

# Proyecto Modelos Lineales

Oscar Gamboa, Gonzalo Mardones, Nicolas Montecinos

2022-10-14

## Introducción

En la actualidad, los servicios públicos han sido objeto de denuncias por parte de la ciudadanía por los temas que competen a las licencias médicas presentadas por los funcionarios de dichas entidades, en ella destacan principalmente las áreas correspondientes a la educación pública y en específico direccionados a los DOCENTES. La ley N° 20.585 de 2012, emitida por el Ministerio de Salud, básicamente indica que la institución estatal se encarga de la tramitación y pago completo de esta, ya sea FONASA o ISAPRE, en desmedro de dejar sin pagos salariales para sus DOCENTES, sin embargo, esto tiende a generar controversia por la gran cantidad de licencias tipo 1 (correspondiente a las psiquiátricas, estrés laboral, medicina general, accidente de trayecto, accidente de trabajo y enfermedad profesional) y el acumulado de días ausentes emitidas por este grupo, ya que indica un problema a los servicios públicos, estos son: búsqueda de reemplazos (por lo general no son docentes debido a la poca oferta que hay hoy en día impactando directamente en la calidad de educación), mayor inversión en las acciones preventivas (mutuales y capacitaciones anti estrés, generando aumentos de presupuestos en las arcas fiscales), uso y abuso de las mismas (compras de licencias médicas que corresponden a un delito), y agotamiento del sistema público.

En el presente proyecto, se direccionará en tratar de explicar el número promedio de días de licencias médicas en Docentes de un servicio local de educación pública de la provincia del Huasco (comunas que comprende: Alto del Carmen, Vallenar, Freirina y Huasco) de la Región de Atacama, en función a características propias de ellos junto a variables relevantes a la demografía y a las educacionales, con el fin de tener una primera mirada ante la problemática ya expuesta, que aporten en gran medida a la toma de decisiones por parte de los establecimientos educacionales estatales.

## Planteamiento del Problema

Dado los antecedentes ya presentados en el acápite anterior, se quiere determinar si el número promedio de días de las licencias médicas (aprobadas) emitidas por los docentes, de enero a octubre del año 2022, puede ser explicado o influido por fenómenos endógenos como la evaluación docente o el tramo docente (para mayor información hacer click en el siguiente link) o por variables exógenas (factores ajenos al DOCENTE) como podría ser el traslado de comuna para ir a su lugar de trabajo, al grupo etario al cual pertenece o si es profesor(a) jefe de un curso (debido a que tiene mayor responsabilidad administrativa) etc. Y en caso de ser significativo el efecto ante cambios en los días promedios de licencias médicas, cual es la magnitud y característica de estas variables.

## Justificación del Problema

Se desea identificar los fenómenos que estén, de alguna manera, relacionada con la emisión de licencias médicas para poder direccionar recursos necesarios, realizar cambios correspondientes y anticipar los problemas que generan las ausencias que puedan afectar la calidad de la educación pública bajo la administración de un servicio local en la provincia del Huasco, Región de Atacama, Chile.

## Estado del Arte

En la actualidad, la problemática descrita anteriormente se ve reflejada tanto en la comunidad docente (ver enlace ) como en la ciudadanía (ver enlace) y (ver enlace)

## Objetivos

- Objetivo General

Se busca explicar la cantidad de días promedios de Licencias Médicas de los docentes contratados, bajo la administración del Servicio de Educación Pública de la provincia del Huasco, con información actualizada entre ENERO a OCTUBRE del 2022, considerando variables exógenas y endógenas del Docente, con el fin poder entregar mayores antecedentes a la hora de tomar decisiones a nivel provincial.

- Objetivos Específicos

- 1.- Definir, obtener, extraer y limpiar los datos que corresponden a las características que más se relacionen y que aporten al posible efecto de los días en promedio de licencias médicas en docentes de la provincia en el periodo mencionado.
- 2.- Descripción uni y bi variada junto a un análisis de asociación de variables predictoras con respecto a la cantidad promedio de días por licencias médicas, para la toma de decisiones respecto a la base de datos a tratar.
- 3.- Validación de la base de datos, modificación de su estructura (en caso de ser necesario) y presentación del modelo completo ante las variables escogidas.
- 4.- Exhibición del modelo escogido, bajo la metodología de selección de variables (Forward y Backward)

## Planteamiento de la hipótesis en el estudio

Se requiere determinar lo siguiente: ¿Que tan cierto es que a mayor evaluación docente y tramo docente (variables endógenas), menor es la cantidad de días de solicitadas por licencias médicas?, y si lo hay, ¿De cuanto es lo que impacta en promedio a los días tomados por licencias medicas tipo 1?

¿Existen variables exógenas que tienen un impacto en los días promedios tomados por las licencias médicas?, si es así, ¿De cuanto es lo que explica este fenómeno?

## Generación de unidades de información

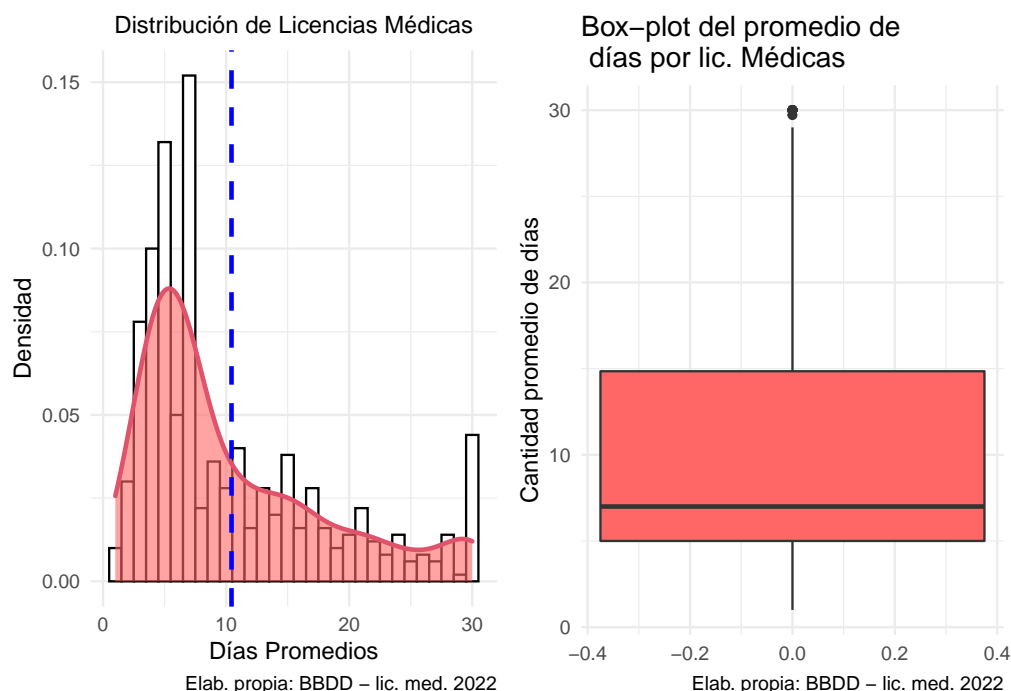
Las unidades de información hace referencia a la dotación de Docentes en la provincia del Huasco que han presentado licencias médicas en el periodo de enero a octubre 2022, que se encuentran bajo la administración del Servicio Local de Educación Pública del Huasco, Región de Atacama, Chile. Por lo tanto se tiene lo siguiente:

- Total de docentes en la provincia: 1.109.
- Cantidad de Establecimientos Educativos: 53.
- Total de docentes que han presentado licencias medicas tipo 1: 657

Para fines de este estudio, se determina una muestra aleatoria de 500 docentes que presentaron licencias médicas tipo 1, bajo la reglamentación impuesta por la entidad evaluadora.

Finalmente se obtienen los siguientes antecedentes de acuerdo a los días promedios de licencias médicas emitidas por docentes de la provincia del Huasco:

Cant. Docentes	Min.	1st Qu.	Mediana	Promedio	3rd Qu.	Max.
500	1.00	5.00	7.00	10.44	14.85	30.00



Se observa que el promedio de días se aproxima a una distribución asimétrica negativa, lo que indica...

### Diseño de estrategias metodológicas

Para el presente proyecto, se plantea metodología a través de regresiones lineales, dando principal énfasis al análisis descriptivo y en el tratamiento de la estructura de la base de datos, tomando decisiones de esta ante una postura de recodificación de variable si es posible dado el caso.

### Estructura de la base de datos

Base de datos de Licencias Médicas		
Variable	Tipo	Descripción
SEXO	Factor	Masculino y Femenino
ESTADO_CIVIL	Factor	Casado(a), Viudo(a), Soltero(a), Divorciado(a)
EDAD	Double	Edad
RENTA_PROMEDIO	Double	Renta Promedio en Pesos Chilenos
SISTEMA_SALUD	Factor	Fonasa, Isapre
JORNADA	Double	Mediadas en horas semanales
NIVEL	Factor	Básica o Media
TIPO_ESTABLECIMIENTO	Factor	Escuela o Liceo
CALIDAD_DESEMPEÑO	Factor	Profesor Titular o no
TRAMO_DOCENTE	Factor	Inicial, Temprano, Acceso, Avanzado, Experto I, Experto II, Sin Encasillamiento
EVALUACION_DOCENTE	Factor	Sin Evaluación, Destacado, Básico, Insatisfactorio, Competente
PROMEDIO_DIAS_LM	Double	Promedio Días de Licencias Médicas
TRASLADO_COMUNA	Double	SI (1) o NO (0)

## Análisis Descriptivo

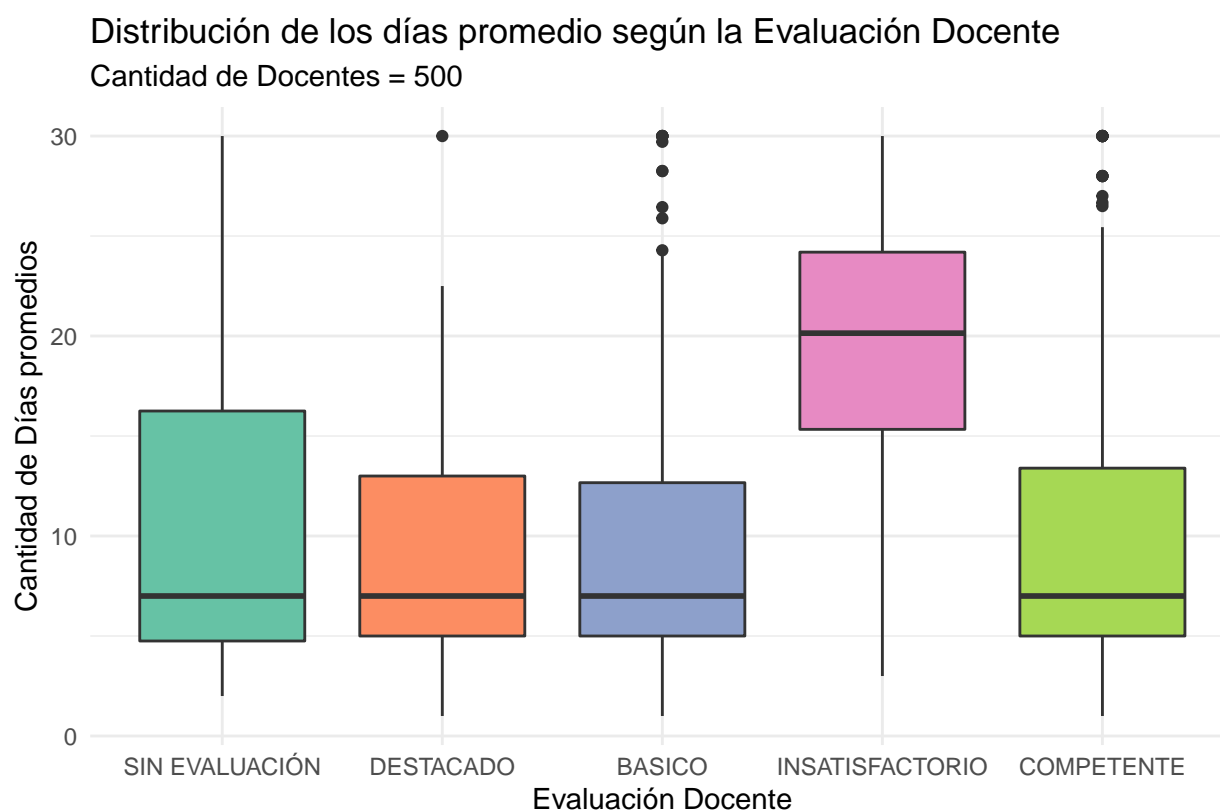
Para el siguiente paso, se definirán las variables a las cuales las llamaremos Exógenas y Endógenas, indicando sus principales indicadores según los días promedios de licencias médicas cursadas.

### a.- Variables Endógenas (indicadores de evaluaciones propias del docentes)

- El Sistema de Evaluación del Desempeño Profesional Docente (o Evaluación Docente) es una evaluación obligatoria para los y las docentes, educadoras y educadores de aula que se desempeñan en establecimientos municipales o estatales a lo largo del país. Su objetivo es fortalecer la profesión docente y contribuir a mejorar la calidad de la educación.<sup>1</sup>

Table 2: Evaluación Docente

EVALUACION_DOCENTE	Promedio	Mediana	DesvEstandar	Mín	Máx	Cantidad_LM
SIN EVALUACIÓN	11.008432	7.00000	9.119703	2	30	110
DESTACADO	9.396442	7.00000	6.866130	1	30	71
BASICO	9.603945	7.00000	7.158842	1	30	478
INSATISFACTORIO	19.417100	20.14286	7.457253	3	30	176
COMPETENTE	9.861203	7.00000	7.169876	1	30	740



Elab. propia: BBDD Lic. Médicas 2022

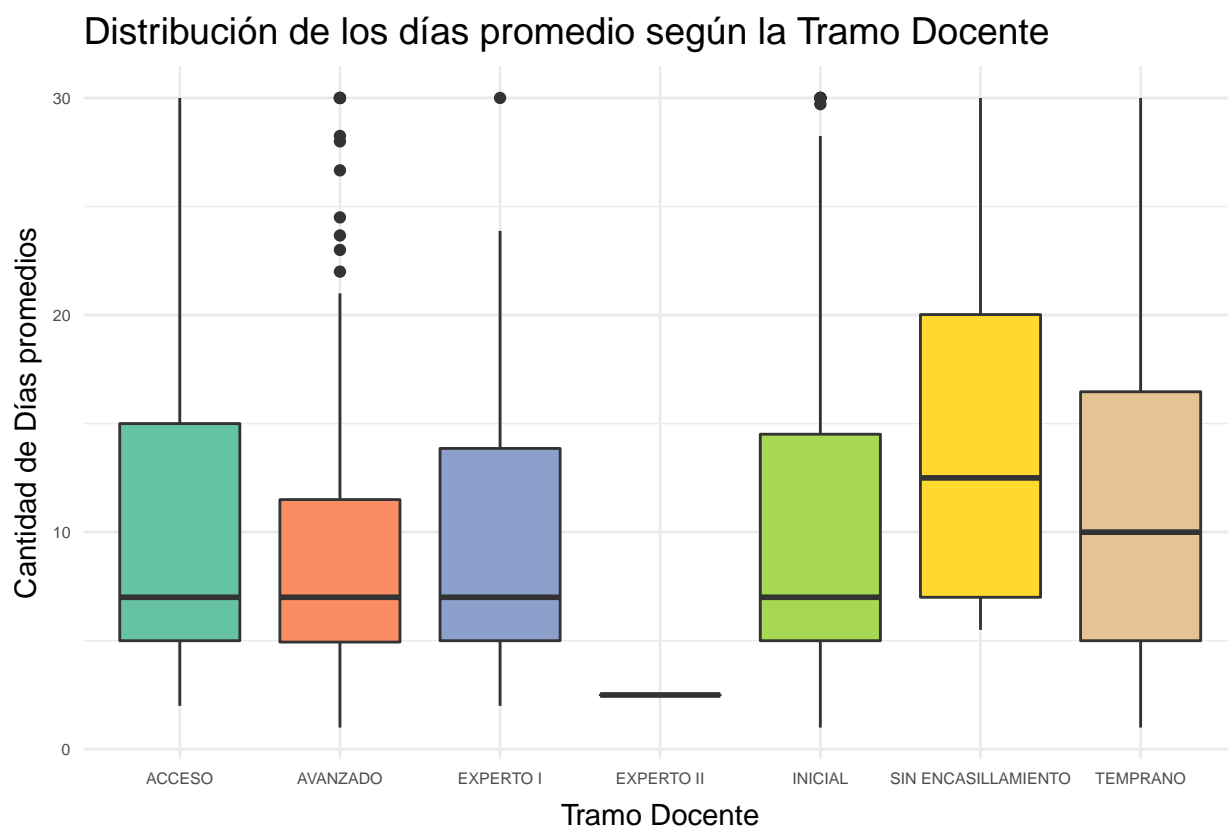
Al observar cada evaluación se aprecia que la categoría Insatisfactorio resalta de las demás por poseer una media y mediana mayor en comparación a las demás categorías de evaluación, por consiguiente, el gráfico de boxplot ratifica lo mencionado e indica que tiene un rango de variación distinto en comparación, sin embargo, las demás categorías de evaluación poseen similares distribuciones al apreciar sus indicadores.

