Homework 5

Charles Ancel

9/28/2023

Homework Instructions

Make sure to add your name to the header of the document. When submitting the assignment on Gradescope, be sure to assign the appropriate pages of your submission to each Exercise.

The point value for each exercise is noted in the exercise title.

For questions that require code, please create or use the code chunk directly below the question and type your code there. Your knitted pdf will then show both the code and the output, so that we can assess your understanding and award any partial credit.

For written questions, please provide your answer after the indicated *Answer* prompt.

You are encouraged to knit your file as you work, to check that your coding and formatting are done so appropriately. This will also help you identify and locate any errors more easily.

Homework Setup

We'll use the following packages for this homework assignment. We'll also read in data from a csv file. To access the data, you'll want to download the dataset from Canvas and place it in the same folder as this R Markdown document. You'll then be able to use the following code to load in the data.

```r
library(ggplot2)
library(MASS)
```

Exercise 1: Formatting [5 points]

The first five points of the assignment will be earned for properly formatting your final document. Check that you have:

- included your name on the document
- properly assigned pages to exercises on Gradescope
- selected **page 1 (with your name)** and this page for this exercise (Exercise 1)
- all code is printed and readable for each question
- all output is printed
- generated a pdf file

Exercise 2: A First Model for Professor Evaluation Scores [12 points]

In the article, "Beauty in the classroom: Instructors' pulchritude and putative pedagogical productivity" in the journal *Economics of Education Review* (http://www.sciencedirect.com/science/article/pii/S0272775 704001165), Hamermesh and Parker explored how course evaluations may be associated with a professor's physical appearance. The data were gathered from end of semester student evaluations for a large sample of professors from the University of Texas at Austin. Later, six undergraduate students were asked to look at each professor's photo and rate their physical beauty from 1 to 10.

We will focus on a few variables from this dataset:

- `score`: the overall student rating that each professor received
- `bty_avg`: is the average score provided by the six student raters
- `age`: the age of the professor
- `cls_students`: the size (number of students) in the class
- `cls_perc_eval`: the proportion of the class who completed the evaluations.

part a

Load the data Prof_Evals.csv into your markdown file as `prof_evals`. To do this, you'll want to download the csv file into the same Folder as your Homework5.Rmd file. Then, use a line of code to call in the csv file. Print the number of professors included in the dataset and the number of variables recorded for each professor.

```r
# Load the necessary library
library(readr)

prof_evals <- read_csv("Prof_Evals.csv", show_col_types = FALSE)

cat("Number of professors:", nrow(prof_evals), "\n")
```

```
## Number of professors: 463
```

```r
cat("Number of variables recorded for each professor:", ncol(prof_evals), "\n")
```

```
## Number of variables recorded for each professor: 19
```

part b

Fit a linear model predicting the `score` of a professor from their beauty rating `bty_avg`. Run the summary of your model.

```r
model <- lm(score ~ bty_avg, data = prof_evals)

summary(model)
```

```
##
## Call:
## lm(formula = score ~ bty_avg, data = prof_evals)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -1.9246 -0.3690  0.1420  0.3977  0.9309
##
## Coefficients:
##             Estimate Std. Error t value Pr(>|t|)
## (Intercept)  3.88033    0.07614   50.96  < 2e-16 ***
## bty_avg      0.06664    0.01629    4.09 5.08e-05 ***
```

```
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.5348 on 461 degrees of freedom
## Multiple R-squared:  0.03502,    Adjusted R-squared:  0.03293
## F-statistic: 16.73 on 1 and 461 DF,  p-value: 5.082e-05
```

part c

What is the estimated value of σ for this model?

```r
sigma_value <- summary(model)$sigma

cat("Estimated value of sigma:", sigma_value, "\n")
```

```
## Estimated value of sigma: 0.5348348
```

Answer: Estimated value of sigma: 0.5348348

part d

Calculate the estimated standard error for the estimated slope. That is, calculate the **standard deviation** for the sampling distribution of the estimated slope based on formulas provided in class.

Then, compare this value to that reported in the R output from Exercise 2b.

```r
x_vals <- prof_evals$bty_avg        # Independent variable (bty_avg)
sigma <- summary(model)$sigma       # Residual standard error from the model
sample_size <- nrow(prof_evals)     # Number of observations

xbar <- mean(x_vals)
Sxx <- sum((x_vals - xbar) ^ 2)
beta_1_sd <- sigma / sqrt(Sxx)

cat("Calculated standard error for the slope:", beta_1_sd, "\n")
```

```
## Calculated standard error for the slope: 0.01629091
```

```r
slope_se <- coef(summary(model))["bty_avg", "Std. Error"]

cat("Standard error for the slope:", slope_se, "\n")
```

```
## Standard error for the slope: 0.01629091
```

Comparison: The standard error for the slope (`bty_avg`) as obtained directly from the R output is 0.01629091.

