ANÁLISIS ESTADÍSTICO DEL PROYECTO "GUÍA DE REQUERIMIENTOS"

Oscar Julian Layton[a]

OSE Bogotá, Laboratorio de transformación regional, bancolombia,

Resumen

Este documento muestra el diseño de un experimento para evaluar el efecto del uso de la herramienta *guía de requerimientos* en el tiempo de solución de procesos, se especifica la toma junto con la selección de los factores más influyentes en la caracterización del modelo y muestra un abordaje desde el punto de vista Estadístico para la toma de decisiones en la Oficina de Servicios Empresariales Bogotá (OSE).

1. Diseño del experimento

Considerando el uso de la herramienta guía de requerimientos en OSE Bogotá, se consideran algunas características propias de la herramienta que son partícipes para un buen diseño del experimento, en estas se destaca dos tipos de factores el primero hace referencia a **herramienta** el cual es un control local y el segundo a **transacción** el cual es un factor de tipo aleatorio. En correspondencia a lo anterior se establece que un diseño apropiado es el denominado **Diseño factorial de dos factores** con tres repeticiones a continuación se presentan cada uno de los niveles de factor:

Transacción (T): Para transacción se consideran 5 de las transacciones más comunes, este es un factor aleatorio y sus niveles están determinados por:

- Recaudos.
- Pagos.
- Cartera.
- Traslados SV.
- CB (corresponsales bancarios).

Herramienta (H): En lo que corresponde al uso de la herramienta se define el factor H quien tiene dos niveles:

- Con la herramienta.
- Sin la herramienta.

El presente diseño está representado por el siguiente diagrama de estructura el cual hace correspondencia DBCA:

[a]Integrantes del laboratorio de transformación regional

$$\begin{tikzcd} & \mu \arrow[dl] \arrow[dr] & \\ H_j & \cdots & \underset{\sim}{T_i} \arrow[dl] \\ & e_{k(ij)} \arrow[from=ul] & \end{tikzcd}$$

Modelo correspondiente al diseño

$$y_{ijk} \;=\; \mu \;+\; \underset{\sim}{T_i} \;+\; H_j \;+ (TH)_{ij} +\; \epsilon_{k(ij)} \qquad con \quad \begin{cases} i = 1,2,3,4,5 \\ j = 1,2 \\ k = 1,2,3 \end{cases}$$

Por consiguiente, los supuestos del modelo asociado al diagrama son:

$$e_{ij} \overset{\text{iid}}{\sim} N(0,\sigma_e^2) \quad , \quad T_i \overset{\text{iid}}{\sim} N(0,\sigma_T^2) \quad , \quad TH_{ij} \overset{\text{iid}}{\sim} N(0,\sigma_{TH}^2)$$

$$cov(T_i, e_{k(ij)}) = cov(H_j, e_{k(ij)}) = \quad ... \quad = cov((TH)_{ij}\ ,\ e_{k(ij)}) = 0 \qquad \forall_{ijk}$$

$$sa : \left\{ \sum_j H_j = 0 \ ; \ \sum_i (TH)_{ij} = 0 \ ; \ \sum_j (TH)_{ij} = 0 \right\}$$

- y_{ijk} : Es el tiempo requerido para la finalización del proceso i, con el uso j de la herramienta en la k-ésima persona.
- T_i : Hace referencia a la i-ésima transacción.
- H_j : Caracteriza al j-ésimo uso de la herramienta.
- $e_{k(ij)}$: Error aleatorio propio del experimento (se pretende controlar)

Unidades experimentales: Personas a las que se les realizan las pruebas.

