

TECNOLÓGICO DE COSTA RICA

ESCUELA DE INGENIERÍA EN COMPUTACIÓN

ASEGURAMIENTO DE LA CALIDAD DE SOFTWARE

Análisis de datos de proyectos de software

Prof. MSc. Saúl Calderón Ramírez

Jose Rodolfo Garita Chaves *2013038397*

José Mario Naranjo Leiva *2013034348*

Manuel Calderón Campos *201023758*

8 de Septiembre de 2016

Índice

1. Introducción	1
2. Coeficiente de Pearson para dos variables aleatorias	1
3. Análisis de los datos	2
3.1. Diagrama de los datos	2
3.2. Análisis de los resultados y conclusiones	45
3.3. Sugerencias	47
4. Uso de otros coeficientes de correlación	48
4.1. Coeficiente de la correlación de Spearman (ρ)	49
4.2. Coeficiente de Kendall (τ)	50
4.3. Correlación de Spearman y Kendall en Matlab	50
5. Referencias	51

Índice de figuras

1.	Diagrama acumulada ciclomático acumulados.	2
2.	Numero de fallas acumuladas.	3
3.	Numero de fallos mayores acumuladas.	4
4.	Numero de fallos menores acumuladas.	5
5.	Numero de fallos triviales acumuladas.	6
6.	Numero de funciones acumuladas.	7
7.	Numero de líneas acumuladas.	8
8.	Tiempo de auditoría diseño con los requerimientos acumuladas.	9
9.	Tiempo de auditoría implementación con los requerimientos acumuladas.	10
10.	Tiempo de diseño acumuladas.	11
11.	Tiempo de requerimientos acumuladas.	12
12.	Diagrama ciclomático.	13
13.	Numero de fallas críticas.	14
14.	Numero de fallos mayores.	15
15.	Numero de fallos menores.	16
16.	Numero de fallos triviales.	17
17.	Numero de funciones.	18
18.	Numero de líneas.	19
19.	Tiempo de auditoría diseño con los requerimientos.	20
20.	Tiempo de auditoría implementación con los requerimientos.	21
21.	Tiempo de diseño.	22
22.	Tiempo de requerimientos.	23
23.	Diagrama ciclomático.	24
24.	Numero de fallas críticas.	25
25.	Numero de fallos mayores.	26
26.	Numero de fallos menores.	27
27.	Numero de fallos triviales.	28
28.	Numero de funciones.	29
29.	Numero de líneas.	30
30.	Tiempo de auditoría diseño con los requerimientos.	31
31.	Tiempo de auditoría implementación con los requerimientos.	32
32.	Tiempo de diseño.	33
33.	Tiempo de requerimientos.	34
34.	Diagrama ciclomático.	35

35.	Numero de fallas criticas.	36
36.	Numero de fallos mayores.	37
37.	Numero de fallos menores.	38
38.	Numero de fallos triviales.	39
39.	Numero de funciones.	40
40.	Numero de lineas.	41
41.	Tiempo de auditoría diseño con los rquerimientos.	42
42.	Tiempo de auditoría implementación con los requerimientos. .	43
43.	Tiempo de diseño.	44
44.	Tiempo de requerimientos.	45

1. Introducción

Para elaborar el siguiente reporte se utilizaron las herramientas de matriz de covarianza y el coeficiente de correlación de Pearson para analizar un conjunto de datos, recolectados durante entre los años 2010-2011 y 2013-2014, de proyectos de software de la empresa *Sport Analytics*.

A partir del uso de las herramientas estadísticas mencionadas y el análisis de los datos recolectados, se asesora la toma de decisiones relacionadas a los tiempos de desarrollo. Asimismo, en este reporte se enumera un conjunto de sugerencias para los procesos de aseguramiento de la calidad del software aplicados proyectos futuros de la empresa.

2. Coeficiente de Pearson para dos variables aleatorias

El coeficiente de correlación de dos variables es la medida de la fuerza entre las dependencias lineales de las variables. En otras palabras, este coeficiente es un índice utilizado para medir el grado de relación entre dos variables aleatorias cuantitativas.

El coeficiente de Pearson puede tomar el rango de valores entre +1 a -1. Un valor de 0 indica que no hay asociación entre las dos variables, un valor mayor a 0 indica una asociación positiva; lo que significa, conforme el valor de una de las variables aumenta, también lo hace el valor de la otra variable. Un valor menor a 0 indica una asociación negativa; esto es, conforme el valor de una variable incrementa, el valor de la otra variable también decrementa. [2]

El cálculo del coeficiente de correlación de Pearson $\rho_{A,B}$ para dos variables X_A y X_B aleatorias está dado por:

$$\rho_{A,B} = \frac{cov(X_A, X_B)}{\sigma_{X_A} \times \sigma_{X_B}} = \frac{1}{N} \sum_{i=1}^N \left(\frac{X_A - \mu_{X_A}}{\sigma_{X_A}} \right) \left(\frac{X_B - \mu_{X_B}}{\sigma_{X_B}} \right)$$

Nótese $cov(X_A, X_B)$ como la covarianza entre las variables aleatorias X_A y X_B .

Esta fórmula es la referente para la implementación manual de la función *pearson* con la cuál se utilizó para verificar los resultados retornados la

función *corrcoef* implementada para MATLAB. [1]

La función *pearson* implementada se encuentra en el archivo *pearson.m* adjunto a este documento.

3. Análisis de los datos

3.1. Diagrama de los datos

A continuación se presentan los diagramas para ilustrar los datos numéricos recolectados en los proyectos de desarrollo de software.

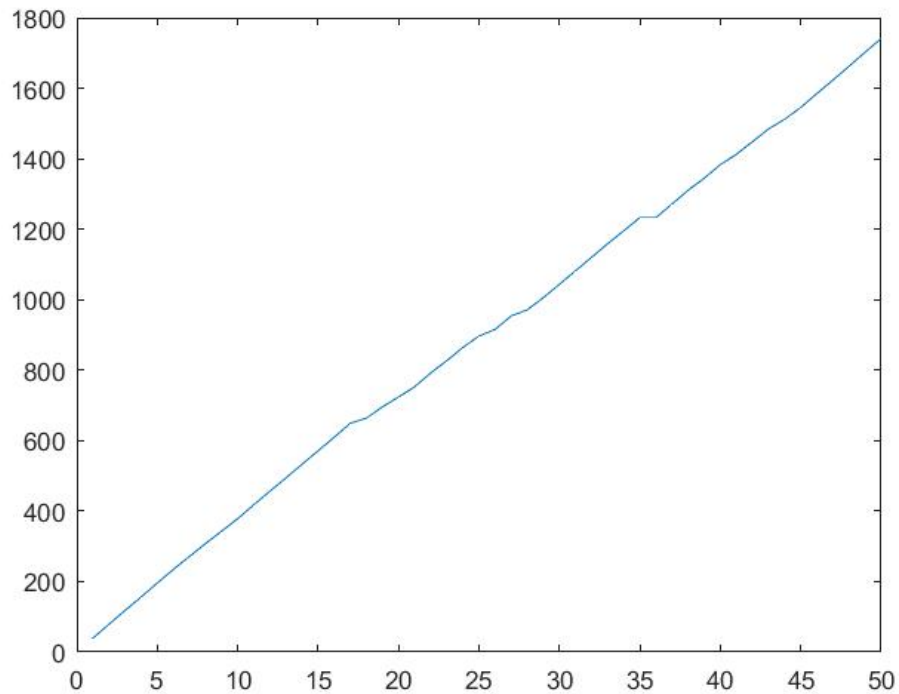


Figura 1: Diagrama acumulada ciclomático acumulados.

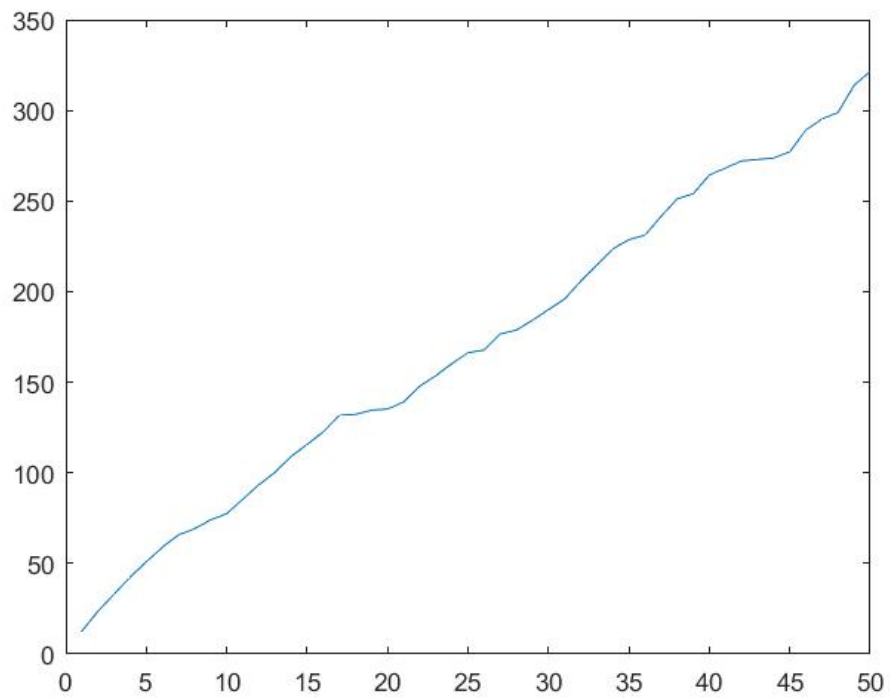


Figura 2: Numero de fallas acumuladas.

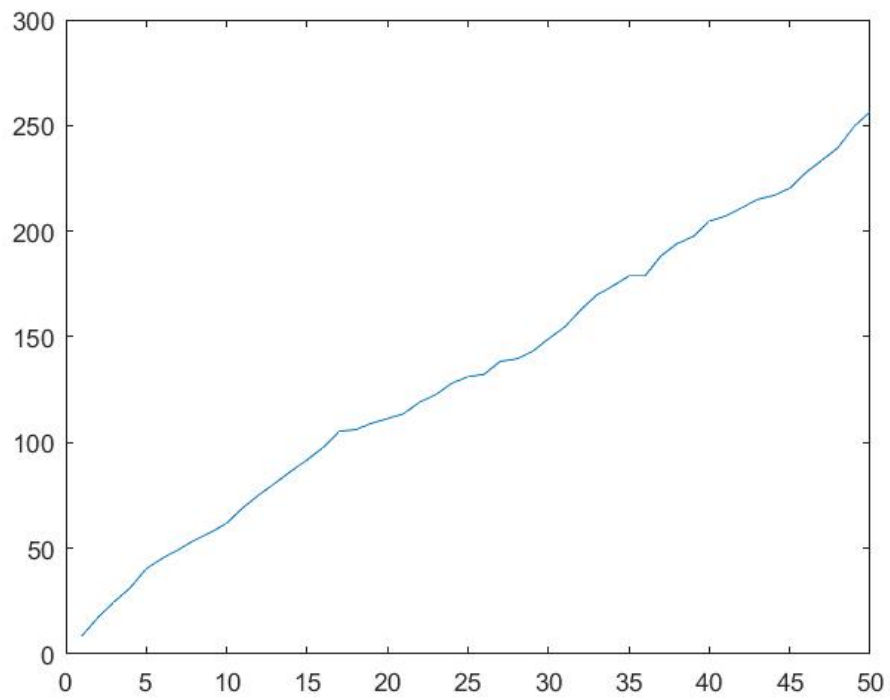


Figura 3: Numero de fallos mayores acumuladas.

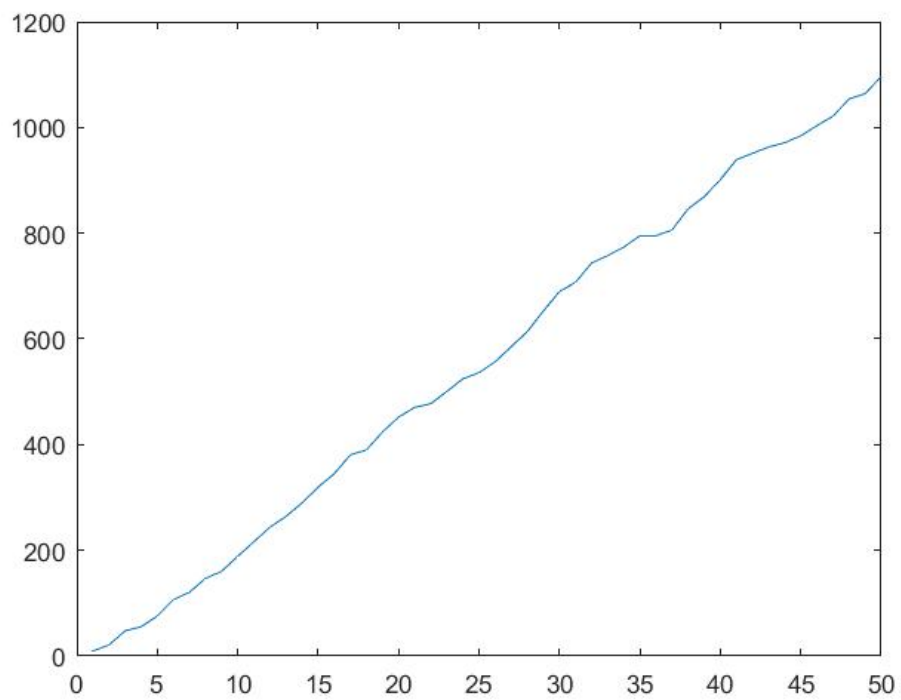


Figura 4: Numero de fallos menores acumuladas.