<sup>1</sup>Evaluación Docente, para mayor información visite [www.cpeip.cl](http://www.cpeip.cl)

- La Carrera Docente cuenta con cinco tramos en que los profesores pueden progresar demostrando las competencias y conocimientos alcanzados. Además, existe un tramo provisorio, denominado Acceso, para aquellos docentes que tienen más de cuatro años de experiencia, pero que no cuentan con resultados en evaluaciones previas del Ministerio de Educación.<sup>2</sup>

Table 3: Tramo Docente

TRAMO_DOCENTE	Promedio	Mediana	DesvEstandar	Mín	Máx	Cantidad_LM
ACCESO	10.851940	7.0	8.732914	2.0	30.0	127
AVANZADO	9.318187	7.0	6.978255	1.0	30.0	359
EXPERTO I	9.712430	7.0	6.860585	2.0	30.0	97
EXPERTO II	2.500000	2.5	NA	2.5	2.5	4
INICIAL	10.227714	7.0	7.709113	1.0	30.0	461
SIN	14.604567	12.5	8.893114	5.5	30.0	55
ENCASILLAMIENTO						
TEMPRANO	11.750642	10.0	7.958442	1.0	30.0	472



Elab. propia: BBDD Lic. Médicas 2022

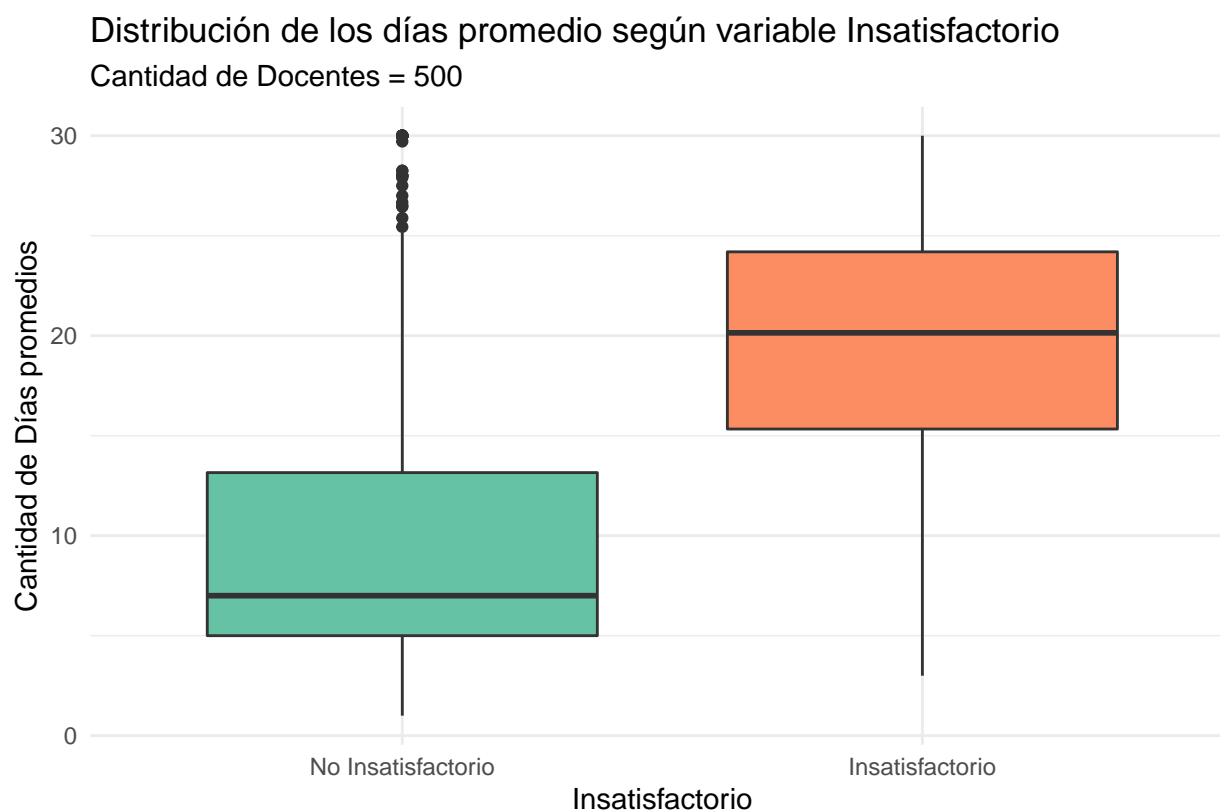
Se instuye que a diferencia de los demás tramos, el EXPERTO II no presenta un rango de variabilidad que la caracterice, por otro lado, la categoría que más se destaca, en términos de sus indicadores, es aquel tramo donde no presenta encasillamiento, pero esto se puede deber a un efecto producido por la cantidad de licencias presentadas de tal tramo. Además es interesante indica que el TRAMO INICIAL y el TEMPRANO poseen la mayor cantidad de licencias medicas liderando así, tanto en media como en mediana este último mencionado respectivamente. Finalmente, y descartando la categoría EXPERTO II, al parecer y en base a un criterio gráfico, todas los tramos poseen un similar comportamiento en funcion a los días promedios por licencias medicas emitidas.

<sup>2</sup>Tramo Docente, para mayor información visite [www.docentemas.cl](http://www.docentemas.cl)

A modo de conclusión de este análisis, se toma la decisión de recodificar la variable EVALUACIÓN DOCENTE, en una nueva variable la que se llamará INSATISFACTORIO, quedando los siguiente indicadores junto con su distribución:

Table 4: Insatisfactorio

INSATISFACTORIO	Promedio	Mediana	DesvEstandar	Mín	Máx	Cantidad_LM
No Insatisfactorio	9.846386	7.00000	7.318830	1	30	1399
Insatisfactorio	19.417100	20.14286	7.457253	3	30	176



Elab. propia: BBDD Lic. Médicas 2022

#### b.- Variables Exógenas (indicadores ajenas al docentes)

En este apartado, se analizarán las variables que son ajenas a las capacidades evaluativas propias del docentes, dado el número promedio de días por licencia médica emitida, las cuales son: SEXO, EDAD, JORNADA, RENTA SALARIAL, ESTADO\_CIVIL, SISTEMA\_SALUD, NIVEL, TIPO\_ESTABLECIMIENTO, CALIDAD\_DESEMPEÑO, COMUNA\_ESTABLECIMIENTO, ESTAMENTO y TRASLADO\_COMUNA.

- La variable SEXO <sup>3</sup> donde el 74.4% equivale a las mujeres y el 25.6% a Hombres, sus indicadores como el promedio y las medianas son similares en torno al 10 y 7 días promedios en licencias médicas respectivamente, además sus varianzas también lo son, lo que da indicios de que posiblemente tengan distribuciones similares y que el efecto respecto a los días promedios de licencias médicas registradas según sexo, a nivel descriptivo, no tenga mayor relevancia dentro de un análisis.
- El ESTADO CIVIL <sup>4</sup> del funcionario(a), presenta diferencias solo con una categoría que es DIVORCIADO(A) con respecto a las demás, presentando un promedio y mediana en torno a 19 y 20 días promedio de licencias médicas respectivamente, en comparación a las otras situaciones que fluctúan

<sup>3</sup>Ver Tabla 5 en apartado Anexo Tablas, junto al gráfico 1 en el apartado Anexo Gráficos.

<sup>4</sup>Ver Tabla 6 en apartado Anexo Tablas, junto al gráfico 2 en el apartado Anexo Gráficos.

en el promedio, entre 9 y 12 días, y 7 días en mediana. En el caso de las desviaciones estandar, se aprecia que la mayor variabilidad también es asociada a la categoría DIVORCIADO(A), por lo que puede ser un indicio de que esta situación en particular tenga una relevancia en el estudio. Sin embargo, se esta hablando de una cantidad minima de 4 funcionarios, que equivale a un 0.8% de la muestra, lo que tiende a minimizar el efecto de la hipótesis recién mencionada.

- En el SISTEMA DE SALUD <sup>5</sup>, el que indica si pertenece a FONASA o ISAPRE, el 75% pertenece a la primera y el 25% restante a la segunda mencionada, donde sus promedios y medianas de días son similares al 10 y 7 días promedio respectivamente y su indicador de variabilidad, que es la desviación estandar, si se diferencian en 2 días promedio siendo superior los afiliados de FONASA. A nivel descriptivo, se podría indicar que se esta en presencia de distribuciones similares y que el impacto dentro de un modelo de regresión lineal posiblemente no sea significativo.
- La variable NIVEL <sup>6</sup>, que indica a que nivel el docente le hace clases, 72.4% pertenece al nivel BASICO y el 27.6% al nivel MEDIO, sus promedios, medianas y desviaciones estandar son similares de un, aproximadamente, 10, 7 y 7 días promedios de licencias medicas presentadas respectivamente. A nivel descriptivo puede dar indicación de que poseen misma distribución en tornos a las magnitudes ya comentadas, y que no debe generar un cambio significativo muy ampli en torno al efecto de los días promedios de licencias medicas presentado por los funcionarios(as).
- El TIPO DE ESTABLECIMIENTO <sup>7</sup> indica si el o la docente pertenece o no a un liceo, colegio, escuela diferencial o a la administración central, donde su distribución de cantidades se mueve en torno al 25%, 72.2%, 1.4% y 0.6% respectivamente, donde ya sus diferencias pueden generar algún tipo de problemas a nivel inferencial. Sus promedio, medianas y desviación estandar en días son similares pero sus cantidades generaran inconsistencias en la interpretación.
- Con respecto a la CALIDAD DE DESEMPEÑO <sup>8</sup>, la cual indica que si un docente tiene un curso de manera titular (profesor jefe) versus si no, en el que se observa que 77.2% de ellos tienen jefatura de curso versus un 22.8% que no tiene jefatura. Donde acá su promedio de dias son 10 y 8 respectivamente (se genera una diferencia), sus medianas de dias son similares en torno al 6 y 7 días promedio respectivamente al igual que su indicador de la variabilidad. lo que puede incurrir descriptivamente hablando, que si bien tienden a tener misma distribución, lo mas probable es que la diferencia de promedio genere un efecto significativo a la hora de ajustar un modelo.
- El TRASLADO DE COMUNA <sup>9</sup> que indica si el o la funcionario(a) se traslada de una comuna a otra para ir a su lugar de trabajo, el 85.4% de ellos no se traslada en comparación del 14.6% que si, donde poseen un promedio, mediana y desviacion estandar de dias promedios similares entre ellos con magnitudes al rededor de 10, 7 y 7 días promedios de licencias medicas emitidas respectivamente, da la indicios de que poseen misma distribución sin ningún tipo de diferencias que puedan ser significativas entre ellas, lo que implicaría a nivel observacional de que al ajustar un modelo no se generaría un efecto relevante ante esta situación respecto a los días promedios de licencias médicas presentadas.

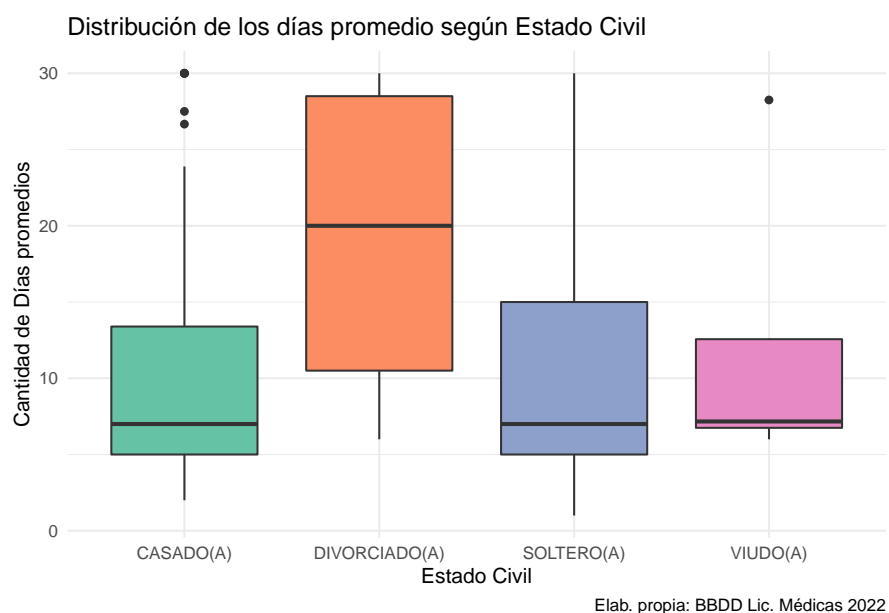
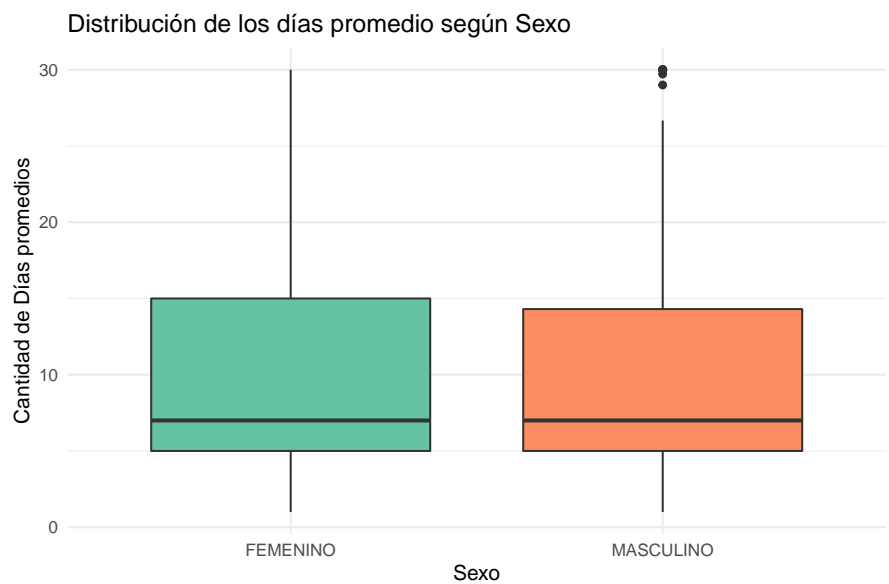
<sup>5</sup>Ver Tabla 7 en apartado Anexo Tablas, junto al gráfico 3 en el apartado Anexo Gráficos.

