

Índice

1. Exploratory Data Analysis (EDA) Report	1
1.1. Objetivo del análisis	1
1.2. Tablas de contingencia	1
1.3. ANOVAs	1
1.4. Prueba no paramétrica de Kruskal Wallis	4
1.5. Gráficos	5
1.6. Conclusiones	6

1. Exploratory Data Analysis (EDA) Report

1.1. Objetivo del análisis

Analizar el comportamiento en los precios del producto Huevos en Pakistán para obtener la concentración de los precios en las diferentes provincias, mediante el lenguaje de programación R.

1.2. Tablas de contingencia

Tabla 1: Distribución de los productos según su categoría

	Freq	% Valid	% Valid Cum.	% Total	% Total Cum.
cereals and tubers	3457	45.118768	45.11877	45.118768	45.11877
meat, fish and eggs	690	9.005482	54.12425	9.005482	54.12425
milk and dairy	265	3.458627	57.58288	3.458627	57.58288
miscellaneous food	597	7.791699	65.37458	7.791699	65.37458
non-food	958	12.503263	77.87784	12.503263	77.87784
oil and fats	782	10.206212	88.08405	10.206212	88.08405
pulses and nuts	913	11.915949	100.00000	11.915949	100.00000
Total	7662	100.000000	100.00000	100.000000	100.00000

Tabla 2: Distribución del producto Eggs - Retail según su Provincia

	Freq	% Valid	% Valid Cum.	% Total	% Total Cum.
Balochistan	1453	18.96372	18.96372	18.96372	18.96372
Khyber Pakhtunkhwa	1555	20.29496	39.25868	20.29496	39.25868
Punjab	3100	40.45941	79.71809	40.45941	79.71809
Sindh	1554	20.28191	100.00000	20.28191	100.00000
Total	7662	100.00000	100.00000	100.00000	100.00000

1.3. ANOVAs

El p-value para el ANOVA en la tabla 3 es de **8.8795435396865e-10**, por lo que **se rechaza la hipótesis nula al 10 % de significancia y si hay diferencias significativas** entre las medias.

Tabla 3: ANOVA del precio contra las provincias

term	df	sumsq	meansq	statistic	p.value
admname	3	10888.31	3629.4353	16.2122	0
Residuals	294	65817.95	223.8706	NA	NA

El p-value del ANOVA en la Tabla 4 es de **8.74132388000631e-06**, por lo que **se rechaza la hipótesis nula al 10 % de significancia y si hay diferencias significativas** entre las medias.

Tabla 4: ANOVA del precio contra los años

term	df	sumsq	meansq	statistic	p.value
year	6	8267.08	1377.8466	5.858536	8.7e-06
Residuals	291	68439.17	235.1862	NA	NA

1.3.1. Supuestos del ANOVA

El p-value de la prueba de shapiro wilk en la Tabla 5 es de **0.00015245398283457**, por lo que **se rechaza la hipótesis nula al 5 % de significancia y no se cumple el supuesto de normalidad de los errores**.

Tabla 5: Prueba de shapiro para el supuesto de normalidad de los errores

statistic	p.value	method
0.9780001	0.0001525	Shapiro-Wilk normality test

El p-value de la prueba de levene en la Tabla 6 es de **0.687343805537898**, por lo que **no se rechaza la hipótesis nula al 5 % de significancia y se cumple el supuesto de homogeneidad de las varianzas**.

Tabla 6: Prueba de levene para el supuesto de homogeneidad de las varianzas

statistic	p.value	df	df.residual
0.4930873	0.6873438	3	294

Seguidamente se presentan los gráficos correspondientes a cada uno de los supuestos.

