

Análisis de Datos Longitudinales

Maestría en Demografía y Estudios de Población

Alejandra Marroig

Instituto de Estadística, Universidad de la República, Uruguay

August 2025

Actividad

Se quiere estudiar la trayectoria del tiempo de reacción de un conjunto de personas en un estudio de privación del sueño.

Datos

Analizamos datos sobre un estudio de sueño.

- Subject: etiqueta de cada individuo
- Reaction: tiempo promedio de reacción (en milisegundos)
- Days: días de privación de sueño
- Female: 1 si mujer; 0 si otro

Datos

```
## # A tibble: 6 x 4
##   Reaction  Days Subject Female
##   <dbl> <dbl>   <dbl>   <dbl>
## 1    250.     0    308     1
## 2    259.     1    308     1
## 3    251.     2    308     1
## 4    321.     3    308     1
## 5    357.     4    308     1
## 6    415.     5    308     1
```

- 1 Generar las siguientes variables con clase factor y llamar al nuevo objeto datos
 - a. para la variable que identifica a cada individuo.
 - b. para el sexo al nacer, incluyendo la etiqueta "Otro" y "Mujer".

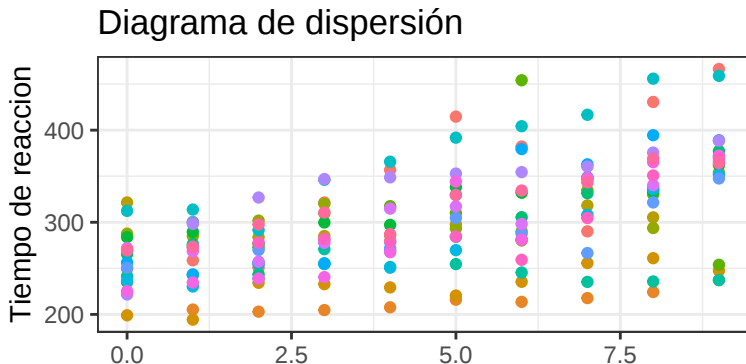
```
datos <- sleepstudymod %>%  
  mutate(Subjectf=factor(Subject),  
         Femalef=factor(Female))  
  
levels(datos$Femalef) <- c("Otro", "Mujer")
```

```
## # A tibble: 4 x 6  
##   Reaction Days Subject Female Subjectf Femalef  
##   <dbl> <dbl>   <dbl>   <dbl> <fct>    <fct>  
## 1     250.     0     308       1 308     Mujer  
## 2     259.     1     308       1 308     Mujer  
## 3     251.     2     308       1 308     Mujer  
## 4     321.     3     308       1 308     Mujer
```

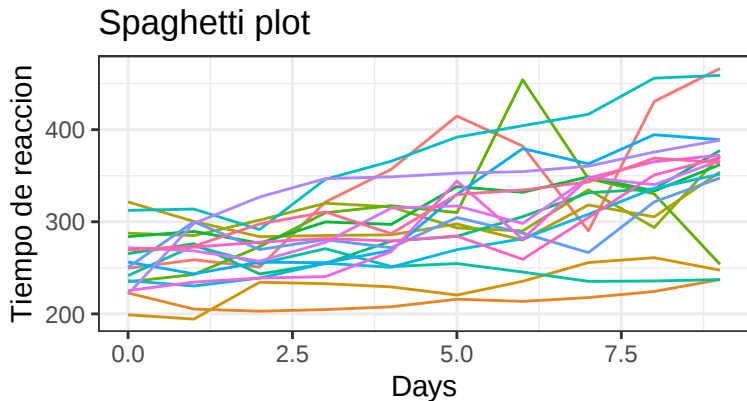
2 ¿Cuál es la métrica del tiempo en este estudio?

- 3 Realizar el diagrama de dispersión y el spaghetti plot de estos datos.

```
ggplot(data = datos) +  
  geom_point(aes(x = Days, y = Reaction, color = Subjectf)) +  
  labs(y = "Tiempo de reaccion",  
        title = "Diagrama de dispersión") +  
  theme(legend.position = "none")
```



```
ggplot(data = datos) +  
  geom_line(aes(x = Days, y = Reaction, color = Subjectf)) +  
  labs(y = "Tiempo de reaccion",  
       title = "Spaghetti plot") +  
  theme(legend.position = "none")
```