<sup>6</sup>Ver Tabla 8 en apartado Anexo Tablas, junto al gráfico 4 en el apartado Anexo Gráficos.

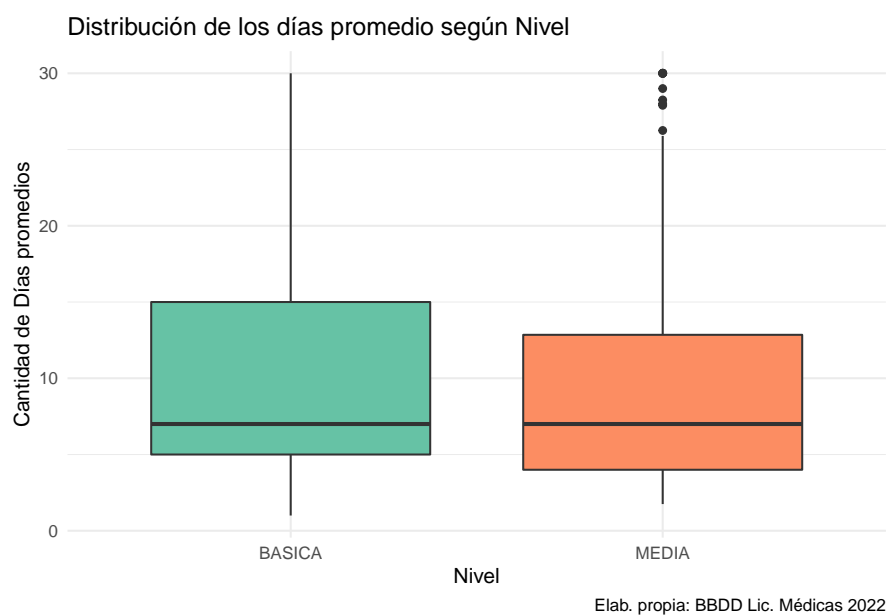
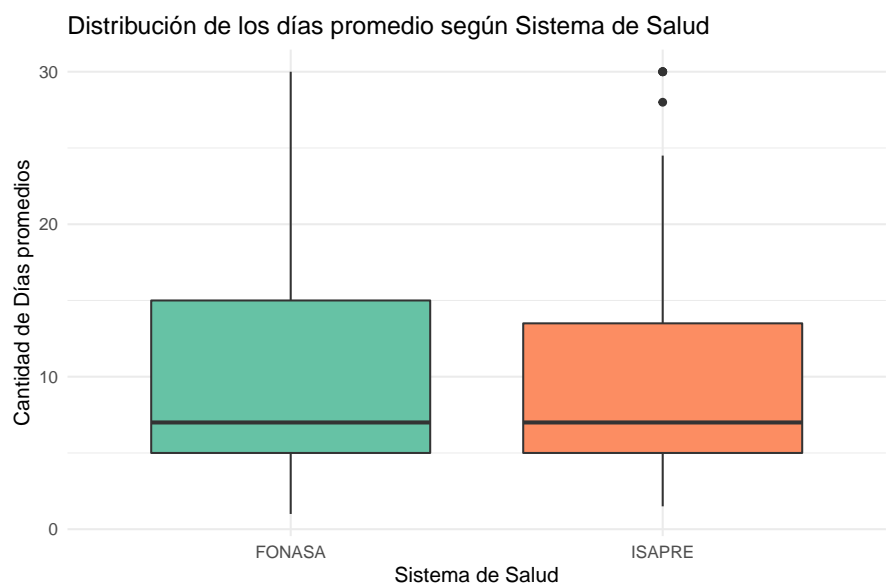
<sup>7</sup>Ver Tabla 9 en apartado Anexo Tablas, junto al gráfico 5 en el apartado Anexo Gráficos.

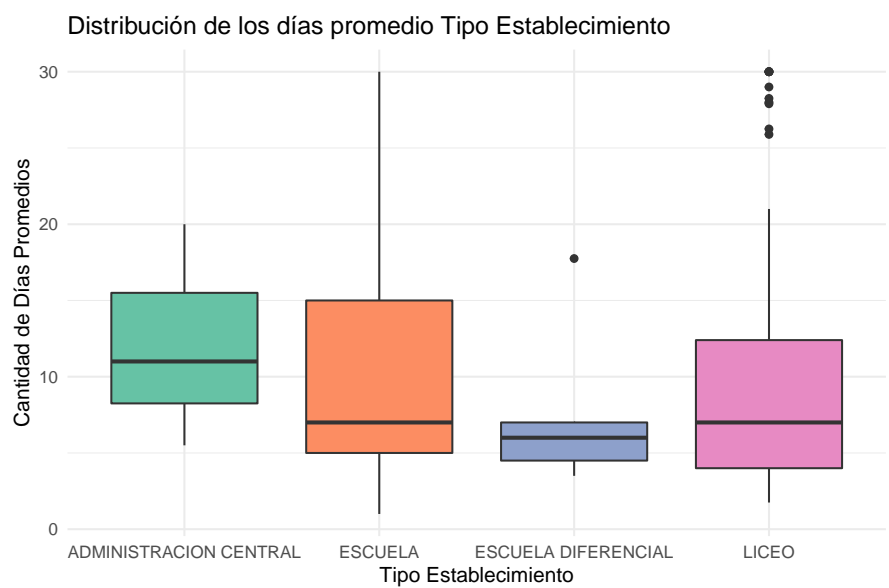
<sup>8</sup>Ver Tabla 9 en apartado Anexo Tablas, junto al gráfico 5 en el apartado Anexo Gráficos.

<sup>9</sup>Ver Tabla 11 en apartado Anexo Tablas, junto al gráfico 7 en el apartado Anexo Gráficos.

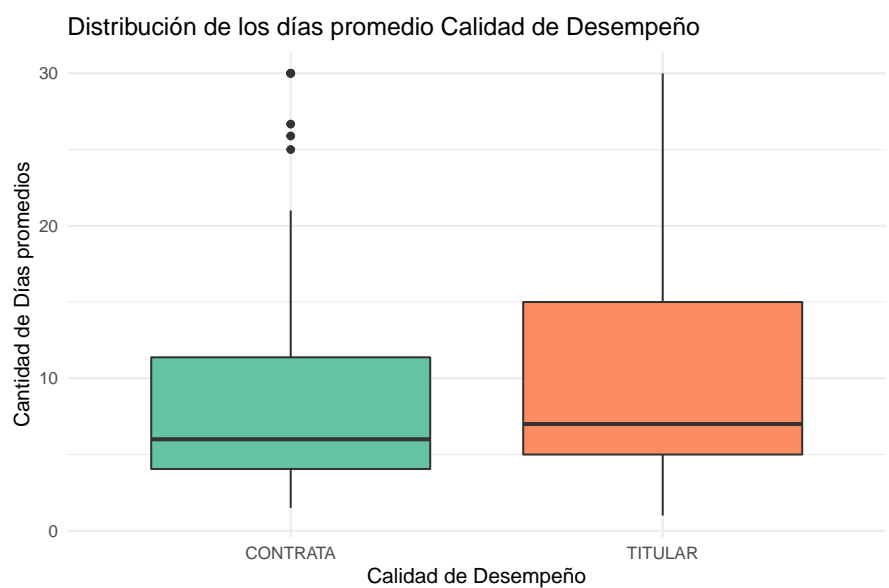








Elab. propia: BBDD Lic. Médicas 2022



Elab. propia: BBDD Lic. Médicas 2022



Table 5: Sexo

SEXO	n	Promedio	Mediana	DesvEstandar	Mín	Máx	Cantidad_LM
FEMENINO	372	10.49411	7	7.596903	1	30	1189
MASCULINO	128	10.28185	7	7.929107	1	30	386

Table 6: Estado Civil

ESTADO_CIVIL	n	Promedio	Mediana	DesvEstandar	Mín	Máx	Cantidad_LM
CASADO(A)	91	9.979785	7.000000	7.489062	2	30.00	283
DIVORCIADO(A)	4	19.000000	20.000000	11.832160	6	30.00	26
SOLTERO(A)	401	10.441749	7.000000	7.625920	1	30.00	1252
VIUDO(A)	4	12.145833	7.166667	10.751050	6	28.25	14

Table 7: Sistema de Salud

SISTEMA_SALUD	n	Promedio	Mediana	DesvEstandar	Mín	Máx	Cantidad_LM
FONASA	375	10.740568	7	8.002497	1.0	30	1216
ISAPRE	125	9.537377	7	6.546181	1.5	30	359

Table 8: Nivel

NIVEL	n	Promedio	Mediana	DesvEstandar	Mín	Máx	Cantidad_LM
BASICA	362	10.860529	7	7.761376	1.00	30	1158
MEDIA	138	9.336039	7	7.360331	1.75	30	417

Table 9: Tipo Establecimiento

TIPO_ESTABLECIMIENTO	n	Promedio	Mediana	DesvEstandar	Mín	Máx	Cantidad_LM
ADMINISTRACION CENTRAL	3	12.166667	11	7.320064	5.50	20.00	4
ESCUELA	361	10.849557	7	7.786245	1.00	30.00	1179
ESCUELA DIFERENCIAL	7	7.178571	6	4.857799	3.50	17.75	14
LICEO	129	9.429805	7	7.421881	1.75	30.00	378

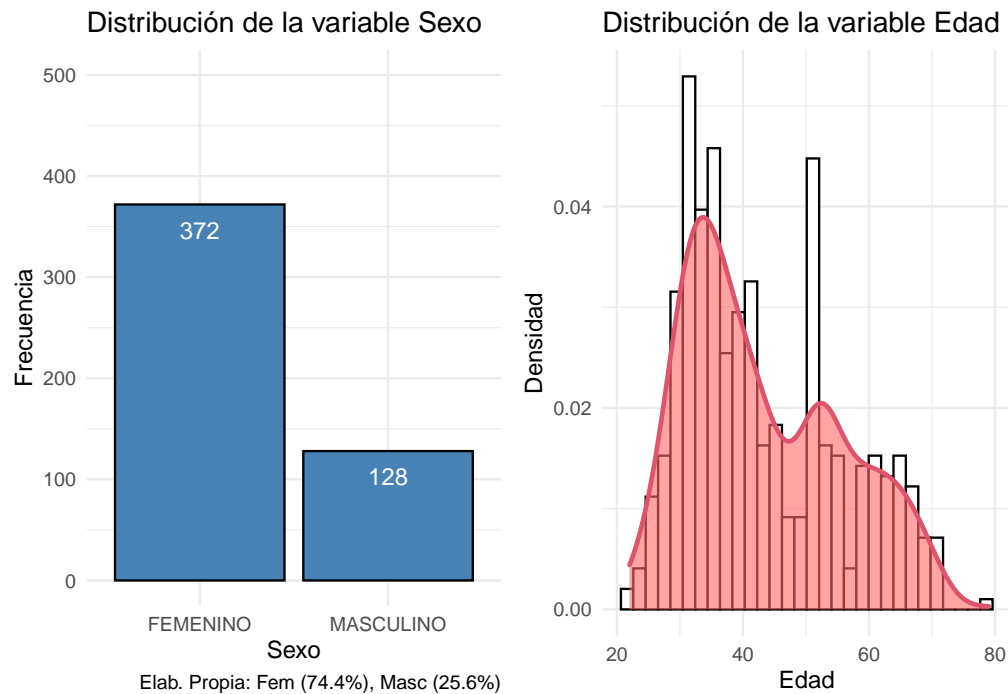
Table 10: Calidad Desempeño

CALIDAD_DESEMPEÑO	n	Promedio	Mediana	DesvEstandar	Mín	Máx	Cantidad_LM
CONTRATA	114	8.747007	6	6.813832	1.5	30	329
TITULAR	386	10.939705	7	7.851021	1.0	30	1246

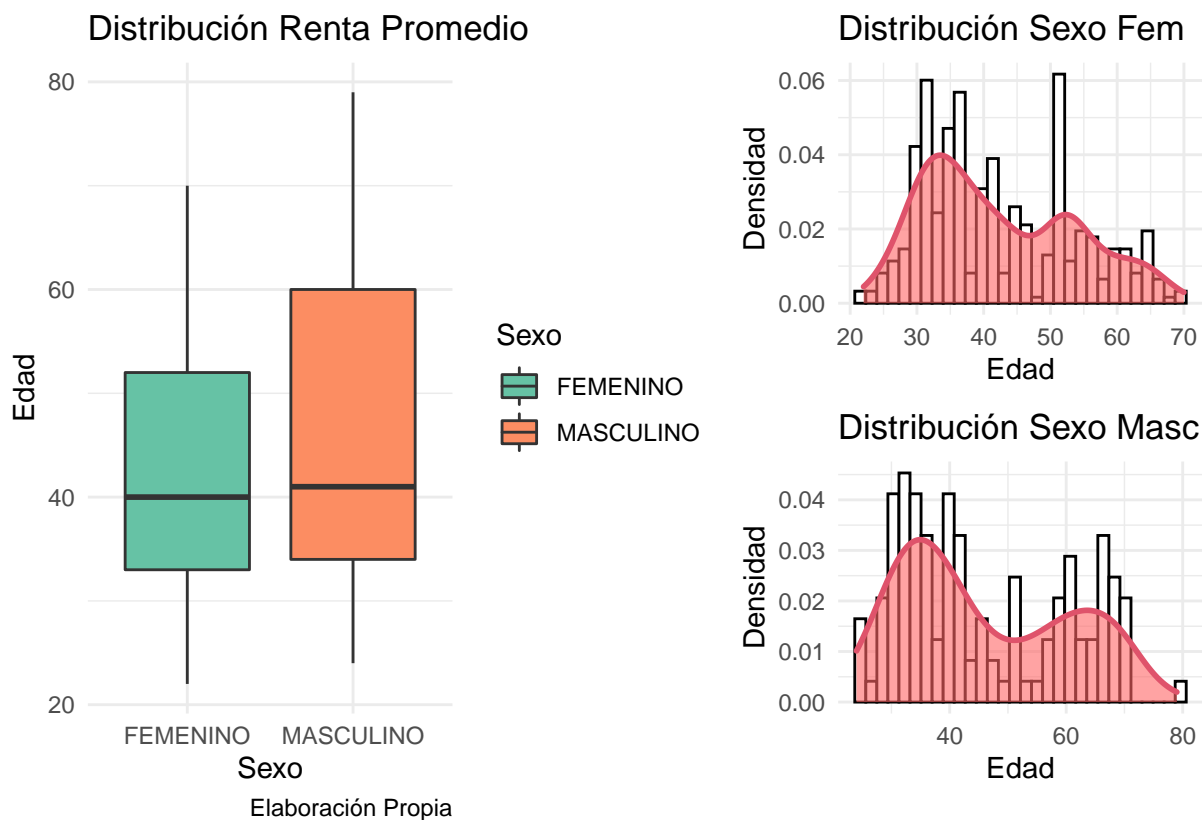
Table 11: Traslado Comuna

TRASLADO_COMUNA	n	Promedio	Mediana	DesvEstandar	Mín	Máx	Cantidad_LM
No	427	10.38741	7	7.704561	1	30	1343
Si	73	10.74601	7	7.550649	2	30	232

