Soc 103M

Causation

Meanings of Causation

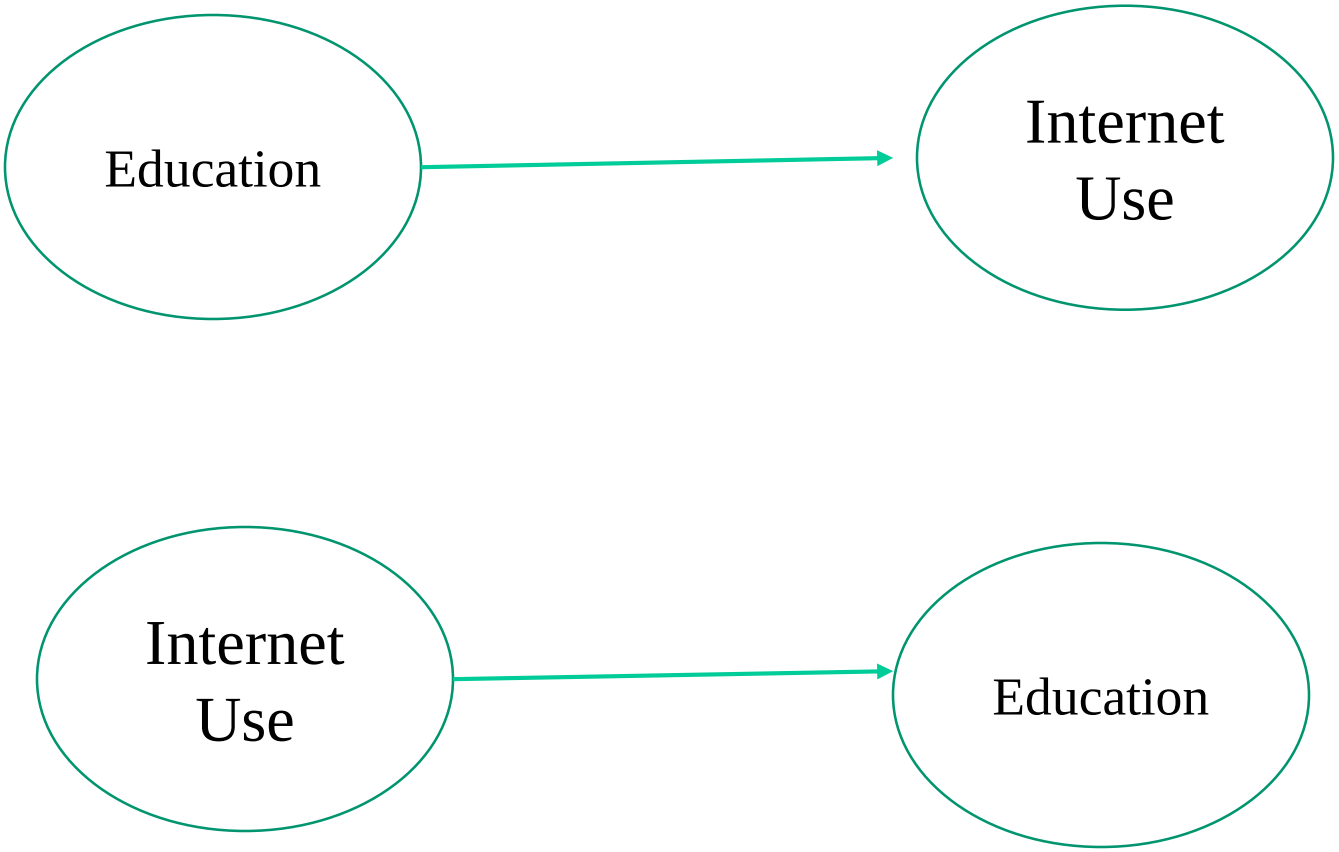
- ***Idiographic Causes***
-
- ***Nomothetic Causes***
-
  - **Counterfactual**
-
- ***Synthetic Causal Explanations***

Criteria for Causal Explanation

- **John Stuart Mill's 3 Main Criteria of Causation**
- ***Empirical Association***
- ***Appropriate Time Order***
- ***Non-Spuriousness (Excluding other Forms of Causation)***

- **2 Additional Criteria of Causal Explanation**
- ***Causal Mechanisms***
- ***Specifying Context***

