



**Aplicación de metodologías estadísticas clásicas e innovadoras
asociadas a la elaboración de informes de validación en
medicamentos de liberación prolongada, en Laboratorio
Andrómaco.**

Roberto Felipe Ureta Díaz

Trabajo de titulación para optar al título de Ingeniero Estadístico

Profesor guía: Dra. María Pilar Sánchez Olavarria

Co-guía: QF. Claudio Godoy Cortes

Enero, 2019



**Aplicación de metodologías estadísticas clásicas e innovadoras
asociadas a la elaboración de informes de validación en
medicamentos de liberación prolongada, en Laboratorio
Andrómaco.**

Roberto Felipe Ureta Díaz

Enero, 2019

Agradecimientos

A mis padres Ximena y Roberto, por haberme forjado como la persona que soy en la actualidad; muchos de mis logros se los debo a ustedes entre los que se incluye este. Me formaron con reglas y con algunas libertades, pero al final, me apoyaron constantemente en esto.

A mis hermanas Ximena y Pamela, que al igual que mis padres siempre se preocuparon y apoyaron por mi bienestar, sus palabras de liento y correctas en cada momento difícil.

A mi profesora María Pilar Sánchez por haberme enseñado muchas cosas que hoy en día puedo aplicar como profesional, donde estoy ahora es gracias a usted. Por su paciencia con mi persona en momentos en los que no tuve la madurez. A usted profesora le debo mucho y la recordaré por siempre.

A mi polola Natalia, por apoyarme en los últimos momentos de este trabajo de tesis, sus palabras de apoyo, su amor incondicional, sus "vamos que se puede".

Finalmente a la Universidad, por darme las herramientas para ser un profesional, enseñarme algunos valores que puedo llevar a cabo.

Gracias totales!!

Al fin...

Resumen

En este proyecto de titulación se presenta una problemática en el área de validación de procesos, asociados a la industria farmacéutica. Hoy en día existe un protocolo y una guía técnica elaborada por el Instituto de Salud Pública (ISP), específicamente por la Agencia Nacional del Medicamento (ANAMED), que entrega las metodologías estadísticas que se deben aplicar para la aprobación del proceso de validación de la fabricación de un medicamento oral sólido.

Dentro del proceso de validación de procesos se consideran algunas metodologías estadísticas, cuyo proceso de aplicación representa puntos críticos dentro del análisis: condición de normalidad de la variable evaluada, la transformación de variables, el cálculo del índice de capacidad de proceso y la detección y eliminación de observaciones atípicas (outliers).

Por lo anterior, se justifica para este proyecto de tesis aplicar algunos métodos estadísticos innovadores, que actualmente no se aplican en la validación del medicamento, que permitan resolver en forma adecuada estos puntos críticos, en el marco de la regulación ya existente exigida por la Agencia Nacional del Medicamento (ANAMED).

Por último, el tipo de medicamento con el que se trabajará en esta tesis refiere a aquellos con formas farmacéuticas de liberación prolongada.

Abreviaciones

ISP: Instituto de Salud Pública.

ANAMED: Agencia Nacional del Medicamento.

EMA: European Medicines Agency.

FDA: Food and Drug Administration.

GMP: Good Manufacturing Practice.

INVIMA: Instituto Nacional de Vigilancia de Medicamentos.

FFLM: Formas Farmacéuticas de Liberación Modificada.

LIE: Límite Inferior de Especificación.

LSE: Límite Superior de Especificación.

LIC: Límite Inferior de Control.

LSC: Límite Superior de Control.

Índice

1	Introducción	10
1.1	Objetivos	14
1.1.1	General	14
1.1.2	Específicos	14
1.2	Regulación actual según Agencia Nacional de Medicamento ANAMED	15
2	Materiales y métodos	16
2.1	Estadística descriptiva	17
2.2	Pruebas de normalidad	20
2.2.1	Prueba de Shapiro-Wilk	20
2.2.2	Prueba de Kolmogorov-Smirnov	21
2.2.3	Prueba de Anderson-Darling	22
2.2.4	Prueba de Lilliefors	23
2.2.5	Prueba de Cramer-von Mises	23
2.3	Capacidad del proceso productivo	24
2.3.1	Índice de Capacidad Cp	25
2.3.2	Índice de Capacidad Cpk	25
2.4	Transformación de variables	26
2.5	Datos atípicos	28
2.5.1	Prueba de Grubb's	29
2.5.2	Prueba de Dixon	30
2.6	Métodos gráficos	31
3	Resultados	36
3.1	Análisis	37
3.1.1	Estadística descriptiva	37
3.1.2	Normalidad	38
3.1.3	Estudio gráfico	39
3.1.4	Transformación de variables	49
3.1.5	Prueba para detección de valores atípicos	50

3.1.6	Capacidad de Proceso	51
3.1.7	Análisis de varianza ANOVA de una vía	52
3.1.8	Conclusión	54
4	Anexos	56
4.1	Códigos Stata Corp	56
4.2	Códigos R-project	68
5	Bibliografía	71

Índice de figuras

Figura 2.6.1 Histograma para variable de calidad Peso	32
Figura 2.6.2 Gráfico Cuantil-Cuantil para variable de calidad Peso	32
Figura 2.6.3 Gráfico de aceptación para variable de calidad Peso	34
Figura 2.6.4 Gráfico de Caja para variable de calidad Peso	35
Figura 3.1.1 Histograma para variable de calidad Peso Lote 1	40
Figura 3.1.2 Histograma para variable de calidad Peso Lote 2	41
Figura 3.1.3 Histograma para variable de calidad Peso Lote 3	41
Figura 3.1.4 Gráfica Cuantil-Cuantil para variable de calidad Peso Lote 1	42
Figura 3.1.5 Gráfica Cuantil-Cuantil para variable de calidad Peso Lote 2	43
Figura 3.1.6 Gráfica Cuantil-Cuantil para variable de calidad Peso Lote 3	43
Figura 3.1.7 Gráfico de Caja para variable de calidad Peso Lote 1	45
Figura 3.1.8 Gráfico de Caja para variable de calidad Peso Lote 2	45
Figura 3.1.9 Gráfico de Caja para variable de calidad Peso Lote 3	46
Figura 3.1.10 Gráfica de Control para variable de calidad Peso Lote 1	47
Figura 3.1.11 Gráfica de Control para variable de calidad Peso Lote 2	48
Figura 3.1.12 Gráfica de Control para variable de calidad Peso Lote 3	48
Figura 3.1.13 Gráfica de Control para variable de calidad Peso, 3 lotes	52

Índice de tablas

Tabla 2.1.1	Análisis descriptivo según formulario.	17
Tabla 3.1.1	Análisis descriptivo de peso.	37
Tabla 3.1.2	Valor-p prueba de normalidad de variable de calidad Peso	39
Tabla 3.1.3	Valor-p para prueba de normalidad Shapiro-Wilk de variables transformadas	49
Tabla 3.1.4	Valores detectados como atípicos mediante test de Grubb's	50
Tabla 3.1.5	Valor-p para prueba de Grubb's	50
Tabla 3.1.6	Valor-p para prueba de normalidad Shapiro-Wilk	51
Tabla 3.1.7	Cálculo Índices de Capacidad de Proceso	51
Tabla 3.1.8	Análisis de varianza para variable de calidad Peso	53
Tabla 3.1.9	Análisis mediante prueba de Bartlett para variable de calidad Peso	53

Capítulo 1

Introducción

Hoy en día, todo tipo de empresa farmacéutica desarrolla diversos productos para el consumo o utilización de los fármacos, los cuales deben pasar por un proceso de prueba. Esto, para corroborar que el producto que se ofrece cumpla con todas las exigencias solicitadas por su ente regulador, entendiendo que uno de los principales pilares de las industrias de los fármacos es la validación.

Según el *D.S.N°3/2010*, un producto farmacéutico o medicamento es cualquier sustancia, natural o sintética, o mezcla de ellas, que se destine al ser humano con fines de curación, atenuación, tratamiento, prevención o diagnóstico de las enfermedades o sus síntomas, para modificar sistemas fisiológicos o el estado mental en beneficio de la persona a quien le es administrado (Decreto Supremo *N3*, 2010).

De acuerdo a lo anterior, según la Food and Drug Administration (FDA, USA) cada producto farmacéutico debe ser diseñado y concebido integralmente bajo los conceptos clásicos de calidad, seguridad y eficacia, asegurando que el producto que se está elaborando posea una solidez tal que sea 100 % confiable hacia los pacientes.

Es por lo anterior, que adquiere relevancia la comprensión del término de validación. Éste se considera como un proceso intermedio entre la fabricación y comercialización de un medicamento, permitiendo controlar todos aquellos procedimientos necesarios para la fabricación del medicamento, y verificar que cada una de sus etapas se repita de la misma manera.

A nivel nacional, el campo de la producción de medicamentos está normado por el *D.S. N°3/2010*.

Reglamento del Sistema Nacional de Control de Productos Farmacéuticos de uso humano, el cual exige a los productores de medicamentos seguir el Manual de las Buenas Prácticas de Manufactura o Good Manufacturing Practice (GMP) propuestas por la Organización Mundial de la Salud (OMS), como lineamientos básicos de los procedimientos destinados a garantizar la calidad uniforme y satisfactoria de los productos farmacéuticos, dentro de los límites aceptados y vigentes para cada uno de ellos. (Decreto Supremo N°3, 2010).

Hoy en día, no sólo la OMS sino que todas las agencias normativas sanitarias importantes a nivel mundial, tales como la FDA y la Agencia Europea del Medicamento o European Medicines Agency (EMA), han incluido anexos exclusivos a sus (GMP) para el desarrollo del proceso de Validación, en donde se trata como un apéndice la validación de procesos no estériles.

Para el D.S. N°3/2010, la validación se entiende como la acción documentada efectuada en concordancia con los principios de las Buenas Prácticas de Manufactura (GMP) y Buenas Prácticas de Laboratorio, que demuestran que los procedimientos, procesos, sistemas de control y producción, son conducentes a resultados dispuestos dentro de los límites establecidos, en las especificaciones propias del medicamento lo que garantiza que los productos cumplen con los atributos de calidad del diseño. La validación de procesos para la fabricación de medicamentos es muy importante en la industria farmacéutica, ya que se encarga de entregar un producto que fue fabricado bajo control y con un proceso fiable, lo que garantiza y entrega seguridad de la producción a los pacientes (Decreto Supremo N°3, 2010).

Validación como definición según el Instituto de Salud Pública (ISP); es la verificación de determinados parámetros de un método, en donde se plantean los requisitos especificados para éstos, demostrando que el método es idóneo para su uso previsto. También proporciona un alto grado de seguridad para que un proceso específico resulte ser consistente, lote a lote, y que representa sus especificaciones predeterminadas y sus características de calidad. (ISP, 2010, p. 7)

Existen tres tipos de validaciones, siendo la utilizada en la industria farmacéutica en Chile es la validación concurrente, es decir, que es la validación de tres lotes que están producidos en forma consecutiva, esta se lleva a cabo durante la manufactura de rutina de un producto a comercializar.

Según la Real Farmacopea Española, las formas farmacéuticas de liberación modificada (FFLM), o también conocidas como formas farmacéuticas de liberación controlada, son aquéllas en las que la velocidad y el lugar de liberación de la sustancia o sustancias activas son diferentes de la

forma farmacéutica de liberación convencional, administrada por la misma vía. Estos medicamentos realizan un control sobre la liberación del principio activo en el organismo de las personas, donde se puede escoger el lugar donde se ejercerá la liberación y también en qué ritmo. Así mismo, las formas farmacéuticas de liberación prolongada tienen como finalidad garantizar una liberación más lenta de la sustancia o sustancias activas, en comparación a una forma farmacéutica de liberación convencional administrada por la misma vía. (Real Farmacopea Española, 2002)

La FDA en su guía de proceso de validación, destaca y enfatiza la importancia que tiene la estadística para este proceso, donde en las GMP ya se aplican normalmente en el proceso de fabricación de medicamentos en sus tres etapas, en las cuales se debe decidir cuál herramienta estadística es la que más se adecua al producto y proceso que se esté llevando a cabo. Entre otros factores se considera el tamaño del conjunto de datos (lote), la selección de variables, de los atributos y parámetros que se utilizan para realizar inferencias sobre el rendimiento del proceso.

Durante la Validación de Procesos de Manufactura (VPM), se estudian todos los pasos de ejecución del procedimiento de fabricación del medicamento. Para evaluar el proceso se toman muestras según el protocolo de validación, las cuales se someten a análisis, entre las cuales destacan dos tipos de pruebas:

- Las pruebas físicas:
 - Humedad del granulado y la mezcla.
 - Peso de los comprimidos.
 - Dureza de los comprimidos.
 - Dimensiones de los comprimidos (diámetro, espesor, etc).
 - Friabilidad.

- Las pruebas químicas:
 - Uniformidad de mezcla para comprimir.
 - Valoración de los comprimidos.
 - Uniformidad de dosis de los comprimidos.
 - Disolución de los comprimidos.

Cada una de estas pruebas permiten recoger datos acerca de los medicamentos, los que deberán ser procesados bajo algunos métodos estadísticos, para así obtener la información que se incluye en el dossier de validación que se presenta a la Agencia Nacional del Medicamento ANAMED,

ISP.

