



# PRUEBA#2G

## ANALÍTICA ARLM/ST

**SABER + SABER HACER + SABER SER**

Métodos Cuantitativos (N) | Facultad de Ingeniería | UTEM  
Viernes 19 08:30 horas, Noviembre de 2021, Semestre Primavera

Nombre del Equipo		Yakuza
#	Nombre del/la Integrante	Participación (%)
1	Tamara Cabrera	100%
2	Lucas Campos	100%
3	Sebastián Villagra	100%
4	Agustín Magnere	100%
5	Matías Aránguiz	100%

### ■ Rúbrica

- Elaboración del informe técnico:** Sobre este mismo documento desarrolle las tablas, gráfico y memoria de cálculo. Debe asegurarse un informe ordenado respetando el número de 9 páginas.
- Fecha de entrega del informe técnico de desarrollo (formato Word y PDF), 4 Excel Data Filtrada ARLM (reloj del profesor):** Jueves 25 08:00 horas. Desde las 08:00 horas en adelante no se recibirá, revisará y evaluará el informe. El envío debe realizarlo el/la líder del equipo al correo [luís.solar@utem.cl](mailto:luís.solar@utem.cl). La no entrega de la Excel (variables del proyecto) se penalizará con un 30% de la nota final. En el mail de envío se debe registrar en "Asunto", como ejemplo, el siguiente texto: **P2A UCHIHA MCUA UTEM**, donde "A" es el código del proyecto desarrollado y "UCHIHA" el nombre del equipo, según corresponda. **El nombre de las Excel deberán tener el nombre "uchihaXX" (minúsculas), según corresponda el nombre del equipo donde XX es 05, 10, 15 y 20.**
- Fecha de defensa del proyecto de investigación el día Viernes 26. Todos los integrantes de cada equipo defienden el trabajo con micrófono (sin cámara).
- Nota final (NF) se calculará como:  

$$NF = \text{Porcentaje de Participación} \times [40\% (\text{Informe Técnico Grupal}) + 60\% (\text{Defensa Individual})]$$

Finalmente, un estudiante con menos de un 75% de participación en el desarrollo o no asiste a la defensa queda excluido del Informe y defensa. Podrá ser recuperada con la Prueba Recuperativa.

Para las estadísticas de Rezago Social en México en los años 2005, 2010 y 2015, se solicita que estudie y analice el mejor modelo ARLM para predecir el Índice de Rezago para la Entidad Federativa de mayor población durante los años 2005,2010 y 2015 para la entidad federativa “**Jalisco**”.

Como entregable se solicita:

- (1) Resumen de la Metodología utilizada para el cálculo del IRS (800 palabras) presentada en el link: [https://www.coneval.org.mx/Medicion/IRS/Paginas/Indice\\_Rezago\\_Social\\_2020.aspx](https://www.coneval.org.mx/Medicion/IRS/Paginas/Indice_Rezago_Social_2020.aspx)
- (2) Histogramas (análisis gráfico) y Boxplot análisis estadístico de caja) del IRS de los años 2005,2010 y 2015
- (3) Matriz de Correlación: Diagramas de dispersión (análisis gráfico), Histogramas (análisis gráfico) y coeficientes de correlación (análisis estadístico) para los años 2005,2010 y 2015
- (4) Modelos “**A la Fuerza**” (todas las regresoras) para los años 2005,2010 y 2015
- (5) Análisis Relacional Inter-Variables
- (6) Análisis de los Modelos ARLM Óptimos RStudio para los años 2005,2010 y 2015
- (7) Proyección para 2020 con el modelo de mayor rendimiento y posteriormente la comparación con el real 2020
- (8) Conclusión final y recomendaciones
- (9) Análisis Adicional de Valor Agregado (1.0 puntos extras)
- (10) Secuencia de comandos realizados en Rstudio (título y escritura Word)
- (11) Memoria de Cálculo de Proyección Paso (7).

La Excel deberá contener solo la data filtrada para este estudio (con nombre de columna y sin gráficos) y en el mismo orden de la base de datos original.

Se adjunta los documento “**IRS\_entidades\_mpios\_20XX.xls**” con la data de la población total, indicadores, índice y grado de rezago social según municipio para los años 2005, 2010, 2015 y 2020.

Nota: Solo para la modelación ARLM no considerar la siguiente información (pero sí para el análisis inter-variables):

- Clave de la entidad
- Entidad federativa
- Clave del Municipio
- Municipio
- Grado de rezago social
- Lugar que ocupa en el contexto nacional.

## SMART “ANÁLISIS ARLM”: Resumen de la Metodología

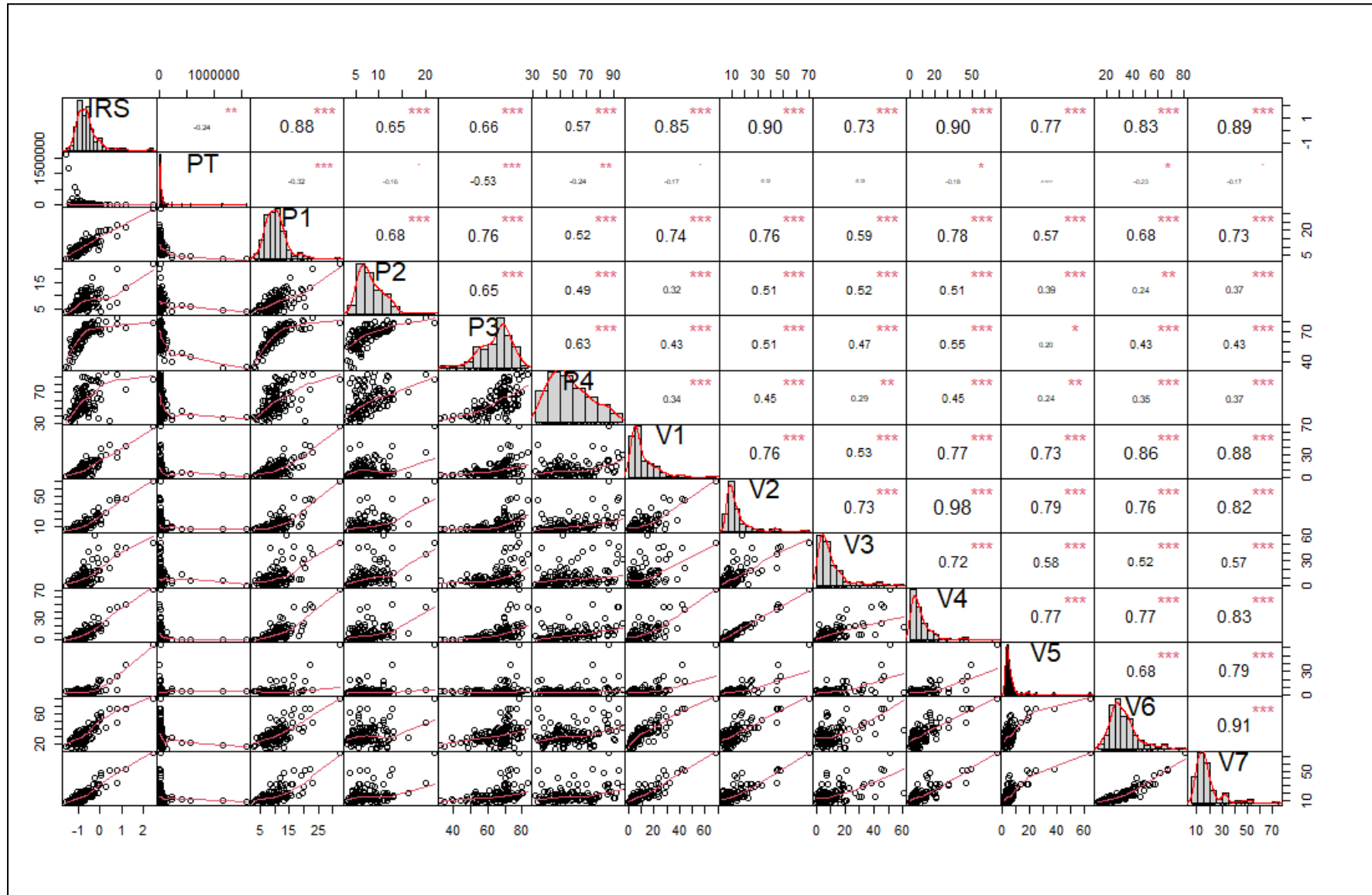
La metodología de los Mapas de Pobreza en México se basa en entregar dos medidas de carencia, que se denominan “la pobreza por ingresos” y “el índice de rezago social”, tanto a nivel municipal como estatal.