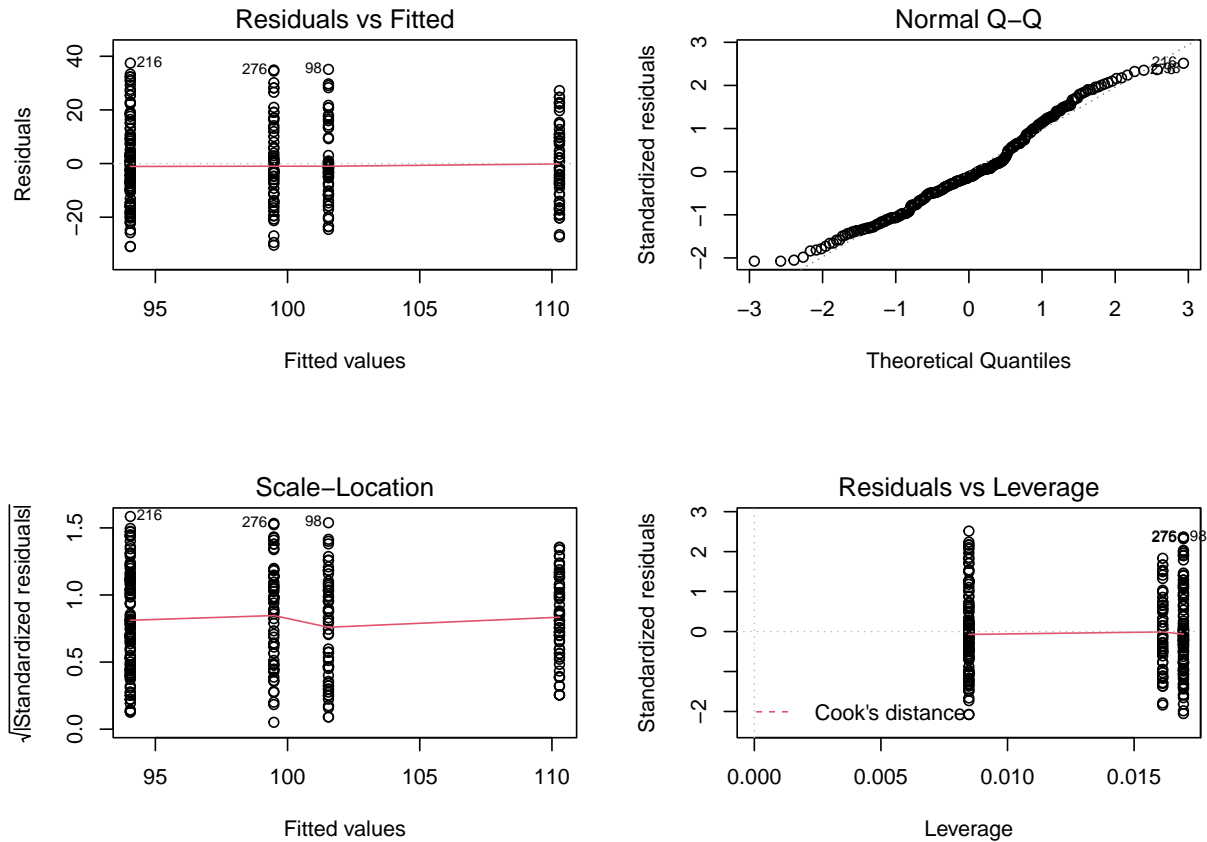


Figura 1: Normalidad de los errores, homocedasticidad de las varianzas e independencia de las observaciones para el producto Huevos según el precio y las provincias

El p-value de la prueba de shapiro wilk en la Tabla 7 es de **3.53922197054256e-05**, por lo que **se rechaza la hipótesis nula al 5% de significancia y no se cumple el supuesto de normalidad de los errores.**

Tabla 7: Prueba de shapiro para el supuesto de normalidad de los errores

statistic	p.value	method
0.9743246	3.54e-05	Shapiro-Wilk normality test

El p-value de la prueba de levene en la Tabla 8 es de **0.687343805537898**, por lo que **no se rechaza la hipótesis nula al 5% de significancia y se cumple el supuesto de homogeneidad de las varianzas.**

Tabla 8: Prueba de levene para el supuesto de homogeneidad de las varianzas

statistic	p.value	df	df.residual
0.4930873	0.6873438	3	294

A continuación se pueden observar los diferentes gráficos que corresponden a su respectivo supuesto.