El formulario que se presenta ante ANAMED recoge algunos aspectos estadísticos que se dedican a las siete variables de calidad más importantes en estudio, que son: Peso, Dureza, Diámetro, Espesor, Uniformidad de dosis, Uniformidad de mezcla y Valoración de los comprimidos. En ocasiones se producen ciertas complejidades dentro del proceso del lote, las cuales dan origen a la necesidad de aplicar alguna metodología distinta que no necesariamente está dentro de dicho formulario. Esto, debido a que no existen acepciones del ente regulador ante estudios de procesos que no cumplan estadísticamente con los estándares mínimos de aceptación.

Un proceso de control, para cualquier tipo de fabricación debe estar dentro de rangos establecidos, es por esto que se exige realizar prueba de normalidad, cálculo de capacidades de procesos, estudios de varianzas y análisis gráficos a todos los atributos en estudio en la industria farmacéutica, permitiendo demostrar que la media muestral y su variabilidad se encuentran dentro de los rangos permitidos por sus propios límites de especificación, y que los datos se distribuyen simétricamente. La principal problemática que puede llegar a existir la principal es el no cumplimiento de la prueba de normalidad de un atributo en estudio, ya que no es posible probar que el proceso está bien ejecutado influyendo en las otras pruebas estadísticas, debido a que la condición para poder calcular capacidades e interpretación de varianzas, es la aprobación de que el atributo en estudio se asemeja a una distribución normal.

Dentro de todos los desarrollos que se quieren aplicar en Laboratorio Andrómaco, la validación toma un rol importante para el desarrollo y posterior venta de sus productos, percatándose de que se debe realizar una revisión de la utilización de la estadística dentro de los informes de validación que son entregados al ISP, por consiguiente la empresa se ha abierto a buscar nuevos procedimientos que permitan evitar estos problemas y entregar un informe de validación más completo del que existe hoy en día.

Es por esto que la presente tesis será una guía técnica para solucionar los problemas estadísticos que se observan hoy en día, en cuanto a los pasos que se deben seguir para cumplir con los parámetros solicitados para presentar ante el ente regulador. Además, considera un estudio de la bibliografía existente para obtener soluciones estadísticas a lo ya conocido.

1.1. Objetivos

1.1.1. General

El objetivo general es aplicar metodologías estadísticas innovadoras en el campo de la validación en la industria farmacéutica, que permitan resolver los puntos críticos, que en la actualidad no tienen solución con la regulación existente según ANAMED, ISP.

1.1.2. Específicos

- Estudiar las pruebas de normalidad existentes, que permitan observar y demostrar las alternativas viables para trabajar las que no se encuentran reglamentadas según el ISP.
- Identificar los métodos estadísticos existentes para realizar la detección y gestión de outliers, los cuales son un gran problema para conseguir normalidad.
- Recopilar diferentes tipos de transformaciones para las variables de calidad dentro del proceso de validación y cálculos del índice de capacidad de proceso.
- Elaborar una nota técnica de las nuevas metodologías estadísticas aplicadas con los resultados finales obtenidos de esta tesis, la que quedará disponible para el laboratorio Andrómaco para validaciones en este tipo de formulación.

1.2. Regulación actual según Agencia Nacional de Medicamento ANAMED

Hoy en día existe un guiamiento por parte de ANAMED en su nota técnica N°4 versión 03, que indica los análisis estadísticos que deben efectuarse para la entrega de informes de validación, siendo una parte fundamental en la validación de los procesos productivos obtenidos y en las diferentes mediciones donde se evalúan los atributos.

Estos análisis estadísticos, con el fin de unificar criterios, son los siguientes:

1. Estadística descriptiva de las variables de calidad, para el caso de validación son necesarios el promedio, desviación estándar, los tres cuartiles, valor mínimo, valor máximo y coeficiente de variación.
2. Prueba Shapiro Wilk, utilizada para determinar si los datos poseen una distribución normal, que es la que exige ANAMED.
3. Intervalo de confianza del 95 % para determinar donde se puede encontrar el verdadero valor del parámetro en estudio con definida probabilidad.
4. Estudio de capacidad de cada etapa, a través del índice de capacidad (Cp-Cpk).
5. Cartas de control con sus límites de especificación definidos con anterioridad, y límites de control calculados gracias a la estadística descriptiva calculada.
6. Homogeneidad de medias entre los lotes en estudio, utilizando prueba de Análisis de Varianza ANOVA de 1 vía.
7. Homogeneidad de varianzas entre los lotes en estudio, utilizando prueba de Bartlett.

Capítulo 2

Materiales y métodos

El área de validación de laboratorios Andrómaco, para cumplir con las exigencias requeridas por ANAMED, realiza la validación concurrente (análisis de tres lotes de producción consecutivos) para llevar a cabo el proceso correspondiente.

En este proceso existen siete atributos o hablando estadísticamente "variables" en estudio de determinado medicamento, donde los denominados "más críticos" son el peso y uniformidad de contenido.

Para realizar la validación de un medicamento se trabaja con tres lotes fabricados consecutivamente, donde por cada uno de los lotes se cuenta con 100 muestras.

En estadística existen variadas pruebas de normalidad que se pueden aplicar. Se realizará un análisis de cada uno para observar cuál es el que mejor se adecúa para la aplicación en la industria farmacéutica, más específicamente en validación de medicamentos.

Hoy en día la prueba de normalidad utilizado en este campo es el Shapiro-Wilk, por muchos autores y especialistas es nombrado como el indicado para su aplicación. No obstante ello, adquiere relevancia el realizar un análisis de cada prueba para corroborar su adecuada aplicación y el porqué los demás se descartan o no son utilizados.

2.1. Estadística descriptiva

La estadística descriptiva es el primer paso para comenzar a informar sobre el proceso que se está llevando a cabo.

ANAMED solicita, según su guía de validación de medicamentos, las siguientes estadísticas para comenzar a observar como han resultado los análisis.

Parámetros	Lote 1	Lote 2	Lote 3
N			
Valor Mínimo			
Cuantil 1			
Cuantil 2			
Cuantil 3			
Valor Máximo			
Promedio			
Desviación estándar			
Coefficiente de variación (%)			
IC 95 %			

Tabla 2.1.1: Análisis descriptivo según formulario.

- Media

La media es una medida de tendencia central que es utilizada para observar el comportamiento que resulta de un proceso en estudio, donde a veces existen valores que ni se acercan al valor de la media de la muestra, denominados outliers. Para esto se necesitan estadísticas complementarias para fortalecerla.

- Valores mínimo y máximo

La importancia de estos valores es que nos da una primer imagen a priori de si pudiesen existir valores atípicos dentro de cada lote, ya que ha haciendo el cálculo de $\pm 3\sigma$ en torno a la media de la muestra, Podemos determinar en pequeña medida si podríamos estar en presencia de datos extraños en el proceso.

- Cuartiles

Los cuartiles son una medida de posición. Estos nos dan la información por ejemplo, en el caso del cuantil 1, contiene 25 % de los datos en estudio están igual o bajo determinado

valor que es calculado.

- Desviación estándar

Es una medida que indica lo disperso que están los valores en estudio con respecto a la media, una de las problemáticas que afecta su cálculo son los outliers, que son valores extremos que se posicionan alejados de la media. Está definido según la siguiente fórmula:

$$S = \sqrt{\frac{1}{(n-1)} \sum_{i=1}^n (X_i - \bar{X})^2}$$

- Coeficiente de variación

Define cuán homogéneo son los datos, Sirve para realizar comparaciones entre grupos, en este caso poder comparar entre los lotes como anduvo la dispersión de los valores en estudio donde es el cociente entre la desviación estándar y la media. Cuando este coeficiente es mayor

- Análisis de Varianza

La prueba de ANOVA es utilizada para determinar si existe una diferencia significativa entre las medias de los lotes en estudio (son tres lotes y se denomina validación concurrente).

El modelo para utilizar esta prueba es el siguiente (Tejedor, 1999):

$$Y_{ij} = \mu + \tau_i + \epsilon_{ij}$$

donde,

Y_{ij} : Variable respuesta obtenida desde la j -ésima variable del i -ésimo tratamiento.

τ_i : Constante del efecto del tratamiento.

μ : Media aritmética.

ϵ_{ij} : Error de estimación.

La hipótesis de estudio ANOVA:

$H_0 : \tau_1 = \tau_2 = \dots = \tau_v$ No existe diferencia significativa entre los tratamientos.

$H_0 : \tau_1 \neq \tau_2 \neq \dots \neq \tau_v$ En al menos un tratamiento existe diferencia.

ANOVA requiere de los siguientes cumplimientos para su correcta utilización:

- Las población en estudio se destruye normal.
- Las K muestras sobre las que se aplican los tratamientos son independientes.
- Existe semejanza de varianza entre los grupos en estudio (homocedasticidad).

■ Prueba de Bartlett

Esta prueba introducida por Bartlett (1937) es utilizada para corroborar de que los lotes en estudio provienen de una población con misma varianza, denominado homocedasticidad u homogeneidad de varianzas. Los n_i en cada tratamiento no necesariamente deben ser iguales; sin embargo, se estima que no sean menores qque 3. La dócima de hipótesis está definida como:

$H_0: \sigma_1^2 = \sigma_2^2 = \dots = \sigma_n^2$ Las varianzas son iguales.

H_1 : Al menos una varianza es distinta.

El estadístico se define como:

$$U = \frac{1}{C} \left[(N - k) \ln(s^2) - \sum_{i=1}^k (n_i - 1) \ln s_i^2 \right]$$

donde,

$$C = 1 + \frac{1}{3(k-1)} \left(\sum_{i=1}^k \frac{1}{n_i - 1} - \frac{1}{(N-k)} \right)$$

con S_i^2 como la varianza muestral de la i -ésima población.

Cuando la hipótesis H_0 es cierta, el estadístico tiene aproximadamente una distribución ji-cuadrado con $k-1$ grados de libertad, esto para cuando se trabaja en base a supuesto de normalidad, este estadístico funciona para muestras pequeñas (Layard, 1973).

$$X_0^2 > X_{\alpha, k-1}^2$$

donde $X_{\alpha, k-1}^2$ es el punto porcentual α superior de la distribución ji-cuadrado con $k-1$ grados de libertad.

2.2. Pruebas de normalidad

2.2.1. Prueba de Shapiro-Wilk

Algunos textos sostienen de que la prueba de Shapiro-Wilk tiene una funcionalidad importante en muestras de hasta 50 datos, sin embargo al pasar el tiempo se ha considerado que esta prueba funciona para valores que van entre $4 < n < 2000$, a pesar que, al aumentar la muestra este no es muy estable (Shapiro, 1965). Al pasar los años se ha ido aclarando este tema, ya que en el caso de la validación de medicamentos generalmente se trabaja con 100 valores por cada lote y atributo, ya sea físico o químico. Esto nos lleva a ir en busca de alguna prueba alternativa para confirmar normalidad más adecuada que el utilizado actualmente.

Existen varias pruebas no paramétricas que se utilizan para analizar si un conjunto de datos proviene de una distribución continua determinada. Consiste en características generales, como la independencia y que la variable que se está estudiando sea continua. Esto a diferencia de la estadística paramétrica nos abre un abanico más amplio para poder trabajar y encontrar una distribución adecuada para algún estudio en donde no se tiene certeza del comportamiento de los

datos.

El estadístico de Shapiro-Wilk (Shapiro, 1965) está definido como:

$$W = \frac{(\sum_{i=1}^n a_i y_i)^2}{\sum_{i=1}^n (y_i - \bar{y})^2},$$

donde,

$$a_i = \frac{m^2 V^{-1}}{(m^T V^{-1} V^{-1} m)^2},$$

con $m = (m_1 \dots m_n)^T$ que corresponden al vector de valores esperados de estadísticas de distribución normal estándar y V se identifica como la matriz de Covarianza.

2.2.2. Prueba de Kolmogorov-Smirnov

Esta prueba no paramétrica es de las más reconocidas, de donde viene la prueba de Anderson-Darling y Lilliefors. A diferencia de estas, contiene sensibilidad en las colas. Esta prueba es muy interesante analizarlo ya que con muestras grandes es bastante potente y confiable, basada en la diferencia vertical más grande entre la hipótesis y la distribución observada. También sólo es válido para trabajar con variables continuas, donde en el caso de Laboratorio Andrómaco, generalmente se trabaja con 100 muestras por cada uno de los tres lotes y para cada atributo de los medicamentos (Conover, 1999).

Requiere ingresar la media y la desviación estándar para su cálculo, estas corresponden a la población, se ocupa para muestras grandes por la posible distorsión de los resultados al utilizar pequeñas cantidades de valores.

Para X_n variable aleatoria, el estadístico propuesto por Kolmogorov (1933) de esta prueba como:

$$D = \text{máx} |F_t(x) - F_n(x)|$$

donde $F_n(x)$ es la función de distribución empírica, $F_t(x)$ representa a la función de distribución de hipótesis, en este caso una distribución normal con media μ y σ como valores conocidos. Con *máx* se busca el mayor valor resultante de la función.

Las dójimas de hipótesis que se utilizan en este estadístico son:

H_0 : Los datos analizados siguen una distribución normal.

H_1 : Los datos analizados no siguen una distribución normal.

Nota: Como se sabe la prueba de Kolmogorov-Smirnov es una prueba no paramétrica para determinar si un conjunto de datos pertenece a una distribución específica, en este caso una distribución normal.

Mediante la tabla de cuantiles utilizada para la prueba de Kolmogorov-Smirnov se determina con el nivel de significación establecido y el tamaño de muestra que se está utilizando.

