

# Detección de factores no productivos con análisis de regresión múltiple

Mikio Tanabe<sup>1</sup>,  
Julio Rojo Hernández<sup>1</sup>,  
José Luis Martínez Godoy<sup>1</sup>

<sup>1</sup>Universidad Politécnica de Querétaro (UPQ), Querétaro, México  
Correspondencia: [mikio.tanabe@upq.mx](mailto:mikio.tanabe@upq.mx) , [julio.rojo@upq.mx](mailto:julio.rojo@upq.mx) , [jose.martinez@upq.mx](mailto:jose.martinez@upq.mx)

## Resumen

A lo largo de toda la fábrica ocurren distintas pérdidas de productividad y éstas, a su vez, crean costos innecesarios para muchos departamentos.

En este trabajo se presenta la aplicación del análisis de regresión múltiple para el estudio de estructuras de pérdida, las cuales generan grandes costos en desperdicios. Como resultado, se pretende probar e identificar los principales factores que dificultan la reducción de costos en los entornos de producción.

Una vez que los factores negativos están claramente identificados, las medidas de mejora continua se pueden implementar de manera más efectiva para reducirlos o eliminarlos. Al hacer esto, se espera aumentar efectivamente el impacto de la reducción de costos en el sitio de producción.

**Palabras Clave:** reducción de desperdicios, estructura de pérdidas, correlación

## Abstract

Various productivity losses occur throughout the factory, and these in turn create unnecessary costs for many departments.

This paper presents the application of multiple regression analysis for the study of loss structures, which generate large waste costs. As a result, it is intended to test and identify the main factors that hinder cost reduction in production environments in the factory.

Once negative factors are clearly identified, continuous improvement measurements could be more effectively implemented to reduce or eliminate them. By doing this, it is expected to effectively increase the impact of cost reduction at the production site.

**Keywords:** waste reduction, loss structure, correlation

---

### Artículo arbitrado

---

Recibido:

15 de octubre de  
2022

Aceptado:

11 de noviembre de 2022

## Introducción

En el sitio de producción, el problema del aumento de costos innecesario debido a la pérdida de productividad es reconocido por la gente como un problema común (Ishikawa et al., 2017).

Los análisis de las relaciones causales de correlación entre la pérdida de productividad y los aumentos de costos se han realizado durante mucho tiempo. Sin embargo, es difícil señalar si en efecto se han analizado desde un punto de vista estadístico y, como consecuencia, no resulta claro si los factores negativos se han identificado cabalmente.

Hay muchos factores negativos que conducen a un aumento de los costos en la planta de producción. Dentro de la estructura de pérdida, los factores humanos incluyen pérdida de gestión, pérdida operativa y pérdida de ajuste.

Los factores relacionados con el equipo comprenden la pérdida por defectos, la pérdida por reprocesamiento, la pérdida de tiempo de trabajo no estándar, la pérdida por micro-paro de máquinas, la pérdida por arranque, la pérdida por falla y la pérdida por configuración.

Desafortunadamente, incluso hoy en día, con tecnología avanzada, es difícil reducir simultáneamente y de manera eficiente varios de los problemas de aumento en los costos de producción.

Por lo tanto, se espera que el método estadístico identifique los principales factores que dificultan la reducción de costos en el entorno de producción de una manera más precisa y específica.

Para aclarar el orden de prioridad y tratar de identificar los verdaderos factores negativos, se utiliza una técnica llamada "visualización", una herramienta esencial para una mejora más eficiente de la productividad en el sitio de producción.

En este documento, al examinar la correlación entre una cierta pérdida de productividad y el aumento de

costos mediante un análisis de regresión múltiple, la variable objetivo es el aumento de costos y la variable explicativa son las diversas pérdidas que ocurren dentro del sitio de producción.

Aumentar el número de variables explicativas en el análisis de regresión múltiple crea inevitablemente correlaciones complejas entre varias pérdidas de productividad, lo cual dificulta el procesamiento de un proceso de análisis adecuado.

Por lo tanto, en este trabajo, el número de variables explicativas es dos. Examinaremos en qué medida el cambio en el número de muestras en el análisis de regresión múltiple afecta la confiabilidad y estabilidad de los resultados del análisis final (Iwasaki, 2021).

## Método

El análisis de regresión simple sólo utiliza una variable explicativa. Sin embargo, para el análisis de regresión múltiple (usado en este trabajo) se requieren de múltiples variables explicativas.