### Distribución de Licencias Médicas



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.	Mediana	Promedio	3rd Qu.	Max.
\$705.804	\$1.377.071	\$1.609.868	\$1.668.484	\$1.931.119	\$3.420.643

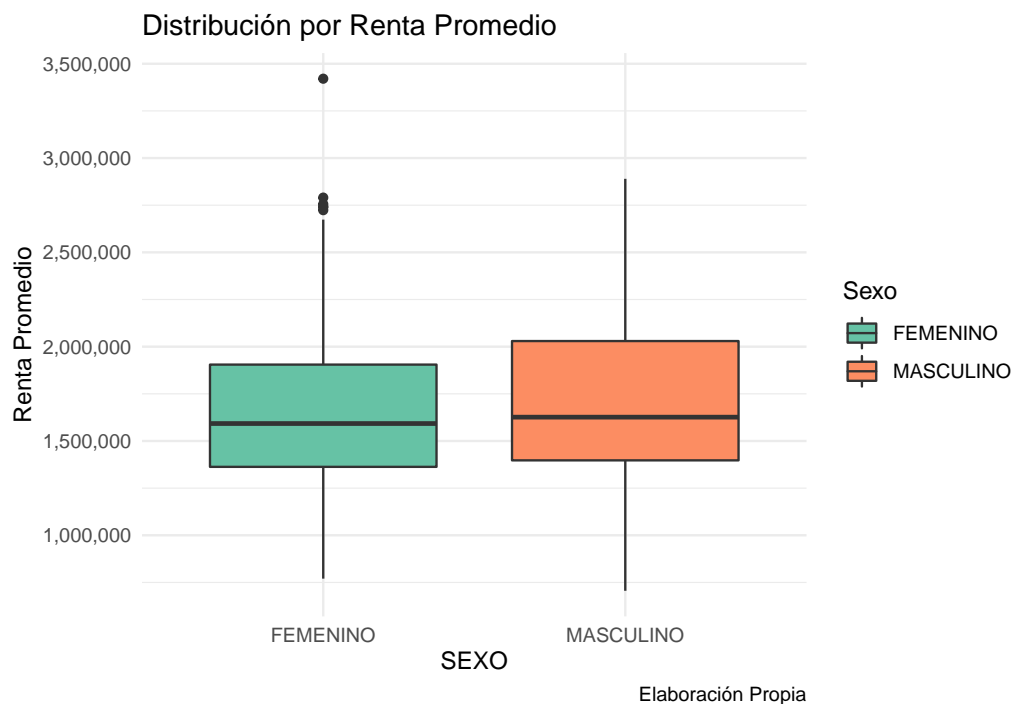
En este caso la distribución de la renta está centrada alrededor de 1,7 millones de pesos, donde el 25% más bajo de la distribución de renta está por debajo de los 1.37 millones, mientras que el 25% de los casos con mayores rentas se encuentran entre 1.93 y 3.42 millones de pesos.

#### Femenino

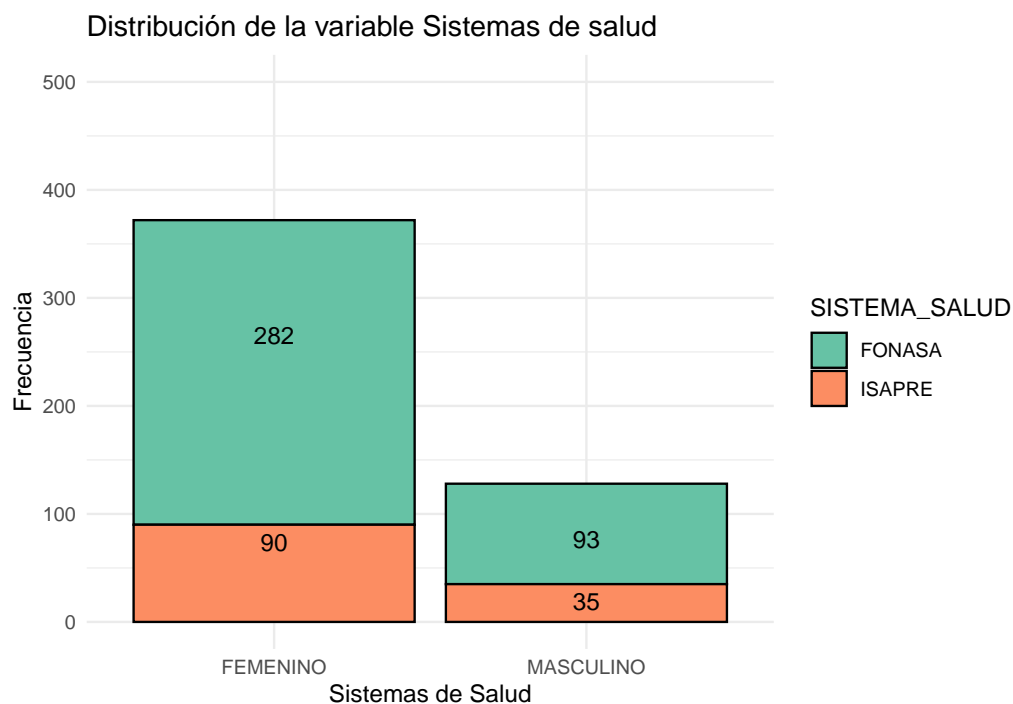
Min.	1st Qu.	Mediana	Promedio	3rd Qu.	Max.
\$770.197	\$1.362.827	\$1.592.526	\$1.652.053	\$1.904.939	\$3.420.643

#### Masculino

Min.	1st Qu.	Mediana	Promedio	3rd Qu.	Max.
\$705.804	\$1.397.245	\$1.626.350	\$1.741.390	\$2.029.702	\$3.233.292

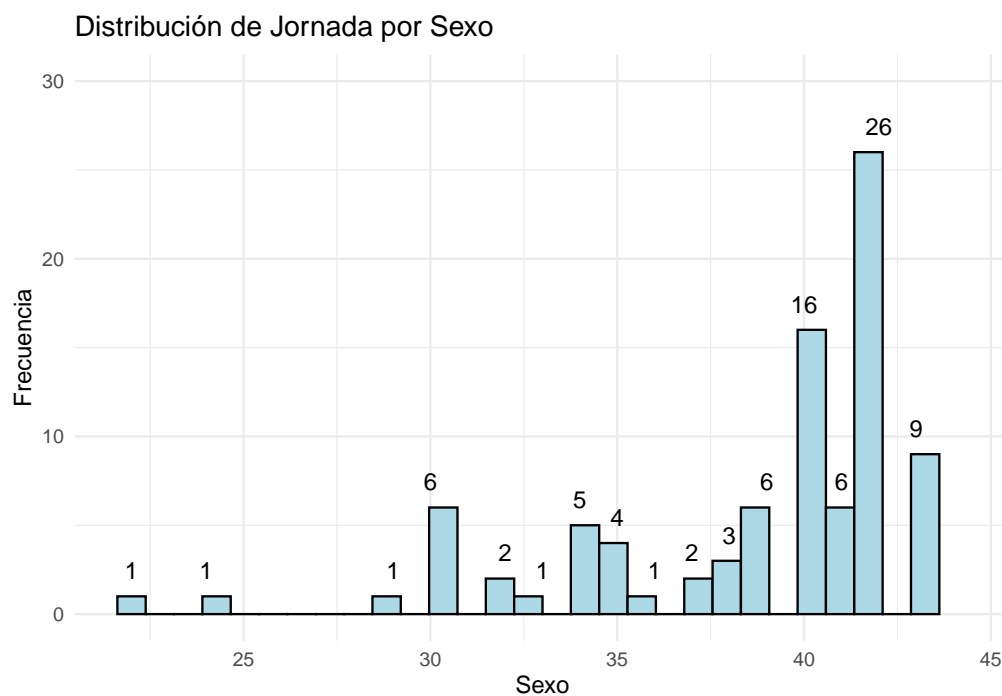


Comparando los niveles de renta por sexo se puede notar una cierta similitud en el rango de montos promedio, más aún, las estadísticas de posición (cuartiles) son bastante similares, por lo que podría no haber un efecto del sexo sobre los ingresos promedio. Para las mujeres la mediana de las rentas promedio es 1.58 millones, mientras que para los hombre bordea los 1.62 millones de pesos.

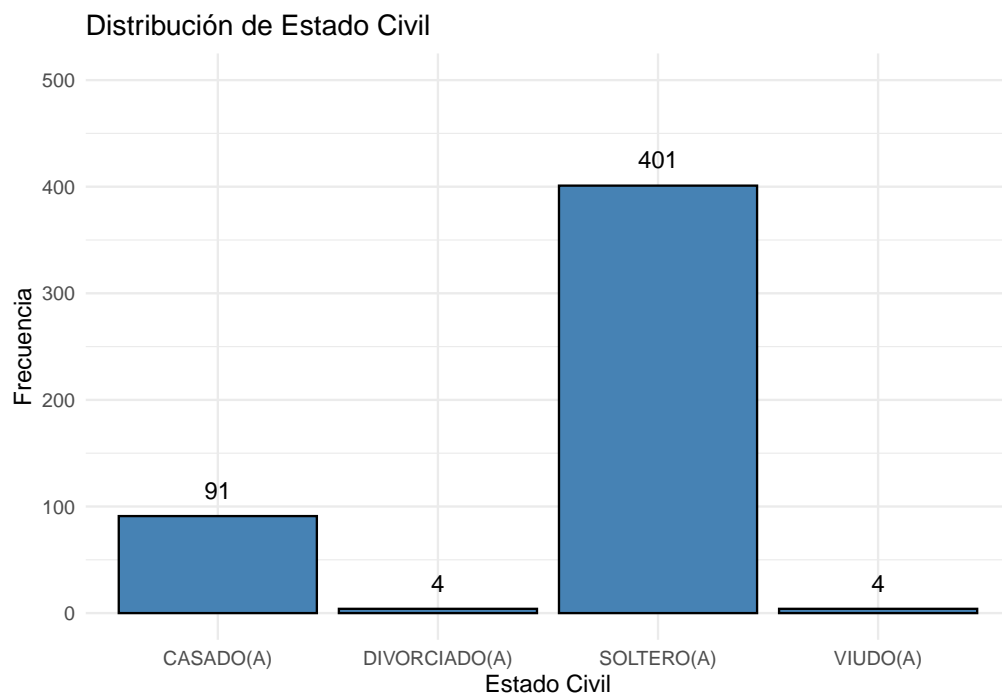


Según los datos, la proporción de personas que pertenecen al sistema de salud fonasa es alto tanto en hombres como en mujeres, esto tiene sentido desde el punto de vista que es una base de información de la provincia de Huasco donde existe una alta mayoría de zonas rurales donde el sistema fonasa es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

Min.	1st Qu.	Mediana	Promedio	3rd Qu.	Max.
22.00	44.00	44.00	43.05	44.00	44.00

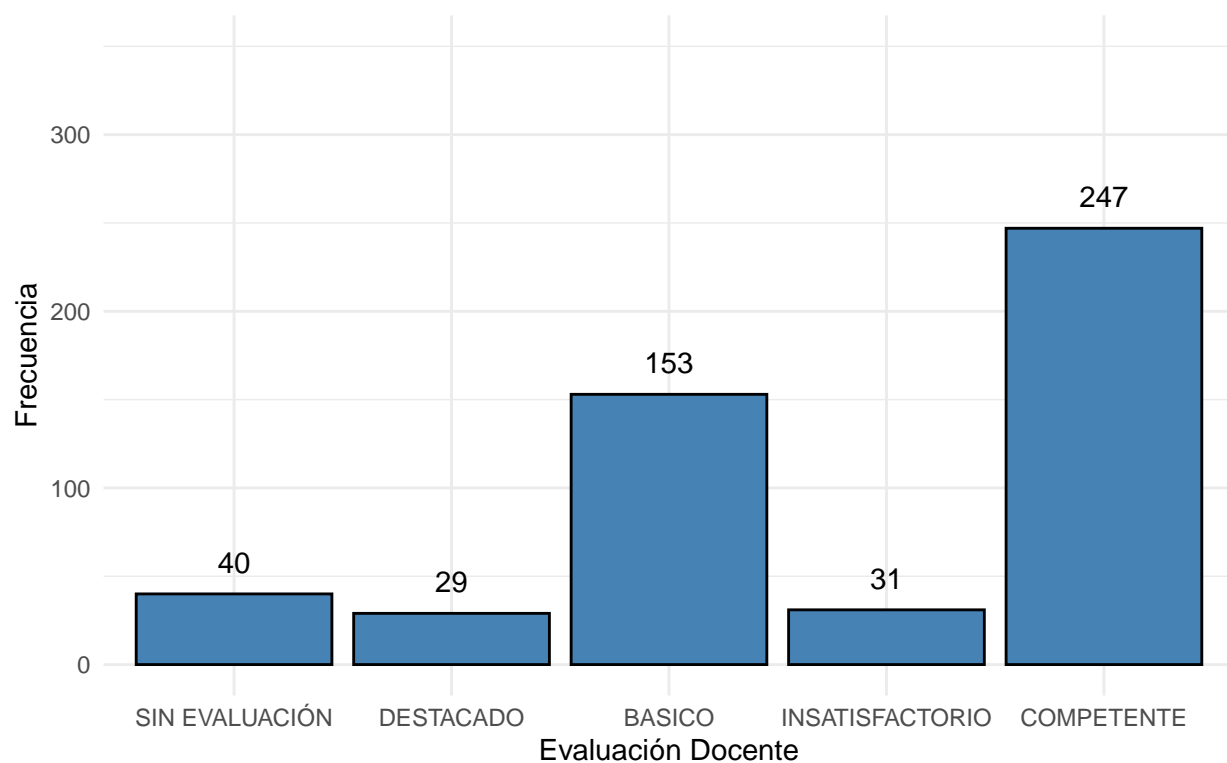


Podemos notar que prácticamente el 75% de los las personas de la muestra trabajan jornadas de 44 o menos horas semanales, mientras que sólo un caso muestra 88 horas laborales semanales, este caso se debe a .... (justificar el caso)