Factores y tratamientos:

- Transacción: (Recaudos, Pagos, Cartera, Traslados SV, Corresponsales Bancarios).
- Herramienta: (Con la herramienta, Sin la herramienta)

2. Toma de tiempos - Recopilación de información

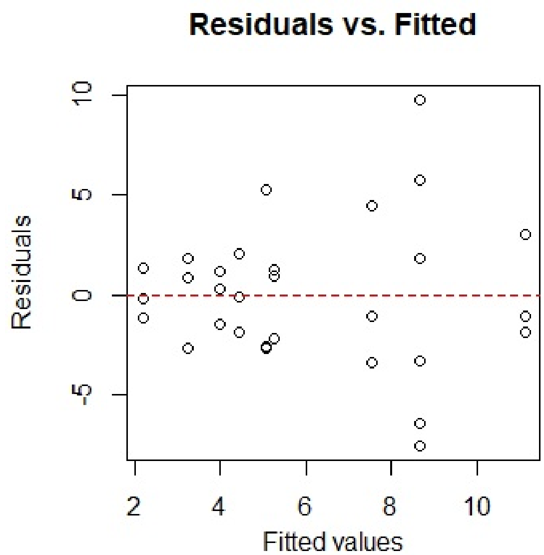
Tabla 1: Tiempos tomados en el área transaccional

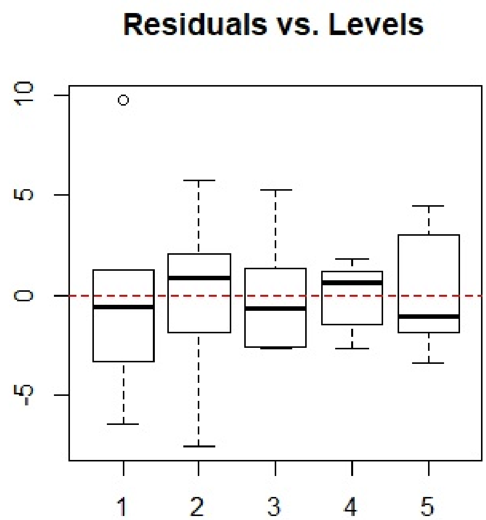
	Proceso				
Uso guia requerimientos	**Pagos**	**Recaudos**	**Cartera**	**Traslados**	**Correspo banc**
	6.19	2.55	1.03	4.14	4.12
Con Herramienta	6.52	4.31	3.50	5.04	6.50
	3.05	6.51	2.03	1.57	11.00
	19.47	10.49	2.37	5.13	9.21
Sin Herramienta	5.34	14.42	9.39	2.50	14.11
	2.21	2.07	2.48	4.30	10.05

Con respecto al anterior modelo propuesto, se hace necesario verificar si los supuestos se cumplen, en este orden de ideas se propone el siguiente conjunto de gráficas:

3. Verificación de supuestos del modelo.

Para verificar el supuesto de homogeneidad de las varianzas ha de observarse la gráfica de residuales vs valores ajustados la cual no debe mostrar ningún patrón, además de que los datos se encuentren en una misma franja, esto da señales de homoceasticidad.





Por consiguiente, la gráfica de los residuales v.s nivel del factor debe mostrar que todos los datos estén alrededor de cero (mostrándose la no relación entre residuales y las transacciones), en miras a confirmar este hecho de homogeneidad se realiza la prueba la prueba de Levene, presentándose la siguiente salida:

```
> levene.test(anova$residuals, data$bloque)

Modified robust Brown-Forsythe Levene-type test based on
the absolute deviations from the median

data:  anova$residuals
Test Statistic = 2.0359, p-value = 0.1647
```

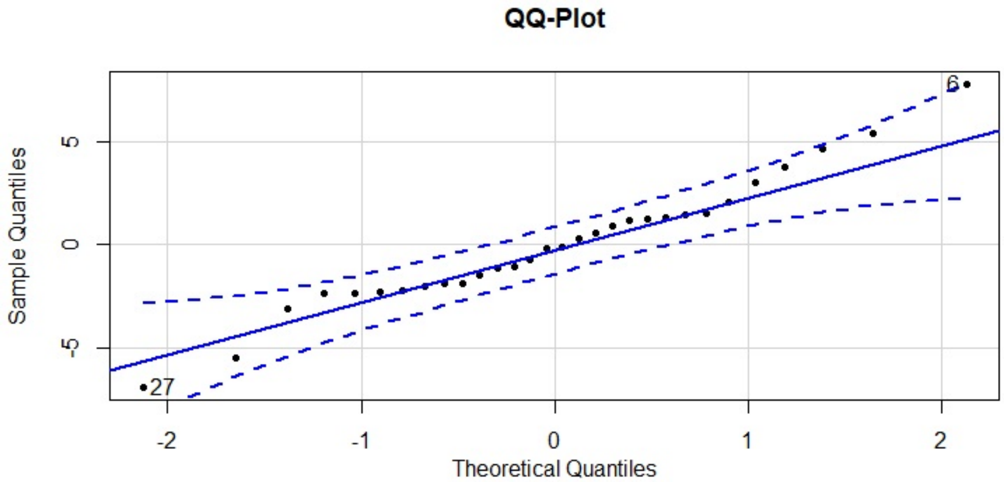
Considerando el p-value= 0.1647 de la prueba Levene se determina no rechazar la hipótesis nula H_0 de varianza constante, y tomando en consideración las gráficas anteriormente descritas se confirma que el supuesto de homogeneidad en las varianzas se cumple.

```
> dwtest(anova)

Durbin-Watson test

data:  anova
DW = 2.0335, p-value = 0.5928
alternative hypothesis: true autocorrelation is greater than 0
```

