

Informe de Análisis Exploratorio de **swiss**

Melani Forsythe Matos
Daniela Guerrero Álvarez
Rubén Martínez Rojas

1 Introducción

El dataset “swiss” contiene datos sobre 47 cantones suizos en 1888. Este análisis proporciona una visión detallada de las variables relacionadas con la fertilidad, la agricultura, y la educación.

2 Descripción de las Variables

El dataset “swiss” incluye las siguientes variables:

- **Fertility:** Tasa de fertilidad (número de hijos por mujer).
- **Agriculture:** Proporción de trabajadores en agricultura.
- **Examination:** Proporción de jóvenes con educación secundaria.
- **Education:** Proporción de población con educación secundaria.
- **Catholic:** Proporción de población católica.
- **Infant.Mortality:** Tasa de mortalidad infantil (número de muertes por 1000 nacidos vivos).

Análisis Descriptivo

1. Fertility:

- La media es 70.14, indicando el promedio del índice de fertilidad.
- La mediana es 70.4, lo que sugiere que la mitad de los valores están por debajo de este nivel.
- La moda es 65, el valor más frecuente en los datos de fertilidad.
- La simetría negativa (-0.46) sugiere una ligera inclinación hacia la izquierda, lo que implica que hay más valores altos que bajos.

Media	Mediana	Moda	Simetría	Curtosis
70.1425532	70.4000000	65.0000000	-0.4556871	3.2599542

Figure 1: Medidas centrales de Fertilidad

Varianza	Desviación Estándar	Rango	Coefficiente de Variación
156.04250	12.49170	57.50000	17.80901

Figure 2: Medidas de dispersión de Fertilidad

- La curtosis de 3.26 indica una distribución ligeramente más apuntada que una distribución normal.
- La varianza es 156.04, reflejando una dispersión moderada de los datos.
- La desviación estándar es 12.49, mostrando una variabilidad significativa en los índices de fertilidad.
- El rango es 57.5, lo que indica una diferencia considerable entre el valor más alto y el más bajo.
- El coeficiente de variación es 17.81, lo que indica una variabilidad del 17.8% respecto a la media.

2. Agriculture:

- La media es 50.66, lo que representa el porcentaje promedio de personas dedicadas a la agricultura.
- La mediana es 54.1, lo que sugiere que la mitad de las observaciones tienen un valor menor o igual a 54.1.
- La moda es 1.2, el valor más común en los datos de agricultura.
- La simetría negativa (-0.32) indica una ligera inclinación hacia la izquierda.
- La curtosis de 2.11 sugiere una distribución algo más plana en comparación con una distribución normal.

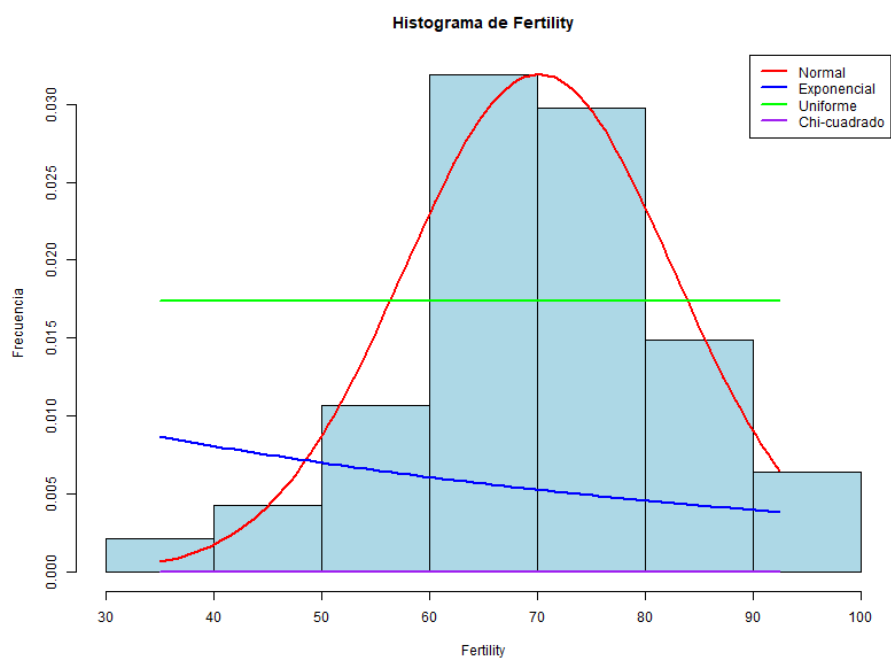


Figure 3: Distribución de la Tasa de Fertilidad

Media	Mediana	Moda	Simetría	Curtosis
50.6595745	54.1000000	1.2000000	-0.3203637	2.1144729

