

# Proyecto Modelos Lineales

Oscar Gamboa, Gonzalo Mardones, Nicolas Montecinos

2022-10-12

## Introducción

- bla bla

## Planteamiento del Problema

- Fenomeno: Licencias medicas en funcionarios de educación en Huasco - 2022 han llegado con mucha frecuencia. . .

## Justificación del Problema

Las razones que conducen a investigar el fenómeno.

- En funcion del impacto de las licencias medicas. - caracterizacion de la licencia medica (numeros promedios de días )

## Estado del Arte

BUSCAR UNA REFERENCIA A NIVEL NACIONAL PARA DECLARAR EL IMPACTO (DOCENTES Y MEDICOS) METODOLOGIA DE CONTEO (DESCRIPTIVO)

## Objetivo General y Objetivo Especifico

- Obj General: se busca caracterizar la catnidad de dias promedios de LM de los docentes contratados bajo la administración del servicio de ed. publica de la provincia de Huasco, con info actualizada ENE-2022 a OCT-2022 con el fin poder entregar mayores antecedentes a la hora de tomar decisiones a nivel provincial
- Obj Especificos:
  - Obtener, recopilar y extraer las características más importantes que aporten al estudio de la investigación
  - Descripción y asociación de variables predictoras con respecto a la cantidad promedio de días.
  - validación la base de datos y modificar su estructura para los fines de este estudio

## Plateamiento de la hipótesis en el estudio

- Se quiere determinar si la evaluación docente esta indirectametne relacionada con el promedio de días tomados por licencias medicas de los mismos docentes

## Generación de unidades de información

### Diseño de estrategias metodológicas

Debido a lo que se quiere explicar, se plantea metodologia a traves de regresiones lineales debido a que se quiere explocar . . . .

## Procesamiento de las unidades de información

### Descripción de los datos

Base de datos de Licencias Médicas		
Variable	Tipo	Descripción
SEXO	Factor	Sexos encontrados: Masculino o femenino
ESTADO_CIVIL	Factor	Relación de familia, provenientes al matrimonio
EDAD	Double	...
RENTA_PROMEDIO	Double	...
SISTEMA_SALUD	Factor	...
JORNADA	Double	...
NIVEL	Factor	...
TIPO_ESTABLECIMIENTO	Factor	...
CALIDAD_DESEMPEÑO	Factor	profe jefe o no
TRAMO_DOCENTE	Factor	...
EVALUACION_DOCENTE	Factor	...
PROMEDIO_DIAS_LM	Double	...
TRASLADO_COMUNA	Double	...

