

QW

US

La adaptacion en R

- importar
- renombrar, etiquetas,
- nuevas variables, por grupos
- tablas
- estadística descriptiva
- gráficos(caja, histogramas, densidad, correlaciones, dispersión),
- regresión, predicción
- visualización de heterocedasticidad(aunque no vi la corrección de esta)
- comparación de modelos

El trabajo

Importar datos

Explorar Datos

Por lo que se desconoce las etiquetas de los distritos pero sirve para hacer comparaciones

Distrito	D	sexo	N()	Estudiante_Anemia	Atencion	Peso	Talla	imc
1	secundaria	Masculino	2	0	44	38	1	20
1	secundaria	Femenino	4	0	41	41	2	19
1	primaria	Masculino	9	0	37	29	16	16
1	primaria	Femenino	8	0	37	32	1	18
2	primaria	Masculino	4	2	37	33	1	16
2	primaria	Femenino	9	4	38	40	1	19
3	primaria	Masculino	35	0	34	38	1	18
3	primaria	Femenino	44	1	35	38	1	18
4	primaria	Masculino	17	2	32	38	1	20
4	primaria	Femenino	19	3	34	36	1	20
5	secundaria	Masculino	18	0	35	34	1	17
5	secundaria	Femenino	12	0	35	36	1	17
5	primaria	Masculino	12	0	34	34	1	18
5	primaria	Femenino	19	0	36	38	1	19
6	secundaria	Masculino	5	5	38	38	1	20
6	secundaria	Femenino	15	0	33	36	1	19

D	sexo	n_estudiantes	Estudiante_Anemia	Atencion	Peso	Talla	std_error
secundaria	Masculino	25	5	36	35	1	0.9
secundaria	Femenino	31	0	35	37	1	0.8
primaria	Masculino	77	4	34	36	3	0.7
primaria	Femenino	99	8	36	37	1	0.5

Generar nuevas variables

Etiquetas ya puestas, el analisis se limitara hacer renombrar variables

Clases de variables: cognitivo , condicion de la persona

- Educacion: memoria a corto plazo, atencion, nivel cognitivo y asistencias
- Salud: IMC, tamizaje(tratamiento promedio)
 - cognitivo: `p[i]_c` i en `c(1:4)`, `mem_ct`, `aten`,
* matematica, comunicacion en base 20
 - intermedia: asistencia, faltas, estas mutuamente excluyentes
 - condicion: peso, talla, `imc`, `d_norm`, `d_sobr`, `tamiza_anem` descarte, `dcnamem`

Las variables excluidas estan implicitamente en otras

Variables relevantes: `sexo`, `edad_mese`, `mem_ct`, `mem_cr_in`, `aten`, `aten_in`, `imc`, `tamiza_anem`,

```
## # A tibble: 2 x 3
##   control pro_aten pro_mem_cp
##   <fct>      <dbl>      <dbl>
## 1 0        -0.00906      0.0431
## 2 1         0.0201     -0.0957

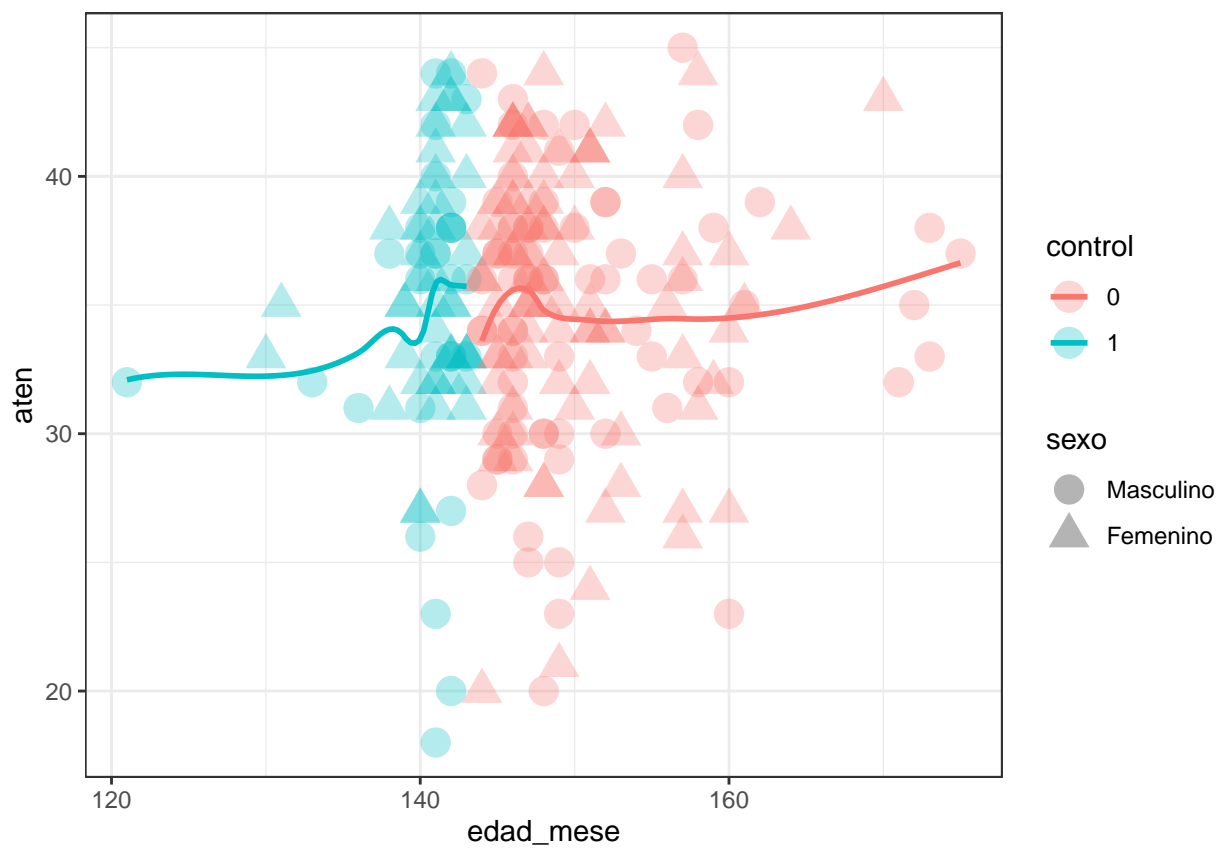
## # A tibble: 2 x 6
##   control mem_ct matematica comunicacion asistencia  imc
##   <fct>      <dbl>      <dbl>      <dbl>      <dbl> <dbl>
## 1 0         35.5      16.4      14.5      99.0  18.3
## 2 1         34.7      17.0      13.9      99.0  18.6
```

