Final Review

SOCI 385

Matt Lawrence

2023-12-07

Reviewing the YAML header

Be sure to add your name (in the author field) and a title (in the title field). That's the bare minimum. Adding a subtitle and date would also be nice.

For your final project, use pdf as the format option.

Note that I have added `execute: echo: false`. This will suppress all the code from appearing in your rendered file. It's an easy way to make publication-ready documents.

Chunk-Specific Options

Turning off the echo keeps code from rendering but does not keep warnings and messages from appearing. The extra output they create is most commonly associated with loading packages, summarise(), and ggplot(). When you have code chunks using any of those functions, it's a good idea to suppress warnings and/or messages like the chunk below does:

The `echo: false` option in the header is a global command. You can change it in individual code chunks using the same syntax that controls messages and warnings. For example, by adding `echo: true` to the code chunk below, the code will render. As a default for your final project, don't render the code (but do submit your qmd file so I can see your code!).

```
round(prop.table(table(final$crimsent)),1)
```

```
About the right amount        Too little time          Too much time
                   0.4                    0.3                    0.3
```

Table 1: Distribution of Responses to Cancel Culture Question

Response	Proportion
Accountable	0.59
Punish	0.41

Often when you are working on a project, there will be output that you want to see that the reader of the final project does not need to see. You can suppress that output in the rendered document by adding `output: false` to the code chunk. When this document renders, you won't see the output from the following chunk or the code (since the `echo: false` global command from the header is in operation again).

Before moving on, the examples below use the `crimsent_toomuch` binary variable. Let's create it.

Use Kable For Most Tables

The `kableExtra` package is a nice way to format most tables. It's a good package to load at the top of a notebook so you can use it later. Recall that you have to laod it before using any of its functions for a document to render correctly.

Kable will work with most tables. The simplest syntax adds the `kbl()` function after a pipe, like in this example following a proportion table:

Var1	Freq
Accountable	0.59
Punish	0.41

There are a few standard options to improve the kable output:

Pander Formats Other Output

The `pander` package is great for cleaning up output from tests like t.test() and chisq.test(). Install it, remembering to put a hashtag in front of the install line after doing so or else the file won't render.

Example of pander with t.test results:

Table 2: Welch Two Sample t-test: `final$weight_w92[final$gender == "Man"]` and `final$weight_w92[final$gender == "Woman"]` (continued below)

Test statistic	df	P value	Alternative hypothesis	mean of x
2.907	6298	0.00366 * *	two.sided	1.019

mean of y
0.9346

Example of pander with chisq.test results:

Table 4: Pearson's Chi-squared test: `final$ideology` and `final$cancul1`

Test statistic	df	P value
880.4	4	2.911e-189 * * *

Regression Output With Huxtable

The raw output of summary tables and regression output can be messy in rendered files. We have previously seen how to use `kable` to format some tables. It works well for simple tables but not as well for regression output. Fortunately, there is a different package - `huxtable` - that works well for . The package is highly customizable; we'll go over just a few options today. For more details about the package, click here.

Install `huxtable` then load it using `library()`.

Summaries of linear regression models have a lot of information and can be messy in rendered files. One option to clean them up is to use the `huxreg()` function in the `huxtable` package.

To see how huxreg compares to the output we are used to, regress `crimsent_toomuch` on `cancul1`, save the model as `model1`, and look at the summary:

Now that you are comfortable interpreting summary outputs, you should continue using the traditional summary outputs to get a sense of your models. But replace `summary()` with `huxreg()` when you want to render your file.

	(1)
(Intercept)	0.358 ***
	(0.007)
cancul1Punish	-0.161 ***
	(0.011)
N	6823
R2	0.030
logLik	-4195.778
AIC	8397.557

*** p < 0.001; ** p < 0.01; * p < 0.05.

Here are some options that are good to add to regression tables:

- Change the title with `set_caption()`. Note that all tables titled with this function will be numbered consecutively in the final report.
- The `huxtable` package has several different themes. The `theme_article()` and `theme_compact()` are the two I use the most. Try `theme_article()` first; it should work well with short tables. If your table is too long for a single page (or single slide), switch to `theme_compact()`.
- We don't need all the model statistics at the bottom of the table. Let's only keep the number of observations, which we can do with the `statistics = c()` option. The package saves the number of observations as an object called `nobs`; we'll rename it `N. obs.` in the table.
- We can also rename the coefficient names using the `coefs = c()` option. Each coefficient in the model (including the Intercept) needs to be listed here or else it will be omitted from the table.

Table 5: A Better Title

	(1)
(Intercept)	0.358 ***
	(0.007)
Cancel culture = punish	-0.161 ***
	(0.011)
Number of Observations	6823
***** p < 0.001; ** p < 0.01; * p < 0.05.**	

We can include more than one model in a table too. Let's save a model with a control variable and a model with an interaction.

And then include our three models in a huxreg function that also includes several of the changes described above.

Table 6: Individuals Convicted of Crimes Spend Too Much Time in Prison

	(1)	**(2)**	**(3)**
(Intercept)	0.358 ***	0.124 ***	0.121 ***
	(0.007)	(0.012)	(0.015)
Cancel culture = Punish	-0.161 ***	-0.034 **	-0.030
	(0.011)	(0.011)	(0.019)
Ideology = Liberal		0.374 ***	0.375 ***
		(0.016)	(0.020)
Ideology = Moderate		0.163 ***	0.165 ***
		(0.013)	(0.018)
Ideology = Very Conservative		-0.009	-0.007
		(0.019)	(0.033)
Ideology = Very Liberal		0.624 ***	0.640 ***
		(0.020)	(0.024)
Punish X Liberal			0.006
			(0.035)
Punish X Moderate			-0.001
			(0.026)
Punish X Very Conservative			-0.003
			(0.041)
Punish X Very Liberal			-0.078
			(0.049)
Number of Observations	6823	6823	6823

*** $p < 0.001$; ** $p < 0.01$; * $p < 0.05$.