Visualization

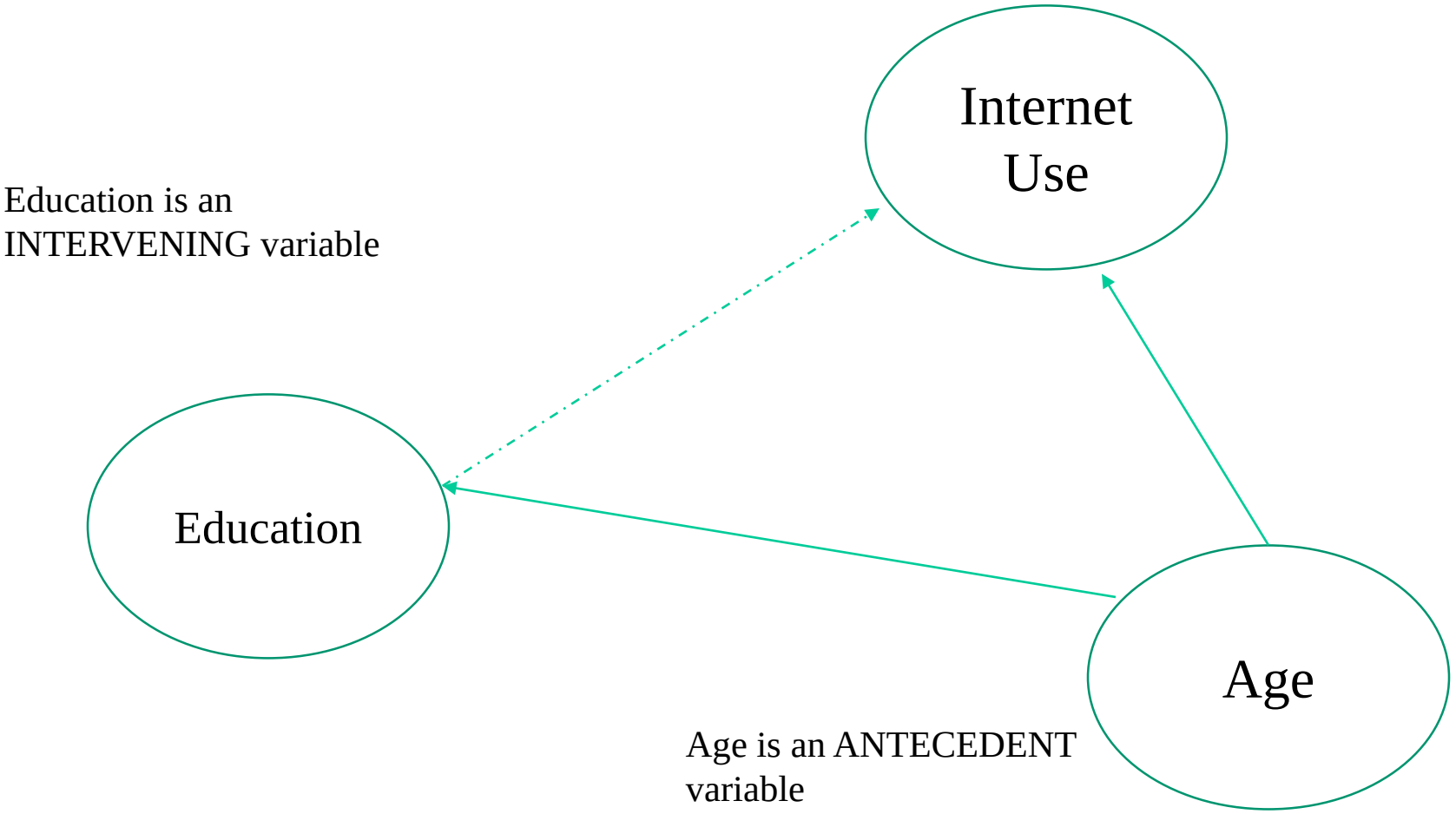


Question

- Suppose you find that better educated people spend more time on the internet than less educated people. You conclude that education increases internet use.
- Someone comes and points out that younger people are both more educated and they are more open to new technology like the internet. Education is not the cause but age is.
- Which one of the causal criteria is not fulfilled according to your critic?

- A. Empirical association
- B. Appropriate time order
- C. Non-spuriousness

Spuriousness is created by a third, antecedent variable
Antecedent means: it is earlier in time than both the supposed cause and effect



No bivariate relationship between Education and internet use.
Also very litte relationship between Age and Education (not significant at the .05 level)

Correlations

		Hours of internet use on weekdays	Highest year of school completed	Age of respondent
Hours of internet use on weekdays	Pearson Correlation	1	.025	-.158**
	Sig. (2-tailed)		.377	.000
	N	1220	1218	1216
Highest year of school completed	Pearson Correlation	.025	1	-.032
	Sig. (2-tailed)	.377		.090
	N	1218	2858	2848
Age of respondent	Pearson Correlation	-.158**	-.032	1
	Sig. (2-tailed)	.000	.090	
	N	1216	2848	2857

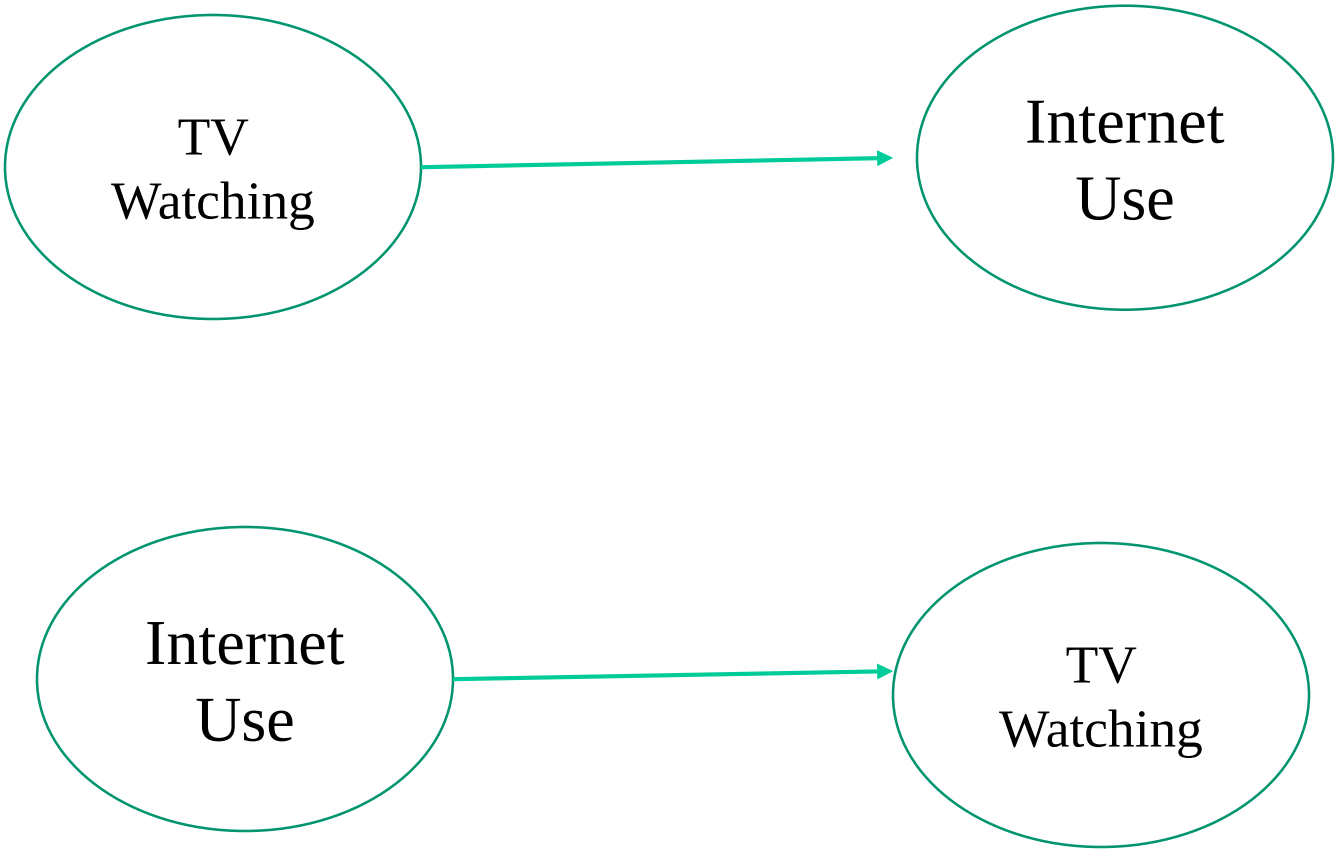
**. Correlation is significant at the 0.01 level (2-tailed).