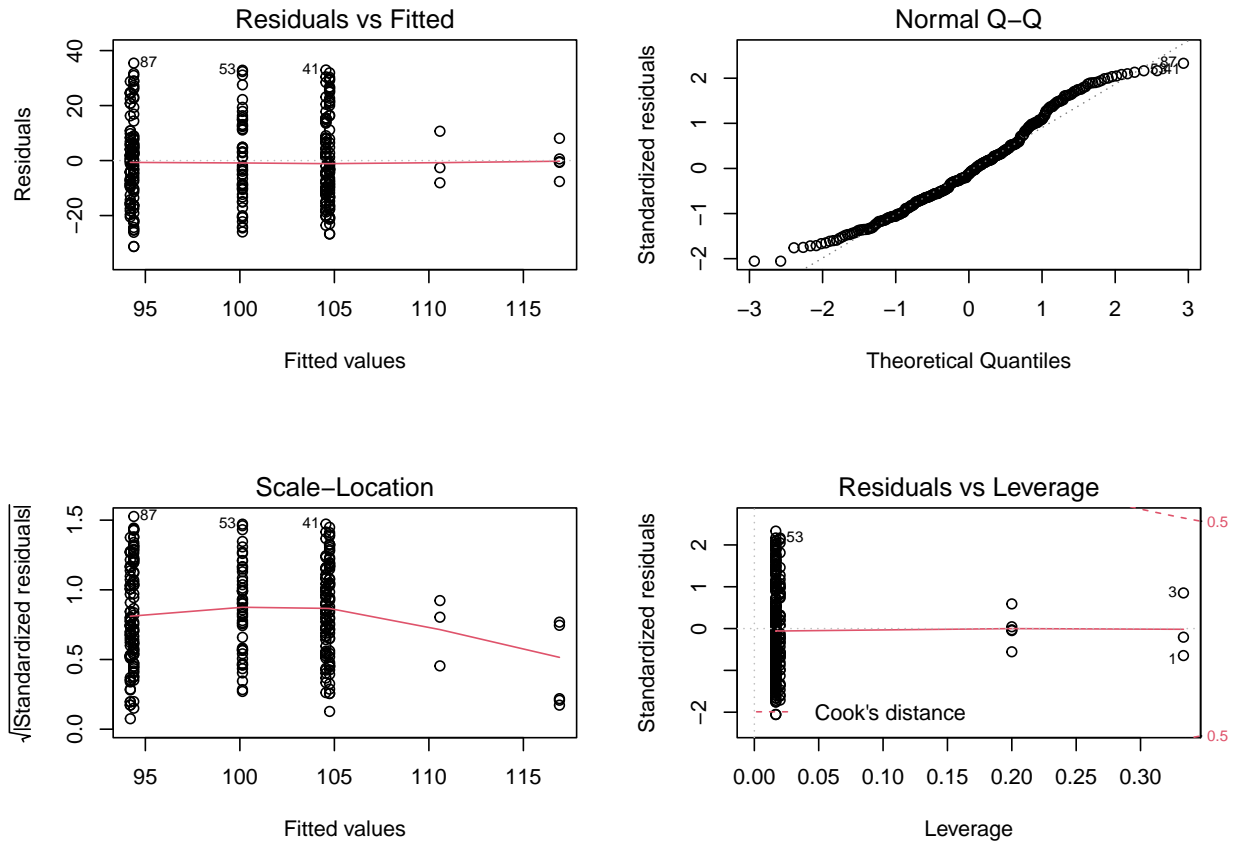


Figura 2: Normalidad de los errores, homocedasticidad de las varianzas e independencia de las observaciones para el producto Huevos según el precio y los años

1.4. Prueba no paramétrica de Kruskal Wallis

Ya que no se cumplen los supuestos de los ANOVAs planteados se procede a la realización de la prueba no paramétrica de Kruskal-Wallis, con la cuál se obtienen los siguientes resultados:

El p-value para la prueba de Kruskal-Wallis en la tabla 9 es de **2.20342710134953e-09**, por lo que **se rechaza la hipótesis nula al 10% de significancia y si hay diferencias significativas** entre las medias.

Tabla 9: Kruskal Wallis del precio contra las provincias

statistic	p.value	parameter	method
43.22613	0	3	Kruskal-Wallis rank sum test

El p-value para la prueba de Kruskal-Wallis en la tabla 10 es de **3.28801220271788e-05**, por lo que **se rechaza la hipótesis nula al 10% de significancia y si hay diferencias significativas** entre las medias.

Tabla 10: Kruskal Wallis del precio contra las provincias

statistic	p.value	parameter	method
30.40757	3.29e-05	6	Kruskal-Wallis rank sum test

1.5. Gráficos

1.5.1. Gráfico de líneas

El promedio más alto entre las provincias es de **117** PKR para el año **2014** con una desviación estándar de **6**. Por otro lado, el más bajo es de **94** PKR para el año **2015** con una desviación estándar de **13**.

En la Figura 3, A=Balochistan, B=Khyber Pakhtunkhwa, C=Punjab y D=Sindh.

