignore those violations and simply report the estimates.

Question 1 (2pts)

Suppose I am interested in knowing the expected difference in salary between two individuals who differ by 1 year in years since phd. Using the `lm` function, write out the code for the model that directly estimates that quantity. Write one sentence to interpret your finding to a collaborator.

Answer

Question 2 (2pts)

Suppose I am interested in knowing the expected difference in salary between two individuals who differ by 1 year in years since phd, but are the same rank and discipline. Using the `lm` function, write out the code for the model that directly estimates that quantity. Write one sentence to interpret your finding to a collaborator.

Answer

Question 3 (2pts)

Suppose I am interested in knowing the % difference in expected salary between two individuals who differ by 1 year in years since phd. Using the `lm` function, write out the code for the model that directly estimates that quantity. Write one sentence to interpret your finding to a collaborator.

Answer

Question 4 (2pts)

Suppose I am interested in knowing the % difference in expected salary between two individuals who differ by 1 year in years since phd, but are in the same discipline. Using the `lm` function, write out the code for the model that directly estimates that quantity. Write one sentence to interpret your finding to a collaborator.

Answer

Question 5 (2pts)

Write a short explanation which interprets the results of model below to a collaborator.

```
summary(lm(salary ~ relevel(rank, ref = "AssocProf"), data = Salaries))

##
## Call:
## lm(formula = salary ~ relevel(rank, ref = "AssocProf"), data = Salaries)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -68972 -16376  -1580   11755  104773
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)          93876      2954  31.777 < 2e-16
## relevel(rank, ref = "AssocProf")AsstProf  -13100      4131  -3.171  0.00164
## relevel(rank, ref = "AssocProf")Prof       32896      3290   9.997 < 2e-16
##
## (Intercept)          ***
## relevel(rank, ref = "AssocProf")AsstProf **
## relevel(rank, ref = "AssocProf")Prof      ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 23630 on 394 degrees of freedom
```

```
## Multiple R-squared:  0.3943, Adjusted R-squared:  0.3912
## F-statistic: 128.2 on 2 and 394 DF,  p-value: < 2.2e-16
```

Answer

Question 6 (3pts)

I think the association of years since phd and salary is different in discipline A than the association in discipline B. Fit a model which would allow the associations to differ. What is the estimate of the expected difference in salary between two individuals who differ by 1 year in years since phd and are both in discipline A? What is the estimate of the expected difference in salary between two individuals who differ by 1 year in years since phd and are both in discipline B?

Answer

Question 7 (8pts)

The following data records data across 55 mammals. It is available in the R package `mice`, but I've taken out some species with missing data and only have a subset of the original variables. In particular, we have:

- species: Species of animal
- gt: gestational time
- bw: body weight (kg)
- brw: brain weight (g)

Suppose we want to fit a model which explores the relationship between the dependent variable gestational time (gt) and the covariates body weight (bw) and brain weight (brw). Choose an appropriate model to fit by considering potential transformations of the data(it doesn't have to be perfect, but it shouldn't be terrible either). Using various plots, explain why the model you choose is appropriate and discuss whether the linear model assumptions hold. For simplicity, don't consider models with interactions.

```
fileName <- "https://raw.githubusercontent.com/ysamwang/btry6020_sp22/main/lab3/mammal_data.csv"
mammals <- read.csv(fileName)
```

Answer

Question 8 (4pts)

Give an interpretation of the estimated coefficients for covariates corresponding to body weight and brain weight.

Answer