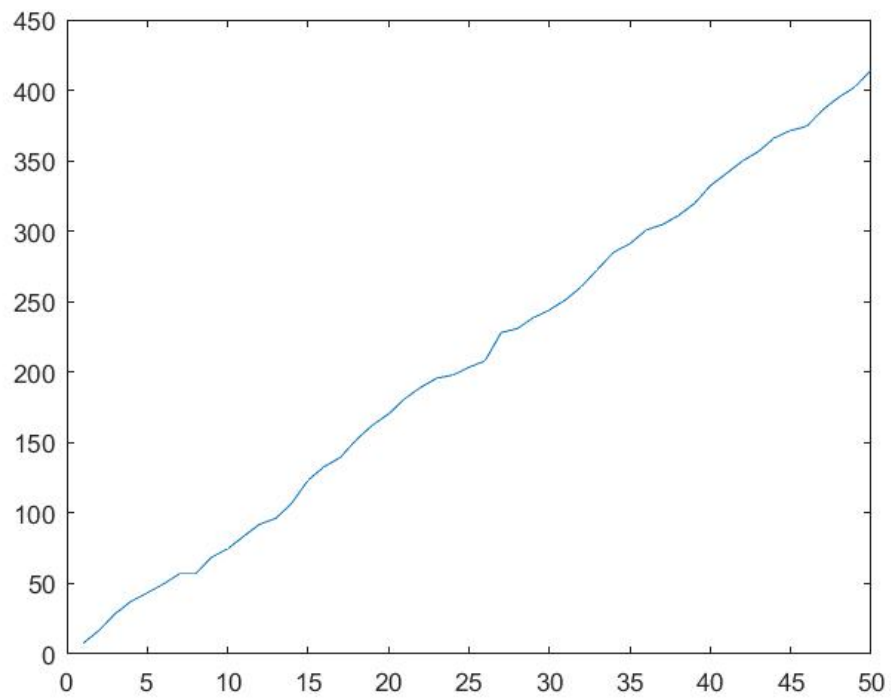


Figura 5: Numero de fallos triviales acumuladas.

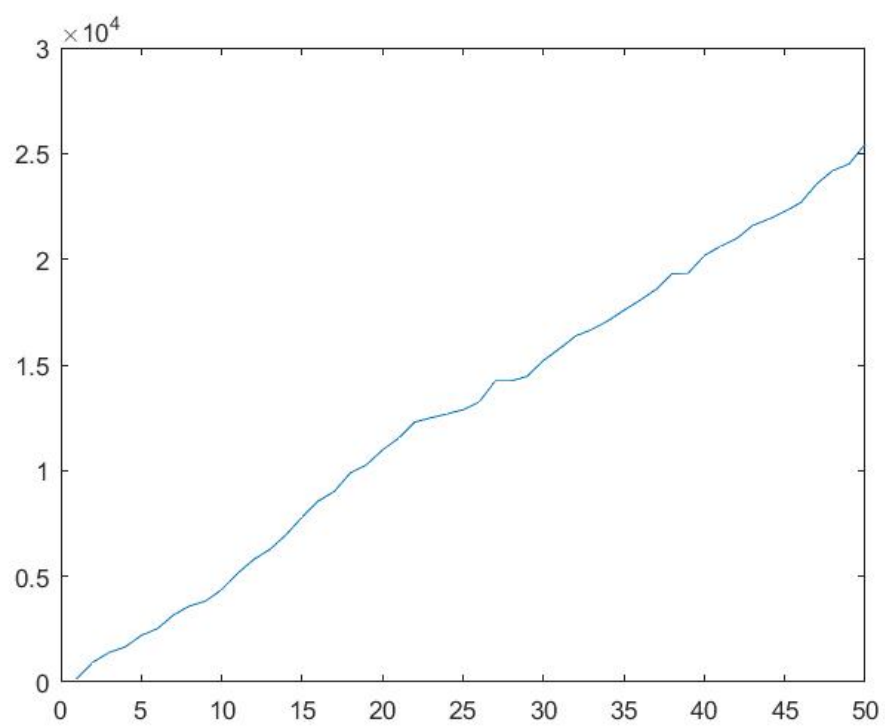


Figura 6: Numero de funciones acumuladas.

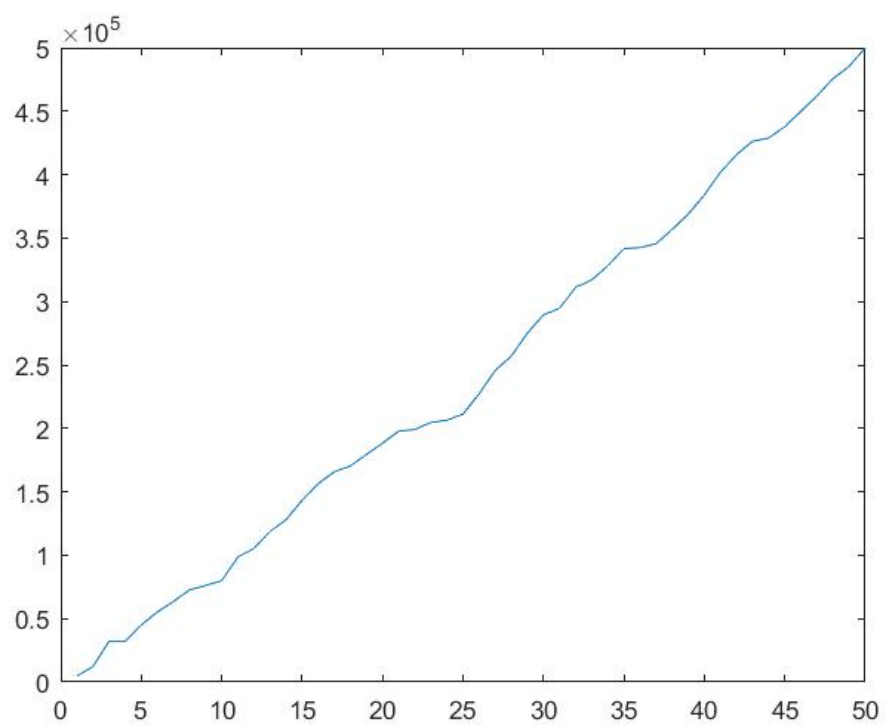


Figura 7: Numero de lineas acumuladas.

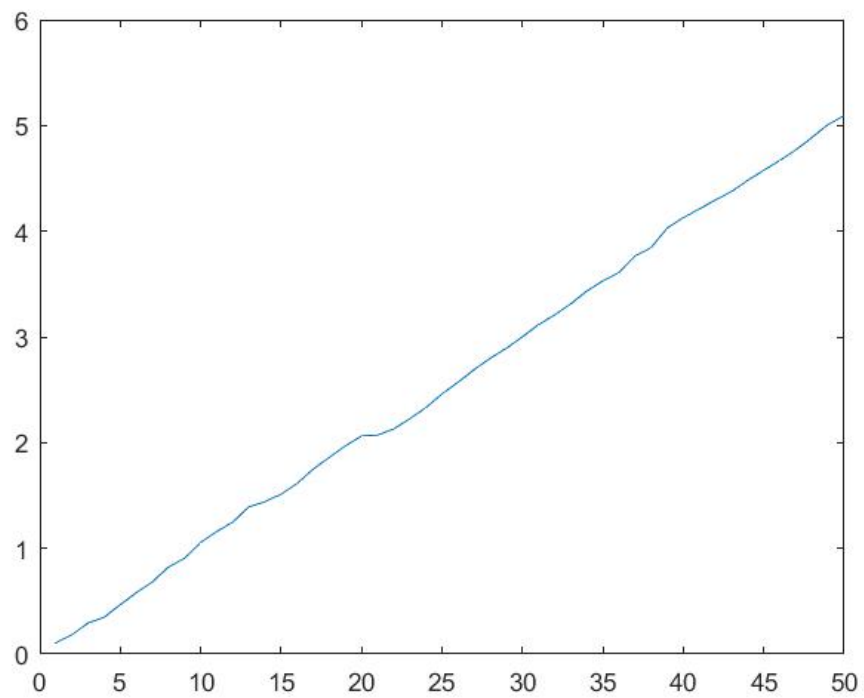


Figura 8: Tiempo de auditoria diseño con los rquerimientos acumuladas.

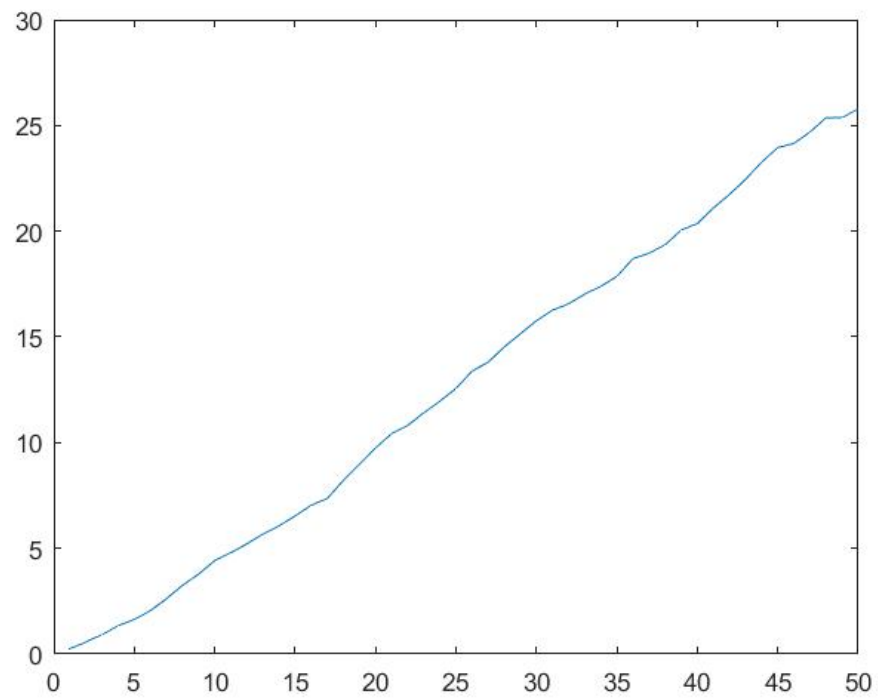


Figura 9: Tiempo de auditoria implementación con los requerimientos acumuladas.

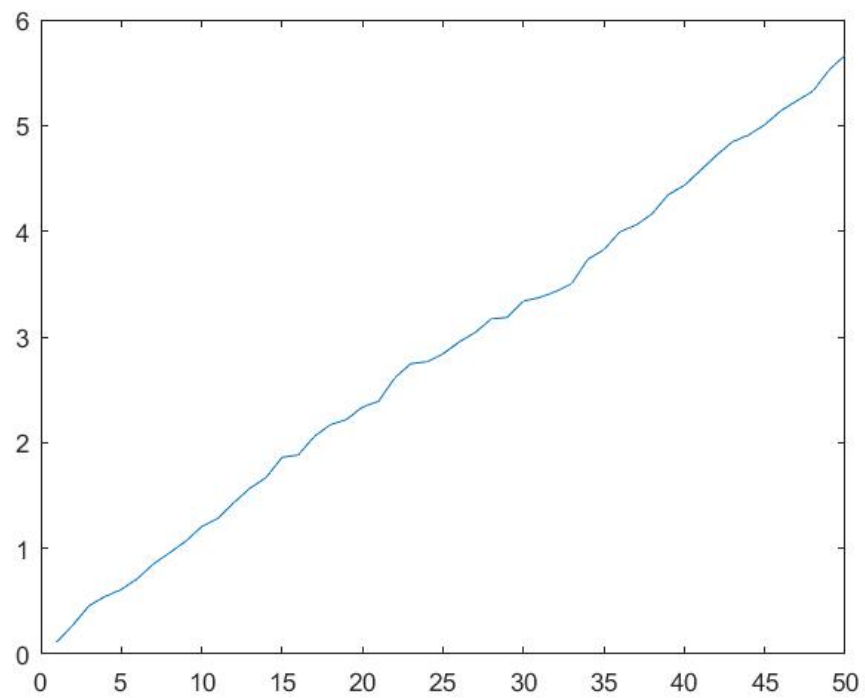


Figura 10: Tiempo de diseño acumuladas.

