

TERCERA PARTE DE COMANDOS A APLICAR

Manricon

Clear all

```
rm(list = ls())
```

Use 'path'

```
library(tidyverse)
haven::read_dta('dat/qw.dta') %>% # leer qw.dta
  saveRDS('dat/qw.rds') # guardar data .r
qw <- read_rds('dat/qw.rds') # leer qw.ds
```

Limpieza de datos

Nombre de las variables de la data original nombre, Distrito, id, sexo, AÑOS, meses, meses_t, D, p1, p1_c, p1_, p2, p2_c, p2_, p3, p3_c, p3_, p4, p4_c, p4_, mem_ct, mem_ct_in, aten, aten_in, matematica, comunicacion, asistencia, faltas, peso, talla, imc, d_norm, d_bajo, d_sobr, DIAGNÓSTICO, tamiza_ane, descarte, d_cnane, d_snane

```
qw %>%
  drop_na() %>% #eliminamos los 'missing values'
  head(6) %>%
  knitr::kable()
```

nombre	Distrito	id	sexo	AÑOS	meses	meses_t	D	p1	p1_c	p1_	p2	p2_c
Jhovisa Araceli Santana Roman	1	14	1	12	4	148	1	7	11	0.4	14	17
Xiomara Flor Cahuana Feirer	1	14	1	12	4	148	1	8	12	0.4	17	20
Rosmery Domitila Delzo Antonio	1	14	1	12	9	153	1	6	9	0.3	8	10
Josef Albert Antonio Justo	1	14	0	12	1	145	1	8	12	0.4	14	17
Jennifer Yadyra Canchari Poma	1	14	1	12	4	148	1	9	14	0.5	15	18
Jorge Luis Sierra Justo	1	14	0	12	5	149	1	5	8	0.3	10	12

Se trabaja con estos datos:aten, sexo, tamiza_ane(nivel de hemoglobina), p3(planteamiento), p4(cancelacion de pares), IMCC, reasignando la como la base de datos qw1

```
qw1 <- qw %>%
  drop_na() %>%
  select(aten, sexo, tamiza_ane, p3, p4, imc) %>%
  mutate(sexo = factor(sexo, labels = c("Masculino", "Femenino")))
qw1 %>% head(6) %>% knitr::kable()
```

aten	sexo	tamiza_ane	p3	p4	imc
28	Femenino	13	13	12	17
35	Femenino	12	17	14	18
28	Femenino	12	14	10	18
30	Masculino	12	13	14	16
38	Femenino	13	17	17	19
23	Masculino	12	13	7	15

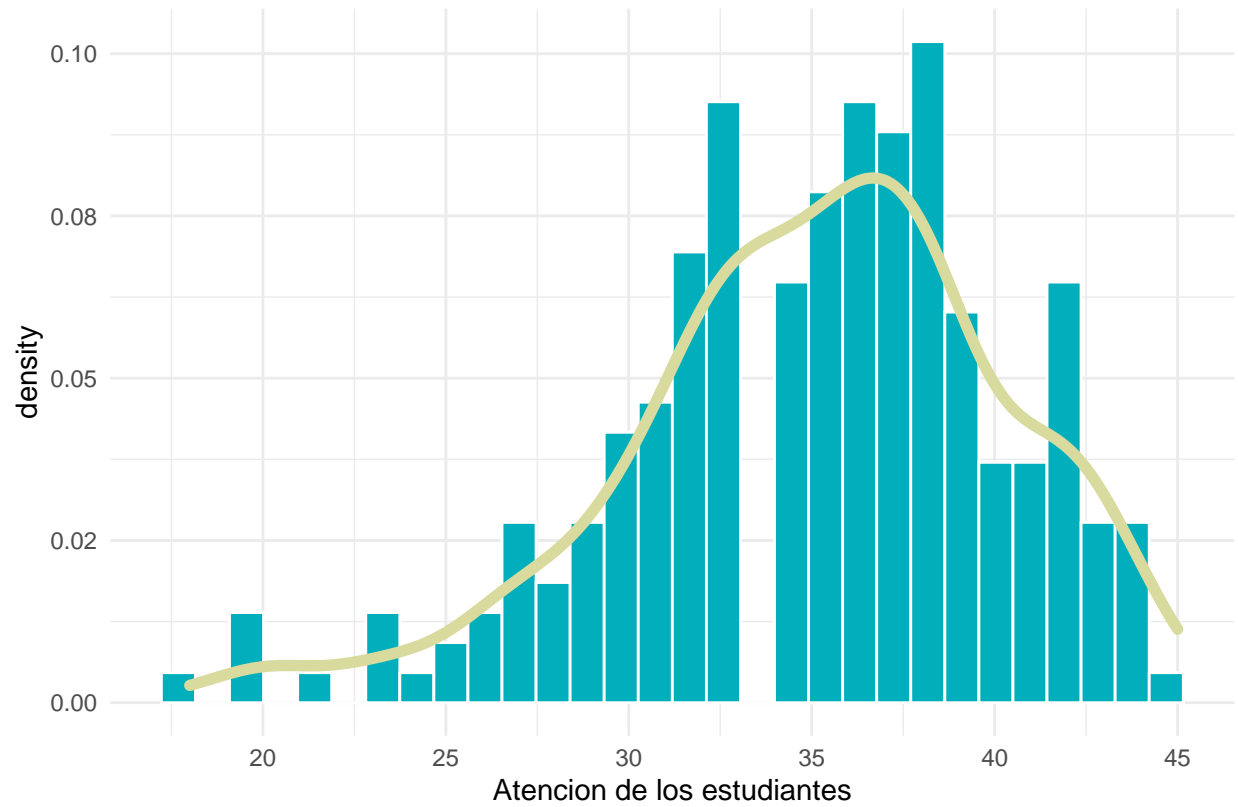
Descripcion de las variables

```
summary(qw1) %>% #resumen
knitr::kable()
```