Using the formula provided in class, the calculated standard error for the slope is also 0.01629091.

These values are identical, indicating perfect consistency in the estimation of the standard error for the slope. The match between the two values confirms the accuracy of the R output in estimating this metric and validates the correctness of the formula and calculations provided in class.

Exercise 3: Predicting Professor Evaluation Scores [16 points]

part a

Use a 99% confidence interval to estimate the mean evaluation score for professors with a beauty rating of 6.25 as well as a beauty rating of 9.5.

Which of the two intervals is wider? How might you use the variance formula for $\hat{y}(x)$ to know beforehand which would be wider?

```
beauty_ratings <- c(6.25, 9.5)

predicted_scores <- predict(model, newdata = data.frame(bty_avg = beauty_ratings), interval = "confiden

lower_bounds <- predicted_scores[, "lwr"]
upper_bounds <- predicted_scores[, "upr"]

intervals <- data.frame(Beauty_Rating = beauty_ratings, Lower_Bound = lower_bounds, Upper_Bound = upper_
intervals
```

```
##   Beauty_Rating Lower_Bound Upper_Bound
## 1          6.25    4.196351    4.397277
## 2          9.50    4.289798    4.736970
```

Answer: For a beauty rating of **6.25**, the 99% confidence interval for the mean evaluation score spans from **4.196351** to **4.397277**.

For a beauty rating of **9.50**, the interval is from **4.289798** to **4.736970**.

The interval's width is influenced by the term $(x - \bar{x})^2$ in the variance formula for $\hat{y}(x)$. The interval will be wider for values of x that are farther from the mean $\bar{x}$ of the beauty ratings. This is because as x gets farther from $\bar{x}$, the uncertainty in predicting y increases, leading to a wider confidence interval.

Given that 9.50 is likely farther from the average beauty rating than 6.25, it makes sense that its confidence interval is wider.

part b

Use a 99% prediction interval to predict the evaluation scores for professors with a beauty rating of 7.75. Is this prediction interval reliable? Explain why or why not.

```
predicted_score_7_75 <- predict(model, newdata = data.frame(bty_avg = 7.75), interval = "prediction", le

lower_bound_7_75 <- predicted_score_7_75[1, "lwr"]
upper_bound_7_75 <- predicted_score_7_75[1, "upr"]

c(Lower_Bound = lower_bound_7_75, Upper_Bound = upper_bound_7_75)
```

```
## Lower_Bound Upper_Bound
##    3.004808    5.788731
```

Answer: The 99% prediction interval for an evaluation score with a beauty rating of 7.75 ranges from 3.004808 to 5.788731. This wide interval suggests significant variability in potential scores. Its reliability hinges on the model's assumptions. Given that 7.75 is within our observed beauty ratings, the prediction is relatively trustworthy. However, the broad range implies caution in relying solely on this interval for specific predictions.

part c

A confidence interval for the evaluation scores is reported as: (4.0425, 4.1624). Determine the beauty rating that this confidence interval corresponds to.

```
beta_0 <- coef(model)["(Intercept)"]
beta_1 <- coef(model)["bty_avg"]

y_hat <- (4.0425 + 4.1624) / 2

beauty_rating <- (y_hat - beta_0) / beta_1
beauty_rating
```

```
## (Intercept)
##    3.333244
```

Answer: The confidence interval $(4.0425, 4.1624)$ for the evaluation scores corresponds to a beauty rating of approximately 3.333244. ***

Exercise 4: Professor Evaluation Scores from Multiple Predictors [27 points]

We will fit a multiple linear regression model in this exercise to predict professor evaluation scores from a few additional variables.

part a

Fit a linear model that predicts the evaluation `score` from the following variables:

- `bty_avg`, the average beauty rating given by 6 independent students
- `age`, the age of the professor
- `cls_students`, the size (number of students) in the class
- `cls_perc_eval`, the proportion of the class who completed the evaluations

```
multi_model <- lm(score ~ bty_avg + age + cls_students + cls_perc_eval, data = prof_evals)

summary(multi_model)
```

```
##
## Call:
## lm(formula = score ~ bty_avg + age + cls_students + cls_perc_eval,
##     data = prof_evals)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -1.9590 -0.3426  0.1220  0.3851  1.1556
##
## Coefficients:
##                 Estimate Std. Error t value Pr(>|t|)
## (Intercept)    3.5934345  0.2071146  17.350  < 2e-16 ***
## bty_avg        0.0489457  0.0172216   2.842 0.004682 **
## age           -0.0024375  0.0026346  -0.925 0.355364
## cls_students   0.0005651  0.0003545   1.594 0.111606
## cls_perc_eval  0.0060699  0.0016024   3.788 0.000172 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```
##
## Residual standard error: 0.5276 on 458 degrees of freedom
## Multiple R-squared:  0.06708,    Adjusted R-squared:  0.05893
## F-statistic: 8.233 on 4 and 458 DF,  p-value: 2.033e-06
```

part b

Write out the fitted linear model. Make sure that the variables are clearly defined for your written model.

Answer: Based on the provided output for the multiple linear regression model, you can write out the fitted model as follows:

Fitted Linear Model:

$$\text{score} = 3.5934 + 0.0489 \times \text{bty\_avg} - 0.0024 \times \text{age} + 0.0006 \times \text{cls\_students} + 0.0061 \times \text{cls\_perc\_eval}$$

Where: - score is the evaluation score of the professor. - bty_avg is the average beauty rating given by 6 independent students. - age is the age of the professor. - cls_students is the size (number of students) in the class. - cls_perc_eval is the proportion of the class who completed the evaluations.

part c

Write out interpretations for the following coefficients in the model:

- intercept
- slope for beauty average

Answer: - The intercept value of 3.5934 is the expected evaluation score when all predictor variables are zero (though this interpretation isn't practically meaningful for all variables). - For each unit increase in bty_avg, the evaluation score is expected to increase by 0.0489 units, holding all other variables constant. - For each additional year in age, the evaluation score is expected to decrease by 0.0024 units, holding all other variables constant. - For each additional student in cls_students, the evaluation score is expected to increase by 0.0006 units, holding all other variables constant. - For each unit increase in the proportion of cls_perc_eval, the evaluation score is expected to increase by 0.0061 units, holding all other variables constant.

part d

Does the intercept provide a meaningful value? Is it reliable? Explain.

Answer: The intercept value of 3.5934 does not provide a meaningful interpretation in the context of this model. It's not practical or meaningful to consider a scenario where a professor has an age of 0 years, a beauty rating of 0, a class size of 0 students, and a 0 proportion of class evaluations. Therefore, while the intercept might be statistically significant, its real-world interpretation is not meaningful.

In terms of reliability, the intercept is reliable in a statistical sense, as indicated by its significance level (p-value < 0.05). However, its practical relevance is questionable given the unrealistic scenario it represents.

part e

We'd like to learn if the beauty score is a significant linear predictor for the overall evaluation score, after controlling for age, class size, and percentage of evaluations completed. For the hypothesis test described above, report the:

- null and alternative hypotheses
- test statistic
- p-value

```
model_summary <- summary(multi_model)

t_statistic_bty_avg <- model_summary$coefficients["bty_avg", "t value"]

p_value_bty_avg <- model_summary$coefficients["bty_avg", "Pr(>|t|)"]

c(Test_Statistic = t_statistic_bty_avg, P_Value = p_value_bty_avg)
```

```
## Test_Statistic        P_Value
##     2.842115908      0.004681627
```

Answer: ### Hypothesis Test:

Let's set up the hypotheses:

- $\beta_{\text{bty\_avg}}$ is the coefficient corresponding to the beauty score in the linear model.

Null Hypothesis (H_0): The beauty score is not a significant linear predictor for the evaluation score after controlling for the other variables.

$$H_0 : \beta_{\text{bty\_avg}} = 0$$

Alternative Hypothesis (H_a): The beauty score is a significant linear predictor for the evaluation score after controlling for the other variables.

$$H_a : \beta_{\text{bty\_avg}} \neq 0$$

Test Statistic and p-value:

From the provided output of the regression model:

- Test statistic for `bty_avg`: **2.842**
- p-value for `bty_avg`: **0.004682**

Answer:

For testing the significance of the beauty score in predicting the evaluation score:

- Null Hypothesis (H_0): The beauty score is not a significant predictor.
- Alternative Hypothesis (H_a): The beauty score is a significant predictor.
- Test statistic for `bty_avg`: 2.842
- p-value for `bty_avg`: 0.004682

Given the low p-value (less than 0.05), we reject the null hypothesis, suggesting that the beauty score is a significant linear predictor for the evaluation score after controlling for age, class size, and percentage of evaluations completed. ## part f

Interpret the R^2 value for this model.

