Effect of Education on Savings

ECON 2509: Econometrics w/ R

Harshada Pujari, Khushboo Surana, Mabisa Chhetry

Data Description

- Data is from 1980’s, and it originally comes from an introductory econometrics' textbook:

  Textbook

- R Documentation:

  saving {wooldridge}

  sav: annual savings, $

  inc: annual income, $

  size: family size

  educ: years educ, household head

  age: age of household head

  black: =1 if household head is black

  cons: annual consumption, $

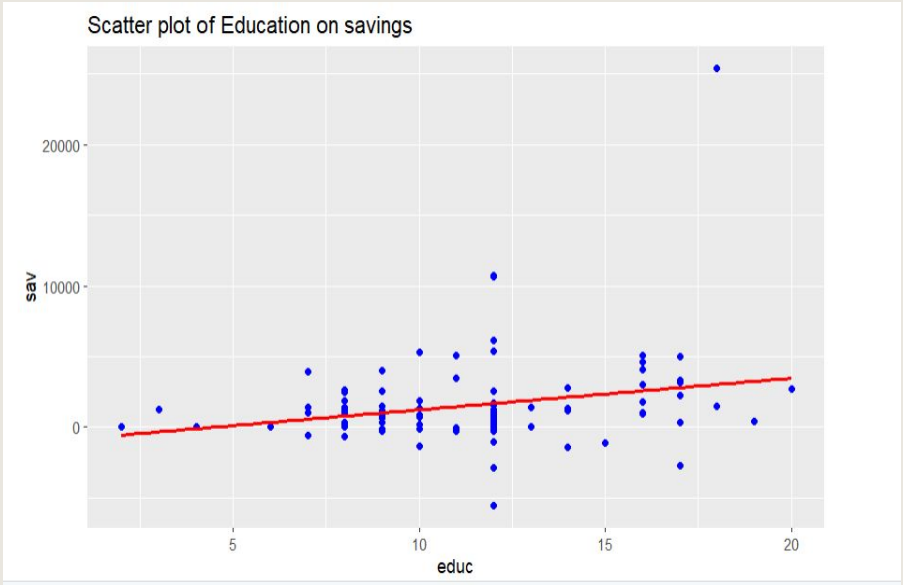
Data Description

Key points:

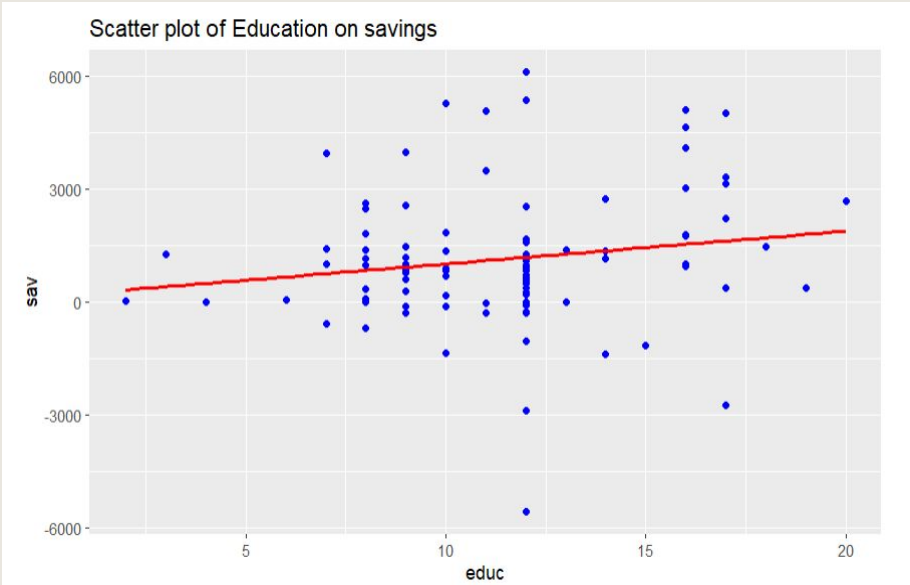
- The variables (income, age, and consumer spending) all have higher mean than the median indicating a right-skewed distribution
- There is a large variation of in how much people are willing to spend
- Relatively homogeneous demographic composition in terms of race

Table for descriptive statistics.