Figure 4: Medidas centrales de Agricultura

Varianza	Desviación Estándar	Rango	Coefficiente de Variación
515.79942	22.71122	88.50000	44.83105

Figure 5: Medidas de dispersión de Agricultura

- La varianza es 515.8, reflejando una alta dispersión en los datos de agricultura.
- La desviación estándar es 22.71, mostrando una alta variabilidad respecto a la media.
- El rango es 88.5, indicando una gran diferencia entre el valor más alto y el más bajo.
- El coeficiente de variación es 44.83, lo que indica una alta variabilidad respecto a la media.

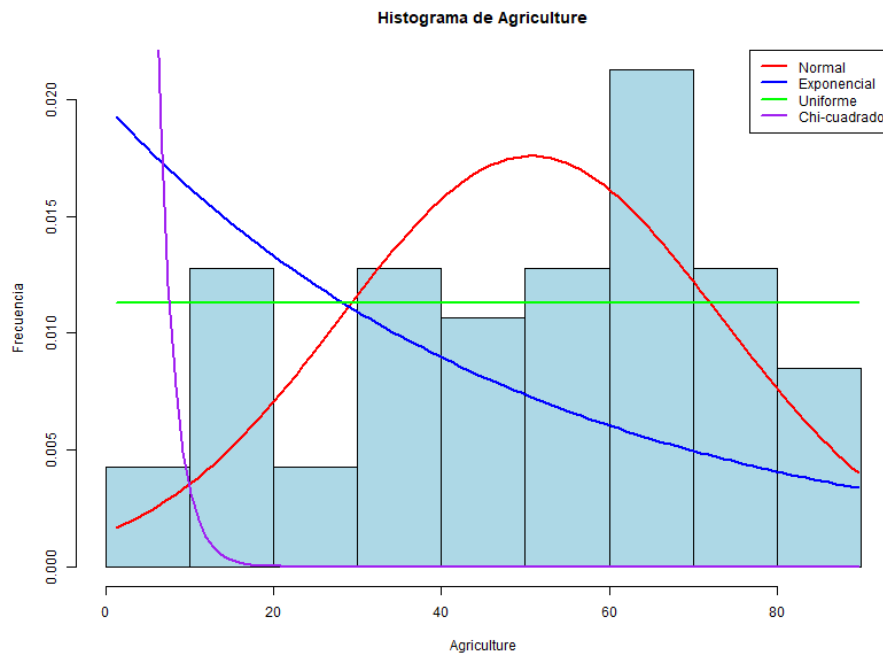


Figure 6: Distribución de la Proporción de Trabajadores en Agricultura

3. Examination:

- La media es 16.49, representando el promedio de la tasa de aprobación en exámenes.
- La mediana es 16, lo que sugiere que la mitad de los valores están por debajo de este nivel.
- La moda es 14, el valor más frecuente.
- La simetría positiva (0.45) sugiere una ligera inclinación hacia la derecha.

Media	Mediana	Moda	Simetría	Curtosis
16.4893617	16.0000000	14.0000000	0.4463996	2.8630636

Figure 7: Medidas centrales de Exámenes

Varianza	Desviación Estándar	Rango	Coefficiente de Variación
63.646623	7.977883	34.000000	48.382003

Figure 8: Medidas de dispersión de Exámenes

- La curtosis de 2.86 indica una distribución ligeramente más apuntada que la normal.
- La varianza es 63.65, reflejando una dispersión considerable.
- La desviación estándar es 7.98, indicando una variabilidad moderada.
- El rango es 34, reflejando una gran diferencia entre el valor más alto y el más bajo.
- El coeficiente de variación es 48.38, lo que indica una alta variabilidad respecto a la media.

4. Education:

- La media es 10.98, lo que indica el promedio del nivel de educación.
- La mediana es 8, sugiriendo que la mitad de los valores están por debajo de este nivel.
- La moda es 7, el valor más frecuente.
- La simetría positiva (2.27) indica una fuerte inclinación hacia la derecha, con una cola larga hacia los valores altos.
- La curtosis de 9.14 sugiere una distribución extremadamente apuntada.
- La varianza es 92.46, reflejando una dispersión considerable en los niveles de educación.