Answer: The R^2 value for this model is 0.06708, indicating that only about 6.7% of the variability in professor evaluation scores is explained by the model's predictors (beauty score, age, class size, and evaluation completion percentage). The remaining 93.3% of variability is due to other factors not captured in the model.

Exercise 5: Predictions for Professors A & Z [20 points]

part a

Calculate the expected evaluation score values for the following two professors with the given features:

- Professor Z, who has an average beauty score of 6.25, an age of 52, a class size of 61, and 83% of the class who completed the evaluations.
- Professor A, who has an average beauty score of 9.5, an age of 34, a class size of 270, and 96% of the class who completed the evaluations.

Print the answers for Professor Z and Professor A, and complete the following statements.

```
bty_avg_Z <- 6.25
age_Z <- 52
cls_students_Z <- 61
cls_perc_eval_Z <- 0.83

score_Z <- coef(multi_model)["(Intercept)"] +
           coef(multi_model)["bty_avg"] * bty_avg_Z +
           coef(multi_model)["age"] * age_Z +
           coef(multi_model)["cls_students"] * cls_students_Z +
           coef(multi_model)["cls_perc_eval"] * cls_perc_eval_Z

bty_avg_A <- 9.5
age_A <- 34
cls_students_A <- 270
cls_perc_eval_A <- 0.96

score_A <- coef(multi_model)["(Intercept)"] +
           coef(multi_model)["bty_avg"] * bty_avg_A +
           coef(multi_model)["age"] * age_A +
           coef(multi_model)["cls_students"] * cls_students_A +
           coef(multi_model)["cls_perc_eval"] * cls_perc_eval_A

c(Professor_Z_Score = score_Z, Professor_A_Score = score_A)
```

```
## Professor_Z_Score.(Intercept) Professor_A_Score.(Intercept)
##                      3.812105                     4.133947
```

Answer:

- Professor Z has an expected evaluation score of: 3.812105
- Professor A has an expected evaluation score of: 4.133947

part b

Suppose Professor Z has an evaluation score of 4.6, and Professor A has an evaluation score of 3.7. Calculate and report the residual for each professor.

```
residual_Z <- 4.6 - score_Z

residual_A <- 3.7 - score_A

c(Residual_Professor_Z = residual_Z, Residual_Professor_A = residual_A)
```

```
## Residual_Professor_Z.(Intercept) Residual_Professor_A.(Intercept)
##                        0.7878952                       -0.4339474
```

Answer:

- Professor Z has a residual of: 0.7878952
- Professor A has a residual of: −0.4339474

part c

The point estimates from **part a** may have some error, so let's instead create intervals for the evaluation score that allow for some wiggle room.

First, create an 85% interval that we anticipate should capture the mean evaluation score for professors with the same characteristics as Professor A.

```
prof_A_data <- data.frame(bty_avg = bty_avg_A,
                          age = age_A,
                          cls_students = cls_students_A,
                          cls_perc_eval = cls_perc_eval_A)

conf_interval_A <- predict(multi_model, newdata = prof_A_data,
                           interval = "confidence", level = 0.85)

conf_interval_A
```

```
##        fit     lwr      upr
## 1 4.133947 3.91644 4.351455
```

part d

Calculate a 75% interval that should contain the true evlaution scores for 75% of professors with the same characteristics as Professor Z.

```
prof_Z_data <- data.frame(bty_avg = bty_avg_Z,
                          age = age_Z,
                          cls_students = cls_students_Z,
                          cls_perc_eval = cls_perc_eval_Z)

pred_interval_Z <- predict(multi_model, newdata = prof_Z_data,
                           interval = "prediction", level = 0.75)

pred_interval_Z
```

```
##        fit     lwr      upr
## 1 3.812105 3.18642 4.437789
```

Exercise 6: Comparing Professor Models [20 points]

For this exercise, we will compare the simple linear model from Exercise 2 to the multiple linear model from Exercise 4 on this assignment.

part a

For the two professor models, which do you expect to have a higher R^2 value (if either)? Explain your answer. Report and compare the actual R^2 values for these two models. No need to recompute these values, although you can if helpful.

```
model_r2 <- summary(model)$r.squared
multi_model_r2 <- summary(multi_model)$r.squared

c(Model_R2 = model_r2, Multiple_Model_R2 = multi_model_r2)
```

```
##         Model_R2 Multiple_Model_R2
##        0.03502322        0.06708106
```