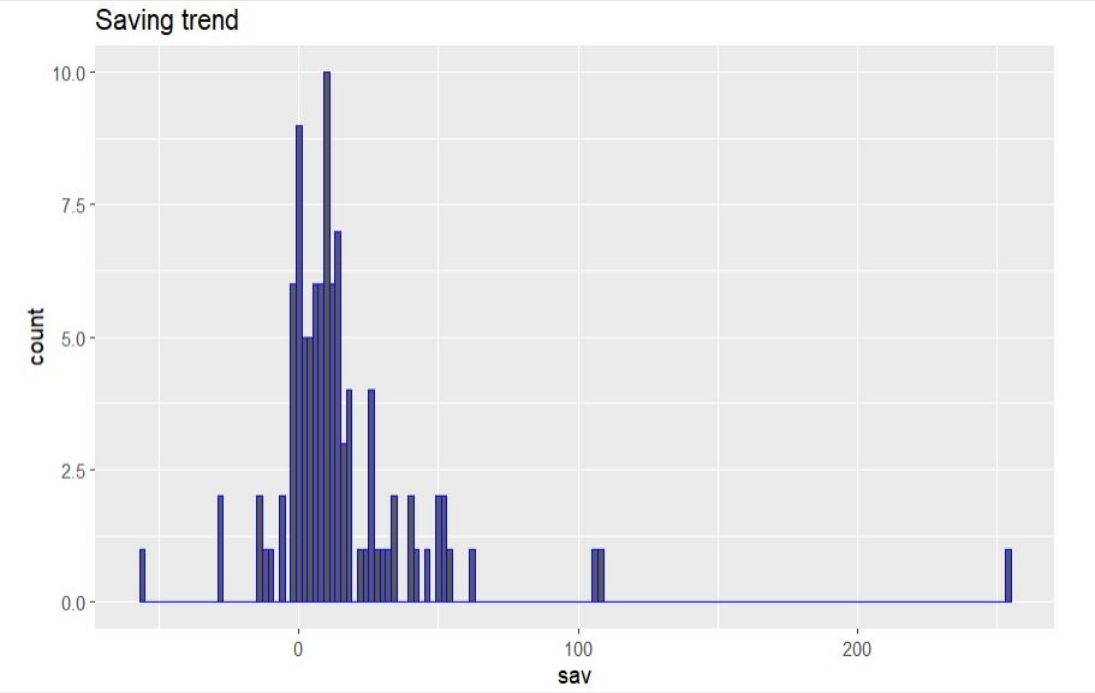
Statistic	N	Mean	Median	St. Dev.
inc	100	9,941.24	8,776.5	5,584.00
size	100	4.35	4	1.49
educ	100	11.58	12	3.44
age	100	38.77	38.5	7.40
black	100	0.07	0	0.26
cons	100	8,358.73	7,561.5	5,729.53



Without outlier



- Clear linear relationship between education and savings
- Removing outlier makes the fitted line more flatter
- Data points are tightly concentrated around the fitted red line
- Specifically focusing on individuals with approximately 12 years of education, corresponding to high school completion, we find a divergence in financial outcomes.
- Diverse impact of career choices



- Majority of individuals fall within a certain range of savings
- Around $2000 in annual savings, which is not the smallest amount considering this is 1980's data
- Still on the lower side
- Symmetrical distribution of data

Multiple regression

	Dependent variable:				
	sav				
	(1)	(2)	(3)	(4)	(5)
educ	223.720	356.495	426.496	434.973	440.421
	(134.078)	(172.765)	(185.568)	(186.609)	(188.275)
cons		-0.256	-0.301	-0.295	-0.294
		(0.157)	(0.163)	(0.159)	(0.160)
age			101.586	275.183	277.871
			(45.467)	(95.509)	(95.939)
I(age * size)				-45.514	-46.063
				(17.368)	(17.369)
size				1,731.619	1,753.937
				(670.651)	(672.647)
black					302.124
					(1,037.600)
Constant	-1,008.167	-406.516	-4,780.997	-11,622.300	-11,822.960
	(1,335.289)	(899.181)	(2,383.851)	(4,488.296)	(4,521.977)
Observations	100	100	100	100	100
R2	0.055	0.235	0.279	0.300	0.300
Adjusted R2	0.045	0.219	0.257	0.263	0.255
Residual Std. Error	3,209.982	2,903.126	2,832.199	2,820.601	2,834.678
F Statistic	5.675	14.875	12.393	8.055	6.658
Note:					NA