Question

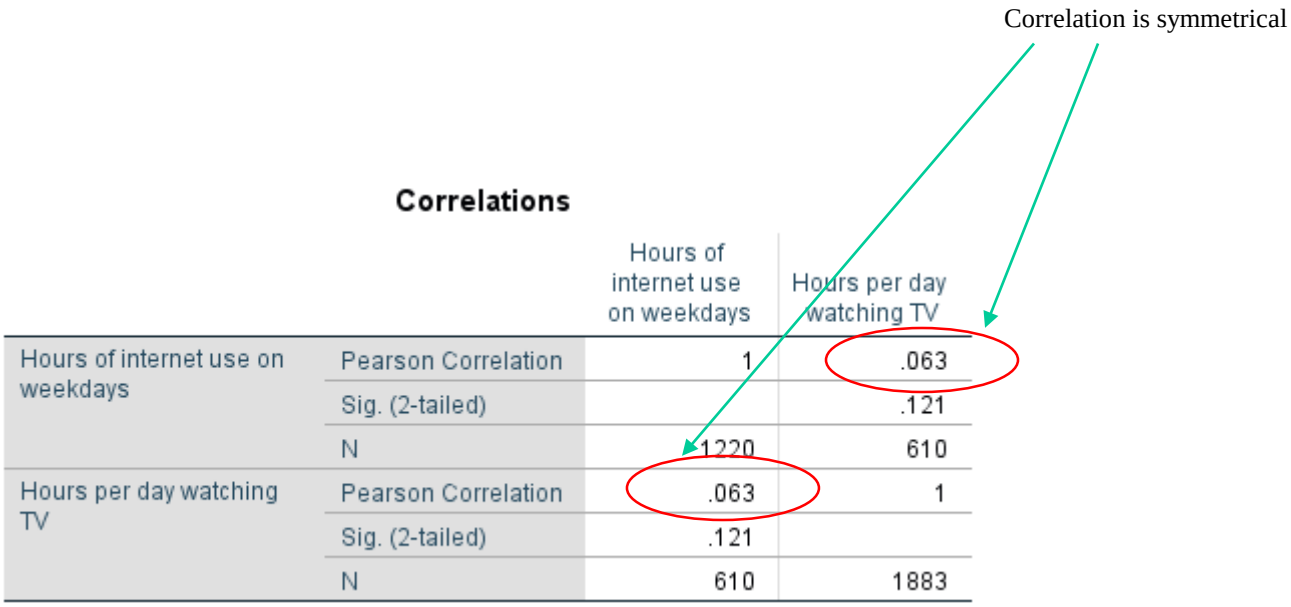
- Suppose you find that people who watch less TV spend more time on the internet. You conclude that TV watching decreases internet use.
- Someone comes and points out that maybe it is internet use that decreases TV watching.
- Which one of the causal criteria is not fulfilled according to your critic?

