

Proyecto Modelos Lineales

Oscar Gamboa, Gonzalo Mardones, Nicolas Montecinos

2022-10-11

Objeto de investigación

Planteamiento del Problema

¿Cuáles son los hechos o fenómenos?; ¿Qué dicen los datos?

Justificación del Problema

Las razones que conducen a investigar el fenómeno.

- Actualmente no se conoce el estado e impacto de licencias médicas de funcionarios de la educación en la provincia de Huasco durante el año 2022 ...

Estado del Arte

¿Qué se ha dicho del tema en cuestión?; ¿Cómo lo han investigado?; ¿Qué tipo de metodología utilizan?

NICO

Objetivo General y Obbjetivo Especifico

¿Para qué y cómo abordar la investigación? - Obj General: La investigación surge por el interes de comprender el alto volumen de licencias medicas por funcionarios ...

- Obj Especificos:
 - Obtener, recopilar y extraer las características más importantes que aporten al estudio de la investigación
 - Creación, definición y validación de un procedimiento exploratorio para las variables más importantes a estudiar ...

Es la fase procedimental que conduce o guía a todo proyecto.

Plateamiento de la hipótesis en el estudio

¿Qué es lo que quiero problematizar o comprobar?

Generación de unidades de información

Diseño de estrategias metodológicas: ¿Qué metodologías utilizaré?;

¿Cómo voy a alcanzar los objetivos que he construido?

Procesamiento de las unidades de información

Análisis de datos: ¿Qué dicen los datos?

La base de datos esta compuesta por 1.109 observaciones y 24 columnas detalladas a continuación:

Descripción de los datos

Base de datos de Licencias Médicas	
Variable	Descripción
ESCALAFON	[STRING] ...
RUT_FUNCIONARIO	[STRING] ...
NOMBRE	[STRING] ...
AP_P	[STRING] ...
AP_M	[STRING] ...
SEXO	[FACTOR] ...
ESTADO_CIVIL	[FACTOR] ...
EDAD	[DOUBLE] ...
COMUNA_RESIDENCIA	[STRING] ...
RENTA_PROMEDIO	[DOUBLE] ...
SISTEMA_SALUD	[FACTOR] ...
JORNADA	[DOUBLE] ...
NIVEL	[FACTOR] ...
ESTABLECIMIENTO	[STRING] ...
TIPO_ESTABLECIMIENTO	[FACTOR] ...
COMUNA_ESTABLECIMIENTO	[FACTOR] ...
CALIDAD_DESEMPEÑO	[FACTOR] ...
ESTAMENTO	[FACTOR] ...
TRAMO_DOCENTE	[FACTOR] ...
EVALUACION_DOCENTE	[FACTOR] ...
CANTIDAD_LM	[DOUBLE] ...
PROMEDIO_DIAS_LM	[DOUBLE] ...
SUMA_DIAS	[DOUBLE] ...
TRASLADO_COMUNA	[DOUBLE] ...

Detalle de ETL de los datos

-NOTA: Agregar el histograma incluyendo los casos de profesionales sin licencias médicas

