Excel Literacy

Adam Patterson

2023-02-22

Used log +1 to account for students with 0 percent

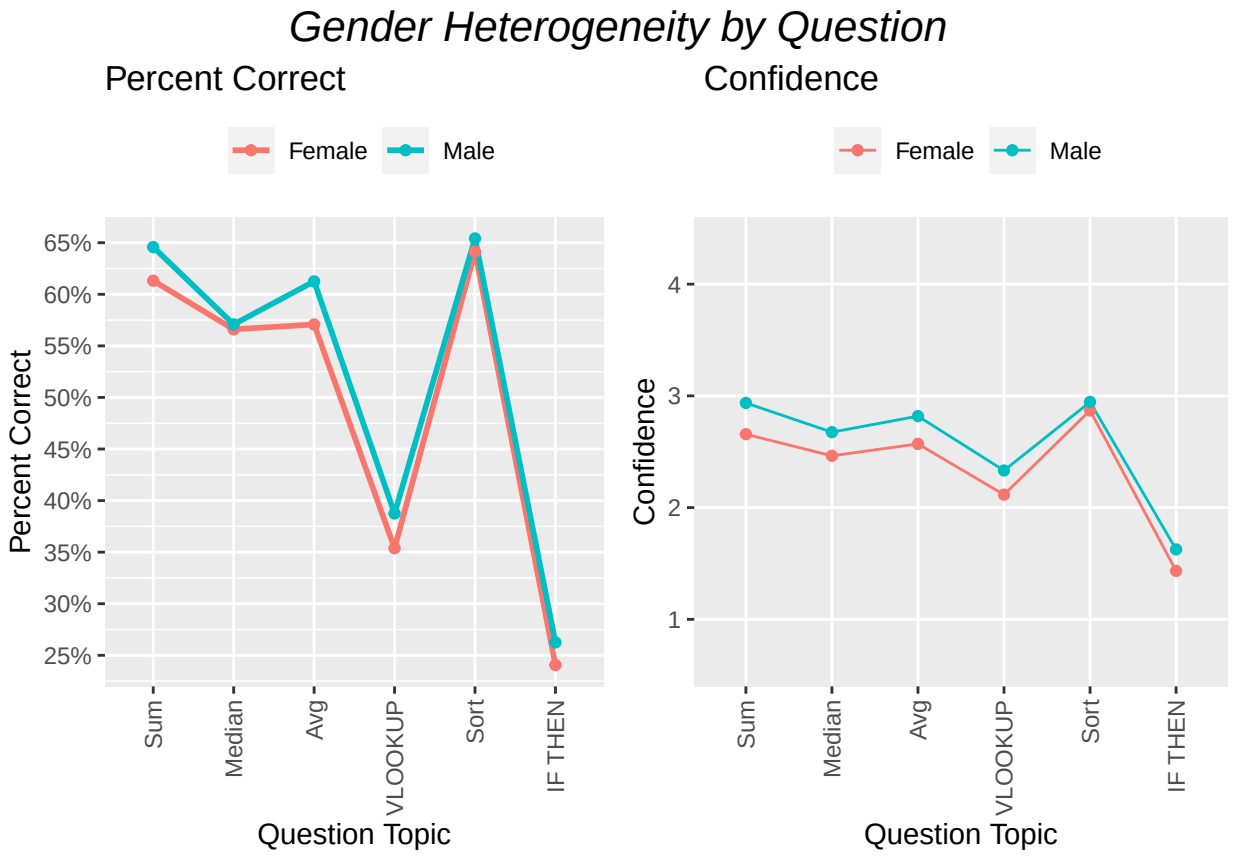
```
stargazer(means, type="latex", title = "Descriptive Statistics of Variables")
```

% Table created by stargazer v.5.2.3 by Marek Hlavac, Social Policy Institute. E-mail: marek.hlavac at gmail.com % Date and time: Wed, Feb 22, 2023 - 17:32:10

