

# Resultados de Percepción sobre Organismos Genéticamente Modificados

*Daniel Quiroz*

## Contents

Cargar Librerías	1
Importación de datos	1
Limpieza de datos	1
Carreras repetidas . . . . .	2
Rangos de edad . . . . .	2
Semestres . . . . .	2
Descripción del espacio muestral	2
Medios de Comunicación . . . . .	3

Este documento tiene como finalidad analizar los resultados obtenidos a partir de las encuestas realizadas en la Universidad Regional Amazónica Ikiam.

## Cargar Librerías

Como primer paso, se cargarán las librerías necesarias para realizar los análisis.

```
#### Load Packages ####  
library(readxl) # Read Excel Files  
library(tidyverse) # Easily Install and Load the 'Tidyverse'  
library(stringr) # Simple, Consistent Wrappers for Common String Operations  
library(ggalluvial)  
library(magrittr)  
library(ggsci)
```

## Importación de datos

Como primer paso se importarán los datos crudos, para esto se usará una función iterativa, map.

```
raw_data <- map(3:11, .f = function(x){  
  read_xlsx(path = 'data/Encuestas.xlsx', sheet = x) %>%  
    mutate(Sheet = x)  
}) %>% bind_rows()
```

## Limpieza de datos

Como todos los datos del mundo real, tienen errores de tabulación los cuales deben ser corregidos. En esta sección se irán recopilando los pasos para limpiar los datos.

## Carreras repetidas

En los datos crudos se encontraron errores de tipeo, los cuales llevan a condierar como diferentes carreras.

```
clean_data <- raw_data %>% mutate(P4a = tolower(P4a) ) %>%  
  # Clean Data  
  # Repeated careers  
  mutate(P4a = ifelse(P4a %in% 'ingenieria en ecosistema',  
    "ingeniería en ecosistemas", P4a)) %>%  
  mutate(P4a = ifelse(P4a %in% c("ingeniería en ciencias del agua",  
    "ingeniería en ciencias del agua"),  
    "ingeniería en ciencias del agua", P4a)) %>%  
  mutate(P4a = ifelse(grepl("ingeniería", P4a), "Ingeniería", P4a) ) %>%  
  mutate(P4a = ifelse(!grepl("Ingeniería", P4a) ),  
    "Licenciaturas", P4a) )
```

Ahora, asignaremos las etiquetas de *masculino* y *femenino* para los valores 1 y 2 tabulados

```
clean_data %<>% mutate(P2 = factor(P2, levels = c(1,2),  
  labels = c("Masculino", "Femenino")))
```

## Rangos de edad

En este paso se asignarán rangos de edad a los valores numéricos

```
clean_data %<>%  
  mutate(P3 = ifelse(P3 == 1, "17-20", P3)) %>%  
  mutate(P3 = ifelse(P3 == 2, "21-25", P3)) %>%  
  mutate(P3 = ifelse(P3 == 3, "26-30", P3)) %>%  
  mutate(P3 = ifelse(P3 == 4, "31-40", P3)) %>%  
  mutate(P3 = ifelse(P3 == 5, "41-50", P3)) %>%  
  mutate(P3 = ifelse(P3 == 1, "> 51", P3))
```

