

### 1. Análisis de Tendencia Central y Dispersión:

Calcule las medidas de tendencia central (media, mediana) y dispersión (varianza, desviación estándar) para las variables IngresoMensual, GastoMensual, y EdadPromedioHogar. ¿Los valores encontrados son consistentes con lo que esperarías en un contexto económico real?

**Ingreso Mensual:** Media = 91.720, Mediana = 91.056. Dispersión = 0.355.

**Gasto Mensual:** Media = 62.151, Mediana = 62.352. Dispersión = 0.402.

**Edad Promedio Hogar:** Media = 40.19, Mediana = 38, con dispersión = 0.41.

En general, las medidas son consistentes con un escenario realista, salvo algunos valores extremos

### 2. Verificación de Máximos y Mínimos:

Identifique los valores máximos y mínimos para IngresoMensual, GastoMensual, y EdadPromedioHogar. ¿Estos valores son realistas? Discuta cualquier valor que parezca inusual o fuera de lo esperado.

- **Ingreso Mensual:** Mín = 40.607, Máx = 148.557.

Son altos, pero no imposibles para un rango amplio de hogares.

- **Gasto Mensual:** Mín = 20.623, Máx = 127.691.

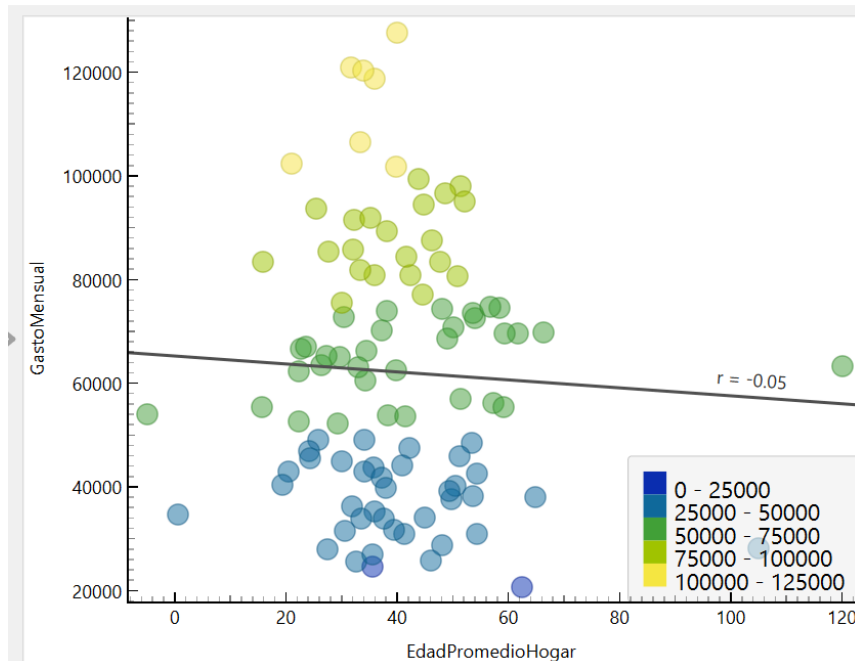
Gastar más de 120.000 puede considerarse muy alto, pero puede ser en hogares de mayor ingreso.

- **Edad Promedio Hogar:** Mín = -5 valor sin sentido y Máx = 120 poco real a menos que hay un hogar de personas muy ancianas.

