

Impacto del Cambio Metodológico Post-COVID en la Inserción Laboral

Carlos Pérez, Daniel Limón, Joaquín Vidal

2025-12-04

Table of contents

1	Resumen	1
2	Introducción	1
3	Metodología: Construcción de la Base de Datos	2
3.1	Fuente de Datos	2
3.2	Procesamiento y Harmonización	2
4	Resultados Descriptivos	5
4.1	Caracterización de la Muestra	6
4.2	Análisis Exploratorio Visual:	6
5	Análisis Inferencial	6
5.1	Especificación del Modelo	6
5.2	Resultados del Modelo	6
6	Discusión y Conclusiones	6
6.1	Interpretación	6
6.2	Matices por rama	6
6.3	Limitaciones	6

1 Resumen

(Pendiente de redacción final) Este estudio analiza la inserción laboral de XX,XXX egresados de universidades andaluzas, comparando la cohorte pre-pandemia (2018-2019) con las cohortes afectadas por la adaptación metodológica COVID-19.

2 Introducción

- Contexto: Breve descripción del cambio metodológico en la universidad andaluza durante la pandemia.
- Hipótesis: El cambio forzoso a la docencia online/híbrida ha impactado negativamente en la calidad de la inserción laboral (mayor precariedad) en las cohortes afectadas.

3 Metodología: Construcción de la Base de Datos

3.1 Fuente de Datos

Los datos primarios proceden de la *Encuesta de Inserción Laboral de los Egresados de Universidades Públicas de Andalucía*, facilitados por el Instituto de Estadística y Cartografía de Andalucía (IECA). El estudio abarca un periodo longitudinal de cinco cohortes académicas, desde el curso 2018-2019 hasta el 2022-2023.

3.2 Procesamiento y Harmonización

Dado que la estructura de los microdatos originales presentaba variaciones anuales en el esquema de columnas y separadores de archivo, se implementó un algoritmo de armonización para consolidar los ficheros anuales en un único dataframe maestro.

El proceso de limpieza y filtrado siguió las siguientes reglas de negocio estrictas:

1. **Universo de Estudio:** Se seleccionaron exclusivamente los egresados de Grado (`tipo_estudio = 1`), excluyendo másteres y doctorados para garantizar la homogeneidad de la muestra.
2. **Tratamiento de Valores Perdidos:**
 - Se estandarizaron los códigos de “No consta” o “No aplica” (99, Q0) como valores perdidos (NA).
 - Se aplicó un criterio de *casos completos* para las variables críticas del estudio: Cohorte, Rendimiento Académico, Rama de Enseñanza y Situación Laboral al primer año.
3. **Definición de Variables Derivadas:**
 - **Rama de Enseñanza (GRUPO):** Se reclasificó la variable `AMBITO` original (códigos de 4 dígitos) en 10 grandes grupos funcionales.
 - **Calidad de la Inserción:** Se generó una variable binaria `CalidadContrato` basada en la jornada laboral al primer año (`TJORNADAA1T3`), donde la jornada parcial se considera un indicador de precariedad (1) frente a la jornada completa (0).

```
files <- c(
  "./microdatos/MicrodatILEUPA2018.csv",
  "./microdatos/MicrodatILEUPA2019.csv",
  "./microdatos/MicrodatILEUPA2020.csv",
  "./microdatos/MicrodatILEUPA2021.csv",
  "./microdatos/MicrodatILEUPA2022.csv"
)

# --- Lector que autodetecta separador ("," o ";") ---
read_csv_auto <- function(path) {
  if (!is.character(path) || length(path) != 1) stop("Ruta no válida.")
  if (!file.exists(path)) stop(paste("No existe el archivo:", path))
  first_line <- readLines(path, n = 1, warn = FALSE)
  if (grepl(";", first_line)) {
    df <- read.csv2(path, stringsAsFactors = FALSE, check.names = FALSE)
  } else {
    df <- read.csv(path, stringsAsFactors = FALSE, check.names = FALSE)
  }
}
```

```

}
df
}

# --- Unir data.frames con distinto set de columnas (rellena faltantes con NA) ---
rbind_union <- function(dflist) {
  all_names <- unique(unlist(lapply(dflist, names)))
  aligned <- lapply(dflist, function(x) {
    miss <- setdiff(all_names, names(x))
    if (length(miss) > 0) {
      for (m in miss) x[[m]] <- NA
    }
    x[, all_names, drop = FALSE]
  })
  do.call(rbind, aligned)
}

# --- Cargar todos los ficheros ---
lst <- lapply(files, read_csv_auto)
datos <- rbind_union(lst)

