

UNIVERSIDAD NACIONAL DE INGENIERIA
FACULTAD DE INGENIERIA ECONOMICA Y CCSS
Escuela Profesional de Ingeniería Estadística



**CONCEPCIÓN DE LA VARIABILIDAD EN EL ESTUDIANTE DE
LA CARRERA PROFESIONAL DE INGENIERÍA ESTADÍSTICA -
UNI: UN ESTUDIO EXPLORATORIO**

INFORME DE SUFICIENCIA

**Para obtener el Título de Licenciado en Estadística por la
Modalidad de Actualización de Conocimientos**

Elaborado por:

SILVIA LUZ PAJUELO ROJAS

LIMA - PERU

2011

A mis padres Benigno y Lucila, por
su apoyo y comprensión.

Resumen

La presente investigación tiene como objetivo estudiar la interiorización de la concepción de la variabilidad en el estudiante de la carrera de Ingeniería Estadística de la UNI. Para alcanzar este propósito se planteó la hipótesis de que el estudiante de la carrera profesional de ingeniería estadística de la UNI no tiene interiorizado el concepto de la variabilidad a lo largo de sus estudios universitarios. Se diseñó un cuestionario, basado en tres tipos de preguntas o test que influyen en la interiorización de la concepción de la variabilidad, el primero mide la capacidad del estudiante de evaluar la variabilidad en los gráficos, el segundo mide la capacidad de resolver problemas experimentales basados en la variabilidad y el tercero mide la capacidad de aplicar conceptos de variabilidad, hacer cálculos e interpretar resultados. Se tomó este cuestionario a una muestra aleatoria de 50 estudiantes, y se aplicó el modelo de diseño experimental de medidas repetidas. Se obtuvo como resultado que la influencia de los tres tipos de prueba es significativa (diferentes), concluyéndose que el estudiante de la carrera de ingeniería estadística no tiene interiorizado el concepto de la variabilidad.

Palabras clave.- Medidas repetidas, ANOVA de un factor, efectos principales variabilidad.

INDICE

RESUMEN	3
INTRODUCCIÓN	6
PROBLEMA DE LA INVESTIGACIÓN	8
CAPÍTULO I JUSTIFICACIÓN, OBJETIVOS E HIPÓTESIS	9
CAPÍTULO II ANTECEDENTES	12
CAPÍTULO III MARCO TEÓRICO	
3.1 Variabilidad	15
3.2 Visualización de la Variabilidad en los gráficos	17
3.3 Variabilidad en los problemas experimentales	24
3.4 Medidas de dispersión	26
3.5 Técnicas ideales para la investigación	29
3.6 Conceptos generales	30
3.7 Diseños con mediciones repetidas	47
3.8 Comparaciones por parejas (Efectos principales)	53
3.9 ANOVA de un factor	55
CAPÍTULO IV METODOLOGÍA	
4.1 Unidad de análisis	57
4.2 Población	57
4.3 Muestra	58
4.4 Instrumento de Recolección de Datos	59
4.5 Diseño del estudio para la hipótesis general	60
4.6 Diseño del estudio para la primera hipótesis específica	62
4.7 Diseño del estudio para la segunda hipótesis específica	63
CAPÍTULO V ANÁLISIS Y RESULTADOS	
5.1 Análisis de la hipótesis general	64
5.2 Supuestos del modelo de medidas repetidas	65
5.3 Análisis de varianza para la hipótesis general	68
5.4 Análisis de la primera hipótesis específica	69

5.5	Supuestos del ANOVA	70
5.6	Análisis de varianza para la primera hipótesis específica	72
5.7	Análisis de la segunda hipótesis específica	73
CONCLUSIONES		76
RECOMENDACIONES		77
BIBLIOGRAFÍA		78
ANEXO I: CUESTIONARIO		80

INDICE DE GRÁFICOS

Gráfico 3.1:	Áreas de los Proyectos de Dos Constructoras de Viviendas.	19
Gráfico 3.2:	Distribución de las Puntuaciones en un Examen en las Aulas A y B.	21
Gráfico 3.3:	Distribución de Barras del Tipo de Sangre de Tres Grupos Étnicos.	23

INDICE DE CUADROS

Cuadro 3.1:	Datos para el Diseño Unifactorial con Medidas Repetidas.	47
Cuadro 4.1:	Tamaños Poblacionales por Estrato.	57
Cuadro 4.2:	Tamaños de Muestras.	58
Cuadro 4.3:	Operacionalización de Variables para la Hipótesis General.	60
Cuadro 4.4:	Operacionalización de Variables para la 1ra. Hipótesis Específ.	62
Cuadro 5.1:	Test de Kolmogorov Smirnov (Hipótesis General).	66
Cuadro 5.2:	Test de Durbin Watson (Hipótesis General).	67
Cuadro 5.3:	Test de Esfericidad (Hipótesis General).	68
Cuadro 5.4:	Análisis de Varianza para la Hipótesis General.	69
Cuadro 5.5:	Test de Kolmogorov Smirnov (1ra. Hipótesis Específica).	70
Cuadro 5.6:	Test de Durbin Watson (1ra. Hipótesis Específica).	71
Cuadro 5.7:	Test de Levene (1ra. Hipótesis Específica).	72
Cuadro 5.8:	ANOVA de un Factor (1ra. Hipótesis Específica).	73
Cuadro 5.9:	Comparaciones por Pares (Componentes Principales).	75

