

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

```
library(tidyverse)
```

```
## -- Attaching packages -----  
## v ggplot2 3.3.0      v purrr  0.3.4  
## v tibble  3.0.1      v dplyr  0.8.5  
## v tidyr   1.0.3      v stringr 1.4.0  
## v readr   1.3.1      v forcats 0.5.0  
  
## -- Conflicts -----  
## x dplyr::filter() masks stats::filter()  
## x dplyr::lag()    masks stats::lag()
```

```
library(kableExtra)
```

```
##  
## Attaching package: 'kableExtra'  
  
## The following object is masked from 'package:dplyr':  
##  
##      group_rows
```

```
haven::read_dta('dat/qw.dta') %>%  
  saveRDS('dat/qw.rds')  
qw <- read_rds('dat/qw.rds')
```

Explorar Datos

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

```

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()

```

Distrito	D	sexo	N()	Estudiante_Anemia	Atencion	Peso	Talla	imc
1	secundaria	Masculino	2	0	43.50000	38.50000	1.375000	20.35000
1	secundaria	Femenino	4	0	41.00000	40.75000	1.472500	18.75000
1	primaria	Masculino	9	0	37.22222	29.22222	16.515556	16.35556
1	primaria	Femenino	8	0	37.00000	31.50000	1.340000	17.45375
2	primaria	Masculino	4	2	37.25000	33.25000	1.390000	16.50000
2	primaria	Femenino	9	4	38.44444	39.77778	1.444444	19.11111
3	primaria	Masculino	35	0	33.91429	38.00000	1.436000	18.25971
3	primaria	Femenino	44	1	34.81818	37.86364	1.437045	18.17136
4	primaria	Masculino	17	2	31.94118	38.41176	1.402941	19.54118
4	primaria	Femenino	19	3	34.52632	36.52632	1.366842	19.55789
5	secundaria	Masculino	18	0	34.72222	34.00000	1.400000	17.32944
5	secundaria	Femenino	12	0	35.41667	36.50000	1.449167	17.30833
5	primaria	Masculino	12	0	33.66667	34.16667	1.372500	18.08226
5	primaria	Femenino	19	0	36.31579	38.26316	1.416842	19.07067
6	secundaria	Masculino	5	5	38.40000	37.60000	1.380000	19.76000
6	secundaria	Femenino	15	0	33.13333	36.26667	1.367333	19.41333

```

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

  kable()

```

D	sexo	N()	Estudiante_Anemia	Atencion	Peso	Talla
secundaria	Masculino	25	5	36.16000	35.08000	1.394000
secundaria	Femenino	31	0	35.03226	36.93548	1.412581
primaria	Masculino	77	4	34.00000	36.22078	3.178961
primaria	Femenino	99	8	35.55556	37.34343	1.412525

Generar nuevas variables

Etiquetas ya puestas, el analisis se limitara hacer renombrar variables

Clases de variables: cognitivo , condicion de la persona

- 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

```
qw1 <- qw %>%
  select(-nombre, -Distrito, -id, -faltas, -d_bajo, -d_sname) %>%
  mutate(edad_mese = meses_t, grado = D,
         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
```

tablas

```
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()
```

grado	sexo	promedio p1-p4	promedio indice p	promedio atencion	promedio edad en meses
secundaria	Masculino	114.7200	0.8028364	36.16000	157.0800
secundaria	Femenino	112.5484	0.7887666	35.03226	151.4516
primaria	Masculino	113.7013	0.7900688	34.00000	144.7662
primaria	Femenino	115.4444	0.8004050	35.55556	145.3636

suponemos que la atencion es y

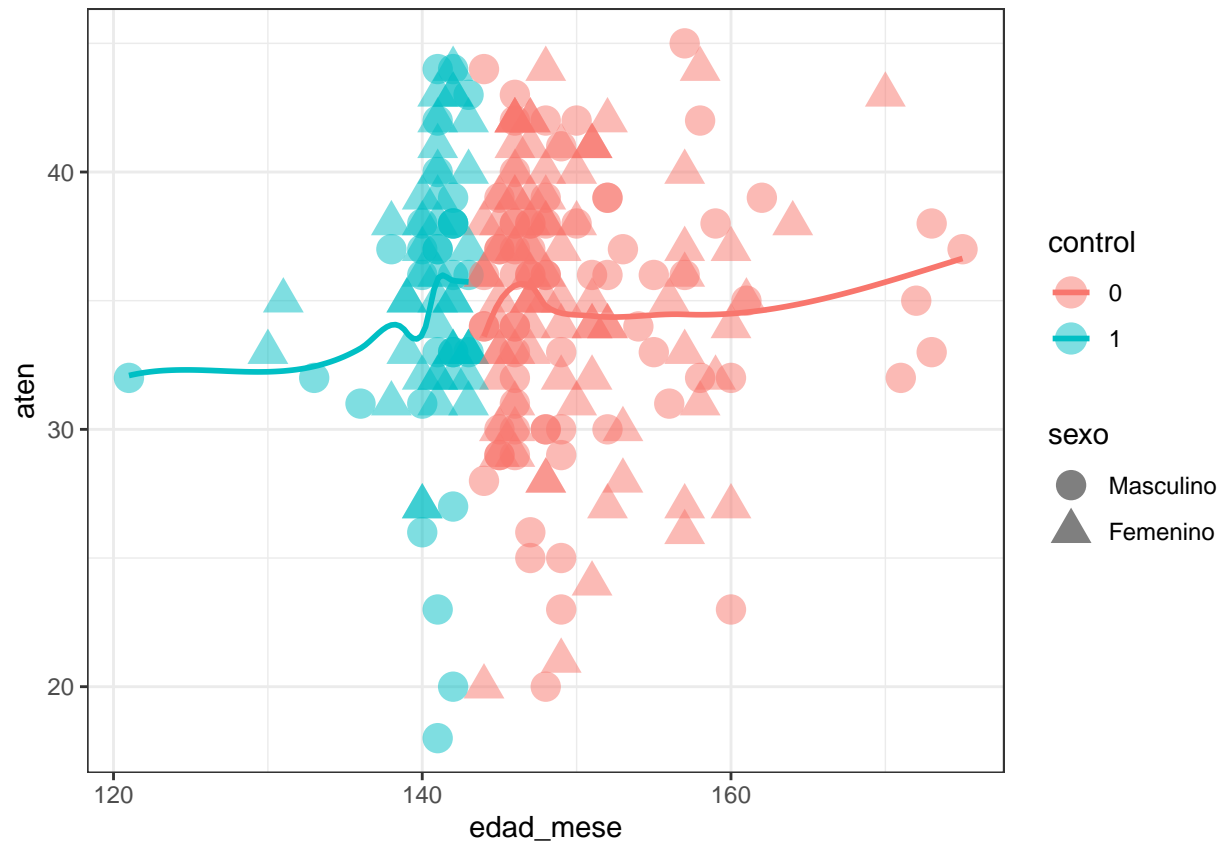
estadística descriptiva

Usar el comando gather pero con las variables relevantes seleccionadas para no perder tiempo gb, y summarise

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

```
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 = .5) + theme_bw() + geom_smooth(se = F)
```

```
## `geom_smooth()` using method = 'loess' and formula 'y ~ x'
```



regresion, prediccion

visualizacion de heterocedasticidad(aunque no vi la correccion de esta)

comparacion de modelos