Provident Regressions

Dario Trujano Ochoa

Enero 2018

Data Base

I imported the data to R; then, I cleaned for the cases with missing values and reconfigured some variables:

```r
# implort with more data
ProvidentDF_complete<- read.csv("Encuesta-EducacionFinancieraDario.csv")

ProvidentDF_complete <- ProvidentDF_complete %>% filter(Status=="Completa") %>%
  filter(!is.na(p9_batepelota)) # question 9 and 12 have missing values
# measure grit as the average of the points made in each item
ProvidentGRIT <- data.frame(
  ProvidentDF_complete$X1..Los.proyectos.o.ideas.nuevas.me.distraen.de.proyectos.o.ideas.que.ten.a.desde
  ProvidentDF_complete$X2..Los..bstaculos.no.me.desaniman.,
  ProvidentDF_complete$X3..Estuve.concentrado.en.una.idea.o.proyecto.por.un.corto.tiempo..pero.despu.s.p
  ProvidentDF_complete$X4..Trabajo.duro.,
  ProvidentDF_complete$X5..Con.frecuencia.me.propongo.un.objetivo..pero.luego.trato.de.cumplir.un.objeti
  ProvidentDF_complete$X6..Me.resulta.dificil.mantener.mi.atenci.n.en.proyectos.que.duran.m.s.all..de.al
  ProvidentDF_complete$X7..Soy.chambeador.a.)
ProvidentGRIT <- apply(substr(apply(ProvidentGRIT, MARGIN = 2,FUN = as.character),start = 1,stop = 1),FU
ProvidentGRIT[,c(1,3,5,6)] <- -ProvidentGRIT[,c(1,3,5,6)] + 4
ProvidentDF_nna <- ProvidentDF_complete %>%  # there is no missing values in this DF
  mutate(grit=rowMeans(ProvidentGRIT))
ProvidentDF_nna <- ProvidentDF_nna[!is.na(ProvidentDF_nna$p9_batepelota),]

# Education
escol <- as.character(ProvidentDF_nna$escol)
escol[escol=="No estudi¢"] <- 0
escol[escol=="1ro Primaria"] <- 1
escol[escol=="2do Primaria"] <- 2
escol[escol=="3ro Primaria"] <- 3
escol[escol%in%c("4to primaria","4to Primaria")] <- 4
escol[escol=="5to Primaria"] <- 5
escol[escol%in%c(" 6to Primaria","6to Primaria")] <- 6
escol[escol=="1ro Secundaria"] <- 7
escol[escol%in%c("2do Secundaria","2do secundaria")] <- 8
escol[escol%in%c("3ro secundaria","3ro Secundaria ","3ro Secundaria")] <- 9
escol[escol=="1ro Preparatoria"] <- 10
escol[escol=="2do Preparatoria"] <- 11
escol[escol=="3ro Preparatoria"] <- 12
escol[escol=="Licenciatura"] <- 13
escol[escol=="Posgrado"] <- 14
escol <- as.numeric(escol)

#original measure
escol1 <- escol
escol1[escol1%in%c(1:6)]<- 0 # Primaria
escol1[escol1%in%c(7:9)]<- 1 # Secundaria
```

```r
escol1[escol1%in%c(10:12)]<- 2 # Preparatoria
escol1[escol1%in%c(13)]<- 3 # Licenciatura
escol1[escol1%in%c(14)]<- 4 # Posgrado



#some order for ordinal data
ProvidentDF_nna$grupo <- factor(ProvidentDF_nna$grupo, levels=c("Current", "Low Arrear","High arrear"),

attach(ProvidentDF_nna)
```

