



Guía de uso de datos abiertos DEMRE

Departamento de Evaluación, Medición y Registro Educativo
Vicerrectoría de Asuntos Académicos

Noviembre de 2024

Guía de uso de datos DEMRE

1. Consideraciones iniciales sobre el uso de datos DEMRE

Desde 2024, el DEMRE ha publicado en su sitio web las bases de datos de los distintos procesos de admisión universitaria, disponibles desde el año 2004 en adelante.

Aunque estos datos ya podían solicitarse previamente, la creación de este portal de descarga representa un avance significativo en la política de transparencia activa del Departamento. Esta iniciativa facilita el acceso a la información generada anualmente, con el propósito de impulsar su utilización y estimular una mayor producción científica que contribuya a mejorar el sistema de acceso universitario.

En línea con este objetivo, el presente documento busca establecer un nuevo precedente en la democratización del acceso a los datos del Sistema de Admisión. Proporciona recomendaciones y ejemplos prácticos que ayudarán a las y los usuarios a comprender mejor estos recursos y, en consecuencia, realizar sus propios análisis conforme a los objetivos de sus investigaciones.

Es fundamental resaltar que la PAES es un instrumento diseñado para evaluar las competencias necesarias para la educación universitaria en aquellos individuos/as que se inscriben en el sistema centralizado de admisión, que no corresponden a la totalidad de las y los estudiantes que egresan de Enseñanza Media en Chile. Por lo tanto, los resultados de la PAES no informan sobre la calidad de la educación secundaria, ni a nivel nacional ni a nivel de instituciones educativas individuales.

Los Estándares para Pruebas Educativas y Psicológicas (AERA, APA & NCME, 2014) establecen que cada interpretación y uso de los resultados de un instrumento de medición debe estar debidamente justificado mediante un proceso de validación. En el caso de las PAES, el DEMRE realiza este proceso de validación específicamente para los fines con los que la prueba ha sido desarrollada, no para otras interpretaciones.

Por lo tanto, se desaconseja el uso de los datos proporcionados por el DEMRE para realizar interpretaciones fuera del contexto de la selección universitaria.

Por último, según la legislación vigente, está prohibido realizar cualquier acción que busque individualizar o identificar a las y los postulantes cuyos datos se encuentran en estas bases de datos.

Si necesita datos identificados mediante MRUN para vincularlos con otras fuentes de información de organismos públicos, puede solicitar esto por correo a solicituddatos@demre.cl. Alternativamente, puede acceder al portal de Datos Abiertos del Centro de Estudios del Ministerio de Educación.

2. Datos disponibles del proceso de admisión

El DEMRE proporciona diversas bases de datos que abarcan las distintas etapas del proceso de admisión universitaria:

- **Base de Inscritos/as (Archivo B):** Contiene información demográfica y educativa de todos los y las postulantes inscritos para rendir las pruebas PAES.
- **Base de Rendición (Archivo C):** Incluye los puntajes obtenidos por los y las postulantes en las diferentes pruebas PAES.
- **Base de Postulación (Archivo D):** Contiene las preferencias de carrera y universidad de los y las postulantes, junto con sus puntajes ponderados para cada opción.
- **Base de Matrícula:** Incluye información sobre los y las postulantes matriculados en cada carrera y universidad.

Es crucial destacar que estas bases de datos contienen un identificador único (ID_aux) para cada postulante, lo que permite unificar la información de las diferentes etapas del proceso de admisión.

Además, las bases de datos de puntajes incluyen información de los últimos dos procesos de admisión, abarcando tanto las aplicaciones regulares como las de invierno, lo que facilita análisis comparativos entre diferentes períodos.

El acceso a estas bases de datos está sujeto a estrictos protocolos de seguridad y confidencialidad. Su uso debe cumplir con la legislación vigente sobre protección de la privacidad de los postulantes, prohibiendo su identificación individual. Para garantizar esta protección, se han eliminado ciertos registros que podrían permitir la identificación de personas específicas. Consecuentemente, las cifras derivadas de estos datos podrían no coincidir exactamente con las estadísticas oficiales publicadas. Esta diferencia tiene un único propósito: salvaguardar la identidad de los y las postulantes, manteniendo la integridad general de los datos para fines de análisis e investigación.