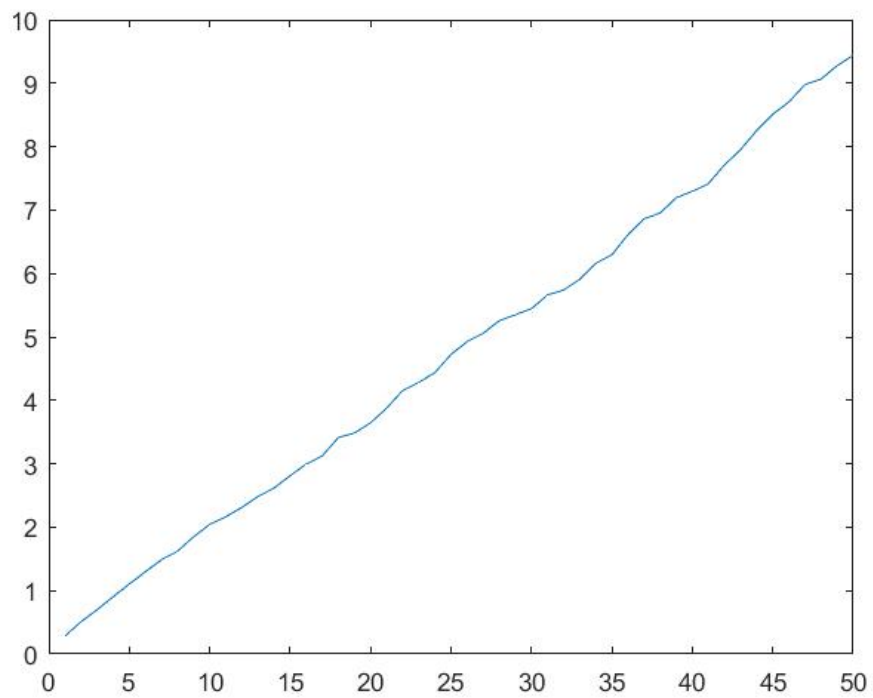


Figura 11: Tiempo de requerimientos acumuladas.

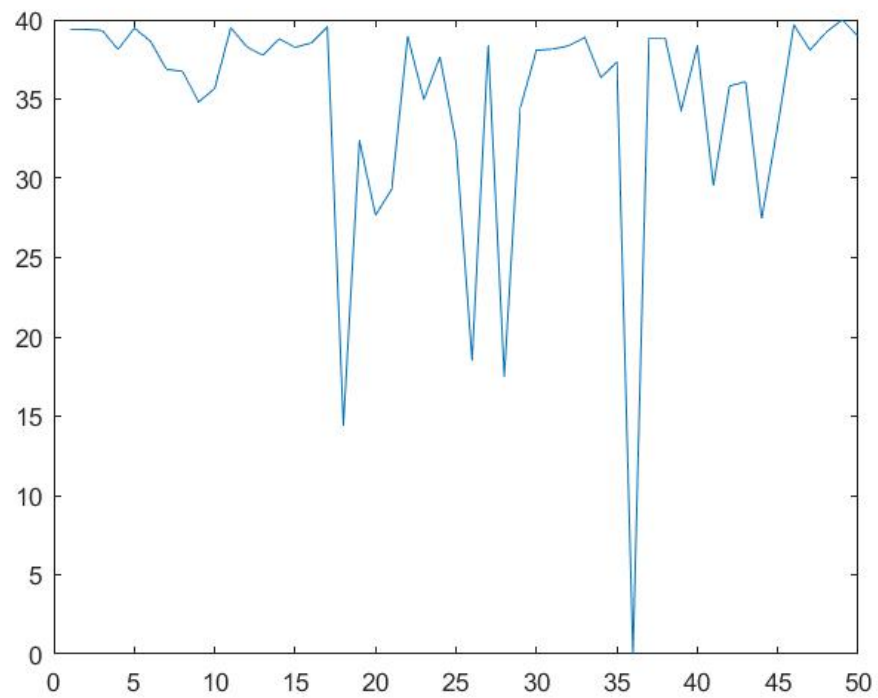


Figura 12: Diagrama ciclomatico.

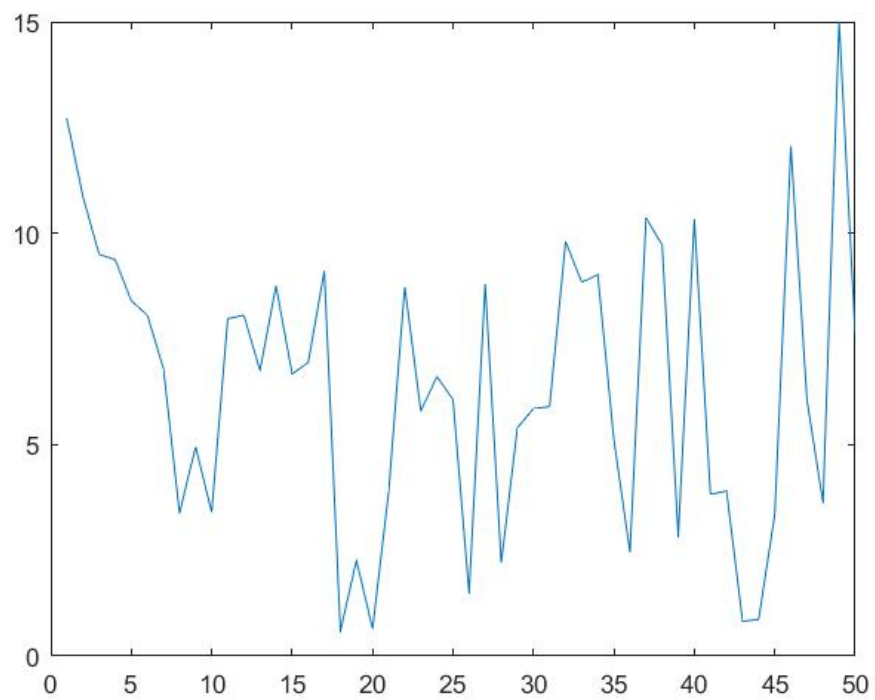


Figura 13: Numero de fallas criticas.

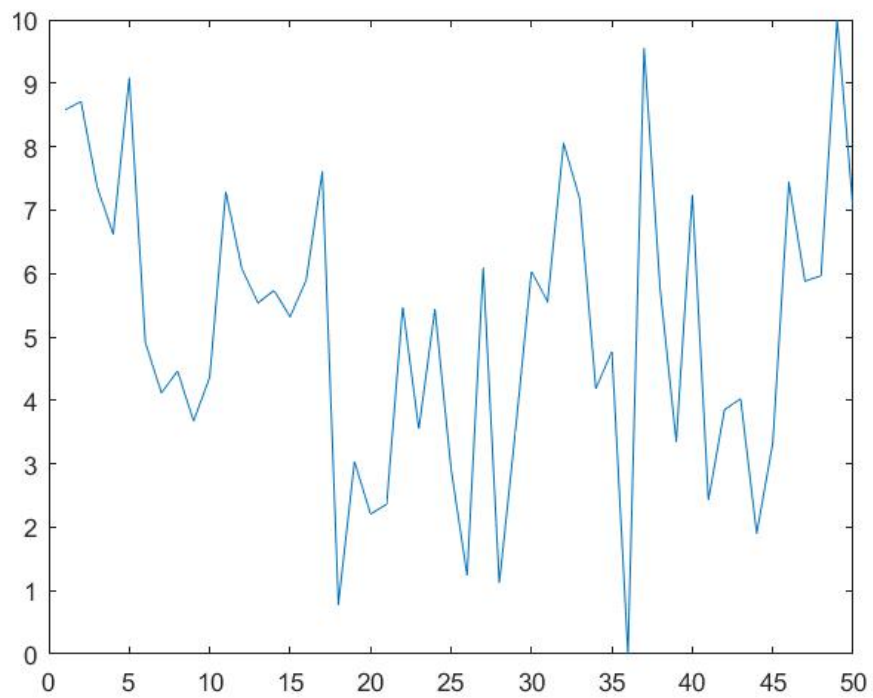


Figura 14: Numero de fallos mayores.

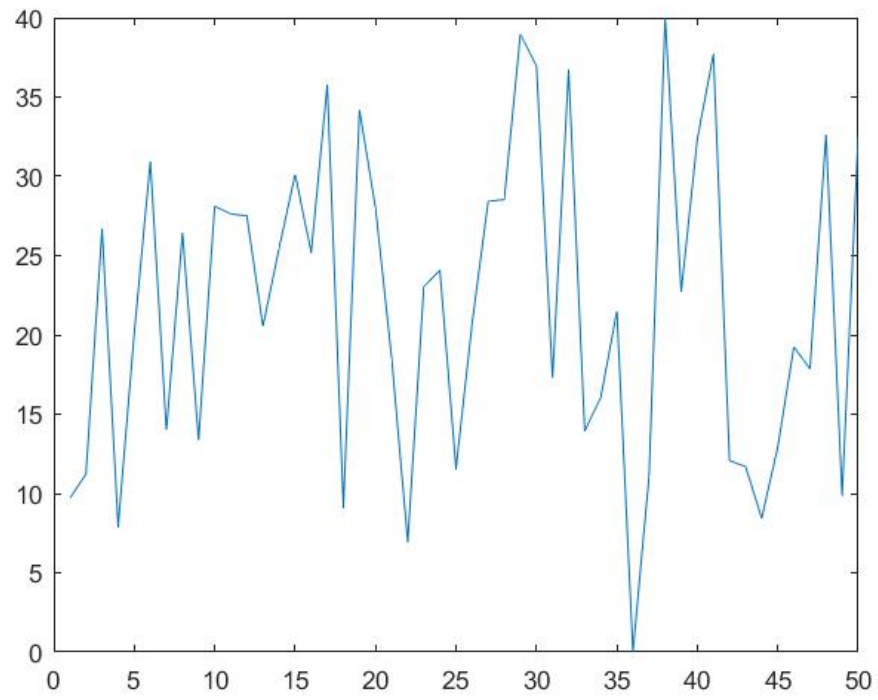


Figura 15: Numero de fallos menores.

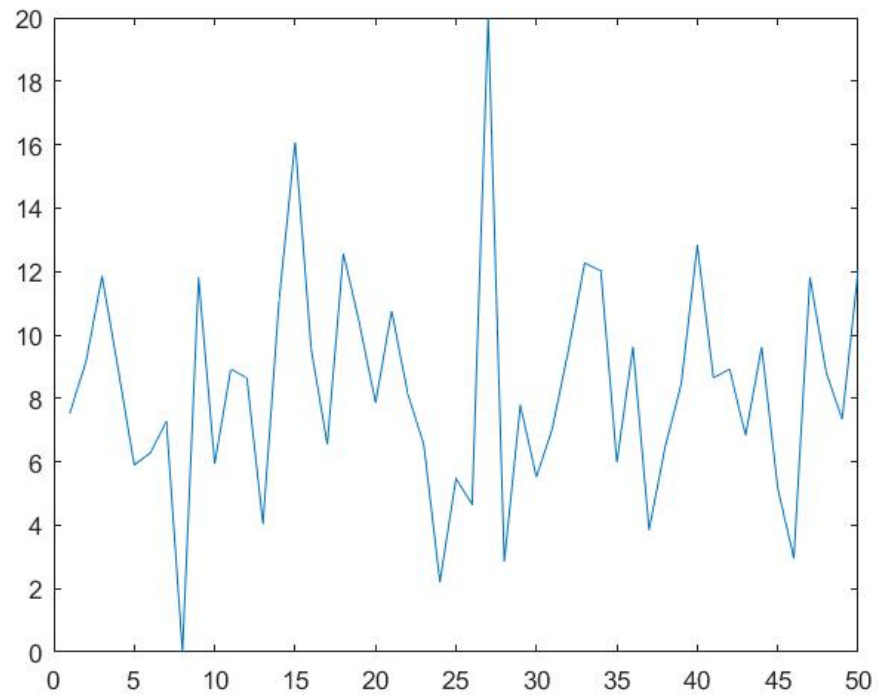


Figura 16: Numero de fallos triviales.

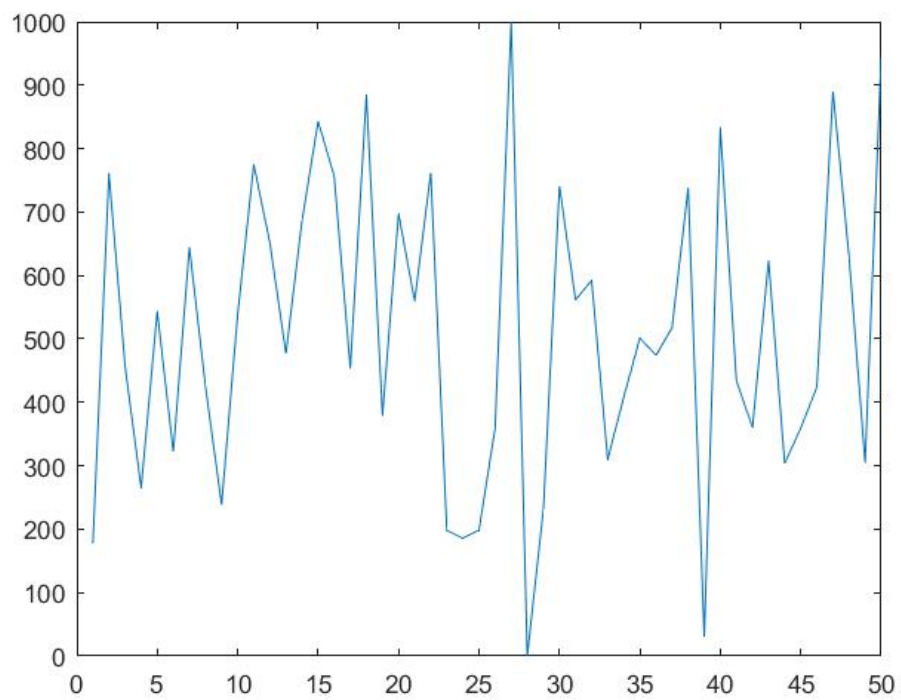


Figura 17: Numero de funciones.