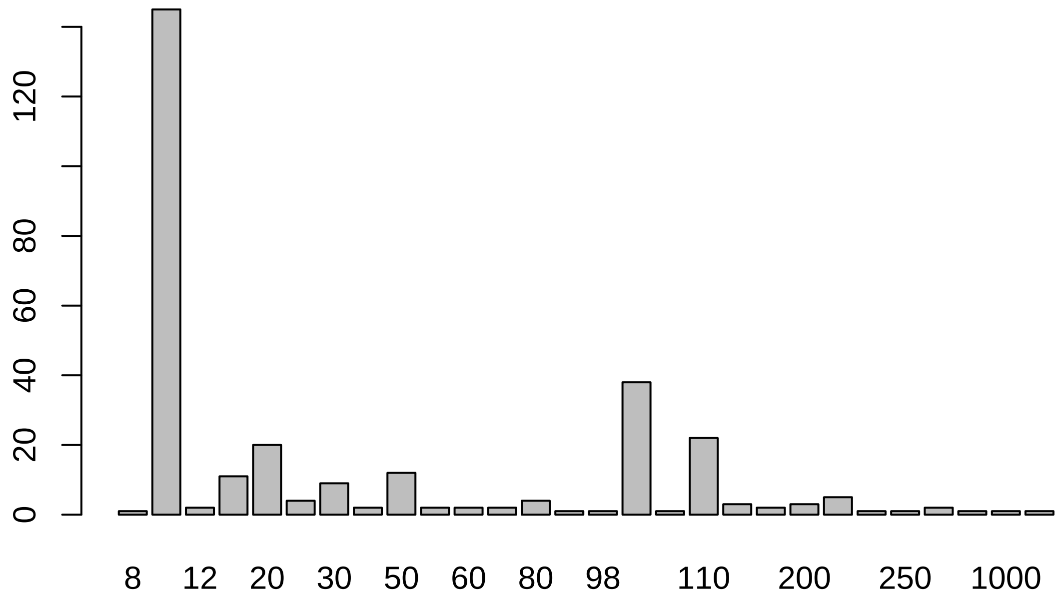
```
## The following object is masked _by_ .GlobalEnv:
##
##     escol
```

- Just married and not married
- Married = 1
- Employment as an ordinal variable.
- Full-time employment = 3
- Half-time employment = 2
- Self-employment = 1
- No earnings = 0

```r
p16_dependents<- p16_1menores+p16_2adulmay+p16_3otros # total number of dependents
casado<-{} # to capture later de degree of responsability, we just need married or unmarried people
casado[p15_estadocivil=="Casado"]<- 1
casado[p15_estadocivil!="Casado"]<- 0
empleoRemunerado <- {} # consider non paid work together
empleoRemunerado[statusempleo%in%c("Desempleados","Ama de Casa","Retirado")]<- 0
empleoRemunerado[statusempleo=="Auto empleado"]<- 1
empleoRemunerado[statusempleo=="Empleado Medio tiempo"]<- 2
empleoRemunerado[statusempleo=="Empleado tiempo completo"]<- 3
```

- To masure the responses in the Finantial literacy we take the distance from the correct answers:

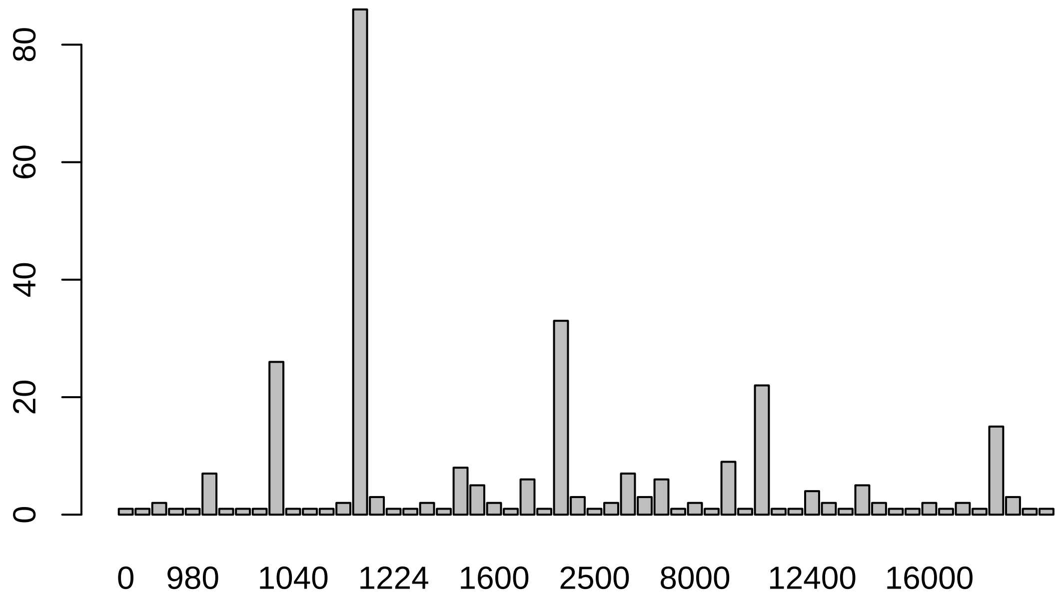
```r
# pregunta del bat y la pelota:
# We can see that most people answered predictably bad: $ 10. But no one answered correctly.
barplot(table(p9_batepelota))
```




```
#9. Un bate y una pelota de béisbol cuestan en total $110 pesos.
#El bate cuesta $100 más que la pelota,
#¿Cuánto cuesta la pelota? (Indique el costo en pesos de la pelota)
## Respuesta correcta: $5

p9_batepelota<- abs(p9_batepelota-5)

# pregunta sobre el Interés
barplot(table(p12_retornodeposito))
```




```
#12. Imagínese que usted deposita $1,000 pesos al inicio del año en una cuenta de ahorro
#con un interés garantizado del 2% al año y la cuenta no tiene ningún costo por mantenerla.
#Además, suponga que usted no saca dinero de esa cuenta.
#¿Cuánto dinero tendría en la cuenta después de un año incluyendo el pago de los intereses?
## Respuesta Correcta: 1020
p12_retornodeposito <- abs(p12_retornodeposito-1020)
```

PCA analysis

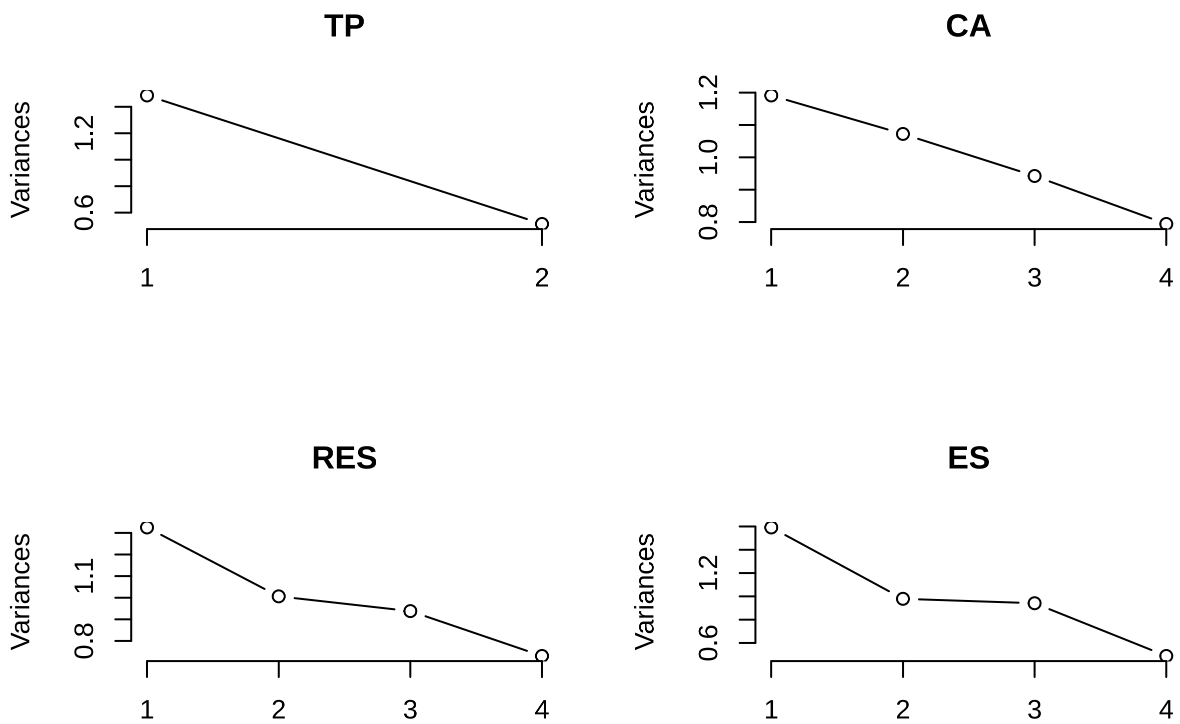
First of all I have to highlight that there is an important error in the way indeces were constructed in the article: eigenvalues were calculated using no centered data which lead towards a biased estimate of the index. In this case the variable with the hights mena will attract all the weight.