3. Guía Práctica de Análisis de Datos

Las bases de datos, disponibles en formato .csv, pueden analizarse con cualquier software estadístico. En esta nota, presentaremos ejemplos utilizando el software estadístico R (R Core Team, 2020).

Específicamente, emplearemos el conjunto de herramientas tidyverse (Wickham & Grolemund, 2016), con énfasis en dos paquetes clave: dplyr (Wickham et al., 2023) — que facilita operaciones como la selección, filtrado y resumen de datos— y ggplot2 (Wickham et al., 2020), que permite la visualización de datos de manera efectiva.

3.1 Importación de base de datos

El entorno R ofrece funciones predeterminadas para la lectura de datos. Los archivos proporcionados por el DEMRE vienen en formato .csv y pueden importarse mediante el comando `read.csv2()`.

A continuación, se muestra un ejemplo de cómo leer los datos de las personas inscritas en la PAES para la Admisión 2024.

Descargue los archivos de la página web del DEMRE, en la sección de [Portal Base de Datos Abiertos](#). Como ejemplo usaremos los archivos:

- [Inscritos de Admisión 2024](#).
- [Rendición de Admisión 2024](#).

Las siguientes líneas de código, abrirá los archivos en el entorno R.

```
# Abrir datos con .csv que están guardados en su disco duro, en la ruta "C:/".  
# En caso de descargar los archivos en otra ruta, esta debe ser a  
Ruta <- "C:/ "  
inscritos <- read.csv2(paste0(Ruta, "ArchivoB_Adm2024.csv"))  
rinden <- read.csv2 (paste0(Ruta, "ArchivoC_Adm2024.csv"))  
  
## Inspeccionar las dimensiones de los archivos (filas x columnas)  
dim(inscritos)  
  
## [1] 306497      20  
  
dim(rinden)  
  
## [1] 306497      35
```

Con lo anterior podemos empezar a trabajar y realizar diferentes análisis sobre las distintas bases de datos disponibles.

3.2 Número de postulantes que rindieron PAES-Regular Admisión 2024 agrupados por Sexo registral

A continuación, presentamos un ejemplo que ilustra cómo calcular la cantidad de postulantes que rindieron la prueba, agrupados según sexo registral.

Las siguientes líneas de código unen los archivos de inscritos y rendición utilizando la variable `ID_aux`, que funciona como identificador único del postulante. Las bases a utilizar son:

- Inscritos de Admisión 2024
- Rendición de Admisión 2024.

Es importante mencionar que, dado que los puntajes PAES permanecen vigentes durante dos procesos de admisión consecutivos, las bases de datos de puntajes (rendición), contienen información de todas las instancias de aplicación del proceso de admisión actual y el inmediatamente anterior. Los puntajes se pueden encontrar según la siguiente nomenclatura:

- **[PRUEBA]_REG_ACTUAL:** Puntaje obtenido en la prueba para la aplicación Regular del Proceso de Admisión actual.
- **[PRUEBA]_INV_ACTUAL:** Puntaje obtenido en la prueba para la aplicación de Invierno del Proceso de Admisión actual.
- **[PRUEBA]_REG_ANTERIOR:** Puntaje obtenido en la prueba para la aplicación Regular del Proceso de Admisión inmediatamente anterior.
- **[PRUEBA]_INV_ANTERIOR:** Puntaje obtenido en la prueba para la aplicación de Invierno del Proceso de Admisión inmediatamente anterior.

Para efectos de análisis, se define que una persona rinde en el proceso regular si tiene un puntaje de entre 100 y 1.000 puntos en al menos una de las pruebas. Si un o una postulante no tiene puntaje en alguna prueba—ya sea por inasistencia o por no haberla inscrito— éste se registra como 0 en las bases de datos. Esto permite identificar a los y las postulantes que efectivamente participaron en el proceso de aplicación PAES, independientemente de si rindieron todas las pruebas o solo algunas de ellas.

```
# Abrir datos con .csv
inscritos <- read.csv2(paste0(Ruta,"ArchivoB_Adm2024.csv")) %>%
  dplyr::select(ID_aux,SEXO)

rinden <- read.csv2 (paste0(Ruta,"ArchivoC_Adm2024.csv")) %>%
  dplyr::mutate(Flag_Rinden=if_else(CLEC_REG_ACTUAL>0|MATE1_REG_ACTUAL>0|
MATE2_REG_ACTUAL>0|HCSOC_REG_ACTUAL>0|CIEN_REG_ACTUAL>0,1,0))

datos_rinden_sexo = rinden %>%
  dplyr::left_join(inscritos, by = c("ID_aux"))

# Descriptivos
datos_rinden_sexo %>%
  dplyr::group_by(SEXO) %>%
  dplyr::summarise(Inscritos = n(),
                  Rinden = sum(Flag_Rinden,na.rm=TRUE),
                  Porcentaje_Rinden=round((Rinden/Inscritos)*100,1)) %>%
  #dplyr::filter(REGIMEN != "") %>%
  knitr::kable(format.args = list(big.mark = ".",dec=","))
```

