

# **ESTUDIO Y ANÁLISIS DE LA POTENCIA DE PRUEBA Y TAMAÑO DE MUESTRA DE LA PRUEBA B CONTRA C Y LA PRUEBA T DE DOS MUESTRAS CON SIMULACIÓN MONTECARLO**

*STUDY AND ANALYSIS OF THE TEST POWER AND SAMPLE SIZE OF TEST "B AGAINST C" AND TEST "T" OF TWO SAMPLES WITH MONTECARLO SIMULATION*

**Carla Alessia Montejo Arcos**

Tecnológico Nacional de México / IT de Celaya, México  
*alessmontejo@gmail.com*

**Manuel Darío Hernández Ripalda**

Tecnológico Nacional de México / IT de Celaya, México  
*dario.hernandez@itcelaya.edu.mx*

**José Alfredo Jiménez García**

Tecnológico Nacional de México / IT de Celaya, México  
*alfredo.jimenez@itcelaya.edu.mx*

**Recepción:** 23/octubre/2020

**Aceptación:** 30/noviembre/2020

## **Resumen**

La aplicación de técnicas estadísticas para mejora de procesos tiene vertientes. Una es la aplicación del Análisis Exploratorio de Datos (EDA) como lo utiliza el Sistema Shainin y otra es la Inferencia Estadística Clásica.

El Sistema Shainin realiza la prueba B contra C para comparar la localización de dos distribuciones suponiendo que una es mayor que la otra. Basada en una técnica no-paramétrica de EDA. En Inferencia Estadística esta comparación se hace con la prueba T de dos muestras para un extremo.

El objetivo de este documento es presentar el estudio y análisis de la potencia de prueba y tamaño de muestra de la prueba B contra C y la prueba t de dos muestras, para un extremo. El método empleado es Simulación Montecarlo de varias situaciones seleccionadas de la literatura. La conjetura a prueba es que para tamaños de muestra semejantes las dos pruebas tendrán potencia similar.

**Palabras Clave:** Monte Carlo, Potencia de Prueba, Prueba B contra C, Prueba T, Tamaño de Muestra.

## **Abstract**

*The application of statistical techniques for process improvement has differences. One is the application of Exploratory Data Analysis (EDA) as used by the Shainin System and another is the Classical Statistical Inference.*

*The Shainin System performs the B vs C test to compare the location of two distributions if one is greater than the other. Based on a non-parametric EDA technique. In Statistical Inference this comparison is made with the T test of two samples for one end.*

*The objective of this document is to present the study and analysis of the test power and sample size of the B vs C test and the two-sample t-test, for one end. The method used is the Monte Carlo Simulation of various situations selected from the literature of practical application. The conjecture being tested is that for similar sample sizes the two tests will have similar power.*

**Keywords:** *Monte Carlo, Sample Size, Test B vs. C, Test Power, Test T.*

## **1. Introducción**

El Análisis Exploratorio de Datos (EDA) es un paradigma para el análisis de datos que emplea una variedad de técnicas (principalmente gráficas) para maximizar el conocimiento de un conjunto de datos, descubrir la estructura subyacente, extraer variables importantes, detectar valores atípicos y anomalías, probar supuestos subyacentes, desarrollar modelos parsimoniosos, y determinar la configuración óptima del factor. EDA es un enfoque para el análisis de datos que pospone los supuestos habituales sobre qué tipo de modelo siguen los datos con el enfoque más directo de permitir que los datos revelen su estructura y modelo subyacentes [Tukey J. W., 1977]. La aplicación en la industria es el llamado Sistema Shainin o Metodología Shainin X roja.

En este artículo se realiza la comparación entre el paradigma clásico y el exploratorio. Para las técnicas de análisis, estos dos paradigmas son similares en el sentido de que todos comienzan y arrojan conclusiones con un problema general de ciencia o ingeniería. La diferencia es la secuencia y el enfoque de los pasos intermedios.

