Regression tables with `{gtsummary}`

On to Table 2!

Univariate regressions

Fit a series of univariate regressions of income on other variables.

```
tbl_uvregression(
  nlsy,
  y = income,
  include = c(sex_cat,
              eyesight
  method = lm)
```

Characteristic	N	Beta	95% CI [1]	p-value
sex_cat	10,195			
Male		—	—	
Female		-358	-844, 128	0.15
race_eth_cat	10,195			
Hispanic		—	—	
Black		-1,747	-2,507, -988	<0.001
Non-Black, Non-Hispanic		3,863	3,195, 4,530	<0.001
eyesight_cat	6,789			
Excellent		—	—	
Very good		-578	-1,319, 162	0.13

[1] CI = Confidence Interval

Characteristic	N	Beta	95% CI [1]	p-value
Good		-1,863	-2,719, -1,006	<0.001
Fair		-4,674	-5,910, -3,439	<0.001
Poor		-6,647	-9,154, -4,140	<0.001
age_bir	4,773	595	538, 652	<0.001

[1] CI = Confidence Interval

Can also do logistic regression

```
tbl_uvregression(
  nlsy,
  y = glasses,
  include = c(sex_cat,
              eyesight
  method = glm,
  method.args = list(f
  exponentiate = TRUE)
```

Characteristic	N	OR [1]	95% CI [1]	p-value
sex_cat	8,450			
Male		—	—	
Female		1.97	1.81, 2.15	<0.001
race_eth_cat	8,450			
Hispanic		—	—	
Black		0.76	0.67, 0.86	<0.001
Non-Black, Non-Hispanic		1.34	1.19, 1.50	<0.001
eyesight_cat	8,444			
Excellent		—	—	
Very good		0.93	0.84, 1.03	0.2
Good		0.95	0.84, 1.07	0.4
Fair		0.81	0.68, 0.96	0.016
Poor		1.15	0.81, 1.63	0.4
age_bir	5,813	1.02	1.01, 1.03	<0.001

[1] OR = Odds Ratio, CI = Confidence Interval

We probably want to do some multivariable regressions

```
linear_model <- lm(income ~ sex_cat + age_bir + race_eth_c
                   data = nlsy)
```

```
linear_model_int <- lm(income ~ sex_cat*age_bir + race_eth
                   data = nlsy)
```

```
logistic_model <- glm(glasses ~ eyesight_cat + sex_cat + i
                      data = nlsy, family = binomial())
```

gtsummary::tbl_regression()

```
tbl_regression(
  linear_model,
  intercept = TRUE,
  label = list(
    sex_cat ~ "Sex",
    race_eth_cat ~ "Race/ethnicity",
    age_bir ~ "Age at first birth"
  ))
```

Characteristic	Beta	95% CI [1]	p-value
(Intercept)	2,147	493, 3,802	0.011
Sex			
Male	—	—	
Female	25	-654, 705	>0.9
Age at first birth	438	381, 495	<0.001
Race/ethnicity			
Hispanic	—	—	
Black	-772	-1,714, 171	0.11
Non-Black, Non-Hispanic	7,559	6,676, 8,442	<0.001

[1] CI = Confidence Interval

gtsummary::tbl_regression()

```
tbl_regression(
  logistic_model,
  exponentiate = TRUE,
  label = list(
    sex_cat ~ "Sex",
    eyesight_cat ~ "Eyesight",
    income ~ "Income"
  ))
```

Characteristic	OR [1]	95% CI [1]	p-value
Eyesight			
Excellent	—	—	
Very good	0.92	0.82, 1.03	0.2
Good	0.92	0.80, 1.05	0.2
Fair	0.80	0.66, 0.98	0.028
Poor	1.03	0.69, 1.53	0.9
Sex			
Male	—	—	
Female	2.04	1.85, 2.25	<0.001
Income	1.00	1.00, 1.00	<0.001

[1] OR = Odds Ratio, CI = Confidence Interval

Arguments

Argument	Description
`label=`	modify variable labels in table
`exponentiate=`	exponentiate model coefficients
`include=`	names of variables to include in output. Default is all variables
`show_single_row=`	By default, categorical variables are printed on multiple rows. If a variable is dichotomous and you wish to print the regression coefficient on a single row, include the variable name(s) here.
`conf.level=`	confidence level of confidence interval
`intercept=`	indicates whether to include the intercept
`estimate_fun=`	function to round and format coefficient estimates
`pvalue_fun=`	function to round and format p-values
`tidy_fun=`	function to specify/customize tidier function

You could put several together

```r
tbl_no_int <- tbl_regression(
  linear_model,
  intercept = TRUE,
  label = list(
    sex_cat ~ "Sex",
    race_eth_cat ~ "Race/ethnicity",
    age_bir ~ "Age at first birth"
  ))

tbl_int <- tbl_regression(
  linear_model_int,
  intercept = TRUE,
  label = list(
    sex_cat ~ "Sex"
```

You could put several together

```
tbl_merge(list(tbl_no_int, tbl_int),
          tab_spanner = c("**Model 1**", "**Model 2**"))
```

Characteristic	Model 1			Model 2		
	Beta	95% CI [1]	p-value	Beta	95% CI [1]	p-value
(Intercept)	2,147	493, 3,802	0.011	4,064	1,884, 6,245	<0.001
Sex						
Male	—	—		—	—	
Female	25	-654, 705	>0.9	-3,635	-6,432, -838	0.011
Age at first birth	438	381, 495	<0.001	364	285, 443	<0.001
Race/ethnicity						
Hispanic	—	—		—	—	
Black	-772	-1,714, 171	0.11	-759	-1,701, 183	0.11
Non-Black, Non-Hispanic	7,559	6,676, 8,442	<0.001	7,550	6,668, 8,433	<0.001
Sex/age interaction						

[1] CI = Confidence Interval

Characteristic	Model 1			Model 2		
	Beta	95% CI [1]	p-value	Beta	95% CI [1]	p-value
Female * Age at first birth				149	39, 260	0.008

[1] CI = Confidence Interval

Exercises

1. Open the script with some examples.
2. Run the examples.

3-6. You're on your own again!

Extra time? Start a table using the data you downloaded for your final project! Make sure you switch to that R project!

15:00