### Detalle de ETL de los datos

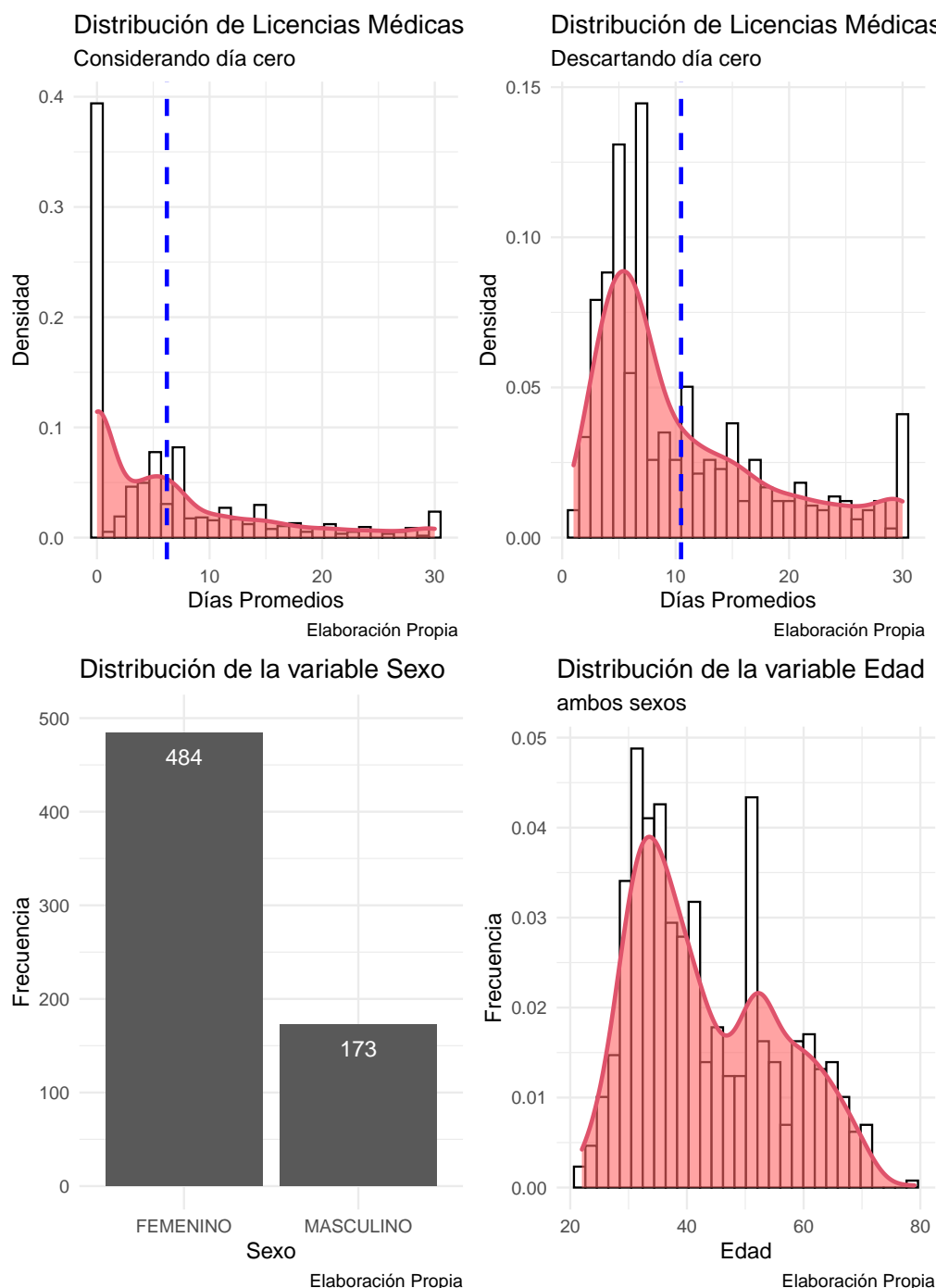
```
BBDD_proyecto <- BBDD_proyecto %>% mutate_at(c("SEXO", "ESTADO_CIVIL",  
                                                  "SISTEMA_SALUD", "NIVEL",  
                                                  "TIPO_ESTABLECIMIENTO",  
                                                  "CALIDAD_DESEMPEÑO",  
                                                  "COMUNA_ESTABLECIMIENTO",  
                                                  "ESTAMENTO", "TRAMO_DOCENTE"), factor)  
  
BBDD_proyecto$EVALUACION_DOCENTE <- factor(BBDD_proyecto$EVALUACION_DOCENTE,  
                                             levels = c("SIN EVALUACIÓN",  
                                                  "DESTACADO",  
                                                  "BASICO",  
                                                  "INSATISFACTORIO",  
                                                  "COMPETENTE"))  
  
BBDD_proyecto_sin_ceros <- BBDD_proyecto %>% filter(PROMEDIO_DIAS_LM != 0)
```