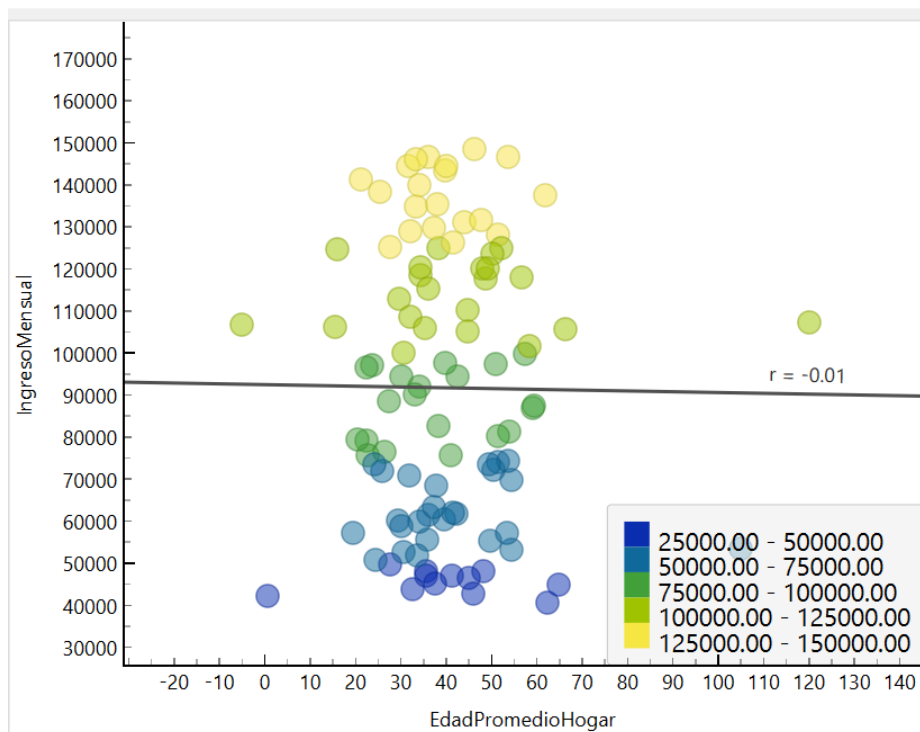
Los extremos en Edad Promedio Hogar son poco realistas y sugieren datos atípicos o errores.

### 3. Relación entre Variables:

Explore la relación entre el EdadPromedioHogar y otras variables numéricas como IngresoMensual y GastoMensual. ¿Hay alguna tendencia visible que indique cómo la edad promedio del hogar afecta los ingresos y gastos?



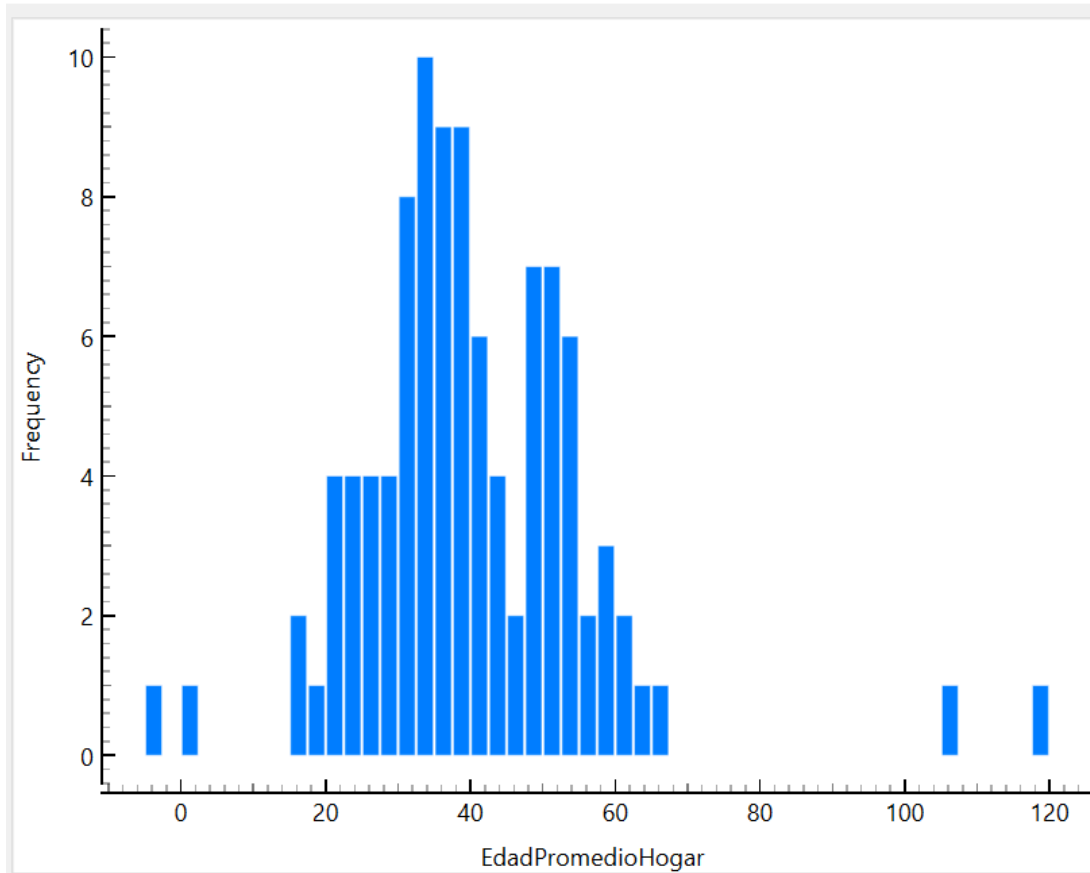
El coeficiente de correlación es de -0.05 que indica una **relación negativa demasiado débil**, casi inexistente. Este valor, que es muy cercano a cero, significa que **no existe** una relación lineal significativa entre GastoMensual y EdadPromedioHogar.