Como no hay límite para el número de variables explicativas si se aumenta su número indiscriminadamente, existe la posibilidad de que haya una alta correlación entre las variables explicativas, lo cual provocará una probabilidad de que no se puedan captar las influencias con precisión entre la variable objetivo y cada variable explicativa.

Por lo tanto, en este artículo, considerando la confiabilidad y la estabilidad, usamos dos variables explicativas y examinamos sus efectos sobre la variable objetivo. Un análisis general de regresión múltiple se modela de la siguiente manera (Sato, 2017):

$$y = a_1x_1 + a_2x_2 + a_3x_3 + \dots + a_nx_n + b \quad (1)$$

Aquí,  $y$  es la variable objetivo del modelo, mientras  $x$  es la variable explicativa de los datos medidos. Los datos medidos se distribuyen en el plano  $xy$ .

Al usar dos variables explicativas, el modelo se puede expresar de la siguiente manera:

$$y = a_1x_1 + a_2x_2 + b \tag{2}$$

En el análisis de regresión múltiple cuando hay dos variables explicativas, los coeficientes  $a_1$  y  $a_2$  de las dos variables explicativas se pueden obtener, a partir de ecuaciones simultáneas, usando las siguientes tres sumas de cuadrados y las tres sumas de los productos (3), (4), (5), (6), (7) y (8) como a continuación se observa:

$$S_{x1} = \sum_{i=1}^n (x_{1i} - x_{1i \text{ promedio}})^2 \tag{3}$$

$$S_{x2} = \sum_{i=1}^n (x_{2i} - x_{2i \text{ promedio}})^2 \tag{4}$$

$$S_y = \sum_{i=1}^n (y_i - y_i \text{ promedio})^2 \tag{5}$$

$$S_{x1y} = \sum_{i=1}^n (x_{1i} - x_{1i \text{ promedio}})(y_i - y_i \text{ promedio}) \tag{6}$$

$$S_{x2y} = \sum_{i=1}^n (x_{2i} - x_{2i \text{ promedio}})(y_i - y_i \text{ promedio}) \tag{7}$$

$$S_{x1x2} = \sum_{i=1}^n (x_{1i} - x_{1i \text{ promedio}})(x_{2i} - x_{2i \text{ promedio}}) \tag{8}$$

Las ecuaciones simultáneas se muestran en (9), (10), (11) y (12) de la siguiente manera:

$$S_{x1} \times a_1 + S_{x1x2} \times a_2 = S_{x1y} \tag{9}$$

$$S_{x1x2} \times a_1 + S_{x2} \times a_2 = S_{x2y} \tag{10}$$

$$a_1 = \frac{S_{x1y}}{S_{x1}} - \frac{S_{x1x2}}{S_{x1}} \times a_2 \tag{11}$$

$$a_1 = \frac{S_{x2y}}{S_{x1x2}} - \frac{S_{x2}}{S_{x1x2}} \times a_2 \tag{12}$$

Ahora, si consideramos como hipótesis el proceso de una operación bajo ambiente productivo normal, entonces tomamos dos factores para dos variables explicativas.

De esta manera, el factor  $x_1$  es la pérdida por parada de una máquina, mientras el factor  $x_2$  es la pérdida por movimiento innecesario de un operador que maneja la máquina.

De ambas pérdidas se mide el tiempo en unidad de minutos. Las pérdidas en minutos se registran por día. Finalmente, la variable objetivo  $y$  es la pérdida del costo de producción por día.

Para este trabajo, el tamaño de la muestra fue de 50. En general, hay más diferencias y errores dentro de los datos medidos cuando el tamaño de la muestra es pequeño. Si el tamaño de muestra es grande, habrá menos afectación no favorable debido a las características generales de la estadística.

## Resultados

A continuación, se presentan los siguientes datos para un análisis de regresión múltiple para, en primer lugar, la detección de factores no productivos y, en segundo lugar, como prueba de la teoría antes de su implementación en piso de producción. En la Tabla 1 se muestran ejemplos de datos de pérdidas en una operación:

Tabla 1. Los datos de pérdidas en una operación

DIA	PERDIDAS		
	COSTO POR DIA [\$/X1000	PARO MAQUINA POR DIA [MIN]	MOVIMIENTO INNECESARIO OPERADOR POR DIA [MIN]
1	40	35	5
2	30	20	6
3	20	15	6
4	90	75	5
5	90	60	2
6	40	30	5
7	100	90	8
8	30	30	7
9	40	30	2
10	30	40	3
11	40	35	5
12	30	20	6
13	20	15	6
14	90	75	5
15	90	60	2
16	40	30	5
17	100	90	8
18	30	30	7
19	40	30	2
20	30	40	3
21	40	35	5
22	30	20	6
23	20	15	6
24	90	75	5
25	90	60	2
26	40	30	5
27	100	90	8
28	30	30	7
29	40	30	2
30	30	40	3
31	20	15	5
32	20	40	6
33	30	40	6
34	40	50	5
35	40	20	2
36	40	20	5
37	120	60	8
38	40	10	7
39	30	50	2
40	20	40	3
41	40	35	5
42	30	20	6
43	20	15	6
44	90	75	5
45	90	60	2
46	40	30	5
47	100	90	8
48	30	30	7
49	40	30	2
50	30	40	3