## Estadística Descriptiva

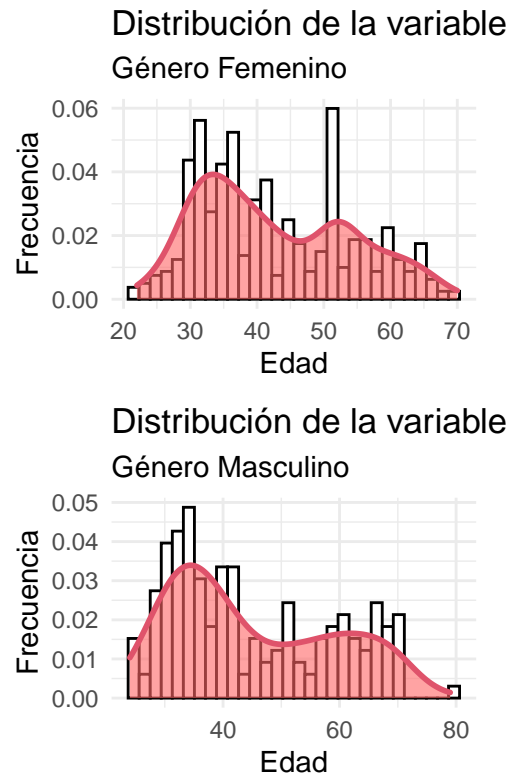
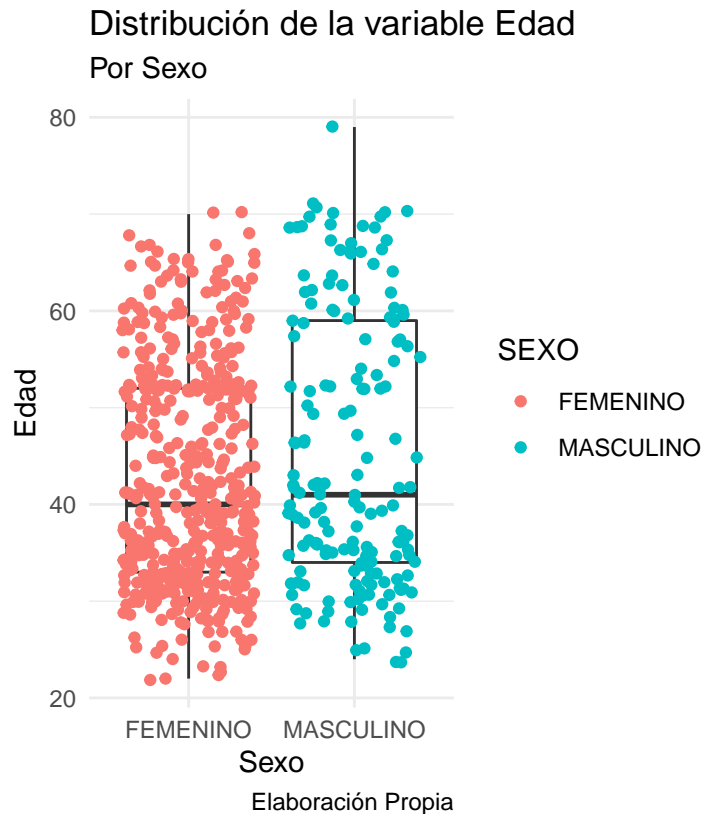
A continuación se presenta análisis descriptivos para las variables más relevantes del estudio, de las cuales destacamos días promedios de licencias médicas, .....

### Distribución de Licencias Médicas

El gráfico de la izquierda muestra la distribución de licencias médicas que incluyen a los docentes/funcionarios que no han presentado licencias médicas hasta la fecha, en el gráfico de la derecha se muestra la distribución de licencias médicas pero descartando los casos de docentes/funcionarios que no han presentado licencias médicas, siendo este el grupo de personas empleadas para el análisis