- A. Empirical association
- B. Appropriate time order
- C. Non-spuriousness

Visualization
Direction of causation



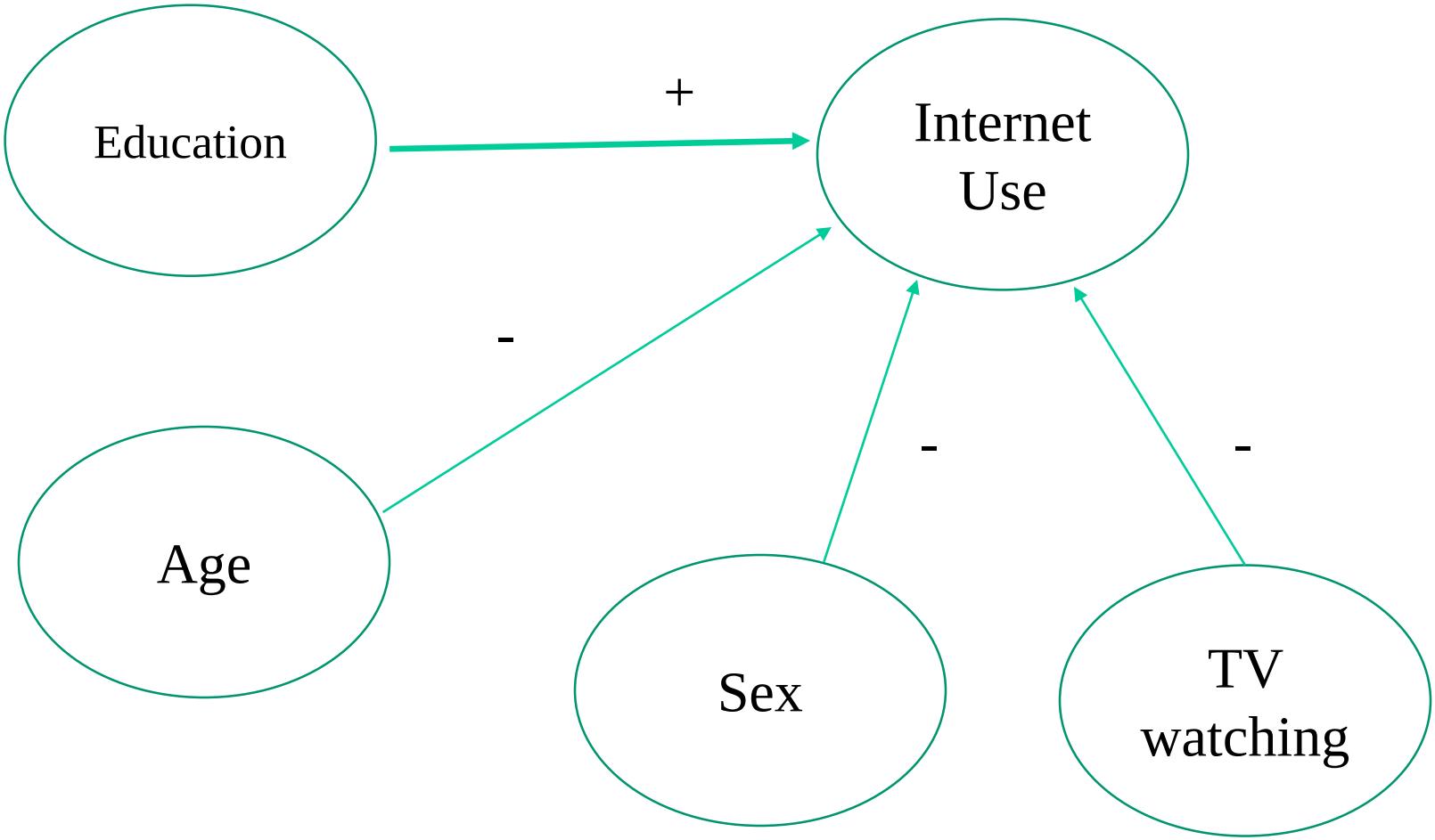
Bivariate Correlation



Correlations

		Hours of internet use on weekdays	Hours per day watching TV
Hours of internet use on weekdays	Pearson Correlation	1	.063
	Sig. (2-tailed)		.121
	N	1220	610
Hours per day watching TV	Pearson Correlation	.063	1
	Sig. (2-tailed)	.121	
	N	610	1883