Para el análisis clásico, la secuencia es:

- Problema → Datos → Modelo → Análisis → Conclusiones
- Para EDA, la secuencia es
- Problema → Datos → Análisis → Modelo → Conclusiones

## **B contra C**

B vs C es un plan experimental simple que se usa para comparar dos tratamientos o condiciones de proceso representados por las letras B y C. En la aplicación de validación, las letras B y C denotan las condiciones "mejores" (Best) y "actuales" (Current). Un uso en el Sistema Shainin [Kosina, 2015], [Steiner, MacKay, & Ramberg, 2007], es verificar que una causa identificada es dominante después de que otras herramientas de generación de pistas hayan conducido a un solo sospechoso. Un segundo uso es validar una solución cuando el objetivo es cambiar el centro de proceso o reducir una tasa de defectos.

En el plan recomendado más simple, se producen tres unidades bajo el tratamiento B y tres bajo el tratamiento C. Para [Bhote & Bhote, 2000] se llama la "prueba six-pack". Los niveles de la causa dominante sospechosa para las corridas B y C se seleccionan en los extremos de la variación del sospechoso en la producción normal. El orden de las seis corridas es aleatorio.

## **Prueba T de Student para 2 muestras independientes**

Una prueba T de dos muestras para un extremo es un análisis que se usa para comparar si dos muestras tienen la misma media o una es mayor que otra, es una técnica del paradigma clásico de la estadística inferencial, supone que las poblaciones que generan las muestras se distribuyen con distribución normal, como la aplicación de esta prueba se estudia en las aplicaciones de mejora de procesos, se supondrá además que las diferencias de medias se pueden deber a cambios de un proceso original y que por lo tanto tienen la misma dispersión medida como varianza. La prueba consiste en la extracción de dos muestras de tamaño similar, ( $n = n_1 = n_2$ ), a cada una se le calcula el promedio aritmético para estimar la media del proceso, y la desviación estándar muestral para estimar la desviación estándar

de la población [Anderson, Sweeney, & Williams, 2011], [Minitab 17 Statistical Software, 2010].

### **Simulación Montecarlo de pruebas**

La simulación Montecarlo es una técnica basada en el muestreo sistemático de variables aleatorias. Abarcando técnicas que permiten obtener soluciones de problemas matemáticos o físicos por medio de pruebas aleatorias repetidas. Con este tipo de simulación se agrupan una serie de procedimientos que analizan distribuciones de variables aleatorias usando simulación de números aleatorios. Permite dar solución a una gran variedad de problemas matemáticos haciendo experimentos con muestreos estadísticos en una computadora. La simulación Montecarlo es utilizada para cualquier esquema que emplee números aleatorios, usando variables aleatorias con distribuciones de probabilidad conocidas, de esta manera se resuelven ciertos problemas estocásticos y determinísticos donde el tiempo no juega un papel importante [Kroese, Taimre, & Botev, 2013], [Montilla & Kromrey, 2010].

## **2. Método**

El método utilizado en este trabajo consta de 4 pasos las cuales se muestran en la figura 1:

- Paso 1: Generar 1,000 pares de muestras con una diferencia de medias conocida y una desviación estándar determinada, estas parejas se hacen siguiendo una distribución de probabilidad.
- Paso 2: La prueba consiste en la extracción de dos muestras de tamaño similar, ( $n = n_1 = n_2$ ), a cada una se le calcula el promedio aritmético para estimar la media del proceso, y la desviación estándar muestral para estimar la desviación estándar de la población. Se pone como hipótesis nula que las medias de los procesos son iguales y por lo tanto su diferencia será cero, la hipótesis alternativa es que una de las medias es diferente en una medida significativa y por lo tanto la diferencia entre las dos será bastante diferente de cero. Si el estadístico cae en una región de rechazo, de tamaño alfa ( $\alpha$ )

dentro de la distribución se rechaza la hipótesis nula, de lo contrario se declara que no hay suficiente evidencia para poder rechazar la hipótesis nula. Otra forma es calcular el área de probabilidad que limita el estadístico calculado y si este es menor que alfa, se rechaza la hipótesis nula, este es el procedimiento que casi todos los programas de cómputo estadístico adoptan reportando el valor calculado como valor P. Este procedimiento se tomó de [Anderson, Sweeney, & Williams, 2011].