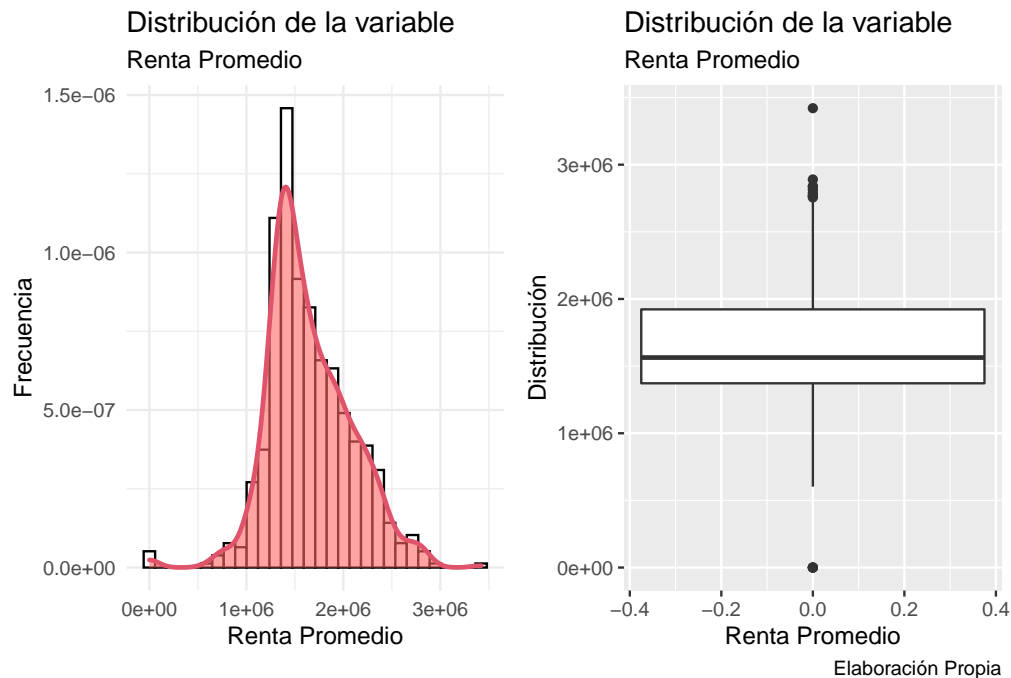


De acuerdo a los resultados, la variable edad parece tener una distribución bimodal, de hecho, nos da luces de una cierta mezcla de distribuciones, por lo cual se hace necesario el poder analizar esta variable de acuerdo a variables categóricas como es el sexo.



En este sentido, la distribución de edades en mujeres y hombres es bastante similar, en cada sexo se puede apreciar una distribución bimodal.

##	Min.	1st Qu.	Median	Mean	3rd Qu.	Max.
##	0	1372121	1563785	1651645	1922665	3420643

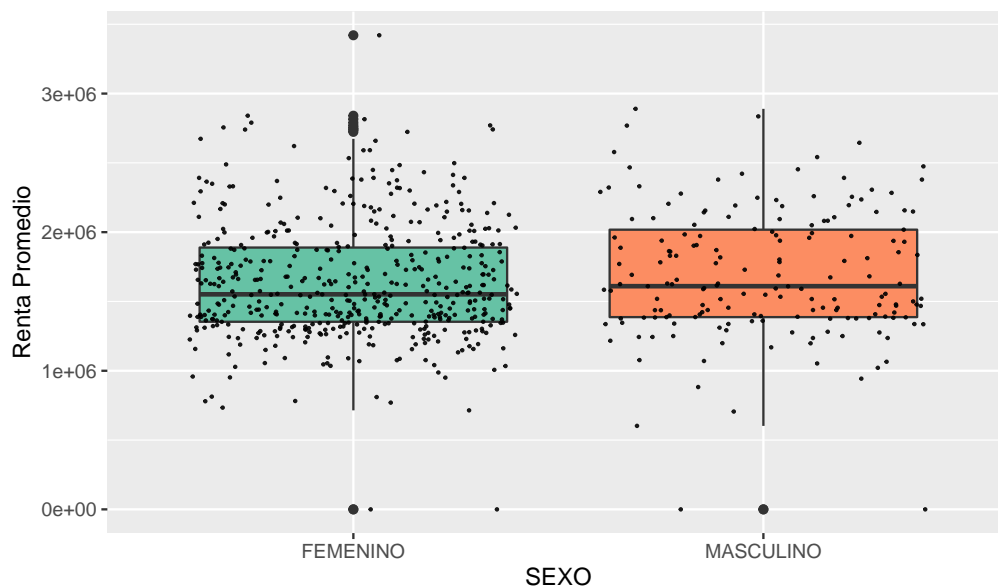


En este caso la distribución de la renta está centrada alrededor de 1,7 millones de pesos, donde el 25% más bajo de la distribución de renta está por debajo de los 1.37 millones, mientras que el 25% de los casos con mayores rentas se encuentran entre 1.93 y 3.42 millones de pesos.

```
##      Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##      0 1353619 1551608 1637245 1888925 3420643
```