**Tabla 1.** El tamaño de muestra es 50. Son datos propios generados al azar. Fuente: elaboración propia.

Si se desea verificar cuál de las múltiples variables explicativas tiene el mayor impacto en la variable objetivo, se debe tener cuidado el manejo de los coeficientes de regresión, esto tanto en el análisis de regresión simple como en el análisis de regresión múltiple.

Dado que el análisis de regresión múltiple utiliza múltiples variables explicativas, los coeficientes de regresión deben estandarizarse debido a las diferencias en las medias y las varianzas entre las variables explicativas.

Para calcular los resultados de análisis de regresión múltiple, se utiliza una función de análisis de datos dentro de Excel.

En la Tabla 2 se muestra el resultado del análisis de regresión múltiple para la cual se usaron los datos de la Tabla 1:

**Tabla 2.** Resultados de análisis de regresión múltiple con los datos de la Tabla 1

Resumen								
Estadísticas de la regresión								
Coefficiente de correlación múltiple	0.87073843							
Coefficiente de determinación R <sup>2</sup>	0.75855436							
R <sup>2</sup> ajustado	0.74832291							
Error típico	14.5343493							
Observaciones	50							
ANÁLISIS DE VARIANZA								
	Grados de libertad	Suma de cuadrados	Promedio de los cuadrados	F	Valor crítico de F			
Regresión	2	31199.51308	15599.75654	73.84693504	3.12262E-15			
Residuos	47	9928.486924	211.244926					
Total	49	41128						
	Coefficientes	Error típico	Estadístico t	Probabilidad	Inferior 95%	Superior 95%	Inferior 95.0%	Superior 95.0%
Intercepción	1.95901558	6.46473072	0.30457356	0.762033723	-11.09635913	14.97496227	-11.09635913	14.97496227
Variable X1	1.119556953	0.092982299	12.04161405	5.73843E-16	0.992600704	1.306713222	0.992600704	1.306713222
Variable X2	0.211628498	1.077693842	0.196382363	0.845157538	-1.956265772	2.379662768	-1.956265772	2.379662768

**Tabla 2.** Tamaño de muestra es 50. Fuente: elaboración propia.

Si observamos el resultado del coeficiente de correlación múltiple  $r$  (0.870), encontramos que salió muy alto.

La razón principal de este resultado, es que las dos variables explicativas son las pérdidas de tiempo en una operación de máquina. Es obvio: las pérdidas de tiempo están relacionadas directamente a los costos innecesarios.

Existe otro indicador, el cual se llama VIF: Variance Inflation Factor (Factor de Inflación de Varianza). La ecuación de VIF se muestra en (13) a continuación:

$$VIF = 1 / (1 - r^2) \quad (13)$$

Se utiliza VIF para determinar si existe un problema con la multi-colinealidad entre las variables explicativas. Si el valor VIF rebasa 10.0, tenemos evidencia de que existe un problema de multi-colinealidad, el cual impedirá el análisis de predicción.

Por otro lado, con pocas variaciones de los datos de una variable explicativa, las variables objetivos variaran mucho de manera inesperada, lo cual se sumará al problema de multi-colinealidad.

En este análisis, VIF dio como resultado lo siguiente:  $VIF = 1 / (1 - r^2) = 1 / (1 - 0.7585^2) = 4.1424$ . Lo cual está dentro del rango y, por lo tanto, podemos determinar que seleccionamos correctamente las dos pérdidas de tiempo como variables explicativas.

Además, el valor de significación  $F$  (grados de libertad) en la tabla de análisis de varianza, es extremadamente pequeño. Los cálculos de regresión múltiple fueron adecuados sin afectación de varianza. Por lo tanto, se puede determinar que las dos variables explicativas influyen en la variable objetivo.

Ahora, si verificamos coeficientes de regresión, el valor de  $a_1$ : coeficiente de variable  $x_1$  (pérdida por paro de máquina) fue 1.1196.