Esta metodología viene a solucionar los problemas de investigación sobre la pobreza, ya que otras metodologías entregan estimaciones de distintos grupos y subdivisiones geográficas, pero no abarcan todas como esta metodología.

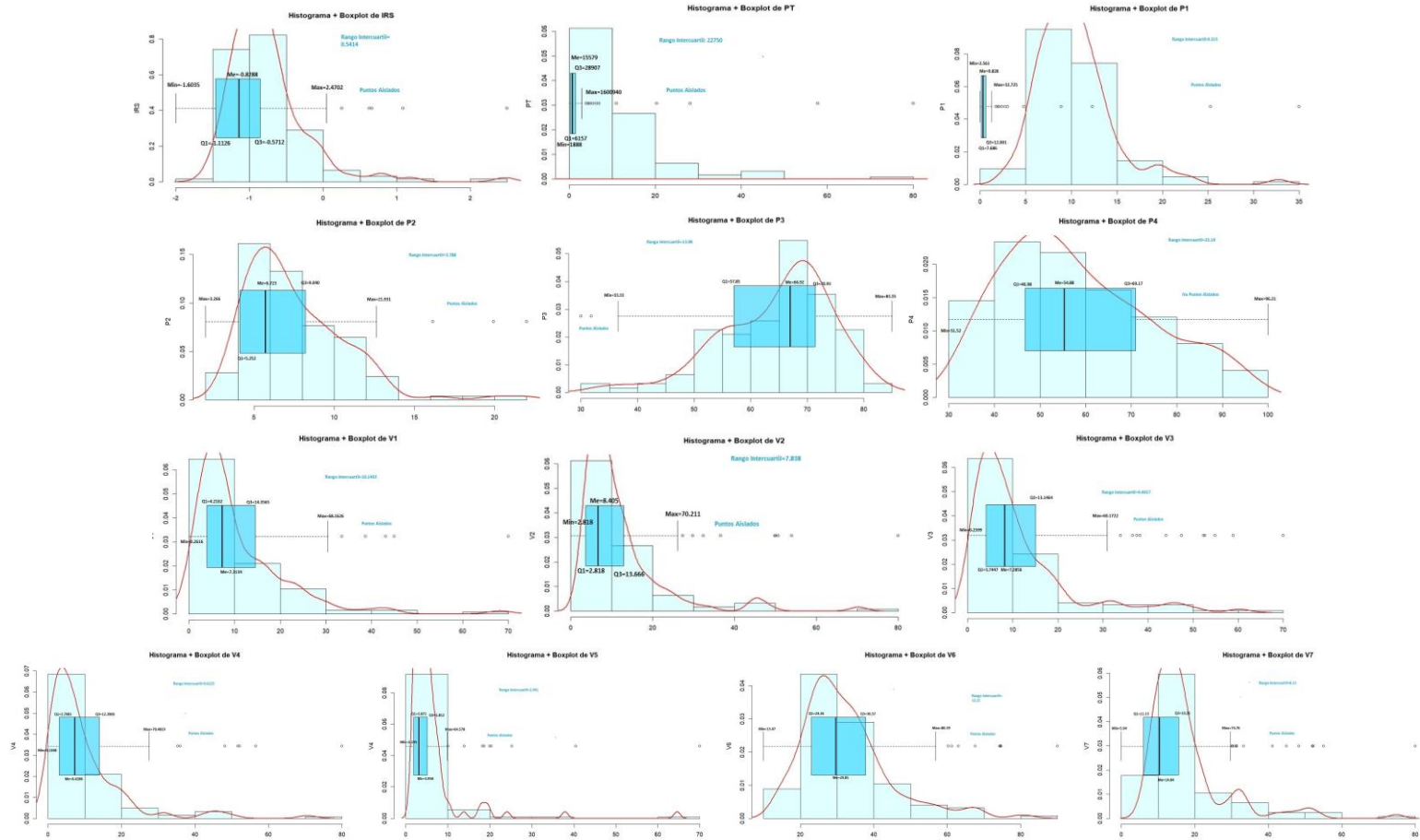
Pasos:

- 1.-Definir una variable ( $W$ ) como un indicador de bienestar de la población.
- 2.-Estimar la distribución conjunta de  $y_h$  y de las covariables  $x_h$ . Utilizando información de la encuesta de hogares.
- 3.-Se restringen las covariables  $x_h$  de la encuesta al subconjunto de ellas que fue captado de manera apropiada y comparable en el censo de población.
- 4.-Generar la distribución de  $y_h$  utilizando la distribución obtenida en el punto 2.
- 5.- Utilizar y restringir a las variables que se utilizaron tanto en la encuesta como en el censo, para obtener una adecuada distribución.
- 6.-Distribución condicional de  $W$ .
- 7.-Se generan estimadores de incidencia de pobreza.
- 8.-Las estimaciones resultantes son suficientemente precisas y confiables, por lo que podemos obtener estimaciones factibles de incidencia de la pobreza a nivel estatal y municipal.

## SMART “ANÁLISIS ARLM”: Matriz de Correlación (Modelo 2005)



# SMART “ANÁLISIS ARLM”: HISTOGRAMAS Y BOXPLOT (Modelo 2005)



# SMART “ANÁLISIS ARLM”: Análisis 2005

## Análisis Matriz de Correlación

### Diagramas de dispersión:

- Cuatro de los doce diagramas de dispersión pares respecto de la variable “IRS” (Índice de rezago social), presentan puntos aislados. PT, P2, P4 y V3.
- Todos los diagramas de dispersión pares respecto de la variable “IRS” presentan una tendencia lineal positiva, a excepción de la potencial variable predictora “PT” (Población total) que presenta una tendencia lineal negativa.
- Los diagramas de dispersión pares respecto de la variable “IRS”, en general no presentan una alta variación.
- Los diagramas de dispersión pares respecto de la variable “IRS” que presentan una moderada variación, corresponden a las potenciales variables predictoras: “P2”, “P4” y “V3”.
- Los diagramas de dispersión pares respecto de la variable “IRS” que presentan una baja variación, corresponde a la potencial variable predictora: “P1”, “P3”, “V1”, “V2”, “V4”, “V5”, “V6” y “V7”.
- Todos los diagramas de dispersión pares respecto de la variable “IRS” presentan homocedasticidad a excepción de la relación con la potencial variable predictora “P2”, “P4” y “V3” que presentan un moderado nivel de heterocedasticidad.