	aten	sexo	tamiza_ane	p3	p4	imc
	Min. :18	Masculino:102	Min. :11	Min. : 2	Min. : 7	Min. :15
	1st Qu.:32	Femenino :130	1st Qu.:13	1st Qu.:12	1st Qu.:17	1st Qu.:17
	Median :36	NA	Median :14	Median :14	Median :19	Median :18
	Mean :35	NA	Mean :14	Mean :14	Mean :18	Mean :18
	3rd Qu.:38	NA	3rd Qu.:15	3rd Qu.:16	3rd Qu.:19	3rd Qu.:20
	Max. :45	NA	Max. :20	Max. :20	Max. :20	Max. :25

Atencion

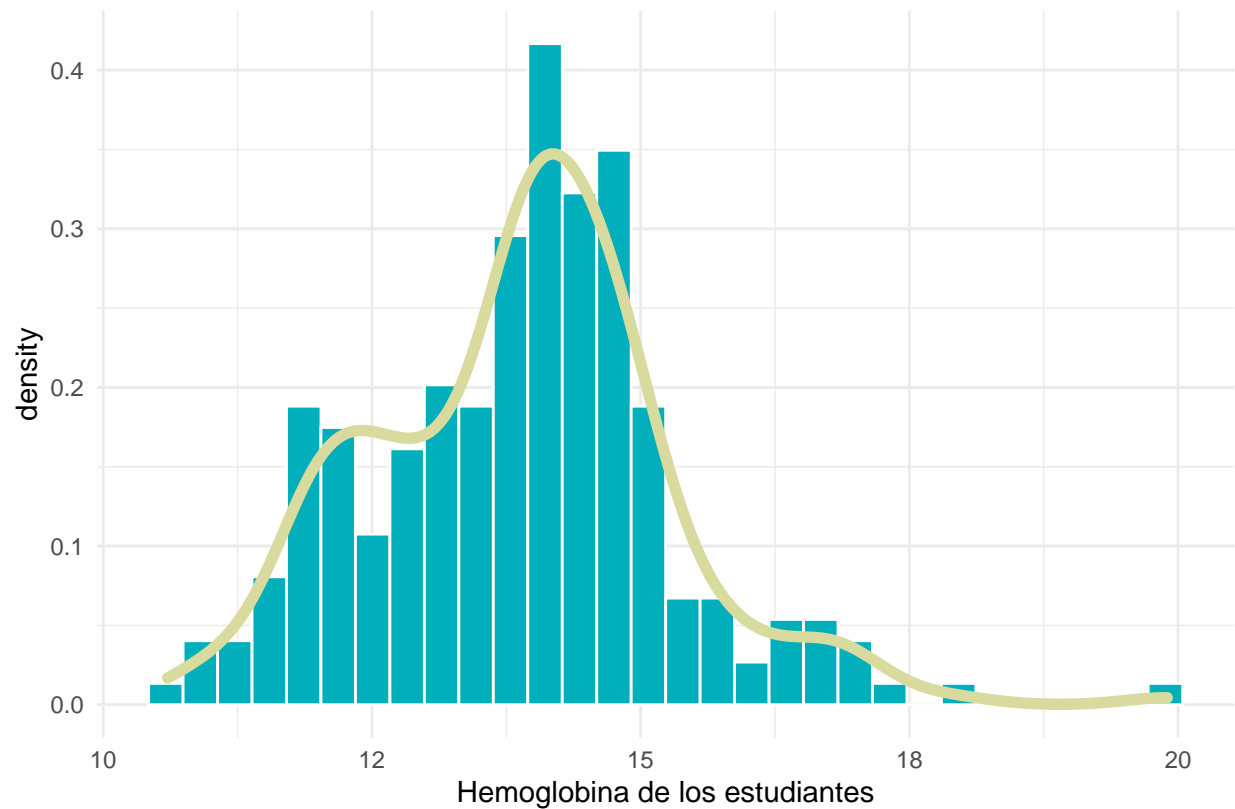
```
qw1 %>%
  ggplot(aes(aten)) +
  geom_histogram(aes(y = ..density.. ), fill = "#00AFBB", color = "white") +
  geom_density(size = 2, color = "#D8DA9E") +
  theme_minimal() +
  labs(x = "Atencion de los estudiantes",
       caption = "Datos: QW.dta")
```