2.2.3. Prueba de Anderson-Darling

Es una prueba no paramétrica que proviene de una mejora de la prueba de Cramer-von Mises, el cuál es utilizada para ver si desde un conjunto de valores muestrales probar si provienen de una distribución específica, donde generalmente se utiliza para estudiar la normalidad (Anderson y Darling, 1954).

El estadístico de Anderson-Darling está dado por:

$$A_n^2 = -n - \frac{1}{n} \sum_{k=1}^n (2k - 1) [\log F(x_k) + \log (1 - F(x_{n+1-k}))]$$

Los datos $x_1 \dots x_n$ deben ordenarse de menor a mayor para poder aplicar el estadístico necesario. $F(x_k)$ corresponde a la distribución acumulada especificada.

Las dójimas de hipótesis que se utilizan en este estadístico son:

H_0 : Los datos analizados siguen una distribución normal.

H_1 : Los datos analizados no siguen una distribución normal.

Nota: Como se sabe la prueba de Anderson-Darling es una prueba no paramétrica para determinar si un conjunto de datos pertenece a una distribución específica, en este caso una distribución

normal.

2.2.4. Prueba de Lilliefors

Este test es en esencia una variante a la prueba de Kolmogorov-Smirnov (K-S). Se desea probar si una muestra proviene de una distribución de una familia normal ya sea con media y varianza de su población desconocida, esto hace que la prueba de Lilliefors sea más potente, a diferencia de (K-S), donde se supone que los parámetros de la distribución ya se conocen con anterioridad. Sin embargo, a veces es complicado definir los parámetros, ya que la distribución es desconocida. En el caso de la prueba de Lilliefors se estiman en base a la muestra (*Lilliefors*, 1967).

El estadístico de Lilliefors está definido por:

$$D = \max_x |F(x) - S_n(x)|$$

Donde $S_n(x)$ es la función distribución acumulativa de la muestra y $F^*(x)$ es la función de distribución normal con media $\mu = \bar{X}$ como la media de la muestra y $\sigma^2 = S^2$ como la varianza de la muestra.

Las dójimas de hipótesis que se utilizan en este estadístico son:

H_0 : Los datos analizados siguen una distribución normal.

H_1 : Los datos analizados NO siguen una distribución normal.

Nota: Como se sabe la prueba de Lilliefors es una prueba no paramétrica para determinar si un conjunto de datos pertenece a una distribución específica, en este caso una distribución normal.

2.2.5. Prueba de Cramer-von Mises

Al igual que las pruebas no paramétricas anteriores, esta prueba es otra alternativa desde el test (K-S). Según Conover (1999) esta prueba fue desarrollada por Cramer (1928), Von Mises (1931)

y Smirnov (1936).

Sean x_1, x_2, \dots, x_n los valores observados en orden de menor a mayor, el estadístico está definido como:

$$\omega^2 = \frac{1}{12n} + \sum i = 1n \left[\frac{2i-1}{2n} - F(x_i) \right]^2$$

Las dójimas de hipótesis que se utilizan en este estadístico son:

H_0 : Los datos analizados siguen una distribución normal.

H_1 : Los datos analizados no siguen una distribución normal.

Nota: Como se sabe la prueba de Lilliefors es una prueba no paramétrica para determinar si un conjunto de datos pertenece a una distribución específica, en este caso una distribución normal.

2.3. Capacidad del proceso productivo

Dentro de cualquier proceso industrial, en este caso farmacéutico, se encuentra variabilidad dentro de este proceso irremediamente, es decir, los datos obtenidos en la fabricación de un medicamento no son tan semejantes como se espera y esto lleva a la industria a buscar ciertos métodos para poder corregir este factor externo para eliminar o mas bien reducir lo que mas se pueda la diferencia entre los datos de los atributos obtenidos al obtener la muestra.

La capacidad de proceso se encarga de medir que tan capaz es un proceso de cumplir con las especificaciones que se determinan en un principio. La principal principal condición para poder realizar el cálculo de las capacidades de proceso es que los datos en estudio sigan una distribución normal.

Los índices de capacidad de proceso son dos y se describen a continuación:

2.3.1. Índice de Capacidad Cp

Este índice mide la capacidad de un proceso para producir un resultado esperado que cumpla con sus límites predefinidos desde un principio, este concepto es primordial para demostrar que el proceso es confiable y se encuentra bajo control estadístico. Para esto necesitamos calcular la media y desviación estándar, donde este último cumple un rol fundamental, ya que la variabilidad, la que es inevitable por múltiples factores incontrolables dentro de cualquier proceso industrial (Montgomery, 2004).

El cálculo de este índice está dado por:

$$Cp = \frac{LES - LEI}{6\sigma}$$

donde LES y LEI corresponden a los límites superior e inferior de especificación respectivamente.

Los criterios para determinar según este índice si el proceso cumple según las especificaciones son:

- Si $Cp > 1,33$ el proceso en estudio es muy capaz.
- Si $1 \leq Cp \leq 1,33$ el proceso en estudio es capaz, pero requiere de una revisión.
- Si $Cp < 1$ el proceso requiere de un mejoramiento en su procedimiento para obtener mejores resultados.

2.3.2. Índice de Capacidad Cpk

Este índice está definido como el valor que representa la relación entre la media y la distancia que existe hacia los límites de especificación, es decir, se mide si el proceso se ajusta a las tolerancias o en otras palabras, si la media del proceso está centrada en relación a la media nominal

determinada desde un principio (Montgomery, 2004).

El cálculo de este índice está dado por:

$$Cpk = \text{Min} \left\{ \frac{LES - \mu}{3\sigma}, \frac{\mu - LEI}{3\sigma} \right\}$$

Donde LES y LEI corresponden a los límites superior e inferior de especificación respectivamente.

Para determinar si el proceso está centrado se debe comparar el valor Cp con Cpk:

- Si $Cp = Cpk$ aproximadamente el proceso está centrado.
- Si $Cp > Cpk$ el proceso no está centrado.

Se realizó también la búsqueda del código para su cálculo en el programa estadístico Stata Corp.

2.4. Transformación de variables

Luego de aplicar las pruebas para determinar si las muestras en estudio siguen una distribución normal y obteniendo resultados negativos para el estudio, se deben buscar soluciones alternativas para normalizar los valores. Estas alternativas lo que quieren conseguir es tanto corregir cuando existen varianzas muy desiguales entre sí y mejorar la linealidad entre las variables que se encuentran en estudio.

En este sentido la transformación más típica y convencional que se utiliza hoy en día tanto por el ISP como por los procedimientos de la gran mayoría de los procesos es la logarítmica, sin embargo al no obtener resultados siempre positivos, es interesante observar otro tipo de transformaciones tan viables como esta.

Cabe mencionar que con respecto a las transformaciones no existe bibliografía especializada en

el tema, sino que en cuanto al razonamiento matemático se indican sus aportes al disminuir la variabilidad de la variable o atributo que se estudia.

Las transformaciones conocidas son las siguientes:

- Transformación Box-Cox

Mediante métodos de máxima verosimilitud, se propone la familia de transformaciones (Box y Cox, 1964), sea Y una variable aleatoria:

$$Y(\lambda) = \begin{cases} \frac{(y+m)^\lambda - 1}{\lambda}, & \text{si } \lambda \neq 0 \\ \ln(y + m), & \text{si } \lambda = 0 \end{cases}$$

con λ y m parámetros de la transformación de la familia Box-Cox, estos como dicho anteriormente, se estiman con el método de máxima verosimilitud. Es utilizada para obtener homocedasticidad, linealidad y normalidad en modelos estadísticos.

- Transformación logarítmica

La función de esta transformación es obtener la estabilidad y disminuir la variación de la varianza. En el caso de que la muestra analizada posea valores 0 lo que se recomienda es $\ln(x + 1)$, así el cálculo puede ser obtenido con cualquier tipo de valor. Esta transformación corresponde a ser la más conocida y utilizada en todo tipo de estudio, está reconocida por el ISP como el método de transformación a seguir.

- Transformación fraccionaria

Este tipo de transformación de la forma $\frac{1}{x}$, donde x corresponde a los datos provenientes del atributo de calidad en estudio, lo que consigue es obtener valores entre 0 y 1, al igual que la transformación logarítmica si es que existen valores 0 originalmente es correcto precisar que se puede utilizar la fracción $\frac{1}{x+1}$.

- Transformación cuadrática

Con esta transformación x^2 lo que se consigue es expandir los valores originales, con este método lo que se intenta es reducir la variabilidad existente.

- Transformación raíz cuadrática

Mediante \sqrt{x} se achican los valores que se están transformando, al igual que la transformación cuadrática se quiere principalmente reducir la variabilidad.

Cada una de las transformaciones presentadas anteriormente son viables para poder normalizar los datos, aunque puede resultar que con ninguno se consigan resultados positivos, es una técnica muy válida. Los valores atípicos pueden influir en que estos métodos no sean efectivos.

2.5. Datos atípicos

Generalmente cuando se trabaja con cualquier tipo de conjunto de datos está el problema de los datos atípicos u outliers se consideran como aquellos que se encuentran a una distancia numérica considerable al resto de los datos, este fenómeno es muy difícil de evitar ya que existen múltiples condiciones externas que pueden provocarlos durante el proceso de fabricación de medicamentos.

Es por esto que, es interesante realizar un análisis de los datos atípicos para un reconocimiento del problema presente y la búsqueda de una solución eficiente.

Estos métodos estadísticos corresponden a las siguientes pruebas: prueba de Grubb's, prueba de Dixon y la prueba de Tukey.

2.5.1. Prueba de Grubb's

La prueba de Grubb's primordialmente se basa en el supuesto de normalidad, es decir, se debe corroborar que los datos se aproximen a una distribución normal (Grubbs, 1950).

Esta prueba está descrita por la detección de un valor atípico a la vez, en este caso de la industria farmacéutica es la gente especializada la que debe decidir si el valor atípico detectado es extraño al proceso normal que se está llevando a cabo y eliminarlo con fundamentos (Grubbs, 1950).

Según Grubbs (1950) el estadístico está definido como:

$$G = \frac{\max_{i=1, \dots, n} |Y_i - \bar{Y}|}{s}$$

donde \bar{Y} como la media de la muestra en estudio y s como la desviación estándar.

Para buscar si el valor mínimo de la muestra es un dato atípico el estadístico de prueba se define por:

$$G = \frac{\bar{Y} - Y_{min}}{s}$$

donde Y_{min} representa al valor mínimo. Para buscar si el valor máximo de la muestra es un dato atípico el estadístico de prueba se define por:

$$G = \frac{Y_{max} - \bar{Y}}{s}$$

donde Y_{max} representa al valor máximo.

La hipótesis estadística para esta prueba es la siguiente:

H_0 : El dato en estudio es un valor atípico.

H_1 : El dato en estudio no es un valor atípico.

2.5.2. Prueba de Dixon

Al igual que la prueba de Grubb's se debe suponer que los datos en estudio se aproximan a una distribución normal, en general es recomendable para cuando se poseen muestras no mayores a 30. Para su realización los datos deben ordenarse en orden creciente (Dixon, 1950).

$x_1 < x_2 < \dots < x_n$, el estadístico de Dixon (1950) se define como:

Para x_1 como posible valor atípico:

$$Q = \frac{x_2 - x_1}{x_n - x_1}$$

Para x_n como posible valor atípico:

$$Q = \frac{x_n - x_{n-1}}{x_n - x_1}$$

La hipótesis estadística para esta prueba es la siguiente:

H_0 : El dato en estudio es un valor atípico.

H_1 : El dato en estudio no es un valor atípico.

Otro métodos de detección de datos atípicos:

- Criterio de Chauvenet

Es un método no tan utilizado, su fundamento es utilizar la probabilidad. Se necesita calcular la media y desviación estándar de los datos en estudio, donde se utiliza la función de distribución normal mediante su tabla. Se desea determinar la probabilidad de que un dato dado sea del valor que se está dudando si es atípico o no, entonces se multiplica la probabilidad obtenida por el número de datos de la muestra escogida, si el resultado es inferior a 0,5 el dato que se pone en duda se puede concluir que es un dato atípico.

2.6. Métodos gráficos

Los métodos gráficos son de gran ayuda para tener una observación a priori de los datos que se estudian. Son una ran alternativa de estudio ya que es de utilidad para corroborar o anticiparse a los datos concretos mediante la estadística descriptiva. Estos métodos ofrecen alternativas para la identificación del comportamiento de valores en estudio, observación de ourliers, variabilidad y cumplimiento de límites de especificación entre otros. Se presentan los gráficos a continuación:

- Histograma

El histograma es de gran ayuda para visualizar como se comportan los datos, y lo mas importante, analizar si poseen forma de campana, con sus colas respectivas, representando a la distribución normal que es la que se busca.

Se describen las abreviaturas siguientes:

- LIE: Límite Inferior de Especificación.
- LSE: Límite Superior de Especificación.
- LIC: Límite Inferior de Control.
- LSC: Límite Superior de Control.

Como se observa en Figura 2.6.1, se dibuja la distribución normal aproximada para realizar una comparación y definir si los datos cumplen con esta condición. Claramente las barras que se encuentran cerca de la media no tienen una tendencia a disminuir y por ende se define según el gráfico que no se cumple con normalidad.