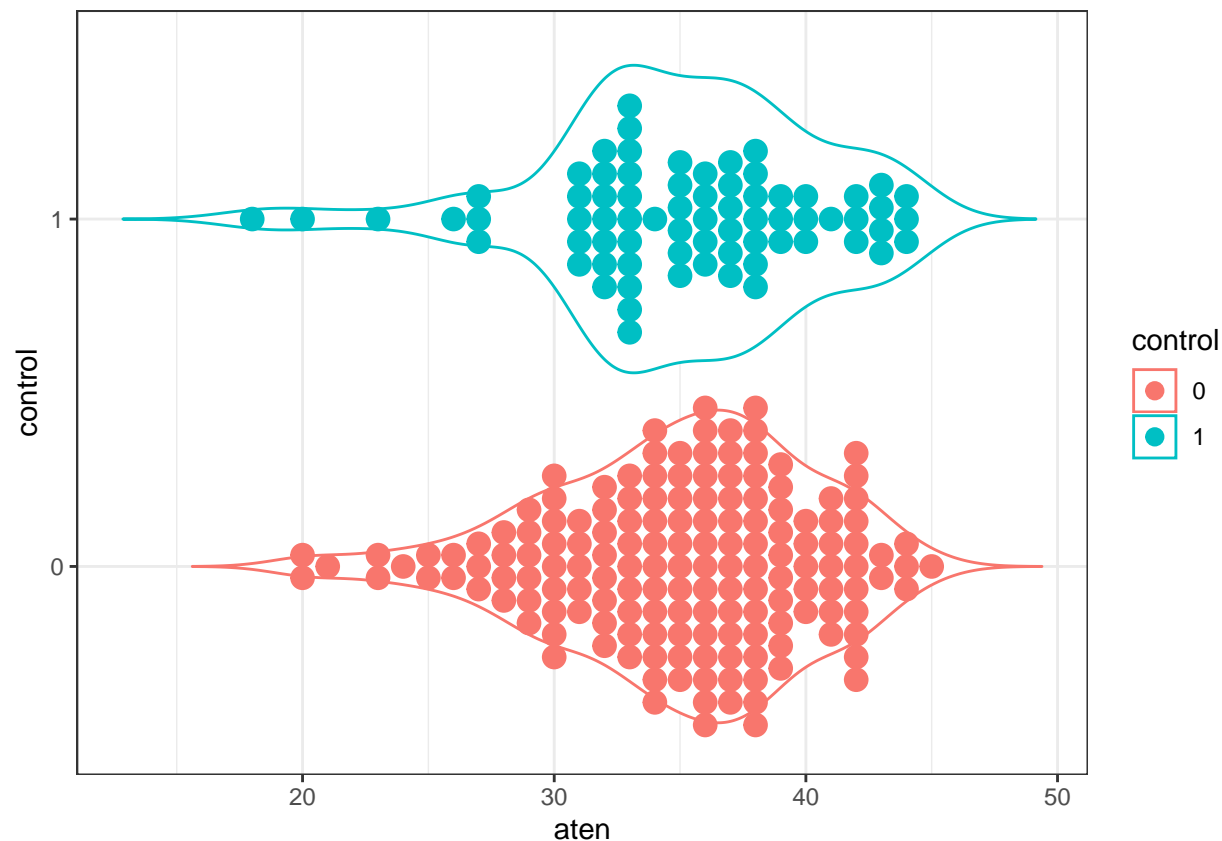
Diferencias significativas

tablas

grado	sexo	promedio p1-p4	promedio indice p	promedio atencion	promedio edad en meses
secundaria	Masculino	115	0.8	36	157
secundaria	Femenino	112	0.8	35	152
primaria	Masculino	114	0.8	34	145
primaria	Femenino	115	0.8	36	145

Graficos(caja, histogramas, densidad, correlaciones, dispersion),





Estimacion

Y: Control por grupos (Atencion, Mem_ct)

comparacion de modelos

Apendice

```
knitr::opts_chunk$set(fig.pos = "H", # Fijar posicion de las figuras
echo = F, # si es FALSE Correr el codigo pero mostrarlo
message = F, # si es FALSE Mensajes omitidos
warning = F, # si es FALSE Advertencias omitidas
out.extra = "",
include = T, # si es FALSE
cache = F #
)
options(knitr.duplicate.label = "allow", # permitir Chunk names repetidos
knitr.table.format = function() { # funcione de acuerdo a la salida del texto
if (knitr::is_latex_output()) 'latex' else 'pandoc'
}, digits = 1,
kableExtra.auto_format = FALSE
)
library(tidyverse)
library(knitr)
library(kableExtra)
```

```

haven::read_dta('dat/qw.dta') %>%
  saveRDS('dat/qw.rds')
qw <- read_rds('dat/qw.rds')
options(kableExtra.auto_format = FALSE)
qw <- qw %>%
  mutate(sexo = factor(sexo, labels = c("Masculino", "Femenino")),
         D = factor(D, labels = c('secundaria', "primaria")),

         ) %>% drop_na()
qw %>%
  group_by(Distrito, D, sexo ) %>%
  summarise('N()' = n(),
            Estudiante_Anemia = sum(d_cnane),
            Atencion = mean(aten),
            Peso = mean(peso),
            Talla = mean(talla),
            imc = mean(imc)
            ) %>%
  kable()

qw %>%
  group_by(D, sexo ) %>%
  summarise(n_estudiantes = n(),
            Estudiante_Anemia = sum(d_cnane),
            Atencion = mean(aten),