### Histogramas:

- Los histogramas que presentan sesgo a la izquierda corresponden a las potenciales variables predictoras: “P1”, “P2”, “V1”, “V2”, “V3”, “V4”, “V5”, “V6” y “V7”
- Los histogramas que presentan sesgo a la derecha corresponden a la potencial variable predictora “P3”.
- 

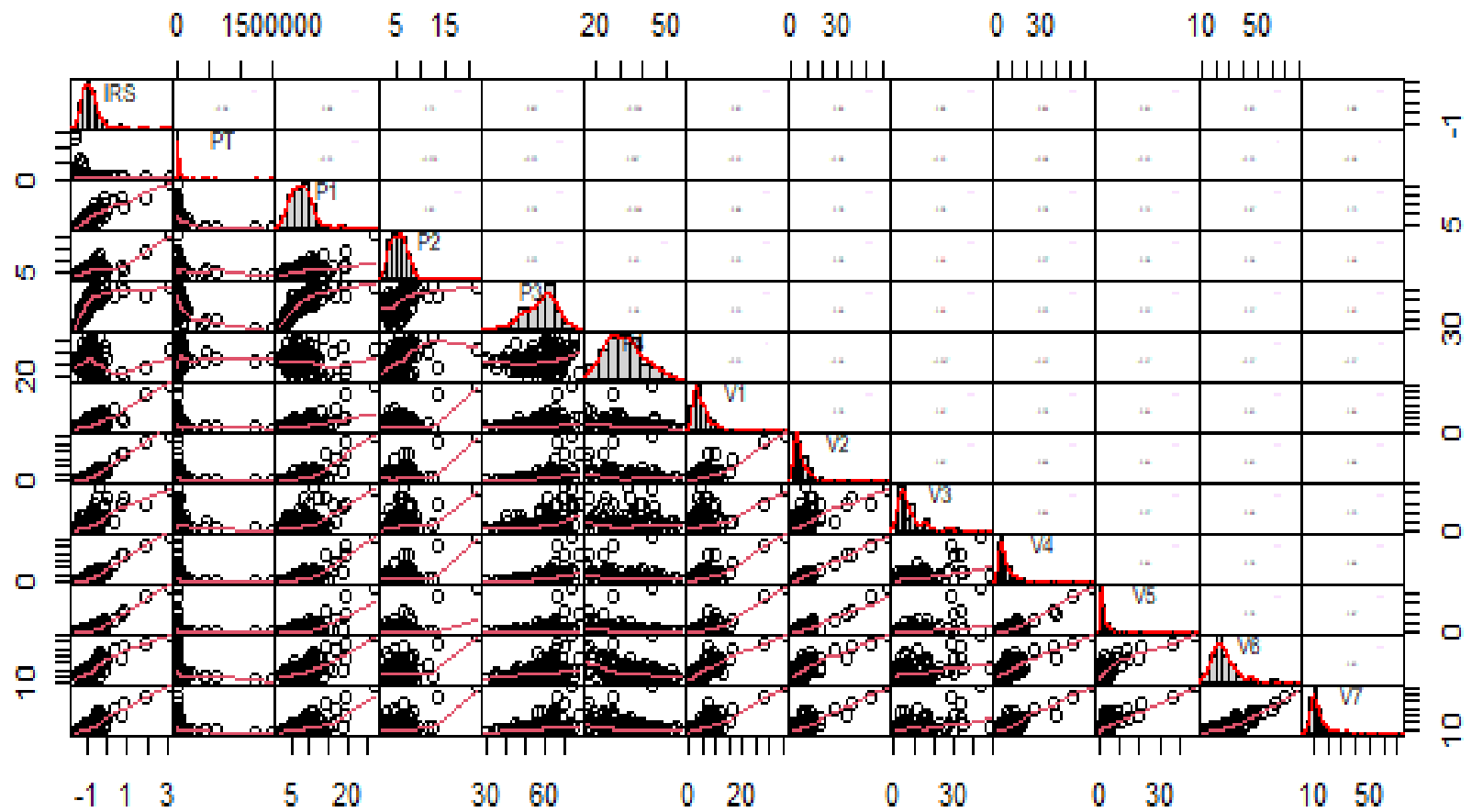
### Coefficientes de correlación:

- La potencial variable predictora “PT” presenta el menor valor de coeficiente de correlación por tendencia lineal horizontal.
- Las potenciales variables predictoras con sesgo a la izquierda presentan valores de coeficiente de correlación moderadamente altos.

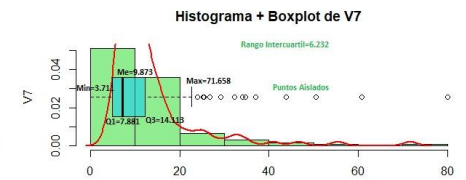
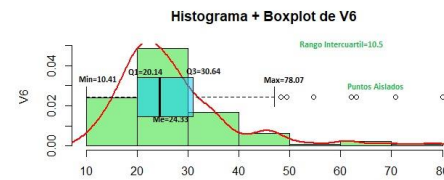
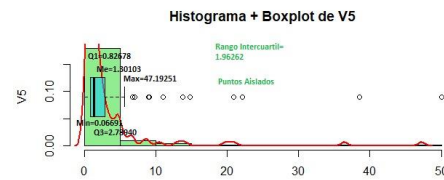
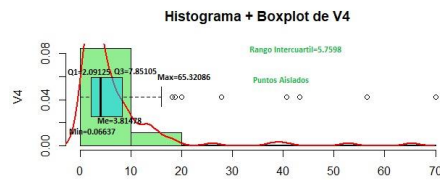
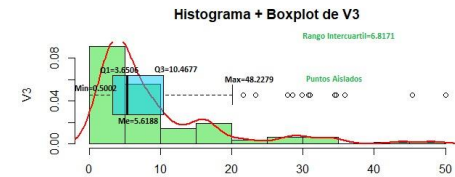
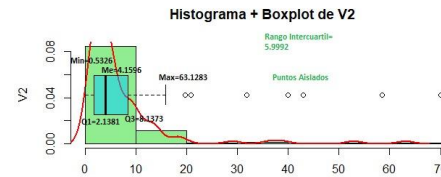
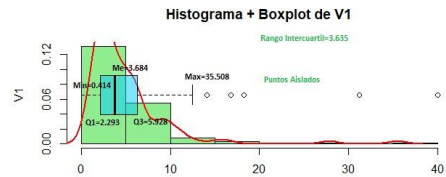
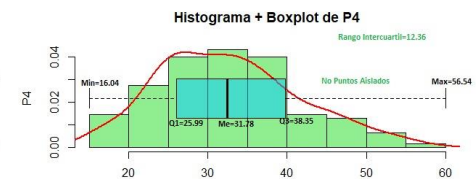
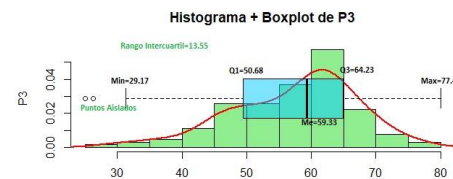
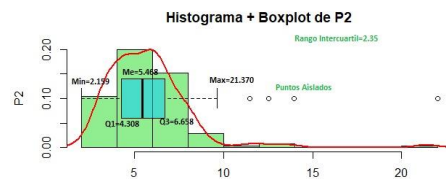
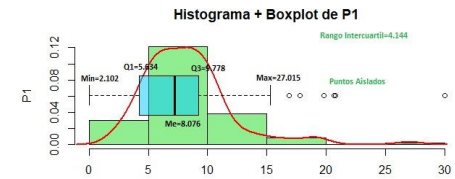
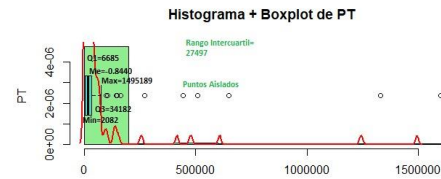
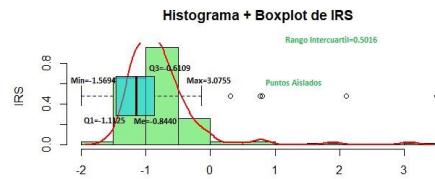
### Boxplot:

- Para los boxplot del año 2005 se tiene para casi todas las variables una distribución de datos asimétricas y con sesgo hacia la izquierda ( $Me < \bar{x}$ ), es decir los datos se concentran en la parte derecha de la caja, con excepción de la variable P3 la cual presenta una asimetría positiva, con un sesgo hacia la derecha ( $Me > \bar{x}$ ), es decir sus datos están concentrados a la izquierda de la caja.
- Los Rangos Intercuartiles para cada variable son los siguientes: IRS=0.5414, PT=22750, P1=4.315, P2= 3.788, P3= 13.08, P4=22.19, V1=10.1403, V2=7.838, V3=9.4017, V4= 9.6225, V5=2.941, V6=12.21, V7=8.13

## SMART “ANÁLISIS ARLM”: Matriz de Correlación (Modelo 2010)



# SMART “ANÁLISIS ARLM”: HISTOGRAMAS Y BOXPLOT (Modelo 2010)





# SMART “ANÁLISIS ARLM”: Análisis 2010

## Análisis Matriz de Correlación

### Diagramas de dispersión:

La mayoría de los diagramas de dispersión respecto de la variable predictora IRS no presentan puntos aislados, a excepción de PT, V3 y P4.

Todos los diagramas de dispersión pares respecto a la variable IRS con excepción de PT y P4, presentan una tendencia lineal positiva.

- Los diagramas de dispersión pares respecto de la variable “IRS”, en general no presentan una alta variación, a excepción de P4.
- El diagrama de dispersión par respecto de la variable “IRS” que presentan una moderada variación, corresponden a la potencial variable predictora “V3”.
- Los diagramas de dispersión pares respecto de la variable “IRS” que presentan una baja variación, corresponden a las potenciales variables predictoras: P1, P2, P3, V1, V2, V4, V5, V6 y V7.
- Todos los diagramas de dispersión pares respecto de la variable “IRS” presentan homocedasticidad a excepción de la relación con la potencial variable predictora “P2, “P4” Y “V3” que presentan un moderado nivel de heterocedasticidad.

### Histogramas:

- Todos los histogramas presentan sesgo a la izquierda.

### Coefficientes de Correlación:

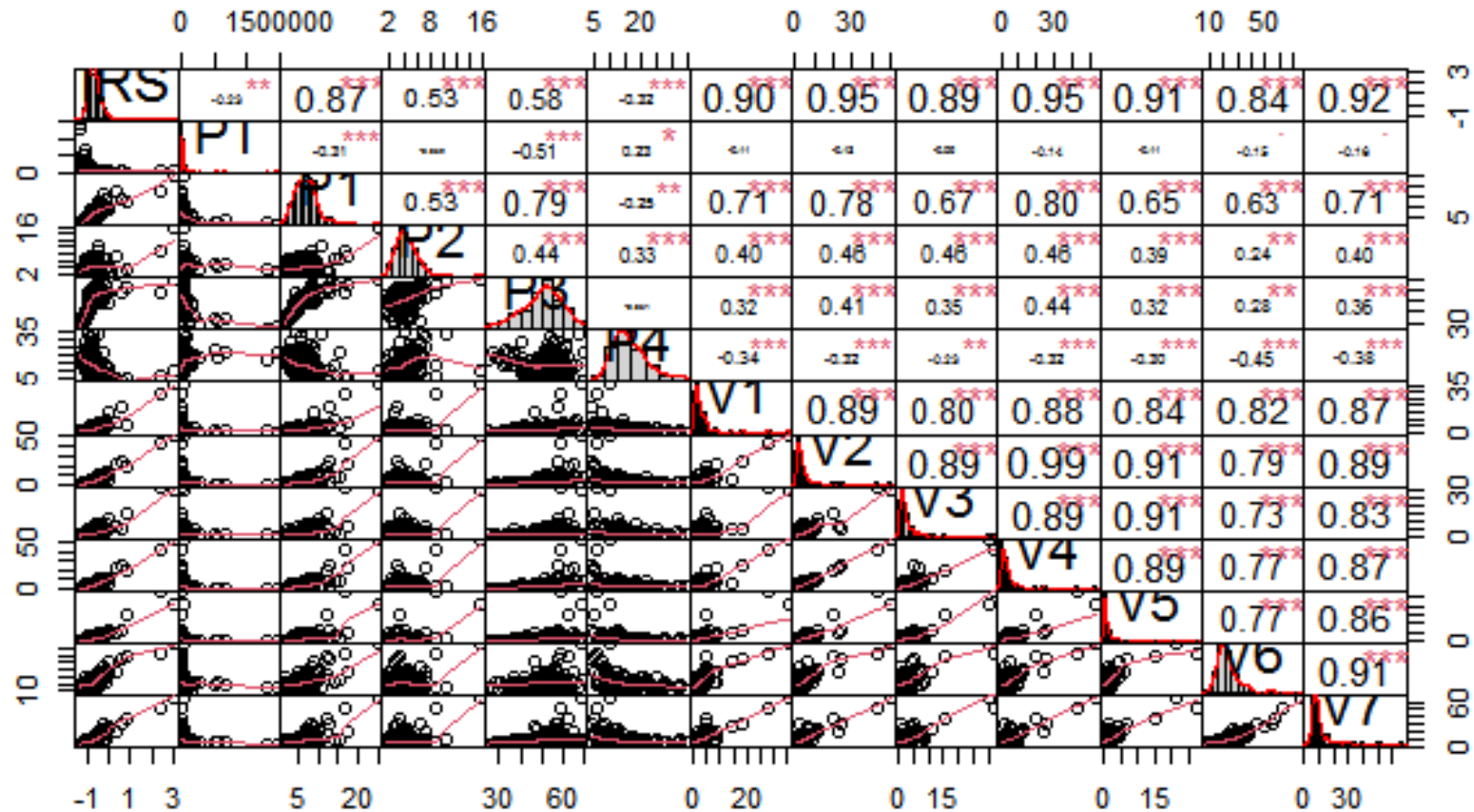
- La potencial variable predictora “P4” presenta el menor valor de coeficiente de correlación por tendencia lineal horizontal.
- Las potenciales variables predictoras con sesgo a la izquierda presentan valores de coeficiente de correlación moderadamente altos.