Considering that all the regression made over the those indices are wrong.

I'll consider the ideces as constructed in the article:

```
# time preference
TP <- prcomp(na.omit(cbind(p10_tandacorto,p11_tandalargo)),center = T,scale. = T)
#cognitive hability
CA <- prcomp(na.omit(cbind( p9_batepelota, p12_retornodeposito,p13_preciosbajan,escol)),center = T,scal
# cognitive ability with less levels for escol
CA1 <- prcomp(na.omit(cbind( p9_batepelota, p12_retornodeposito,p13_preciosbajan,escol1)),center = T,sca
# responsability
RES <- prcomp(na.omit(cbind(edad,hijos,casado, p16_dependents)),center = T,scale. = T)
# economic sucess
ES <- prcomp(na.omit(cbind(empleoRemunerado,SCelular,STelCasa,SResidencia)),center = T,scale. = T)
```

PCA visual analysis shows that there is no a big difference in the variance explained by the components:



PCA weigths

```r
TP$rotation
```

```
##                        PC1        PC2
## p10_tandacorto 0.7071068  0.7071068
## p11_tandalargo 0.7071068 -0.7071068
```

```r
CA$rotation
```

```
##                             PC1          PC2         PC3        PC4
## p9_batepelota       -0.1725868  0.692953358 -0.63869914 -0.2865185
## p12_retornodeposito -0.1795121  0.622998104  0.75772040 -0.0742197
## p13_preciosbajan    -0.6510778 -0.362828070  0.07923137 -0.6619485
## escol                0.7169985 -0.006693184  0.10791477 -0.6886383
```

```r
RES$rotation
```

```
##                        PC1         PC2        PC3         PC4
## edad           -0.3751918  0.64613531 -0.5833343  0.31853004
## hijos           0.5919374  0.19803781 -0.4925299 -0.60646964
## casado          0.2569157  0.73105085  0.6316692 -0.02351686
## p16_dependents  0.6654589 -0.09410012 -0.1346455  0.72813470
```

```r
ES$rotation
```

```
##                          PC1        PC2          PC3       PC4
## empleoRemunerado  0.3109411 -0.3418428  0.876328114 0.1360446
## SCelular          0.6481033  0.1812993 -0.266362713 0.6900316
## STelCasa         -0.6436841 -0.3333109 -0.008564225 0.6888405
## SResidencia      -0.2625804  0.8597545  0.401281193 0.1756334
```

Final Regressions with PCA

```r
DV <- cbind.data.frame(
       Grit = grit,
       TP1 = TP$x[,1], TP2 = TP$x[,2],
       CA1 = CA1$x[,1], CA2 = CA1$x[,2],
       ES1 = ES$x[,1], ES2 = ES$x[,2],
       edad, hijos,casado, p16_dependents,
       ProvidentDF_nna$Prestamos.activos,
       Genero)

final_regression <- polr(grupo ~ Grit+TP2+CA2+edad, data = DV,Hess=TRUE,na.action = na.omit) # model
  stargazer(final_regression)
```

% Table created by stargazer v.5.2 by Marek Hlavac, Harvard University. E-mail: hlavac at fas.harvard.edu
% Date and time: lun., ene. 15, 2018 - 09:30:28 p. m.

