

Equation 1: $y = B_0 + B_1X_1 + B_2X_2$

Equation 2: $y = B_0 + B_1X_1 + B_2X_2 + B_3X_1X_2$

Hypothesis 1: Democratic countries will have higher levels of gender equality.

Hypothesis 2: Though democracies may provide greater gender equality, democratic institutions still do little to ensure gender equality. Even strong democracies, those with the highest value (10) on Democ11, will still have a great deal of inequality, greater than .5 on the Gender_Unequal scale.

Word_Test_Pres (available on the course website) contains two variables for testing this relationship: Gender_Unequal and Democ. Gender_Unequal is a scale of how much gender inequality a country has. And ranges from 0 to 1. A 0 on the scale is perfect gender equality. A 1 on the scale is perfect gender inequality. Democ11 is a variable ranging from 0-10 measuring how democratic a country is. A 10 on the scale is a democracy and 0 is an authoritarian regime. Using these two variables, the hypothesis above.

1) Given variables and hypothesis above, what kind of regression should you use?

a) Linear

b) Binary Logistic Regression

2) Explain your answer to the previous question

The dependent variable, Gender_Unequal is an interval variable that ranges from 0 to 1. Since it is an interval variable, a linear regression is most appropriate.

	B	t-statistic	p-value
Constant	.645		
Democ11	-.002	-5.352	.000
Adjusted R ²	.171		

3) Is there support for Hypothesis 1?

a) Yes

b) No

4) Explain your answer to the previous question, referencing specific statistics from the SPSS regression output.

The coefficient on Democ11 is negative, indication that as values on Democ11 increase (become more democratic) values on Gender_Unequal decrease (become more equal). The p-value associated with Democ11 is .000, which is smaller than .05, and the t-value is greater than 2 (-5.352), so this is a significant relationship.

5) Is there support for Hypothesis 2?

- a) Yes
- b) No

6) Explain your answer to question 5 (you must show your work for this problem).

Using Equation 1 and substituting in Beta coefficients we start with

$$y = B_0 + B_1X_1 + B_2X_2 \text{ (drop the } B_2X_2 \text{ since there is only one IV)}$$
$$\text{Gender_Unequal} = .645 + (-.002 * X_1)$$

The hypothesis tells us that the most democratic country, a 10 on the Democ11 scale will still have lots of gender inequality. So we need to substitute in a 10 for X_1 .

$$\text{Gender_Unequal} = .645 + (-.002 * 10)$$
$$\text{Gender_Unequal} = .645 - .02$$
$$\text{Gender_Unequal} = .625$$

The most democratic country is estimated at .625 on Gender_Unequal. This is substantially above the .5 as predicted in the hypothesis and the p-value (.000) is below .05 and the t-statistic is greater than 2 (-5.352).

Hypothesis 3: Controlling for regime type (democratic or authoritarian), countries that spend more money on education will have greater gender equality.

In addition to Gender_Unequal and Democ, World_Full also contains SpendEduc, a variable that measure spending on education as a percentage of GDP. Higher values denote greater public spending on education relative to GDP.

Controlling for regime type,

	B	t-statistic	p-value
Constant	.724		
Democ11	-.001	-4.892	.000

SpendEduc	-.021	-2.763	.007
Adjusted R ²	.200		

7) Is there support for the hypothesis?

- a) Yes
- b) No

8) How do you know there is/there is not support for the hypothesis? Explain your answer references specific statistics from the SPSS regression output.

The negative coefficient on SpendEduc means indicates that as public spending on education, relative to GDP, increases, Gender_Unequal will decrease (become more equal). The p-value for SpendEduc is less than .05 (.007) and the t-statistic is greater than 2 (-2.763).

9) Imagine that the United States (a 10 on the Democ11 scale) increased it spending on education from 5% to 10%. Explain how would this affect gender equality, references specific statistics from the SPSS regression output.

We have to solve the equation for SpendEduc at 5% and SpendEduc at 10% and look at the difference. Using Equation 1 and substituting in Beta coefficients we start with

$$y = B_0 + B_1X_1 + B_2X_2$$

$$\text{Gender_Unequal}(5\%) = .724 + (-.001 \cdot 10) + (-.021 \cdot 5)$$

$$\text{Gender_Unequal}(5\%) = .724 - .01 - .105$$

$$\text{Gender_Unequal}(5\%) = .609$$

$$\text{Gender_Unequal}(10\%) = .724 + (-.001 \cdot 10) + (-.021 \cdot 10)$$

$$\text{Gender_Unequal}(10\%) = .724 - .01 - .21$$

$$\text{Gender_Unequal}(10\%) = .504$$

When education spending is 5% of GDP, then the U.S. is at .609 on in inequality scale. This number falls to .504 if the U.S. would spend 10% of its GDP on education, a decrease in gender inequality of .105.