Multi-casual model



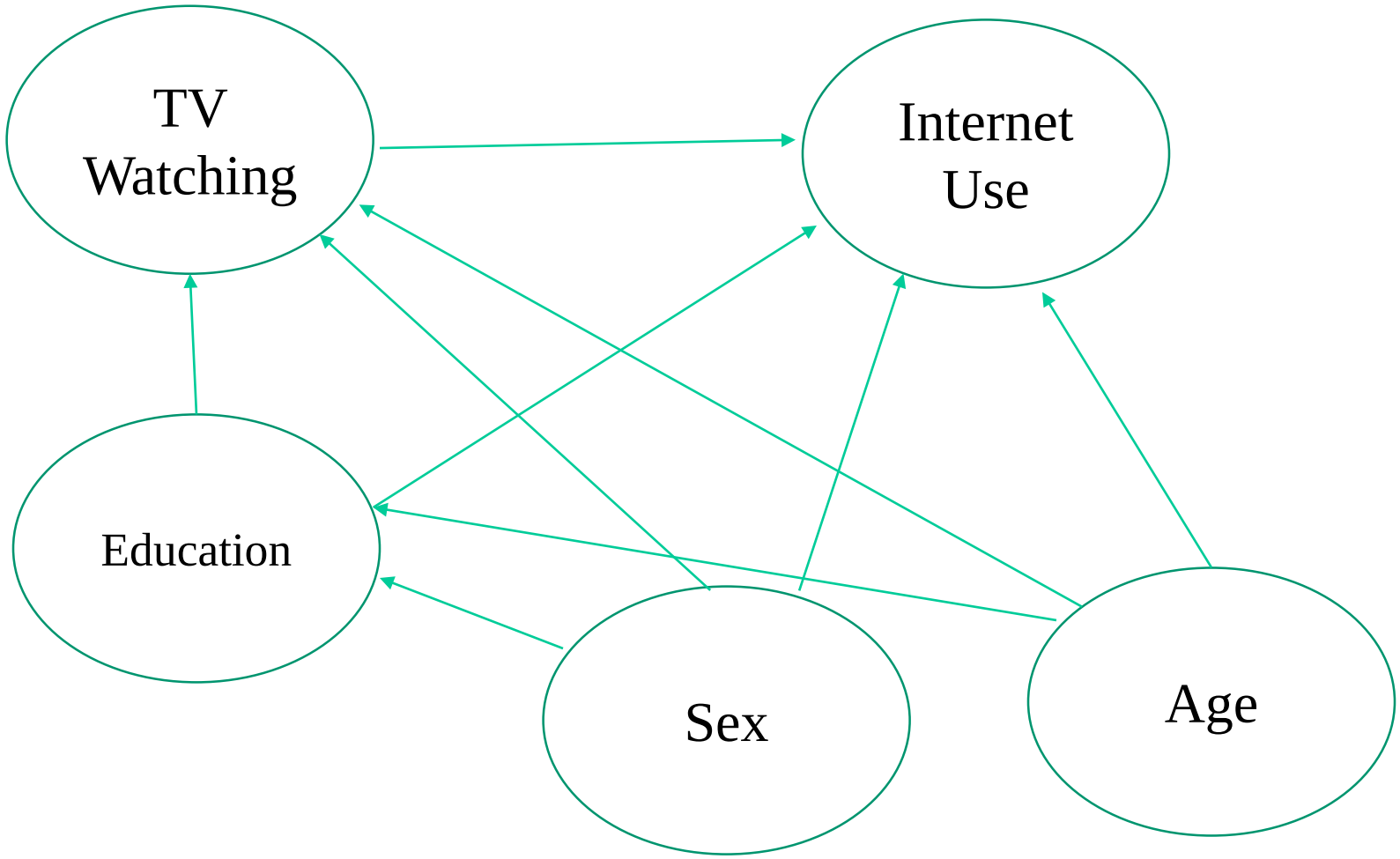
Correlations

		Hours of internet use on weekdays	Highest year of school completed	Age of respondent	Respondents sex	Hours per day watching TV
Hours of internet use on weekdays	Pearson Correlation	1	.025	-.158**	-.045	.063
	Sig. (2-tailed)		.377	.000	.115	.121
	N	1220	1218	1216	1220	610
Highest year of school completed	Pearson Correlation	.025	1	-.032	.012	-.205**
	Sig. (2-tailed)	.377		.090	.513	.000
	N	1218	2858	2848	2858	1880
Age of respondent	Pearson Correlation	-.158**	-.032	1	.042*	.259**
	Sig. (2-tailed)	.000	.090		.025	.000
	N	1216	2848	2857	2857	1878
Respondents sex	Pearson Correlation	-.045	.012	.042*	1	.035
	Sig. (2-tailed)	.115	.513	.025		.132
	N	1220	2858	2857	2867	1883
Hours per day watching TV	Pearson Correlation	.063	-.205**	.259**	.035	1
	Sig. (2-tailed)	.121	.000	.000	.132	
	N	610	1880	1878	1883	1883

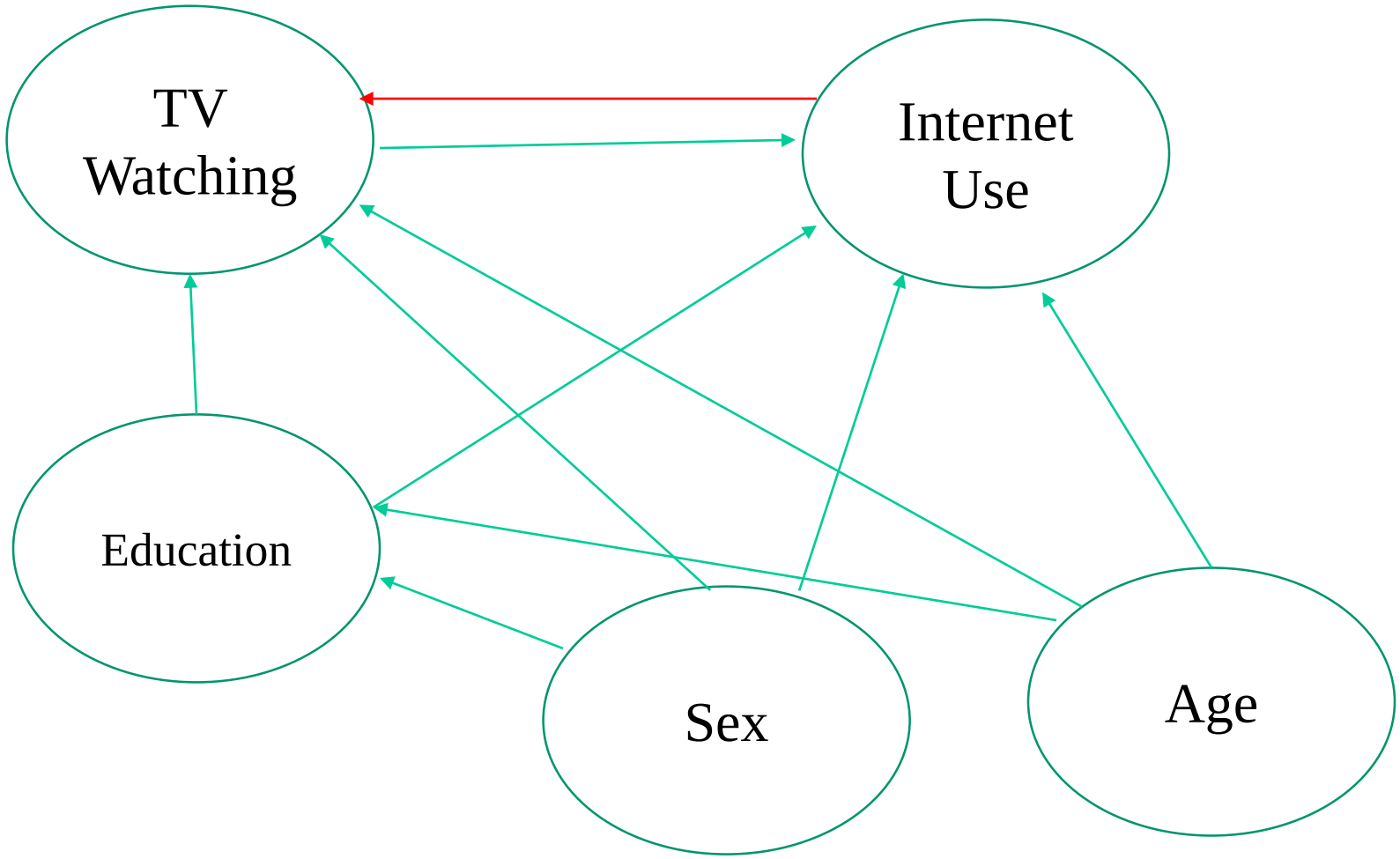
**. Correlation is significant at the 0.01 level (2-tailed).