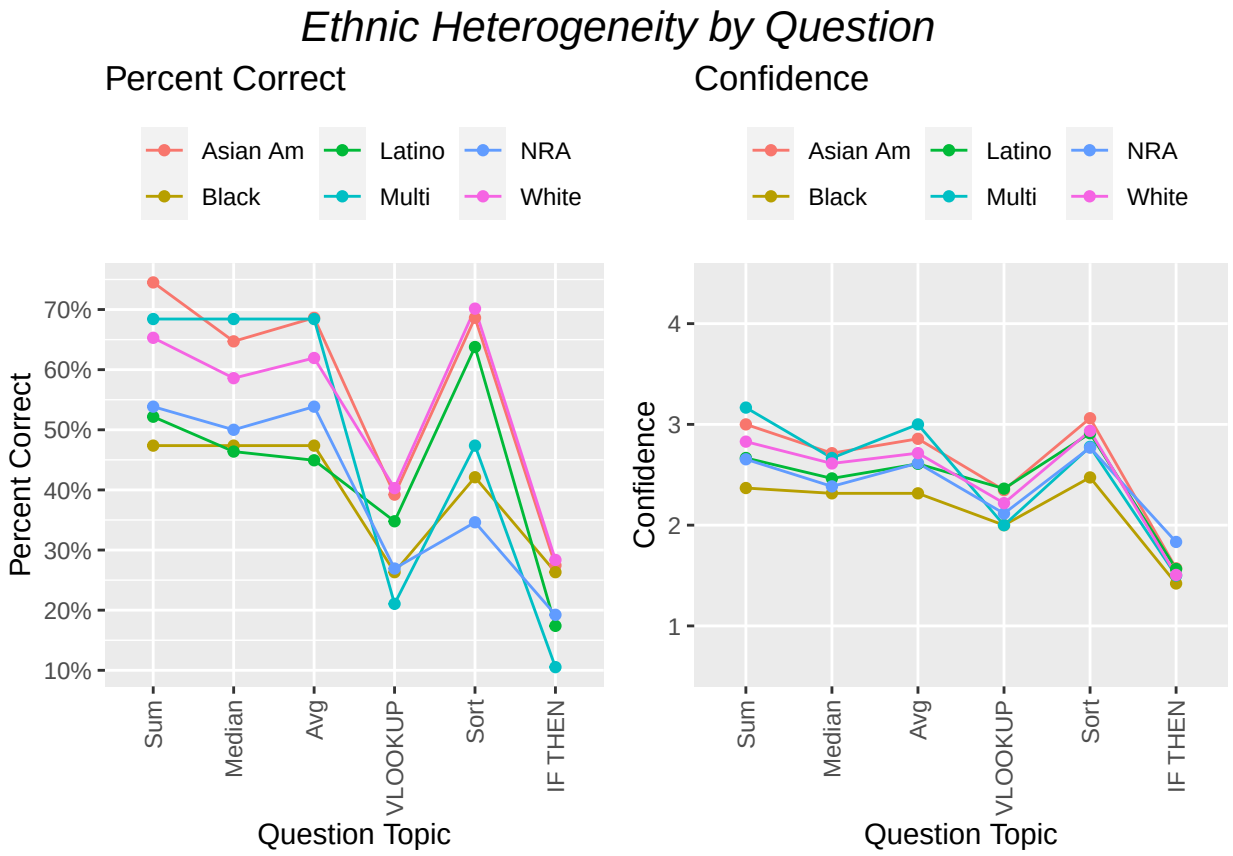
Table 1: Descriptive Statistics of Variables

Statistic	N	Mean	St. Dev.	Min	Max
Percent Correct	452	0.511	0.341	0.000	1.000
Percent Correct Log	452	0.385	0.242	0.000	0.693
Average Confidence	446	2.458	0.784	1.000	4.000
Cumulative GPA	452	3.406	0.573	0.690	4.000
Male	452	0.531	0.500	0	1
Female	452	0.469	0.500	0	1
Asian	452	0.113	0.317	0	1
Black	452	0.042	0.201	0	1
Hispanic	452	0.153	0.360	0	1
Multiracial	452	0.042	0.201	0	1
Non-Resident Alien	452	0.058	0.233	0	1
Age	452	0.593	0.492	0	1
age	452	19.980	0.995	19	26

```
grid.arrange(a,a1,ncol=2,top = textGrob("Gender Heterogeneity by Question",gp=gpar(fontsize=16,font=3)))
```