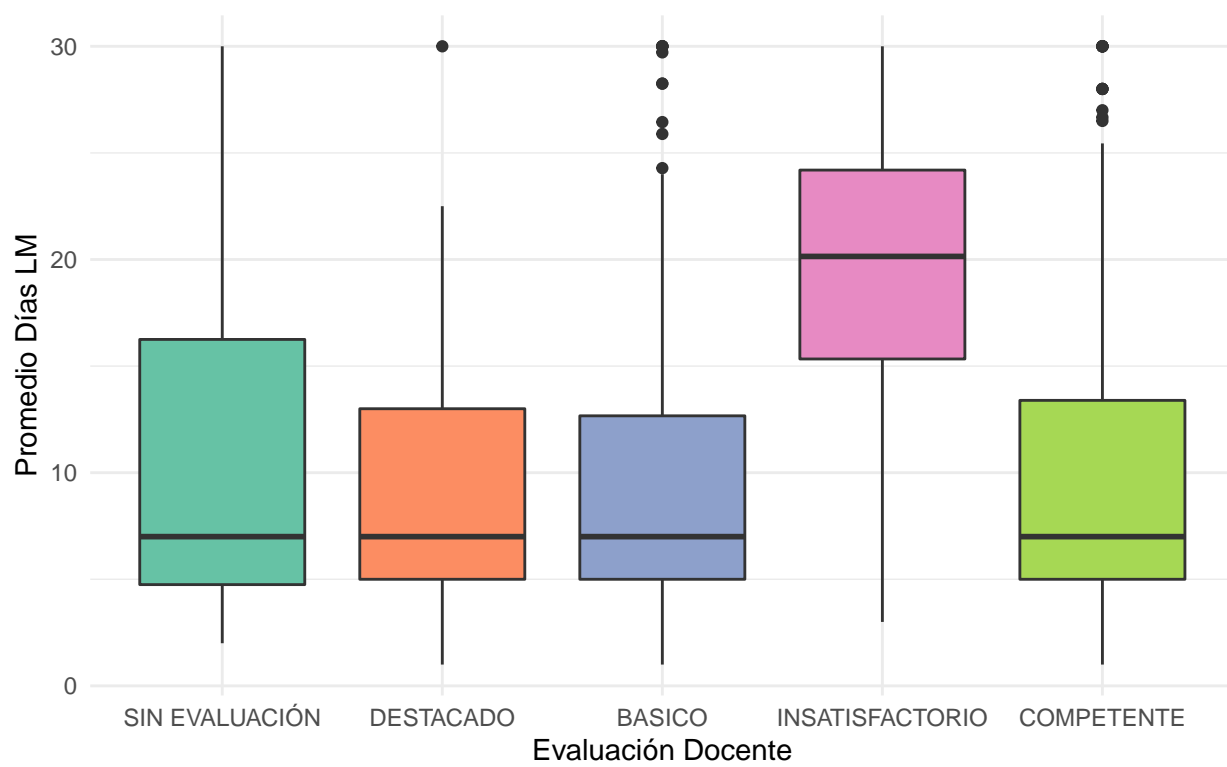
texto

Distribución de Evaluación Docente



Elaboración Propia

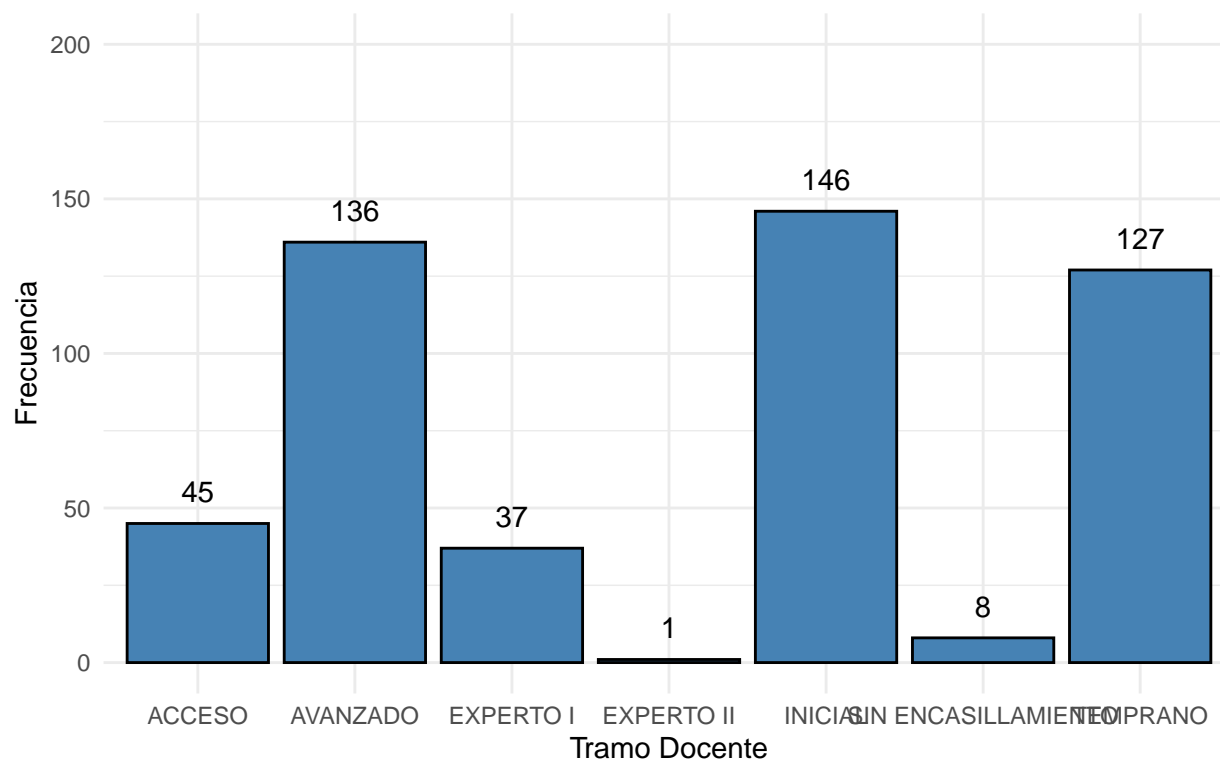
Distribución de Evaluación Docente



Elaboración Propia



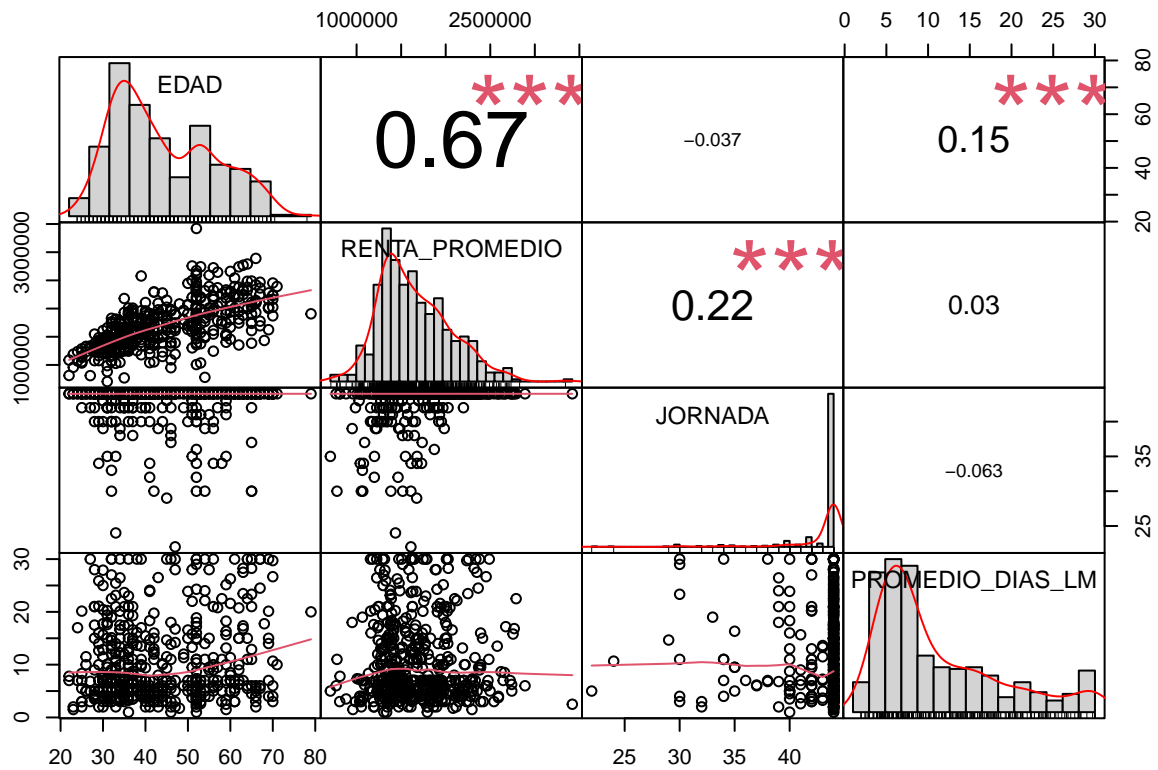
## Distribución de la variable Tramo Docente



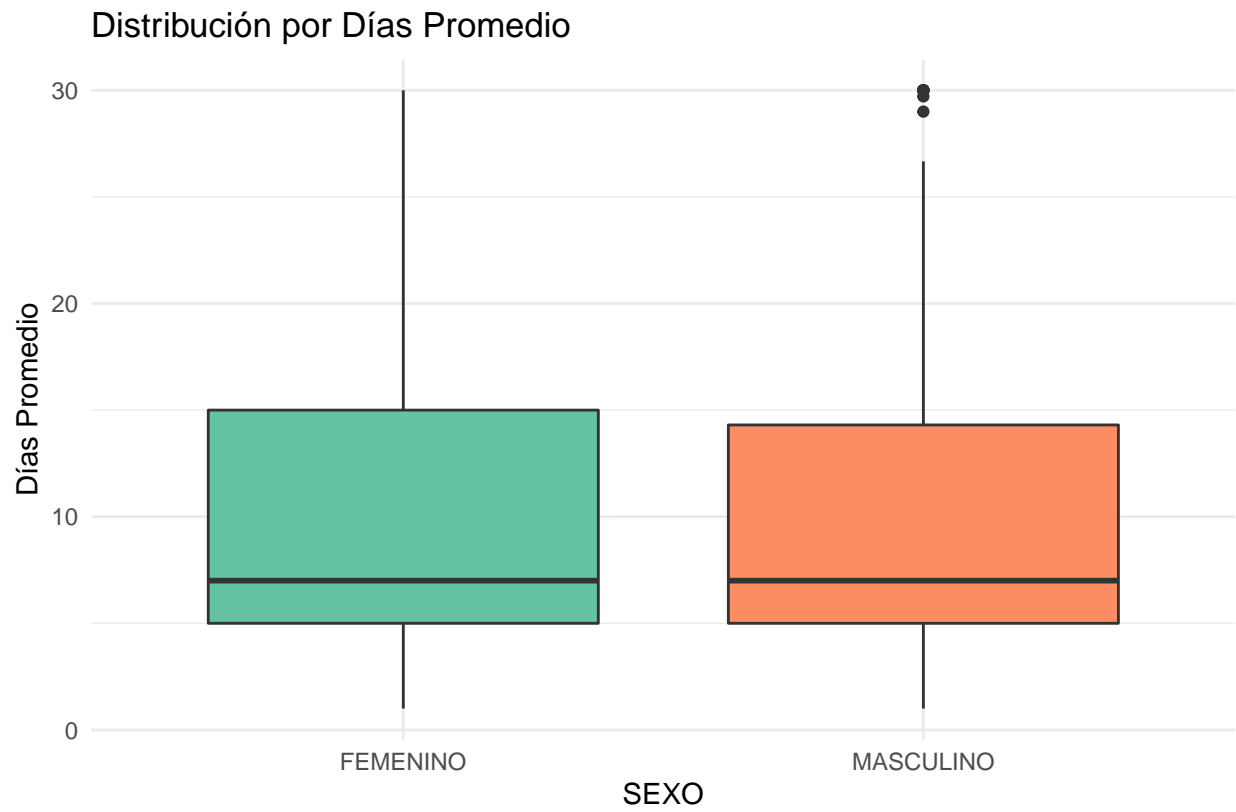
Elaboración Propia

## Correlación

```
## # A tibble: 6 x 25
##   ESCALAFON RUT_FUNCIONARIO NOMBRE      AP_P AP_M SEXO ESTADO_CIVIL EDAD
##   <chr>      <chr>          <chr>    <chr> <chr> <fct> <fct>      <dbl>
## 1 DOCENTE  16184521-3    MARCELA ALEJAN~ MEND~ GALL~ FEME~ SOLTERO(A)    36
## 2 DOCENTE  12940875-8    ROBERTO CESAR  GOYA  TORR~ MASC~ CASADO(A)    46
## 3 DOCENTE  8240042-7     NORMA ISABEL  COVA~ DIAZ  FEME~ SOLTERO(A)    64
## 4 DOCENTE  16111923-7    CLAUDIA STEPHA~ OJEDA ANGU~ FEME~ SOLTERO(A)    37
## 5 DOCENTE  9267343-K     TATIANA DE LOU~ LEDE~ ROJAS FEME~ CASADO(A)    61
## 6 DOCENTE  7680751-5     ESTELA EDIT   VELA~ ALVA~ FEME~ SOLTERO(A)    63
## # ... with 17 more variables: COMUNA_RESIDENCIA <chr>, RENTA_PROMEDIO <dbl>,
## #   SISTEMA_SALUD <fct>, JORNADA <dbl>, NIVEL <fct>, ESTABLECIMIENTO <chr>,
## #   TIPO_ESTABLECIMIENTO <fct>, COMUNA_ESTABLECIMIENTO <fct>,
## #   CALIDAD_DESEMPEÑO <fct>, ESTAMENTO <fct>, TRAMO_DOCENTE <fct>,
## #   EVALUACION_DOCENTE <fct>, CANTIDAD_LM <dbl>, PROMEDIO_DIAS_LM <dbl>,
## #   SUMA_DIAS <dbl>, TRASLADO_COMUNA <fct>, INSATISFACTORIO <fct>
```