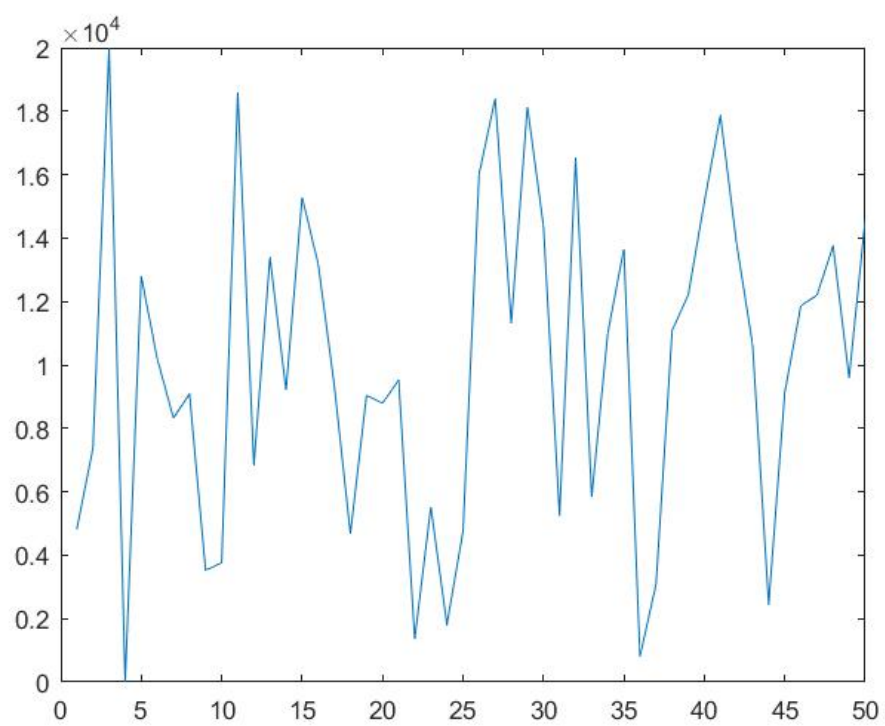


Figura 18: Numero de lineas.

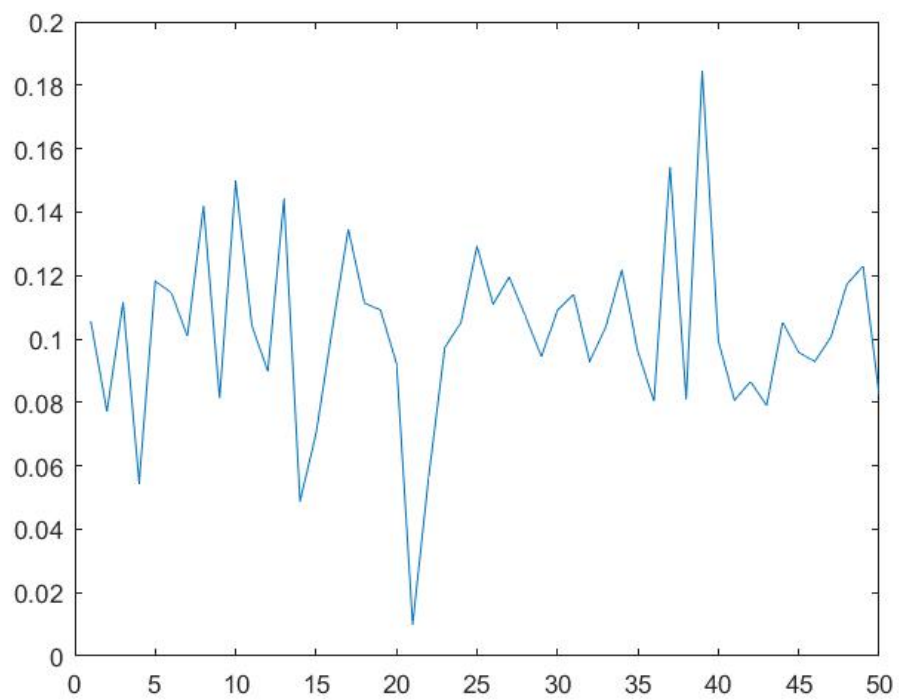


Figura 19: Tiempo de auditoria diseño con los rquerimientos.

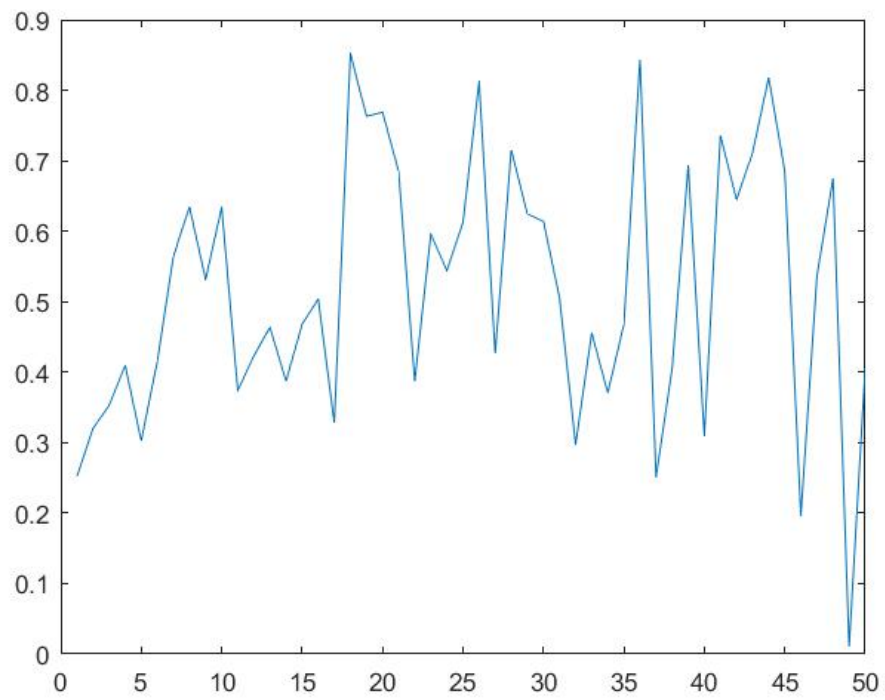


Figura 20: Tiempo de auditoria implementación con los requerimientos.

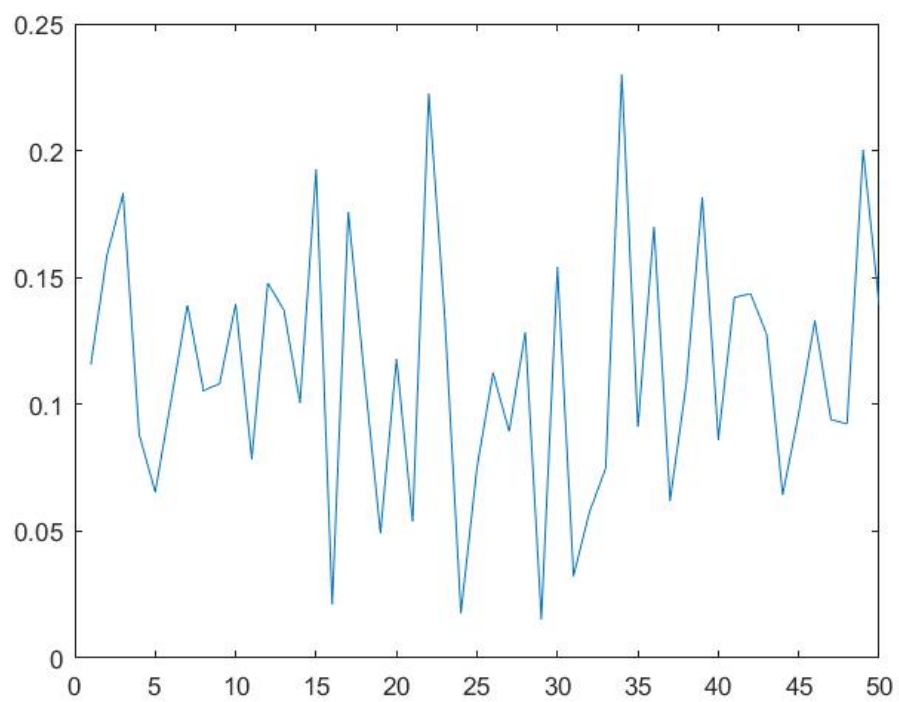


Figura 21: Tiempo de diseño.

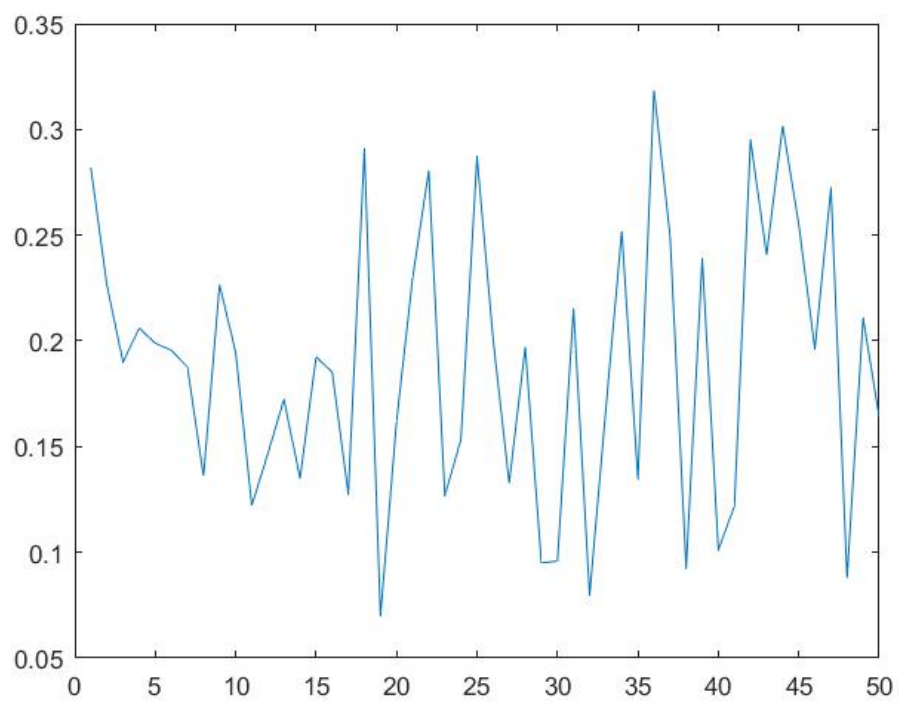


Figura 22: Tiempo de requerimientos.

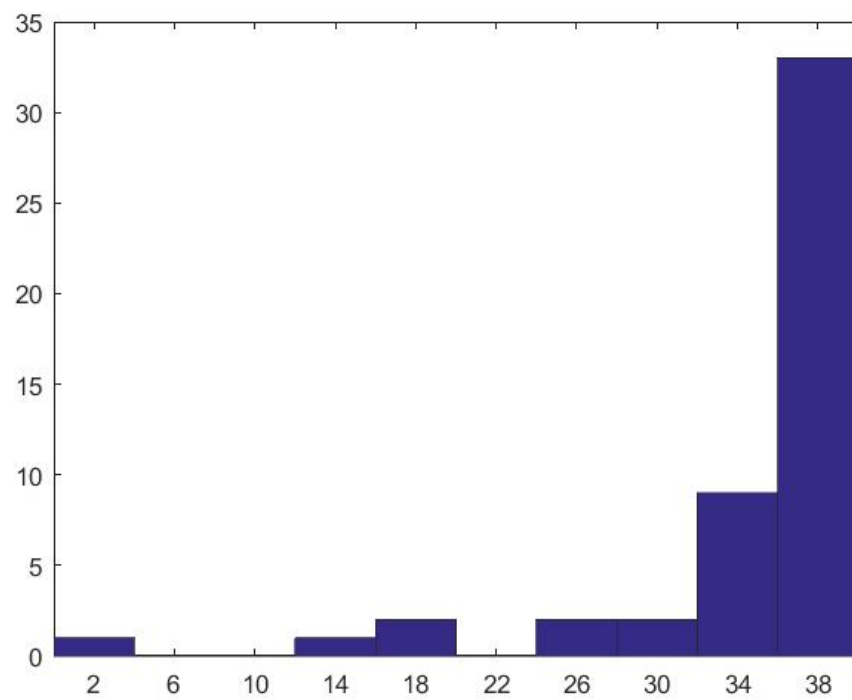


Figura 23: Diagrama ciclomatico.

