Quiz 1 Honors Opportunity for 432 in Spring 2019

Thomas E. Love

due at 2 PM Monday 2019-03-25. Version 2019-03-19

Instructions

This honors opportunity requires the completion of a set of 5 questions to our satisfaction. All students are eligible for this opportunity, and we strongly recommend that everyone complete the activity. After your submission, your responses to each of the 5 questions listed below will be scored as either satisfactory (in which case you will receive a point for that question) or not satisfactory (0 points.) To receive the reward for this honors opportunity, you must achieve a score of 4 or higher.

1. You will get two attempts, assuming your work on the initial attempt is received on time (by 2 PM on Monday 2019-03-25), **and** contains a reasonable effort to generate a response for each of the 5 questions.
   - We will grade your work and return a score to you by Thursday's class on 2019-03-28, indicating your score on each question (which will be 0 or 1) and your overall score on the first attempt (out of a possible 5).
   - If this score is 4 or higher (out of a possible 5) you will have completed the opportunity successfully.
   - If this score is below 4, but you made an honest effort at each of the five questions (as we will indicate on our response to your work), then we will let you redo the work once (and that revision will be due at 2 PM on Monday 2019-04-01). A score of 4 or more points on the revision will complete the opportunity successfully.
2. You can discuss this assignment with the teaching assistants and with Dr. Love, and **no one else** (until after 2019-04-01). You are permitted to ask questions through `431-help` or in person, and to do during the preparation of your first or second attempt.
3. You will submit the HTML result of your R Markdown work for this assignment to Canvas. You do not need to submit anything else, including your R Markdown code - just the HTML will be sufficient. Be sure to submit your initial attempt by the deadline of 2 PM on Monday 2019-03-25.

Reward for Completing the Opportunity Successfully

- If your score on Quiz 1 was 46 or higher (out of 60), then successful completion of this opportunity will cause Dr. Love to raise your grade on Quiz 1 by 5 points.
- If your score on Quiz 1 was below 46 (out of 60), then successful completion of this opportunity will cause Dr. Love to raise your grade on Quiz 1 by 80% of the distance between your current score and a score of 51/60 on the Quiz[1].

The packages used in preparing this activity are listed below.

```
library(janitor)
library(leaps)
library(rms)
library(broom)
library(tidyverse)
```

The `honors` data file used in preparing this activity is available on the Data and Code page of our web site.

---

[1]Note that because of the initial distribution of grades, this means that all students who scored below 46 initially are eligible for an increase of more than 5 points.

Setup for Questions 1-3

The data in the `honors.csv` file contain information for 255 subjects on:

- a binary `outcome` (Good or Bad),
- a `size` (quantitative, between 60 and 200, in millimeters),
- an indicator of whether a `treatment` was used (1 = treatment was used or 0 = treatment was not used), and
- a specification as to which of five ordered groups (1 = lowest, 5 = highest) by socio-economic status (`ses_group`) the subject falls in, along with
- a subject ID.

Import the data into the `honors` frame, and then fit a logistic regression model to predict the log odds of a Good `outcome` using the subject's `size`, `treatment` status and `ses_group`, treating the `ses_group` as a categorical variable. Questions 1-3 use the data as they are, without imputation. One such analysis yields these results.

```
honors <- read_csv("honors.csv") %>% clean_names() %>%
  mutate(goodoutcome = ifelse(outcome == "Good", 1, 0),
         ses_group = factor(ses_group))
```

Note that the fitting of the actual `m1` is not shown.

```
tidy(m1, conf.int = TRUE, conf.level = 0.95,
    exponentiate = TRUE) %>% knitr::kable(digits = 3)
```

term	estimate	std.error	statistic	p.value	conf.low	conf.high
(Intercept)	0.155	0.822	-2.267	0.023	0.030	0.751
size	1.010	0.006	1.677	0.093	0.998	1.022
treatment	0.557	0.291	-2.010	0.044	0.314	0.985
ses_group2	1.226	0.601	0.339	0.735	0.378	4.087
ses_group3	1.384	0.523	0.622	0.534	0.510	4.051
ses_group4	1.110	0.558	0.188	0.851	0.378	3.442
ses_group5	1.416	0.511	0.680	0.496	0.535	4.068

The output below comes from another approach to fitting the identical logistic regression model that we saw previously. I'll call this model `m1a`, to emphasize that it contains the same outcome and predictors, put together in the same way, again without any imputation.

```
summary(m1a) %>% knitr::kable(digits = 3)
```

	Low	High	Diff.	Effect	S.E.	Lower 0.95	Upper 0.95	Type
size	100.525	135.5	34.975	0.346	0.207	-0.058	0.751	1
Odds Ratio	100.525	135.5	34.975	1.414	NA	0.943	2.120	2
treatment	0.000	1.0	1.000	-0.585	0.291	-1.156	-0.015	1
Odds Ratio	0.000	1.0	1.000	0.557	NA	0.315	0.986	2
ses_group - 1:5	5.000	1.0	NA	-0.348	0.511	-1.350	0.655	1
Odds Ratio	5.000	1.0	NA	0.706	NA	0.259	1.924	2
ses_group - 2:5	5.000	2.0	NA	-0.144	0.481	-1.086	0.798	1
Odds Ratio	5.000	2.0	NA	0.866	NA	0.337	2.220	2
ses_group - 3:5	5.000	3.0	NA	-0.023	0.379	-0.765	0.720	1
Odds Ratio	5.000	3.0	NA	0.978	NA	0.465	2.054	2
ses_group - 4:5	5.000	4.0	NA	-0.243	0.425	-1.075	0.589	1
Odds Ratio	5.000	4.0	NA	0.784	NA	0.341	1.802	2

An Important Note

Note that every Question requires a response in complete English sentences, supplemented by the code you develop in R Markdown. No Question will be marked as satisfactory unless a complete response (code + English sentences) is provided in a clear way in your HTML output.

1 Question 1 (1 point)

What do you conclude from the `m1a` summary about the odds ratio and confidence interval associated with the `treatment` variable? To answer this question, provide a complete description (in complete English sentences) of the odds ratio effect associated with `treatment` in the `summary(m1a)` output. This should require two or three sentences.

2 Question 2 (1 point)

Why is the odds ratio shown in the `m1a` output for `size` different from that shown in the earlier presentation using `tidy` for the `m1` model? Keep your answer to two or three sentences.

3 Question 3 (1 point)

Using the `honors` data (again without imputing any missing values), obtain a Spearman ρ^2 plot and use it to identify a good way to add a single additional non-linear term to this model (**you may spend only a single additional degree of freedom**). What addition would you make? This should be explained in one or more complete English sentences.

Setup for Questions 4 and 5

Using the `honors` data, fit the model you specified in Question 3 (including the non-linear term), while also accounting for missing data using **multiple imputation**. Set your seed to be `432432`, and impute the predictors that need imputation using all available observations on all available variables, with 20 imputations. Be sure to show the code you used to fit your imputation model and your outcome model in your HTML file. Call the imputation model `model_imp` and the outcome model `m2`.

4 Question 4 (1 point)

If Harry was size 100 mm and fell into group 4 in socio-economic status and Sally was size 120 mm and fell into group 3 in socio-economic status, and both Harry and Sally received the treatment, which of the two would have a larger probability of a Good outcome according to your model? How do you know? Your answer should be given in complete English sentences.

5 Question 5 (1 point)

Write a few English sentences describing how the addition of imputation and a non-linear term changes (or doesn't change) the conclusions that you draw in `m2` from what you saw in the `m1` (or, equivalently, the `m1a`) model examined earlier.