## Semestres

Se diferenciará mayormente entre tronco común y semestres de carrera

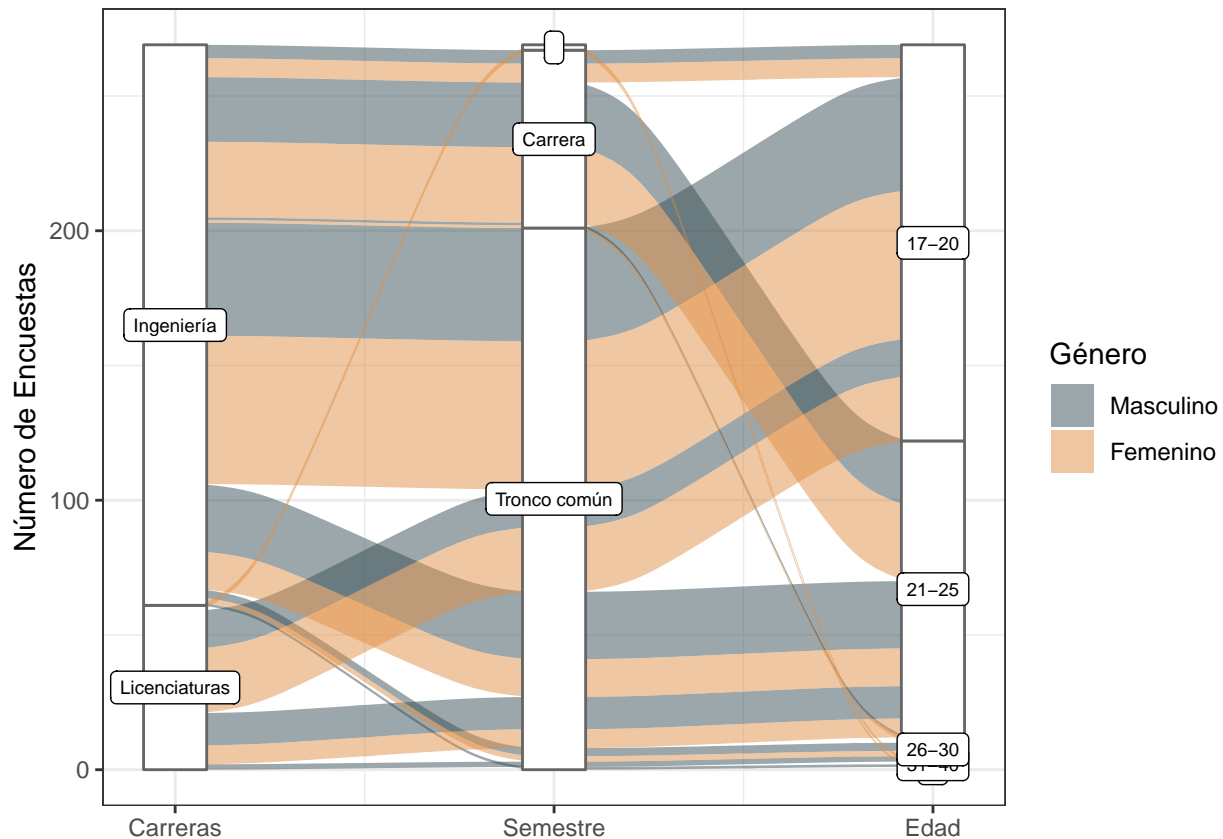
```
clean_data %<>%  
  mutate(`5` = ifelse(`5` > 4, "Carrera", "Tronco común"))
```

## Descripción del espacio muestral

Como primera instancia, debemos la composición del espacio muestral. Para esto, emplearemos gráficos descriptivos.

```
# Summarising data  
sample_space <- clean_data %>% group_by(P2, P3, P4a, `5`) %>%  
  summarise(N = n()) %>% arrange(P4a)  
sspace_plot <- ggplot(sample_space, aes(axis1 = P4a, axis2 = `5`, axis3 = P3, y = N)) +  
  geom_alluvium(aes(fill = P2), width = 1/8) +  
  geom_stratum(width = 1/6, fill = "white", color = "grey40") +  
  geom_label(stat = "stratum", label.strata = TRUE, size = 2.5) +  
  scale_x_continuous(breaks = 1:3, labels = c("Carreras", "Semestre", "Edad")) +
```

```
scale_fill_jama() + theme_bw() +
labs(fill = "Género", y = "Número de Encuestas")
sspace_plot
```



## Medios de Comunicación

En este apartado, se analizará el número de medios de comunicación que emplea cada encuestado para informarse.

```
medios <- clean_data %>% select(`6a`:`6g`) %>% mutate(Index = 1:n()) %>%
gather(key = "Medio", value = "Valor", `6a`:`6g`) %>%
mutate(Medio = ifelse(Medio %in% "6a", "Radio AM/FM", Medio)) %>%
mutate(Medio = ifelse(Medio %in% "6b", "Televisión abierta", Medio)) %>%
mutate(Medio = ifelse(Medio %in% "6c", "Televisión por cable", Medio)) %>%
mutate(Medio = ifelse(Medio %in% "6d", "Periódicos/Revistas", Medio)) %>%
mutate(Medio = ifelse(Medio %in% "6e", "Internet/Redes Sociales", Medio)) %>%
mutate(Medio = ifelse(Medio %in% "6f", "Revistas científicas", Medio)) %>%
mutate(Medio = ifelse(Medio %in% "6g", "Otros", Medio)) %>%
filter(Valor <= 1)
```

Una vez tenemos asignado las etiquetas asignadas a cada código, procedemos a obtener **la probabilidad de que determinado medio sea empleado, cuando son empleados  $n$  medios de comunicación.**