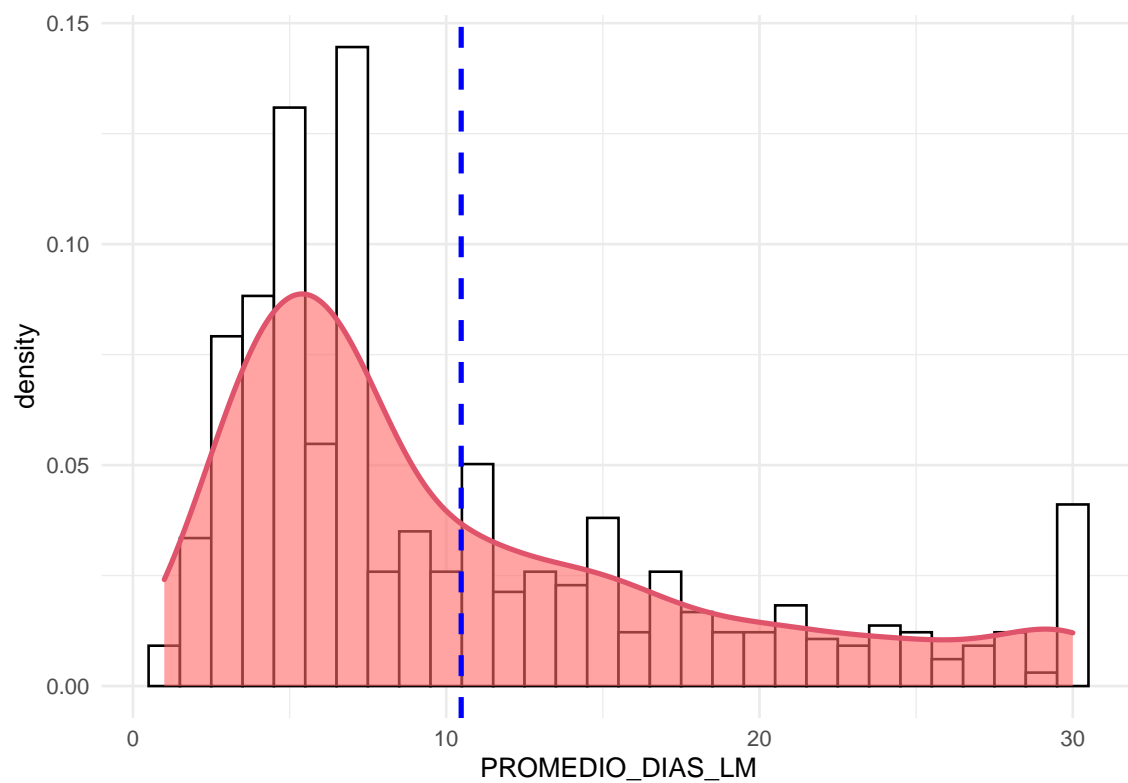
```
BBDD_proyecto <- BBDD_proyecto %>% mutate_at(c("SEXO", "ESTADO_CIVIL",
  "SISTEMA_SALUD", "NIVEL",
  "TIPO_ESTABLECIMIENTO",
  "CALIDAD_DESEMPEÑO",
  "COMUNA_ESTABLECIMIENTO",
  "ESTAMENTO", "TRAMO_DOCENTE"), factor)

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)

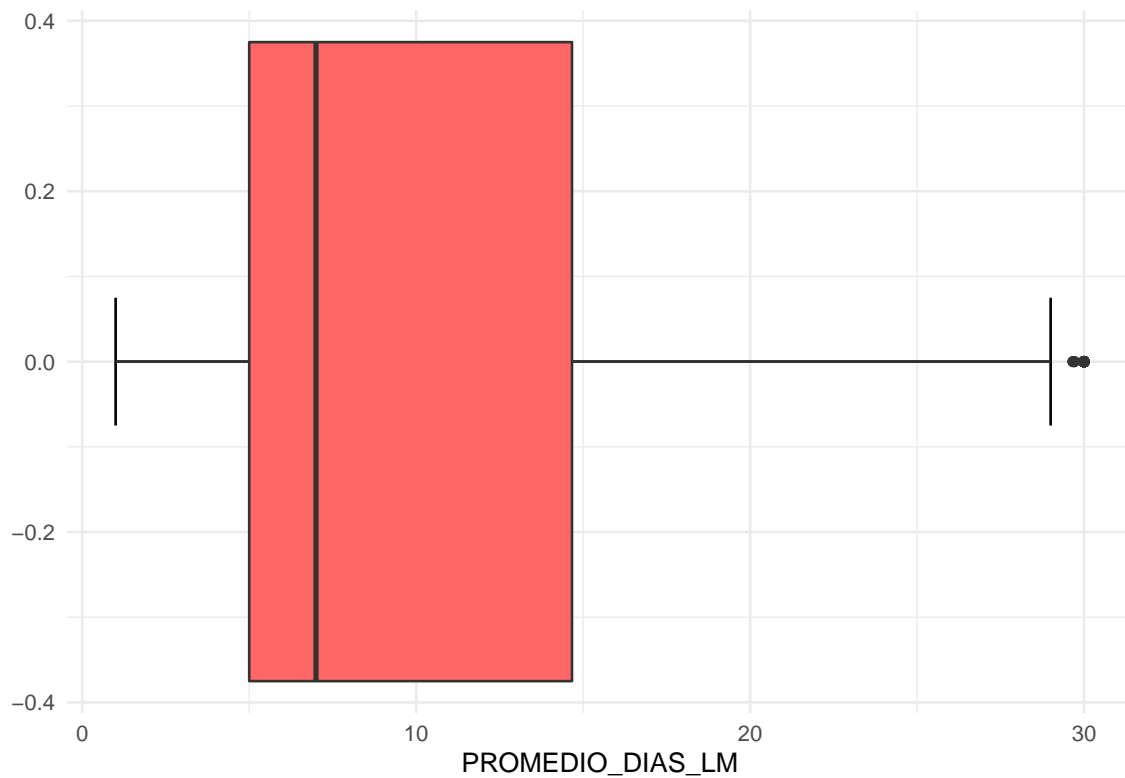
ggplot(BBDD_proyecto_sin_ceros, aes(x = PROMEDIO_DIAS_LM)) +
  geom_histogram(aes(y = ..density..), colour = "black", fill = "white") +
  geom_density(alpha = 0.6, fill = "#FF6666", lwd = 1, colour = 2) +
  geom_vline(aes(xintercept = mean(PROMEDIO_DIAS_LM)),
```

```
color = "blue", linetype = "dashed", size = 1)+
theme_minimal()
```