Y como el valor de la probabilidad de su error fue extremadamente bajo ( $5.73343E-16$ ), por lo tanto, fue evidente que el uso de la variable  $x_1$  (pérdida por paro de máquina) fue muy confiable para este análisis de regresión múltiple.

Para el valor de  $a_2$ : coeficiente de variable  $x_2$  (pérdida por movimiento innecesario del operador), el resultado fue 0.2116. Además, el valor de la probabilidad de su error fue 0.8451.

Como el valor de probabilidad de error debe ser menor que 0.05 para la confiabilidad, es posible que la causa de esto sea el valor alto de  $r$  (coeficiente correlación múltiple)=0.870.

El coeficiente de intercepción  $b$  obtuvo el valor de 1.969 con la probabilidad de error de 0.7620, resultado poco favorable por ser mayor a 0.05.

En este trabajo se utilizaron  $a_1$  (pérdida por paro de máquina) y  $a_2$  (pérdida por movimiento innecesario del operador) como variables explicativas.

Como variable objetivo, se utilizó  $y$  (la pérdida de costes). Con base en lo anterior, la fórmula de predicción para la variable objetivo se puede expresar de la siguiente manera en (14):

$$y (\text{pérdida de costeo}) = 1.1196 a_1 + 0.2116 a_2 + 1.969 \quad (14)$$

La fórmula de predicción indica que la variable explicativa  $a_1$  tiene la mayor influencia en el resultado. La influencia de  $a_1$  es 5.29 veces más fuerte que la de  $a_2$ .

Como método comparativo, ahora aumentamos el tamaño de muestra hasta 120 en lugar de 50, lo cual se puede observar a continuación en la Tabla 3.

Tabla 3. Los datos de pérdidas en una operación

DIA	PERDIDAS			DIA	COSTO POR DIA [\$]X1000	PARO MAQUINA POR DIA[MIN]	MOVIMIENTO INNECESARIO OPERADOR POR DIA[MIN]
1	40	35	5	61	40	35	5
2	30	20	6	62	30	20	6
3	20	15	6	63	20	15	6
4	90	75	5	64	90	75	5
5	90	60	2	65	90	60	2
6	40	30	5	66	40	30	5
7	100	90	8	67	100	90	8
8	30	30	7	68	30	30	7
9	40	30	2	69	40	30	2
10	30	40	3	70	30	40	3
11	40	35	5	71	20	15	5
12	30	20	6	72	20	40	6
13	20	15	6	73	30	40	6
14	90	75	5	74	40	50	5
15	90	60	2	75	40	20	2
16	40	30	5	76	40	20	5
17	100	90	8	77	120	60	8
18	30	30	7	78	40	10	7
19	40	30	2	79	30	50	2
20	30	40	3	80	20	40	3
21	40	35	5	81	40	35	5
22	30	20	6	82	30	20	6
23	20	15	6	83	20	15	6
24	90	75	5	84	90	75	5
25	90	60	2	85	90	60	2
26	40	30	5	86	40	30	5
27	100	90	8	87	100	90	8
28	30	30	7	88	30	30	7
29	40	30	2	89	40	30	2
30	30	40	3	90	30	40	3
31	20	15	5	91	40	35	5
32	20	40	6	92	30	20	6
33	30	40	6	93	20	15	6
34	40	50	5	94	90	75	5
35	40	20	2	95	90	60	2
36	40	20	5	96	40	30	5
37	120	60	8	97	100	90	8
38	40	10	7	98	30	30	7
39	30	50	2	99	40	30	2
40	20	40	3	100	30	40	3
41	40	35	5	101	20	15	5
42	30	20	6	102	20	40	6
43	20	15	6	103	30	40	6
44	90	75	5	104	40	50	5
45	90	60	2	105	40	20	2
46	40	30	5	106	40	20	5
47	100	90	8	107	120	60	8
48	30	30	7	108	40	10	7
49	40	30	2	109	30	50	2
50	30	40	3	110	20	40	3
51	20	15	5	111	40	35	5
52	20	40	6	112	30	20	6
53	30	40	6	113	20	15	6
54	40	50	5	114	90	75	5
55	40	20	2	115	90	60	2
56	40	20	5	116	40	30	5
57	120	60	8	117	100	90	8
58	40	10	7	118	30	30	7
59	30	50	2	119	40	30	2
60	20	40	3	120	30	40	3

Tabla 3. El tamaño de muestra es 120. Son datos propios generados al azar. Fuente: elaboración propia.