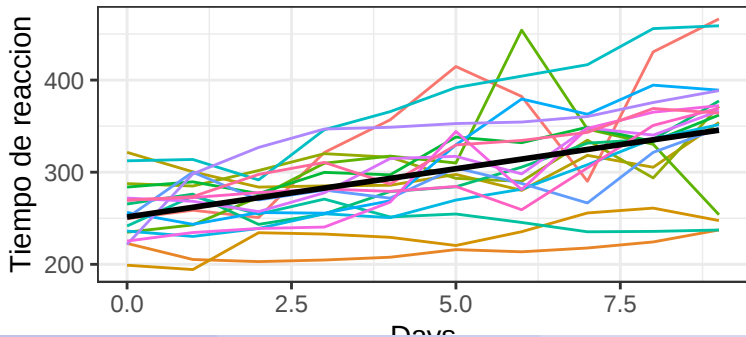


- 4 Realizar la estimación por MCO de la trayectoria promedio con fines exploratorios y agregarla al spaghetti plot.

```
##  
## Call:  
## lm(formula = Reaction ~ Days, data = datos)  
##  
## Coefficients:  
## (Intercept)          Days  
##      251.41         10.47
```

```
ggplot(data = datos) +
  geom_line(aes(x = Days, y = Reaction, color = Subjectf)) +
  geom_smooth(method='lm', se=FALSE,
             aes(x = Days, y = Reaction), color="black") +
  labs(y = "Tiempo de reaccion",
       title = "Trayectoria del tiempo de reaccion") +
  theme(legend.position = "none")
```

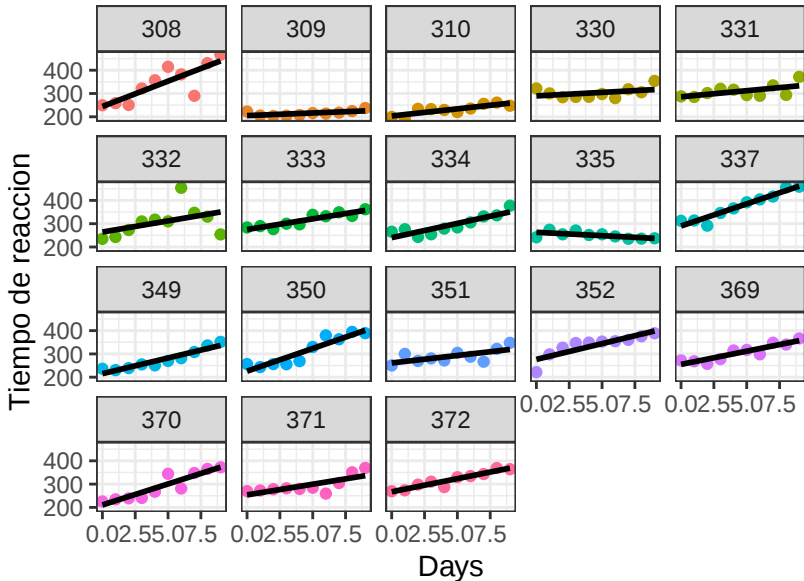
Trayectoria del tiempo de reaccion



- 5 Realizar un gráfico que permita visualizar
- a. la heterogeneidad individual en las trayectorias observadas del tiempo de reacción
 - b. la heterogeneidad individual en las trayectorias estimadas del tiempo de reacción

```
ggplot(data = datos) +  
  geom_point(aes(x = Days, y = Reaction, color = Subjectf)) +  
  geom_smooth(method='lm', se=FALSE,  
             aes(x = Days, y = Reaction), color="black") +  
  facet_wrap(~Subjectf) +  
  labs(y = "Tiempo de reaccion",  
       title = "Trayectoria del tiempo de reaccion") +  
  theme(legend.position = "none")
```

Trayectoria del tiempo de reaccion



- ⑥ Estimar las trayectorias individuales del tiempo de reacción mediante un modelo lineal.

```
indiv <- unique(datos$Subject)
beta0 <- NULL
beta1 <- NULL

for (j in 1:length(unique(datos$Subject))) {
  beta0[j] <- coef(lm(Reaction ~ Days,
                     data = datos %>%
                       filter(Subject==indiv[j])))[1]
  beta1[j] <- coef(lm(Reaction ~ Days,
                     data = datos %>%
                       filter(Subject==indiv[j])))[2]
}
```

- 7 Calcule el promedio, varianzas, desvíos y correlación de los coeficientes de las trayectorias individuales estimados. Interprete los resultados obtenidos. ¿Qué aportan las estimaciones individuales con respecto a la obtenida por MCO utilizando el conjunto de datos individuales agrupados?

```
data.frame(beta0=beta0,beta1=beta1) %>%  
  summarise(  
    int_m = mean(beta0),  
    int_d = sd(beta0),  
    pen_m = mean(beta1),  
    pen_d = sd(beta1),  
    cor    = cor(data.frame(beta0=beta0,beta1=beta1))[1,2]  
  ) %>%  
  mutate_if(is.numeric, ~ round(., 2))
```

```
##      int_m int_d pen_m pen_d   cor  
## 1 251.41 28.95 10.47  6.56 -0.14
```

- 8 Muestre una visualización exploratoria para analizar diferencias sistemáticas entre individuos en las trayectorias del tiempo de reacción según sexo.

```
ggplot(data=datos) +  
  geom_smooth(aes(x = Days, y = Reaction, color = Subjectf), method="lm") +  
  geom_smooth(aes(x = Days, y = Reaction), color="black", method="lm") +  
  facet_grid(~Femalef) +  
  labs(y = "Tiempo de reaccion") +  
  theme(legend.position = "none")
```