SEXO	Inscritos	Rinden	Porcentaje_Rinden
1	142.150	114.863	80,8

SEXO	Inscritos	Rinden	Porcentaje_Rinden
2	164.347	135.204	82,3

De lo anterior, se observa que el 82,3% de las mujeres inscritas rindieron al menos una prueba en la rendición regular en el proceso de admisión 2024, mientras que en el caso de los hombres fue un 80,8%.

3.3 Cambio de puntajes de PAES de Competencia Matemática 1 Admisión 2023 versus 2024

A continuación, veremos un ejemplo que nos muestra de manera gráfica el cambio de la distribución de puntajes entre los postulantes que rindieron la prueba de Competencia Matemática 1 en la Admisión 2023 y 2024.

Las bases a utilizar son:

- [Rendición de Admisión 2024.](#)
- [Rendición de Admisión 2023.](#)

```
M1_2024 = read.csv2(paste0(Ruta, "ArchivoC_Adm2024.csv")) %>%
  dplyr::filter(MATE1_REG_ACTUAL > 0) %>%
  dplyr::mutate(Distribucion = case_when(is.na(MATE1_REG_ACTUAL) ~ "99",
    MATE1_REG_ACTUAL < 300 ~ "MENOS DE 300",
    MATE1_REG_ACTUAL < 400 ~ "300-399",
    MATE1_REG_ACTUAL < 500 ~ "400-499",
    MATE1_REG_ACTUAL < 600 ~ "500-599",
    MATE1_REG_ACTUAL < 700 ~ "600-699",
    MATE1_REG_ACTUAL < 800 ~ "700-799",
    MATE1_REG_ACTUAL < 900 ~ "800-899",
    900 <= MATE1_REG_ACTUAL ~ "900 o MAS"))
%>%
  dplyr::group_by(Distribucion) %>%
  dplyr::summarise(n = n()) %>%
  dplyr::mutate(Porcentaje = round((n / sum(n)) * 100, 3)) %>%
  dplyr::rename("M1_2024" = "Porcentaje")

M1_2023 = read.csv2(paste0(Ruta, "ArchivoC_Adm2023.csv")) %>%
  dplyr::filter(MATE1_REG_ACTUAL > 0) %>%
  dplyr::mutate(Distribucion = case_when(is.na(MATE1_REG_ACTUAL) ~ "99",