- Gráfico de probabilidad Cuantil-Cuantil (QQ-plot)

Este gráfico permite observar mediante una linea trazada de manera diagonal si los valores

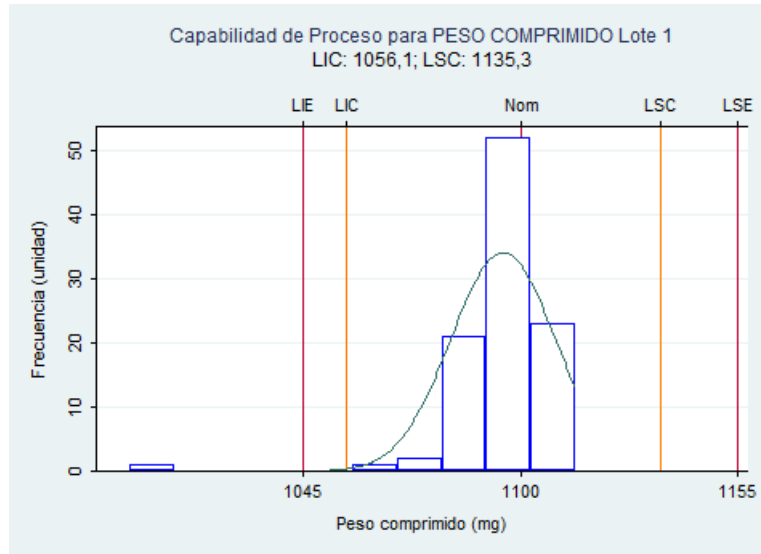


Figura 2.6.1: Histograma para variable de calidad Peso

que estoy estudiando se asemejan o se acercan a una distribución normal, también sirve de gran ayuda para detectar posibles datos atípicos que podrían estar influyendo en que el conjunto de datos no cumpla con la condición normal.

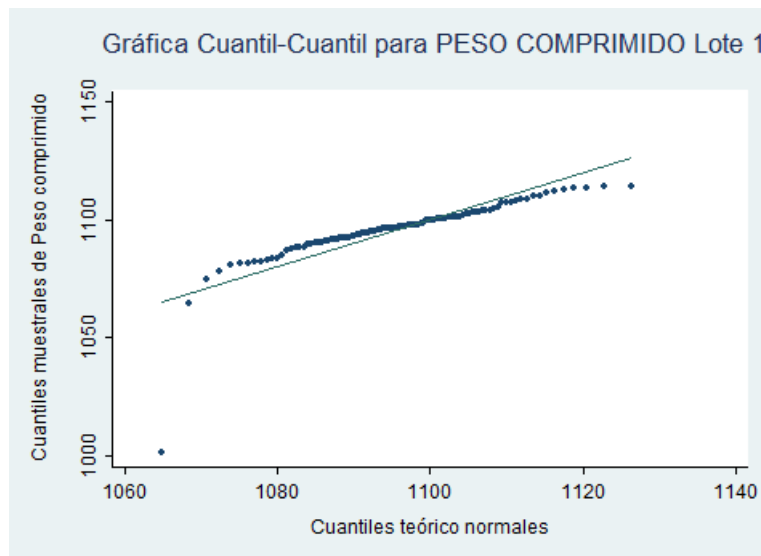


Figura 2.6.2: Gráfico Cuantil-Cuantil para variable de calidad Peso

Se puede observar en Figura 2.6.2 que los valores se tornan a la línea de normalidad, sin embargo también se puede observar que existen al menos dos valores máximos los cuales pudiesen ser datos atípicos que influyen en la no normalidad de los datos en estudio, por

ende se debe proceder a analizar si es posible que cumplen con las condiciones de ser datos atípicos.

■ Gráfica de Control

Este gráfico lo que ayuda es a visualizar claramente con sus límites de especificación y límites de control calculados si el proceso esta bajo control estadístico.

También en algunas ocasiones ayuda para observar que si algún dato se encuentra fuera de un límite de control, analizarlo por si existe a posibilidad de que sea un dato atípico influyendo de alguna manera en el proceso.

Se describen las abreviaturas siguientes:

- LIE: Límite Inferior de Especificación.
- LSE: Límite Superior de Especificación.
- LIC: Límite Inferior de Control.
- LSC: Límite Superior de Control.

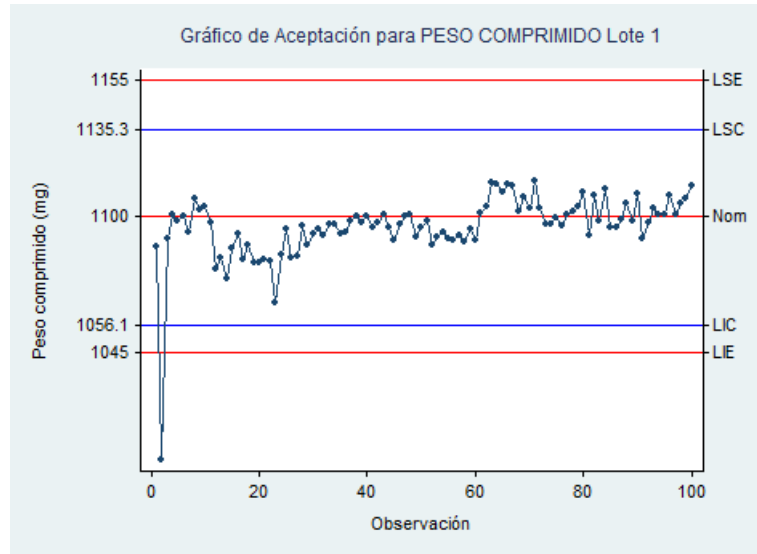


Figura 2.6.3: Gráfico de aceptación para variable de calidad Peso

En Figura 2.6.3 se identifica un posible valor atípico que se encuentra alejado del resto, los cuales se observan dentro de los límites de especificación.

■ Gráficos Box-Plot

Este gráfico también denominado como Gráfico de Caja, nos permite obtener una visión acerca de la simetría de la distribución de los datos. También un aporte importante de este tipo de gráfico es la detección visual de datos atípicos, esto al existir puntos fuera de la caja y los límites de los bigotes correspondientes.

Se describen las abreviaturas siguientes:

- LIE: Límite Inferior de Especificación.
- LSE: Límite Superior de Especificación.
- LIC: Límite Inferior de Control.
- LSC: Límite Superior de Control.

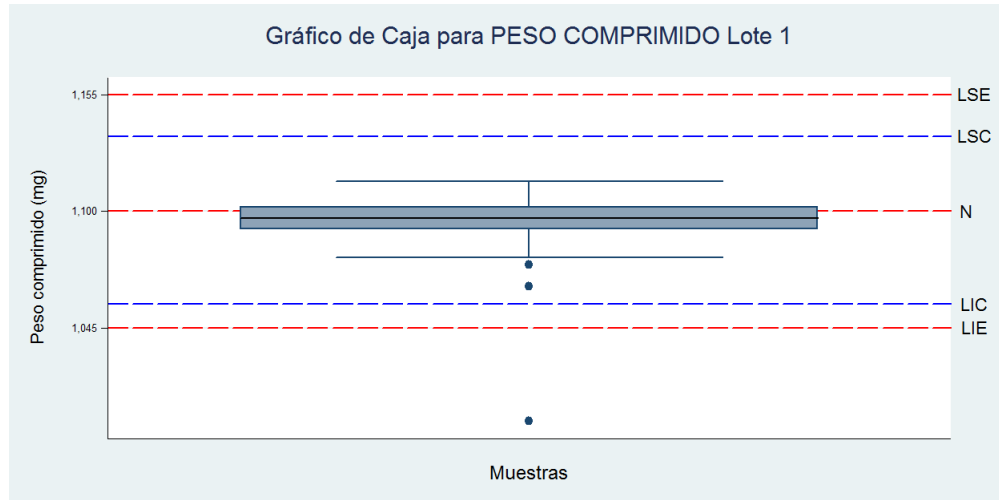


Figura 2.6.4: Gráfico de Caja para variable de calidad Peso

El Gráfico de Caja de gran ayuda para observar variabilidad del conjunto de datos, visión de valores atípicos, analizar la mediana para identificar la simetría y hacia que límite de especificación se encuentra la tendencia.

Los métodos gráficos mencionados sirven como apoyo para las estadísticas robustas y determinar los pasos a seguir en este proceso de cumplimiento que se quiere lograr.

Capítulo 3

Resultados

Finalmente, los resultados obtenidos se presentarán en este capítulo con los 3 lotes en estudio los que corresponden a un medicamento de liberación prolongada de laboratorios Andrómaco, donde se expondrán todos los análisis requeridos por la entidad de salud y su guía de validación anteriormente detallada.

El objetivo principal a conseguir en este estudio es mediante estadística alternativa ya conocida, poder solucionar distintos problemas que se presentan dentro de un análisis estadístico en un informe de validación de proceso, el cuál como se ha indicado anteriormente es primordial para la posterior venta al mercado y favorecer la salud del paciente.

Como presentación inicial, la fabricación de estos 3 lotes que se estudiarán fueron realizados con las condiciones óptimas para asegurar independencia entre los datos de peso y que se cumpla con las mismas condiciones en cada lote, ya sea maquinarias para fabricar, punzones, entre otros. Es importante indicar que la cantidad de N valores de Peso por cada uno de los lotes no son la totalidad de comprimidos fabricados, sino que está compuesta por N muestras determinadas en el protocolo de fabricación.

3.1. Análisis

Antes de todo tipo de análisis se deben determinar los límites de especificación del comprimido de liberación prolongada para el atributo peso que se encuentra en estudio, los cuales son:

- Límite de especificación inferior es de 1045 miligramos.
- Valor nominal es de 1100 miligramos.
- Límite de especificación superior es de 1155 miligramos.

Los límites de control del proceso de cada lote serán determinados luego de haber realizado los análisis descriptivos.

3.1.1. Estadística descriptiva

Se presenta en la siguiente Tabla el análisis descriptivo realizado a los tres lotes que se encuentran en estudio.

Parámetros	Lote 1	Lote 2	Lote 3
N	100	100	100
Valor Mínimo	1001,1	1061,1	1000,1
Cuantil 1	1091,4	1094,2	1085,3
Cuantil 2	1096,8	1101,8	1093,8
Cuantil 3	1102,0	1112,5	1100,9
Valor Máximo	1114,0	1136,6	1124,3
Promedio	1095,7	1102,6	1092,5
Desviación estándar	13,2	15,6	15,1
Coefficiente de variación (%)	1,20	1,41	1,38
IC 95 %	1093,05 - 1098,28	1099,47 - 1105,66	1089,50 - 1095,50

Tabla 3.1.1: Análisis descriptivo de peso.

- Se observa en el cuadro 3,1,1 que tanto los valores mínimos como los máximos de los tres

lotes no sobrepasan los límites de especificación que están determinados.

- El promedio de los lotes esta extremadamente cercano a la media nominal determinada que es de 1100 miligramos, esto se repite en el segundo cuartil más conocido como la mediana.
- Se debe tomar en cuenta que los valores de desviación estándar son grandes ya que el peso de los comprimidos es una cantidad importante de miligramos.
- Según los IC en los lotes 1 y 2 con un 95 % de confianza el valor nominal se encuentra dentro del intervalo, no así para lote 3 el cual su intervalo se encuentra por debajo del valor nominal.

3.1.2. Normalidad

Como se ha mencionado anteriormente, para realizar un análisis estadístico más detallado sobre los atributos de validación se debe cumplir con la d^ocima de hipótesis de normalidad, es decir, demostrar que los datos se distribuyen normal y así cumplir con las exigencias de las entidades de salud en este caso ANAMED.

Se presentarán a continuación las hipótesis estadísticas con los test de normalidad existentes para determinar si se cumple con la condición principal para poder realizar los estudios siguientes.

Pruebas utilizadas mediante Stata Corp:

- Shapiro-Wilk, Shapiro-Francia y Skewness/Kurtosis.

Pruebas utilizadas mediante R project:

- Kolmogorov-Smirnov, Anderson-Darling, Cramer-von Mises y Lilliefors.

Test	Lote 1	Lote 2	Lote 3
Shapiro-Wilk	< 0,0001	0,2864	< 0,0001
Shapiro-Francia	< 0,0001	0,3226	< 0,0001
Skewness/Kurtosis	< 0,0001	0,6659	< 0,0001
Kolmogorov-Smirnov	0,0375	0,6937	0,3399
Anderson-Darling	< 0,0001	0,1702	< 0,0001
Cramer-von Mises	< 0,0001	0,1708	0,0011
Lilliefors	< 0,0001	0,2458	0,0296

Tabla 3.1.2: Valor-p prueba de normalidad de variable de calidad Peso

En lote 1 se observa que en todas las pruebas de normalidad utilizadas se corrobora el no cumplimiento de normalidad, por el contrario en lote 2 se indica el cumplimiento de condición de normalidad y en lote 3 sólo en prueba de Kolmogorov-Smirnov se consigue cumplir con normalidad.

3.1.3. Estudio gráfico

Histogramas

Otra manera de observar si un conjunto de datos sigue una distribución normal es mediante histogramas, donde gracias a estos gráficos se puede analizar si poseen la forma de campana mencionada anteriormente, que es la forma ideal que se busca. Se presentan para cada uno de los tres lotes de validación de proceso en estudio.

Antes de analizar los gráficos se describen las abreviaturas siguientes:

- LIE: Límite Inferior de Especificación.
- LSE: Límite Superior de Especificación.
- LIC: Límite Inferior de Control.
- LSC: Límite Superior de Control.