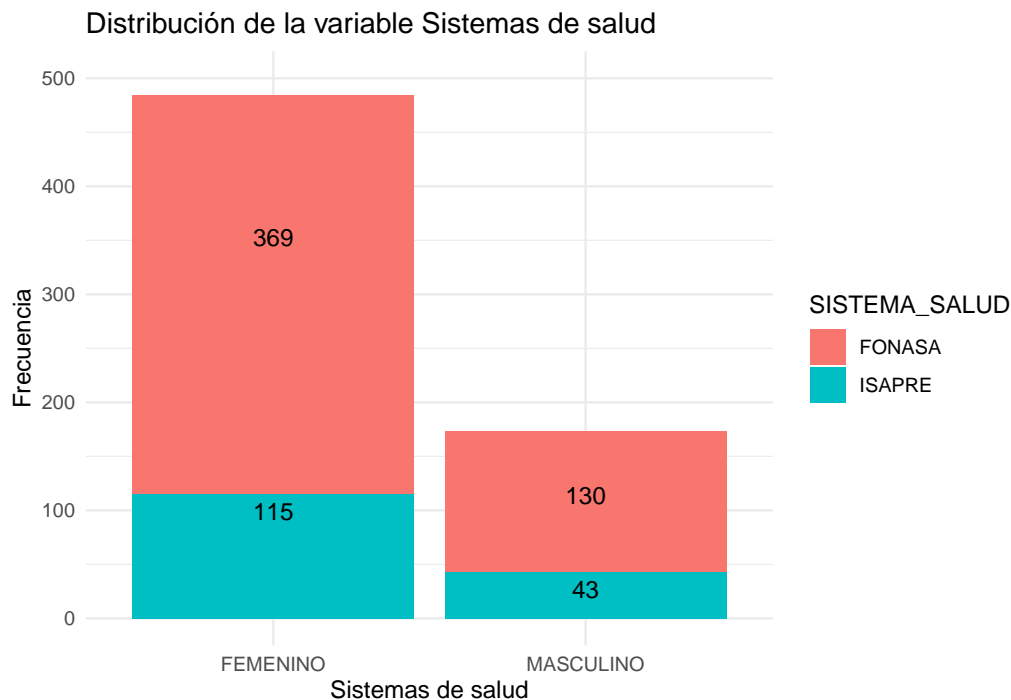
```
##      Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##      0 1387003 1610283 1691930 2018010 2889913
```

Distribución de variables por la renta promedio  
Boxplot por sexo



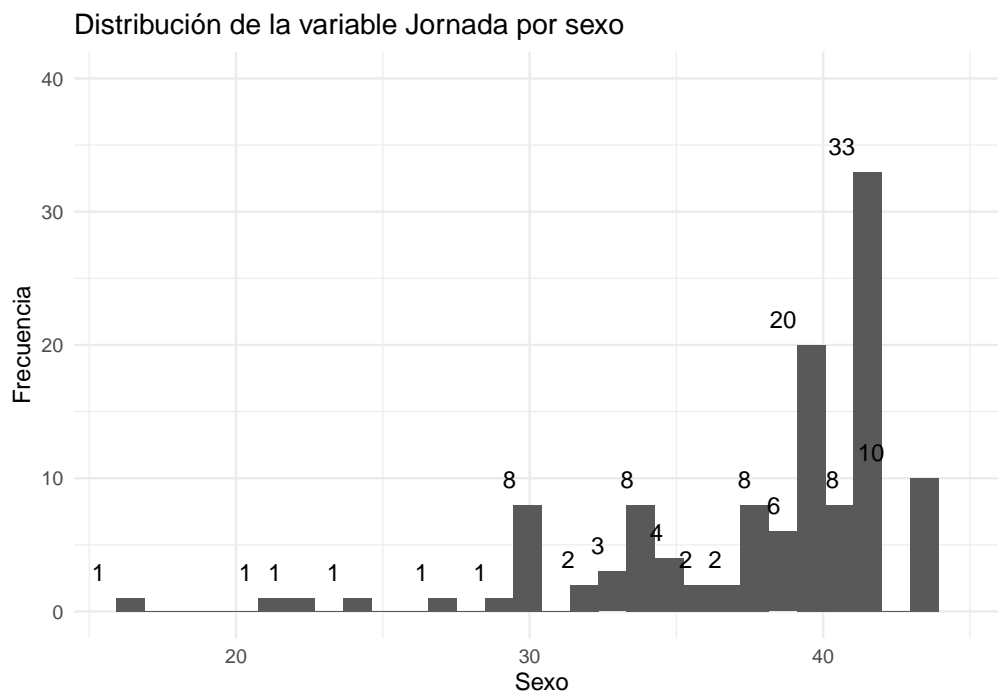
Comparando los niveles de renta por sexo se puede notar una cierta similitud en el rango de montos promedio, más aún, las estadísticas de posición (cuartiles) son bastante similares, por lo que podría no haber un efecto

del sexo sobre los ingresos promedio. Para las mujeres la mediana de las rentas promedio es 1.58 millones, mientras que para los hombre bordea los 1.62 millones de pesos.



