

Análisis Bootstrap de las Variables P901 (Crédito) y P905 (Seguro)

Análisis Estadístico de Datos de Acceso a Crédito y Seguro (ENA 2014-2024)

Elaborado con datos de `data_seleccionada.csv`

1 Descripción de los Datos

1.1 Resumen del Conjunto de Datos

El conjunto de datos, obtenido de `data_seleccionada.csv`, contiene 14,220 observaciones y seis variables: P901 (Crédito), P905 (Seguro), CCPP, NOMBREPV, CCDI y NOMBREDI. Las variables de interés, P901 y P905, son binarias, donde 1 indica “Sí” y 2 indica “No” para el acceso a crédito y seguro, respectivamente.

1.1.1 Estadísticas Descriptivas

- **P901 (Crédito):**

- Mínimo: 1
- 1er Cuartil: 2
- Mediana: 2
- Media: 1.912
- 3er Cuartil: 2
- Máximo: 2
- Valores NA: 2

- **P905 (Seguro):**

- Mínimo: 1
- 1er Cuartil: 2
- Mediana: 2
- Media: 1.979
- 3er Cuartil: 2
- Máximo: 2
- Valores NA: 2

1.1.2 Distribución de Frecuencias

- **P901 (Crédito):**

- 1 (Sí): 1,250
- 2 (No): 12,968
- NA: 2

- **P905 (Seguro):**

- 1 (Sí): 294
- 2 (No): 13,924
- NA: 2

2 Análisis Bootstrap

2.1 Bootstrap Simple

Se realizó un bootstrap simple con 1,000 remuestreos para estimar la variabilidad de las medias muestrales de P901 y P905. A continuación, se presenta el código en R utilizado:

```
1 bootstrap_simple <- function(data, n_bootstrap = 1000) {  
2   n <- nrow(data)  
3   bootstrap_samples <- list()  
4   for (i in 1:n_bootstrap) {  
5     indices <- sample(1:n, n, replace = TRUE)  
6     bootstrap_samples[[i]] <- data[indices, ]  
7   }  
8   return(bootstrap_samples)  
9 }
```

2.2 Bootstrap para Proporciones (P901 - Crédito)

Se calculó la proporción de individuos con crédito (P901 = 1) para cada muestra bootstrap. El código en R es el siguiente:

```
1 bootstrap_proporcion_credito <- function(data, n_bootstrap = 1000) {  
2   proporciones <- numeric(n_bootstrap)  
3   n <- nrow(data)  
4   for (i in 1:n_bootstrap) {  
5     indices <- sample(1:n, n, replace = TRUE)  
6     muestra <- data[indices, ]  
7     proporciones[i] <- mean(muestra$P901, na.rm = TRUE)  
8   }  
9   return(proporciones)  
10 }
```

2.3 Bootstrap para Proporciones (P905 - Seguro)

De manera similar, se calculó la proporción de individuos con seguro (P905 = 1):

```
1 bootstrap_proporcion_seguro <- function(data, n_bootstrap = 1000) {  
2   proporciones <- numeric(n_bootstrap)  
3   n <- nrow(data)  
4   for (i in 1:n_bootstrap) {  
5     indices <- sample(1:n, n, replace = TRUE)  
6     muestra <- data[indices, ]  
7     proporciones[i] <- mean(muestra$P905, na.rm = TRUE)  
8   }  
9   return(proporciones)  
10 }
```

2.4 Resultados Bootstrap con el Paquete boot

Se utilizó el paquete boot para calcular las estadísticas bootstrap de ambas variables:

```
1 estadistico_credito <- function(data, indices) {  
2   return(mean(data[indices, ]$P901, na.rm = TRUE))  
3 }  
4 estadistico_seguro <- function(data, indices) {  
5   return(mean(data[indices, ]$P905, na.rm = TRUE))  
6 }  
7 boot_credito <- boot(datos, estadistico_credito, R = 1000)  
8 boot_seguro <- boot(datos, estadistico_seguro, R = 1000)
```

2.4.1 Intervalos de Confianza

Se calcularon los intervalos de confianza del 95% para las medias de P901 y P905 utilizando varios métodos (Normal, Básico, Percentil y BCa):

- **P901 (Crédito):**

- Normal: (1.907, 1.917)
- Básico: (1.907, 1.917)
- Percentil: (1.907, 1.917)
- BCa: (1.908, 1.917)

