Linearity and bias script in R.

Ing. Mario Iván Salas Domínguez.
[a] Universidad Tecnológica de Aguascalientes, msalas@utags.edu.mx, Aguascalientes, Aguascalientes, México.

Resumen

Este trabajo tiene como objetivo desarrollar un código en R que lleve a cabo el estudio de linealidad y sesgo de un sistema de medición de variables bajo los lineamientos del manual MSA de AIAG 4° edición y en una presentación similar a la de Minitab®.
Como primera parte se establece la elección de R como programa estadístico y los antecedentes de estudios de linealidad usando R u otro programa estadístico.
A continuación, se describen los conceptos estadísticos básicos en los que se sustenta un estudio de linealidad en el sistema de medición según MSA 4° edición y la relación que guardan con la lógica del código.
Por último, se describe un bosquejo de la lógica utilizada en el código, los resultados del procesamiento de los datos y los siguientes pasos a seguir en el desarrollo del proyecto.
Palabras clave— Linealidad, MSA, Minitab, R

Abstract

This work aims to develop an R scrip that carries out the study of linearity and bias of a variable measurement system under the guidelines of the AIAG 4th edition MSA manual and in a presentation similar as Minitab®.
As a first part, the choice of R as a statistical program and the history of linearity studies using R or another statistical program are established.
The following describes the basic statistical concepts underlying a linearity study in the measurement system according to MSA 4th edition and their relationship to code logic.
Finally, it describes an outline of the logic used in the code, the results of data processing and the next steps to follow in the development of the project.
Keywords— *Linearity, MSA, Minitab,R*

1. INTRODUCCIÓN

La linealidad es un parámetro dentro del estudio de exactitud de un sistema de medición o también llamado MSA por su acrónimo en inglés (measurment system analysis), siendo MSA una de las herramientas eje dentro del sector automotriz o también conocidas como core tools, esta herramienta es de vital importancia en el desarrollo de procesos de mejora de calidad como es señalado por [1]
Para [2] la evaluación de la calidad de un sistema de medición se da en la validación estadística de los datos que este arroja, de aquí la importancia de contar con los programas estadísticos robustos, confiables y accesibles que permitan realizar estos estudios.
Dentro de las alternativas se encuentra R programa para evaluación estadística y de graficación que puede ser usado de forma libre y debido a su amplia difusión a nivel mundial cuenta con una gran comunidad que lo alimenta de nuevas capacidades y proporciona soporte para su uso [3].
Para lograr el objetivo deseado, se conjuntaron los principios estadísticos señalados en MSA 4°edición y las funciones y paquetes básicos con los que cuenta R, para así lograr gráficos y estadísticos lo más afines a Minitab®, programa estadístico comercial con el cual la industria está ampliamente familiarizado.

2. CONTENIDO

2.1. Antecedentes

Con el crecimiento en el volumen de datos que genera la sociedad actual, combinado con el desarrollo de dispositivos de computo accesibles en sus dimensiones y costos, trajo como consecuencia que los programas estadísticos también crecieran en número y capacidades de procesamiento.
Ante esta gama de programas, siempre es importante seleccionar la opción que mejor se ajustes a las necesidades del uso que se va a dar.
Para [4] un programa estadístico se debe de evaluar en los siguientes aspectos amigable con el usuario, manipulación de datos, calidad de gráficos, control del proceso,
costo, código fuente disponible, variedad de análisis estadísticos, documentación, soporte técnico y
uso en diversos sistemas operativos, de estos diez aspectos, solo el que corresponde con amigable con el usuario, es evaluado como bajo-regular, destacando que R es gratuito, al contrario de otros programas comerciales como lo es Minitab®, SAS® o SPSS® que requieren de licencias de un alto costo sobre todo cuando se hace un uso profesional o comercial.
Estas características no han pasado desapercibidas en usuarios que desean realizar evaluaciones del sistema de medición utilizando a R como herramienta y han desarrollado paquetes usando R, como lo es el caso de qualityTools, el cual es un aglomerado de herramientas para el control de calidad y la metodología seis sigma[5] que en su última versión 1.55 incluye una función para llevar a cabo el estudio de linealidad[6], sin embargo, lamentablemente al inicio del año 2023 ya no se encuentra disponible dentro del repositorio de paqueterías de R[7]

2.2. Fundamento estadístico de la linealidad y el sesgo

Una forma de evaluar la exactitud de un sistema de medición es mediante el estudio de linealidad y sesgo.
Para [8, p. 213] citado por [2], el sesgo es “La diferencia entre el promedio de las mediciones observadas (intentos sobre condiciones de repetibilidad) y el valor referencia; históricamente referido como exactitud el sesgo es evaluado y expresado como un solo punto dentro del rango de operación del sistema de medición” y la linealidad como “La diferencia en errores de sesgos sobre el rango de operación esperado del sistema de medición.”
Con base en esta definición, podemos ver que todo estudio de linealidad trae implícito una valoración del sesgo para cada valor del instrumento dentro de su rango de operación.