Según los datos, la proporción de personas que pertenecen al sistema de salud fonasa es alto tanto en hombres como en mujeres, esto tiene sentido desde el punto de vista que es una base de información de la provincia de Huasco donde existe una alta mayoría de zonas rurales donde el sistema fonasa es predominante.

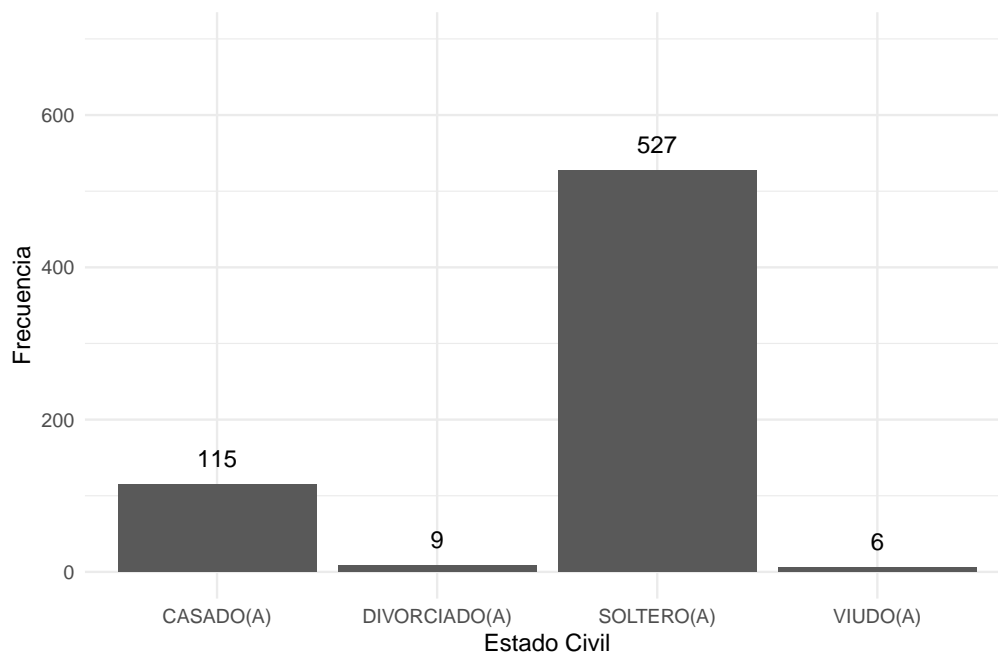
```
##      Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##  16.00   44.00   44.00   42.94   44.00   44.00
```



Podemos notar que prácticamente el 75% de los las personas de la muestra trabajan jornadas de 44 o menos