El coeficiente de correlación es **-0.01** prácticamente nulo, por lo tanto no hay correlación entre las variables EdadPromedioHogar e IngresoMensual.

#### 4. Análisis de Outliers:

Utilizando técnicas gráficas, identificar datos atípicos en las variables numéricas. Describa por qué estos puntos podrían considerarse atípicos y qué factores podrían explicar estos valores.

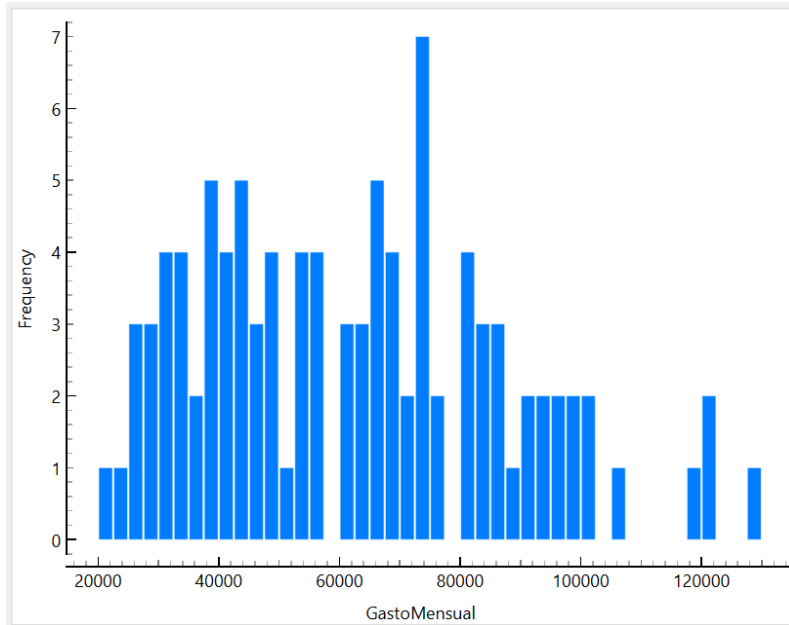


La mayoría de los hogares tienen edades promedio entre 20 y 65 años, con un pico alrededor de los 35.

También, aparecen barras aisladas en valores negativos -5 y muy altos entre 100 y 120, que son claros **outliers**, los cuales son valores atípicos.

Los valores negativos son claramente errores graves de codificación, ya que son imposibles. Estos deben ser corregidos o eliminados urgentemente.