Boxplot de días promedios de licencias medicas

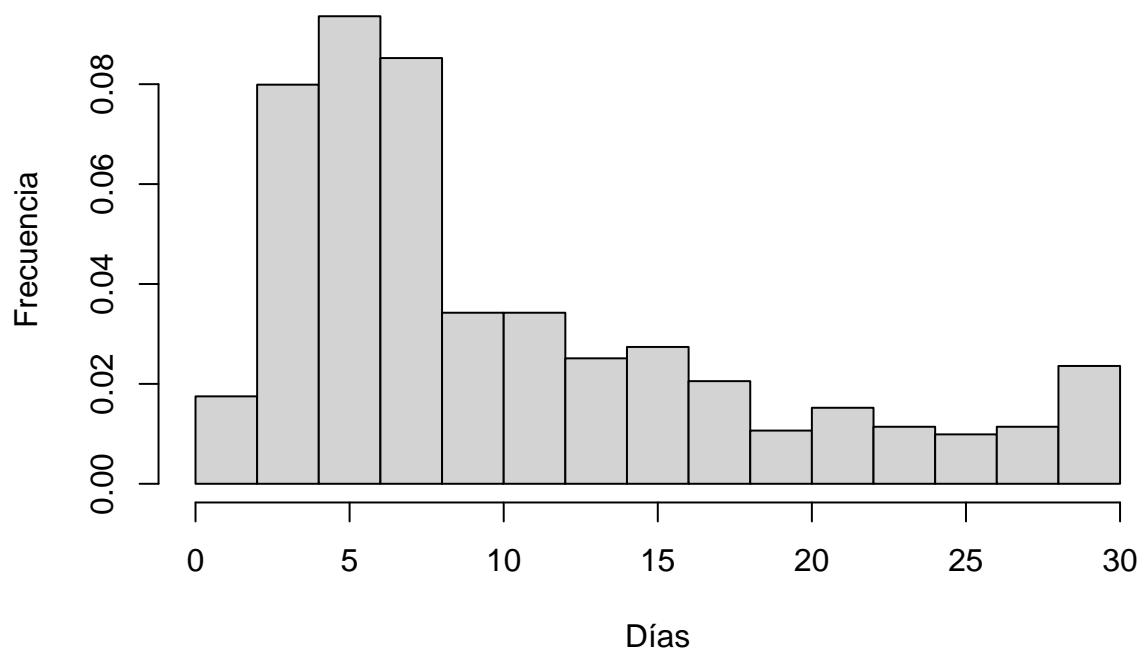
```
ggplot(BBDD_proyecto_sin_ceros, aes(x = PROMEDIO_DIAS_LM))+
  stat_boxplot(geom = "errorbar",width = 0.15) +
  geom_boxplot(fill="#FF6666")+
  theme_minimal()
```



Explorar las variables más importantes

```
hist(BBDD_proyecto_sin_ceros$PROMEDIO_DIAS_LM,freq = FALSE,
     main = "Histograma de Licencia Días Promedio",
     xlab = "Días", ylab = "Frecuencia")
```

Histograma de Licencia Días Promedio



Selección Formal de Modelo

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	13.4	38244	2674.1	0.2301	0.6315751
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

modelo0 = lm(PROMEDIO_DIAS_LM ~ EVALUACION_DOCENTE, 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 ~ EVALUACION_DOCENTE
##
```

	Df	Sum of Sq	RSS	AIC	F value	Pr(>F)
<none>			34816	2618.4		
SEXO	1	63.98	34752	2619.2	1.1986	0.2740068
ESTADO_CIVIL	3	0.54	34815	2624.4	0.0033	0.9997332
EDAD	1	694.67	34121	2607.1	13.2538	0.0002936 ***
RENTA_PROMEDIO	1	188.81	34627	2616.8	3.5498	0.0599976 .
SISTEMA_SALUD	1	158.81	34657	2617.4	2.9831	0.0846134 .
JORNADA	1	1.30	34814	2620.3	0.0243	0.8760602
NIVEL	1	454.18	34361	2611.8	8.6048	0.0034707 **
CALIDAD_DESEMPEÑO	1	1177.13	33638	2597.8	22.7809	2.244e-06 ***
ESTAMENTO	1	11.51	34804	2620.2	0.2153	0.6428212
TRAMO_DOCENTE	6	880.73	33935	2613.5	2.7944	0.0108571 *
TRASLADO_COMUNA	1	43.86	34772	2619.6	0.8211	0.3651882