Para comenzar a describir los valores de peso mediante los histogramas se debe ser minucioso en cuanto a observar ciertas características que nos indicarán lo que se busca, esto ayudará a una mejor comprensión del estudio, la estadística descriptiva y test de normalidad calculados.

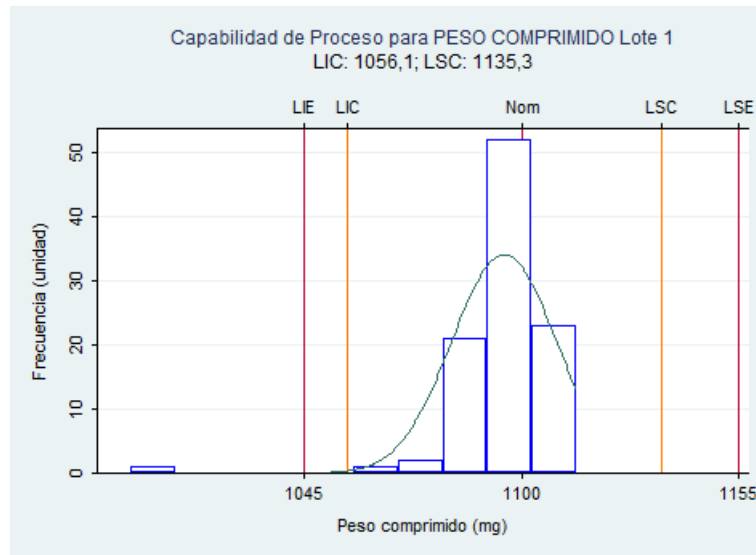


Figura 3.1.1: Histograma para variable de calidad Peso Lote 1

- En el caso de la Figura 3.1.1 se observa que los datos no se asemejan a una distribución normal, las barras del histograma indican una mayor cantidad de valores de peso por debajo del valor nominal, esto refleja una asimetría hacia el límite inferior del peso del comprimido de liberación prolongada. Finalmente se identifica un posible dato atípico que es menor a los límites establecidos, esto se debe corroborar con un test para la detección de dato atípicos.

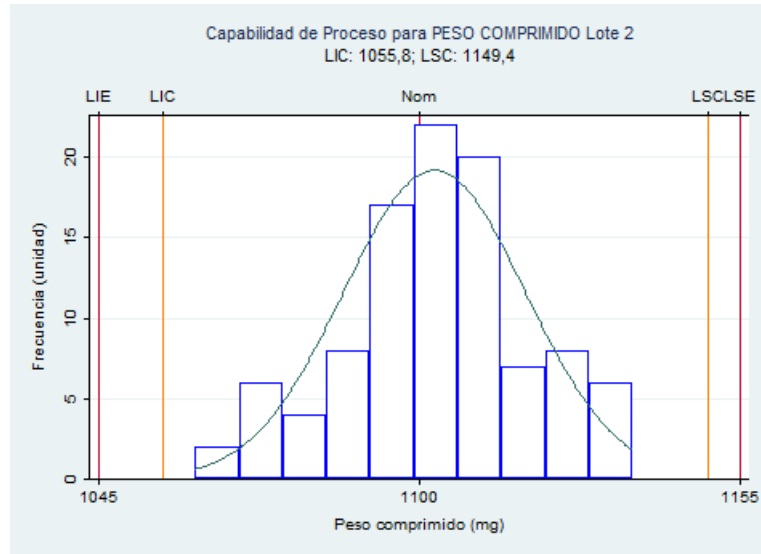


Figura 3.1.2: Histograma para variable de calidad Peso Lote 2

- En Figura 3.1.2 se aprecia una simetría de los datos de peso ya que a medida que se alejan las barras del histograma hacia los límites de especificación la frecuencia va disminuyendo de manera similar. Todos los valores de peso se encuentran dentro de sus límites de especificación y por último no hay indicios de algún valor atípico que pueda entorpecer el proceso.

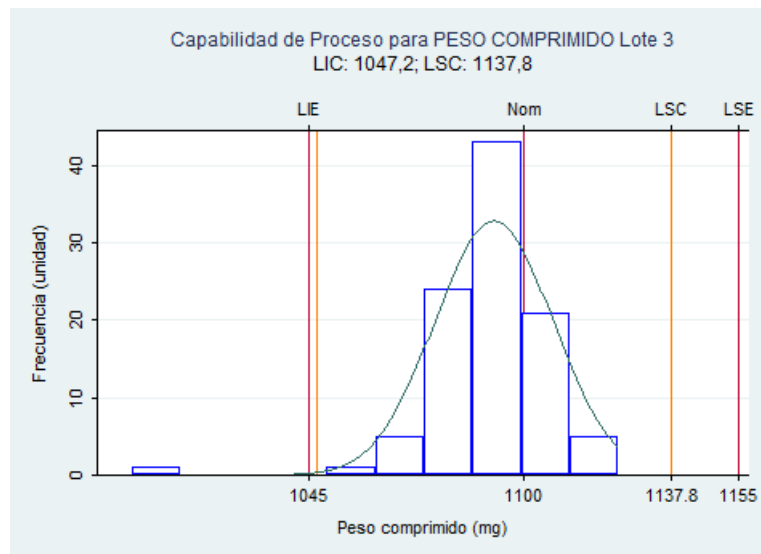


Figura 3.1.3: Histograma para variable de calidad Peso Lote 3

- Al igual que en Figura 3.1.1 no se advierte en el tercero una forma de campana la cual reflejaría una condición semejante a la distribución normal, las barras del histograma también se encuentran en mayor cantidad por debajo del valor nominal de 1100 miligramos. Una asimetría se asemeja más a este gráfico por lo expuesto anteriormente, esto se acrecenta al ver la posibilidad de que exista un valor atípico que se encuentra fuera de especificación inferior.

Gráfica Cuantil-Cuantil

Este modelo de gráfico permite observar al igual que un histograma si los valores que se estudiarían se asemejan o se acercan a una distribución normal, también a la detección de posibles datos atípicos que pueden influir en el proceso. Se hace una interpretación de cada lote a continuación:

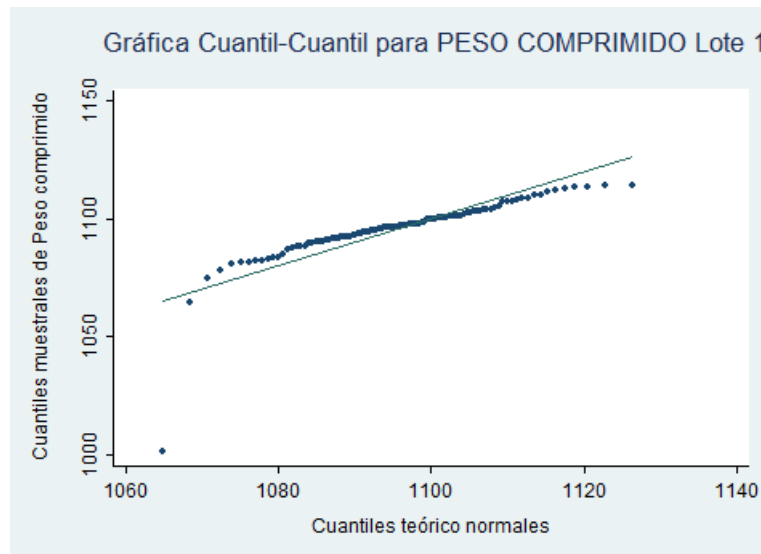


Figura 3.1.4: Gráfica Cuantil-Cuantil para variable de calidad Peso Lote 1

- Al observar la Figura 3.1.4 se corrobora la información que entrega el histograma, en el primer lote hay indicios de posible valor atípico que está afectando la no normalidad de los valores en estudio. Mediante la recta empírica de la distribución normal se identifican

valores que no se comportan según condición de normalidad.

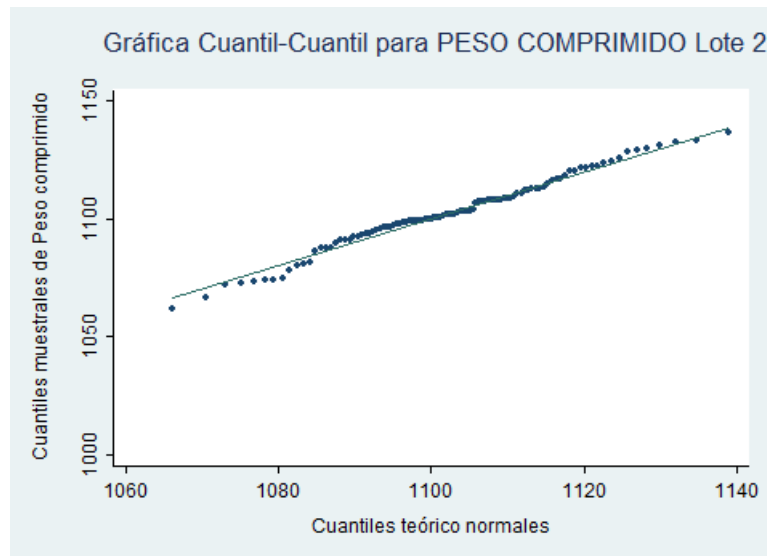


Figura 3.1.5: Gráfica Cuantil-Cuantil para variable de calidad Peso Lote 2

- En lote 2 se interpreta una semejanza a una distribución normal mediante la línea diagonal que este gráfico traza, donde los puntos deben ir rodeándola. No existen evidencias de posibles valores outliers.

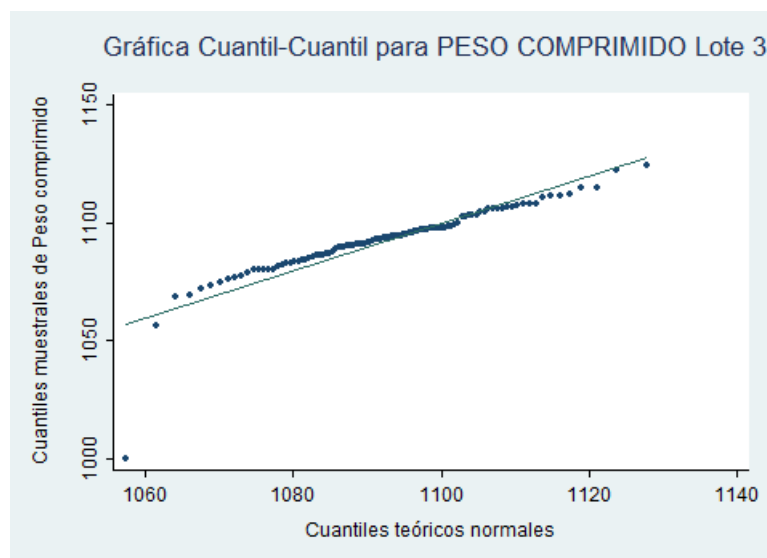


Figura 3.1.6: Gráfica Cuantil-Cuantil para variable de calidad Peso Lote 3

- Se corrobora la información que entrega el histograma, en el tercer lote existen indicios de posible valor atípico que está afectando la no normalidad de los valores. Mediante la recta empírica de la distribución normal se identifican valores que no se comportan según condición de normalidad.

Gráfico Box-Plot

Se realiza este gráfico para principalmente analizar la dispersión y simetría de los lotes en estudio donde esto ayuda a corroborar lo estudiado en los distintos tipos de gráficos ya vistos. Lo distinto de este modelo con los demás es los detalles de este, la forma de la caja mediante los cuartiles, el largo de los extremos de la caja hasta el bigote, la mediana si es que está centrada, etc. Se presentan los tres y se realizará una descripción detallada de sus características:

Antes de analizar los gráficos se describen las abreviaturas siguientes:

- LIE: Límite Inferior de Especificación.
- LSE: Límite Superior de Especificación.
- LIC: Límite Inferior de Control.
- LSC: Límite Superior de Control.

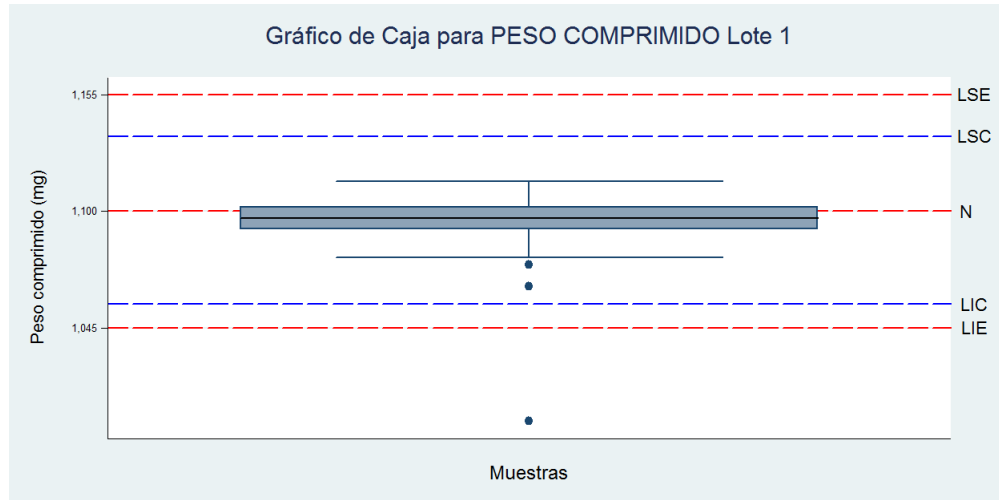


Figura 3.1.7: Gráfico de Caja para variable de calidad Peso Lote 1

- En la Figura 3.1.7 se observa en lote 1 una baja dispersión de los datos, donde alrededor del 80 % de los datos se encuentran por debajo de la media nominal establecida. Existe una asimetría entre los pesos y se identifica un punto por debajo del límite inferior de especificación correspondiente a 1045, se debe analizar si corresponde a un outlier.

