

# Examen Final

Bruno C. Gonzalez

4/12/2019

## Final

### 1. Inferencia gráfica

#### 1. Preparación de los datos.

```
wages <- read_csv("data/wages.csv",
  col_types = cols(
    id = col_integer(),
    lnw = col_double(),
    exper = col_double(),
    ged = col_double(),
    postexp = col_double(),
    black = col_integer(),
    hispanic = col_integer(),
    hgc = col_integer(),
    hgc.9 = col_integer()
  ))
```

- Selecciona los sujetos con grado de estudios completado igual a 9, 10 u 11.

```
tidy_wages <- wages %>%
  filter(hgc >= 9, hgc <=11) %>%
  select(1:8)
```

- Elimina las observaciones donde el logaritmo del salario (lnw) es mayor a 3.5.

```
tidy_wages <- tidy_wages %>%
  filter(lnw <= 3.5)
```

- Crea una variable correspondiente a raza, un sujeto es de raza hispana si la variable hispanic toma el valor 1, de raza negra si la variable black toma el valor 1 y de raza blanca si las dos anteriores son cero.

```
tidy_wages <- tidy_wages %>%
  mutate(raza = ifelse(hispanic==1, 'hispana', ifelse(black==1, 'negra', 'blanca')))
```

- Crea un subconjunto de la base de datos de tal manera que tengas el mismo número de sujetos distintos en cada grupo de raza. Nota: habrá el mismo número de sujetos en cada grupo pero el número de observaciones puede diferir pues los sujetos fueron visitados un número distinto de veces.

Primero creamos una función que tomo como variable el número de sujetos que se desean incluir

```
f_wages_sample <- function(n=10){
  sample <- tidy_wages %>%
    group_by(id, raza) %>%
    summarize() %>%
    ungroup() %>%
    group_by(raza) %>%
    sample_n(n)

  tidy_wages %>%
```

```
semi_join(sample)
}
```

Posteriormente creamos el subconjunto de datos.

```
sample_wages <- f_wages_sample(80)
```

## 2 Prueba de hipótesis visual

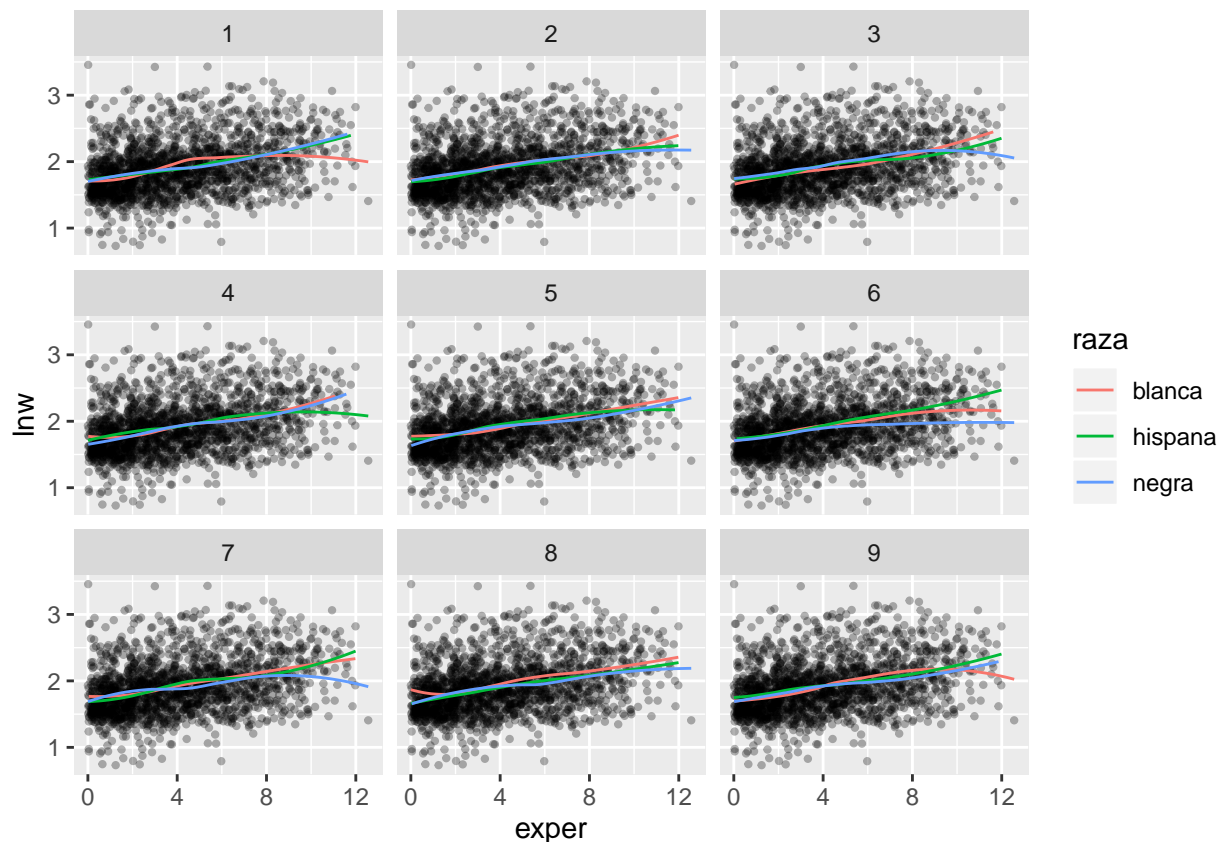
- El escenario nulo consiste en que no hay diferencia entre las razas. Para generar los datos nulos, la etiqueta de raza de cada sujeto se permuta, es decir, se reasigna la raza de cada sujeto de manera aleatoria (para todas las mediciones de un sujeto dado se reasigna una misma raza). Genera 9 conjuntos de datos nulos y para cada uno ajusta una curva *loess* siguiendo la instrucción de la gráfica de arriba. Crea una gráfica de paneles donde incluyas los 9 conjuntos nulos y los datos reales, estos últimos estarán escondidos de manera aleatoria.

Primero creamos los datos nulos

```
null_wages <- lineup(null_permute('raza'),
  n = 9,
  sample_wages)
```

A continuación hacemos las gráficas correspondientes

```
ggplot(null_wages, aes(x = exper, y = lnw)) +
  geom_point(alpha = 0.3, size = 0.8) +
  geom_smooth(aes(group = raza, color = raza), method = "loess", se = FALSE, size=0.5) +
  facet_wrap(~.sample)
```



```
theme_light()
```

- Realiza la siguiente pregunta a una o más personas **que no tomen la clase**:

*Las siguientes 10 gráficas muestran suavizamientos de  $\log(\text{salarios})$  por años de experiencia laboral. Una de ellas usa datos reales y las otras 9 son datos nulos, generados bajo el supuesto de que no existe diferencia entre los subgrupos. ¿Cuál es la gráfica más distinta?*

Reporta si las personas cuestionadas pudieron distinguir los datos.

- ¿Cuál es tu conclusión de la prueba de hipótesis visual?
- ¿A cuántas personas preguntaste y cuál es el valor p de la prueba?