    MATE1_REG_ACTUAL < 300 ~ "MENOS DE 300",
    MATE1_REG_ACTUAL < 400 ~ "300-399",
    MATE1_REG_ACTUAL < 500 ~ "400-499",
    MATE1_REG_ACTUAL < 600 ~ "500-599",
    MATE1_REG_ACTUAL < 700 ~ "600-699",
    MATE1_REG_ACTUAL < 800 ~ "700-799",
    MATE1_REG_ACTUAL < 900 ~ "800-899",
    900 <= MATE1_REG_ACTUAL ~ "900 o MAS"))
%>%
```

```

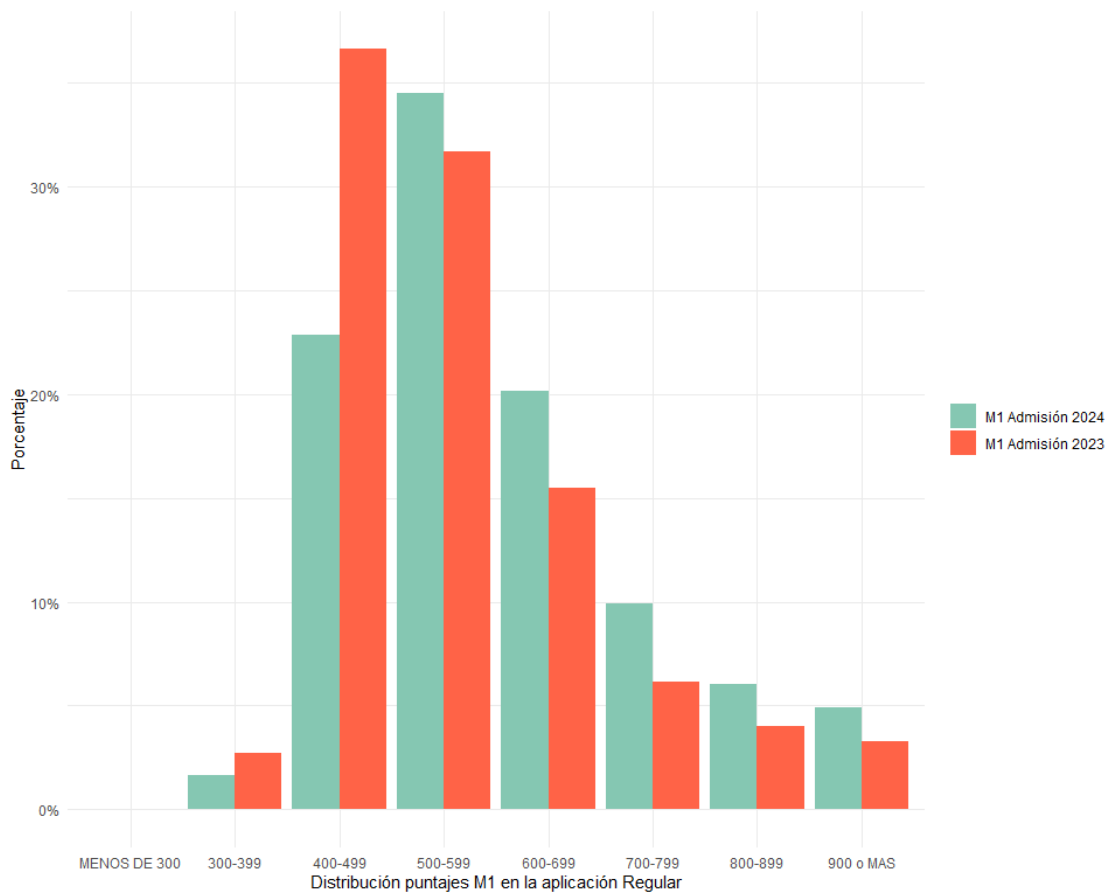
dplyr::group_by(Distribucion) %>%
dplyr::summarise(n = n()) %>%
dplyr::mutate(Porcentaje = round((n / sum(n)) * 100, 3)) %>%
dplyr::rename("M1_2023" = "Porcentaje")

comportamiento = M1_2024 %>%
dplyr::left_join(M1_2023,by = "Distribucion") %>%
dplyr::select(-n.x,-n.y)

data_long <- comportamiento %>%
tidyr::pivot_longer(cols = c(M1_2024, M1_2023), names_to = "Prueba", va
lues_to = "Porcentaje") %>%
dplyr::mutate(Prueba = factor(Prueba, levels = c("M1_2024", "M1_2023"))
,
      Distribucion = factor(Distribucion,levels = c("99",
      "MENOS DE 300",
      "300-399",
      "400-499",
      "500-599",
      "600-699",
      "700-799",
      "800-899",
      "900 o MAS")))

ggplot(data_long, aes(x = factor(Distribucion), y = Porcentaje, fill = Pr
ueba)) +
geom_bar(stat = "identity", position = "dodge") +
labs(x = "Distribución puntajes M1 en la aplicación Regular",
      y = "Porcentaje")+
scale_fill_manual(values = c("M1_2024" = "#85C7B2", "M1_2023" = "#FF634
7"),
      labels = c("M1_2024" = "M1 Admisión 2024", "M1_2023"
= "M1 Admisión 2023"),
      name = NULL) +
scale_y_continuous(labels = scales::percent_format(scale = 1)) +
theme_minimal()