```

Selección del Universo y Variables de Interés

Sobre el conjunto consolidado, se aplicó un filtrado para restringir el análisis a las columnas relevantes (Renombrado de CURSO a anio) y al universo de egresados de Grado (tipo_estudio == "1"), excluyendo otros niveles formativos para garantizar la comparabilidad.

```

# --- Selección de columnas deseadas ---
# nombres base tal cual vienen en los microdatos
base_keep <- intersect(c("CURSO", "TIPOESTUDIO", "AMBITO", "RENDIMIENTO"), names(datos))

# patrones de columnas adicionales (insensible a mayúsculas/minúsculas)
nm <- names(datos)
keep_extra <- nm[
  grepl("^ACTIVIDAD", nm, ignore.case = TRUE) |
  grepl("^GRUPO_?COT", nm, ignore.case = TRUE) | # GRUPOCOT o GRUPO_COT
  grepl("^TCONTRATO", nm, ignore.case = TRUE) |
  grepl("^TJORNADA", nm, ignore.case = TRUE)
]

cols_to_keep <- unique(c(base_keep, keep_extra))
tabla <- datos[, cols_to_keep, drop = FALSE]

# --- Renombrar CURSO -> anio, TIPOESTUDIO -> tipo_estudio (si existen) ---
if ("CURSO" %in% names(tabla)) names(tabla)[names(tabla) == "CURSO"] <- "anio"
if ("TIPOESTUDIO" %in% names(tabla)) names(tabla)[names(tabla) == "TIPOESTUDIO"] <- "tipo_es

datosdef=tabla[tabla$tipo_estudio=="1",]

```

Tratamiento de Valores Ausentes (NAs)

Se implementó una estrategia de limpieza estricta. Los códigos administrativos “99” (No consta) y “Q0” (Sin rendimiento) fueron recodificados como valores perdidos (NA). Posteriormente, se evaluó la tasa de completitud de los registros.

```
# Imputamos los NA a los 99
variables_objetivo_a_modificar <- c("GRUPOCOTA1T3","GRUPOCOTA2T3","GRUPOCOTA3T3","GRUPOCOTA4T3",
                                     "TCONTRATO1T3","TCONTRATO2T3","TCONTRATO3T3","TCONTRATO4T3",
                                     "TJORNADAA1T3","TJORNADAA2T3","TJORNADAA3T3","TJORNADAA4T3")

for (col in variables_objetivo_a_modificar) {
  datosdef[[col]][datosdef[[col]] == 99] <- NA
}

# Imputamos los NA a los Q0
datosdef[["RENDIMIENTO"]][datosdef[["RENDIMIENTO"]] == "Q0"] <- NA

# 1) Conteo de NO-NA solo en las variables objetivo (+ RENDIMIENTO)
vars_target <- unique(c("RENDIMIENTO", variables_objetivo_a_modificar))
vars_target <- vars_target[vars_target %in% names(datosdef)]

# Tratamos cadenas vacías como NA
datosdef[datosdef == ""] <- NA
```

Generación del Dataset Analítico (datos_analisis)

Finalmente, se construyeron las variables derivadas necesarias para la contrastación de hipótesis:

1. Rama de Enseñanza (GRUPO): Normalización de la variable AMBITO.
2. Calidad del Contrato (CalidadContrato): Variable dicotómica donde la jornada parcial (TJORNADAA1T3 == 2) se codifica como 1 (baja calidad) y el resto como 0.