Para [9] citado por [2] la evaluación de la linealidad se sustenta en probar que:

$$H_0\ a = 0[1]$$
$$H_1\ a \neq 0[2]$$

Para la ecuación de regresión lineal

$$\bar{y}_i = ax_i + b[3]$$

Donde :

a es la pendiente de la recta de regresión.

b es la ordenada en el origen

$\bar{y}_i$ son los i valores promedio de los sesgos para cada una de las referencias

x_i. Son los i valores de referencias

Para [8], [10] la magnitud de la linealidad se calcula mediante

$$Linealidad = |a|(Variación\ del\ proceso\ o\ Tolerancia)[4]$$

El porcentaje de linealidad se calcula como

$$\%Linealidad = 100 * Linealidad/ Variación\ del\ proceso\ o\ Tolerancia) \text{ [5]}$$

Para [9, p. 96] El valor del sesgo para cada una de las referencias usadas en el estudio está dado por la ecuación

$$\bar{y}_i = \sum_{i=1}^{g}\sum_{j=1}^{m}(z_{i,j} - x_i)/\sum_{i=1}^{g} m_i[6]$$

Donde:

$\bar{y}_i$ Son los promedios de sesgo de la i esima referencia.

x_i i esima referencia.

g Es el número de referencias usadas en el estudio

m Número de réplicas para cada una de las g referencias.

$z_{i,j}$ j ésima medición de la i ésima referencia

Con base en el concepto de sesgo, se desea que las lecturas del instrumento de medición sean iguales a las de la referencia de tal forma que la hipótesis que se plantea es

$$H_0\ Sesgo = 0[2]$$
$$H_1\ Sesgo \neq 0[3]$$

La hipótesis nula es probada en su significancia estadística usando una prueba t de la media del sesgo y su correspondiente pvalor, para medir la magnitud del sesgo, [10] y [11] proponen un índice con respecto a la variación del proceso o tolerancia con la siguiente fórmula

$$\%sesgo = 100 \times \frac{|Promedio\ de\ sesgo|}{Variación\ del\ proceso\ o\ Tolerancia}[4]$$

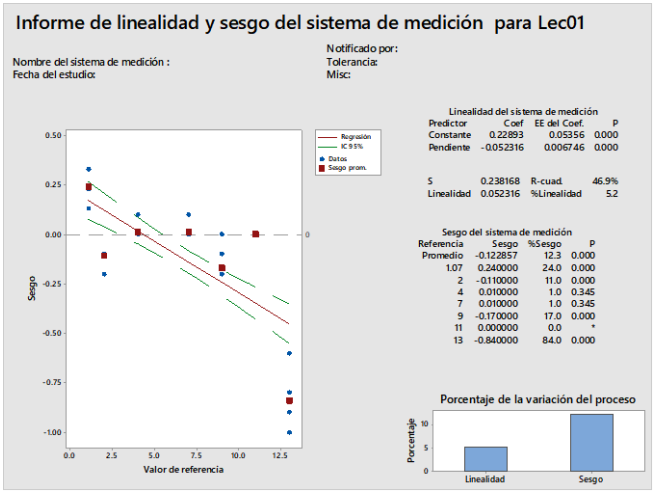
2.3. Lógica del código

La lógica del código está dirigida a obtener los reportes y gráficos que genera Minitab® para el estudio de linealidad y sesgo.

Fuente: Elaboración propia muestra los gráficos y reportes que se busca obtener por el código, estos gráficos y reportes son los siguientes.

- Reporte de la regresión lineal de los sesgos por referencia.
- Reporte de sesgos promedios y p valores por referencia.
- Gráfico de regresión lineal con intervalos de confianza de los sesgos por referencia.
- Gráfico de barras del porcentaje de linealidad y sesgo con respecto a las tolerancias o variación del proceso.

Fig. 1 Reporte linealidad y sesgo Minitab®

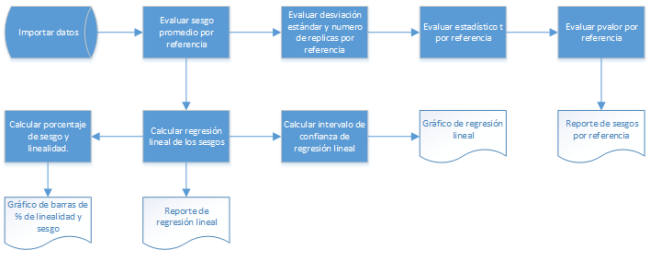


Fuente: Elaboración propia usando Minitab®

El código se desarrolló evitando en la medida de lo posible el uso de paqueterías que no se agregan por default al instalar R como podría ser ggplot, y con esto evitar algún conflicto por la actualización de R como sucedió en el caso de qualityTools, mencionado en los antecedentes.