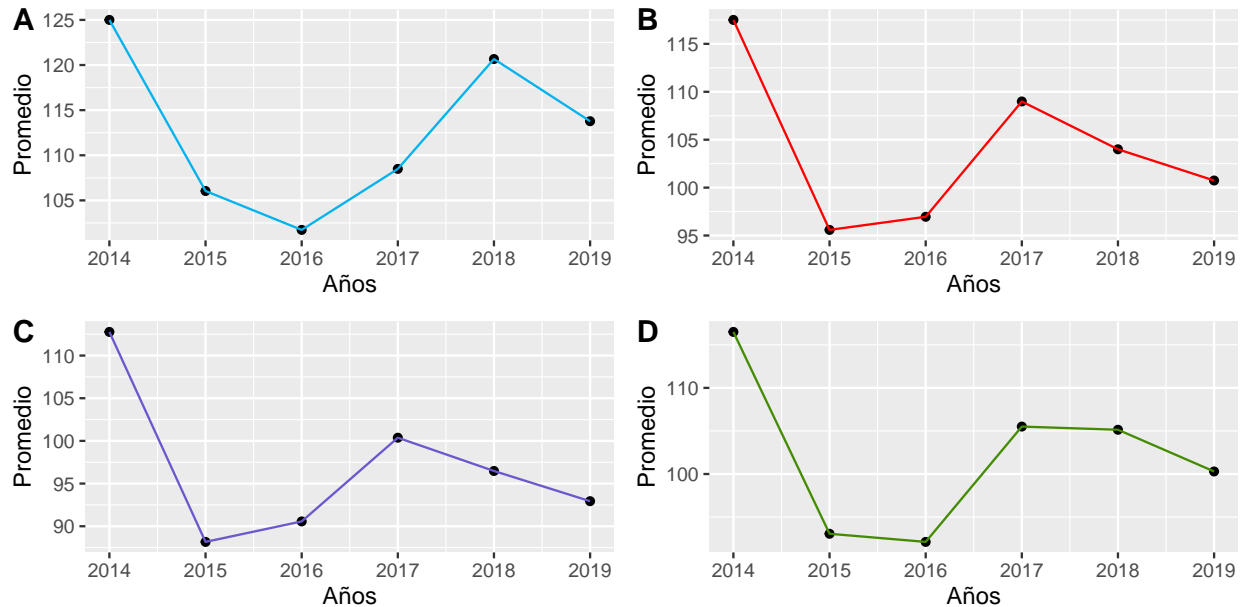


Figura 3: Promedio del precio (PKR) de Huevos según las provincias y el año

1.5.2. Gráfico de densidad

En el gráfico se observa que la provincia Balochistan posee una mayor variabilidad en el precio y tiene el promedio más alto (**110PKR**) entre las cuatro provincias, mientras que Khyber Pakhtunkhwa posee la menor variabilidad, contando con el segundo promedio más alto (**102PKR**).