Regressions

Regression 1:

- **sav = -1008.167 + 223.720*educ**
- The t-stat for education is 1.66>1.64
- The regression model explains just 4.5% variation in savings

Regression 2:

- **sav = -406.516+356.495*educ - 0.256*cons**
- The t-stat for education is 2.06 >1.96
- The regression model explains 21.9% variation in savings

Regression 3:

- **sav = -4,780.997+426.496*educ -0.301*cons+ 101.586*age**
- The t-stat for education is 2.3 >1.96
- The regression model explains 25.7% variation in savings

Multiple regression

	Dependent variable:				
	sav				
	(1)	(2)	(3)	(4)	(5)
educ	223.720	356.495	426.496	434.973	440.421
	(134.078)	(172.765)	(185.568)	(186.609)	(188.275)
cons		-0.256	-0.301	-0.295	-0.294
		(0.157)	(0.163)	(0.159)	(0.160)
age			101.586	275.183	277.871
			(45.467)	(95.509)	(95.939)
I(age * size)				-45.514	-46.063
				(17.368)	(17.369)
size				1,731.619	1,753.937
				(670.651)	(672.647)
black					302.124
					(1,037.600)
Constant	-1,008.167	-406.516	-4,780.997	-11,622.300	-11,822.960
	(1,335.289)	(899.181)	(2,383.851)	(4,488.296)	(4,521.977)
Observations	100	100	100	100	100
R2	0.055	0.235	0.279	0.300	0.300
Adjusted R2	0.045	0.219	0.257	0.263	0.255
Residual Std. Error	3,209.982	2,903.126	2,832.199	2,820.601	2,834.678
F Statistic	5.675	14.875	12.393	8.055	6.658
Note:					NA

Regressions

Regression 4:

- **sav = -11,622.300 + 434.973*educ - 0.295*cons+ 275.183*age - 45.514(age*size) + 1,731.619*size**
- The t-stat for education is 2.33 >1.96
- For a family size of 2, 1 year increase in the age, the savings increases by 275.183-45.514*2= $184.155 keeping education, size and consumption constant.
- For a household head of 30 years, an increase in family size by 1 increases savings 1731.619-45.514(30)= $366.199 keeping education, age and consumption constant.
- The regression model explains 26.3% variation in savings

Regression 5:

- **sav = -11822.960 + 440.421*educ - 0.294*cons+ 277.871 *age + 1753.937 *size -46.063(age*size)+ 302.124*black**
- The t-stat for education is 2.339 >1.96
- The regression model explains 25.5% variation in savings

Multiple regression

	Dependent variable:			
	sav			
	(1)	(2)	(3)	(4)
educ	175.020	179.950	184.391	-86.493
	(56.154)	(57.066)	(57.092)	(202.900)
I(educ2)				11.793
				(8.916)
cons	-0.109	-0.108	-0.113	0.048
	(0.062)	(0.063)	(0.061)	(0.090)
age	35.569	34.556	118.819	88.194
	(25.765)	(25.353)	(57.721)	(54.825)
I(age * size)			-21.544	-13.964
			(13.790)	(13.566)
I(cons * size)				-0.042
				(0.028)
size		-39.534	770.427	822.285
		(101.968)	(514.839)	(500.196)
black		331.757		
		(913.298)		
Constant	-1,305.911	-1,180.034	-4,370.692	-3,101.245
	(1,163.712)	(1,189.999)	(2,282.058)	(2,459.529)
Observations	97	97	97	97
R2	0.090	0.093	0.106	0.135
Adjusted R2	0.061	0.043	0.057	0.067
Residual Std. Error	1,759.226	1,775.303	1,762.940	1,753.488
F Statistic	3.061	1.868	2.151	1.979
Note:				NA

Regressions (without the outlier)`

Regression 1:

- **sav = -1305.911+175.020*educ -0.109*cons+ 35.569*age**
- t-Stat for education is 3.11 >2.56
- The regression model explains 6.1% variation on savings

Regression 2:

- **sav = -1180.034+179.950*educ -0.108*cons+ 34.556*age -39.534size+331.757black**
- t-Stat for education is 3.15 >2.56
- The regression model explains 4.3% variation on savings

Multiple regression

	Dependent variable:			
	sav			
	(1)	(2)	(3)	(4)
educ	175.020	179.950	184.391	-86.493
	(56.154)	(57.066)	(57.092)	(202.900)
I(educ2)				11.793
				(8.916)
cons	-0.109	-0.108	-0.113	0.048
	(0.062)	(0.063)	(0.061)	(0.090)
age	35.569	34.556	118.819	88.194
	(25.765)	(25.353)	(57.721)	(54.825)
I(age * size)			-21.544	-13.964
			(13.790)	(13.566)
I(cons * size)				-0.042
				(0.028)
size		-39.534	770.427	822.285
		(101.968)	(514.839)	(500.196)
black		331.757		
		(913.298)		
Constant	-1,305.911	-1,180.034	-4,370.692	-3,101.245
	(1,163.712)	(1,189.999)	(2,282.058)	(2,459.529)
Observations	97	97	97	97
R2	0.090	0.093	0.106	0.135
Adjusted R2	0.061	0.043	0.057	0.067
Residual Std. Error	1,759.226	1,775.303	1,762.940	1,753.488
F Statistic	3.061	1.868	2.151	1.979
Note:				NA

Regressions (without the outlier)

Regression 3: (Baseline Regression)

- We are using this regression as our baseline regression since the variable of interest (education), as well as other variables are statistically significant in this regression model and shows true picture without the outlier
- **sav = -4370.69+184.391*educ-0.113*cons+118.819*age -21.544(age*size) + 770.427size**
- For 1 year increase in years of education, the savings increases by $184.391 keeping consumption, size and age constant.
- t-Stat for education is 3.22 >2.56
- For 1 year increase in education saving increases by $184.391,holding consumption, age and size constant.
- For a family size of 2, 1 year increase in age, the savings increases by **(117.859-21.463*2)= $74.933** keeping education, age, size,consumption and black constant.
- For a household head of 30 years, an increase in family size by 1, increases the savings by **(779.661-21.463(30))= $135.771** keeping education, age, consumption and black constant.
- The regression model explains 5.7% variation on savings

Multiple regression

	Dependent variable:			
	sav			
	(1)	(2)	(3)	(4)
educ	175.020	179.950	184.391	-86.493
	(56.154)	(57.066)	(57.092)	(202.900)
I(educ2)				11.793
				(8.916)
cons	-0.109	-0.108	-0.113	0.048
	(0.062)	(0.063)	(0.061)	(0.090)
age	35.569	34.556	118.819	88.194
	(25.765)	(25.353)	(57.721)	(54.825)
I(age * size)			-21.544	-13.964
			(13.790)	(13.566)
I(cons * size)				-0.042
				(0.028)
size		-39.534	770.427	822.285
		(101.968)	(514.839)	(500.196)
black		331.757		
		(913.298)		
Constant	-1,305.911	-1,180.034	-4,370.692	-3,101.245
	(1,163.712)	(1,189.999)	(2,282.058)	(2,459.529)
Observations	97	97	97	97
R2	0.090	0.093	0.106	0.135
Adjusted R2	0.061	0.043	0.057	0.067
Residual Std. Error	1,759.226	1,775.303	1,762.940	1,753.488
F Statistic	3.061	1.868	2.151	1.979
Note:				NA

Regressions (without the outlier)

Regression 4:

- **sav=-3101.245-86.493*educ+11.793(educ^2)-0.048*cons+88.194*age -13.964(age*size) -0.042(cons*size) + 822.285size**
- For change in years of education from 12 years to 13 years, savings increases from **(-86.493(13-12)+11.793(169-144))=$208.332**,holding consumption, age and size constant.
- For a family size of 2, 1 year increase in age, the savings increases by **(88.194-13.964*2)= $60.266** keeping education, age, size,consumption and black constant.
- For a household head of 30 years, an increase in family size by 1, increases the savings by **(822.285-13.964(30))= $403.365** keeping education, age, consumption and black constant.
- For a family consumption of $100, an increase in family size by 1, increases the savings by **(822.285-0.042(100))= $818.085** keeping education, age, consumption and black constant.
- For a family size of 2, $1 increase in consumption, the savings decreases by **(0.048-0.042*2)= $0.036** keeping education, age, size,consumption and black constant.
- For every $1 increase in consumption, saving will increase by **(0.048-0.042)=$0.006** holding education, age and size constant
- The regression model explains 6.7% of variations in savings.