La **¡Error! No se encuentra el origen de la referencia.**, muestra los procesos y reportes que se ejecutan. Para la importación de los datos se usó la librería readxl. Lo correspondiente a la evaluación de los sesgos promedio, deviación estándar y numero de replicas por referencia se usó la función tapply. Para el proceso de la regresión lineal se usó la función lm y el intervalo de confianza con la función predict. Para los gráficos se usó la función barplot para el porcentaje de linealidad y sesgo, plot para los puntos correspondientes a los valores de sesgo individuales y promedio y abline para la línea de regresión, para las líneas de los intervalos de confianza de la regresión se usó la función lines.[12].

Fig. 2 Lógica del código



Fuente: elaboración propia

2.4. Resultados

Los resultados obtenidos son coincidentes con un procesamiento en Minitab de los mismos datos usados en R, ver Fuente: Elaboración propia

El reporte del sesgo por consola representado por la variable bias_ref_report, es coincidente con el reporte de Minitab® como se muestra en Fig. 3 sesgos por referencias.

Fig. 3 sesgos por referencias

```
Console  Terminal
R 4.0.4 · C:/Users/msalas/OneDrive/Documentos/Projects R/linealidad/
>   bias_ref_report<-cbind(ref_values,avge_bias,p_values)#matrix r
eferences bias report
> bias_ref_report
     ref_values avge_bias     p_values
1.07       1.07      0.24 3.050481e-07
2          2.00     -0.11 1.609932e-06
4          4.00      0.01 3.434364e-01
7          7.00      0.01 3.434364e-01
9          9.00     -0.17 2.292673e-05
11        11.00      0.00          NaN
13        13.00     -0.84 1.738608e-08
>
```

Fuente: elaboración propia usando R

Lo que corresponde al reporte de regresión, en la Fig. 4 Reporte de regresión en amarillo, se destacan los valores que muestra el reporte de Minitab®, en este caso aún falta calcular el valor S que corresponde con la desviación estándar sobre la regresión y extraer y ajustar los valores de linealidad a como los presenta Minitab®.

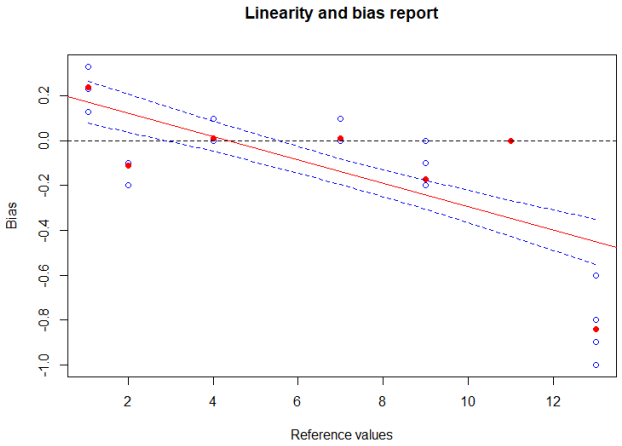
Fig. 4 Reporte de regresión

```
Console  Terminal
R 4.0.4 · C:/Users/msalas/OneDrive/Documentos/Projects R/linealidad/
> summary(regresion)

Call:
lm(formula = Datos$sesgo_i ~ Datos$ref_values)

Residuals:
     Min       1Q   Median       3Q      Max
-0.54882 -0.20543  0.04191  0.13728  0.34655

Coefficients:
                Estimate Std. Error t value Pr(>|t|)
(Intercept)     0.228932   0.053556   4.275 6.11e-05 ***
Datos$ref_values -0.052316   0.006746  -7.755 6.11e-11 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 0.2382 on 68 degrees of freedom
Multiple R-squared:  0.4693,    Adjusted R-squared:  0.4615
F-statistic: 60.14 on 1 and 68 DF,  p-value: 6.11e-11
```

Fuente: elaboración propia usando R

Lo que corresponde con el gráfico de regresión que muestra **¡Error! No se encuentra el origen de la referencia.**, se cuenta con todos los elementos solo falta agregar las leyendas.