```



3.4 Análisis descriptivo de puntajes de la PAES de Competencia Matemática 1

Presentaremos un ejemplo gráfico que ilustra la distribución total y por dependencia de los puntajes obtenidos por los y las postulantes en la prueba de Competencia Matemática 1 durante la aplicación Regular del Proceso de Admisión 2024. La base a utilizar es:

- [Rendición de Admisión 2024.](#)

```
punt_mate = read.csv2 (paste0(Ruta,"ArchivoC_Adm2024.csv")) %>%
  dplyr::filter(MATE1_REG_ACTUAL>0) %>%
  dplyr::select(MATE1_REG_ACTUAL,GRUPO_DEPENDENCIA) %>%
  dplyr::mutate(Prueba = "MATE",
               Dependencia = case_when(GRUPO_DEPENDENCIA == 3 | GRUPO_DEPENDENCIA == 4 ~ "Mun.",
                                       GRUPO_DEPENDENCIA == 2 ~ "PS",
                                       GRUPO_DEPENDENCIA == 1 ~ "PP",
                                       TRUE ~ NA_character_),
               Dependencia = factor(Dependencia, levels = c("PP","PS","Mun.")))

media_puntajes <- punt_mate %>%
  dplyr::group_by(Prueba) %>%
```



```

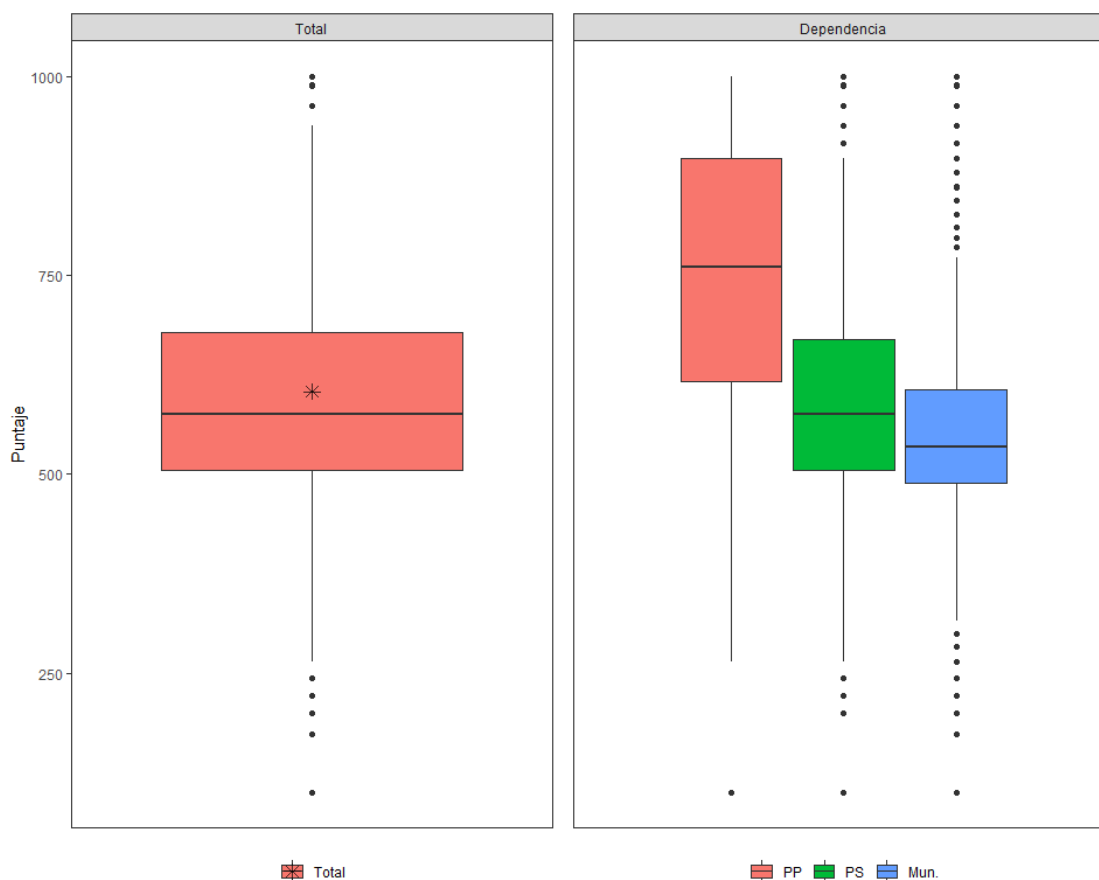
dplyr::summarise(Media = mean(MATE1_REG_ACTUAL))

punt_Mate1_tot = ggplot(punt_mate ,
                        aes(x = Prueba, y = MATE1_REG_ACTUAL, fill = Prueb
a)) +
  geom_boxplot() +
  facet_grid(~factor(Prueba), scales = "free_x", labeller = as_labeller(c("
MATE" = "Total")))+
  geom_point(data = media_puntajes, aes(x = Prueba, y = Media),
            color = "black", size = 3, shape = 8, alpha = 0.7) +
  theme_bw() +
  labs(x = NULL,
       y = "Puntaje")+
  scale_fill_discrete(labels=c('Total'))+
  theme(legend.position = "bottom", legend.direction = "horizontal",
        legend.title=element_blank(), panel.grid = element_blank(),
        axis.title.x = element_blank(),
        axis.text.x = element_blank(),
        axis.ticks.x = element_blank())

Mate1_dep = ggplot(punt_mate %>%
                  dplyr::filter(!is.na(Dependencia)),
                  aes(x = Prueba, y = MATE1_REG_ACTUAL, fill = Dependenci
a)) +
  geom_boxplot() +
  scale_x_discrete(labels = NULL) +
  facet_grid(~factor(Prueba), scales = "free_x", labeller = as_labeller(c("
MATE" = "Dependencia")))+
  theme_bw() +
  theme(legend.position = "bottom", legend.title=element_blank(), panel.gri
d = element_blank(),
        axis.text.x = element_blank(),
        axis.ticks.x = element_blank(),
        axis.text.y=element_blank(),
        axis.ticks.y=element_blank()) +
  labs(x = NULL, y = NULL)

cuadrricula_con_leyenda <- gridExtra::grid.arrange(punt_Mate1_tot, Mate1_de
p, nrow=1)