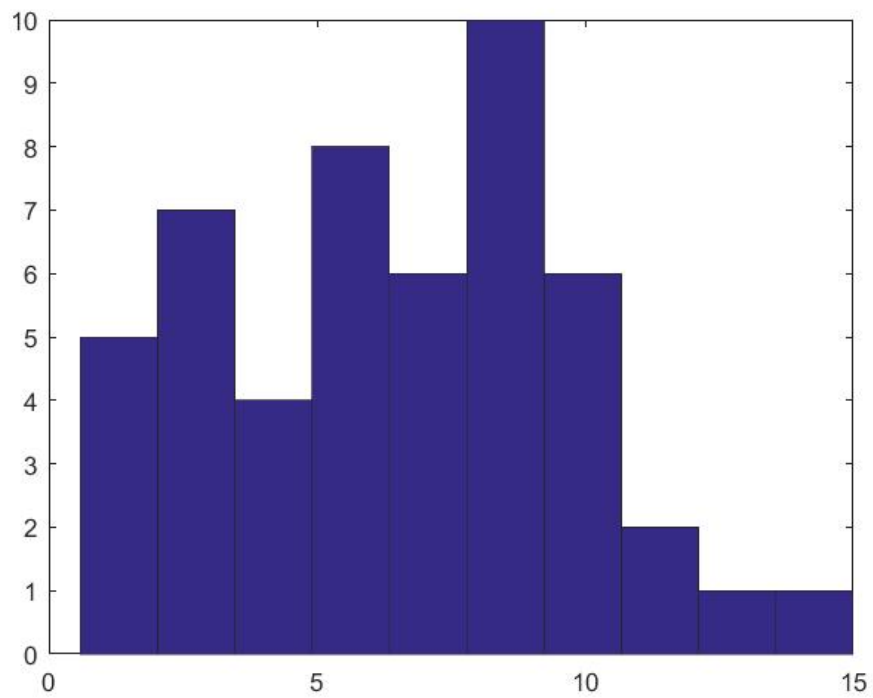


Figura 24: Numero de fallas criticas.

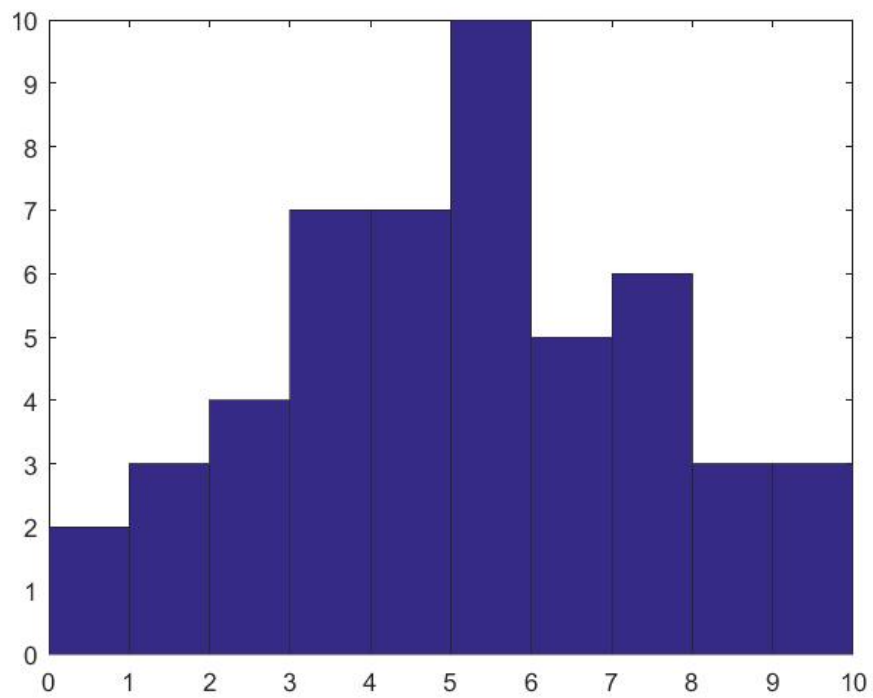


Figura 25: Numero de fallos mayores.

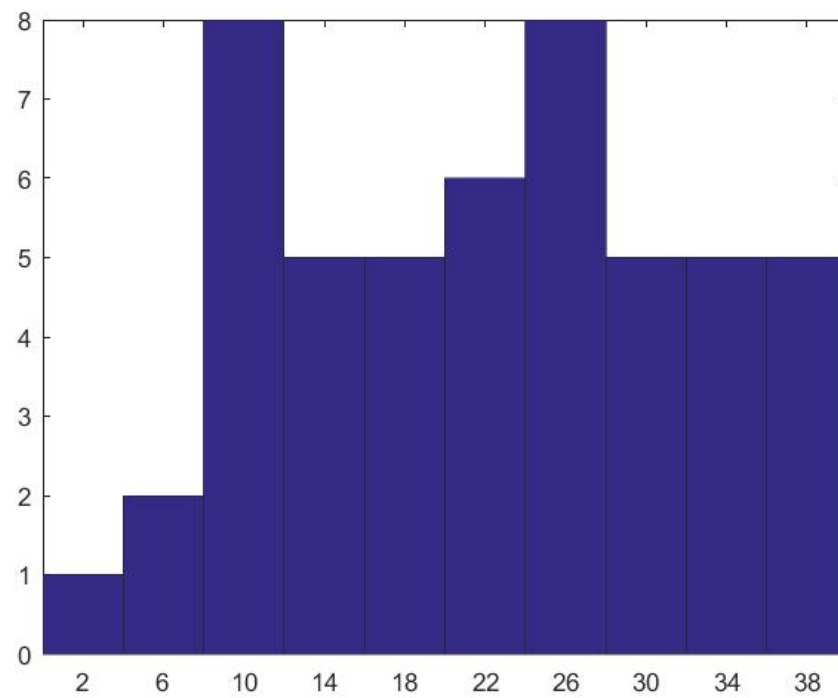


Figura 26: Numero de fallos menores.

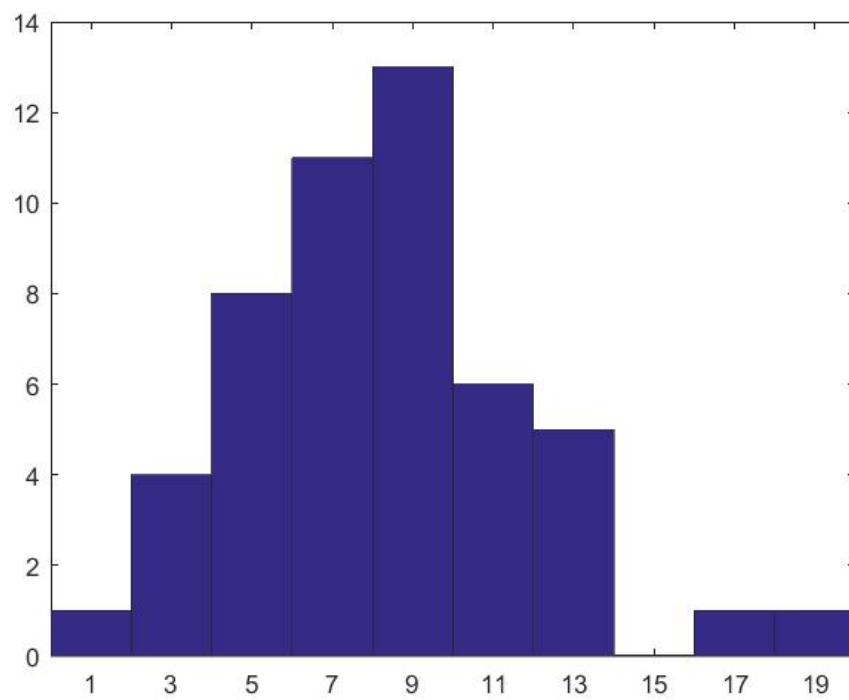


Figura 27: Numero de fallos triviales.

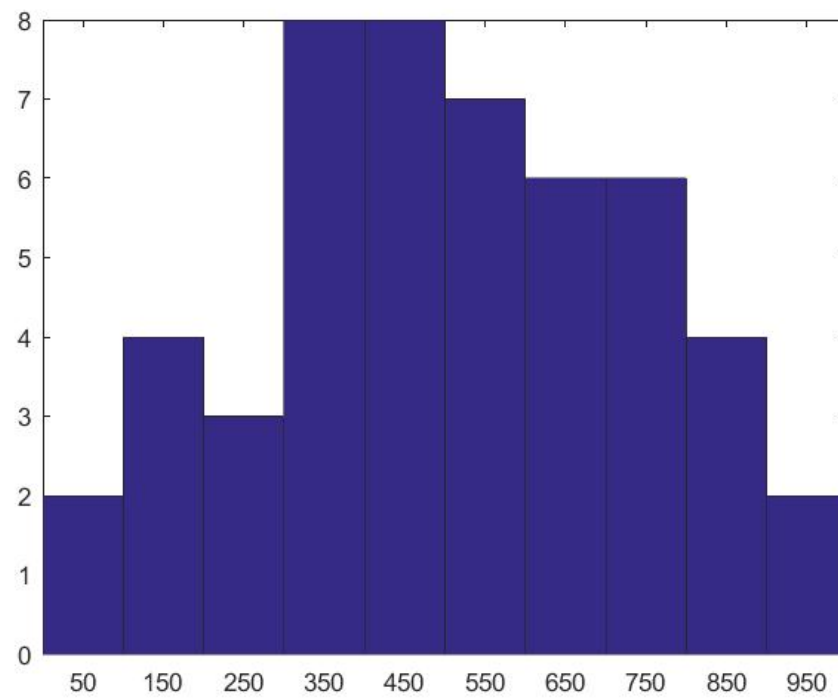


Figura 28: Numero de funciones.

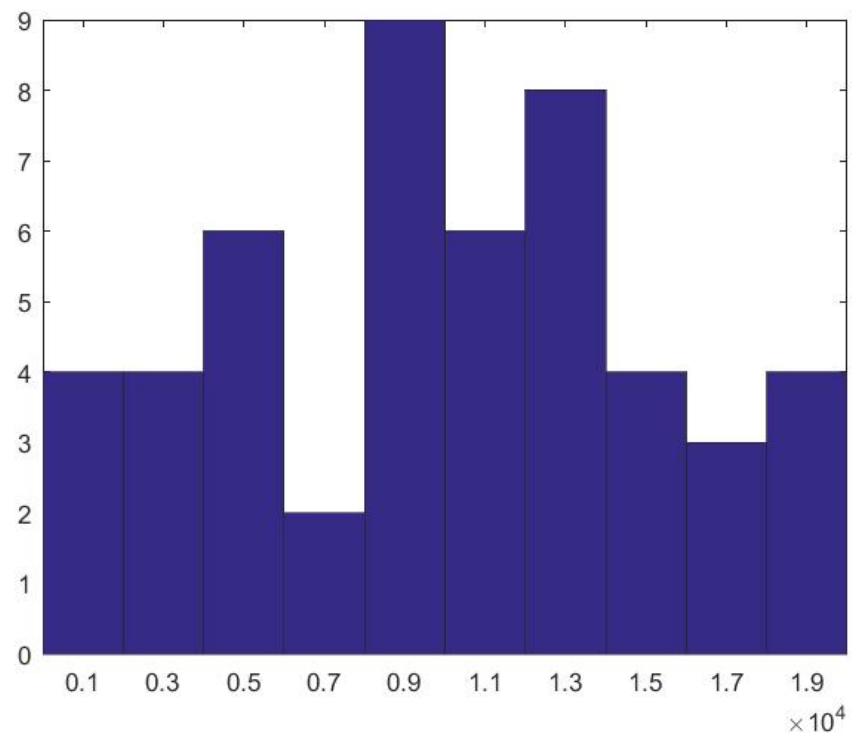


Figura 29: Numero de lineas.

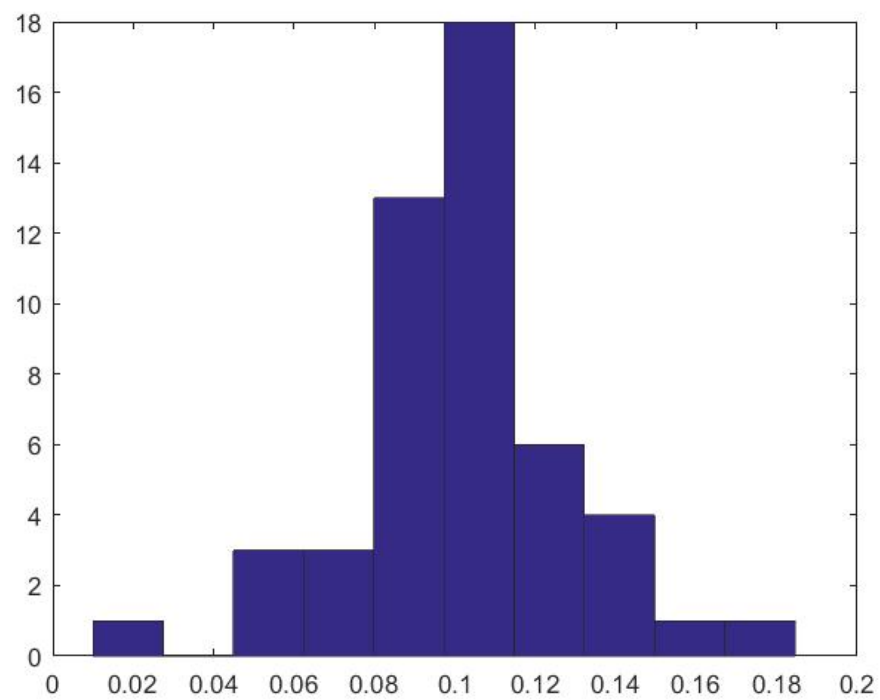


Figura 30: Tiempo de auditoría diseño con los rquerimientos.