Datos: QW.dta

Nivel de Hemoglobina

```
qw1 %>%
  ggplot(aes(tamiza_ane)) +
  geom_histogram(aes(y = ..density..), fill = "#00AFBB", color = "white") +
  geom_density(size = 2, color = "#D8DA9E") +
  theme_minimal() +
  labs(x = "Hemoglobina de los estudiantes",
       caption = "Datos: QW.dta")
```

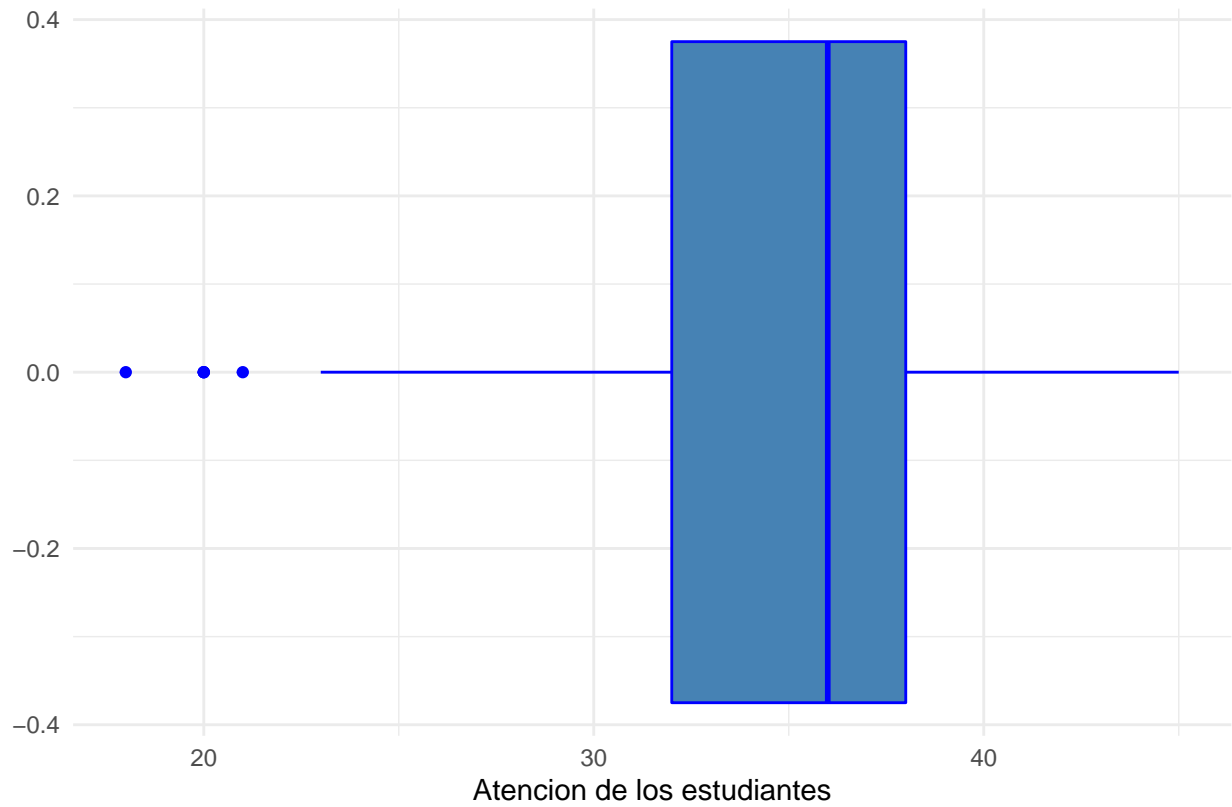