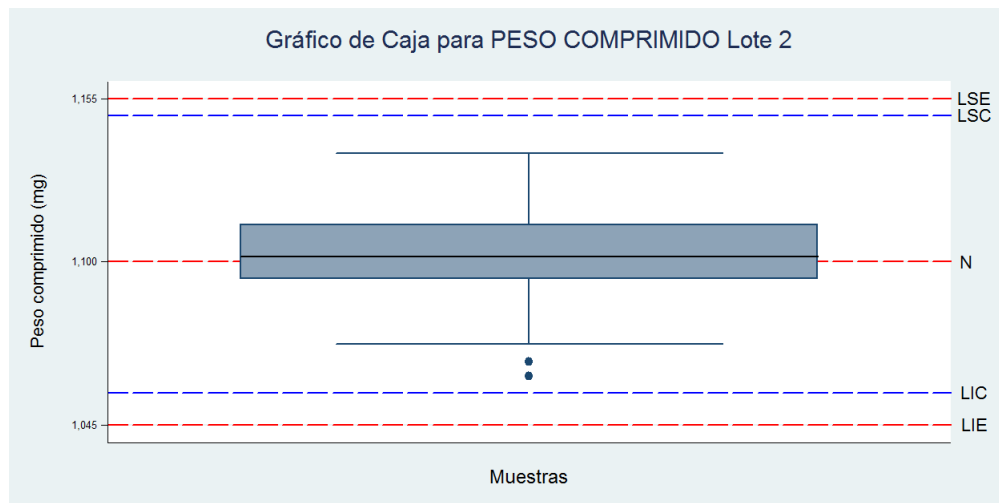


Figura 3.1.8: Gráfico de Caja para variable de calidad Peso Lote 2

- La Figura 3.1.8 presenta una simetría ya que la distancia entre el cuartil 1 hacia el valor mínimo y cuartil 3 hacia el valor máximo no se diferencian, aproximadamente el 50 % de

los datos esta en la media nominal para peso. Todos los valores se encuentran dentro de los límites de especificación.

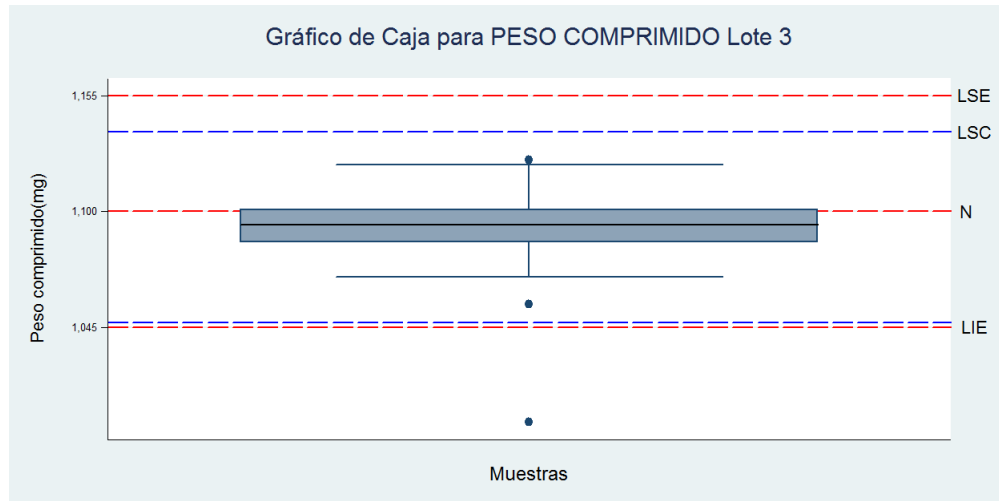


Figura 3.1.9: Gráfico de Caja para variable de calidad Peso Lote 3

- En cuanto a la Figura 3.1.9 existe mucha semejanza al primer lote, ya que existe asimetría entre los valores de peso y se reconoce un posible valor atípico que puede estar influyendo en el no cumplimiento de normalidad.

Gráfica de Control

Otra importante gráfica para poder realizar una observación detallada de los valores en estudio es la de control,

Antes de analizar los gráficos se describen las abreviaturas siguientes:

- LIE: Límite Inferior de Especificación.
- LSE: Límite Superior de Especificación.

- LIC: Límite Inferior de Control.

- LSC: Límite Superior de Control.

Se analizan los gráficos de control para los tres lotes a continuación:

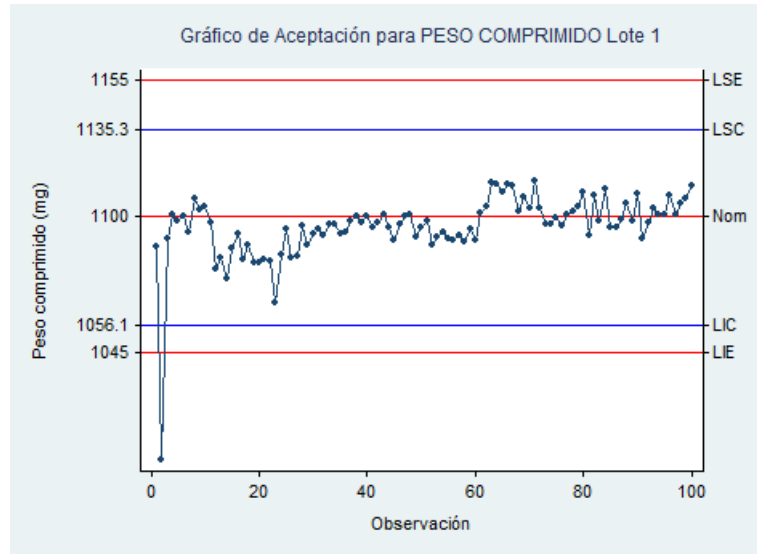


Figura 3.1.10: Gráfica de Control para variable de calidad Peso Lote 1

Se identifica Figura 3.1.10 al igual que en Figuras anteriores la posible presencia de un valor atípico que puede estar influyendo en el cálculo de prueba de normalidad. Se observa el resto de los datos dentro de los límites de especificación y de control.

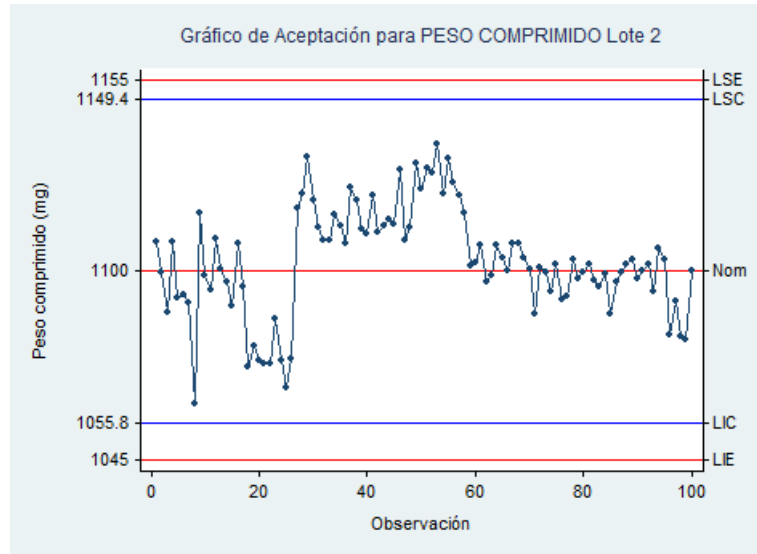


Figura 3.1.11: Gráfica de Control para variable de calidad Peso Lote 2

En la Figura 3.1.11 se observan todos los valores dentro de los límites de especificación y de control, una variabilidad controlada y una variabilidad controlada.

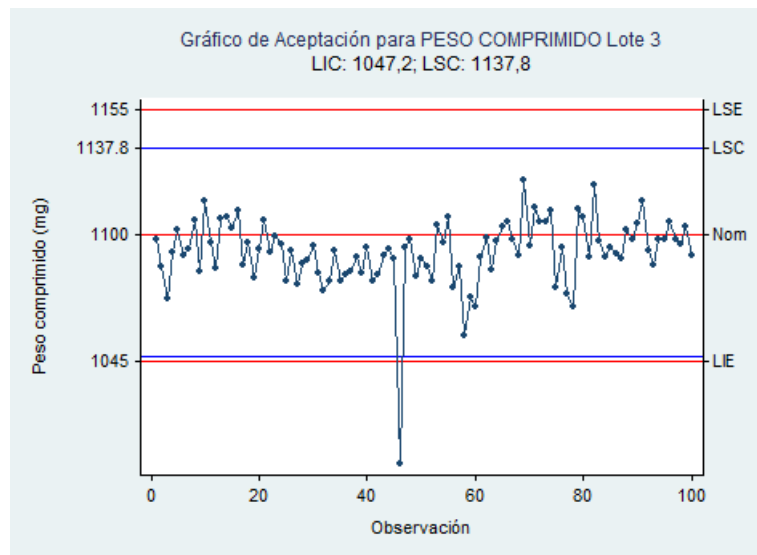


Figura 3.1.12: Gráfica de Control para variable de calidad Peso Lote 3

Se identifica en Figura 3.1.12 al igual que en estudios gráficos anteriores la posible presencia de un valor atípico que puede estar influyendo en el cálculo de prueba de normalidad.

Se observa el resto de los datos dentro de los límites de especificación y de control.

3.1.4. Transformación de variables

Luego de corroborar mediante pruebas y gráficos de que tanto en lote 1 como en lote 3 existe un proceso no controlado y también la posible existencia de valores atípicos, se realizan transformaciones para intentar normalizar los valores de Peso y obtener una menor variabilidad de las observaciones. Hoy en día la única transformación aceptada por el ISP es la de logaritmo natural, es por esto que se ofrecen más opciones de encontrar resultados favorables dentro de un proceso controlado.

Se muestra a continuación la Tabla con todas las transformaciones que se aplicaron, su valor-p mediante prueba de Shapiro-Wilk para observar si se obtuvo alguna mejora con respecto a los valores originales del atributo Peso:

Transformación	Lote 1	Lote 2	Lote 3
N	100	100	100
$\frac{1}{x}$	< 0,0001	0,1954	< 0,0001
ln(x)	< 0,0001	0,240	< 0,0001
\sqrt{x}	< 0,0001	0,263	< 0,0001
x^2	<0, 0001	0,330	< 0,0001
ln(y+m)	< 0,0001	0,141	< 0,0001
Box-Cox	< 0,0001	0,244	< 0,0001

Tabla 3.1.3: Valor-p para prueba de normalidad Shapiro-Wilk de variables transformadas

- Se observa en Tabla 3.1.3 que a pesar de aplicar las transformaciones no se obtienen resultados satisfactorios, tanto en lote 1 como en lote 3 no se logra cumplir con test de normalidad Shapiro-Wilk. Mediante la visualización gráfica en la sección anterior se pudo visualizar probables valores atípicos en estos dos lotes, estos pueden influir también en que las transformaciones no solucionen el problema de normalidad.

Luego de descartar este método se debe realizar el estudio sobre la detección de posibles datos atípicos.

3.1.5. Prueba para detección de valores atípicos

Para poder realizar una correcta detección de posibles valores atípicos y outliers se utiliza la prueba de Grubb's (Grubbs, 1950), donde su metodología corresponde a la detección de un valor a la vez y calcular el valor-p para determinar si se cumple con la dócima de hipótesis definida por:

H_0 : El dato en estudio es un valor atípico.

H_1 : El dato en estudio NO es un valor atípico.

Se presenta la Tabla para observar los valores atípicos detectados en cada uno de los tres lotes en estudio:

Lote 1	Lote 2	Lote 3
1001,1	-	1000,1

Tabla 3.1.4: Valores detectados como atípicos mediante test de Grubb's

Observando la Tabla 3.1.4 se detecta un valor en Lote 1 y un valor en Lote 3 como posibles valores atípicos, se presenta a continuación la tabla con la aplicación de prueba de Grubb's y su respectivo valor-p para definir si se rechaza la hipótesis nula H_0 .

Lote 1	Lote 2	Lote 3
< 0,0001	-	< 0,0001

Tabla 3.1.5: Valor-p para prueba de Grubb's

Por lo tanto, ya que se obtiene un valor-p menor a 0,05 no se está en condiciones de rechazar la hipótesis nula H_0 , es decir, no existe evidencia estadística suficiente para determinar que tanto para Lote 1 como para Lote 3 su valor detectado como atípico no lo sea. Luego de definir si es factible la eliminación de estos dos valores se debe proceder al cálculo nuevamente de Prueba de normalidad.

3.1.6. Capacidad de Proceso

La última etapa del análisis estadístico solicitado por el ente regulador es el cálculo del índice de capacidad de proceso, el cual tiene como condición fundamental para su utilización el cumplimiento de normalidad de los valores en estudio, en este caso para la variable de calidad Peso. Es por esto que antes de todo se calcula mediante prueba de Shapiro-Wilk los tres lotes de proceso sin los valores detectados por prueba de Grubb's.