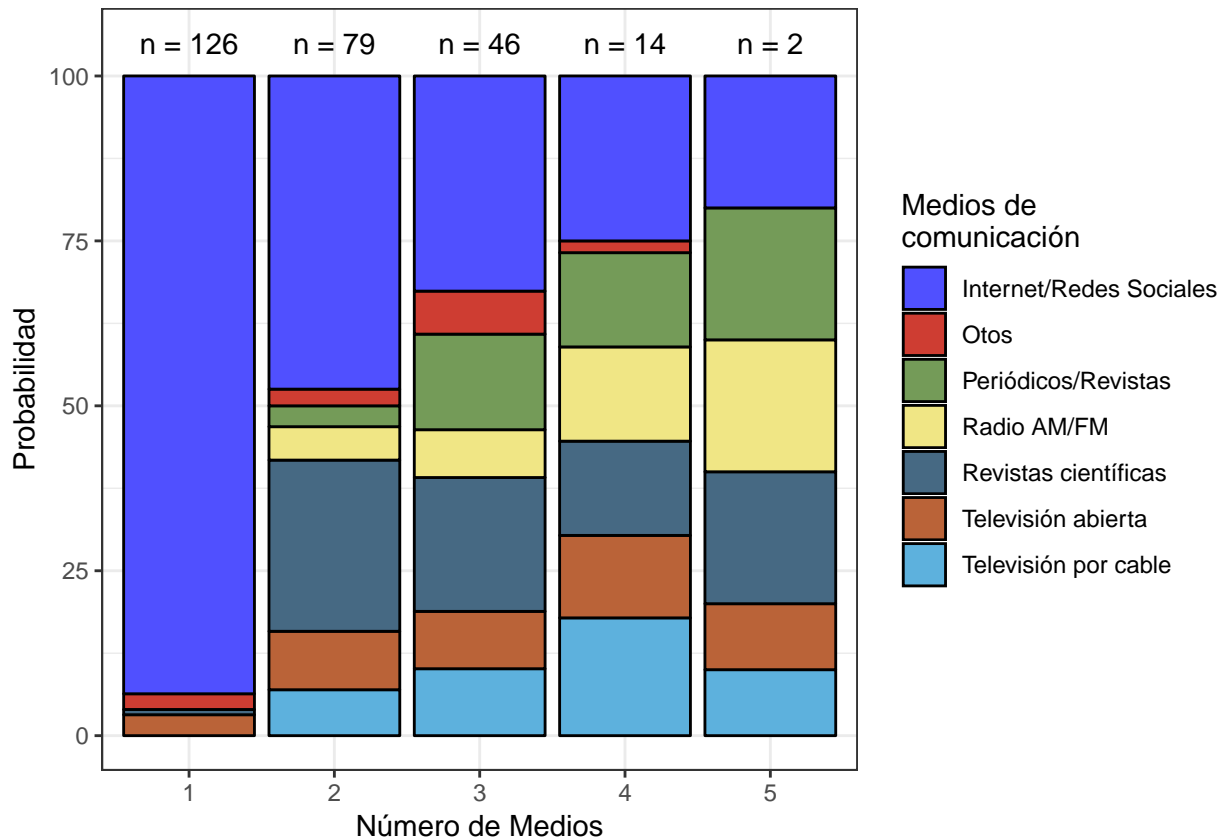
```
labels <- medios %>% filter(Valor != 0) %>%
arrange(Index) %>%
group_by(Index) %>% mutate(total = sum(Valor)) %>%
count(total) %>% ungroup() %>% count(n) %>%
```

```

mutate(nn = paste0("n = ", nn))

medios %>% filter(Valor != 0) %>% arrange(Index) %>%
  group_by(Index) %>% mutate(total = sum(Valor)) %>%
  select(-Valor) %>%
  ggplot() +
  geom_bar(aes(factor(total), fill = Medio),
    position = "fill", color = "black") +
  geom_text(data = labels, aes(x= factor(n), y = 1.05, label = nn) ) +
  ggsci::scale_fill_igv() + theme_bw() +
  scale_y_continuous(breaks = seq(0, 1, by = 0.25), labels = seq(0, 100, by = 25)) +
  scale_x_discrete(limits = 1:5) +
  labs(x = "Número de Medios", y = "Probabilidad", fill = "Medios de\ncomunicación")

```



Por ejemplo, de los 126 encuestados que solo emplean un medio de comunicación, al rededor del 90% emplea solo Internet y/o redes sociales. Por otro lado, cuando emplean dos medios de comunicación, el uso de internet decrece a un 50% y las revistas científicas tienen un 20% de probabilidad de ser empleadas como medio de comunicación.