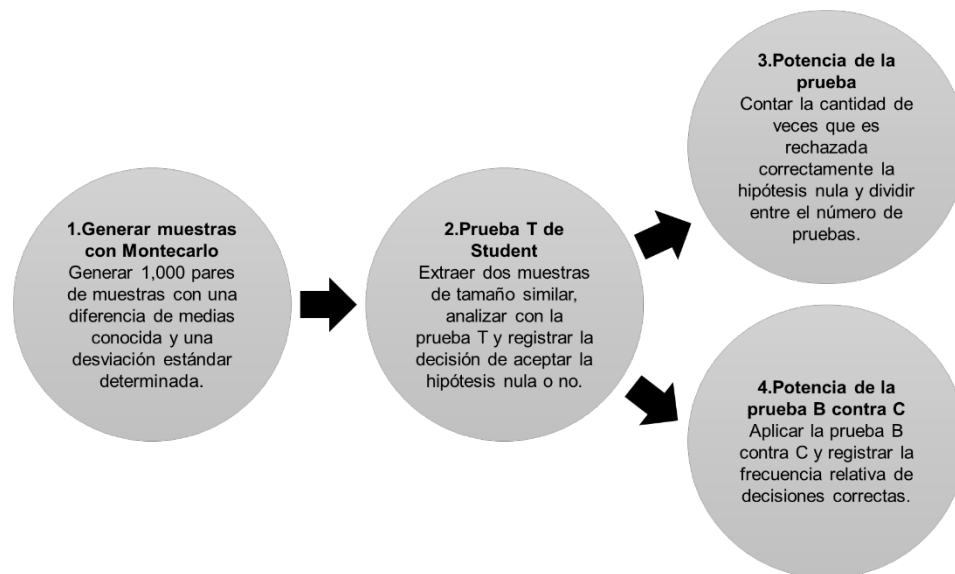


Figura 1 Pasos del método utilizado.

Alfa ( $\alpha$ ) en prueba de hipótesis es la probabilidad de rechazar una hipótesis nula cuando esta es cierta, el valor de alfa se fija al inicio de la prueba de hipótesis, también se considera la probabilidad de quedarse con una hipótesis nula falsa, la probabilidad de ese error se denomina beta ( $\beta$ ), y depende del tamaño de la muestra empleada, pues al aumentar la muestra la curva de la distribución T se hace más estrecha y la región complementaria a la región de rechazo disminuye. Para análisis de una prueba estadística se usa el complemento de beta,  $(1-\beta)$ , denominado “potencia de prueba”. Técnicamente, la potencia de prueba es la probabilidad de rechazar la hipótesis nula cuando la hipótesis alternativa específica es verdadera. Esto es, una decisión correcta, se muestra en la figura 2.

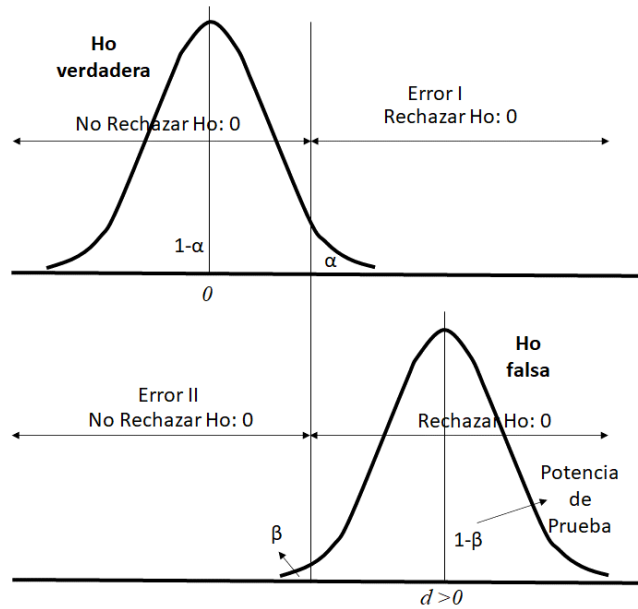


Figura 2 Relación entre hipótesis nula Ho verdadera y falsa y valores de alfa, beta y potencia de prueba.