Table 1:

	Dependent variable:
	grupo
Grit	-0.503^{**}
	(0.241)
TP2	0.500^{***}
	(0.161)
CA2	-0.213^{*}
	(0.116)
edad	-0.028^{***}
	(0.009)
Observations	299
Note:	*p<0.1; **p<0.05; ***p<0.01

In order to maintain the significance of CA2, it was required that escol was constructed according with the first categorization, otherwise the component is significative just at 10%. It is remarkable that this categorization consider no instruction and elementary instruction in the same level of escolarity.

Robustez

% Table created by stargazer v.5.2 by Marek Hlavac, Harvard University. E-mail: hlavac at fas.harvard.edu
% Date and time: lun., ene. 15, 2018 - 09:30:33 p. m.

Table 2:

	Dependent variable:						
	grupo						
	(1)	(2)	(3)	(4)	(5)	(6)	(7)
Grit	-0.503^{**}	-0.517^{**}	-0.520^{**}	-0.428^{*}	-0.503^{**}		-0.448^{*}
	(0.241)	(0.242)	(0.243)	(0.238)	(0.241)		(0.239)
TP1		0.132	0.127			0.125	0.113
		(0.087)	(0.087)			(0.086)	(0.086)
TP2	0.500^{***}	0.507^{***}	0.515^{***}	0.460^{***}	0.500^{***}	0.504^{***}	0.466^{***}
	(0.161)	(0.162)	(0.162)	(0.158)	(0.161)	(0.162)	(0.159)
CA2	-0.213^{*}	-0.231^{**}	-0.239^{**}	-0.210^{*}	-0.213^{*}	-0.198^{*}	
	(0.116)	(0.116)	(0.117)	(0.115)	(0.116)	(0.114)	
edad	-0.028^{***}	-0.027^{***}	-0.027^{***}		-0.028^{***}	-0.025^{***}	-0.027^{***}
	(0.009)	(0.009)	(0.009)		(0.009)	(0.009)	(0.009)
casado			-0.155				
			(0.220)				
AIC	645.35	645.04	646.54	653.02	645.35	647.65	647.37
BIC	667.55	670.94	676.14	671.52	667.55	669.86	669.57
Observations	299	299	299	299	299	299	299
Note:					*p<0.1; **p<0.05; ***p<0.01		

Considering debt history

% Table created by stargazer v.5.2 by Marek Hlavac, Harvard University. E-mail: hlavac at fas.harvard.edu
% Date and time: lun., ene. 15, 2018 - 09:30:33 p. m.

Table 3:

	Dependent variable:		
	grupo		
	(1)	(2)	(3)
Grit	-0.517^{**}	-0.514^{**}	-0.514^{**}
	(0.242)	(0.261)	(0.261)
TP1	0.132	0.121	
	(0.087)	(0.092)	
TP2	0.507^{***}	0.352^{**}	
	(0.162)	(0.170)	
p10_tandacorto			0.818^{**}
			(0.342)
p11_tandalargo			-0.411
			(0.336)
CA2	-0.231^{**}	-0.194	-0.194
	(0.116)	(0.124)	(0.124)
edad	-0.027^{***}	-0.017^{*}	-0.017^{*}
	(0.009)	(0.010)	(0.010)
Prestamos.activos		-1.744^{***}	-1.744^{***}
		(0.222)	(0.222)
AIC	645.04	573.36	573.36
BIC	670.94	602.96	602.96
Observations	299	299	299
Note:	*p<0.1; **p<0.05; ***p<0.01		