INTRODUCCION

El presente trabajo tiene como objetivo general estudiar, a nivel exploratorio, las características de la conceptualización de la variabilidad que el estudiante de Ingeniería Estadística de la UNI ha construido a lo largo de sus estudios universitarios en su especialidad, y como objetivos específicos determinar cuantitativamente la influencia del avance de sus estudios universitarios en la interiorización del concepto de variabilidad; y además determinar cuantitativamente la influencia de la importancia que se da a los cálculos, en la interiorización del concepto de la variabilidad. Este estudio es importante porque permitirá a la escuela profesional de ingeniería estadística disponer de información exploratoria basada en las conclusiones del presente, lo que desde hace mucho tiempo los académicos perciben, de que sus estudiantes no tienen interiorizado la conceptualización de la variabilidad; además, de infundir directrices en el escenario curricular, y los ejes temáticos que circunscriben este contexto. Otro aspecto de importancia es que un profesional en estadística sin una interiorización del concepto de variabilidad, trasciende negativamente en la esencia de lo que es un estadístico, esto es, su naturaleza de investigador. Se ha encontrado muchas investigaciones relativas al tema de la variabilidad, como “Los Efectos de los Datos y Tipo de Gráfico en los Conceptos y Visualizaciones de la Variabilidad” de Cooper y Shore (2010). Esta investigación permitió un aporte importante al diseño del cuestionario del presente trabajo, porque estudia la conceptualización de la variabilidad en la comparación de los diferentes tipos de gráficos. Otras investigaciones están dirigidas a plantear la importancia de la Variabilidad en el campo de la Estadística, como indica Pfannkuch, (1999): *“La pieza central de la calidad y de*

las definiciones de la American Statistical Association (ASA) sobre el pensamiento estadístico es "variación" o "variabilidad". Cualquier discusión seria del pensamiento estadístico debe examinar el papel de la "variación". La terminología de "Variación" y el mensaje parecen haber surgido en una pequeña área de aplicación estadística, es decir, en el concepto de calidad, y su interiorización en otras áreas parece no trascender. Si la "variación" (como una gran fuente de incertidumbre) es de hecho una norma para todo proceso estadístico y se circunscribe alrededor de ella, entonces se necesita llegar a una concepción común de las estadísticas en términos de la "variación". Los tres primeros mensajes de la "variación" son: la variación es omnipresente, la variación puede tener graves consecuencias prácticas, y las estadísticas nos dan una forma de entender un mundo acosado por la variación. Mensajes posteriores se refieren a cómo las estadísticas tratan de hacer eso."

En el aspecto metodológico, para el propósito del estudio se consideró a todos los estudiantes regulares de la escuela profesional de ingeniería estadística de la FIECS UNI, inscritos en el periodo 2011 – I, haciendo una población de 298 estudiantes, y mediante un muestreo estratificado, se extrajo una muestra total de 50 estudiantes, a quienes se les tomó un cuestionario dividido en tres grupos secuenciales de preguntas. El primer grupo, que representa el primer nivel del factor, mide la interiorización de la concepción de la variabilidad que tienen los estudiantes al comparar la variabilidad en los gráficos; el segundo grupo, que representa el segundo nivel del factor, mide la interiorización de la concepción de la variabilidad que tienen los estudiantes de resolver problemas experimentales que implícitamente se relacionan con la variabilidad; el tercer grupo, que representa el tercer nivel del factor, mide la Interiorización de la concepción de la variabilidad que tienen los estudiantes de aplicar conceptos, hacer cálculos e interpretar resultados sobre la variabilidad. Se planteó como hipótesis principal de que el estudiante de la carrera profesional de ingeniería estadística – UNI no tiene interiorizado el concepto de la variabilidad a lo largo de sus estudios universitarios. Esta hipótesis se validó con la técnica de diseño experimental de medidas repetidas, determinando efectivamente la validez de ésta hipótesis, con un margen de error del 5%.