Para una prueba de la diferencia de dos medias de muestras independientes, suponiendo que las desviaciones estándar son desconocidas se estima un estadístico t con la fórmula de la ecuación 1.

$$t = \frac{(\bar{x}_1 - \bar{x}_2)}{\sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}}} \quad (1)$$

Donde t es el estadístico estimado,  $\bar{x}_1$  es el promedio de la primera muestra,  $\bar{x}_2$  es el promedio de la segunda muestra,  $s_1^2$  es la varianza estimada de la primera muestra,  $n_1$  es el número de datos de la primera muestra,  $s_2^2$  es la varianza estimada de la segunda muestra,  $n_2$  es el número de datos de la segunda muestra. El estadístico se supone que tiene una distribución t de Student, con un número de grados de libertad, estos se estiman redondeando el resultado de la ecuación 2.

$$gl = \left\lfloor \frac{\left(\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}\right)^2}{\frac{1}{n_1 - 1} \left(\frac{s_1^2}{n_1}\right)^2 + \frac{1}{n_2 - 1} \left(\frac{s_2^2}{n_2}\right)^2} \right\rfloor \quad (2)$$

Donde:  $g$  son los grados de libertad estimados redondeando hacia el entero inferior más cercano la expresión (entre corchetes de piso),  $s_1^2$  es la varianza estimada de la primera muestra,  $n_1$  es el número de datos de la primera muestra,  $s_2^2$  es la varianza estimada de la segunda muestra,  $n_2$  es el número de datos de la segunda muestra [Anderson, Sweeney, & Williams, 2011].

- Paso 3: Se cuenta la cantidad de veces que se rechaza correctamente la hipótesis nula y se divide entre el número de pruebas, para estimar la frecuencia relativa de rechazo correcto, esto como un estimador de la potencia de la prueba.
- Paso 4: Simultáneamente con el paso 3 cada par de muestras también se le aplica la prueba B contra C y se registra la frecuencia relativa de decisiones correctas, para estimar la potencia de prueba de B contra C. Así se tienen resultados comparables para ambas técnicas.

El análisis recomendado se basa en el esquema de conteo final. Solo si los valores de salida para las tres corridas B y las tres corridas C se separan en la dirección esperada, hemos verificado la causa dominante.

[Tukey J. W., 1959] creó esta prueba como una prueba de hipótesis unilateral; ningún cambio versus cambio está en una dirección especificada. La figura 3 muestra algunas situaciones. En caso 1, C y B están encimadas, no hay extremos libres, no se puede distinguir un proceso de otro, en caso 2 una ligera separación, genera extremos libres, en caso 3 los extremos libres son grandes, la separación es evidente.

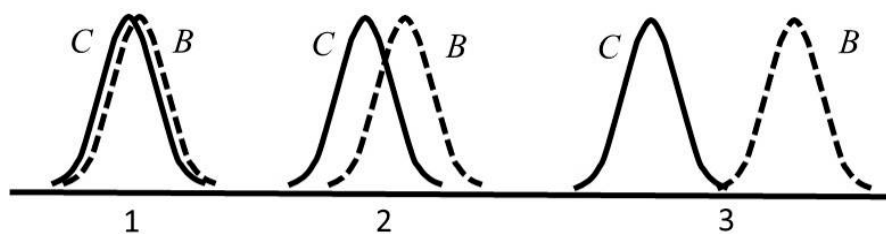


Figura 3 Situaciones de una prueba B contra C

[Sleeper, 2006] resume el procedimiento de Tukey de la siguiente manera:

- Objetivo: ¿Tiene (proceso A) una distribución diferente de (proceso B)?

- Suposiciones:
  - ✓ 0: cada muestra es una muestra aleatoria de observaciones mutuamente independientes del proceso de interés.
  - ✓ 1: Cada proceso es estable.
- Estadística de prueba: calcule el recuento final, que es la suma del recuento de observaciones en la muestra A que son menores que todas las observaciones en la muestra B, más el recuento de observaciones en la muestra B que son mayores que todas las observaciones en la muestra A.