*. Correlation is significant at the 0.05 level (2-tailed).

Multi-casual model



Multi-casual model



Simple Regression

This is the bivariate correlation from the earlier slide

But in multiple regression R is just the squareroot of R-square

Variation explained in the depedent variable by the regression model

Model Summary

Model	R	R Square	Adjusted R Square	Std. Error of the Estimate
1	.025[a]	.001	.000	3.596

a. Predictors: (Constant), Highest year of school completed

ANOVA[a]

Statistical significance of Adjusted R Square

Model		Sum of Squares	df	Mean Square	F	Sig.
1	Regression	10.111	1	10.111	.782	.377[b]
	Residual	15724.457	1216	12.931		
	Total	15734.568	1217			

a. Dependent Variable: Hours of internet use on weekdays

b. Predictors: (Constant), Highest year of school completed

Coefficients[a]

Constant or intercept. The value of the dependent variable when the independent variable(s) is (are all) ZERO

Model		Unstandardized Coefficients B	Unstandardized Coefficients Std. Error	Standardized Coefficients Beta	t	Sig.
1	(Constant)	2.670	.547		4.885	.000
	Highest year of school completed	.033	.037	.025	.884	.377

a. Dependent Variable: Hours of internet use on weekdays

Statistical significance of slope

Slope. The unit change in the dependent variable in one unit change in the independent variable. The effect of the independent variable on the dependent variable

Multiple Regression

Notice that the bivariate correlation between TV Watching and Internet use was not significant. But here the effect of internet use on TV watching IS significant.

THIS CAN HAPPEN!

It is called suppression.

Model Summary

Model	R	R Square	Adjusted R Square	Std. Error of the Estimate
1	.188[a]	.035	.029	3.846

a. Predictors: (Constant), Highest year of school completed, Respondents sex, Hours per day watching TV, Age of respondent

ANOVA[a]

Model		Sum of Squares	df	Mean Square	F	Sig.
1	Regression	327.334	4	81.833	5.532	.000[b]
	Residual	8905.375	602	14.793		
	Total	9232.708	606			

a. Dependent Variable: Hours of internet use on weekdays

b. Predictors: (Constant), Highest year of school completed, Respondents sex, Hours per day watching TV, Age of respondent

Coefficients[a]

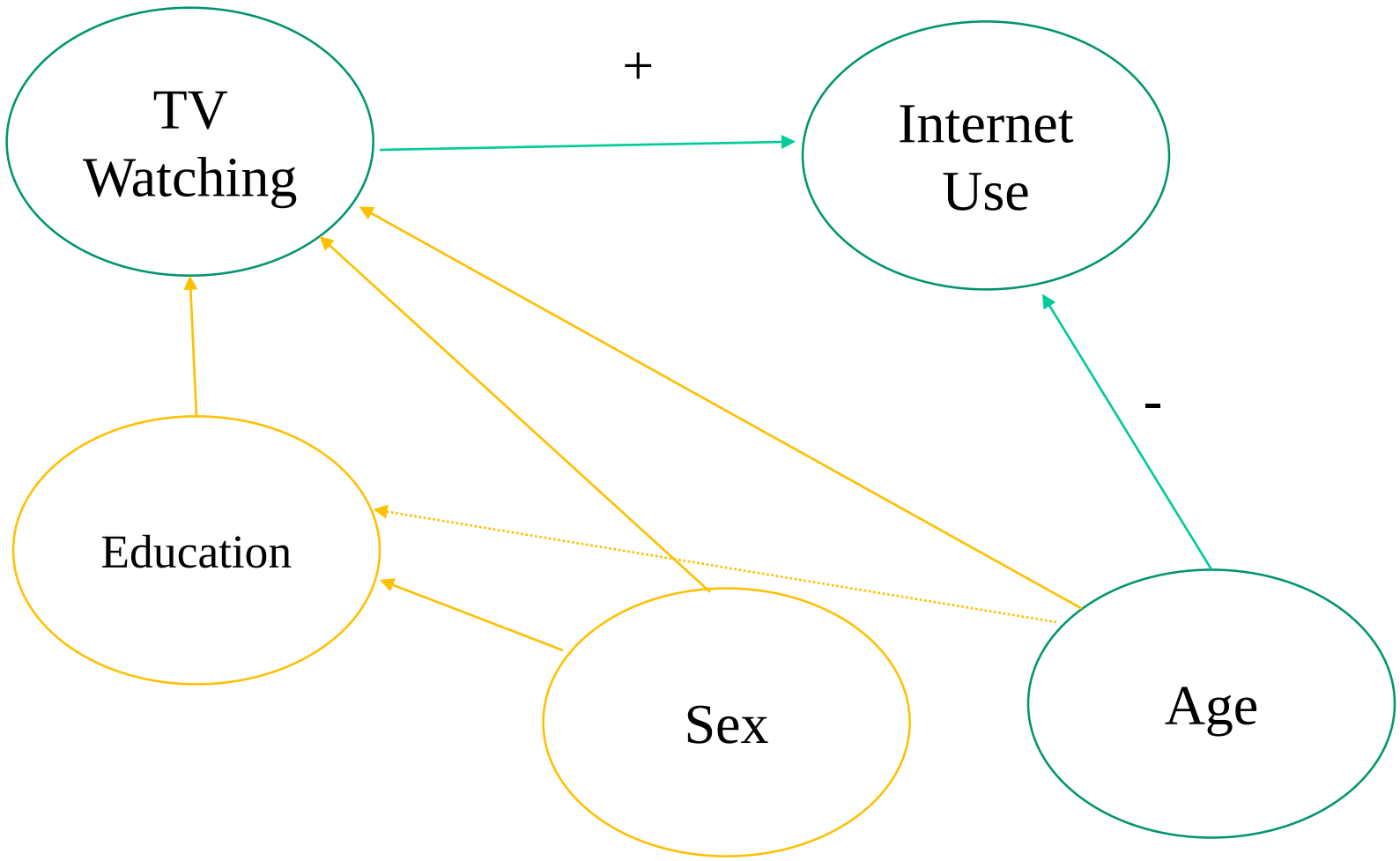
Model		Unstandardized Coefficients B	Unstandardized Coefficients Std. Error	Standardized Coefficients Beta	t	Sig.
1	(Constant)	3.941	1.061		3.715	.000
	Age of respondent	-.044	.010	-.179	-4.354	.000
	Respondents sex	-.216	.316	-.027	-.682	.495
	Hours per day watching TV	.163	.072	.093	2.265	.024
	Highest year of school completed	.062	.062	.041	1.001	.317