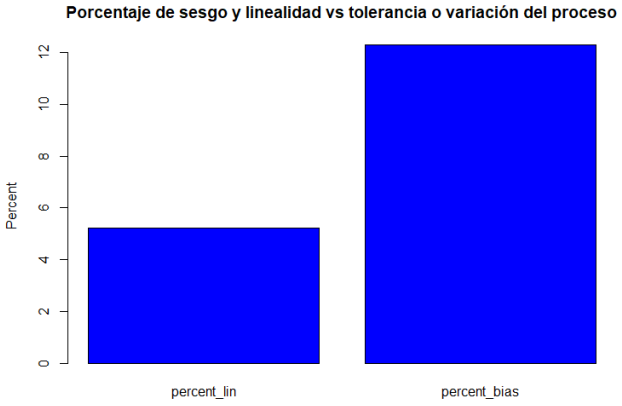
Fig. 5 Gráfico de regresión



Fuente: elaboración propia usando R

Para el gráfico del porcentaje de linealidad y sesgo se cuenta con los mismos elementos que reporta Minitab®

Fig. 6 Porcentaje de linealidad y sesgo.



Fuente: elaboración propia usando R

3. CONCLUSIONES Y RECOMENDACIONES

El código muestra los principales elementos de un estudio de linealidad y sesgo y este es coincidente con los valores que reporta Minitab, sin embargo, aún falta extraer y procesar valores dentro del reporte de regresión y presentarlos de forma amigable, para esto último se recomienda el uso de un reporte con Markdown

Es comprensible considerar, como se mencionó en los antecedentes, que uno se sus puntos débiles es la parte amigable con el usuario, para resolver este inconveniente, se recomienda extender este código aquí presentado al entorno de Shiny y crear una interface que le sea amigable al usuario tanto para la carga de los datos como a los reportes generados y yendo más allá, dar una opinión sobre el estado que guarda el instrumento.

El presente trabajo muestra la excelente opción que representa sobre los programas comerciales, el uso de programas de análisis estadístico de código abierto como lo

es R o también lo podría ser Phyton, los cuales muestran su potencial para la evaluación, en este caso de la linealidad y sesgo en sistemas de medición de variables, sin embargo estas capacidades de pueden extender a otras áreas como lo es el control estadístico del proceso o el diseño de experimentos u otras herramientas de amplio uso dentro del sector industrial.

4. REFERENCIAS

[1] J. A. Doshi and D. A. Desai, "Measurement system analysis for continuous quality improvement in automobile SMEs: multiple case study," *Total Qual. Manag. Bus. Excell.*, vol. 30, no. 5–6, pp. 626–640, Apr. 2019, doi: 10.1080/14783363.2017.1324289.

[2] S. D. M. Iván, "Accuracy study of a pH meter with Minitab.," *Revista Arista ciencia Básica y Aplicada*, vol. 8, no. 15, Tijuana MX, pp. 211–213, Jul-2020.

[3] "R: What is R?" [En línea]. Disponible en: https://www.r-project.org/about.html. [Accedido: 10-Feb-2023].

[4] C. Salas, "¿Por qué comprar un programa estadístico si existe R?" [En línea]. Disponible en: http://www.scielo.org.ar/scielo.php?script=sci_arttext&pid=S1667-782X2008000200007. [Accedido: 09-Feb-2023].

[5] T. Roth, "Working with the qualityTools package," 2017. [En línea]. Disponible en: http://r-qualitytools.org/index.html. [Accedido: 12-Feb-2023].

[6] R. Documentation, "qualityTools package." [En línea]. Disponible en: https://www.rdocumentation.org/packages/qualityTools/versions/1.55. [Accedido: 11-Feb-2023].

[7] C. Team, "Available CRAN Packages By Name." [En línea]. Disponible en: https://cran.r-project.org/. [Accedido: 12-Feb-2023].

[8] AUTOMOTIVE INDUSTRY ACTION GROUP (AIAG), *Measurement System Analysis, Reference Manual*, 4ta ed. 2010.

[9] Automotive Industry Action Group (AIAG)., Measurement Systems Analysis, and American Society for Quality, "MEASUREMENT SYSTEMS ANALYSIS Reference Manual Third Edition," *Automot. Ind. Action Gr.*, vol. 4, p. 241, 2002.

[10] E. J. E. Vázquez, *Seis-Sigma Metodología y Técnicas.*, 1st ed. México D.F.: LIMUSA, 2004.

[11] Minitab INC., "Métodos y fórmulas para el sesgo del sistema de medición - Minitab." [En línea]. Disponible en: https://support.minitab.com/es-mx/minitab/18/help-and-how-to/quality-and-process-improvement/measurement-system-analysis/how-to/gage-study/gage-linearity-and-bias-study/methods-and-formulas/gage-bias/#percent-bias. [Accedido: 01-Nov-2019].

[12] M. I. S. Domínguez, "Código en R para evaluar linealidad y sesgo." [En línea]. Disponible en: https://github.com/mivansalas/linealidad. [Accedido: 13-Feb-2023].