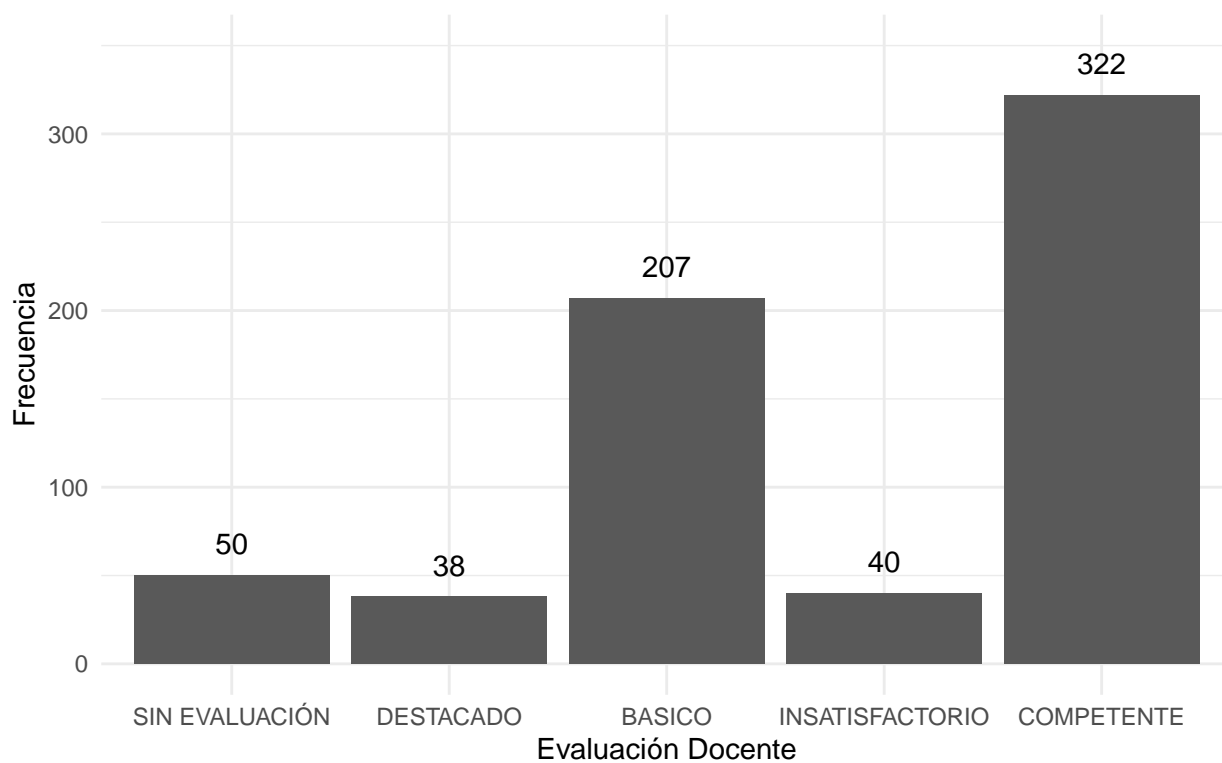
horas semanales, mientras que sólo un caso muestra 88 horas laborales semanales, este caso se debe a ....  
(justificar el caso)