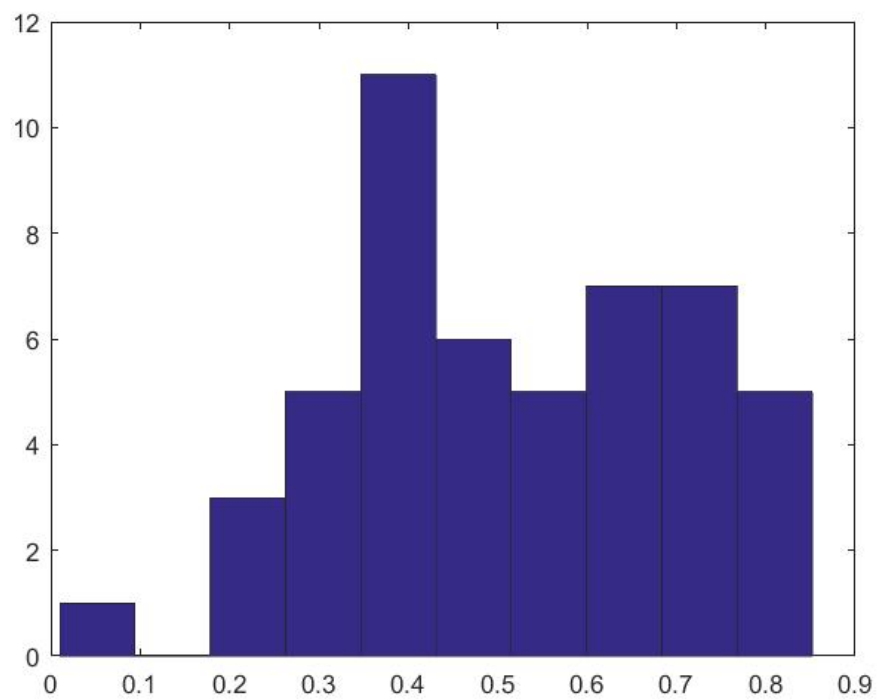


Figura 31: Tiempo de auditoría implementación con los requerimientos.

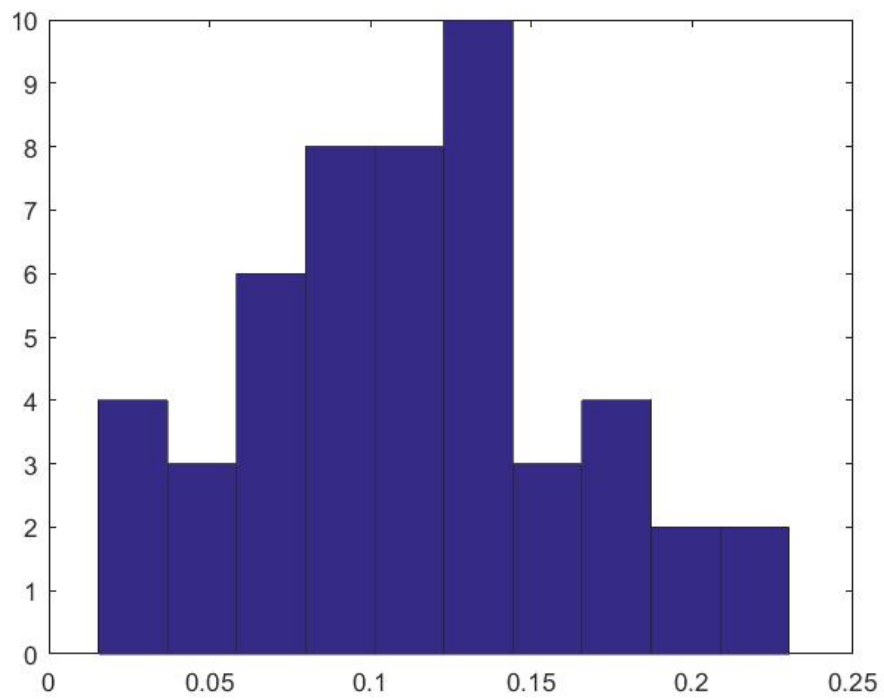


Figura 32: Tiempo de diseño.

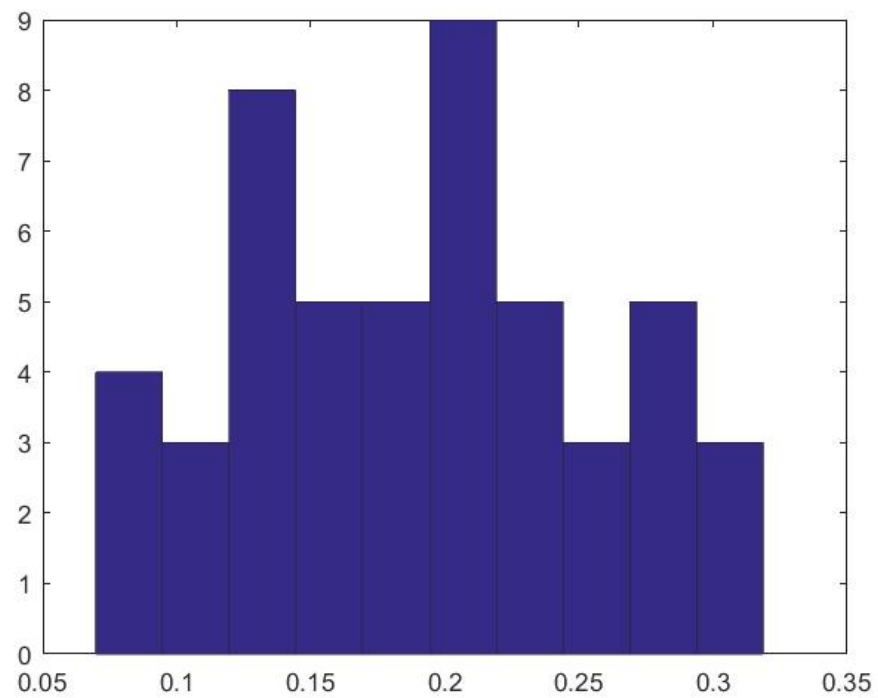


Figura 33: Tiempo de requerimientos.

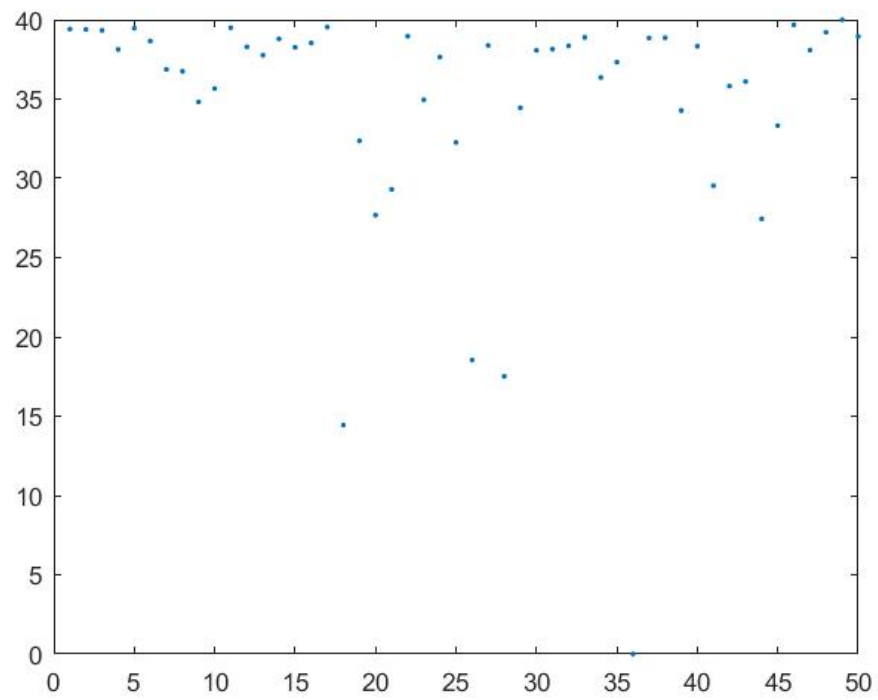


Figura 34: Diagrama ciclomatico.

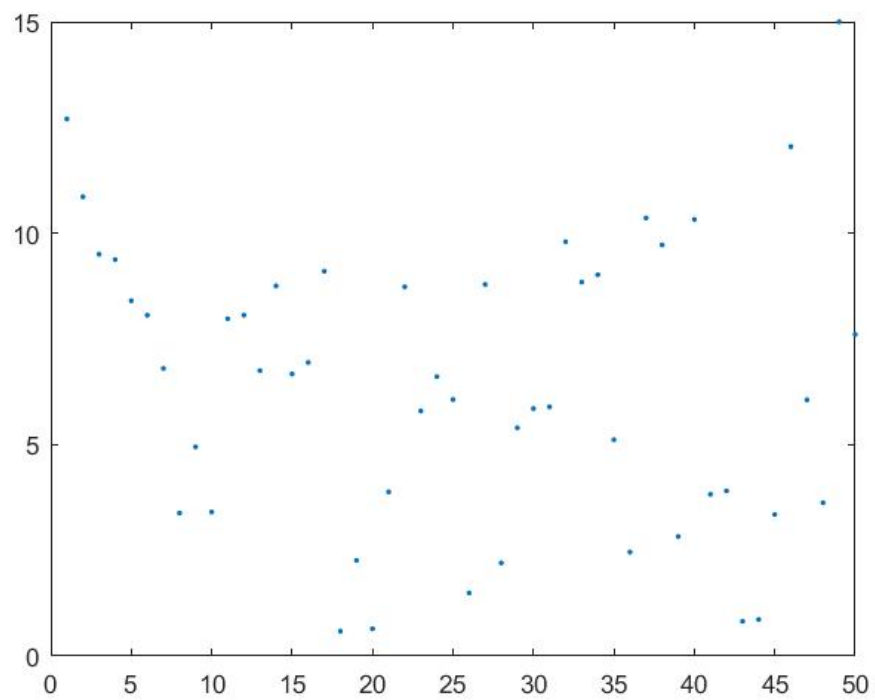


Figura 35: Numero de fallas criticas.

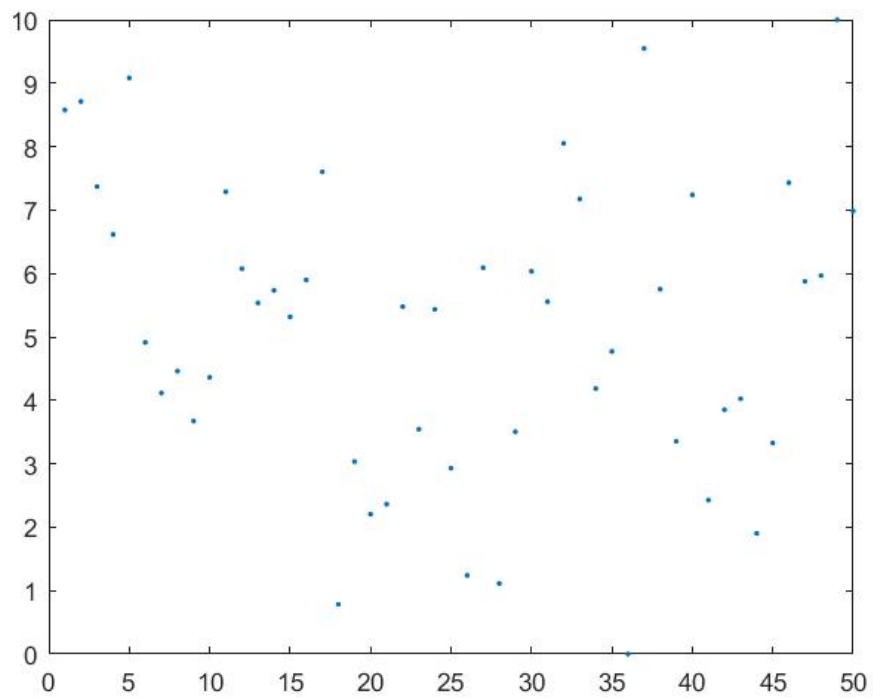


Figura 36: Numero de fallos mayores.