Se considera que la cantidad de 1,000 réplicas es suficiente para poder tener separados los comportamientos de las pruebas, para esto se llevó también una estimación del error estándar de las corridas de simulación y se estimaron intervalos de confianza para las estimaciones de potencia de prueba quedando bastante separadas y claramente mostrado el comportamiento de cada prueba.

En la literatura el estudio de referencia para la comparación de la prueba T de dos medias y la prueba B contra C es el trabajo de [Neave & Granger, 1968], sin embargo, la simulación se limitó a corridas de 500 parejas de muestras con distribuciones normales, suponiendo varianzas iguales y conocidas, y con tamaños de muestras en las parejas de tamaño 20. También uso para el caso de comportamiento no normal una distribución artificial bimodal combinada por dos normales, justificada por haber sido usada antes en la literatura. Parece que un motivo fuerte de su investigación fue mostrar la superioridad de una modificación que le hizo a la prueba original de Tukey [Neave H. , 1966], [Nelson, 1975]. Sin embargo, este procedimiento no prosperó. La prueba B contra C es el procedimiento original de Tukey, y se ha difundido por su adopción en el Sistema Shainin.

Hay un punto adicional de interés, B contra C, está en la base de varios análisis del Sistema Shainin, en especial lo que se llama ventaja o apalancamiento de análisis, donde se toman las tres piezas más malas de un proceso y las tres piezas más buenas para iniciar otros análisis. Esto implica que deben cubrir el requisito mínimo de que las piezas no deben traslapar en sus valores. Esto es el criterio de la prueba de Tukey en un extremo,  $n_1=n_2=3$ , y el criterio de cuenta de extremos es igual a 6.



Es pues, este caso extremo de primordial interés para entender que pasa cuando se aplica el apalancamiento o ventaja de análisis.

### **3. Resultados**

Se generaron 1,000 pares de muestras con varios tamaños de pruebas y con distribuciones de probabilidad normal con desviación estándar igual a 1, tomando como base de comparación la normal estándar con media igual a cero.

La distancia entre medias se expresó estandarizada, usando la práctica de los algoritmos disponibles en software para estimar potencia de prueba [Faul, Erdfelder, Buchner, & Lang, 2009], así para el caso de las pruebas con distribución normal se calculó “d” como la diferencia entre medias dividida entre la raíz cuadrada de la suma de las varianzas dividida entre dos, lo cual resultaba por elegir varianzas igual a uno en simplemente la diferencia entre medias; se mantuvo así para poder comparar con la potencia que teóricamente puede tener la prueba T para dos medias independientes como lo estiman programas disponibles. El tamaño de muestra de 1,000 pares mostró ser bastante suficiente para mostrar la separación entre las respuestas de los experimentos. Se tuvo una precisión aceptable.

#### **Primer experimento**

El primer experimento consistió en utilizar distribuciones normales con varianza 1 con  $n_1=n_2=3$ , usando como criterio en B contra C, 6 objetos sin traslape, lo que proporciona una alfa de prueba de 5%, la prueba T se realizó usando el procedimiento para varianzas desconocidas y diferentes como lo describe [Anderson, Sweeney, & Williams, 2011] que además es el procedimiento previsto en el comando “t.test()” del Lenguaje de programación estadístico R [R Core Team, 2020] y en Minitab [Minitab 17 Statistical Software, 2010]. Se calculo la potencia de prueba con el programa G\*Power [Faul, Erdfelder, Buchner, & Lang, 2009], para tener una comparación teórica, recordando que el procedimiento para estos valores supone varianzas iguales y conocidas.

El primer experimento con distribuciones normales  $n_1=n_2=3$ , se presenta en tabla 1, mostrando resultados con los valores de distancia estandarizada de 0.25 a 4.5.

Tabla 1 Valores de potencia de prueba estimados para la diferencia de medias,  $n_1=n_2=3$ .