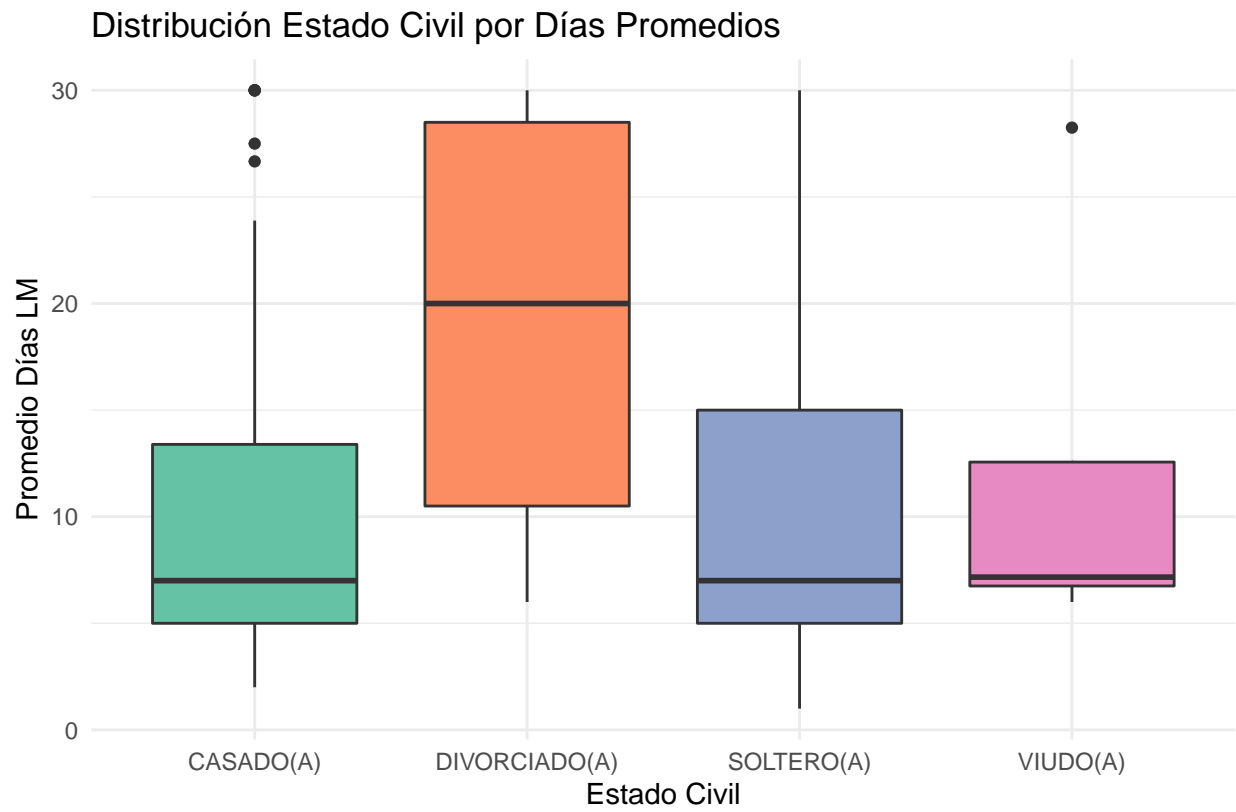


Al analizar las correlaciones entre las variables predictoras continuas y el número promedio de días con licencia se puede apreciar que la mayor correlación la presenta la variable edad con un coeficiente de 0.15 lo que sugiere posible efecto explicativo, seguido por la renta promedio con una correlación de 0.043, mientras que la jornada presente una correlación menor y negativa de -0.063.



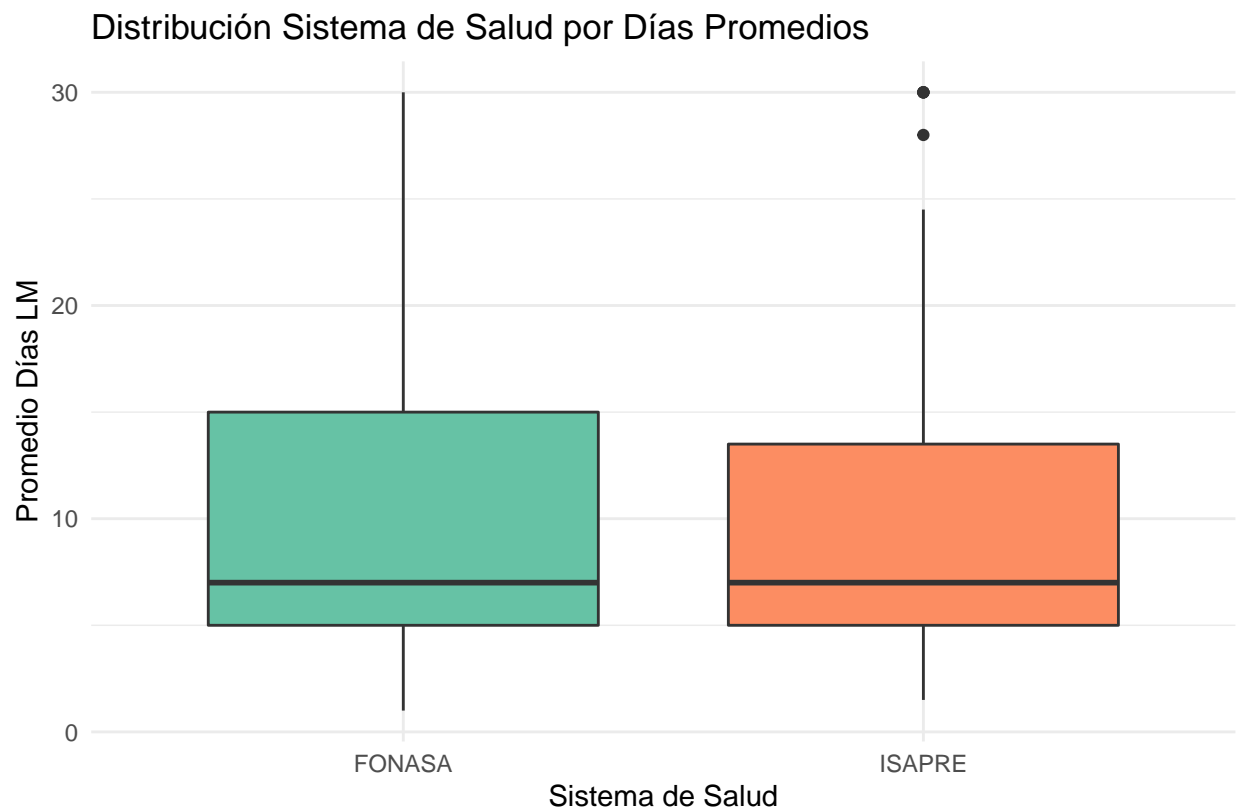
Elaboración Propia

En este caso, los boxplot no sugieren un efecto del sexo sobre la distribución del número de días con licencia promedio, dado que ambos soportes y estadísticas de posición son similares.



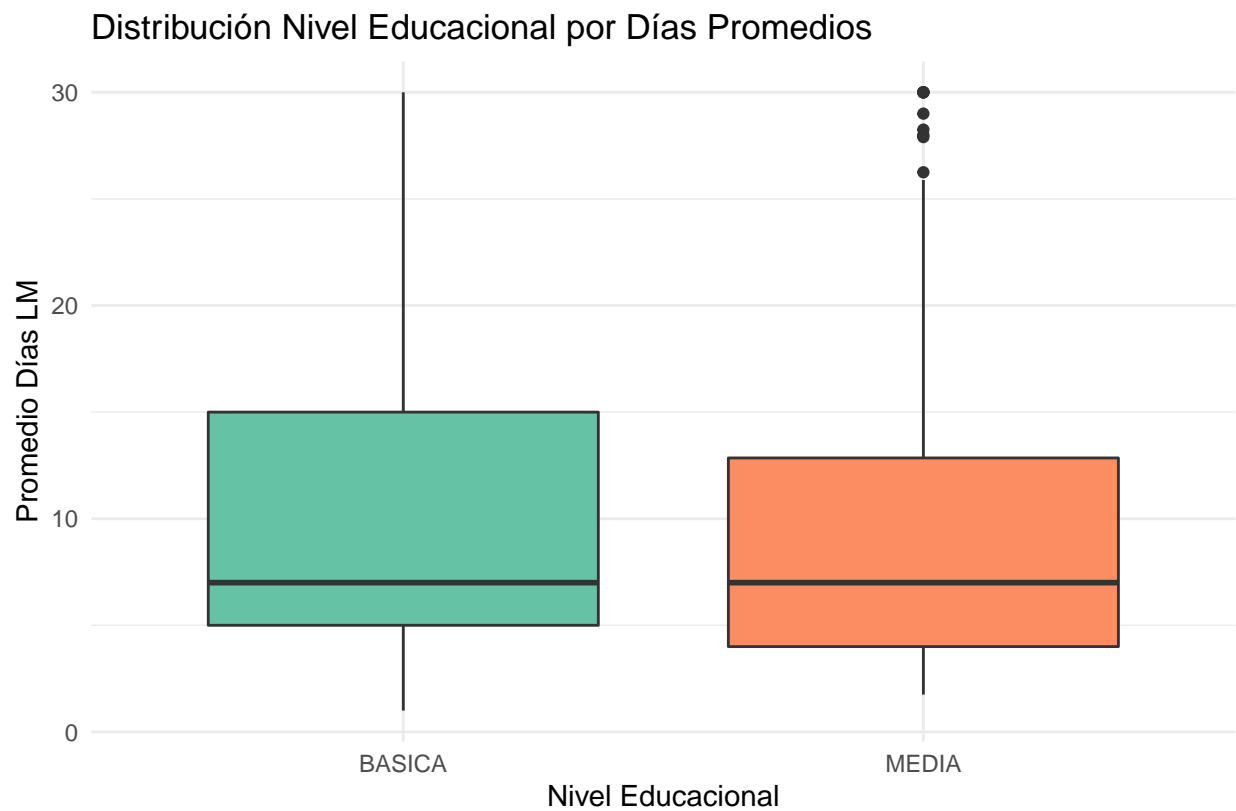
Elaboración Propia

De acuerdo al gráfico los datos sugieren que el nivel “Divorciado” podría tener un efecto en el número promedio de días con licencia, dado que su soporte es levemente superior al de los demás niveles.



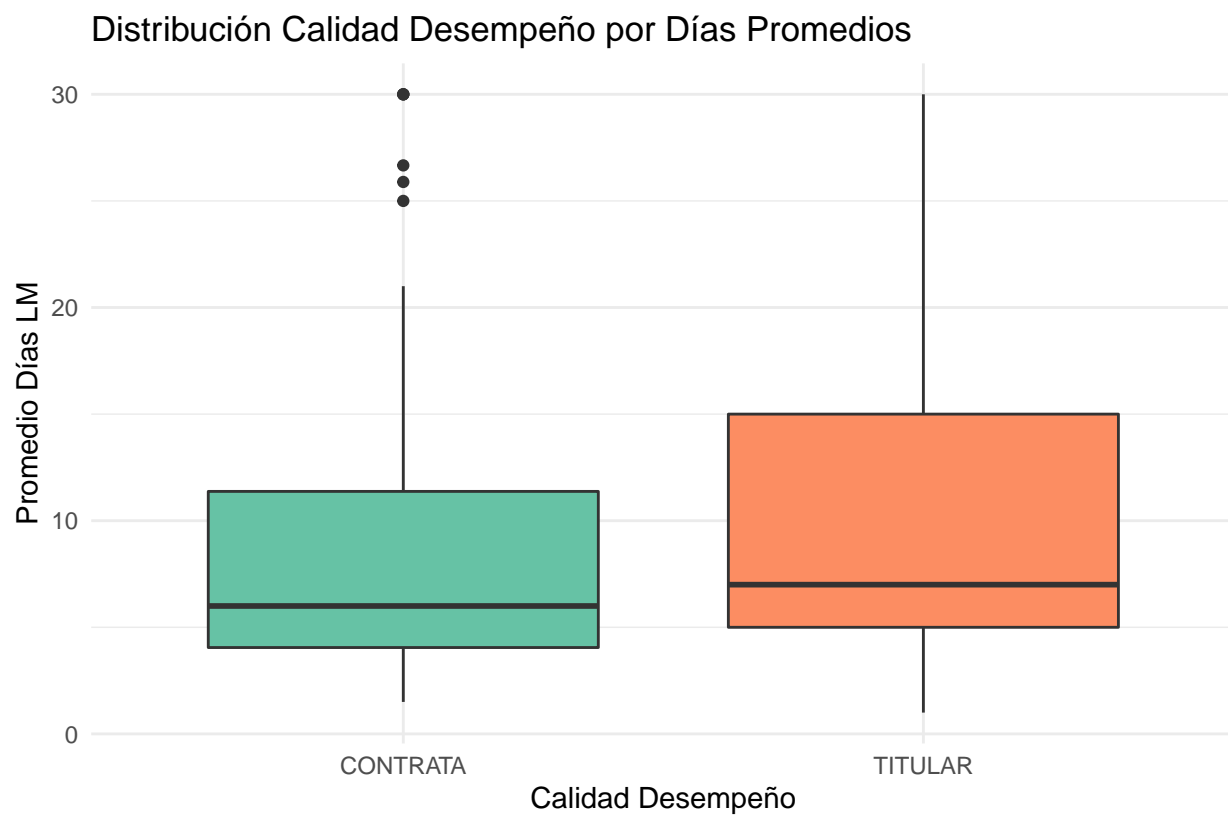
Elaboración Propia

Para el caso de los sistemas de salud, no se observan diferencias importantes en los soportes ni en las estadísticas de posición del número promedio de días con licencia, es decir, el sistema de salud podría no tener un efecto significativo.



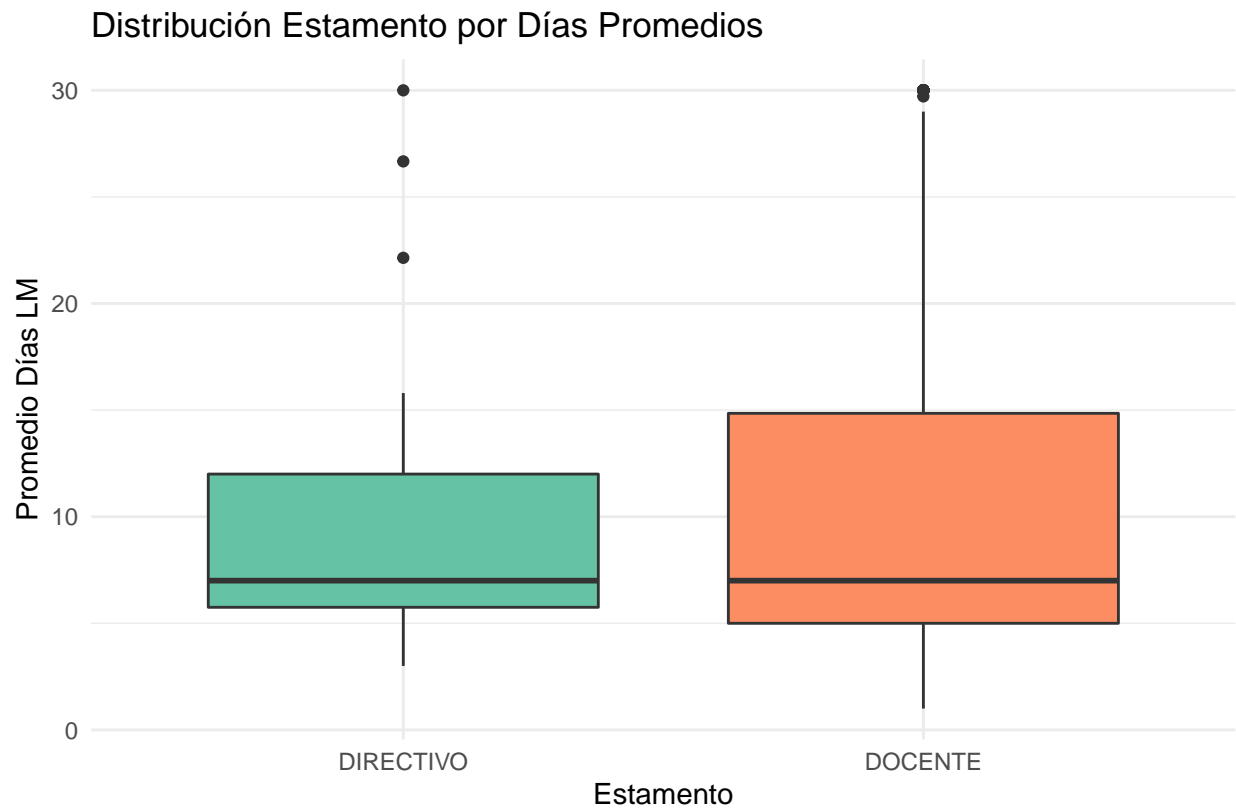
Elaboración Propia

Al observar la distribución del número promedio de días con licencia por nivel educativo no se observan grandes diferencias, lo que sugiere que este factor no presenta un grado de asociación significativo.