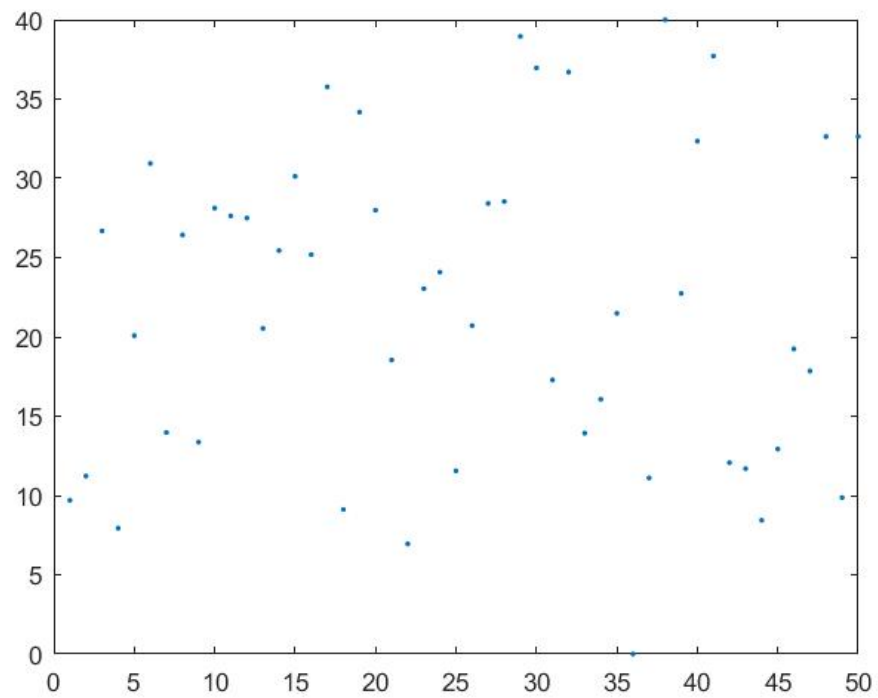


Figura 37: Numero de fallos menores.

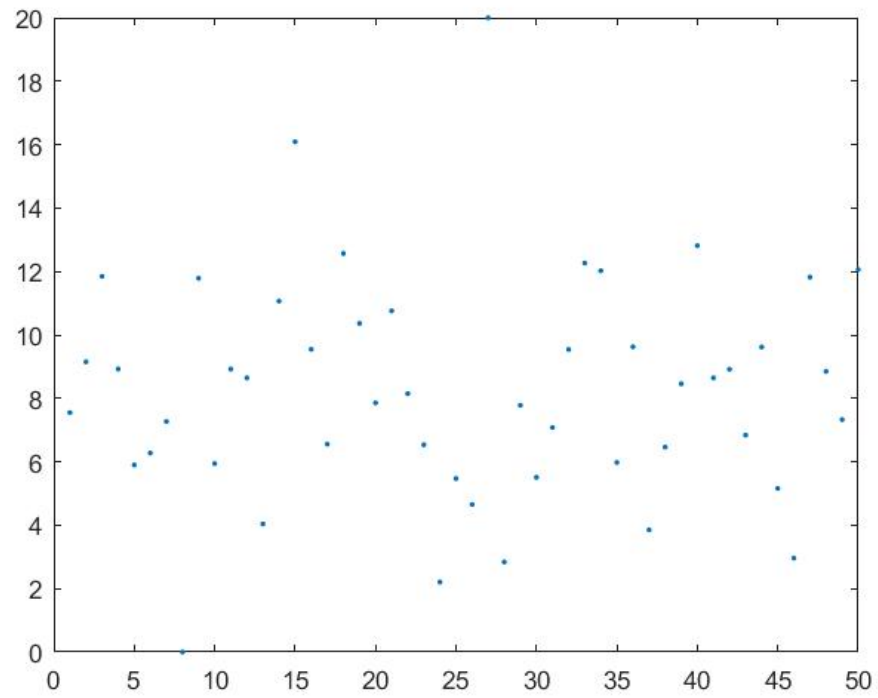


Figura 38: Numero de fallos triviales.

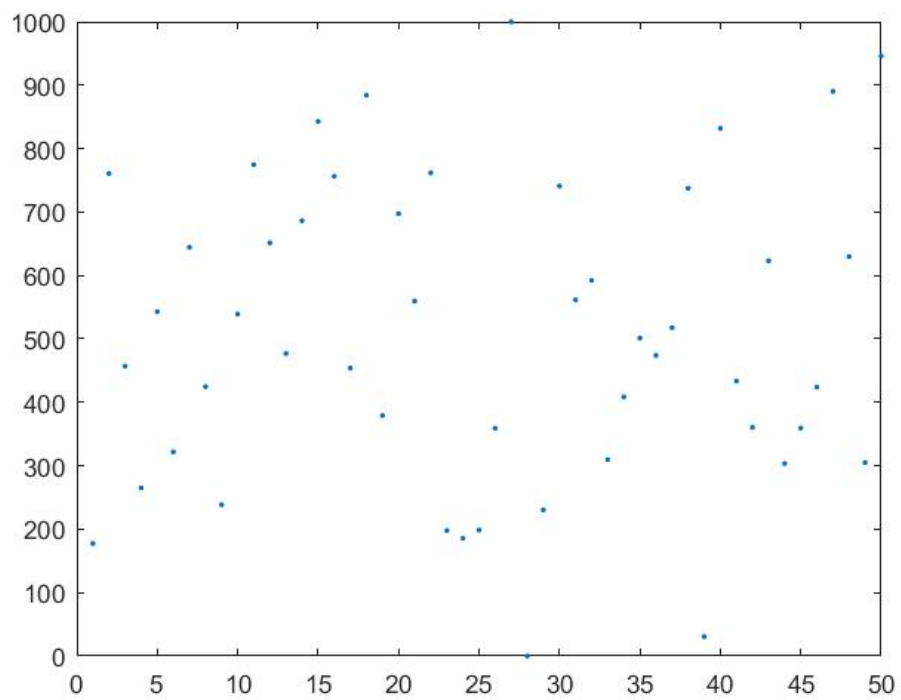


Figura 39: Numero de funciones.

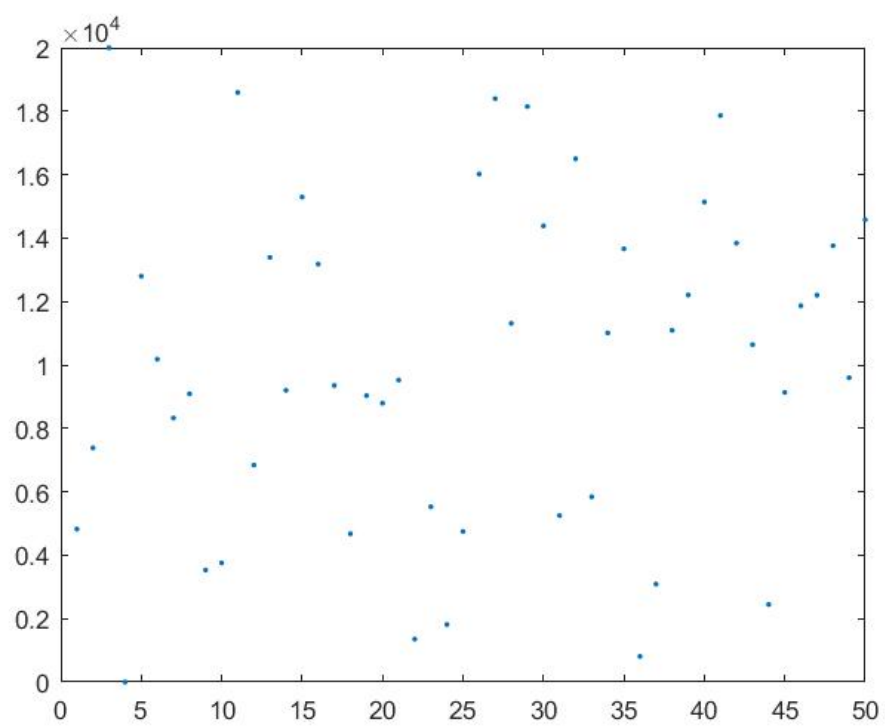


Figura 40: Numero de lineas.

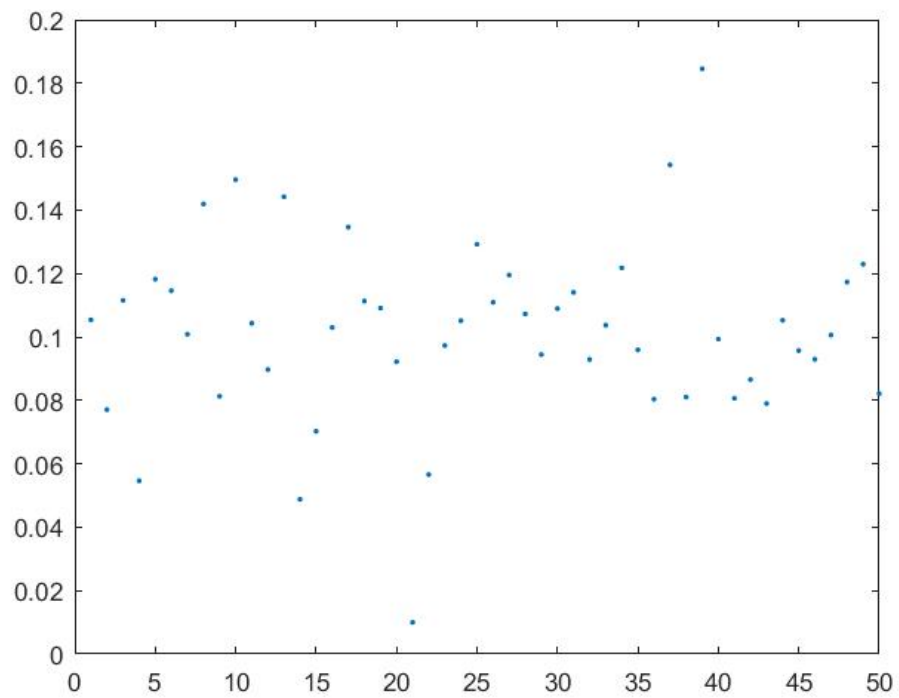


Figura 41: Tiempo de auditoría diseño con los rquerimientos.

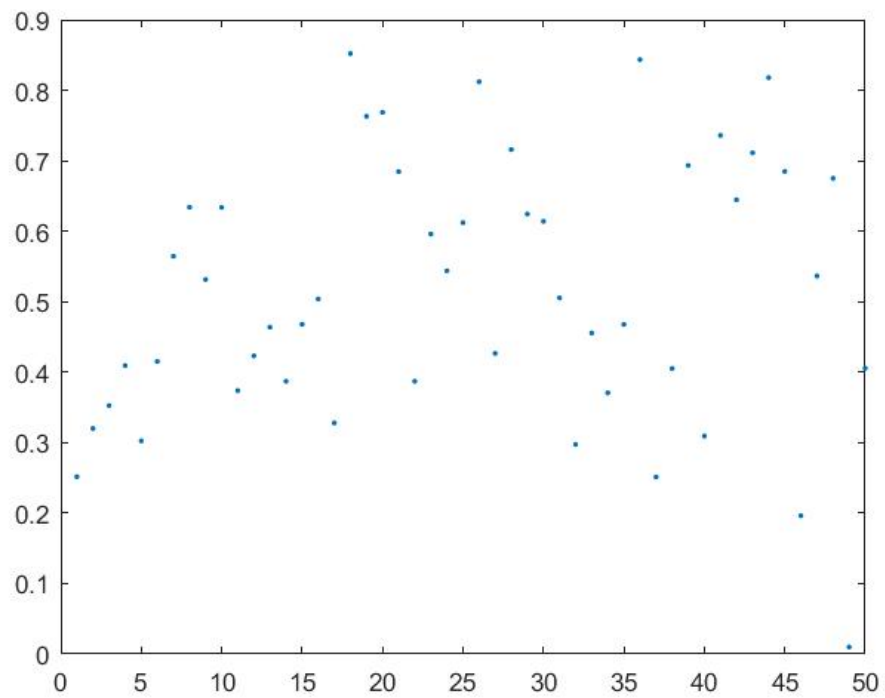


Figura 42: Tiempo de auditoría implementación con los requerimientos.

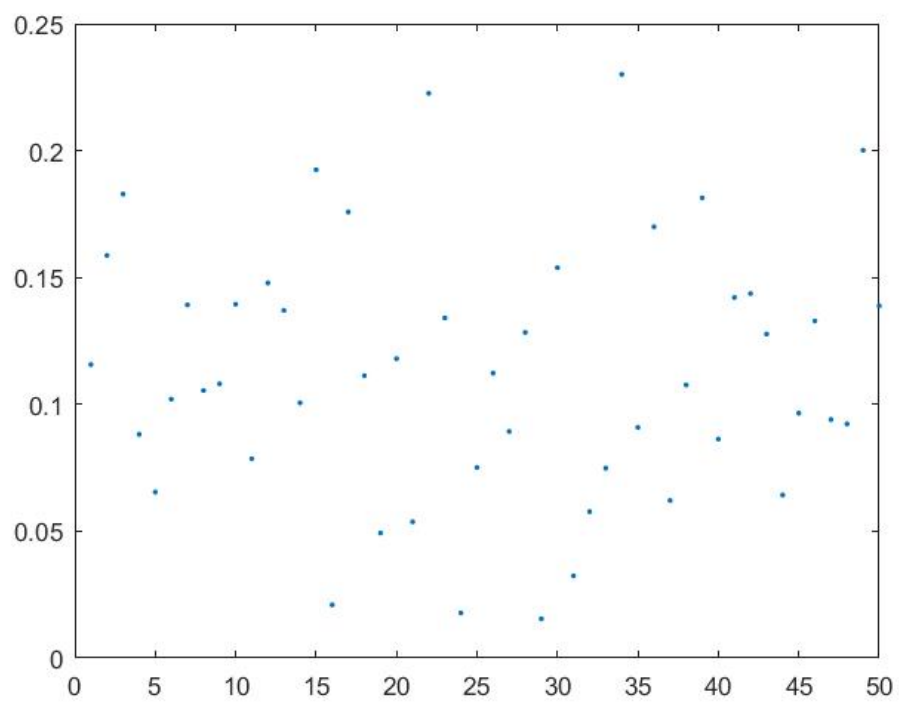


Figura 43: Tiempo de diseño.