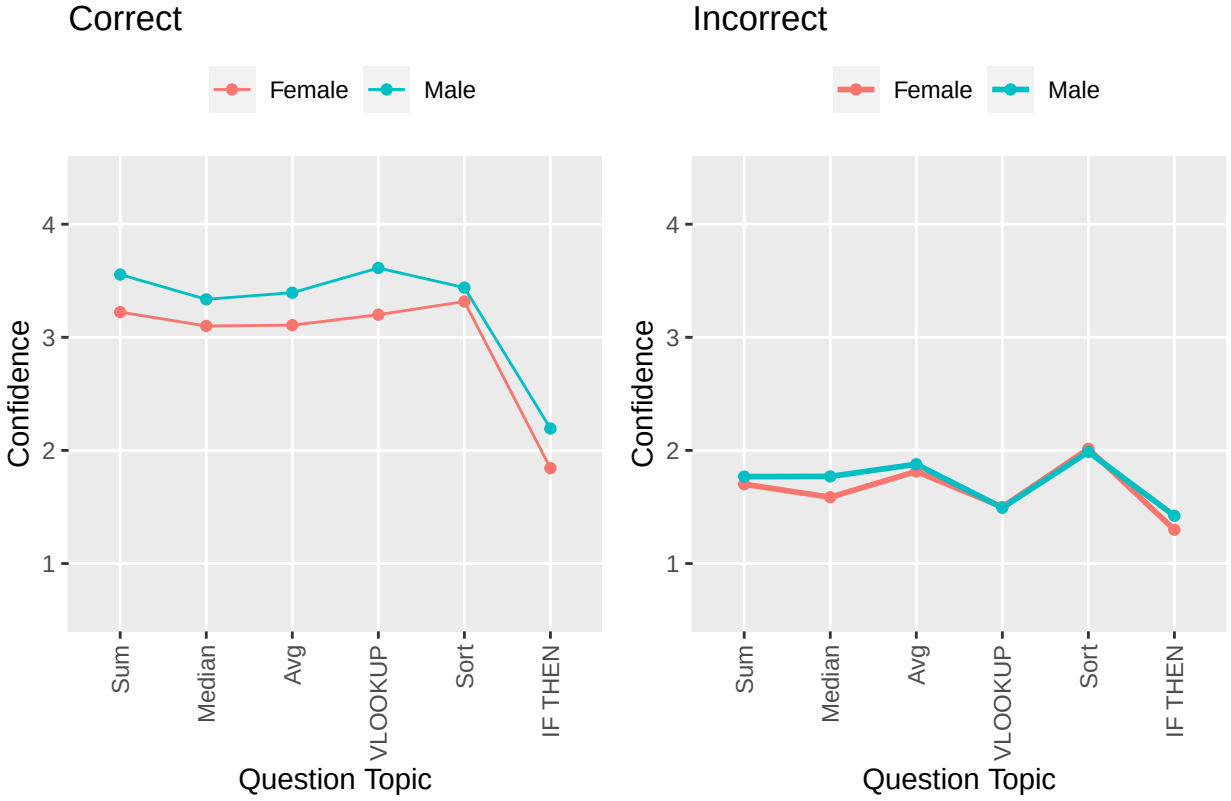



```
grid.arrange(b,b1,ncol=2,top = textGrob("Ethnic Heterogeneity by Question",gp=gpar(fontsize=16,font=3)))
```