Elaboración Propia

En este caso, el ser docente titular presenta un soporte levemente mayor que aquel docente con modalidad contrata, lo que puede sugerir un grado de asociación de este factor con el número promedio de días con licencia.

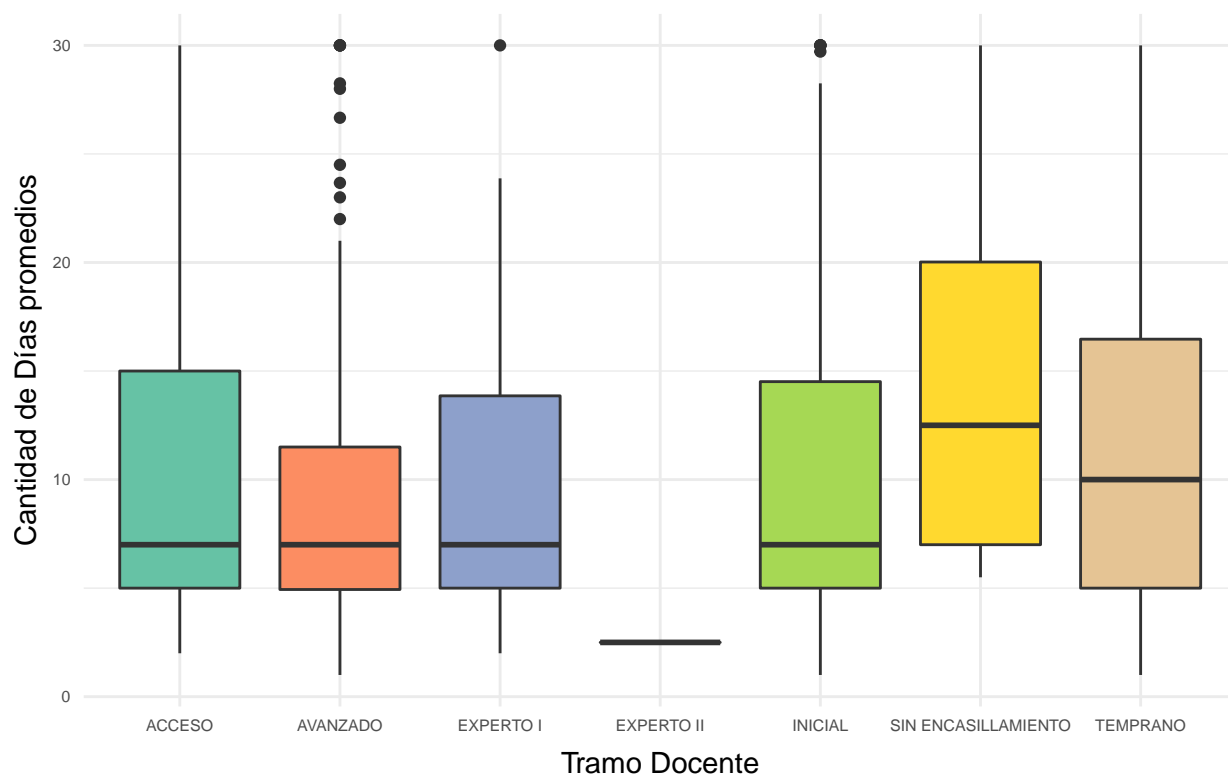


Elaboración Propia

Analizando el factor “Estamento” se puede notar que aquel docente que sólo presneta estamento de docente tiene mayor soporte que aquel que tiene estamento de directivo, lo que sugiere algún grado de asociación de este factor con el número promedio de días de licencias.

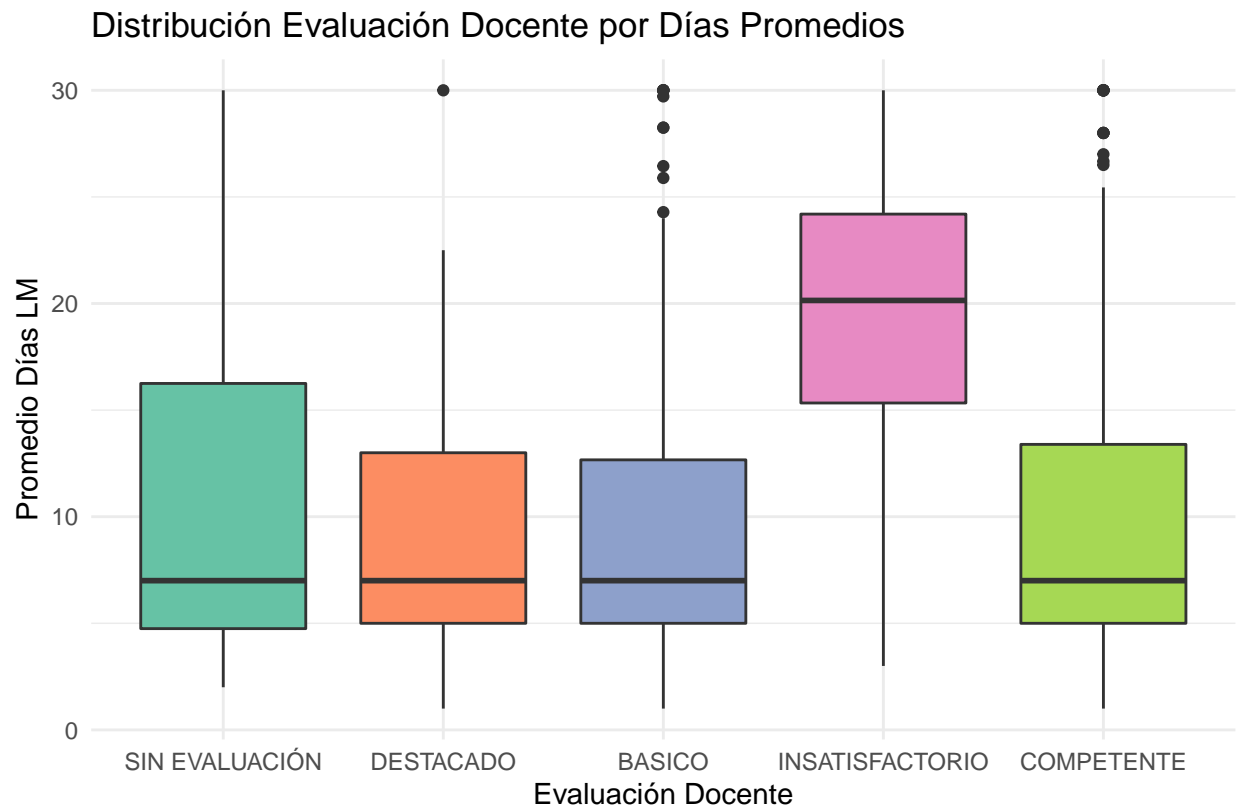


## Distribución de los días promedio según la Tramo Docente



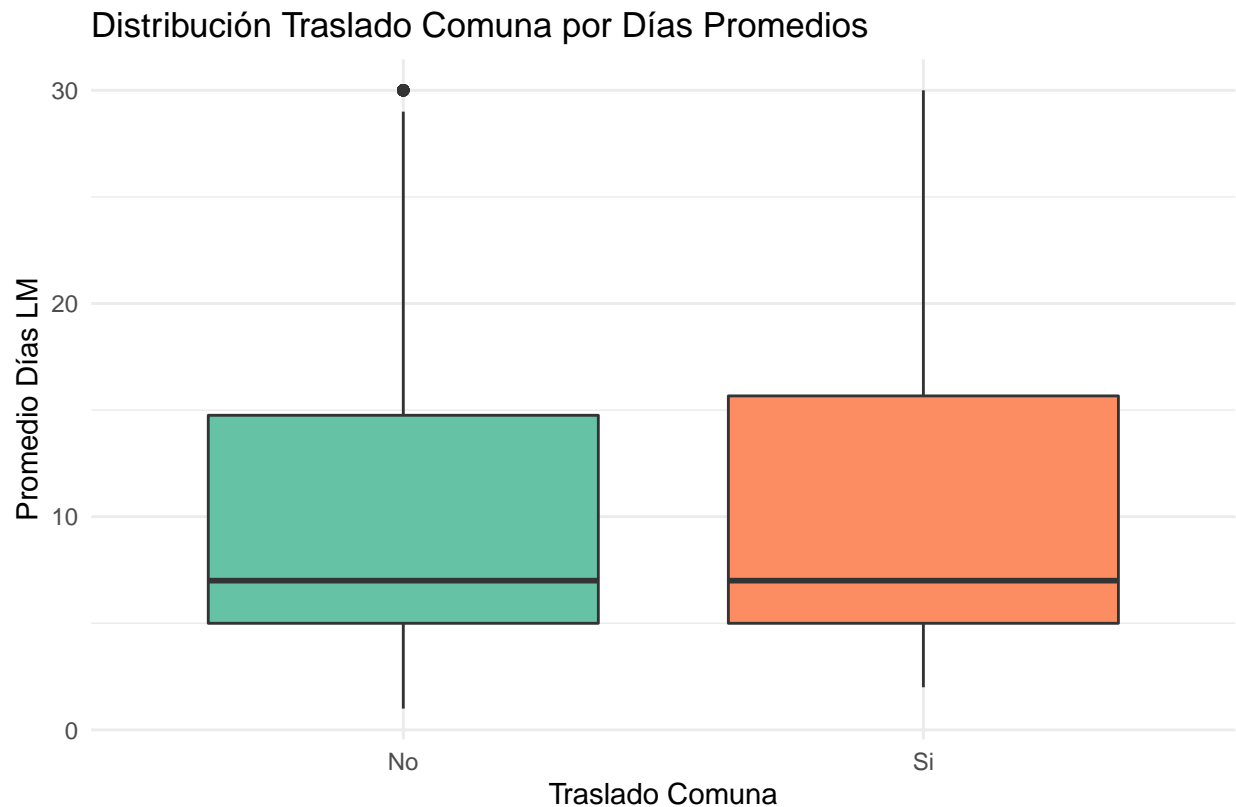
Elab. propia: BBDD Lic. Médicas 2022

En este caso los boxplot de los días promedio de licencias analizados por los niveles del tramo docente muestran cambios en sus soportes lo que puede sugerir algún asociación de este factor.



Elaboración Propia

De acuerdo al gráfico se aprecia una distribución de valores centrales diferente en el nivel “insatisfactorio” del factor “Evaluación docente”, mientras que el resto de los niveles presentan soportes y distribuciones centrales similares, esto sugiere que se podría crear una variable dummy que tome valor uno cuando la evaluación docente es insatisfactorio o cero en otro caso, pero se tendría que evaluar si produce los mismos asociaciones sobre el modelo que se proponga.



Elaboración Propia

Respecto a la distribución del número de días con licencia de acuerdo al traslado, los soportes y distribuciones son similares, por lo cual no se sugiere alguna asociación de este factor sobre la variable respuesta (número de días promedio con licencias)

## Selección Formal de Modelo

### Forward

Para la selección del modelo forward se utilizó una significancia del 5% en cada uno de los tests 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>			29400	2039.1		
SEXO	1	4.29	29396	2041.0	0.0727	0.7875728
ESTADO_CIVIL	3	324.01	29076	2039.5	1.8424	0.1385146
EDAD	1	701.41	28699	2029.0	12.1712	0.0005284 ***
RENTA_PROMEDIO	1	26.17	29374	2040.6	0.4437	0.5056665
SISTEMA_SALUD	1	135.72	29265	2038.8	2.3095	0.1292163

```
## JORNADA          1      116.36 29284 2039.1   1.9789 0.1601349
## NIVEL            1      232.20 29168 2037.1   3.9645 0.0470156 *
## CALIDAD_DESEMPEÑO 1      423.14 28977 2033.8   7.2720 0.0072411 **
## ESTAMENTO        1         0.42 29400 2041.1   0.0072 0.9325123
## EVALUACION_DOCENTE 4      2732.44 26668 1998.3  12.6796 7.858e-10 ***
## TRAMO_DOCENTE     6       624.90 28776 2040.3   1.7844 0.1003511
## TRASLADO_COMUNA   1         8.02 29392 2040.9   0.1358 0.7126186
## ---
## 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

## Categoriza la edad

## Metodología Forward

## Primera variable que entra al modelo

```
## Single term additions
##
## Model:
## PROMEDIO_DIAS_LM ~ 1
##
##      Df Sum of Sq  RSS    AIC F value    Pr(>F)
## <none>                 29400 2039.1
## SEXO          1         4.29 29396 2041.0   0.0727   0.787573
## ESTADO_CIVIL  3       324.01 29076 2039.5   1.8424   0.138515
## TRAMO_EDAD     1       593.46 28807 2030.9  10.2594   0.001447 **
## RENTA_PROMEDIO 1        26.17 29374 2040.6   0.4437   0.505667
## JORNADA         1       116.36 29284 2039.1   1.9789   0.160135
## NIVEL           1       232.20 29168 2037.1   3.9645   0.047016 *
## TIPO_ESTABLECIMIENTO 3       275.60 29125 2040.4   1.5645   0.197107
## CALIDAD_DESEMPEÑO 1       423.14 28977 2033.8   7.2720   0.007241 **
## TRAMO_DOCENTE   6       624.90 28776 2040.3   1.7844   0.100351
## INSATISFACTORIO 1      2663.50 26737 1993.6  49.6103  6.265e-12 ***
## TRASLADO_COMUNA 1         8.02 29392 2040.9   0.1358   0.712619
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

## Segunda variable que entra al modelo

```
## Single term additions
##
## Model:
## PROMEDIO_DIAS_LM ~ INSATISFACTORIO
##
```

	Df	Sum of Sq	RSS	AIC	F value	Pr(>F)
<none>			26737	1993.6		
SEXO	1	16.80	26720	1995.3	0.3125	0.5764099
ESTADO_CIVIL	3	208.80	26528	1995.7	1.2987	0.2741730
TRAMO_EDAD	1	554.49	26182	1985.1	10.5255	0.0012567 **
RENTA_PROMEDIO	1	100.31	26637	1993.7	1.8716	0.1719044
JORNADA	1	22.72	26714	1995.2	0.4227	0.5159075
NIVEL	1	245.36	26492	1991.0	4.6031	0.0323988 *
TIPO_ESTABLECIMIENTO	3	274.52	26462	1994.4	1.7117	0.1636561
CALIDAD_DESEMPEÑO	1	661.75	26075	1983.1	12.6131	0.0004195 ***
TRAMO_DOCENTE	6	609.22	26128	1994.1	1.9120	0.0771500 .
TRASLADO_COMUNA	1	1.92	26735	1995.6	0.0356	0.8503823

```
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

## Tercera variable que entra al modelo