a. Dependent Variable: Hours of internet use on weekdays

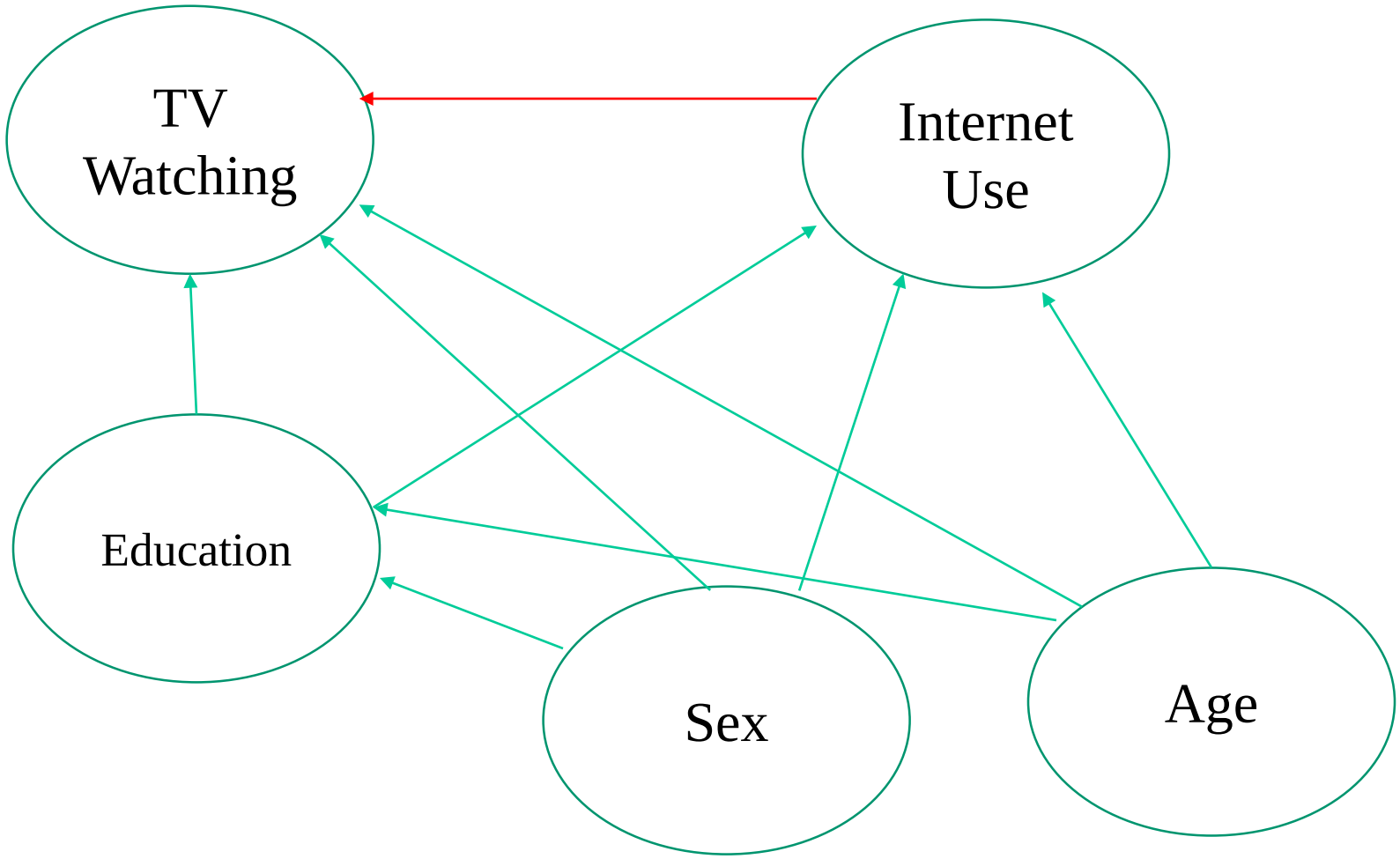
Effect of Education controlling for Age, Sex and TV watching.
We call it the NET EFFECT of Education (net of those other three independent variables).