Para observar la independencia ha de observarse la gráfica de residuales v.s observaciones ordenadas, el no tener ningún patrón en este tipo de gráfica indica que los datos fueron tomados de datos independientes.



En el qqplot puede identificarse que los datos se encuentran alrededor de la línea recta, además se sitúan dentro de las bandas de confianza lo cual da indicios de cumplimiento del supuesto de normalidad en los errores. Para confirmar este hecho se realizan las pruebas de Kolmogorov-Smirnov y Jarque-Bera para contrastar la hipótesis:

$$H_0 : F_n(w) = N(\mu_0, \sigma_0^2) \qquad vs \qquad H_1 : F_n(w) \neq N(\mu_0, \sigma_0^2)$$

A continuación se presenta la salida en R para el desarrollo de los anteriores contrastes:

```
> ks.test(anova$residuals,"pnorm",mean(anova$residuals),sd(anova$residuals))

##  One-sample Kolmogorov-Smirnov test
data:  anova$residuals
D = 0.12956, p-value = 0.6481
alternative hypothesis: two-sided


> jarque.bera.test(anova$residuals)

##  Jarque Bera Test
data:  anova$residuals
X-squared = 0.52185, df = 2, p-value = 0.7703
```

Considerando el p-value= 0.6481 de la prueba Kolmogorov-Smirnov y el p-value=0.7703 de la prueba Jarque-Bera, se determina no rechazar la hipótesis nula H_0, por consiguiente tomando en consideración el qqplot y las pruebas anteriormente descritas se confirma el supuesto de normalidad.

Para realizar el análisis Estadístico se lleva a cabo el ANOVA propio del diseño, especificando los grados de libertad, la suma de cuadrados, las esperanzas de los cuadrados medios y probar la hipótesis de interés. Se presenta la siguiente tabla de varianza mostrándose a su vez las distintas fuentes de variabilidad (ver tabla 2):

Tabla 2: Análisis de varianza del modelo

Causa de Variación	Grados de Libertad	Suma de Cuadrados	Cuadrado Medio	F	P-Valor
T	4	$\frac{1}{6}\sum_i y_{i..}^2 - \frac{1}{30}y_{...}^2 = 160.52$	40.13	2.901	0.0481
H	1	$\frac{1}{15}\sum_j y_{.j.}^2 - \frac{1}{30}y_{...}^2 = 68.95$	68.95	4.984	0.0372
TH	4	$\frac{1}{3}\sum_i y_{ij.}^2 - \frac{1}{6}y_{i..}^2 - \frac{1}{15}y_{.j.}^2 + \frac{1}{30}y_{...}^2 = 22.12$	5.53	0.400	0.8065
Error	20	$\sum_{ijk} y_{ijk}^2 - \frac{1}{2}\sum_{ij} y_{ij.}^2 = 276.68$	13.83		
Total	29	$\sum_{ijk} y_{ijk}^2 - \frac{1}{24}y_{...}^2 = 528.27$			

Tabla 3: Tabla de la esperanza de los cuadrados medios del modelo

Cuadrado Medio	i	j	k	E(CM)
$\frac{1}{4}SC(T)$	1	2	3	$\sigma_e^2 + 6\sigma_T^2$
$SC(H)$	5	0	3	$\sigma_e^2 + 3\sigma_{TH}^2 + 15\sum_j H_j^2$
$\frac{1}{4}SC(TH)$	1	0	3	$\sigma_e^2 + 3\sigma_{TH}^2$
$\frac{1}{20}SCE$	1	1	1	σ_e^2