```
## Single term additions
##
## Model:
## PROMEDIO_DIAS_LM ~ INSATISFACTORIO + CALIDAD_DESEMPEÑO
##
```

	Df	Sum of Sq	RSS	AIC	F value	Pr(>F)
<none>			26075	1983.1		
SEXO	1	26.02	26049	1984.6	0.4955	0.48182
ESTADO_CIVIL	3	202.91	25872	1985.2	1.2914	0.27663
TRAMO_EDAD	1	313.67	25762	1979.0	6.0392	0.01433 *
RENTA_PROMEDIO	1	0.96	26074	1985.0	0.0183	0.89249
JORNADA	1	25.33	26050	1984.6	0.4822	0.48774
NIVEL	1	193.66	25882	1981.3	3.7114	0.05461 .
TIPO_ESTABLECIMIENTO	3	264.42	25811	1984.0	1.6870	0.16888
TRAMO_DOCENTE	6	586.07	25489	1983.7	1.8816	0.08219 .
TRASLADO_COMUNA	1	37.66	26038	1984.3	0.7174	0.39739

```
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

## Cuarta variable que entra al modelo

```
## Single term additions
##
## Model:
## PROMEDIO_DIAS_LM ~ INSATISFACTORIO + CALIDAD_DESEMPEÑO + TRAMO_EDAD
##
```

	Df	Sum of Sq	RSS	AIC	F value	Pr(>F)
<none>			25762	1979.0		
SEXO	1	38.78	25723	1980.3	0.7464	0.38805
ESTADO_CIVIL	3	221.41	25540	1980.7	1.4246	0.23475
RENTA_PROMEDIO	1	120.19	25641	1978.7	2.3202	0.12834
JORNADA	1	17.92	25744	1980.7	0.3446	0.55747

```
## NIVEL 1 207.70 25554 1977.0 4.0234 0.04542 *
## TIPO_ESTABLECIMIENTO 3 230.84 25531 1980.5 1.4859 0.21752
## TRAMO_DOCENTE 6 535.74 25226 1980.5 1.7344 0.11106
## TRASLADO_COMUNA 1 35.19 25726 1980.3 0.6771 0.41099
## ---
## Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

## Finalmente

```
## Single term additions
##
## Model:
## PROMEDIO_DIAS_LM ~ INSATISFACTORIO + CALIDAD_DESEMPEÑO + TRAMO_EDAD +
## NIVEL
## Df Sum of Sq RSS AIC F value Pr(>F)
## <none> 25554 1977.0
## SEXO 1 10.37 25543 1978.8 0.2005 0.6545
## ESTADO_CIVIL 3 189.19 25365 1979.2 1.2233 0.3006
## RENTA_PROMEDIO 1 86.24 25468 1977.3 1.6727 0.1965
## JORNADA 1 18.13 25536 1978.6 0.3508 0.5539
## TIPO_ESTABLECIMIENTO 3 298.04 25256 1977.1 1.9353 0.1229
## TRAMO_DOCENTE 6 528.93 25025 1978.5 1.7226 0.1137
## TRASLADO_COMUNA 1 29.23 25524 1978.4 0.5657 0.4523
```

## Modelo resultante por forward

Para decidir la entrada del segundo predictor se debe evaluar todos los modelos que contienen cuya variable respuesta es PROMEDIO\_DIAS\_LM agregando cada uno de los predictores por separado, para lo cual es necesario desarrollar un test F por predictor

Modelo	p-valor
PROMEDIO_DIAS_LM ~ SEXO	0.787573
PROMEDIO_DIAS_LM ~ ESTADO_CIVIL	0.138515
PROMEDIO_DIAS_LM ~ TRAMO_EDAD	0.001447
PROMEDIO_DIAS_LM ~ RENTA_PROMEDIO	0.505667
PROMEDIO_DIAS_LM ~ JORNADA	0.160135
PROMEDIO_DIAS_LM ~ NIVEL	0.047016
PROMEDIO_DIAS_LM ~ TIPO_ESTABLECIMIENTO	0.197107
PROMEDIO_DIAS_LM ~ CALIDAD_DESEMPEÑO	0.007241
PROMEDIO_DIAS_LM ~ TRAMO_DOCENTE	0.100351
PROMEDIO_DIAS_LM ~ INSATISFACTORIO	6.265e-12
PROMEDIO_DIAS_LM ~ TRASLADO_COMUNA	0.712619

## Primera Variable que ingresa al modelo

Coefficientes	Estimación	Std. Error	t value	Pr(>t)
Intercepto	7.9831	0.7314	10.915	< 2e-16
INSATISFACTORIOInsatisfactorio	9.9027	1.3399	7.390	6.28e-13
CALIDAD_DESEMPEÑOTITULAR	2.1505	0.7967	2.699	0.00718
TRAMO_EDAD[50.5,79.1)	1.7839	0.7080	2.520	0.01207

Coeficientes	Estimación	Std. Error	t value	Pr(>t)
NIVELMEDIA	-1.4459	0.7208	-2.006	0.04542

## Metodología Backward

### Primera variable que sale al modelo

```
## Single term deletions
##
## Model:
## PROMEDIO_DIAS_LM ~ SEXO + ESTADO_CIVIL + TRAMO_EDAD + RENTA_PROMEDIO +
##     JORNADA + NIVEL + TIPO_ESTABLECIMIENTO + CALIDAD_DESEMPEÑO +
##     TRAMO_DOCENTE + INSATISFACTORIO + TRASLADO_COMUNA
##           Df Sum of Sq  RSS   AIC F value    Pr(>F)
## <none>                 24491 1987.7
## SEXO                   1    11.59 24502 1986.0  0.2266  0.634258
## ESTADO_CIVIL           3   148.82 24640 1984.8  0.9702  0.406513
## TRAMO_EDAD              1   171.82 24662 1989.2  3.3605  0.067400 .
## RENTA_PROMEDIO          1     1.47 24492 1985.8  0.0288  0.865250
## JORNADA                 1     5.96 24497 1985.8  0.1165  0.732987
## NIVEL                   1   302.72 24793 1991.9  5.9206  0.015329 *
## TIPO_ESTABLECIMIENTO    3   327.11 24818 1988.3  2.1326  0.095266 .
## CALIDAD_DESEMPEÑO       1   437.75 24928 1994.6  8.5618  0.003596 **
## TRAMO_DOCENTE           6   474.48 24965 1985.3  1.5467  0.161072
## INSATISFACTORIO         1  2717.68 27208 2038.3 53.1536 1.284e-12 ***
## TRASLADO_COMUNA         1    19.82 24510 1986.1  0.3877  0.533822
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

### Segunda variable que sale al modelo

```
## Single term deletions
##
## Model:
## PROMEDIO_DIAS_LM ~ SEXO + ESTADO_CIVIL + TRAMO_EDAD + JORNADA +
##     NIVEL + TIPO_ESTABLECIMIENTO + CALIDAD_DESEMPEÑO + TRAMO_DOCENTE +
##     INSATISFACTORIO + TRASLADO_COMUNA
##           Df Sum of Sq  RSS   AIC F value    Pr(>F)
## <none>                 24492 1985.8
## SEXO                   1    11.32 24504 1984.0  0.2219  0.637826
## ESTADO_CIVIL           3   147.46 24640 1982.8  0.9633  0.409815
## TRAMO_EDAD              1   286.92 24779 1989.6  5.6232  0.018118 *
## JORNADA                 1     4.76 24497 1983.8  0.0933  0.760158
## NIVEL                   1   307.73 24800 1990.0  6.0309  0.014411 *
## TIPO_ESTABLECIMIENTO    3   332.64 24825 1986.5  2.1730  0.090378 .
## CALIDAD_DESEMPEÑO       1   458.17 24950 1993.0  8.9792  0.002872 **
## TRAMO_DOCENTE           6   533.36 25026 1984.5  1.7421  0.109403
## INSATISFACTORIO         1  2718.79 27211 2036.4 53.2830 1.207e-12 ***
## TRASLADO_COMUNA         1    20.07 24512 1984.2  0.3934  0.530822
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

## Tercera variable que sale al modelo

```
## Single term deletions
##
## Model:
## PROMEDIO_DIAS_LM ~ SEXO + ESTADO_CIVIL + TRAMO_EDAD + NIVEL +
##      TIPO_ESTABLECIMIENTO + CALIDAD_DESEMPEÑO + TRAMO_DOCENTE +
##      INSATISFACTORIO + TRASLADO_COMUNA
##
##      Df Sum of Sq  RSS    AIC F value    Pr(>F)
## <none>                24497 1983.8
## SEXO                  1    11.92 24509 1982.1  0.2341  0.628711
## ESTADO_CIVIL          3   147.49 24644 1980.8  0.9653  0.408862
## TRAMO_EDAD             1   292.13 24789 1987.8  5.7359  0.017003 *
## NIVEL                  1   307.24 24804 1988.1  6.0326  0.014396 *
## TIPO_ESTABLECIMIENTO  3   333.31 24830 1984.6  2.1815  0.089381 .
## CALIDAD_DESEMPEÑO     1   457.54 24954 1991.1  8.9838  0.002865 **
## TRAMO_DOCENTE          6   542.25 25039 1982.8  1.7745  0.102443
## INSATISFACTORIO       1  2765.72 27263 2035.3 54.3051 7.536e-13 ***
## TRASLADO_COMUNA       1    20.51 24518 1982.3  0.4026  0.526034
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

## Cuarta variable que sale al modelo

```
## Single term deletions
##
## Model:
## PROMEDIO_DIAS_LM ~ ESTADO_CIVIL + TRAMO_EDAD + NIVEL + TIPO_ESTABLECIMIENTO +
##      CALIDAD_DESEMPEÑO + TRAMO_DOCENTE + INSATISFACTORIO + TRASLADO_COMUNA
##
##      Df Sum of Sq  RSS    AIC F value    Pr(>F)
## <none>                24509 1982.1
## ESTADO_CIVIL          3   157.48 24666 1979.3  1.0323  0.377914
## TRAMO_EDAD             1   287.30 24796 1985.9  5.6502  0.017843 *
## NIVEL                  1   312.95 24822 1986.4  6.1547  0.013447 *
## TIPO_ESTABLECIMIENTO  3   330.55 24839 1982.8  2.1669  0.091098 .
## CALIDAD_DESEMPEÑO     1   447.50 24956 1989.1  8.8008  0.003161 **
## TRAMO_DOCENTE          6   539.89 25049 1981.0  1.7696  0.103470
## INSATISFACTORIO       1  2754.53 27263 2033.3 54.1715 7.988e-13 ***
## TRASLADO_COMUNA       1    17.54 24526 1980.5  0.3449  0.557292
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

## Quinta variable que sale al modelo

```
## Single term deletions
##
## Model:
## PROMEDIO_DIAS_LM ~ ESTADO_CIVIL + TRAMO_EDAD + NIVEL + TIPO_ESTABLECIMIENTO +
##      CALIDAD_DESEMPEÑO + TRAMO_DOCENTE + INSATISFACTORIO
##
##      Df Sum of Sq  RSS    AIC F value    Pr(>F)
## <none>                24526 1980.5
## ESTADO_CIVIL          3   163.56 24690 1977.8  1.0736  0.359856
## TRAMO_EDAD             1   288.83 24815 1984.3  5.6880  0.017466 *
```



```
## NIVEL 1 319.15 24846 1984.9 6.2851 0.012503 *
## TIPO_ESTABLECIMIENTO 3 333.61 24860 1981.2 2.1899 0.088400 .
## CALIDAD_DESEMPEÑO 1 431.44 24958 1987.2 8.4964 0.003724 **
## TRAMO_DOCENTE 6 538.89 25065 1979.3 1.7687 0.103651
## INSATISFACTORIO 1 2786.32 27313 2032.2 54.8712 5.785e-13 ***
## ---
## Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

## Sexta variable que sale al modelo

```
## Single term deletions
##
## Model:
## PROMEDIO_DIAS_LM ~ TRAMO_EDAD + NIVEL + TIPO_ESTABLECIMIENTO +
## CALIDAD_DESEMPEÑO + TRAMO_DOCENTE + INSATISFACTORIO
## Df Sum of Sq RSS AIC F value Pr(>F)
## <none> 24690 1977.8
## TRAMO_EDAD 1 284.70 24975 1981.5 5.6041 0.018309 *
## NIVEL 1 326.39 25016 1982.3 6.4246 0.011567 *
## TIPO_ESTABLECIMIENTO 3 334.85 25025 1978.5 2.1971 0.087573 .
## CALIDAD_DESEMPEÑO 1 440.40 25130 1984.6 8.6689 0.003392 **
## TRAMO_DOCENTE 6 565.74 25256 1977.1 1.8560 0.086672 .
## INSATISFACTORIO 1 2892.17 27582 2031.2 56.9298 2.242e-13 ***
## ---
## Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

## Septima variable que sale al modelo

```
## Single term deletions
##
## Model:
## PROMEDIO_DIAS_LM ~ TRAMO_EDAD + NIVEL + CALIDAD_DESEMPEÑO +
## TRAMO_DOCENTE + INSATISFACTORIO
## Df Sum of Sq RSS AIC F value Pr(>F)
## <none> 25025 1978.5
## TRAMO_EDAD 1 276.36 25301 1982.0 5.4003 0.020542 *
## NIVEL 1 200.89 25226 1980.5 3.9256 0.048117 *
## CALIDAD_DESEMPEÑO 1 472.11 25497 1985.8 9.2254 0.002514 **
## TRAMO_DOCENTE 6 528.93 25554 1977.0 1.7226 0.113745
## INSATISFACTORIO 1 2677.97 27703 2027.3 52.3292 1.825e-12 ***
## ---
## Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

## Octava variable que sale al modelo

```
## Single term deletions
##
## Model:
## PROMEDIO_DIAS_LM ~ TRAMO_EDAD + NIVEL + CALIDAD_DESEMPEÑO +
## INSATISFACTORIO
## Df Sum of Sq RSS AIC F value Pr(>F)
## <none> 25554 1977.0
```

```
## TRAMO_EDAD      1      327.70 25882 1981.3  6.3479  0.012066 *
## NIVEL           1      207.70 25762 1979.0  4.0234  0.045418 *
## CALIDAD_DESEMPEÑO 1      376.16 25930 1982.3  7.2866  0.007185 **
## INSATISFACTORIO 1     2819.58 28373 2027.3 54.6179 6.279e-13 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

## Modelo resultante por backward

Coeficientes	Estimación	Std. Error	t value	Pr(>t)
Intercepto	7.9831	0.7314	10.915	< 2e-16
TRAMO_EDAD[50.5,79.1)	1.7839	0.7080	2.520	0.01207
NIVELMEDIA	-1.4459	0.7208	-2.006	0.04542
CALIDAD_DESEMPEÑO[TITULAR	2.1505	0.7967	2.699	0.00718
INSATISFACTORIO[Insatisfactorio	9.9027	1.3399	7.390	6.28e-13

## Anexos

### Anexo Tablas

### Anexo Gráficos