### Boxplot:

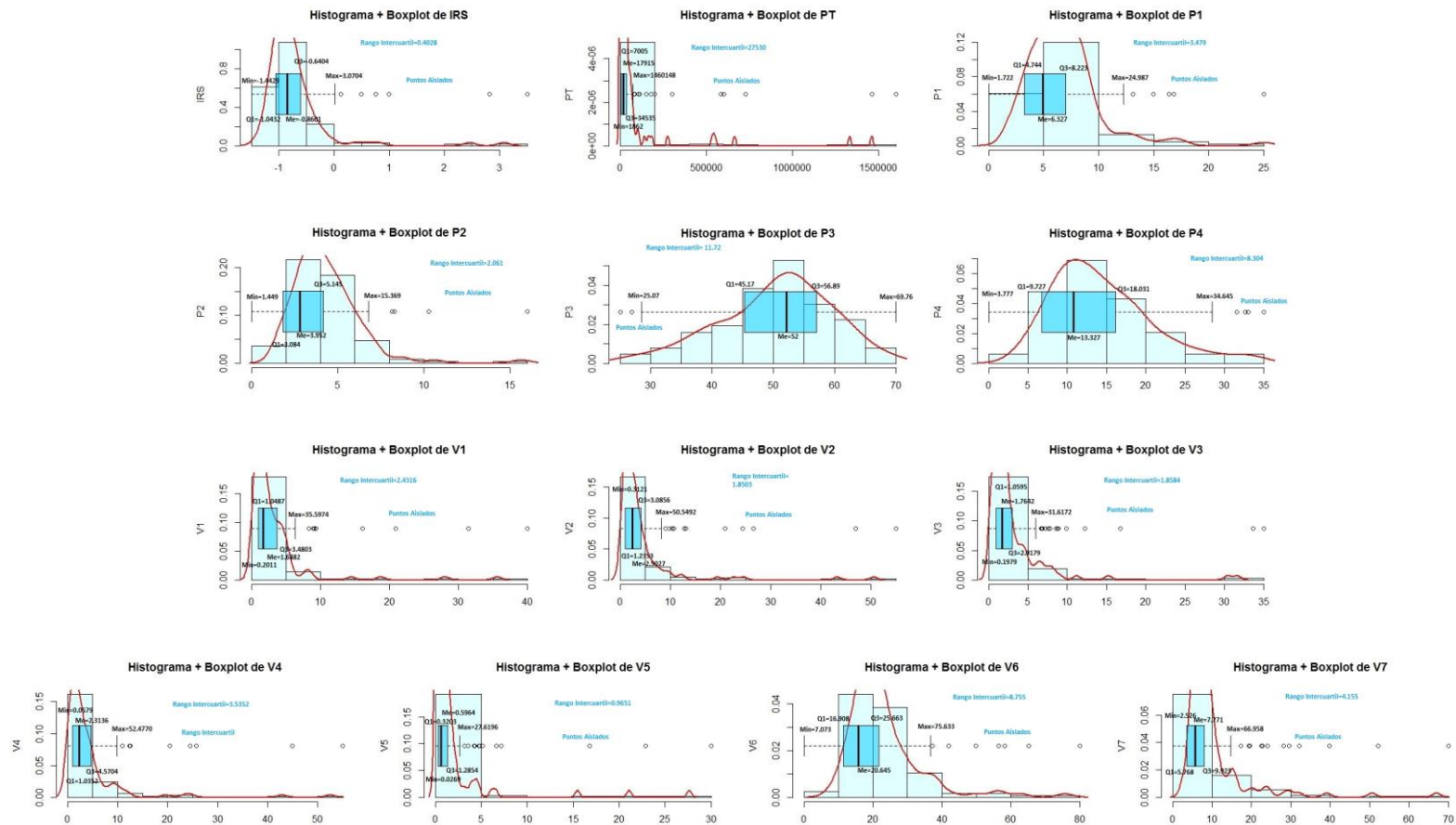
- Asimetría positiva o segada a la derecha si la parte más larga de la caja es la parte superior a la mediana.
- Los datos se concentran en la parte inferior de la distribución. La media suele ser mayor que la mediana.
- Asimetría negativa o sesgada a la izquierda si la parte más larga es la inferior a la mediana.
- Los datos se concentran en la parte superior de la distribución. La media suele ser menor que la mediana.
- Para los boxplot del año 2010 se tiene para casi todas las variables una distribución de datos asimétricas y con sesgo hacia la izquierda ( $Me < \bar{x}$ ), es decir los datos se concentran en la parte derecha de la caja, con excepción de la variable P3 la cual presenta una asimetría positiva, con un sesgo hacia la derecha ( $Me > \bar{x}$ ), es decir sus datos están concentrados a la izquierda de la caja.
- Los Rangos Intercuartiles para cada variable son los siguientes

IRS=0.5016 PT=27497 P1=4.144 P2= 2.35 P3= 13.55 P4=12.36 V1=3.635 V2=5.9992 V3=6.8171 V4=5.7598 V5=1.96262 V6=10.5 V7=

## SMART “ANÁLISIS ARLM”: Matriz de Correlación (Modelo 2015)



# SMART “ANÁLISIS ARLM”: HISTOGRAMAS Y BOXPLOT (Modelo 2015)



## SMART “ANÁLISIS ARLM”: Análisis 2015

### Análisis Matriz de Correlación

Algunos de los diagramas de dispersión pares respecto la variable IRS presentan puntos aislados, estos corresponden a PT, P2 y V5.

Todos los diagramas de dispersión pares respecto la variable IRS con excepción de PT y P4, presentan una tendencia lineal creciente y positiva, en el caso de PT son negativa casi horizontal (pendiente 0) y P4 negativa.

Todos los diagramas de dispersión pares respecto la variable IRS muestran grupos de relación y tendencia a agruparse hacia la izquierda.

Los diagramas de dispersión pares respecto a la variable IRS que presentan una alta variabilidad corresponden a P4, por lo tanto, se considera como variable predictora.

Los diagramas de dispersión pares respecto a la variable IRS que presentan una moderada variabilidad corresponden a P1, P2, P3 por lo tanto se consideran variables Predictoras.

Los diagramas de dispersión pares respecto a la variable IRS que presentan una baja variabilidad pertenecen a PT con -0,99, luego le siguen V1, V2, V4, V5 Y V7.

Todos los diagramas de dispersión pares respecto de la variable IRS muestran un alto nivel de homoscedasticidad con la excepción de la relación entre las variables P1, P2, P3 Y P4 presentando un moderado nivel de heteroscedasticidad.

Los histogramas respecto la variable IRS no presentan poblaciones distintas

La mayoría de las variables de los histogramas presentan un fuerte sesgo hacia la izquierda las cuales son IRS, PT, P1, P2, V1, V2, V3, V4, V5, V6 y V7

Las variables con un bajo sesgo hacia la izquierda es P4

Las variables con un bajo sesgo hacia la derecha es P3

La variable predictora PT presenta el mayor valor de coeficiente de correlación por tendencia lineal horizontal, ya que posee una pendiente negativa cercana a 1

A pesar de la heteroscedasticidad y variabilidad, las variables con mayor número de correlación pertenecen a aquellas que muestran un fuerte sesgo hacia la izquierda

Para los boxplot del año 2015 se tiene para casi todas las variables una distribución de datos asimétricas y con sesgo hacia la izquierda ( $Me < x^-$ ), es decir los datos se concentran en la parte derecha de la caja, con excepción de la variable P3 la cual presenta una asimetría positiva, con un sesgo hacia la derecha ( $Me > x^-$ ), es decir sus datos están concentrados a la izquierda de la caja.

los Rangos Intercuartiles para cada variable son los siguientes:

IRS=0,8025 PT=1460148 P1=3479 P2=3696 P3=11.72 P4=8.759 V1=2.4316 V2=2.8503 V3=1.8584 V5=3.5352 V6=0.9651 8.755 4.155

# SMART “ANÁLISIS ARLM”: Modelos “A la Fuerza” 2005-2015

>2005

```
> summary(model)
```

Call:

```
lm(formula = IRS ~ ., data = yakuza05, na.action = na.exclude)
```

Residuals:

Min	1Q	Median	3Q	Max
-0.189227	-0.048410	0.000875	0.054150	0.136743

Coefficients:

	Estimate	Std. Error	t value	Pr(> t )
(Intercept)	-2.246e+00	8.277e-02	-27.132	< 2e-16 ***
PT	-1.014e-08	4.795e-08	-0.212	0.832862
P1	8.787e-03	4.237e-03	2.074	0.040386 *
P2	3.363e-02	3.985e-03	8.439	1.34e-13 ***
P3	4.157e-03	1.693e-03	2.455	0.015624 *
P4	3.545e-03	5.894e-04	6.014	2.36e-08 ***
V1	8.286e-03	1.696e-03	4.886	3.49e-06 ***
V2	1.119e-02	3.914e-03	2.860	0.005069 **
V3	5.667e-03	9.584e-04	5.913	3.78e-08 ***
V4	-4.319e-03	3.659e-03	-1.180	0.240372
V5	5.582e-03	2.113e-03	2.642	0.009425 **
V6	5.782e-03	1.644e-03	3.518	0.000631 ***
V7	1.205e-02	2.116e-03	5.694	1.03e-07 ***

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

Residual standard error: 0.07703 on 111 degrees of freedom

Multiple R-squared: 0.9827, Adjusted R-squared: 0.9808

F-statistic: 524 on 12 and 111 DF, p-value: < 2.2e-16

>2010

```
> summary(modelo)
```

Call:

```
lm(formula = IRS ~ ., data = yakuza.10, na.action = na.exclude)
```

Residuals:

Min	1Q	Median	3Q	Max
-5.293e-07	-1.126e-07	-3.730e-09	9.324e-08	6.261e-07

Coefficients:

	Estimate	Std. Error	t value	Pr(> t )
(Intercept)	-2.308e+00	2.126e-07	-1.086e+07	<2e-16 ***
PT	-2.156e-13	1.301e-13	-1.657e+00	0.1
P1	1.665e-02	1.394e-08	1.194e+06	<2e-16 ***
P2	2.758e-02	1.465e-08	1.883e+06	<2e-16 ***
P3	1.029e-02	4.396e-09	2.340e+06	<2e-16 ***
P4	4.012e-03	3.149e-09	1.274e+06	<2e-16 ***
V1	1.168e-02	9.615e-09	1.214e+06	<2e-16 ***
V2	7.539e-03	2.278e-08	3.310e+05	<2e-16 ***
V3	5.575e-03	2.927e-09	1.905e+06	<2e-16 ***
V4	5.943e-03	2.113e-08	2.813e+05	<2e-16 ***
V5	1.887e-02	1.280e-08	1.474e+06	<2e-16 ***
V6	6.622e-03	5.207e-09	1.272e+06	<2e-16 ***
V7	6.770e-03	7.003e-09	9.667e+05	<2e-16 ***

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

Residual standard error: 2.147e-07 on 112 degrees of freedom

Multiple R-squared: 1, Adjusted R-squared: 1

F-statistic: 7.62e+13 on 12 and 112 DF, p-value: < 2.2e-16

>2015

Call:

```
lm(formula = IRS ~ ., data = yakuza15, na.action = na.exclude)
```

Residuals:

Min	1Q	Median	3Q	Max
-4.267e-06	-1.188e-06	-3.250e-08	1.151e-06	4.237e-06

Coefficients:

	Estimate	Std. Error	t value	Pr(> t )
(Intercept)	-1.876e+00	1.533e-06	-1.223e+06	<2e-16 ***
PT	-9.830e-13	9.931e-13	-9.900e-01	0.324
P1	1.991e-02	1.491e-07	1.335e+05	<2e-16 ***
P2	2.615e-02	1.239e-07	2.111e+05	<2e-16 ***
P3	1.014e-02	3.933e-08	2.579e+05	<2e-16 ***
P4	-3.352e-03	3.441e-08	-9.742e+04	<2e-16 ***
V1	1.845e-02	8.920e-08	2.069e+05	<2e-16 ***
V2	5.096e-03	1.929e-07	2.641e+04	<2e-16 ***
V3	9.753e-03	1.018e-07	9.579e+04	<2e-16 ***
V4	6.806e-03	1.772e-07	3.842e+04	<2e-16 ***
V5	3.766e-02	1.396e-07	2.698e+05	<2e-16 ***
V6	6.902e-03	3.956e-08	1.745e+05	<2e-16 ***
V7	7.285e-03	6.347e-08	1.148e+05	<2e-16 ***

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

Residual standard error: 1.728e-06 on 112 degrees of freedom

Multiple R-squared: 1, Adjusted R-squared: 1

F-statistic: 1.128e+12 on 12 and 112 DF, p-value: < 2.2e-16

## SMART “ANÁLISIS ARLM”: Análisis Relacional Inter-variables

>2005

- Se tiene que a medida que aumenta el “IRS”, también aumenta V2 (viviendas que no disponen de excusado), V4 (viviendas que no disponen de drenaje), “V7” (Viviendas que no disponen refrigerador) y P1 (Población de 15 años o más analfabeta). Debido a su tendencia lineal y alta correlación positiva 0.88, 0.89 y 0.90.

- V2 (viviendas que no disponen de excusado) y V4 (viviendas que no disponen de drenaje) poseen una correlación muy fuerte de 0,98\*\*\*. Cada variable que aumenta, indica que la otra también crece.

-V1 (Viviendas con piso de tierra) con V6 (Viviendas que no disponen de lavadora) y V7 (Viviendas que no disponen de refrigerador) tienen también una correlación alta que llega a 0,86\*\*\* y 0,88\*\*\* respectivamente.

- A medida que aumenta PT (Población total), V2 (viviendas que no disponen de excusado) y V3 (viviendas que no disponen de agua entubada de la red pública) disminuyen. Estos tienen un bajo nivel de correlación 0,10.

-Se recomienda analizar las variables V2 y V4 debido a su alto nivel de correlación (0,98\*\*\*)

>2010

- Se observa un aumento en la correlación entre las variables V2 y V4, ahora representativo en el 0,99\*\*\* respecto al IRS.

- V4 (Viviendas que no disponen drenaje) y V5 (viviendas que no disponen de energía eléctrica), presenta una fuerte correlación de 0,94\*\*\*.

- Se observa una fuerte correlación en variables de Vivienda respecto al IRS.

- En general las variables P (población) tienen valores más bajo (débiles), lo que es un indicador que en los últimos 5 años ha habido progresos respecto a estas variables entre sí.

- Las variables P y V respecto al IRS por lo general tienen altas correlaciones, sin embargo, la variable PT indica niveles muy bajos de correlación e incluso encontramos negativos que indican que el IRS retrocede a medida que crece la población total.

> 2015

-Se observa una relación creciente entre el Índice de Rezago Social y las variables excepto en "Población Total" y "Población sin derecho habiencia a servicios de salud", que van decreciendo al aumento de IRS.

-A mayor Índice de Rezago Social mayor es la cantidad de viviendas con alguna carencia como falta de drenaje, lavadora entre otras sumado a la población de edad de 15 años con problemas de educación como analfabetismo, ausentismo escolar y educación básica incompleta.