El conjunto final datos_analisis contiene únicamente los registros completos (complete.cases) para las variables críticas, formateados como factores ordenados.

```
# 1) Grupo
datosdef$GRUPO <- ifelse(datosdef$AMBITO == 9999,
                          NA_integer_,
                          datosdef$AMBITO %/% 100)

datosdef$AMBITO <- NULL

# 2) Filtrar filas con claves completas: anio, RENDIMIENTO, GRUPO
datosdef_ok <- datosdef[!is.na(datosdef$anio) &
                        !is.na(datosdef$RENDIMIENTO) &
                        !is.na(datosdef$GRUPO), , drop = FALSE]

# 3) Construir dataset de análisis solo con las variables relevantes
vars_analisis <- c("anio", "RENDIMIENTO", "GRUPO", "TJORNADAA1T3")
vars_analisis <- vars_analisis[vars_analisis %in% names(datosdef_ok)]

datos_analisis <- datosdef_ok[, vars_analisis, drop = FALSE]

# 4) Eliminar casos sin información de jornada
```

```

datos_analisis <- datos_analisis[!is.na(datos_analisis$TJORNADAA1T3), , drop = FALSE]

# 5) Crear CalidadContrato SOLO a partir de TJORNADAA1T3
tmp <- datos_analisis$TJORNADAA1T3
if (is.character(tmp)) {
  suppressWarnings(tmp_num <- as.integer(tmp))
} else {
  tmp_num <- suppressWarnings(as.integer(tmp))
}

# Aquí asumimos (ajústalo si es al revés):
#   CalidadContrato = 1 -> contrato de mala calidad si TJORNADAA1T3 == 2
#   CalidadContrato = 0 -> contrato de buena calidad si TJORNADAA1T3 != 2
datos_analisis$CalidadContrato <- ifelse(tmp_num == 2L, 1L, 0L)

# 6) (Opcional) dejar solo las columnas finales
datos_analisis <- datos_analisis[, c("anio", "RENDIMIENTO", "GRUPO",
                                     "TJORNADAA1T3", "CalidadContrato"),
                                drop = FALSE]

datos_analisis$anio <- factor(datos_analisis$anio,
                              levels = c("2018-2019", "2019-2020", "2020-2021",
                                           "2021-2022", "2022-2023"),
                              ordered = TRUE)

datos_analisis$RENDIMIENTO <- factor(datos_analisis$RENDIMIENTO,
                                     levels = c("Q1", "Q2", "Q3", "Q4"),
                                     ordered = TRUE)

library(dplyr)
library(ggplot2)
# Verificación final
glimpse(datos_analisis)

```

```

Rows: 58,703
Columns: 5
$ anio          <ord> 2018-2019, 2018-2019, 2018-2019, 2018-2019, 2018-2019, ~
$ RENDIMIENTO   <ord> Q4, Q2, Q1, Q4, Q3, Q1, Q4, Q3, Q4, Q1, Q3, Q4, Q1, Q1~
$ GRUPO         <dbl> 4, 9, 9, 9, 9, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, ~
$ TJORNADAA1T3  <int> 1, 2, 2, 2, 1, 2, 2, 1, 2, 2, 1, 2, 1, 1, 2, 2, 1, ~
$ CalidadContrato <int> 0, 1, 1, 1, 0, 1, 1, 0, 1, 1, 0, 1, 0, 0, 1, 1, 1, 0, ~

```

4 Resultados Descriptivos

Objetivo: Validar que los grupos son comparables y observar tendencias “a simple vista”

4.1 Caracterización de la Muestra

- Uso de gtsummary para comparar variables demográficas y académicas (Rendimiento, Rama) entre el grupo Control (2018-19) y Tratamiento (2019-23).
- Test estadístico de diferencias entre grupos (Chi-cuadrado).

4.2 Análisis Exploratorio Visual:

- Gráfico de barras: Tasa de Precariedad (Jornada Parcial) por Rama de Enseñanza.
- Evolución temporal: Línea de tendencia de la precariedad a lo largo de los 5 años (para ver si hubo un “shock” en 2020).

5 Análisis Inferencial

Objetivo: Aislar el efecto COVID de otras variables confusoras

5.1 Especificación del Modelo

- Regresión Logística Binaria (glm).
- VD: Probabilidad de Precariedad (CalidadContrato == 1).
- VIs: Periodo COVID + Rama + Rendimiento Académico.

5.2 Resultados del Modelo

- Presentación de Odds Ratios (OR) e intervalos de confianza.
- Forest Plot: Visualización de los coeficientes para ver qué ramas protegen más contra la precariedad.

6 Discusión y Conclusiones

6.1 Interpretación

¿Se confirma la hipótesis? (Si $OR > 1$ en periodo COVID, sí afectó).

6.2 Matrices por rama

¿Afectó a todas las carreras por igual o sufrieron más las que requieren presencialidad (ej. Ciencias vs Humanidades)?

6.3 Limitaciones

Uso de datos administrativos (sin datos de habilidades blandas). Esto ya lo podemos hablar y consideramos si hacemos mención de las soft skills o no.