10) The adjusted R² of .200 means that of all the variation in gender inequality in countries across the world, 20.0% is explained by the type of regime (Democ11).

Hypothesis 4: Democracies will have higher levels of gender equality regardless of the amount of money spend on education, but authoritarian regimes that spend more on education will have greater gender equality than authoritarian regimes that do not spend money on education.

Note: This is proposing an interactive relationship. Notice that the dependent variable regime type has a caveat to it—a conditional statement.

You will need to begin by creating the interaction term in SPSS. Then run the regression.

not missing(Democ11) & not missing (SpendEduc) not missing(Democ11) & not missing (SpendEduc)

	B	t-statistic	p-value
Constant	.655		
Democ11	.000	-.407	.684
SpendEduc	.006	.586	.559
Democ11*SpendEduc	-.004	-3.825	.000
Adjusted R ²	.285		

11) Is there support for Hypothesis 3?

- a) Yes
- b) No

12) Explain your reasoning, making reference to specific statistics from the SPSS output.

Starting with Equation 2: $y = B_0 + B_1X_1 + B_2X_2 + B_3X_1X_2$ we can substitute 0 in for X_1 and X_2 to find our reference category.

$Gender_Unequal = B_0 + B_1*(Democ11=0) + B_2*(SpendEduc=0) + B_3*(Democ11=0)*(SpendEduc=0)$

If we multiply though, we are left with just B_0 .

When $Democ11=0$ it is authoritarian. When $SpendEduc = 0$, there is no spending on Education.

This means that B_0 must refer to authoritarian regimes that spend no money on education. This is our reference group.

The hypothesis asks about authoritarian regimes that spend money on education compared to authoritarian regimes that do not spend money on education. Keep in mind that our reference group is authoritarian regime (1% of GDP is spent on education, which is more/increased education spending compared to no education spending)

$Gender_Unequal = B_0 + B_1 * (Democ11=0) + B_2 * (SpendEduc=1) + B_3 * (Democ11=0) * (SpendEduc=1)$

If we multiply though, we are left with $B_0 + B_2$. If the reference group is B_0 then the value of B_2 must be the variable that shows how increased spending on education for authoritarian regimes affects gender inequality ($Gender_Unequal$). Here, the coefficient of .006 says for authoritarian regimes, as spending increases values of $Gender_Unequal$ also increase (become more unequal). This is contrary to the expectation in the hypothesis. However, it must also be noted that the p-value is greater than .05 and the t-statistic is less than 2. This means that in authoritarian regimes, the amount of money spent on education does not affect the amount of gender inequality in authoritarian regimes.

We could also multiply though for democracies, but we already have enough information to reject the hypothesis.

Trump proposes that it's not income that explains who will vote, but rather education. In hypothetical dataset, there is a variable $vote2012$ that measures whether someone voted for a presidential candidate in 2012. It is coded 1 if they voted and 0 if they did not vote. In addition there is a variable for income, which is on a scale from 1 to 5. Five on the scale is high income and 1 on the scale is low income. There is also a variable for education. It is measured on a scale from 1 to 4 with 1 being low education and 4 being high education.

13) Given variables and hypothesis above, what kind of regression should you use?

a) Linear

b) Binary Logistic Regression

14) Explain your reasoning to question 13.

Trump is accessing if income or education affects whether a person votes, making voting the DV. Since $vote2012$ a dichotomous variable (0 or 1), we have to use a binary logistic regression.

Here is some hypothetical SPSS output testing Trump's hypothesis.

	Coefficient	p-value	Exp ^(B)
Income	.26	.051	1.308
Education	.68	.002	2.019
Constant	.19		

15) Is there support for Trump's Hypothesis?

a) Yes

b) No

16) Explain your answer to question 15 making specific reference to the statistics from the SPSS output.

Firs it is important to not that the p-value for income is .051, greater than .05, meaning that income does not affect whether a person voted or now. The p-value for education is .002, which is less than .05 and indicates that education does affect whether a person voted. So far the hypothesis is supported but we must also consider the direction of the relationship. The coefficient is positive indicating a positive relationship between education and voting (Alternatively, you could say that the odds ratio of 2.019 is greater than 1, which also indicates a positive relationship). Given that only education is significant and has a positive relationship with voting, Trumps hypothesis is supported.

17) A person who is at 2 on the education scale is 2.019 times more likely than someone at 1 on the education scale to have voted.

18) The odds ratio of 2.019 means that there is a 101.9 percent increase in the odds that a person votes for each unit increase in education.