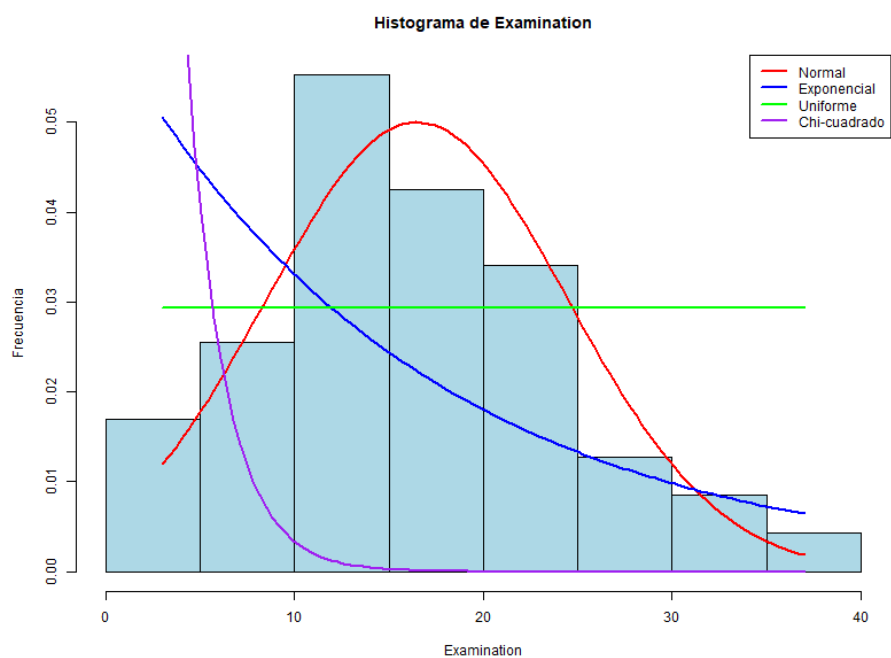


Figure 9: Distribución de la Proporción de Jóvenes con Educación Secundaria

Media	Mediana	Moda	Simetría	Curtosis
10.978723	8.000000	7.000000	2.268439	9.139735

Figure 10: Medidas centrales de Educación

Varianza	Desviación Estándar	Rango	Coeficiente de Variación
92.456059	9.615407	52.000000	87.582199

Figure 11: Medidas de dispersión de Educación

- La desviación estándar es 9.62, mostrando una alta variabilidad.
- El rango es 52, reflejando una gran diferencia entre el valor más alto y el más bajo.
- El coeficiente de variación es 87.58, indicando una variabilidad muy alta respecto a la media.

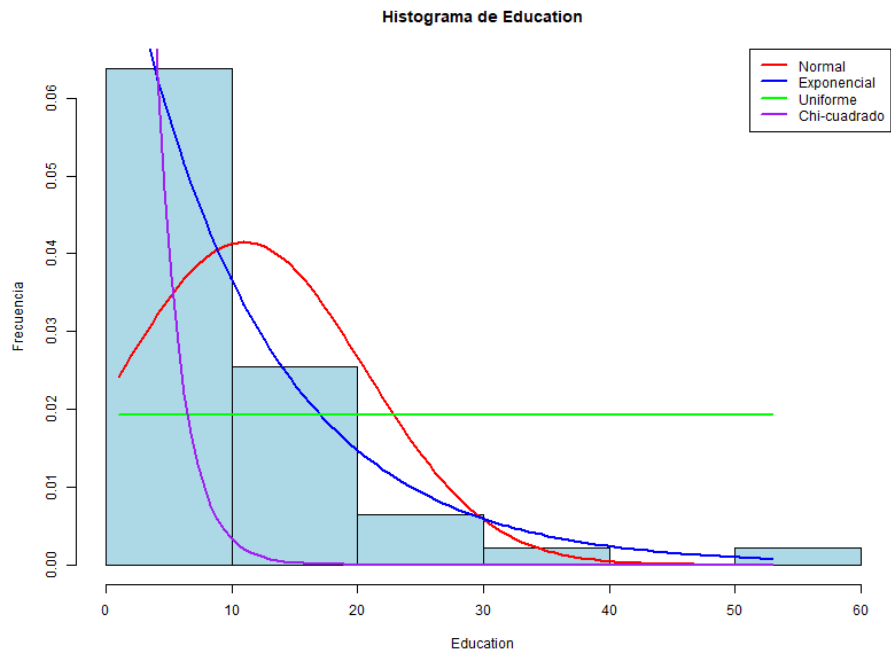


Figure 12: Distribución de la Proporción de Población con Educación Secundaria

5. Catholic:

Media	Mediana	Moda	Simetría	Curtosis
41.1438298	15.1400000	5.2300000	0.4789257	1.3345805