Si utilizamos los datos similares, como la Tabla 3 y aplicamos el método de análisis de regresión múltiple, tenemos como resultados en la Tabla 4:

Tabla 4. Resultados de análisis de regresión múltiple con tamaño de muestra 120 de la Tabla 3

Resumen							
Estadísticas de la regresión							
Coefficiente de correlación múltiple	0.821298433						
Coefficiente de determinación R <sup>2</sup>	0.67451116						
R <sup>2</sup> ajustado	0.668967545						
Error típico	16.60737016						
Observaciones	120						
ANÁLISIS DE VARIANZA							
	Grados de libertad	Suma de cuadrados	Promedio de los cuadrados	F	Valor crítico de F		
Regresión	2	66877.51167	33438.75583	121.2406843	-3.032516-29		
Residuos	117	32269.155	275.8047436				
Total	119	99146.66667					
	Coefficientes	Error típico	Estadístico t	Probabilidad	Inferior 95%	Superior 95%	Inferior 95.0%
Intercepción	-2.291816417	4.79122535	-0.478336177	0.633803555	-11.78053716	7.196554322	-11.78053716
Variable X1	1.077215365	0.07094736	15.18329079	1.571766E-29	0.936707697	1.217722833	0.936707697
Variable X2	-1.370627351	0.794266658	1.725652939	0.087049926	-0.20237475	2.94529864	-0.20237475

Tabla 4. Son datos propios generados al azar. Fuente: elaboración propia.

Al comparar los dos resultados anteriores se observa que fueron muy parecidos. Sólo el valor de  $a_2$ : coeficiente de variable  $x_2$  (pérdida por movimiento innecesario del operador) fue de 1.370, en lugar del anterior resultado de 0.2116. El valor de la probabilidad de error fue 0.08704, el cual mejoró mucho en comparación al anterior resultado de 0.8451.

Finalmente, la fórmula de predicción para la variable objetivo se puede expresar de la siguiente manera en (15):

$$y (\text{pérdida de costeo}) = 1.07721 a_1 + 1.3706 a_2 - 2.2918 \quad (15)$$

En comparación a la fórmula de predicción anterior (14), hay varios cambios con respecto a la fórmula (15). Idealmente, no debe haber variaciones grandes en la fórmula de predicción de aumento de costos.

De todo lo anterior, podemos dictaminar que el análisis de regresión múltiple es muy adecuado para discernir la relación entre múltiples variables explicativas y la variable objetivo.

Sin embargo, todavía es necesario seguir mejorando el método en la aplicación del análisis de regresión múltiple para obtener la fórmula de predicción sin una variación mayor, independientemente del diferente tamaño de las muestras y otros factores de posibles errores en el análisis.

## Discusión y Conclusiones

A través de las pruebas comparativas de cálculo de regresión múltiple realizadas, a lo largo de este trabajo se pudo observar que su aplicación puede ser una estrategia significativa para determinar el verdadero factor de prioridad que afecta el aumento de costos innecesarios dentro de las fábricas.

Principalmente, este método ayuda a clasificar los factores principales que aumentan costos innecesarios.

Este trabajo partió de una hipótesis para comprobar si la técnica de análisis de regresión múltiple se puede utilizar en piso de producción.

La mayoría de los resultados obtenidos fueron favorables para este estudio. Ahora, para mejorar la confiabilidad en la obtención de la fórmula de predicción de aumento de costos, es muy importante seleccionar las variables explicativas de manera más precisa.

Si se seleccionan variables explicativas que no tienen verdaderas relaciones e influencias a la variable objetivo, se puede bajar el nivel de precisión de la fórmula de predicción sobre una variable objetivo.

En el futuro, si se agregan y reducen más variables explicativas distintas, se pueden realizar más pruebas de análisis de regresión múltiples. De esta manera, se podrá optimizar la fórmula de predicción de aumento de costos a través de las pérdidas claves que existen dentro de las fábricas.

El análisis numérico profundo para la detección de factores no productivos, es un método muy válido para contribuir al aumento de productividad en las PyMES mexicanas en distintos giros del sector industrial. Es necesario seguir mejorando las técnicas para la aplicación de este método en piso de producción.

## Referencias

- Ishikawa, H., Ishikawa, R. (2017). *Conceptos básicos y práctica del sistema de producción de Toyota*. Japón: Shuwa System.
- Iwasaki, M. (2021). Multiple Regression Viewed from Causal Inference Perspective. *Revista de la Sociedad Estadística de Japón*. 50, pp. 363-379.
- Sato, H. (2017). *Fundamentos de Estadística*. Japón: Gijutsu-Hyoron publishing.