Se muestra la Tabla a continuación:

Lote 1	Lote 2	Lote 3
0,1084	0,2865	0,9163

Tabla 3.1.6: Valor-p para prueba de normalidad Shapiro-Wilk

Observando la tabla 3.1.6 se destaca que ahora sin los valores detectados como atípicos se cumple con la condición de normalidad según valor-p determinado con la prueba de Shapiro-Wilk. Como consecuencia de lo anterior se puede realizar el cálculo de los índices de Capacidad de Proceso.

El cálculo se realiza mediante el programa estadístico Stata Corp y se presenta en la siguiente tabla:

Índice	Lote 1	Lote 2	Lote 3
Cp	2,0054	1,1759	1,5359
Cpk	1,8822	1,1211	1,3523

Tabla 3.1.7: Cálculo Índices de Capacidad de Proceso

- Se puede observar según los criterios de estudio de capacidad para el caso de capacidad Cp en Lote 1 y Lote 3 el proceso es muy capaz ya que estos son mayores a 1, 33. En tanto para Lote 2 el proceso es capaz pero requiere de una revisión constante.
- Para capacidad Cpk se puede observar que el proceso no está completamente centrado, esto ya que no se obtiene como resultado que $C_p = C_{pk}$, por lo tanto la tendencia de los valores de peso tienen una pequeña tendencia a alguno de los límites de especificación.

Finalmente se muestra a continuación el gráfico de control con todos los valores individuales a la variable de calidad Peso:

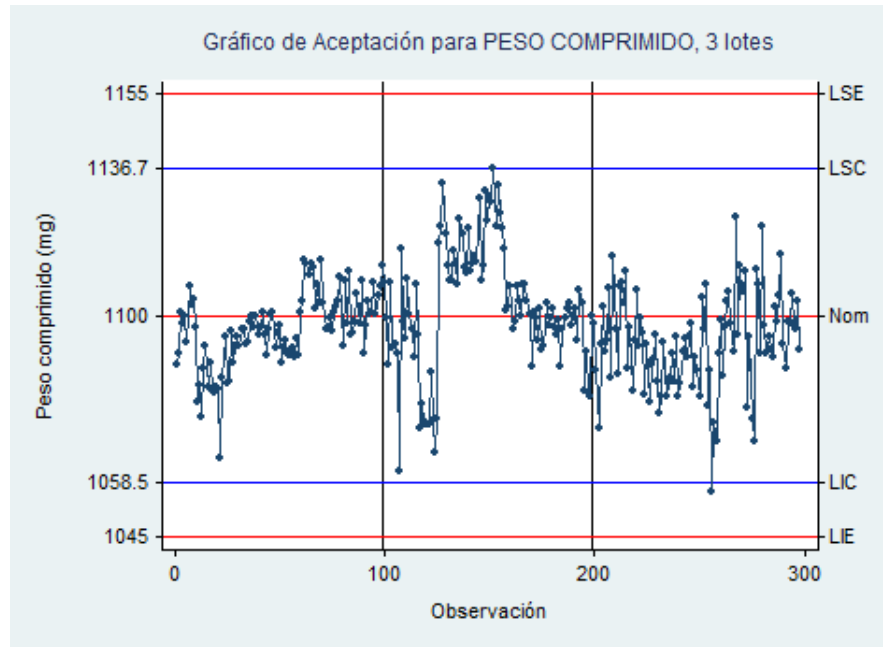


Figura 3.1.13: Gráfica de Control para variable de calidad Peso, 3 lotes

En el gráfico se observan todos los valores de la variable de calidad Peso en sus tres lotes, se puede identificar que todos los valores se encuentran dentro de los límites de especificación, existe un proceso controlado y también se localizan dentro de los límites estadísticos de control

3.1.7. Análisis de varianza ANOVA de una vía

Luego de comprobar que la variable de calidad Peso cumple con la condición de normalidad sin los valores atípicos detectados, es posible también analizar si existe semejanza entre las varianzas y las medias de cada uno de los tres lotes en estudio.

La hipótesis en estudio está definida como:

$$H_0 : \tau_1 = \tau_2 = \dots = \tau_v \text{ No existe diferencia significativa entre los tratamientos.}$$

$H_0 : \tau_1 \neq \tau_2 \neq \dots \neq \tau_v$ En al menos un tratamiento existe diferencia.

Se presenta a continuación con los resultados:

Fuente	Suma de cuadrados	G.L	Cuadrado medio	F	Valor-p
Intergrupo	4280,6807	2	13,66	< 0,0001	
Intragrupo	46219,2319	295	-	-	
Total	1,8822	297	-	-	

Tabla 3.1.8: Análisis de varianza para variable de calidad Peso

Prueba de Bartlett	Valor-p: < 0,0001
--------------------	-------------------

Tabla 3.1.9: Análisis mediante prueba de Bartlett para variable de calidad Peso

Para poder realizar correctamente Anova, es decir, estudio sobre las semejanzas de la media de la población, se debe cumplir con la condición de homocedasticidad mediante la prueba de Bartlett. Como se observa en las tablas, el valor-p para prueba de Bartlett es menor a 0,05, por lo tanto no existe evidencia estadística suficiente para indicar que no hay diferencia entre las varianzas de los lotes en estudio. Esto nos lleva por consiguiente a no poder estudiar un análisis de las medias.

3.1.8. Conclusión

A partir del objetivo de esta tesis, se aplicaron metodologías estadísticas innovadoras en el área de validación de procesos dentro de la industria farmacéutica, las cuales permitirán resolver los problemas que al día de hoy existen.

Se analizaron las pruebas de normalidad existentes, ya que en la actualidad la única aceptada por el organismo regulador ISP, es la prueba de Shapiro-Wilk. Se demostró que hay variadas pruebas igualmente válidas, para poder demostrar que los lotes de validación que se estudian se asemejan a una distribución normal, donde el objetivo principal es demostrar que un proceso de fabricación se encuentra controlado, alrededor de una media nominal establecida con anterioridad y una variabilidad simétrica con respecto a ella.

Luego se identificaron las pruebas para la detección y gestión de datos atípicos u outliers, los cuales, adquieren relevancia al momento de no cumplir con algunas de las pruebas de normalidad estudiadas. Dentro de la variable de calidad *Peso* que se utilizó para este estudio, se encontraron datos atípicos a través de la prueba de Grubb's y se determinó que en Lote 1 y Lote 3 influían en el proceso y se obtenían resultados desproporcionados. Por lo mismo se decidió eliminarlos -los datos atípicos- observándose mejores análisis para continuar con el objetivo solicitado por el organismo regulador, en este caso, demostrar un proceso robusto.

Se recopilaron los diferentes tipos de transformaciones para variables, donde se consiguieron interesantes opciones para poder obtener mejores resultados en el cumplimiento de la condición de normalidad. Luego al realizarlo sobre el estudio de la variable de calidad *Peso* se aplicaron las pruebas de normalidad, sin embargo según valor-p no hubieron mejores.

Para el caso del cálculo de los índices de capacidad C_p y C_{pk} , se lograron determinar sus parámetros de aceptación para identificar si el proceso es capaz y si se encuentra centrado con respecto a la media. Se realizó también la búsqueda del código para su cálculo en el programa estadístico Stata Corp.

Finalmente se elaboró una nota técnica con las metodologías ya existentes y también alternativas trabajadas en esta tesis, donde los pasos a seguir en la aplicación de esta guía técnica, dependerán del proceso que se estudiará en la práctica y todos los factores externos que están sometidos dentro de cualquier fabricación de los medicamentos en Laboratorio Andrómaco. Esta propuesta de

trabajo servirá para tener más opciones cuando se produzcan los problemas críticos que ocurren en la mayoría de los análisis que se realizan.

Como recomendación se debe seguir investigando sobre metodologías estadísticas de mejoramiento para cada una de las distintas etapas del lote de fabricación, donde los factores dentro de ella siempre van a producir variabilidad.

Capítulo 4

Anexos

4.1. Códigos Stata Corp

```
*****  
***** PESO NÚCLEO *****  
*****
```

```
clear all  
version 13
```

```
cd "E:\TESIS\resultados"
```

```
import excel using "resultados glifortex2", firstrow sheet("peso2") clear
```

```
capture log close  
log using "peso", text replace
```

```
/**** reshape *****/  
/*  
gen id=_n  
reshape long comp, i(id) j(COMP)  
rename comp peso
```

```
/** orden datos **/
```

```

sort lote COMP id

save "peso.dta", replace

/** orden datos **/

drop if peso == .

sort lote COMP id

by lote: sample 100, count

*/
/** estadísticas **/

tabstat peso, stat(n min q max mean sd cv range) col(stat) by(lote)

bys lote:ci peso

/** cp **/

pciest 1095.666 13.1899, f(1045) s(1155)
pciest 1102.563 15.59104, f(1045) s(1155)
pciest 1092.495 15.10468, f(1045) s(1155)

*****
/** Test de Normalidad **/
*****

/* Shapiro-Wilk */

bys lote: swilk peso

/* Shapiro-Francia */

bys lote: sfrancia peso

```

```

/* Skewness/Kurtosis */

sktest peso if lote == 1
sktest peso if lote == 2
sktest peso if lote == 3

/* Kolmogorov Smirnov */

ksmirnov peso if lote == 1
ksmirnov peso if lote == 2
ksmirnov peso if lote == 3

*****
/***** Test ANOVA *****/
*****

oneway peso lote
log close

/** Gráficos de capacidad **/
    /** LOTE 1 **/

#delimit ;
histogram peso if lote == 1, xaxis(1 2)
freq normal
xline(1045 1100 1155)
xline(1056.1 1135.3, lcolor(orange))
lcolor(blue) fcolor(gs16) xlabel(60(5)80)
xlabel(1045 "LIE" 1155 "LSE" 1100 "Nom" 1056.1 "LIC" 1135.3 "LSC", angle(0)
xlabel(1045 1100 1155, angle(0) axis(1) grid gmax)
title("Capabilidad de Proceso para PESO COMPRIMIDO Lote 1", size(big))
xtitle(" " "Peso comprimido (mg)")
subtitle("LIC: 1056,1; LSC: 1135,3" " " ", size(big))
ytittle("Frecuencia (unidad)" " " ") scale(0.9)
;
delimit cr

```

```

                /** LOTE 2 **/

#delimit ;
histogram peso if lote == 2, xaxis(1 2)
freq normal
xline(1045 1100 1155)
xline(1055.8 1149.4, lcolor(orange))
lcolor(blue) fcolor(gs16) xlabel(60(5)80)
xlabel(1045 "LIE" 1155 "LSE" 1100 "Nom" 1055.8 "LIC" 1149.4 "LSC", angle(0)
xlabel(1045 1100 1155, angle(0) axis(1) grid gmax)
title("Capabilidad de Proceso para PESO COMPRIMIDO Lote 2", size(big))
subtitle("LIC: 1055,8; LSC: 1149,4" " " ", size(big))
xtitle(" " "Peso comprimido (mg)")
ytittle("Frecuencia (unidad)" " " ") scale(0.8)
;
delimit cr

                /** LOTE 3 **/

#delimit ;
histogram peso if lote == 3, xaxis(1 2)
freq normal
xline(1045 1100 1155)
xline(1047.2 1137.8, lcolor(orange))
lcolor(blue) fcolor(gs16) xlabel(60(5)80)
xlabel(1045 "LIE" 1155 "LSE" 1100 "Nom" 1137.8 "LSC", angle(0) axis(2) grid gmax)
xlabel(1045 1100 1137.8 1155, angle(0) axis(1) grid gmax)
title("Capabilidad de Proceso para PESO COMPRIMIDO Lote 3", size(big))
subtitle("LIC: 1047,2; LSC: 1137,8" " " ", size(big))
xtitle(" " "Peso comprimido (mg)")
ytittle("Frecuencia (unidad)" " " ") scale(0.9)
;
delimit cr

/** Q-Q PLOT 3 lotes ***/

# delimit ;

```

```
qnorm peso if lote == 1 ,
title(" Gráfica Cuantil-Cuantil para PESO COMPRIMIDO Lote 1" " ")
subtitle("")
ytitle("Cuantiles muestrales de Peso comprimido" " " " ")
xtitle(" " "Cuantiles teórico normales")
scheme(vg_past)
msize(small)
;
# delimit cr

# delimit ;
qnorm peso if lote == 2,
title(" Gráfica Cuantil-Cuantil para PESO COMPRIMIDO Lote 2" " ")
subtitle("")
ytitle("Cuantiles muestrales de Peso comprimido" " " " ")
xtitle(" " "Cuantiles teórico normales")
scheme(vg_past)
msize(small)
;
# delimit cr

# delimit ;
qnorm peso if lote == 3,
title(" Gráfica Cuantil-Cuantil para PESO COMPRIMIDO Lote 3" " ")
subtitle("")
ytitle("Cuantiles muestrales de Peso comprimido" " " " ")
xtitle(" " "Cuantiles teóricos normales")
scheme(vg_past)
msize(small)
;
# delimit cr

/**** Carta de Control todos los datos ****/

drop id
bys lote: g id = _n
```

***** LOTE 1 *****

```
#delimit ;
tway connected peso id if lote == 1, yaxis(1 2)
sort yline(1045 1100 1155, lcolor(red))
yline(1056.1 1135.3, lcolor(blue))
ylabel(1045 "LIE" 1155 "LSE" 1100 "Nom" 1056.1 "LIC" 1135.3 "LSC", angle(0) axis(2) grid gmax)
ylabel(1045 1056.1 1100 1135.3 1155, angle(0) axis(1) grid gmax)
xtitle(" " "Observación") ytitle("Peso comprimido (mg)" " ")
title("Gráfico de Aceptación para PESO COMPRIMIDO Lote 1" " ", size(big))
scale(0.9)
;
```