- **P905 (Seguro):**

- Normal: (1.977, 1.982)
- Básico: (1.977, 1.982)
- Percentil: (1.977, 1.982)
- BCa: (1.977, 1.981)

2.5 Bootstrap Estratificado

Se realizó un bootstrap estratificado por año y distrito:

```
1 bootstrap_estratificado_anio <- function(data, n_bootstrap = 1000) {
2   bootstrap_samples <- list()
3   for (i in 1:n_bootstrap) {
4     muestra_estratificada <- data %>%
5       group_by(anio) %>%
6       sample_n(n(), replace = TRUE) %>%
7       ungroup()
8     bootstrap_samples[[i]] <- muestra_estratificada
9   }
10  return(bootstrap_samples)
11 }
12
13 bootstrap_estratificado_distrito <- function(data, n_bootstrap =
14 1000) {
15   bootstrap_samples <- list()
16   for (i in 1:n_bootstrap) {
17     muestra_estratificada <- data %>%
18       group_by(NOMBREDI) %>%
19       sample_n(n(), replace = TRUE) %>%
20       ungroup()
21     bootstrap_samples[[i]] <- muestra_estratificada
22   }
23   return(bootstrap_samples)
24 }
```

2.6 Análisis de Proporciones Bootstrap

Se utilizó la siguiente función para resumir los resultados bootstrap:

```
1 analizar_proporciones <- function(proporciones, variable_name) {
2   resultados <- list(
3     variable = variable_name,
4     media_proporcion = mean(proporciones, na.rm = TRUE),
5     mediana_proporcion = median(proporciones, na.rm = TRUE),
6     desviacion_estandar = sd(proporciones, na.rm = TRUE),
7     error_estandar = sd(proporciones, na.rm = TRUE),
8     intervalo_confianza_95 = quantile(proporciones, c(0.025,
9       0.975), na.rm = TRUE),
10    min_valor = min(proporciones, na.rm = TRUE),
11    max_valor = max(proporciones, na.rm = TRUE)
12  )
13  return(resultados)
14 }
```

2.6.1 Resultados Resumidos

- P901 (Crédito):

- Media de la proporción: 1.912
- Mediana de la proporción: 1.912
- Desviación Estándar: 0.00245
- Error Estándar: 0.00245
- Intervalo de Confianza 95%: (1.907, 1.917)
- Mínimo: 1.905
- Máximo: 1.919

- **P905 (Seguro):**

- Media de la proporción: 1.979
- Mediana de la proporción: 1.979
- Desviación Estándar: 0.00115
- Error Estándar: 0.00115
- Intervalo de Confianza 95%: (1.977, 1.982)
- Mínimo: 1.976
- Máximo: 1.983

2.7 Análisis de Correlación

Se estimó la correlación entre P901 y P905 mediante bootstrap:

```

1 bootstrap_correlacion <- function(data, n_bootstrap = 1000) {
2   correlaciones <- numeric(n_bootstrap)
3   n <- nrow(data)
4   for (i in 1:n_bootstrap) {
5     indices <- sample(1:n, n, replace = TRUE)
6     muestra <- data[indices, ]
7     correlaciones[i] <- cor(muestra$P901, muestra$P905, use = "
      complete.obs")
8   }
9   return(correlaciones)
10 }
```

2.7.1 Resultados de la Correlación

- Variable: Correlación P901-P905
- Media de la proporción: 0.0684
- Mediana de la proporción: 0.0680
- Desviación Estándar: 0.0118
- Error Estándar: 0.0118
- Intervalo de Confianza 95%: (0.0458, 0.0914)