Figure 13: Medidas centrales de Población Católica

Varianza	Desviación Estándar	Rango	Coefficiente de Variación
1739.29454	41.70485	97.85000	101.36356

Figure 14: Medidas de dispersión de Población Católica

- La media es 41.14, representando el porcentaje promedio de población católica.
- La mediana es 15.14, lo que sugiere que la mitad de las observaciones están por debajo de este valor.
- La moda es 5.23, el valor más común.
- La simetría positiva (0.48) indica una ligera inclinación hacia la derecha.
- La curtosis de 1.33 indica una distribución más plana que una distribución normal.
- La varianza es 1739.29, reflejando una dispersión extremadamente alta en los porcentajes de población católica.
- La desviación estándar es 41.70, mostrando una gran variabilidad.
- El rango es 97.85, lo que refleja una gran diferencia entre el valor más alto y el más bajo.
- El coeficiente de variación es 101.36, lo que indica una variabilidad extremadamente alta respecto a la media.

6. Infant Mortality:

- La media es 19.94, lo que indica el promedio de la tasa de mortalidad infantil.
- La mediana es 20, lo que sugiere que la mitad de los valores están por debajo de este nivel.
- La moda es 18, el valor más frecuente.
- La simetría negativa (-0.33) sugiere una ligera inclinación hacia la izquierda.
- La curtosis de 3.78 indica una distribución algo más apuntada que una distribución normal.
- La varianza es 8.48, reflejando una dispersión baja en la tasa de mortalidad infantil.

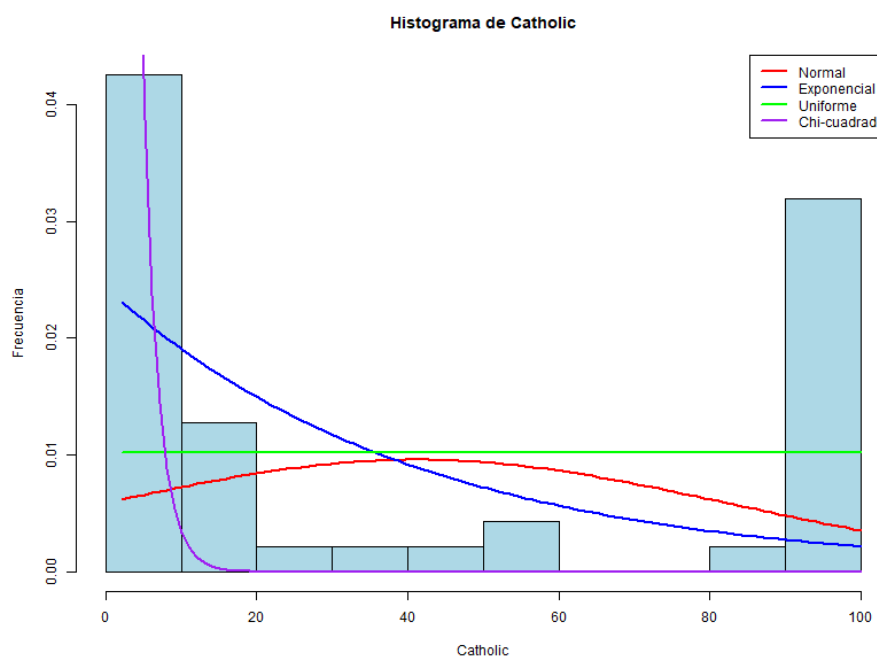


Figure 15: Distribución de la Proporción de Población Católica

Media	Mediana	Moda	Simetría	Curtosis
19.9425532	20.0000000	18.0000000	-0.3314326	3.7772868

Figure 16: Medidas centrales de Mortalidad Infantil

Varianza	Desviación Estándar	Rango	Coeficiente de Variación
8.483802	2.912697	15.800000	14.605435

Figure 17: Medidas de dispersión de Mortalidad Infantil

- La desviación estándar es 2.91, mostrando una baja variabilidad respecto a la media.
- El rango es 15.8, indicando una diferencia moderada entre el valor más alto y el más bajo.
- El coeficiente de variación es 14.61, lo que sugiere una baja variabilidad respecto a la media.