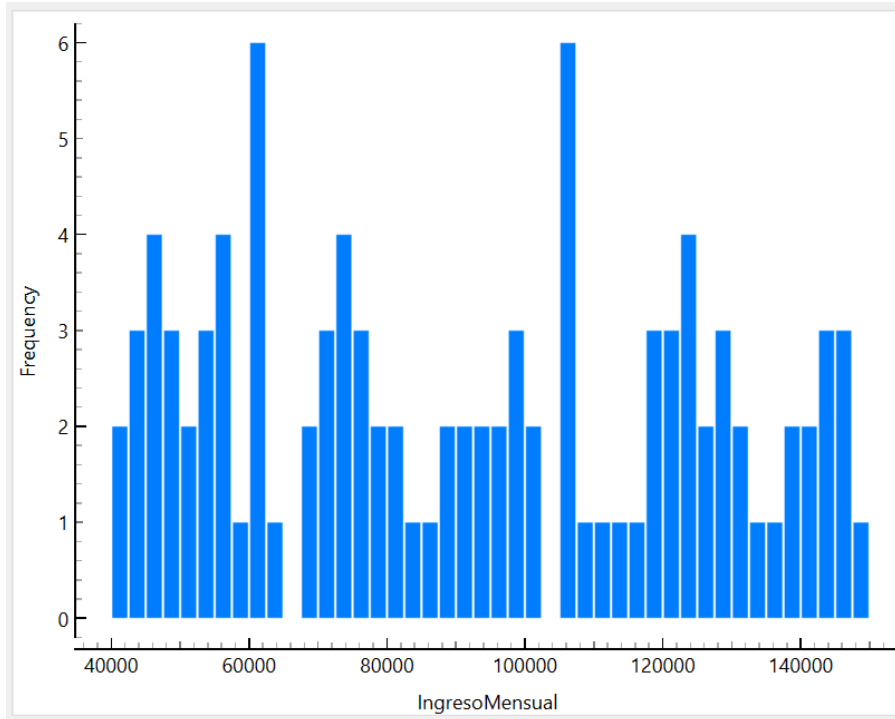
Por otro lado, los valores muy altos que están entre 100 y 120, podrían representar algún error, pero también pueden ser datos reales que representen hogares muy longevos.



El gráfico muestra que la mayor parte de los gastos se concentran entre \$35.000 y \$80.000 pesos, con un pico cercano a los \$75.000. Pero también aparecen valores en los extremos, con gastos muy bajos de alrededor de \$20.000 y hogares con gastos muy altos cerca de \$120.000. Los outliers en este caso son los valores entre \$105.000 y \$140.000.

Estos casos se apartan del comportamiento central y se consideran atípicos pero no son algo imposible.

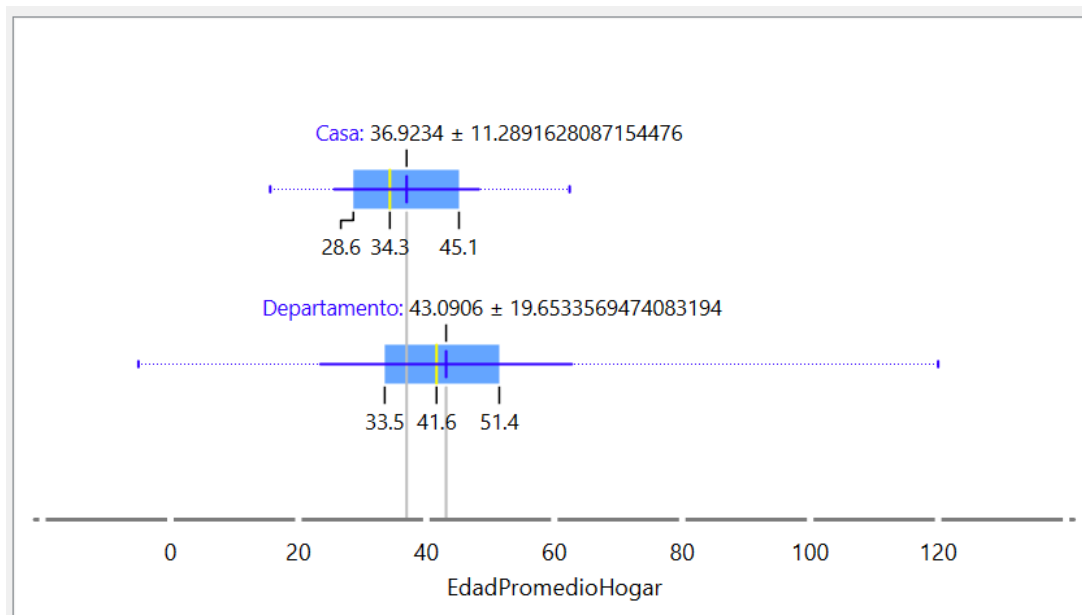
En conclusión, este gráfico muestra una gran variabilidad en los gastos mensuales.