            Peso = mean(peso),
            Talla = mean(talla),
            std_error = sd(aten) / sqrt(n())
            ) %>%
  kable()
qw1 <- qw %>%
  select(-nombre, -Distrito, -id, -faltas, -d_bajo, -d_sname) %>%
  mutate(edad_mese = meses_t, grado = D,
         control = factor(ifelse(edad_mese < 144, 1, 0)),
         correctasT = p1_c + p2_c + p3_c + p4_c, # total de preguntas p correctas
         pregunT = p1_c/p1_ + p2_c/p2_ + p3_c/p3_ + p4_c/p4_, # preguntas p totales aproximacion
         ind_corre = correctasT/pregunT, # indice de preguntas p correctas
         prueba = (aten - mean(aten)) / sd(aten),
         mem_cp = (mem_ct - mean(mem_ct)) / sd(mem_ct)
         )
qw1 %>% # promedi de las varianzas
  group_by(control) %>%
  summarise(pro_aten = mean(prueba),
            pro_mem_cp = mean(mem_cp))
var <- c('mem_ct', 'ate', 'matematica', 'comunicacion', 'asistencia', 'imc')

qw1 %>%
  group_by(control) %>%
  select(one_of(var)) %>%
  summarise_all(funs(mean(., na.rm = T))) %>% tibble()

qw1 %>%

```

```

select(sexo, edad_mese, grado, correctasT, ind_corre, everything()) %>%
group_by(grado, sexo) %>%
summarise(
  "promedio p1-p4" = mean(correctasT),
  "promedio indice p" = mean(ind_corre),
  "promedio atencion" = mean(aten),
  "promedio edad en meses" = mean(edad_mese)
) %>%
kable()

qw1 %>%
mutate(control = factor(ifelse(edad_mese < 144, 1, 0)))
) %>%
ggplot(aes(edad_mese, aten, color = control)) +
geom_point(aes(shape = sexo), size = 5, alpha = .3) + theme_bw() + geom_smooth(se = F)
library(gridExtra); library(GGally)
qw1 %>%
ggplot(aes(control, aten)) +
geom_violin(aes(color = control), trim = FALSE) +
geom_dotplot(aes(color = control, fill = control), binaxis = 'y', stackdir = 'center') + theme_bw()
coord_flip()

#ggpairs(qw1)

```