```



Los resultados de la prueba de Competencia Matemática 1 muestran algunas diferencias entre estudiantes de distintos tipos de colegios. En distribución, los estudiantes de colegios municipales obtienen puntajes más bajos que los de otros tipos de establecimientos. Sin embargo, es importante recordar que estas diferencias no significan necesariamente que la prueba no sea ecuánime. Para saber si la prueba “trata a todos por igual”, se necesitarían estudios más detallados que analicen cómo funciona la prueba para diferentes grupos de estudiantes.

3.5 Análisis de nacionalidades

A continuación, veremos un ejemplo que muestra los 10 países extranjeros con mayor representación de postulantes que rindieron la prueba y que están matriculados, y el porcentaje de matriculados. Las bases a utilizar son:

- [Inscripción de Admisión 2024.](#)
- [Rendición de Admisión 2024.](#)
- [Matrícula de Admisión 2024.](#)

Es importante destacar que, en este ejemplo, para obtener los nombres de los países se debe cruzar la información con el libro de códigos del archivo de Inscritos correspondiente. Este archivo está en formato .xlsx y se debe consultar específicamente la hoja llamada ‘Anexo - País’.

```

rinden_puntaje = read.csv2 (paste0(Ruta, "ArchivoC_Adm2024.csv")) %>%
  dplyr::filter(CLEC_REG_ACTUAL>0|MATE1_REG_ACTUAL>0|MATE2_REG_ACTUAL>0|H
CSOC_REG_ACTUAL>0|
  CIEN_REG_ACTUAL>0)

inscritos_extranjeros = read.csv2 (paste0(Ruta, "ArchivoB_Adm2024.csv"))
%>%
  dplyr::filter(PAIS_NACIMIENTO != 210) %>%
  dplyr::inner_join(rinden_puntaje, by = c("ID_aux" = "ID_aux"))

inscritos = read.csv2("ArchivoB_Adm2024.csv")

matriculados_extranjeros = read.csv2 (paste0(Ruta, "ArchivoMatr_Adm2024.c
sv")) %>%
  dplyr::inner_join(inscritos, by = c("ID_aux" = "ID_aux")) %>%
  dplyr::filter(PAIS_NACIMIENTO != 210) %>%
  dplyr::filter(ID_aux %in% rinden_puntaje$ID_aux)

nacionalidad = read_xlsx(paste0(Ruta,"Libro_CódigosADM2024_ArchivoB.xlsx"
),sheet = "Anexo - País")

inscritos_extranjeros %>%
  dplyr::group_by(PAIS_NACIMIENTO) %>%
  dplyr::summarise(inscritos_extranjeros = n()) %>%
  dplyr::left_join(matriculados_extranjeros %>%
    dplyr::group_by(PAIS_NACIMIENTO) %>%
    dplyr::summarise(matriculados_extranjeros = n()),
    by = c("PAIS_NACIMIENTO" = "PAIS_NACIMIENTO")) %>%
  dplyr::mutate(Porc_matriculados = round(matriculados_extranjeros/inscri
tos_extranjeros,3)*100) %>%
  dplyr::left_join(nacionalidad,by = c("PAIS_NACIMIENTO" = "Código")) %>%
  dplyr::arrange(desc(inscritos_extranjeros)) %>%
  dplyr::select(País,inscritos_extranjeros, matriculados_extranjeros, Por
c_matriculados) %>%
  head(n = 10)%>%
  knitr::kable(format.args = list(big.mark = "."))