The effect of the indepdent variable on the dependent variable CONTROLLING FOR ALL OTHER INDEPENDENT VARIABLES

Multi-casual model of internet use



Multi-casual model



Model Summary

Model	R	R Square	Adjusted R Square	Std. Error of the Estimate
1	.239[a]	.057	.051	2.174

a. Predictors: (Constant), Highest year of school completed, Hours of internet use on weekdays, Respondents sex, Age of respondent

ANOVA[a]

Model		Sum of Squares	df	Mean Square	F	Sig.
1	Regression	171.677	4	42.919	9.078	.000[b]
	Residual	2846.112	602	4.728		
	Total	3017.789	606			

a. Dependent Variable: Hours per day watching TV

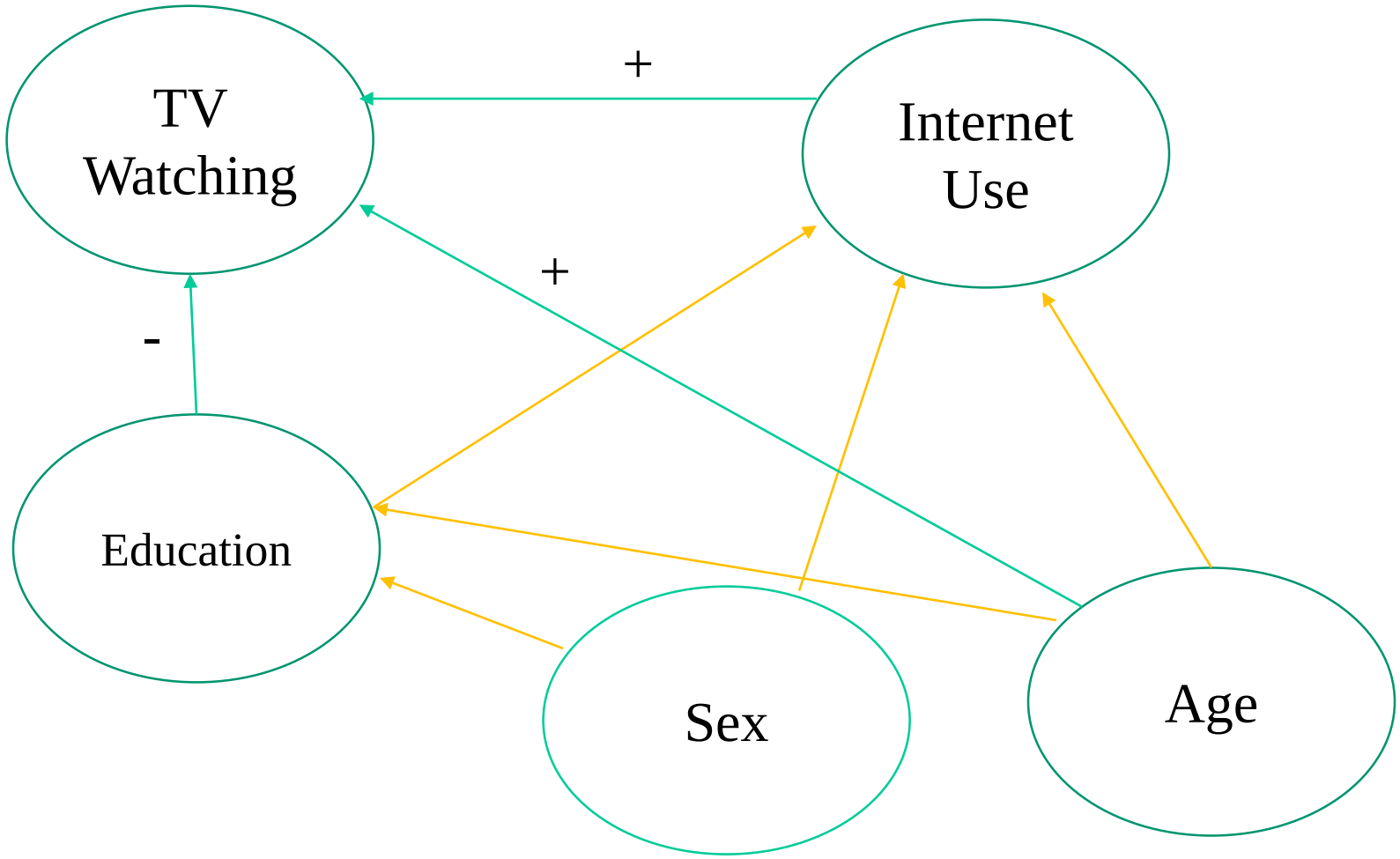
b. Predictors: (Constant), Highest year of school completed, Hours of internet use on weekdays, Respondents sex, Age of respondent

Coefficients[a]

Model		Unstandardized Coefficients		Standardized Coefficients		
		B	Std. Error	Beta	t	Sig.
1	(Constant)	3.470	.590		5.884	.000
	Hours of internet use on weekdays	.052	.023	.091	2.265	.024
	Age of respondent	.025	.006	.180	4.411	.000
	Respondents sex	.037	.179	.008	.208	.836
	Highest year of school completed	-.154	.035	-.179	-4.467	.000

a. Dependent Variable: Hours per day watching TV

Multi-casual model of TV watching



Summary

- Simple Regression
  - Look for Adjusted R-squared
    - this tells you how good your model is overall
  - Look for the significance of the F-statistics
    - this tells you if R-squared is statistically significantly different from 0
  - Look at the intercept
    - This tells you the value of the DV when the IV =0
  - Look at the slope
    - This tells you how many units the DV changes by one unit change in the IV
  - Look at the significance of the slope (t-statistics)
    - This tells you if the IV has an effect that is statistically significant(ly different from 0)

Summary

- Multiple Regression
  - Look for Adjusted R-squared
    - this tells you how good your model is overall
  - Look for the significance of the F-statistics
    - this tells you if R-squared is statistically significantly different from 0
  - Look at the intercept
    - This tells you the value of the DV when ALL the IVs =0
  - Look at each slope
    - This tells you how many units the DV changes by one unit change in that IV controlling for/net of all other IVs in the model
      - What does this mean? Suppose you have a statistically significant effect of an IV. When you explain WHY you cannot utter the name of the other variables. Those cannot be invoked.
  - Look at the significance of each slope (t-statistics)
    - This tells you if the IV has an effect that is statistically significant(ly different from 0) controlling for/net of the other IVs in the model
  - Simplify your model
    - Through out IVs that are not significant UNLESS they are INTERVENING between a significant ANTECEDENT IV and the DV.