Answer: Based on the principle that adding predictors tends to increase the R^2 value, we'd expect the multiple linear regression model from Exercise 4 to have a higher R^2. Upon comparing the actual values, the multiple linear regression model does have a higher R^2 value (0.06708) than the simple linear model (0.03502). This indicates that the multiple linear model explains a greater proportion of the variance in the evaluation scores compared to the simple linear model.

part b

Calculate the SSE and SST values for these two models. Do the observed results match what you would anticipate? Explain.

```
residuals <- residuals(model)
SSE <- sum(residuals^2)
SST <- sum((prof_evals$score - mean(prof_evals$score))^2)

residuals_multi <- residuals(multi_model)
SSE_multi <- sum(residuals_multi^2)
SST_multi <- sum((prof_evals$score - mean(prof_evals$score))^2)  # SST remains the same for both models

c(SSE = SSE, SSE_Multi = SSE_multi, SST = SST_multi)
```

```
##       SSE SSE_Multi       SST
##  131.8683  127.4874  136.6543
```

Answer: - SSE for the simple model (using only `bty_avg`): 131.8683 - SSE for the multiple regression model (using multiple predictors): 127.4874 - SST (common for both models): 136.6543

The SSE quantifies the model's unexplained variability, with a lower value indicating better fit:

- SSE for the simple model: 131.8683
- SSE for the multiple regression model: 127.4874

The SST (Total Sum of Squares) measures the total variability in the response variable. It remains constant at 136.6543 for both models since it's determined by the response variable, not the predictors.

The reduction in SSE from the simple to the multiple regression model shows that the additional predictors in the latter capture more of the total variability (SST), even if the improvement is modest.

part c

For each of these two models, report the dimensions of the X and y matrices that would be used to calculate $\hat{\beta}$. What are the degrees of freedom associated with each of these models?

```
# For the simple model
num_obs <- nrow(prof_evals)
num_parameters <- length(coef(model))
df <- num_obs - num_parameters

# For the multiple regression model
num_parameters_multi <- length(coef(multi_model))
df_multi <- num_obs - num_parameters_multi

c(DF = df, DF_Multi = df_multi)
```

```
##       DF DF_Multi
##      461      458
```

Answer: For the simple model: - Dimensions of y: 463×1 (463 observations) - Dimensions of X: 463×2 (463 observations, 1 intercept + 1 predictor) - Degrees of freedom: 461

For the multiple regression model: - Dimensions of y: 463×1 (463 observations) - Dimensions of X: 463×5 (463 observations, 1 intercept + 4 predictors) - Degrees of freedom: 458

The degrees of freedom represent the number of independent pieces of information that go into the estimation of the parameters (like slopes and intercepts). As more predictors are added to a model, the degrees of freedom decrease, which is reflected in the results here.

part d

Thinking critically about this dataset, do you have any concerns about how it could be used? Any variables in the dataset that you'd like to know more about, or any variables you'd like to have added to the dataset?

You do not need to answer all of these questions, but I am hoping that you will carefully and thoughtfully consider our professor dataset and its applications.

Answer: ### Concerns:

1. **Bias in Beauty Ratings**: Beauty is subjective and can be influenced by cultural, societal, and individual biases. Using it as a predictor might introduce these biases into our model.
2. **Ethical Implications**: Using beauty ratings to predict or evaluate professional performance might be ethically questionable. It could perpetuate stereotypes and bias in academic or professional settings.
3. **Overfitting**: While adding more predictors can improve the R^2 value, it's important to ensure we aren't overfitting the model. An overly complex model might not generalize well to new data.
4. **Causality vs Correlation**: Even if there's a correlation between beauty and evaluation scores, it doesn't imply causation. There could be lurking variables that influence both beauty ratings and evaluation scores.

Variables to Know More About:

1. **Beauty Rating Process**: How were the beauty ratings determined? Was it a consistent process with clear guidelines, or was it subjective?
2. **Evaluation Metrics**: What specific criteria were used in the evaluation? Were evaluations anonymous? Were students given any guidelines?
3. **Demographics**: More details on the demographics of students and professors could help adjust for potential biases. This includes information on race, gender, socioeconomic status, etc.

Desired Additional Variables:

1. **Teaching Experience**: The number of years a professor has been teaching could be a significant predictor of evaluation scores.
2. **Course Difficulty**: It's possible that professors teaching more challenging courses might receive different evaluations.
3. **Feedback Details**: Instead of just a score, detailed feedback or comments from students could provide more context.
4. **Class Time & Format**: Was the class in-person, online, a mix? Was it a morning or evening class? Such factors might influence student evaluations.

In conclusion, while the dataset provides interesting insights, it's essential to approach its conclusions with caution, especially when considering real-world applications or policy changes based on the findings.