```

País	inscritos_extranjeros	matriculados_extranjeros	Porc_matriculados
Venezuela	3.285	1.008	30.7
Perú	2.122	752	35.4
Colombia	1.473	428	29.1
Bolivia	1.128	285	25.3
Haití	460	63	13.7
Ecuador	392	106	27.0
Argentina	278	94	33.8
Dominicana	111	24	21.6

País	inscritos_extranjeros	matriculados_extranjeros	Porc_matriculados
Brasil	64	27	42.2
España	55	28	50.9

3.6 Universidades con mayores postulaciones

Aquí presentamos un ejemplo que muestra las 10 universidades con mayor cantidad de postulaciones activas. La base de datos que utilizaremos es:

- [Postulación de Admisión 2024.](#)

Para obtener los nombres de las universidades, es necesario cruzar la información con el libro de códigos del archivo de postulaciones correspondiente. Este archivo está en formato .xlsx y se debe consultar específicamente la hoja llamada "Anexo - Oferta académica".

```
# Oferta académica
oferta = read_xlsx(paste0(Ruta, "Libro_CódigosADM2024_ArchivoD.xlsx"), sheet = "Anexo - Oferta académica") %>%
  dplyr::select(CODIGO_CARRERA, NOMBRE_UNIVERSIDAD)

postulaciones = read.csv2 (paste0(Ruta, "ArchivoD_Adm2024.csv")) %>%
  dplyr::left_join(oferta, by = c("COD_CARRERA_PREF" = "CODIGO_CARRERA"))
%>%
  dplyr::group_by(NOMBRE_UNIVERSIDAD) %>%
  dplyr::summarise(Cantidad = n()) %>%
  dplyr::arrange(desc(Cantidad)) %>%
  head(10)

postulaciones %>%
  knitr::kable(format.args = list(big.mark = "."))
```

NOMBRE_UNIVERSIDAD	Cantidad
UNIVERSIDAD DE CHILE	110.522
PONTIFICIA UNIVERSIDAD CATOLICA DE CHILE	108.472
UNIVERSIDAD ANDRES BELLO	103.478
UNIVERSIDAD DE CONCEPCION	93.346
UNIVERSIDAD AUTONOMA DE CHILE	92.353
UNIVERSIDAD SANTO TOMAS	82.314
UNIVERSIDAD SAN SEBASTIAN	81.187
UNIVERSIDAD DE SANTIAGO DE CHILE	78.166
UNIVERSIDAD MAYOR	64.166
UNIVERSIDAD DIEGO PORTALES	55.294



Av. José Pedro Alessandri 685, Ñuñoa - Santiago, Chile

 demre.cl  [/demre_uchile](https://twitter.com/demre_uchile)

 [/demre.uchile](https://facebook.com/demre.uchile)  [/DEMREuchile](https://youtube.com/DEMREuchile)

 [DEMRE U.chile](https://whatsapp.com/DEMREUchile)  [/demre.uchile](https://instagram.com/demre.uchile)

 [DEMRE Uchile](https://tiktok.com/DEMREUchile)  [@demre.uchile](https://tiktok.com/@demre.uchile)