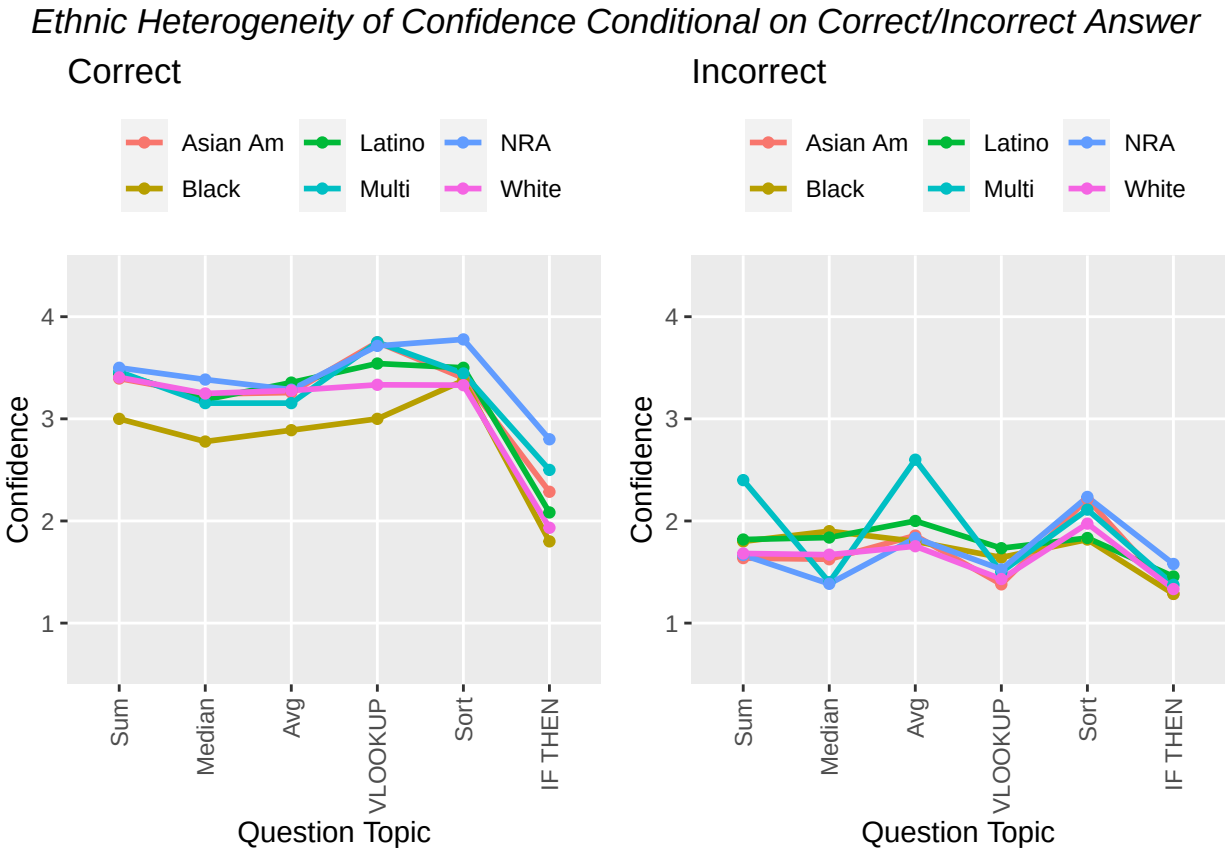



```
grid.arrange(c,g,ncol=2,top = textGrob("Gender Heterogeneity of Confidence Conditional on Correct/Incor
```

Gender Heterogeneity of Confidence Conditional on Correct/Incorrect Answer




```
grid.arrange(c1,h,ncol=2, top = textGrob("Ethnic Heterogeneity of Confidence Conditional on Correct/Inc
```




```
ols1<-lm(m$question_percentLog~m$Cumulative_GPA+as.factor(m$Gender)+as.factor(m$Ethnicity)+as.factor(m$
stargazer(ols1,type="latex", title="Regression of Correct Percent Log on Covariates",covariate.labels =
            c("Cumulative GPA", "Male", "Black", "Latino",
              "Multiracial","NRA", "White", "Confidence 2","Confidence 3","Confidence 4","Age","Intercep
```

% Table created by stargazer v.5.2.3 by Marek Hlavac, Social Policy Institute. E-mail: marek.hlavac at gmail.com % Date and time: Wed, Feb 22, 2023 - 17:32:13