***** LOTE 2 *****

```
#delimit ;
tway connected peso id if lote == 2, yaxis(1 2)
sort yline(1045 1100 1155, lcolor(red))
yline(1055.8 1149.4, lcolor(blue))
ylabel(1045 "LIE" 1155 "LSE" 1100 "Nom" 1055.8 "LIC" 1149.4 "LSC", angle(0) axis(2) grid gmax)
ylabel(1045 1055.8 1100 1149.4 1155, angle(0) axis(1) grid gmax)
xtitle(" " "Observación") ytitle("Peso comprimido (mg)" " ")
title("Gráfico de Aceptación para PESO COMPRIMIDO Lote 2" " ", size(big))
scale(0.9)
;
```

***** LOTE 3 *****

```
#delimit ;
tway connected peso id if lote == 3, yaxis(1 2)
sort yline(1045 1100 1155, lcolor(red))
yline(1047.2 1137.8, lcolor(blue))
ylabel(1045 "LIE" 1155 "LSE" 1100 "Nom" 1137.8 "LSC", angle(0) axis(2) grid gmax)
ylabel(1045 1100 1137.8 1155, angle(0) axis(1) grid gmax)
xtitle(" " "Observación") ytitle("Peso comprimido (mg)" " ")
```

```

title("Gráfico de Aceptación para PESO COMPRIMIDO Lote 3", size(big))
subtitle("LIC: 1047,2; LSC: 1137,8" " " ")
scale(0.9)
;
delimit cr

*****
/***** Gráficos Box-Plot *****/
*****

*****
***** Lote 1 *****/
*****

# delimit ;
graph box peso if lote==1,
/*Nominal*/
scheme(vg_past)
/*scheme(vg_s2m)*/
/*scheme(vg_rose)*/
text(1100 104 "N")
text(1045 105 "LIE")
text(1155 105 "LSE")
text(1135.3 105 "LSC")
text(1056.1 105 "LIC")
yline(1045 1100 1155, lwidth(Medium) lcolor(red) lpattern(longdash))
yline(1056.1 1135.3, lwidth(Medium) lcolor(blue) lpattern(longdash))
ylabel(1045(55)1155, angle(0))
medtype(cline)
medline(lwidth(Medium) lcolor(black))
b1title(" " "Muestras")
r2title(" " " ")
title("Gráfico de Caja para PESO COMPRIMIDO Lote 1" " " ")
yttitle("Peso comprimido (mg)" " " ")
xsize(20) ysize(10)
ylabel(,labsize(vsmall))
scale(1.0)

```

```

;
# delimiter cr

*****
***** Lote 2 *****
*****

# delimiter ;
graph box peso if lote==2,
scheme(vg_past)
text(1100 104 "N")
text(1045 105 "LIE")
text(1155 105 "LSE")
text(1149.4 105 "LSC")
text(1055.8 105 "LIC")
yline(1045 1100 1155, lwidth(Medium) lcolor(red) lpattern(longdash))
yline(1055.8 1149.4, lwidth(Medium) lcolor(blue) lpattern(longdash))
ylabel(1045(55)1155, angle(0))
medtype(cline)
medline(lwidth(Medium) lcolor(black))
bltitle(" " "Muestras")
title("Gráfico de Caja para PESO COMPRIMIDO Lote 2" " ")
yttitle("Peso comprimido (mg)" " ")
r2title(" " " ")
xsize(20) ysize(10)
ylabel(,labsize(vsmall))
scale(1.0)
;
# delimiter cr

*****
***** Lote 3 *****
*****

# delimiter ;
graph box peso if lote==3,
scheme(vg_past)

```

```

text(1100 104 "N")
text(1045 105 "LIE")
text(1155 105 "LSE")
text(1137.8 105 "LSC")
yline(1045 1100 1155, lwidth(Medium) lcolor(red) lpattern(longdash))
yline(1047.2 1137.8, lwidth(Medium) lcolor(blue) lpattern(longdash))
ylabel(1045(55)1155, angle(0))
medtype(cline)
medline(lwidth(Medium) lcolor(black))
bltitle(" " "Muestras")
title("Gráfico de Caja para PESO COMPRIMIDO Lote 3" " ")
ytittle("Peso comprimido(mg)" " ")
r2title(" " " ")
xsize(20) ysize(10)
ylabel(,labsize(vsmall))
scale(1.0)
;
# delimit cr

*****

clear all
version 13

cd "E:\TESIS\resultados"

import excel using "resultados glifortex2", firstrow sheet("peso2") clear

/** APLICACIÓN MÉTODOS PARA NORMALIDAD **/
/** TRANSFORMACIONES **/

g a = 1/peso
g b = ln(peso)
g c = sqrt(peso)
g d = (peso)^2
lnskew0 e = peso
bcskew0 f = peso

```

```
/** SHAPIRO-WILK **/  
  
bys lote: swilk a  
bys lote: swilk b  
bys lote: swilk c  
bys lote: swilk d  
bys lote: swilk e  
bys lote: swilk f  
  
/** ANOVA **/  
  
oneway a lote  
oneway b lote  
oneway c lote  
oneway d lote  
oneway e lote  
oneway f lote  
  
/** GRUBBS **/  
  
drop if peso == 96.9 & lote == 3  
*drop if peso == 98 & lote == 1  
*drop if peso == 98.2 & lote == 1  
*drop if peso == 102.7 & lote == 1  
*drop if peso == 98.5 & lote == 2  
*drop if peso == 103.8 & lote == 2  
drop if peso == 97.8 & lote == 3  
drop if peso == 102.8 & lote == 3  
  
/** Estadísticas **/  
  
tabstat peso, stat(n min q max mean sd cv range) col(stat) by(lote)  
  
/** Nuevamente test de normalidad **/  
  
bys lote: swilk peso
```

```

/*
bys lote: swilk a
bys lote: swilk b
bys lote: swilk c
bys lote: swilk d
bys lote: swilk e
bys lote: swilk f
*/
/** Nuevamente test de ANOVA **/

oneway peso lote
/*
oneway a lote
oneway b lote
oneway c lote
oneway d lote
oneway e lote
oneway f lote
*/
log close

*****

clear all
version 13

cd "E:\TESIS\resultados"

import excel using "resultados glifortex2", firstrow sheet("peso2") clear

/** orden datos **/

sort lote COMP id

save "peso.dta", replace

/** orden datos **/

```

```

drop if peso == .

*/
/** estadísticas **/

tabstat peso, stat(n min q max mean sd cv range) col(stat) by(lote)

drop if peso == 1001.1
drop if peso == 1000.1

bys lote:ci peso

/** cp **/

pciest 1096.621 9.141923, f(1045) s(1155)
pciest 1102.563 15.59104, f(1045) s(1155)
pciest 1093.428 11.93688, f(1045) s(1155)

***** Gráfico de Control para los 3 lotes *****

drop id
g id = _n

#delimit ;
tway connected peso id , yaxis(1 2)
sort yline(1045 1100 1155, lcolor(red))
ylines(1058.5 1136.7, lcolor(blue)) xlines(99 199, lcolor(black))
ylabel(99 199 298)
ylabel(1045 "LIE" 1155 "LSE" 1100 "Nom" 1058.5 "LIC" 1136.7 "LSC", angle(0))
ylabel(1045 1058.5 1100 1136.7 1155, angle(0) axis(1) grid gmax)
xtitle(" " "Observación") ytitle("Peso comprimido (mg)" " ")
title("Gráfico de Aceptación para PESO COMPRIMIDO, 3 lotes" " ", size(big))
scale(0.9)
;
delimit cr

```

4.2. Códigos R-project

```
library(nortest)
library(tseries)
library(outliers)

PES01<-read.table("PES01TESIS.txt",header=T)
PES02<-read.table("PES02TESIS.txt",header=T)
PES03<-read.table("PES03TESIS.txt",header=T)

attach(PES01)
attach(PES02)
attach(PES03)

PES01
PES02
PES03

### TEST DE NORMALIDAD ###

## SHAPIRO WILK ##

shapiro.test(PES01)
shapiro.test(PES02)
shapiro.test(PES03)

## SHAPIRO FRANCIA ##

sf.test(PES01)
sf.test(PES02)
sf.test(PES03)

## LILLIEFOR ##

lillie.test(PES01)
lillie.test(PES01)
lillie.test(PES01)
```

```
## KOLMOGOROV SMIRNOV ##

ks.test(B3BL1VH, "pnorm", mean(B3BL1H), sd(B3BL1VH))
ks.test(B3BL2VH, "pnorm", mean(B3BL2VH), sd(B3BL2VH))
ks.test(B3BL3VH, "pnorm", mean(B3BL3VH), sd(B3BL3VH))

## ANDERSON DARLING ##

ad.test(PESO1)
ad.test(PESO2)
ad.test(PESO3)

## PEARSON ##

pearson.test(PESO1)
pearson.test(PESO2)
pearson.test(PESO3)

## CRAMER-VON MISSES

cvm.test(PESO1)
cvm.test(PESO2)
cvm.test(PESO3)

## JARQUE BERA ##

jarque.bera.test(PESO1)
jarque.bera.test(PESO2)
jarque.bera.test(PESO3)

##### Datos atípicos #####

Lote 1
PESO1<-read.table("PESO1TESIS.txt", header=T)
attach(PESO1)
outlier(peso, opposite = FALSE, logical = FALSE)
```

```
grubbs.test(peso)
```

```
#####
```

```
Lote 2
```

```
PESO2<-read.table("PESO2TESIS.txt",header=T)
```

```
attach(PESO2)
```

```
outlier(peso, opposite = FALSE, logical = FALSE)
```

```
grubbs.test(peso)
```

```
#####
```

```
Lote 3
```

```
PESO3<-read.table("PESO3TESIS.txt",header=T)
```

```
attach(PESO3)
```

```
outlier(peso, opposite = FALSE, logical = FALSE)
```

```
grubbs.test(peso)
```

Capítulo 5

Bibliografía

- Conover, W.J. (1999). Practical Nonparametric Statistics. Third Edition, John Wiley Sons, Inc. New York, 428 – 433(6,1).
- Dixon, W.J., (1950), Analysis of extreme values: Annals of Mathematical Statistics, 21(4), 488–506.
- F. J. Tejedor Tejedor (1999). Análisis de varianza. Schaum. Madrid. La Muralla S.A.
- Grubbs, Frank E. (1950), Sample criteria for testing outlying observations, The Annals of Mathematical Statistics 21(1), 27 – 58.
- López de Maturana, P; Riveros, M (2000). El Concepto de Validación en los Procesos Farmacéuticos. Apuntes del curso de postítulo dictado por la Universidad de Chile.
- Mohd, N. Bee, Y. (2011). Power comparisons of Shapiro-Wilk, Kolmogorov-Smirnov, Lilliefors and Anderson-Darling tests. Journal of Statistical Modeling and Analytics. 21 – 33.
- Montgomery, D. (2005). Introducción al Control Estadístico de la Calidad. México. Limusa.
- Montgomery (2004). Introduction to Statistical Quality Control. Nueva York. John Wiley Sons, Inc.

- Layard, M. (1973), 'Robust Large-Sample Test for Homogeneity of Variances', Journal of the American Statistical Association. 68(341), 195–198.'
- Lilliefors, H.W. (1967). On the Kolmogorov-Smirnov Test for Normality with Mean and Variance Unknown. Journal of American Statistical Association, *Vol*,62, *No*,318, 399 – 402.
- RB Dean y WJ Dixon (1951), Estadísticas simplificado para pequeños números de observaciones. Anal. Chem., 23(4), 636 – 638.
- Peña. D (1994), Un contraste de normalidad basado en la transformación Box Cox. Estadística española, 33 – 46.
- Taylor, John R. (1997), An Introduction to Error Analysis. 2nd edition. Sausalito, California. University Science Books, 166 – 168.
- Decreto Supremo N°3 (2010); Reglamento del Sistema Nacional de Control de los productos farmacéuticos de uso humano.
ISPCH. Recuperdo de [http://www.ispch.cl/sites/default/files/u33/DTO %20N %C2 %B0 %203–2010 %20DIARIO %20OFICIAL.pdf](http://www.ispch.cl/sites/default/files/u33/DTO%20N%C2%B0%203-2010%20DIARIO%20OFICIAL.pdf).
- Food and Drug Administration. FDA. Recuperado de <http://www.fda.gov/>.
- Good Manufacturing Practice. Recuperado de http://www.who.int/biologicals/publications/trs/areas/vaccines/gmp/WHOTRS822_A1.pdf?ua=1.
[https : //www.overleaf.com/project/5c3a662df9b21a10fadb4748](https://www.overleaf.com/project/5c3a662df9b21a10fadb4748)
- Real Farmacopea Española, 2ª edición, Formas farmacéuticas. Recuperado de <http://www.ugr.es/adol-fina/asignaturas/formasfarmaceuticasRFE.pdf>.
- Valdiviezo, Martha, Fermín, José. Estudio comparativo de índices de capacidad de proceso con variables distribuidas no normales.

Recuperado de <http://www.redalyc.org/articulo.oa?id=81619984007>.

- Validación de métodos y determinación de la incertidumbre de la medición; Aspectos generales sobre la validación de métodos.

Recuperado de <http://www.ispch.cl/sites/default/files/documento/2010/12/guiatecnica1validaciondemetodos.pdf>.