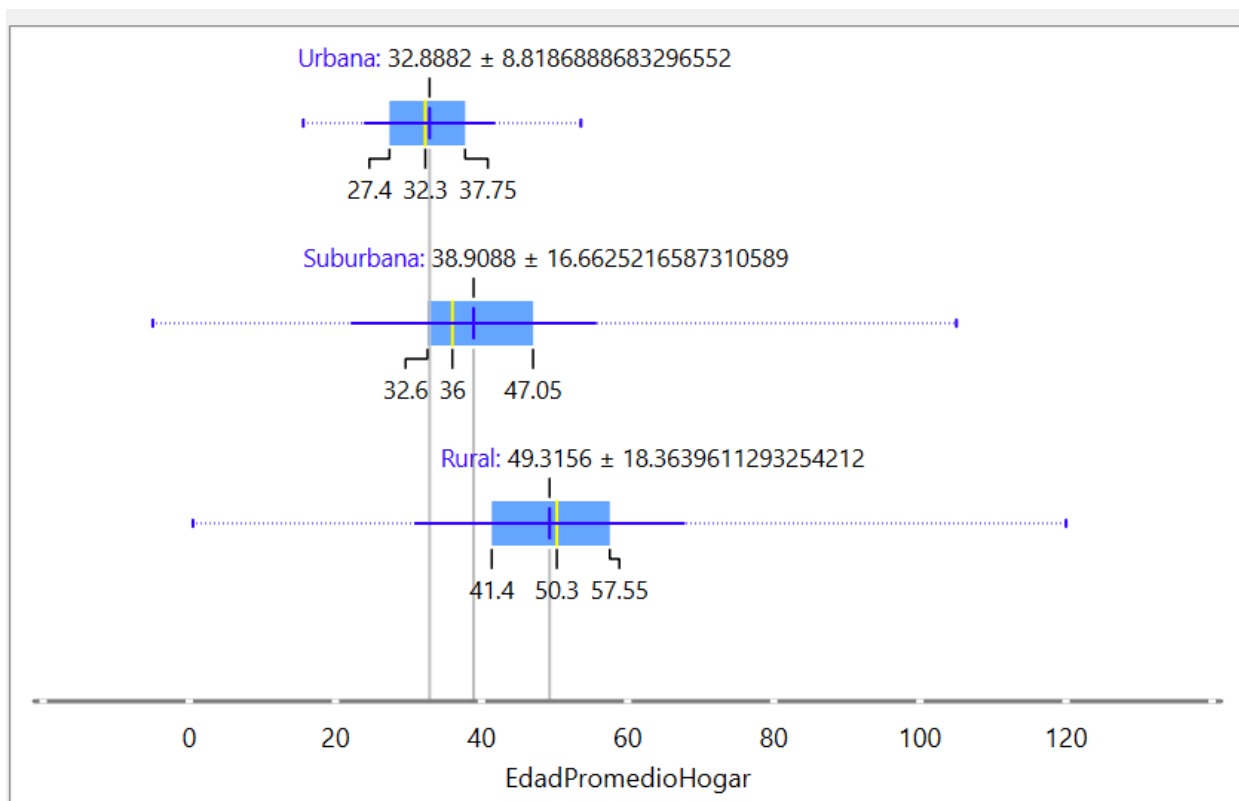
El histograma de Ingreso Mensual es muy irregular, con picos principales cerca de los \$60,000 y \$105,000. Esta forma atípica sugiere que la muestra incluye varios grupos de ingresos distintos. Además, no hay outliers extremos y aislados.

### 5. Comparación entre Grupos:

Analice cómo varía la EdadPromedioHogar entre diferentes tipos de vivienda (Casa, Departamento) y zonas de vivienda (Urbana, Suburbana, Rural). ¿Existen diferencias significativas que puedan sugerir patrones demográficos?