```
## ---
```

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

```
modelo0 = lm(PROMEDIO_DIAS_LM ~ EVALUACION_DOCENTE + CALIDAD_DESEMPEÑO, 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 ~ EVALUACION_DOCENTE + CALIDAD_DESEMPEÑO
```

```
##
```

	Df	Sum of Sq	RSS	AIC	F value	Pr(>F)
<none>			33638	2597.8		
SEXO	1	71.15	33567	2598.4	1.3777	0.240926
ESTADO_CIVIL	3	9.12	33629	2603.6	0.0586	0.981399
EDAD	1	278.75	33360	2594.3	5.4314	0.020083 *
RENTA_PROMEDIO	1	9.57	33629	2599.6	0.1851	0.667199
SISTEMA_SALUD	1	203.14	33435	2595.8	3.9491	0.047315 *
JORNADA	1	0.10	33638	2599.8	0.0020	0.964377
NIVEL	1	388.41	33250	2592.2	7.5929	0.006024 **
ESTAMENTO	1	12.55	33626	2599.5	0.2426	0.622504
TRAMO_DOCENTE	6	747.42	32891	2595.0	2.4429	0.024172 *
TRASLADO_COMUNA	1	3.23	33635	2599.7	0.0624	0.802865

```
## ---
```

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

```
modelo0 = lm(PROMEDIO_DIAS_LM ~ EVALUACION_DOCENTE + CALIDAD_DESEMPEÑO + NIVEL, 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 ~ EVALUACION_DOCENTE + CALIDAD_DESEMPEÑO +
```

```
## NIVEL
```

	Df	Sum of Sq	RSS	AIC	F value	Pr(>F)
<none>			33250	2592.2		
SEXO	1	22.67	33227	2593.7	0.4428	0.50603
ESTADO_CIVIL	3	5.17	33245	2598.1	0.0336	0.99175
EDAD	1	319.53	32930	2587.8	6.2973	0.01233 *
RENTA_PROMEDIO	1	20.10	33230	2593.8	0.3926	0.53115
SISTEMA_SALUD	1	162.22	33088	2590.9	3.1819	0.07493 .
JORNADA	1	0.37	33250	2594.1	0.0073	0.93196
ESTAMENTO	1	10.15	33240	2593.9	0.1982	0.65631
TRAMO_DOCENTE	6	711.93	32538	2589.9	2.3485	0.02984 *
TRASLADO_COMUNA	1	2.19	33248	2594.1	0.0427	0.83641

```
## ---
```

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

```
modelo0 = lm(PROMEDIO_DIAS_LM ~ EVALUACION_DOCENTE + CALIDAD_DESEMPEÑO + NIVEL + EDAD, 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 ~ EVALUACION_DOCENTE + CALIDAD_DESEMPEÑO +
##     NIVEL + EDAD
##           Df Sum of Sq  RSS    AIC F value  Pr(>F)
## <none>                32930 2587.8
## SEXO                1    39.69 32891 2589.0  0.7819 0.37690
## ESTADO_CIVIL        3    73.49 32857 2592.3  0.4816 0.69516
## RENTA_PROMEDIO      1    66.75 32864 2588.5  1.3161 0.25172
## SISTEMA_SALUD       1   168.81 32762 2586.4  3.3390 0.06812 .
## JORNADA             1     0.84 32930 2589.8  0.0166 0.89759
## ESTAMENTO           1     0.08 32930 2589.8  0.0017 0.96756
## TRAMO_DOCENTE       6   720.97 32209 2585.3  2.3988 0.02668 *
## TRASLADO_COMUNA    1     1.21 32929 2589.8  0.0239 0.87717
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```
summary(modelo0)
```

```
##
## Call:
## lm(formula = PROMEDIO_DIAS_LM ~ EVALUACION_DOCENTE + CALIDAD_DESEMPEÑO +
##     NIVEL + EDAD, data = BBDD_proyecto_sin_ceros)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -17.787  -4.830  -2.260   3.468  21.881
##
## Coefficients:
##                Estimate Std. Error t value Pr(>|t|)
## (Intercept)      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 *
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 7.123 on 649 degrees of freedom
## Multiple R-squared:  0.1392, Adjusted R-squared:  0.13
## F-statistic: 15 on 7 and 649 DF, p-value: < 2.2e-16
```

```
contrasts(BBDD_proyecto_sin_ceros$EVALUACION_DOCENTE)
```

```
##           DESTACADO BASICO  INSATISFACTORIO  COMPETENTE
## SIN EVALUACIÓN          0      0              0           0
## DESTACADO               1      0              0           0
## BASICO                  0      1              0           0
## INSATISFACTORIO        0      0              1           0
```

## COMPETENTE	0	0	0	1
---------------	---	---	---	---