## SMART “ANÁLISIS ARLM”: Modelo Optimizado 2005

```
Call:
lm(formula = IRS ~ P1 + P2 + P3 + P4 + V1 + V2 + V3 + V5 + V6 +
    V7, data = yakuza05, na.action = na.exclude)

Residuals:
    Min       1Q   Median       3Q      Max
-0.199209 -0.045935  0.002357  0.054482  0.145325

Coefficients:
            Estimate Std. Error t value Pr(>|t|)
(Intercept) -2.2294222  0.0654537 -34.061  < 2e-16 ***
P1           0.0081229  0.0041874   1.940  0.054892 .
P2           0.0342447  0.0038590   8.874  1.22e-14 ***
P3           0.0039919  0.0014688   2.718  0.007606 **
P4           0.0035884  0.0005840   6.144  1.23e-08 ***
V1           0.0082083  0.0016839   4.874  3.59e-06 ***
V2           0.0070119  0.0016735   4.190  5.56e-05 ***
V3           0.0056253  0.0009545   5.893  3.99e-08 ***
V5           0.0057134  0.0020802   2.747  0.007010 **
V6           0.0060483  0.0016181   3.738  0.000293 ***
V7           0.0116598  0.0020848   5.593  1.58e-07 ***
---
signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 0.07682 on 113 degrees of freedom
(1 observation deleted due to missingness)
Multiple R-squared:  0.9824,    Adjusted R-squared:  0.9809
F-statistic: 632.1 on 10 and 113 DF,  p-value: < 2.2e-16
```

### Análisis:

- Al optimizar con el comando step, se observa que el modelo sugiere la eliminación de las potenciales variables predictoras “PT” (Población Total) y “V4” (Viviendas que no disponen drenaje).
- Se observa las potenciales variables predictoras: “P1”, “P2”, “P3”, “P4”, “V1”, “V2”, “V3”, “V5”, “V6”, “V7”. De ellas, P1, V2, V3 y V7 son las que muestran mayor correlación con la variable “IRS” y las variables restantes presentan una fuerza moderadamente alta.
- Dado la eliminación de las variables señaladas en el primer punto, el R2 ajustado aumenta de 0.9808% a 0.9809%. Esto muestra que el nuevo modelo ARLM propuesto representa una pequeña variación, aunque mayor en representatividad a la predicción.



## SMART “ANÁLISIS ARLM”: Modelo Optimizado 2010

```
> summary(modelo.opt)

Call:
lm(formula = IRS ~ ., data = yakuza10opt, na.action = na.exclude)

Residuals:
    Min       1Q   Median       3Q      Max
-0.07482 -0.01574  0.00028  0.01347  0.06835

Coefficients:
            Estimate Std. Error t value Pr(>|t|)
(Intercept) -2.2154125  0.0200376 -110.563 < 2e-16 ***
P1           0.0150012  0.0016524   9.078 3.88e-15 ***
P2           0.0384359  0.0013999  27.457 < 2e-16 ***
P3           0.0107709  0.0004561  23.616 < 2e-16 ***
V1           0.0119380  0.0011479  10.400 < 2e-16 ***
V2           0.0069751  0.0027159   2.568  0.01151 *
V3           0.0054031  0.0003492  15.471 < 2e-16 ***
V4           0.0074512  0.0025200   2.957  0.00378 **
V5           0.0164498  0.0014871  11.061 < 2e-16 ***
V6           0.0053424  0.0006083   8.783 1.87e-14 ***
V7           0.0066083  0.0008365   7.900 1.92e-12 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 0.02565 on 114 degrees of freedom
Multiple R-squared:  0.9982,    Adjusted R-squared:  0.9981
F-statistic: 6394 on 10 and 114 DF, p-value: < 2.2e-16
```

### Análisis:

- Al optimizar con el comando step, se observa que el modelo no sugiere la eliminación de ninguna variable, ya que el  $r$  ajustado en el modelo inicial es igual a uno, lo que quiere decir que existe una alta correlación entre las variables y que es un buen modelo. Debido a lo anterior se decide eliminar las variables con menos fuerza o menos “\*”, dejando fuera PT y P4.
- Dado la eliminación de las variables señaladas en el primer punto, el  $R^2$  ajustado disminuye de 1% a 0.9981%. Esto muestra que el nuevo modelo ARLM propuesto representa una pequeña variación y disminuye su correlación. Con esto podemos sugerir que no se elimine ninguna variable, ya que entre todas generan un muy buen modelo, donde todas logran una alta correlación con el índice de rezago social.



## SMART “ANÁLISIS ARLM”: Modelo Optimizado 2015

```
Call:
lm(formula = IRS ~ ., data = yakuza15opt, na.action = na.exclude)

Residuals:
    Min       1Q   Median       3Q      Max
-4.317e-06 -1.250e-06 -1.270e-08  1.137e-06  4.317e-06

Coefficients:
            Estimate Std. Error t value Pr(>|t|)
(Intercept) -1.876e+00  1.366e-06 -1373750  <2e-16 ***
P1           1.991e-02  1.463e-07   136099  <2e-16 ***
P2           2.615e-02  1.237e-07   211413  <2e-16 ***
P3           1.014e-02  3.479e-08   291583  <2e-16 ***
P4          -3.352e-03  3.371e-08   -99422  <2e-16 ***
V1           1.845e-02  8.899e-08   207339  <2e-16 ***
V2           5.096e-03  1.929e-07    26413  <2e-16 ***
V3           9.753e-03  1.017e-07    95853  <2e-16 ***
V4           6.806e-03  1.771e-07    38425  <2e-16 ***
V5           3.766e-02  1.396e-07   269779  <2e-16 ***
V6           6.902e-03  3.956e-08   174492  <2e-16 ***
V7           7.285e-03  6.342e-08   114871  <2e-16 ***
---
signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 1.728e-06 on 113 degrees of freedom
Multiple R-squared: 1, Adjusted R-squared: 1
F-statistic: 1.231e+12 on 11 and 113 DF, p-value: < 2.2e-16
```

### Análisis:

- Al optimizar con el comando step, se observa que el modelo no sugiere la eliminación de ninguna variable, ya que el r ajustado en el modelo inicial es igual a uno, lo que quiere decir que existe una alta correlación entre las variables y que es un buen modelo. Debido a lo anterior se decide eliminar las variables con menos fuerza o menos “\*”, dejando fuera PT.
- Dado la eliminación de la variable señalada en el primer punto, el R2 ajustado se mantiene en 1%. Con esto podemos sugerir que PT no es tan relevante para este modelo ya que entre en modelo principal y el optimizado el R2 ajustado se mantiene.

## SMART “ANÁLISIS ARLM”: Proyección 2020

```

Residuals:
    Min       1Q   Median       3Q      Max
-5.728e-07 -2.798e-07 -4.275e-08  2.448e-07  5.951e-07

Coefficients:
            Estimate Std. Error    t value Pr(>|t|)
(Intercept) -1.727e+00  2.545e-07 -6.786e+06 <2e-16 ***
PT           7.824e-14  1.833e-13  4.270e-01  0.67
P1           2.232e-02  2.802e-08  7.966e+05 <2e-16 ***
P2           1.671e-02  1.523e-08  1.098e+06 <2e-16 ***
P3           1.072e-02  7.011e-09  1.530e+06 <2e-16 ***
P4          -3.884e-03  3.353e-09 -1.158e+06 <2e-16 ***
V1           1.874e-02  2.027e-08  9.243e+05 <2e-16 ***
V2           1.491e-02  7.095e-08  2.102e+05 <2e-16 ***
V3           1.147e-02  2.734e-08  4.196e+05 <2e-16 ***
V4           7.704e-03  7.028e-08  1.096e+05 <2e-16 ***
V5           4.390e-02  4.819e-08  9.110e+05 <2e-16 ***
V6           7.227e-03  8.205e-09  8.807e+05 <2e-16 ***
V7           8.217e-03  1.571e-08  5.230e+05 <2e-16 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 3.156e-07 on 112 degrees of freedom
Multiple R-squared: 1, Adjusted R-squared: 1
F-statistic: 4.44e+13 on 12 and 112 DF, p-value: < 2.2e-16

```

Se elige el modelo optimizado del año 2015 debido a que éste posee un R ajustado de 1, indicando ser un modelo preciso y confiable para su uso.