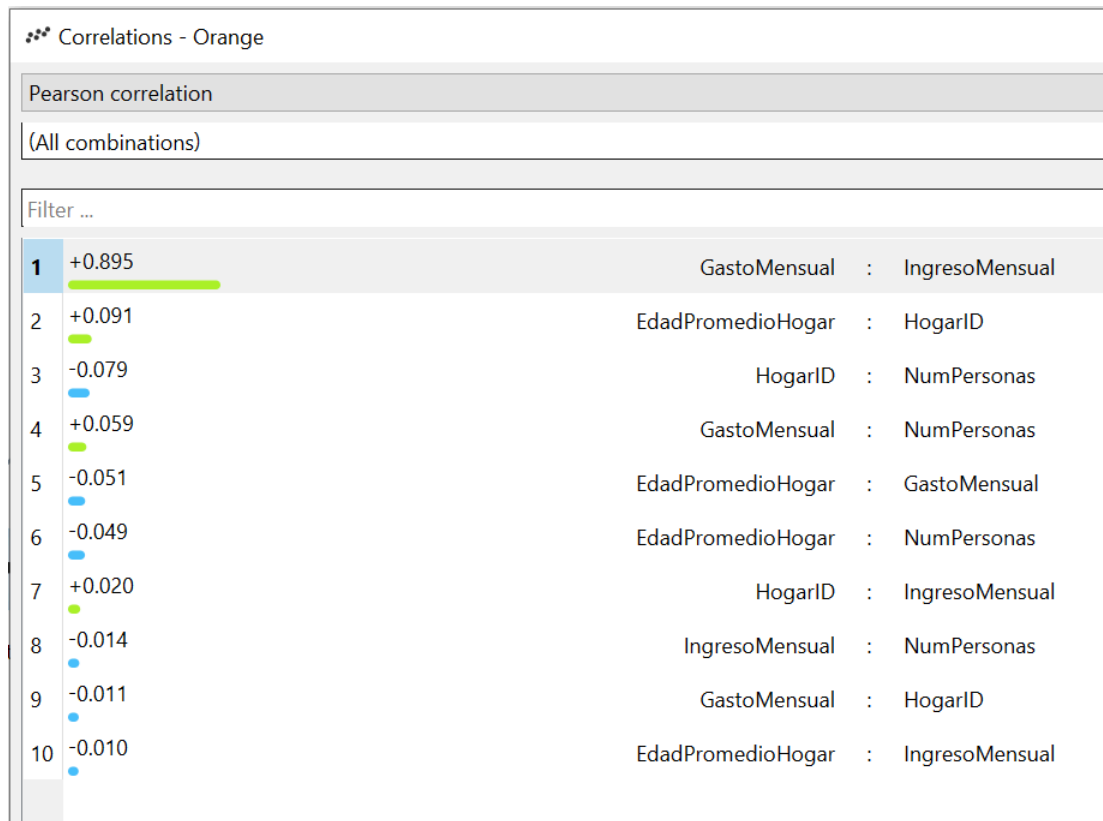
Podría esperarse que las casas estén asociadas a familias establecidas y de mayor edad, pero en estos datos pasa lo contrario: los hogares en departamento presentan mayor edad promedio y mayor dispersión que los de casa.



Se ve un patrón claro: la edad promedio del hogar aumenta al pasar de zonas urbanas → suburbanas → rurales. Esto tiene sentido con la población: zonas urbanas concentran hogares jóvenes, mientras que en áreas rurales predominan hogares más viejos.

## 6. Correlaciones:

Investigue la correlación entre las variables numéricas. ¿Qué relaciones se pueden identificar como relevantes y cómo se interpretan?



Analizando las correlaciones podemos ver:

- La relación más fuerte es entre GastoMensual e IngresoMensual de +0.895, que es la relación más clara y lógica del dataset, ya que el nivel de gasto depende directamente de la capacidad de ingresos del hogar.
- También vemos relaciones muy débiles o casi nulas como
  - EdadPromedioHogar y GastoMensual:  $r = -0.051$
  - EdadPromedioHogar e IngresoMensual:  $r = -0.010$
  - EdadPromedioHogar y NumPersonas:  $r = -0.049$
  - Todas estas anteriores, indican que casi no hay relación entre ellas y por lo tanto la edad promedio del hogar no influye ni en los ingresos, ni en los gastos, ni en el número de integrantes. Esto nos parece un poco raro ya que supusimos que un mayor edad promedio del hogar afectará de alguna forma en el ingreso.