Diferencia "d"	Teoría	T	BvsC
0.25	0.083	0.071	0.084
0.50	0.129	0.080	0.113
0.75	0.191	0.140	0.183
1.00	0.267	0.185	0.266
1.25	0.356	0.248	0.334
1.50	0.452	0.356	0.458
1.75	0.550	0.433	0.557
2.00	0.645	0.546	0.653
2.25	0.731	0.600	0.712
2.50	0.805	0.676	0.773
2.75	0.864	0.767	0.846
3.00	0.910	0.834	0.899
3.25	0.943	0.860	0.922
3.50	0.965	0.913	0.949
3.75	0.980	0.929	0.966
4.00	0.989	0.963	0.983
4.25	0.994	0.975	0.987
4.50	0.997	0.978	0.991

Los valores calculados con un procedimiento analítico teórico en "Teoría", "T" muestra las estimaciones de una prueba T con varianzas desconocidas y no iguales, "BvsC" muestra las estimaciones para una prueba Tukey con 6 elementos sin traslape. La columna teoría son los valores calculados con el software G\*Power, la columna T muestra la frecuencia relativa obtenida de rechazos de la hipótesis nula con diferencia de medias igual a cero, por lo tanto, una estimación de la potencia de prueba, en la columna BvsC se muestra la potencia de prueba estimada a partir de la frecuencia relativa observada en la corrida de simulación con 1,000 pares de muestras.

### Segundo experimento

El segundo experimento consistió en probar otro extremo de la prueba cuando los tamaños de la muestra no son iguales, utilizando las mismas condiciones de distribución normal, y tomando como base de comparación una distribución normal con media cero y desviación estándar uno, se tomó la muestra con la media cero de tamaño  $n_1=30$ , y la muestra con media mayor con  $n_2=9$ . El procedimiento de la

prueba T no cambió. En la prueba B contra C se siguió la recomendación de [Tukey J. W., 1959] para muestras de tamaños diferentes y el valor de cuenta para valores extremos sin traslape se cambió de 7 a 10 para una alfa de 5%. Este experimento refleja una condición común en manufactura, un proceso existente del que se dispone muchas piezas, que se monitorean constantemente contra un nuevo proceso del que se disponen pocas piezas y no se tiene una línea base anterior. Los resultados del segundo experimento de la tabla 2 con una muestra de tamaño  $n_1=30$  para una distribución normal con media de cero, y la muestra de una distribución normal con media mayor con tamaño  $n_2=9$ . Se corrió el experimento en 1,000 pares de muestras en simulación Montecarlo y se calculó para comparación el valor teórico a partir del software G\*Power. El intervalo calculado abarco un rango de distancias entre medias de 0.25 a 4.0.

Tabla 2 Valores estimados de Potencia de prueba para pruebas con muestras de distribuciones normales de varianza igual a 1.

<b>d</b>	<b>Teoría</b>	<b>T</b>	<b>BvsC</b>
0.25	0.159	0.163	0.070
0.50	0.362	0.325	0.145
0.75	0.615	0.601	0.300
1.00	0.826	0.785	0.456
1.25	0.943	0.927	0.615
1.50	0.987	0.982	0.793
1.75	0.998	0.993	0.880
2.00	1.000	1.000	0.934
2.25	1.000	1.000	0.976
2.50	1.000	1.000	0.994
2.75	1.000	1.000	0.997
3.00	1.000	1.000	0.996
3.25	1.000	1.000	1.000
3.50	1.000	1.000	1.000
3.75	1.000	1.000	1.000
4.00	1.000	1.000	1.000

### Tercer experimento

En el tercer experimento se usaron distribuciones normales con varianza 1 y se tomó como media de referencia el valor cero, las muestras fueron  $n_1=n_2=15$ , lo que supone una situación más favorable en cuanto a cantidad de información y la

simetría de los tamaños de las dos muestras debe favorecer a la prueba B contra C, el cálculo teórico se hizo con G\*Power. El rango de estimaciones corrió para valores de diferencia entre 0.25 y 4.0.

Con distribuciones normales de varianza 1 y como media de referencia el valor cero, los tamaños de muestras fueron  $n_1=n_2=15$ , el cálculo teórico se hizo con G\*Power. El rango de diferencia fue entre 0.25 y 4.00. La tabla 3 muestra los resultados que se obtuvieron con la simulación de 1,000 pares de muestras probadas comparando variables normales con varianza 1 y diferencia de medias desde 0.25 hasta 4.0.