Datos: QW.dta

Diagrama de caja y violin

Atencion

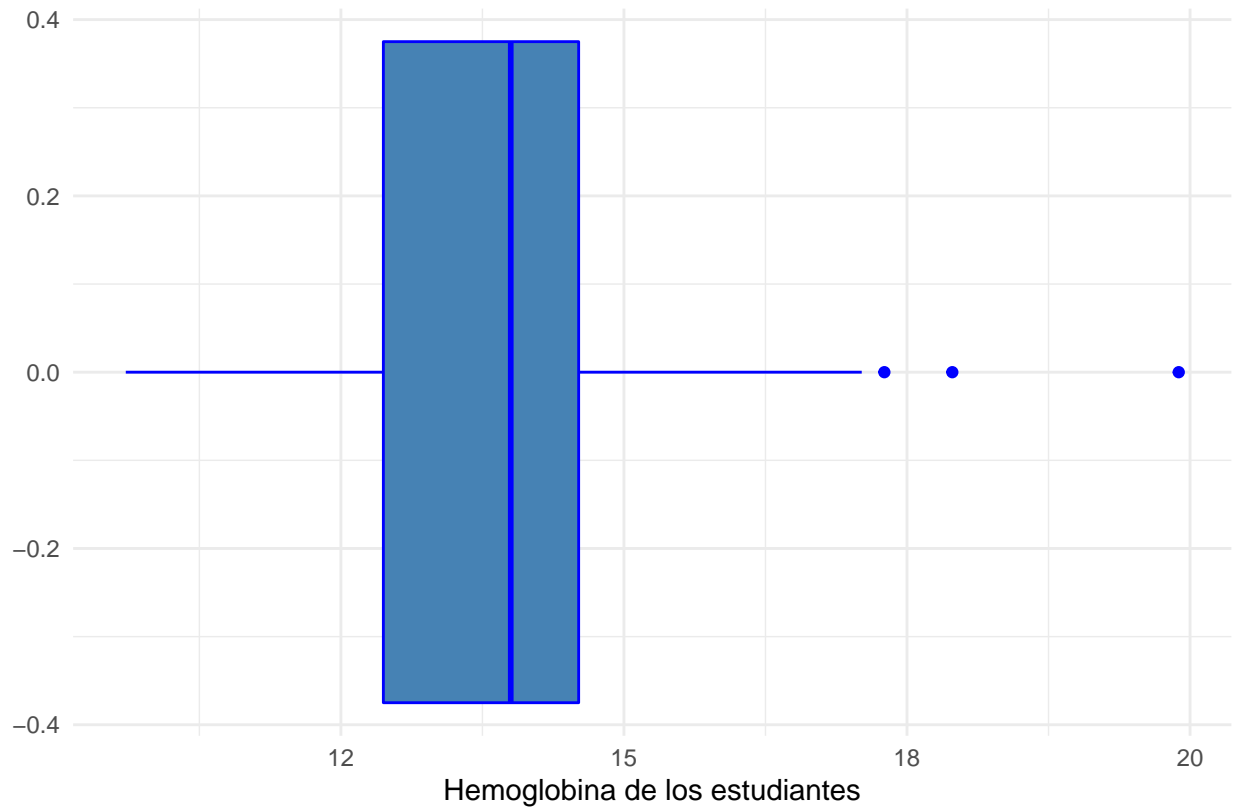
```
qw1 %>%
  ggplot(aes(aten)) +
  geom_boxplot(color = "blue", fill = "steelblue") + scale_color_brewer(palette = "Dark2") +
  theme_minimal() +
  labs(x = "Atencion de los estudiantes",
       caption = "Datos: QW.dta")
```