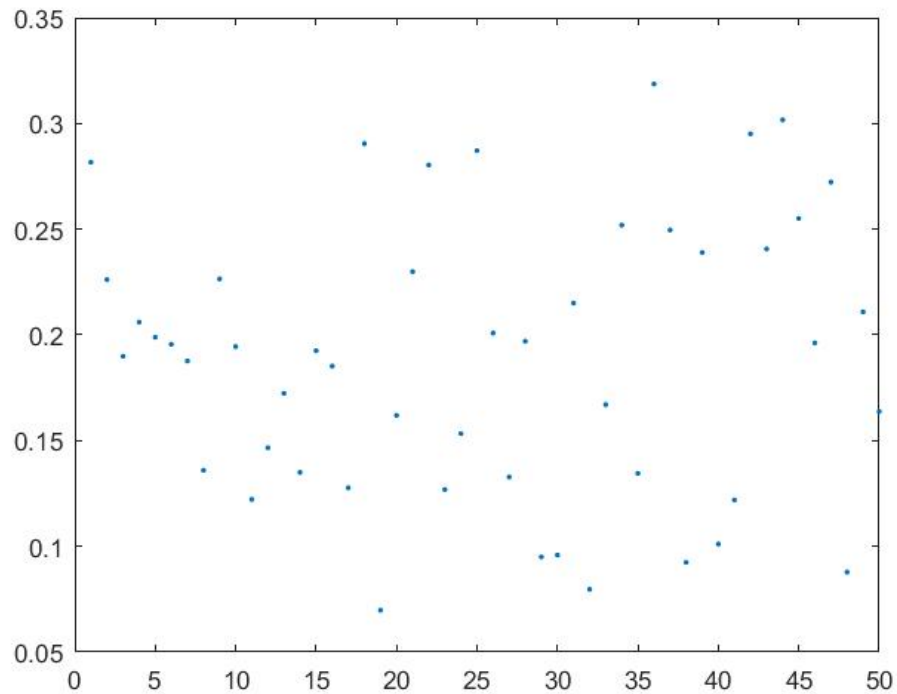


Figura 44: Tiempo de requerimientos.

3.2. Análisis de los resultados y conclusiones

A continuación se presentan los resultados obtenidos al realizar el análisis entre datos relacionados con tiempo y complejidad en proyectos pasados.

El análisis se basó en el uso de coeficiente de correlación de Pearson. Por lo que los resultados obtenidos se encuentra en el rango de valores entre -1 a +1

Correlación de:

- Tiempo de requerimientos figura 22 con numero de fallas criticos figura 13 con resultado de : -0.1406
- Tiempo de requerimientos figura 22 con numero de fallas mayores figura 14 con resultado de : -0.2406

- Tiempo de requerimientos figura 22 con numero de fallas menores figura 15 con resultado de : -0.8529
- Tiempo de requerimientos figura 22 con numero de fallas triviales figura 16 con resultado de : 0.0156
- Tiempo de diseño figura 21 con numero de fallas criticos figura 13 con resultado de : -0.1500
- Tiempo de diseño figura 21 con numero de fallas mayores figura 14 con resultado de : 0.0102
- Tiempo de diseño figura 21 con numero de fallas menores figura 15 con resultado de : -0.1526
- Tiempo de diseño figura 21 con numero de fallas triviales figura 16 con resultado de : 0.1044
- Tiempo de auditoria diseño y requerimientos figura 19 con numero de fallas criticos figura 13 con resultado de : -0.9610
- Tiempo de auditoria diseño y requerimientos figura 19 con numero de fallas mayores figura 14 con resultado de : 0.9045
- Tiempo de auditoria diseño y requerimientos figura 19 con numero de fallas menores figura 15 con resultado de : 0.9045
- Tiempo de auditoria diseño y requerimientos figura 19 con numero de fallas triviales figura 16 con resultado de : 0.9045
- Tiempo de auditoria implementación figura 20 con numero de fallas críticos figura 13 con resultado de : -0.0263
- Tiempo de auditoria implementación figura 20 con numero de fallas mayores figura 14 con resultado de : 0.1207
- Tiempo de auditoria implementación figura 20 con numero de fallas menores figura 15 con resultado de : -0.0233
- Tiempo de auditoria implementación figura 20 con numero de fallas triviales figura 16 con resultado de : -0.0508

- Numero de complejidad ciclomatica 12 con la del tiempo de auditoria de diseño figura 19 con resultado de : 0.0661
- Numero de complejidad ciclomatica 12 con la del tiempo de auditoria de implementación figura 20 con resultado de : -0.1054
- Numero de complejidad ciclomatica 12 con la de los fallos triviales¹⁶ con resultado de : 0.0222
- Numero de complejidad ciclomatica 12 con la de los fallos menores ¹⁵ con resultado de : 0.2778
- Numero de complejidad ciclomatica 12 con la de los fallos mayores¹⁵ con resultado de : 0.7771
- Numero de complejidad ciclomatica 12 con la de los fallos criticos¹⁶ con resultado de : 0.6251

3.3. Sugerencias

Dados los resultados y el análisis realizado sobre los datos recolectados, el equipo de desarrollo de la empresa *Sport Analytics* enumera las siguientes sugerencias para ser tomadas en cuenta en futuros proyectos de desarrollo de software.

1. Hay que dedicar más tiempo en la especificación de requerimientos para disminuir la cantidad de errores críticos. Esto debido a que las correlaciones son negativas. Solo los triviales son positivos.
2. El tiempo dedicado de construcción de artefactos diseño con respecto a la cantidad de errores es suficiente, ya que las relaciones son positivas, no obstante se sugiere dedicarle un poco más ya disminuir la cantidad de errores menores.
3. En auditoria de la conformidad de la implementación respecto a los requerimientos hay que dedicar más tiempo, las fallas críticas son bastante altas (bastante negativas), por ende es urgente la dedicación de

tiempo.

4. La auditoria de conformidad del diseño respecto a los requerimientos al dedicarle más tiempo no va a existir un impacto tan alto, debido a que los números son positivos o muy cercanos a 0.
5. Se recomienda incrementar el tiempo de auditoría de la implementación para reducir líneas de código y mejorar el mismo.
6. Como se suponía, de acuerdo a la complejidad ciclomatica, si aumenta, entonces aumenta los errores de todo tipo.

Se espera que con estas sugerencias y su práctica, se pueda mejorar la aplicación de procesos para el aseguramiento de la calidad del software en las etapas de gestión de proyectos.

4. Uso de otros coeficientes de correlación

Una correlación es la medida del grado de la relación entre dos o más variables, usando coeficientes según la naturaleza de las variables tenemos varios casos que se mencionan y definirán a continuación:

1. Coeficiente de Pearson: usada si ambas variables son intercalares.
2. Coeficiente de Correlación de Spearman: usada si ambas variables ordinales.
3. Coeficiente de Kendall: usada cuando las variables no siguen distribución normal.
4. Coeficiente Phi: usada si ambas variables son dicotómicas o binarias
5. Estadístico Ji-Cuadrado: usado cuando las variables son nominales.
6. Coeficiente Omega Cuadrado: usada si una variable es intervalar y otra nominal.

Para efectos de esta tarea se mencionan 2 de los más usados, y que aparecen en matlab como parte del tipo de la correlación. Para el caso de la correlación de Pearson no se explicará mucho dado que este trabajo se realizó con Pearson y esta sección busca métodos opcionales a ese método. Como comparación Pearson es más usable para distribuciones de datos normales, Kendall y Spearman es mejor en datos no normales.

4.1. Coeficiente de la correlación de Spearman (rho)

Esta es rango-base de la correlación de Pearson, es usada para medir la relación en 2 variables ordinales y aleatorias continuas, este puede ser de resultado más cercano que Pearson o Kendall. Para calcularlo se ocupan los datos ordenados y reemplazados por su respectivo orden. Y usando la siguiente expresión:

$$p = 1 - \frac{6 \sum D^2}{N(N^2 - 1)}$$

Y manejado con la siguiente fórmula:

$$r_s = \frac{\sum_{i=1}^N ((rank(x_i) - \overline{rank(x)})(rank(y_i) - \overline{rank(y)}))}{\sqrt{\sum_{i=1}^N (rank(x_i) - \overline{rank(x)})^2 \sum_{i=1}^N (rank(y_i) - \overline{rank(y)})^2}}$$

Con rank(xi) y rank(yi), rangos de observación en la muestra.

La interpretación del coeficiente de Spearman es igual que la del coeficiente de Correlación de Pearson, este oscila entre -1 y +1, indicándonos asociaciones negativas o positivas respectivamente, o cero. Una correlación de Spearman puede ser calculada con la fórmula de Pearson si antes se transformaron las puntuaciones en rangos.

Las ventajas de usar esta relación es que es muy útil cuando hay cambios en las unidades de medida, ya que no es afectada por esto, otra es que al no ser una técnica paramétrica, es libre de distribución probabilística, también se usa cuando los valores presentan valores extremos, ya que estos valores afectan a Pearson, pero como desventaja la r no puede ser utilizada para análisis de causa y efecto.

Un caso práctico, para usar Spearman es por ejemplo cuando hay 3 o más condiciones, con varios individuos siendo observados en cada condición, y prediciendo que las observaciones tendrán un orden en particular.

4.2. Coeficiente de Kendall (tau)

Esta es similar a la de Spearman, esta es diseñada para capturar asociaciones en medio de 2 variables ordinales. Puede ser usada cuando los datos no siguen distribuciones normales, y sirve para comprobar si 2 variables cuantitativas tienen una relación lineal entre sí, cuando a su vez las múltiples evaluadores evalúan las mismas muestras. Usada en análisis de concordancia de atributos.

Kendall puede variar de -1 a 1. Los valores positivos involucran asociaciones positivas, y negativas una asociación negativa, entre más alta la magnitud más alta la asociación. El coeficiente de correlación de Kendall y sus valores, p para elegir entre 2 hipótesis opuestas, con base en los datos de la muestra:

- H0: No existe asociación entre las calificaciones de todos los evaluadores y el valor estándar conocido.
- H1: Las calificaciones de todos los evaluadores están asociadas con el valor estándar conocido.

El valor p proporciona la probabilidad de obtener su muestra, si la hipótesis nula (H0) es verdadera. Si el valor p es menor que o igual a un nivel predeterminado de significancia, la hipótesis nula es rechazada y se acredita la alterna. Kendall es usable cuando los datos son ordinales y se tenga un valor conocido para cada condición.

Kendall se hace con la siguiente fórmula:

$$t = \frac{\sum_{i=1}^N \sum_{j=1}^N \text{sgn}(x_i - x_j) \text{sgn}(y_i - y_j)}{n(n-1)}$$

4.3. Correlación de Spearman y Kendall en Matlab

Para Matlab la correlación puede ser usada con [RHO, PVAL] = corr(X, Y), esto por default usa Pearson, para usar Spearman o Kendall se usa lo siguiente [RHO, PVAL] = corr(X, Y, name, value) con name= type y con value = Pearson, Kendall, Spearman. Ejemplo:

```
[RHO,PVAL] = corr(a',b','Type','Spearman');
```

5. Referencias

Referencias

- [1] Matlab documentation.(2016). corrcoef. Consultado en <http://www.mathworks.com/help/matlab/ref/corrcoef.html>
- [2] Pearson Product-Moment Correlation. (2013). Consultado en <https://statistics.laerd.com/statistical-guides/pearson-correlation-coefficient-statistical-guide.php>
- [3] Correlación de Kendall.(2015). Consultado en <https://rpro.wikispaces.com/Correlación+de+Kendall>
- [4] Uso de los estadísticos kappa y los coeficientes de Kendall.(2016). Consultado en <http://support.minitab.com/es-mx/minitab/17/topic-library/basic-statistics-and-graphs/tables/other-statistics-and-tests/using-kappa-statistics-and-kendall-s-coefficients/#what-is-kendall-s-correlation-coefficient>
- [5] Análisis de Correlación Simple, Múltiple, Parcial.(2016). Consultado en <http://www.uru.edu/fondoeditorial/libros/pdf/manualdestatistix/cap8.pdf>
- [6] Pearson's versus Spearman's and Kendall's correlation coefficients for continuos data.(2008). Consultado en http://d-scholarship.pitt.edu/8056/1/Chokns_etd2010.pdf
- [7] Coeficientes de correlacion de Pearson y Spearman.(2015). Consultado en <http://es.slideshare.net/PatriciaCastillo31/coeficiente-de-correlacion-de-pearson-y-spearman>
- [8] MathWorks Documentation corr.(2016). Consultado en <http://www.mathworks.com/help/stats/corr.html>