Tabla 3 Resultados de 1,000 pruebas simuladas para comparar la localización de muestras de tamaño 15.

d	teoría	T	BvsC	d	teoría	T	BvsC
0.25	0.164	0.198	0.054	2.25	1.000	1.000	0.993
0.50	0.379	0.401	0.143	2.50	1.000	1.000	0.991
0.75	0.640	0.636	0.267	2.75	1.000	1.000	1.000
1.00	0.848	0.826	0.465	3.00	1.000	1.000	1.000
1.25	0.955	0.949	0.690	3.25	1.000	1.000	1.000
1.50	0.991	0.992	0.826	3.50	1.000	1.000	1.000
1.75	0.999	0.999	0.923	3.75	1.000	1.000	1.000
2.00	1.000	1.000	0.953	4.00	1.000	1.000	1.000

#### 4. Discusión

La prueba con  $n_1=n_2=3$  y distribución normal se realizó para revisar un criterio utilizado en el Sistema Shainin que inicia el procedimiento de búsqueda de variables a partir de 3 piezas muy malas y 3 piezas muy buenas, la potencia estimada señala que se necesita tener una distancia estandarizada de medias mayor a 2.5 para obtener una potencia de al menos 80%. Esto implica que para poder obtener piezas lo bastante separadas, deben estar a 2.5 de distancia estandarizada, si las piezas “buenas” están en la media cero, las “malas” deberán estar más allá de 2.5 desviaciones estándar de distancia, para garantizar encontrar la diferencia, la probabilidad de estar a 2.5 de distancia o más en una normal estándar es 0.00621  $\approx 1/161.39$ . Esto implicaría que en promedio cada 162 piezas producidas se podría encontrar una pieza lo suficientemente “mala” y para encontrar 3 lo suficientemente “malas”, en promedio, se tendrían que producir al menos 486 piezas.

El segundo experimento muestra una mayor potencia de prueba para la prueba T, la prueba B contra C está recomendada para casos cercanos de la simetría en la proporción de los tamaños de muestra, 4:3, por ejemplo, aún con la corrección sugerida no se ve un desempeño que supere a la prueba T.

En el tercer experimento se corrió con valores de muestra normales iguales  $n_1=n_2=15$ , la potencia de prueba de la prueba T domino a la de la prueba B contra C. Más tamaño de muestra le dio ventaja a la prueba T, bajo estas condiciones. Parece que la prueba B contra C gana en condiciones de tamaños de muestra más pequeños e iguales. Los valores teóricos se consideran sólo como una referencia de lo “mejor” que se puede aspirar a lograr para la condición de distancia dada, recordando que los valores suponen condiciones ideales, especialmente respecto a las varianzas.

## **5. Conclusiones**

El presente trabajo se puede ver como una continuación del antiguo trabajo de [Neave & Granger, 1968] pero con la diferencia que en su tiempo no se consideró la existencia del Sistema Shainin de mejora de procesos y la difusión que le daría a la llamada ahora prueba rápida de Tukey-Duckworth o B contra C. Para el experimento con tamaños de muestra diferentes, 30 y 9, la potencia de la prueba T domina desde valores bajos de diferencia “d” hasta 2.5. La prueba B contra C no parece muy favorable cuando hay asimetría en los tamaños de muestra, mostrados en la figura 4.

Para la prueba con muestras de tamaño 3, se encuentra que la potencia de prueba para cualquiera de las dos pruebas empieza a tener un valor de más de 80% más allá de una distancia estandarizada de 2.5, esto implica en una estimación rápida que, si se quiere encontrar un trío de piezas malas para comparar con piezas buenas, se tienen que observar en promedio al menos 486 piezas producidas. Esto merece más investigación por que puede dar un arranque para establecer la muestra necesaria para una establecer una línea base.

En general la prueba B contra C parece recomendable en situaciones donde la velocidad de una decisión es importante y los tamaños de muestra son reducidos y

sin forma de aumentarlos. El segundo experimento señala que es preferible tener tamaños de muestra iguales si se quiere usar la prueba B contra C. Sin embargo, cuando se corre el tercer experimento con tamaños iguales = 15, resulta que la potencia de prueba de la T domina a la de B contra C, esto sólo deja un espacio a la prueba B contra C para pruebas de muestras pequeñas y simétricas.