- Mínimo: 0.0295
- Máximo: 0.1116

2.8 Visualización de Resultados

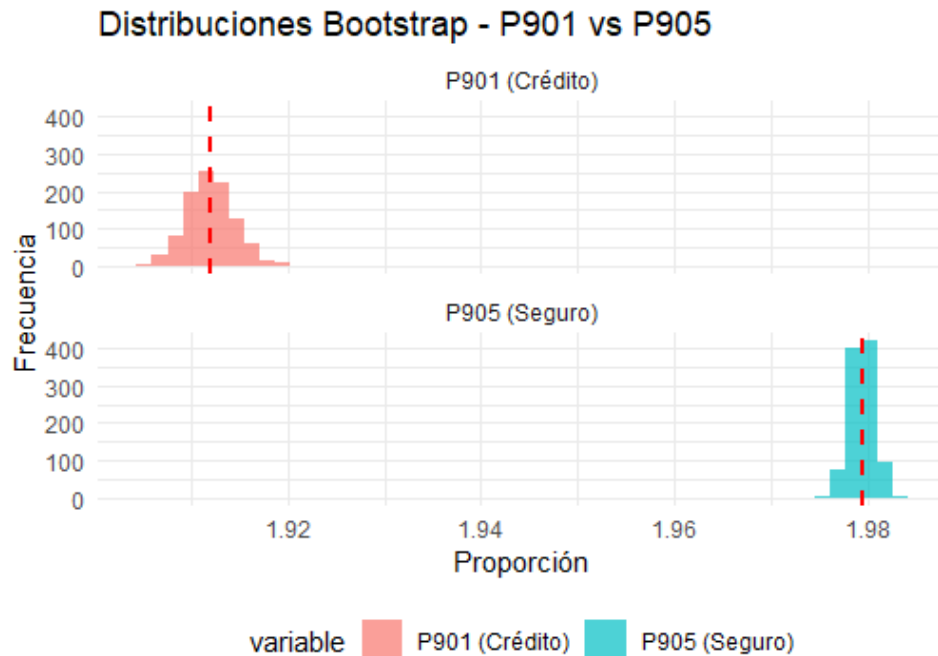


Figure 1: Distribuciones Bootstrap de P901 (Crédito) y P905 (Seguro). La línea roja discontinua indica la media de cada distribución.

2.9 Tabla Resumen

A continuación, se presenta un resumen completo de los resultados bootstrap:

Table 1: Resumen de Resultados Bootstrap

Variable	Media	Error Estándar	Límite Inferior IC 95%	Límite Superior
P901 (Crédito)	1.912	0.00245	1.907	1.917
P905 (Seguro)	1.979	0.00115	1.977	1.982
Correlación P901-P905	0.0684	0.0118	0.0458	0.0914

3 Conclusión

El análisis bootstrap proporciona estimaciones robustas de las proporciones y la correlación entre P901 (Crédito) y P905 (Seguro). Los resultados indican una correlación positiva baja entre las dos variables, con intervalos de confianza estrechos para las medias respectivas, lo que sugiere estimaciones confiables.

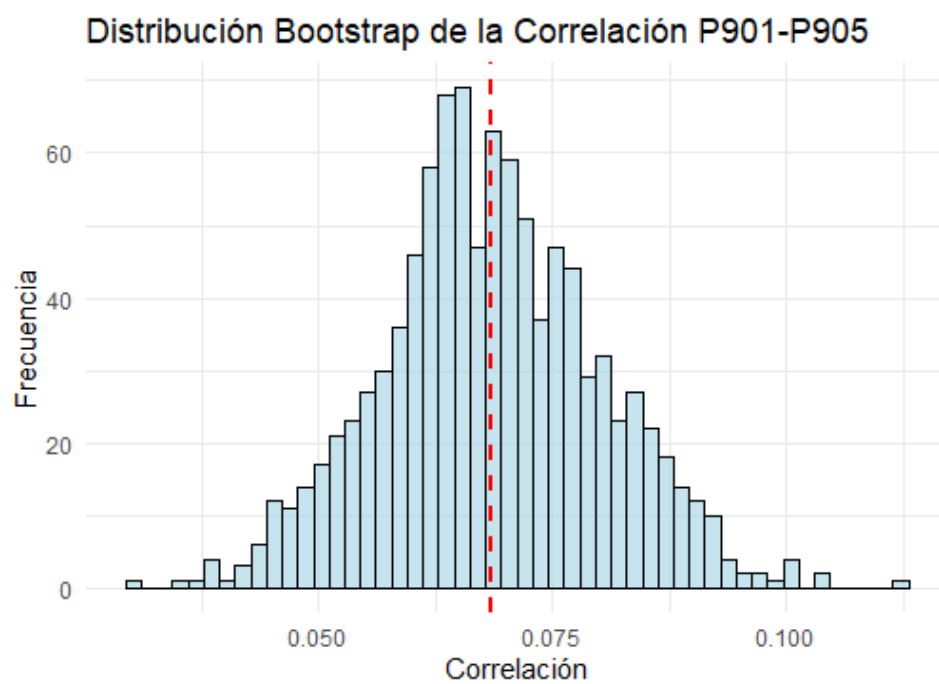


Figure 2: Distribución Bootstrap de la Correlación entre P901 y P905. La línea roja discontinua indica la media de la correlación.