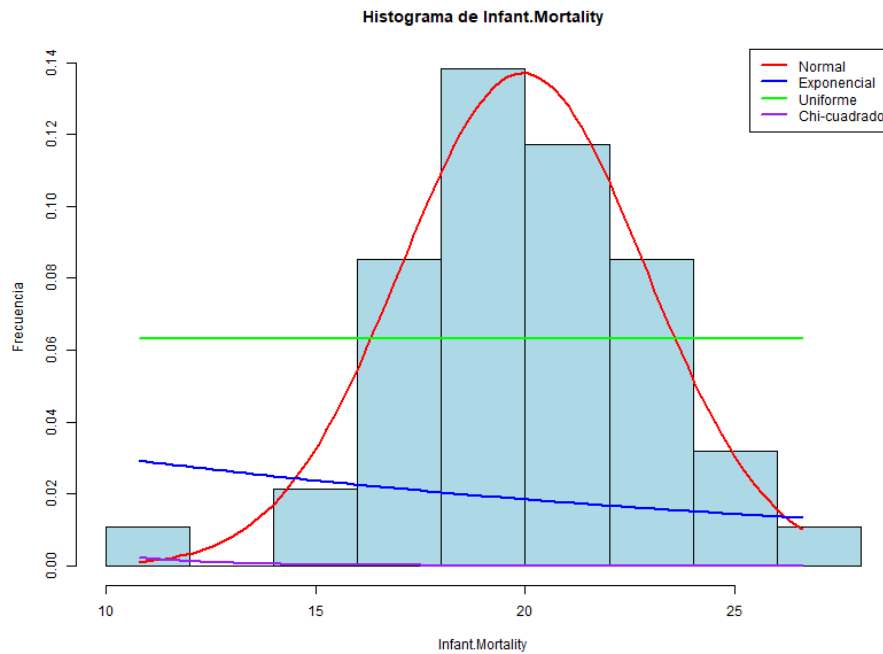


Figure 18: Distribución de la Tasa de Mortalidad Infantil

3 Conclusión

El análisis del dataset “swiss” muestra que las variables relacionadas con la fertilidad, la agricultura, y la educación tienen diferentes patrones de distribución y dispersión. La variabilidad en estas variables sugiere que los cantones suizos en 1888 tenían una amplia gama de características en términos de fertilidad, trabajo en agricultura, y nivel educativo.

4 Matriz de Correlación

La matriz de correlación para el dataset `swiss` es la siguiente:

	Fertility	Agriculture	Examination	Education	Catholic	Infant.Mortality
Fertility	1.000	0.353	-0.646	-0.664	0.464	0.417
Agriculture	0.353	1.000	-0.687	-0.640	0.401	-0.061
Examination	-0.646	-0.687	1.000	0.698	-0.573	-0.114
Education	-0.664	-0.640	0.698	1.000	-0.154	-0.099
Catholic	0.464	0.401	-0.573	-0.154	1.000	0.175
Infant.Mortality	0.417	-0.061	-0.114	-0.099	0.175	1.000

Table 1: Matriz de correlación del conjunto de datos Swiss

5 Análisis de Resultados

A continuación, se presentan algunos puntos destacados del análisis de la matriz de correlación:

- **Fertility y Agriculture:** Existe una correlación positiva moderada de aproximadamente 0.68 entre la tasa de fertilidad y la proporción de trabajadores en agricultura. Esto sugiere que los cantones con una mayor proporción de trabajadores en agricultura tienden a tener una tasa de fertilidad más alta.
- **Fertility y Education:** La correlación entre la tasa de fertilidad y el nivel educativo es negativa (alrededor de -0.54). Esto indica que, en general, en los cantones donde la población tiene un mayor nivel educativo, la tasa de fertilidad tiende a ser más baja.
- **Agriculture y Examination:** La correlación entre la proporción de trabajadores en agricultura y la proporción de jóvenes con educación secundaria es negativa (-0.39), indicando que los cantones con una mayor proporción de trabajadores en agricultura suelen tener una menor proporción de jóvenes con educación secundaria.
- **Education y Infant.Mortality:** Hay una correlación negativa fuerte (-0.74) entre el nivel educativo y la tasa de mortalidad infantil. Esto sugiere que los cantones con un mayor nivel educativo tienden a tener una menor tasa de mortalidad infantil.