Hypothesis Testing

```
Linear hypothesis test

Hypothesis:
age = 0
I(age * size) = 0

Model 1: restricted model
Model 2: sav ~ educ + I(educ^2) + cons + age + I(age * size) + size

Note: Coefficient covariance matrix supplied.

  Res.Df Df      F Pr(>F)
1     92
2     90  2 2.2196 0.1146
```

Since F2(2.21)< 3, we fail to reject the null hypothesis. This indicates that the additional terms related to 'age' and its interaction with 'size' do not significantly improve the model fit compared to the restricted model.

Probit Regression

Y- binary variable (**debt_dummy**)

1- person has negative savings, **0**- otherwise

X- years of education

- Probit regression uses the standard normal cumulative distribution function
- eg. For a person with 16 years of education (bachelor's) the probability of being in debt is:
  Pr(Y=1, x=16) = Φ(3.367+(-0.073*16))
  = Φ(2.199) = 0.98574 = 98.57%
- eg. For a person with 18 years of education (master's) the probability of being in debt is:
  Pr(Y=1, x=18) = Φ(3.367+(-0.073*18))
  = Φ(2.053) = 0.97982 = 97.98%
- Since our sample is less, the individual observations are having a substantial impact on the probability. This is leading to a less stable and more sensitive results.
- Nevertheless, the relationship between two variables savings and education still holds true. There is a **decrease in probability of being in debt with increase in years of education.**

Probit Regression-Effects of Education on debt

	Dependent variable:	
	debt_dummy	
	probit	binary model (marginal effect)
	(1)	(2)
educ	-0.073	-0.017
	(0.057)	(0.014)
cons	0.0001	0.000
	(0.00004)	(0.000)
age	-0.123	-0.029
	(0.075)	(0.017)
I(age * size)	0.025	0.006
	(0.017)	(0.004)
size	-0.825	-0.197
	(0.658)	(0.155)
Constant	3.367	0.803
	(3.008)	(0.710)
Observations	97	97
Log Likelihood	-41.566	

Logit Regression

- The Logit model also calculates the probability of a person being in debt for a certain years of education similar to probit model
- The Logit model uses cumulative standard logistic distribution function
- eg. For a person with 16 years of education (bachelor's) the probability of being in debt using Logit model is: **Pr(Y=1, x=16) = 1/1+e^{-(5.951-0.121*16)}= 0.9822 = 98.22%**
- eg. For a person with 18 years of education (master's) the probability of being in debt using Logit model is: **Pr(Y=1, x=18) =1/1+e^{-(5.951-0.121*18)} = 0.9775 = 97.75%**
- The negative correlation between education and debt is maintained in the Logit model too

Logit Regression-Effect of Education on Debt

	Dependent variable:	
	debt_dummy	
	logistic	binary model (marginal effect)
	(1)	(2)
educ	-0.121	-0.016
	(0.104)	(0.013)
cons	0.0001	0.000
	(0.0001)	(0.000)
age	-0.215	-0.028
	(0.137)	(0.017)
I(age * size)	0.043	0.006
	(0.031)	(0.004)
size	-1.417	-0.185
	(1.194)	(0.152)
Constant	5.951	0.777
	(5.446)	(0.695)
Observations	97	97
Log Likelihood	-41.720	

Internal Validation

- Data Quality
  - Complete and does not have missing values
  - Quality does not seem to affect validity
- Sample Representativeness
  - Data is homogeneous in terms of demographic
  - Sample is somewhat representative of the population but could be better
- Reverse causality
  - It is difficult to investigate whether savings effects education as we do not have additional past data

External Validation

- Macroeconomic Conditions
  - Brief 1980's recession
  - 1980's was also period of rapid technological innovation, with the widespread adoption of personal computers
  - Growth of the information technology industry
- Capitalism
  - Less saving, more investing

Conclusion

In conclusion, education affects savings, positively.

Thank You!