Distribución de la variable Estado Civil



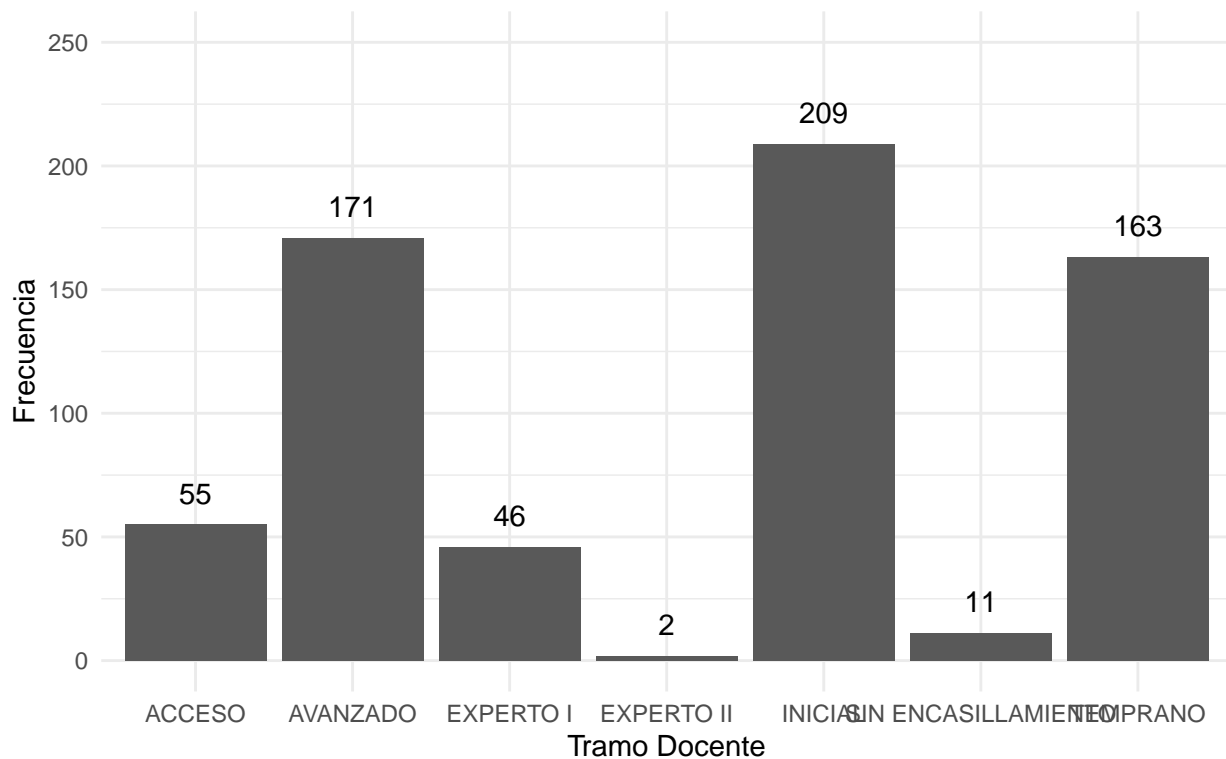
texto

Distribución de la variable Evaluación Docente



Elaboración Propia

## Distribución de la variable Tramo Docente



Elaboración Propia

## Selección Formal de Modelo

### Forward

Para la selección del modelo forward se utilizó una significancia del 5% en cada uno de los test de hipótesis realizados

```
modelo0 = lm(PROMEDIO_DIAS_LM ~ 1, data = BBDD_proyecto_sin_ceros)
add1(modelo0, ~ . + SEXO + ESTADO_CIVIL + EDAD + RENTA_PROMEDIO +
        SISTEMA_SALUD + JORNADA + NIVEL + CALIDAD_DESEMPEÑO +
        ESTAMENTO + EVALUACION_DOCENTE + TRAMO_DOCENTE + TRASLADO_COMUNA,
        test="F")
```

```
## Single term additions
##
## Model:
## PROMEDIO_DIAS_LM ~ 1
##
##      Df Sum of Sq  RSS    AIC F value    Pr(>F)
## <none>                 38257 2672.3
## SEXO      1      40.0 38217 2673.6  0.6860 0.4078231
## ESTADO_CIVIL 3      12.2 38245 2678.1  0.0696 0.9761499
## EDAD      1     658.4 37599 2662.9 11.4703 0.0007495 ***
## RENTA_PROMEDIO 1     117.9 38139 2672.3  2.0241 0.1553012
## SISTEMA_SALUD 1     251.2 38006 2670.0  4.3299 0.0378370 *
## JORNADA      1      38.7 38219 2673.7  0.6635 0.4156263
## NIVEL      1     388.9 37868 2667.6  6.7273 0.0097068 **
## CALIDAD_DESEMPEÑO 1     942.1 37315 2657.9 16.5368 5.352e-05 ***
## ESTAMENTO    1        3.5 38254 2674.2  0.0594 0.8075512
```



```
## EVALUACION_DOCENTE 4      3441.8 34816 2618.4 16.1139 1.364e-12 ***
## TRAMO_DOCENTE      6      909.7 37348 2668.5  2.6388 0.0155134 *
## TRASLADO_COMUNA    1        2.2 38255 2674.3  0.0369 0.8476568
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

Para decidir la entrada del segundo predictor, se debe evaluar todos los modelos que ya contienen al promedio de días de licencias medicas, a lo que se agrega cada uno de los predictores restantes por separado.

De acuerdo al caso anterior, el menor valor-p corresponde al modelo `PROMEDIO_DIAS_LM ~ EVALUACION_DOCENTE`

Coeficientes	Estimación	Std. Error	t value	Pr(>t)
Intercepto	7.59163	1.49024	5.094	4.6e-07
EVALUACION_DOCENTEDESTACADO	-2.69256	1.55890	-1.727	0.08460
EVALUACION_DOCENTEBASICO	-2.08287	1.13656	-1.833	0.06732
EVALUACION_DOCENTEINSATISFACTORIO	7.87294	1.52249	5.171	3.1e-07
EVALUACION_DOCENTECOMPETENTE	-1.91725	1.12365	-1.706	0.08844
CALIDAD_DESEMPEÑOTITULAR	2.65846	0.72190	3.683	0.00025
NIVELMEDIA	-1.84583	0.63467	-2.908	0.00376
EDAD	0.06085	0.02425	2.509	0.01233

## Backward