Datos: QW.dta

Hemoglobina

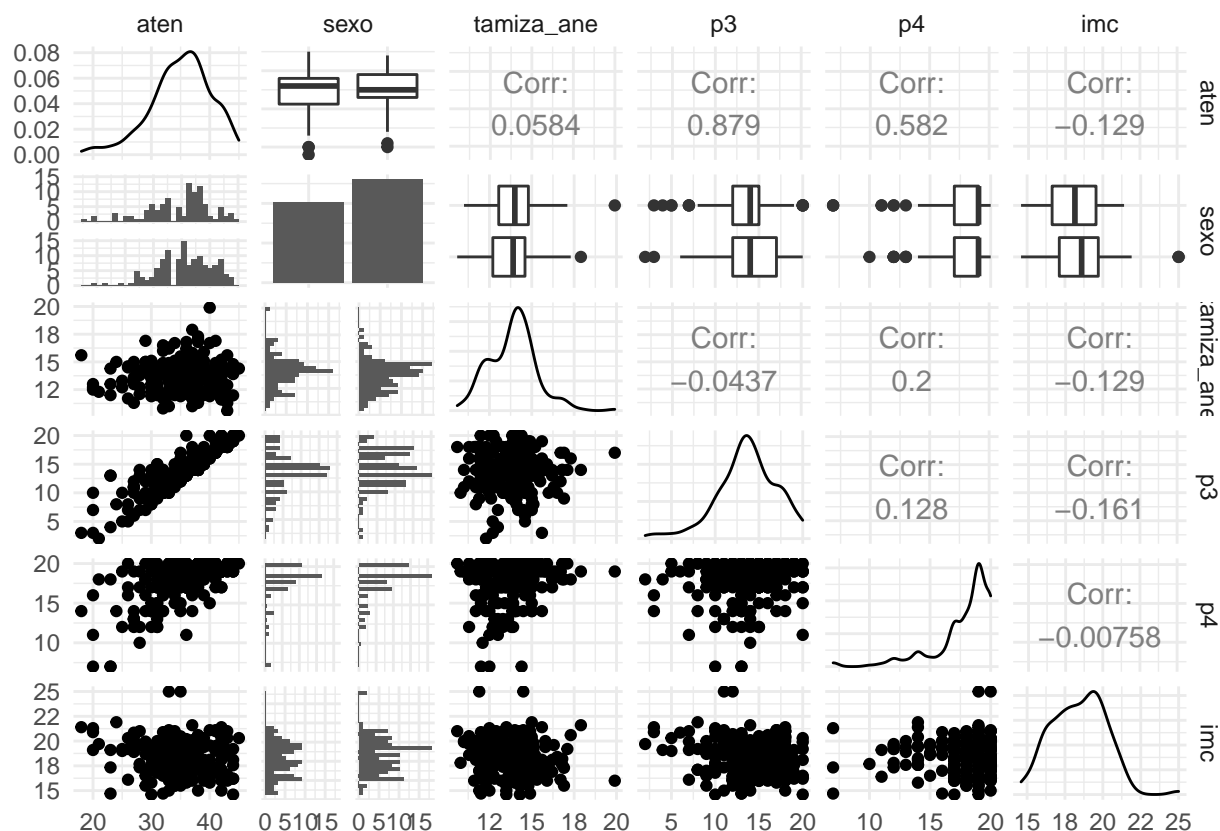
```
qw1 %>%
  ggplot(aes(tamiza_ane)) +
  geom_boxplot(color = "blue", fill = "steelblue") + scale_color_brewer(palette = "Dark2") +
  theme_minimal() +
  labs(x = "Hemoglobina de los estudiantes",
       caption = "Datos: QW.dta")
```



Datos: QW.dta

correlaciones

```
library(GGally)
ggpairs(qw1) + theme_minimal()
```



Estimacion

Modelo 1

```
library(broom)
model <- aten ~ sexo + tamiza_ane + p3 + p4 + imc #definimos el modelo

est1 <- lm(model, data = qw1)
tidy(est1) %>%
  knitr::kable()
```

term	estimate	std.error	statistic	p.value
(Intercept)	-0.3	0.3	-0.9	0.4
sexoFemenino	0.0	0.0	-0.3	0.8
tamiza_ane	0.0	0.0	-0.1	0.9
p3	1.3	0.0	212.4	0.0
p4	1.0	0.0	123.6	0.0
imc	0.0	0.0	1.6	0.1