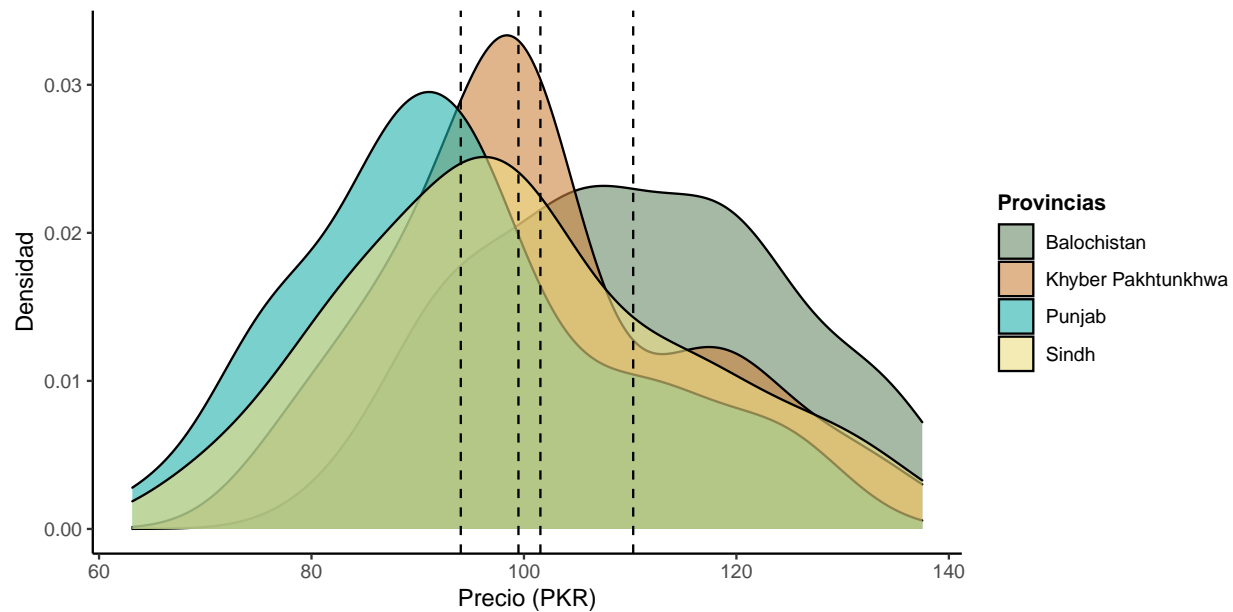


Figura 4: Densidad del precio (PKR) de Huevos en Pakistán según las provincias

1.5.3. Box plot

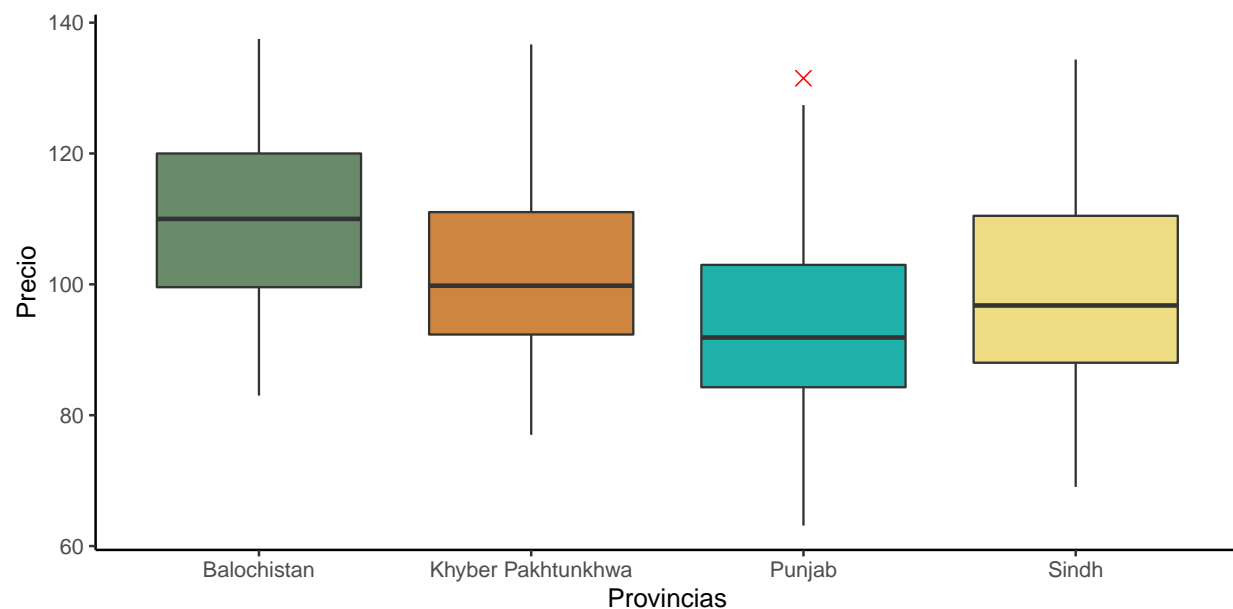


Figura 5: Resumen y comparación en la distribución del precio (PKR) de Huevos según las provincias

1.6. Conclusiones

El comportamiento de precios para el producto **Huevos** mostró gran variabilidad entre las provincias analizadas, siendo la provincia de Balochistan la que presenta mayores cambios de precios y a su vez, mantiene

el promedio del precio por encima de las demás provincias. En contraste, punjab es la provincia con precios promedios más bajos en este mercado.

A nivel nacional, los precios de **Huevos** en Pakistán sufrieron una disminución abrupta para el año 2015, debido a que el promedio de los precios se vio reducido por una sobre oferta a nivel internacional, tendencia que se vio reflejada en las diferentes provincias, sin embargo, para el año 2016 los precios apuntaban al alza, mostrandose para el periodo 2017-2018 crecimiento positivo.