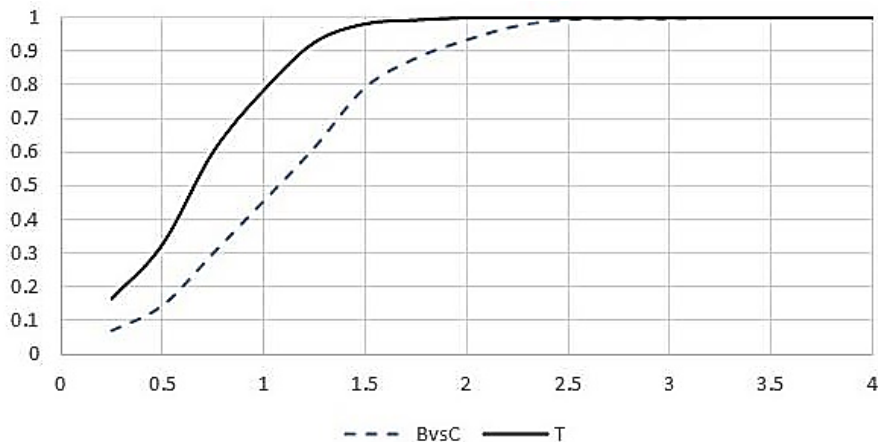


Figura 4 Potencia de prueba  $n_1=30$ ,  $n_2=9$ .

## 6. Bibliografía y Referencias

- [1] Anderson, D. R., Sweeney, D. J., & Williams, T. A. 2011. Essentials of statistics for business and economics, revised. Cengage Learning.
- [2] Bhote, K. R., & Bhote, A. K. 2000. World class quality: using design of experiments to make it happen. New York: AMACOM.
- [3] Faul, F., Erdfelder, E., Buchner, A., & Lang, A. G. 2009. Statistical power analyses using G\*Power 3.1: Tests for correlation and regression analyses. Behavior Research Methods, 41(4), 1149-1160.
- [4] Kosina, J. 2015. Shainin methodology: an alternative or an effective complement to six sigma? Quality Innovation Prosperity, 19(2), 18-31. doi:10.12776/QIP.V19I2.580
- [5] Kroese, D. P., Taimre, T., & Botev, Z. I. 2013. Handbook of monte carlo methods (Vol. 706). John Wiley & Sons.
- [6] Minitab 17 Statistical Software. 2010. Computer software. Retrieved from State College, PA: Minitab, Inc: www.minitab.com

- [7] Montilla, J., Kromrey, J. 2010. Robustez de las pruebas T en comparación de medias, ante violación de supuestos de normalidad y homocedasticidad. *Revista ciencia e ingeniería*. 31(2), 101-108.
- [8] Neave, H. 1966. A development of Tukey's quick test of location. *Journal of the American Statistical Association*, 61(316), 949-964.
- [9] Neave, H. R., & Granger, C. W. 1968. A Monte Carlo study comparing various two-sample tests for differences in mean. *Technometrics*, 10(3), 509-522.
- [10] Nelson, L. S. 1975. A Quick and Powerful Two-Sample Test. *Journal of Quality Technology*, 7(3), 150-152.
- [11] R Core Team. 2020. R: A language and environment for statistical computing. Retrieved from R Foundation for Statistical Computing. R Foundation for Statistical Computing: <https://www.R-project.org/>.
- [12] Sleeper, A. D. 2006. *Design for Six Sigma Statistics*. McGraw-Hill.
- [13] Steiner, S. H., MacKay, R. J., & Ramberg, J. S. 2007. An overview of the Shainin System™ for quality improvement. *Quality Engineering*, 20(1), 6-19.
- [14] Tukey, J. W. 1959. A quick, compact, two-sample test to Duckworth's specifications. *Technometrics*, 1(1), 31-48.
- [15] Tukey, J. W. 1977. *Exploratory Data Analysis*. Addison-Wesley.