Extraccion del modelo

```
library(equatiomatic)
extract_eq(est1)
```

$$\text{aten} = \alpha + \beta_1(\text{sexoFemenino}) + \beta_2(\text{tamiza_ane}) + \beta_3(\text{p3}) + \beta_4(\text{p4}) + \beta_5(\text{imc}) + \epsilon$$

```
extract_eq(est1, use_coefs = T)
```

$$\text{aten} = -0.31 - 0.01(\text{sexo}_{\text{Femenino}}) + 0(\text{tamiza_ane}) + 1.25(\text{p3}) + 1.01(\text{p4}) + 0.02(\text{imc}) + \epsilon$$

Modelo 2

```
qw1 <- qw1 %>%
  mutate(tamiza_ane = tamiza_ane/12)
model <- aten ~ sexo + tamiza_ane + p3 + p4 + imc
est2 <- lm(model, data = qw1)
tidy(est1) %>%
  knitr::kable()
```

term	estimate	std.error	statistic	p.value
(Intercept)	-0.3	0.3	-0.9	0.4
sexoFemenino	0.0	0.0	-0.3	0.8
tamiza_ane	0.0	0.0	-0.1	0.9
p3	1.3	0.0	212.4	0.0
p4	1.0	0.0	123.6	0.0
imc	0.0	0.0	1.6	0.1

Extraccion del modelo

```
extract_eq(est2)
```

$$\text{aten} = \alpha + \beta_1(\text{sexo}_{\text{Femenino}}) + \beta_2(\text{tamiza_ane}) + \beta_3(\text{p3}) + \beta_4(\text{p4}) + \beta_5(\text{imc}) + \epsilon$$

```
extract_eq(est2, use_coefs = T)
```

$$\text{aten} = -0.31 - 0.01(\text{sexo}_{\text{Femenino}}) - 0.02(\text{tamiza_ane}) + 1.25(\text{p3}) + 1.01(\text{p4}) + 0.02(\text{imc}) + \epsilon$$

Normalidad de los residuos

Del ultimo modelo

```
qw1$residuals.est2 <- residuals( est2 ) # residuos
qw1$rstudent.est2 <- rstudent( est2 ) # residuos estudentizados
```

Normalidad de residuos

```
sha <- shapiro.test(qw1$rstudent.est2)
sha
```

Shapiro-Wilk normality test

data: qw1\$rstudent.est2 W = 0.9, p-value = 2e-11

No hay normalidad de residuos

Multicolineadlidad

```
library(caret)
car::vif(est1)
```



```

sexo tamiza_ane      p3      p4      imc
1      1      1      1      1

```

Estimacion robusta

```

library(MASS)
est3 <- rlm(model, data = qw1)
tidy(est3) %>% knitr::kable()

```

term	estimate	std.error	statistic
(Intercept)	-0.3	0.3	-0.9
sexoFemenino	0.0	0.0	-0.3
tamiza_ane	0.0	0.2	-0.1
p3	1.3	0.0	212.4
p4	1.0	0.0	123.6
imc	0.0	0.0	1.6

Modelos juntos

```

stargazer::stargazer(est1, est2, est3, type = "latex", header = F, title = "Todos los modelos")

```

Table 1: Todos los modelos

	<i>Dependent variable:</i>		
	aten		
	<i>OLS</i>	<i>robust linear</i>	
	(1)	(2)	(3)
sexoFemenino	−0.010 (0.040)	−0.010 (0.040)	−0.010 (0.040)
tamiza_ane	−0.002 (0.010)	−0.020 (0.200)	−0.020 (0.200)
p3	1.000*** (0.006)	1.000*** (0.006)	1.000*** (0.006)
p4	1.000*** (0.008)	1.000*** (0.008)	1.000*** (0.008)
imc	0.020 (0.010)	0.020 (0.010)	0.020 (0.010)
Constant	−0.300 (0.300)	−0.300 (0.300)	−0.300 (0.300)
Observations	232	232	232
R ²	1.000	1.000	
Adjusted R ²	1.000	1.000	
Residual Std. Error (df = 226)	0.300	0.300	0.500
F Statistic (df = 5; 226)	14,248.000***	14,248.000***	

Note:

*p<0.1; **p<0.05; ***p<0.01