Checking assumptions of the first OLS model

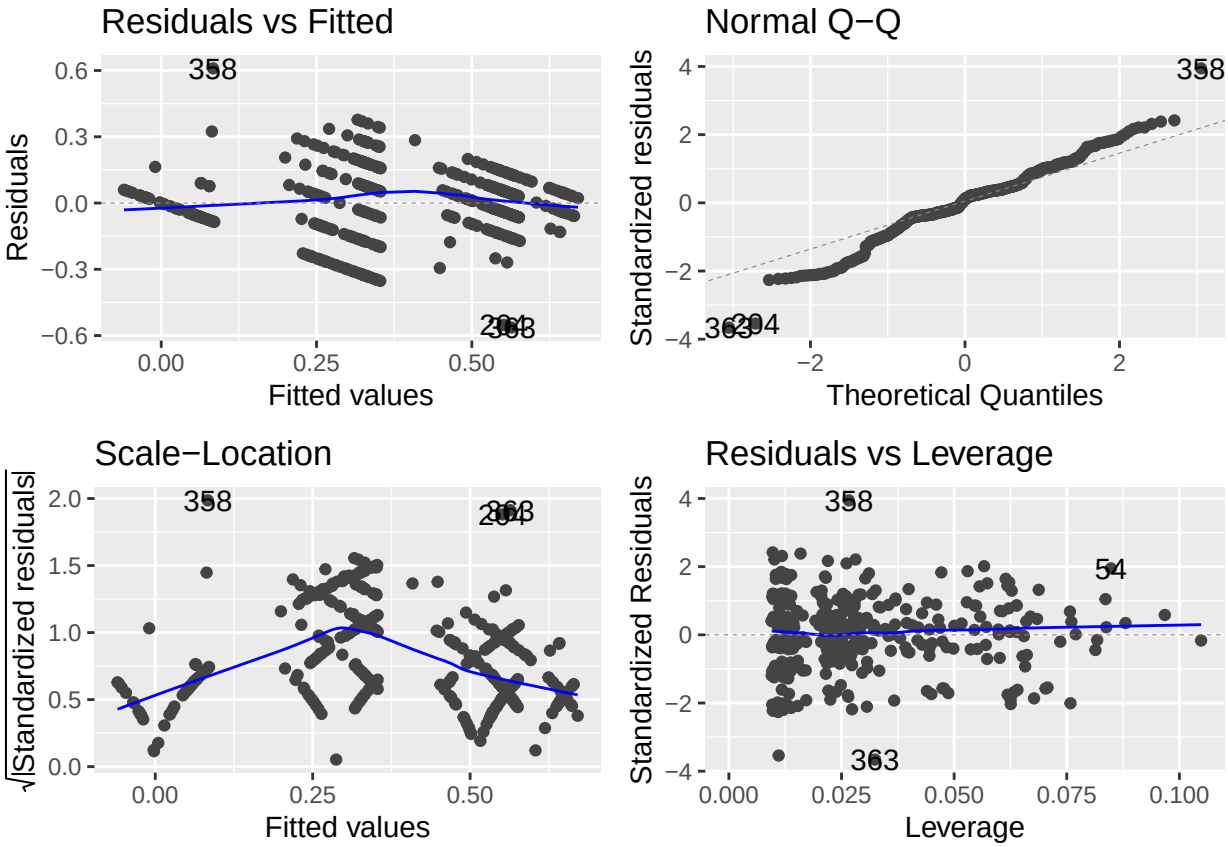
```
vif(ols1)
```

```
##                              GVIF Df GVIF^(1/(2*Df))
## m$Cumulative_GPA         1.104237  1        1.050827
## as.factor(m$Gender)      1.038643  1        1.019138
## as.factor(m$Ethnicity)   1.197449  5        1.018183
## as.factor(m$Confidence) 1.086494  3        1.013922
## m$age                    1.036359  1        1.018017
```

```
autoplot(ols1)
```