Al tratarse de un diseño factorial mixto, la hipótesis a considerar es la igualdad de efectos medios en el uso de la herramienta:

$$H_0 : \mu_1 = \mu_2 \qquad vs \qquad H_1 : \mu_i \neq \mu_j \quad \text{para algún } i \neq j, \quad i, j = 1, 2$$

Teniendo que la estadística de prueba es:

$$F_c = \frac{CM(H)}{CM(TH)} \quad ; \quad F_c \mid_{H_0} \sim F_{[1,4]}$$

Resultado 1

A un nivel de significancia de 5 % se rechaza H_0, ya que $F_c = 12.46835 > 7.70 = F_{[1,4,0.05]}$ con lo cual se determina que el uso de la herramienta guía de requerimientos difiere significativamente en el tiempo requerido para la solución de procesos.

Resultado 2

En la práctica es importante establecer inferencias sobre los resultados obtenidos, para esto se crea un intervalo de confianza para el tiempo destinado al uso de la herramienta guía de requerimientos:

$$\overline{y_{.i}} \pm t_{(glE, \frac{\alpha}{2})} \sqrt{\frac{CME}{r_i}}$$

$$4.537 \pm t_{(20, \frac{0.05}{2})} \sqrt{\frac{13.83}{3}}$$

Un intervalo con el 95 % de confianza para el tiempo requerido al usar la herramienta es $(0.071, 9.002)$ minutos.

Resultado 3

El efecto del NO uso de la herramienta es estimado por medio de $\overline{y}_{.2} - \overline{y}_{..} = 7.569 - 6.053 = 1.516$ con lo cual puede afirmarse que NO usar la guía de requerimientos establece 1 unidad de tiempo por encima de la media. En especial $\overline{y}_{.2} - \overline{y}_{.1} = 7.569 - 4.537 = 3.032$ que significa que en base a la información suministrada en el experimento, se puede establecer que más de 3 minutos se ahorra con el uso de la herramienta

Resultado 4

Conociendo lo anterior se tiene:

$$LS(\overline{y}_{.2} - \overline{y}_{.1}) : \quad \overline{y}_{.2} - \overline{y}_{.1} + t_{(glE, \frac{\alpha}{2})} \sqrt{\frac{CME}{r} \sum_{i=1}^{2} a_i^2}$$

$$3.032 + t_{(20, \frac{0.05}{2})} \sqrt{\frac{2}{3} 13.83}$$

Un intervalo con el 95 % de confianza supone que el tiempo ahorrado al usar la herramienta puede ser menor que 9.34 minutos.

4. Conclusiones

Se realiza un **Diseño factorial mixto de dos factores** con un efecto aleatorio y se crea un estudio en relación a un modelo Estadístico resultante, además se evidencia el cumplimiento de los supuestos de normalidad, heteroceasticidad y no correlación de los efectos. No obstante, se concluye en relación a la información suministrada del diseño y análisis del experimento, estableciéndose que:

- Según el resultado 1, con un error del 5 % se afirma que SI existe diferencias significativas en la reducción del tiempo para la solución de requerimientos en la OSE Bogotá, debida al uso de la herramienta *guía de requerimientos.*
- Según el resultado 2, con un error del 5 % se concluye que el tiempo requerido al usar la *guía de requerimientos* está entre (0.071,9.002) minutos, en otras palabras de 100 veces que se realicen requerimientos, aproximadamente 5 de estos superarán el tiempo de 9 minutos.
- Según el resultado 3, con un error del 5 % se ahorra aproximadamente 3 minutos al usar la herramienta en el área transaccional.
- En relación al resultado 4, con un nivel de significancia del 5 % se concluye que el tiempo ahorrado al usar la herramienta *guía de requerimientos* puede llegar a ser de 9 minutos, Esto quiere decir puede superarse este umbral.
- En relación a la herramienta se observan algunos cambios en los conceptos que hacen referencia a las transacciones, por tal motivo es necesario actualizar la herramienta.