Al estudiar los valores de correlación de los pares de variables respecto al IRS como variable dependiente, en el año 2020, se puede apreciar una disminución de la correlación de las variables P2 P3 Y P4, siendo esta última la con mayor variabilidad, y como es negativa, se deduce que el rango de edad del grupo P4 va disminuyendo al aumentar IRS, y como su relación lineal es baja, estas variables afectan débilmente al índice de rezago social.

Es distinto al caso de las correlaciones desde V1 A V7 las cuales aumentaron, lo que quiere decir que las viviendas con escasez tienen mucho que ver con el IRS, esto se explica debido a un contexto que altera enormemente la economía en el país de México, resultando en la pandemia de covid-19, logrando mayores índices de pobreza y por consecuencia IRS.

El analfabetismo está relacionado con la pobreza y el rezago social, ya que las personas tienen mayores dificultades para encontrar empleo y en la inserción social. Respecto al año 2015, el IRS es de -1.876 y al año 2020 aumenta a -1.727, se ve un avance muy leve en el desarrollo para ser 5 años.

## SMART “ANÁLISIS ARLM”: Conclusión Final&Recomendaciones

-Podemos concluir que del periodo 2005 al 2010 hubo un nulo avance, podemos inferir que es a causa de la crisis del año 2008. El grupo más afectado de manera positiva fue el P3 que dice relación con el acceso a la salud, por otra parte, tenemos las variables V1, V2, V3, V4, V5, V6 y V7, que son menos afectadas por IRS ya que su correlación se ve disminuida, es decir tenemos variables exógenas que están afectando a este modelo, como lo es una crisis económica.

-En el periodo 2015 hubo un avance respecto del periodo anterior analizado, sin embargo, se notan retrocesos en las variables V1, V2, V3, V4, V5, V6 y V7. Además, la correlación de P4 aumenta en relación con el periodo anterior, pero sigue siendo bajo.

Recomendación:

Se sugiere que el gobierno de México se enfoque en las variables de desarrollo de la vivienda, a través de remodelaciones, mejoras y otros proyectos que mejoren las condiciones de estas. Para mejorar la calidad de vida de las personas y bajar la pobreza.

## SMART “ANÁLISIS ARLM”: Comandos RStudio

```
#importar data
view(yakuza05)
attach(yakuza05)
names(yakuza05)
chart.correlation(yakuza05)
model=lm(formula=IRS~., data=yakuza05, na.action = na.exclude)
summary(model)
step(model, direction = "both", trace = 0)
#SE IMPORTA DATA OPTIMIZADA
view(yakuza05v1)
attach(yakuza05v1)
names(yakuza05v1)
chart.correlation(yakuza05v1)
model=lm(formula = IRS~., data=yakuza05v1, na.action = na.exclude)
summary(model)
#HISTOGRAMAS
hist(yakuza05$IRS, freq=FALSE, col="lightcyan", main="Histograma + Boxplot de PT",xlab="",
ylab="IRS")
lines(density(yakuza05$IRS), col="red",lwd="2")
par(new=TRUE)
boxplot(yakuza05$IRS, horizontal = TRUE, axes=FALSE,col=rgb(0,0.8,1,alpha=0.5))
> attach(yakuza10)
> yakuza.10= select(yakuza10,IRS,PT,P1,P2,P3,P4,V1,V2,V3,V4,V5,V6,V7)
> pairs(yakuza.10[,1:13])
> td<-yakuza.10
> m=td[,1:13]
> chart.Correlation(m)
> modelo<-lm(formula = IRS~.,data=yakuza.10, na.action = na.exclude)
> summary(modelo)
> step(modelo, direction = "both", trace = 0)
> yakuza10opt=select(yakuza.10, IRS,P1,P2,P3,V1,V2,V3,V4,V5,V6,V7)
> chart.Correlation(yakuza10opt)
> modelo.opt=lm(formula = IRS~.,yakuza10opt,na.action = na.exclude)
> summary(modelo.opt)
> hist(yakuza.10$IRS, freq=FALSE, col="lightgreen", main="Histograma + Boxplot de PT",xlab="",
ylab="IRS")
> lines(density(yakuza.10$IRS), col="red",lwd="2")
> par(new=TRUE)
> boxplot(yakuza.10$IRS, horizontal = TRUE, axes=FALSE,col=rgb(0,0.8,1,alpha=0.5))
,col=rgb(0,0.8,1,alpha=0.5))
```

## SMART “ANÁLISIS ARLM”: Comandos RStudio

```
> attach(yakuza15)
> yakuza.15= select(yakuza10,IRS,PT,P1,P2,P3,P4,V1,V2,V3,V4,V5,V6,V7)
> pairs(yakuza.15[,1:13])
> td<-yakuza.15
> m=td[,1:13]
> chart.Correlation(m)
> modelo<-lm(formula = IRS~.,data=yakuza.15, na.action = na.exclude)
> summary(modelo)
> step(modelo, direction = "both", trace = 0)
> yakuza15opt=select(yakuza.15, IRS,P1,P2,P3,P4,V1,V2,V3,V4,V5,V6,V7)
> chart.Correlation(yakuza15opt)
> modelo.opt=lm(formula = IRS~.,yakuza15opt,na.action = na.exclude)
> summary(modelo.opt)
> hist(yakuza.15$IRS, freq=FALSE, col="lightcyan", main="Histograma + Boxplot de PT",xlab="",
ylab="IRS")
> lines(density(yakuza.15$IRS), col="red",lwd="2")
> par(new=TRUE)
> boxplot(yakuza.15$IRS, horizontal = TRUE, axes=FALSE,col=rgb(0,0.8,1,alpha=0.5))

> View(yakuza20)
> attach(yakuza20)
> names(yakuza20)
> chart.correlation(yakuza20)
> model=lm(formula = IRS~.,data = yakuza20, na.action = na.exclude)
> summary(model)
```

## SMART “ANÁLISIS ARLM”: Memoria de Cálculo (Proyecciones)

Coefficients: -1.876e+00,

P1      1.991e-02 + P2      2.615e-02 + P3      1.014e-02 + P4      -3.352e-03 + V1      1.845e-02 + V2      5.096e-03 + V3      9.753e-03 + V4      6.806e-03 + V5      3.766e-02 + V6      6.902e-03 + V7      7.285e-03

Coefficients: -1.727e+00 + PT      7.824e-14 + P1      2.232e-02 + P2      1.671e-02 + P3      1.072e-02 + P4      -3.884e-03 + V1      1.874e-02 + V2      1.491e-02 + V3      1.147e-02 + V4      7.704e-03 + V5      4.390e-02 + V6      7.227e-03 + V7      8.217e-03

## SMART “ANÁLISIS ARLM”: Análisis Estadístico Adicional

(Solo en este recuadro y página).

## SMART “ANÁLISIS ARLM”: Análisis Estadístico Adicional

(Solo en este recuadro y página).