Table 2: Regression of Correct Percent Log on Covariates

	Dependent variable:
	Percent Correct Log
Cumulative GPA	0.026^{*}
	(0.014)
Male	-0.020
	(0.015)
Black	-0.038
	(0.044)
Latino	-0.075^{**}
	(0.030)
Multiracial	-0.060
	(0.043)
NRA	-0.081^{**}
	(0.039)
White	-0.003
	(0.024)
Confidence 2	0.268^{***}
	(0.024)
Confidence 3	0.489^{***}
	(0.025)
Confidence 4	0.585^{***}
	(0.030)
Age	-0.001
	(0.008)
Intercept	-0.001
	(0.163)
Observations	446
R^2	0.584
Adjusted R^2	0.573
Residual Std. Error	0.157 (df = 434)
F Statistic	55.372^{***} (df = 11; 434)
Note:	*p<0.1; **p<0.05; ***p<0.01



We can see from the VIF that none of the predictors are correlated with each other. This is a good sign as we can interpret the coefficients and standard errors with external validity. We also check the four major assumptions of linear regression. All assumptions and VIF metrics check out. This model is usable.

1) Residual vs. Fitted plot= Used to check the linear relationship assumptions. We notice that there is no pattern amongst residuals given the fitted value. Assumption 1 satisfied.

2) Normal Q-Q= Used to examine whether the residuals are normally distributed. The residuals mostly follow the straight line. A few curls toward the lower and upper bounds, but generally looks good. Assumption 2 satisifed.

3) Scale-Location= Used to check the homogeneity of variance of the residuals (homoscedasticity). Lower fitted values and higher fitted values tend to have slightly smaller residuals comparative to the median values. Although we see this increase, then decrease in residual value, there is not a distinctive trend or pattern that allows us to reject Assumption 3. Assumption 3 satisfied.

4) Residuals vs Leverage= Used to identify influential cases, that is extreme values that might influence the regression results when included or excluded from the analysis. The trend line is relatively straight. Residuals are consistent even for high leverage observations. Assumption 4 satisfied.

```
# Fit a multinomial logistic regression model
model <- multinom(as.factor(Confidence) ~ age+Ethnicity+Gender+question_percentLog+Cumulative_GPA, data
```

```
## # weights:  44 (30 variable)
## initial  value 618.287285
## iter  10 value 445.457502
## iter  20 value 367.617251
```

```
## iter  30 value 362.445753
## iter  40 value 362.407549
## iter  50 value 362.400646
## iter  60 value 362.400448
## iter  60 value 362.400447
## iter  60 value 362.400447
## final  value 362.400447
## converged
```

```
summary(model)
```

```
## Call:
## multinom(formula = as.factor(Confidence) ~ age + Ethnicity +
##     Gender + question_percentLog + Cumulative_GPA, data = m)
##
## Coefficients:
##   (Intercept)        age EthnicityBlack EthnicityLatino EthnicityMulti
## 2    4.982288 -0.3017122      1.5758732       0.3645258       1.734502
## 3    2.203743 -0.3148605      0.1035163       0.6731923       1.710310
## 4   -4.873772 -0.3132804    -12.1718352       1.6748669       2.264434
##   EthnicityNRA EthnicityWhite GenderMale question_percentLog Cumulative_GPA
## 2    0.6437231     -0.2112685  0.1453327            10.24345    0.205967623
## 3   -0.4382388     -0.6315482  0.5247932            18.66135   -0.008553917
## 4    2.7888854     -0.3747899  1.1053869            28.94022   -0.245696921
##
## Std. Errors:
##   (Intercept)       age EthnicityBlack EthnicityLatino EthnicityMulti
## 2    4.353825 0.2136767   1.089233e+00       0.8049965       1.315454
## 3    5.062392 0.2426087   1.335405e+00       0.9510173       1.494105
## 4    6.100219 0.2875981   5.390931e-05       1.1173008       1.734902
##   EthnicityNRA EthnicityWhite GenderMale question_percentLog Cumulative_GPA
## 2    0.9197423      0.7252458  0.3812251            1.732445      0.3272135
## 3    1.2981840      0.8233355  0.4601778            2.050354      0.4070094
## 4    1.3954270      0.9612015  0.5628050            2.973428      0.5038218
##
## Residual Deviance: 724.8009
## AIC: 784.8009
```