Problema de la investigación

El problema que se presenta en la enseñanza-aprendizaje de la variabilidad a lo largo de la carrera universitaria, es que se le da más importancia a los cálculos matemáticos para encontrar los valores de las mediciones formales, como son: rango, rango intercuartil, desviación estándar y coeficiente de variación y se le da poca relevancia a la conceptualización, comprensión e interpretación de estos resultados. Nuestra formación escolar y preuniversitaria, asentada en la perfección del cálculo, ha exceptuado el desarrollo del criterio y del análisis. Por el mismo contexto, cuando se pide comparar dos o más distribuciones los alumnos se centran en la búsqueda de algún valor que represente dichos grupos como son los valores de posición central (media, mediana o moda) y no toman en cuenta la dispersión de los datos o en el mejor de los casos, llegan a confundir la heterogeneidad de los datos con la variabilidad respecto a una medida de tendencia central (media). Se identifica como problema principal, el desconocimiento de las características de la conceptualización de la variabilidad en los estudiantes de la carrera profesional de ingeniería estadística de la UNI, a lo largo de sus estudios. También se identifican como problemas secundarios la carencia de métricas de la influencia del avance de los estudios universitarios o la importancia que se da a los cálculos en la interiorización del concepto de la variabilidad.

CAPITULO I

Justificación

La concepción de la variabilidad es un componente clave para el entendimiento de otros conceptos importantes de la estadística tales como la distribución de los datos y la inferencia estadística. Esto justifica la necesidad de conocer la situación de los alumnos frente a este tema, para implementar mejoras en la metodología de la enseñanza, en la medida que sea necesario, esto contribuirá al mejor entendimiento de las asignaturas especializadas y el buen desarrollo de toda la carrera de Estadística en la UNI.

La interiorización de la variabilidad como eje temático en la formación del estudiante de ingeniería estadística, permitirá aperturas de desarrollo y entendimiento en el estudiante, en estándares de procesos industriales basados en “seis sigma”.

OBJETIVOS

Objetivo general:

Estudiar, a nivel exploratorio, las características de la conceptualización de la variabilidad que el estudiante de estadística de la EPIES ha construido a lo largo de sus estudios universitarios en su especialidad.

Objetivos específicos:

1. Determinar cuantitativamente la influencia del avance de sus estudios universitarios en la interiorización del concepto de variabilidad en el estudiante de la carrera profesional de ingeniería estadística - UNI
2. Determinar cuantitativamente la influencia de la importancia que se da a los cálculos, en la interiorización del concepto de la variabilidad, en el estudiante de la carrera profesional de ingeniería estadística-UNI

FORMULACION DE LA HIPOTESIS

Para el presente estudio se plantea una hipótesis general y dos hipótesis específicas.

Hipótesis general:

El estudiante de la carrera profesional de ingeniería estadística – UNI no tiene interiorizado el concepto de la variabilidad a lo largo de sus estudios universitarios.

Hipótesis específicas:

1. El avance de sus estudios universitarios influye significativamente en la interiorización del concepto de variabilidad en el estudiante de la carrera profesional de ingeniería estadística – UNI.
2. La importancia que se da a los cálculos influye significativamente en la interiorización del concepto de la variabilidad, en el estudiante de la carrera profesional de ingeniería estadística-UNI.

CAPITULO II

Antecedentes

Carmen Batanero (2001) en su libro: "Didáctica de la Estadística", se refiere a un experimento hecho por Loosen y Cols en 1985 con 154 estudiantes de psicología a quienes se les preguntó cuál de los dos conjuntos A y B que están formados por medidas de longitudes de bloques de madera, presentaba mayor variabilidad. El 36% dijeron que B, 50% dijeron que A y el 14% dijeron que los dos conjuntos presentaban mayor variabilidad, con esto determinaron de que el concepto intuitivo de variabilidad se equipara al de no semejanza. Las mismas dificultades encontró Mevarech (1983) con alumnos universitarios concluyendo que el conjunto de datos junto con la operación de cálculo de la varianza tiene una estructura de grupo.

Un artículo de la Revista Iberoamericana de Educación que presenta un estudio titulado: "Identificación y Análisis de los errores cometidos por los alumnos de Estadística Descriptiva", plantea problemas con preguntas específicas de cálculo e interpretación escrito por Silvia Del Puerto y Silvia Seminara de la Universidad Tecnológica de Argentina y Claudia Minnaard de la Universidad Nacional de Lomas de Zamora de Argentina. El propósito de este trabajo fue identificar y analizar los errores más frecuentes cometidos por los alumnos universitarios en la asignatura Probabilidad y Estadística. Para ello se revisaron 519 exámenes parciales del periodo 2001-2004 de dicha asignatura y se fue indicando la frecuencia en que aparecían todos los posibles errores, de entre los cuales resaltaremos los que conciernen a nuestro tema de variabilidad, estos son: los alumnos confunden desviación estándar con

varianza, confunden variabilidad absoluta y variabilidad relativa y error en la fórmula al calcular la media o la desviación estándar que aparecen con poca frecuencia. Finalmente concluyeron que el seguimiento de los errores a través del tiempo muestran parámetros que pueden ayudar a diseñar actividades especiales que ayuden a los alumnos a que el proceso de aprendizaje sea más fructífero y más aún si se trata de errores relacionados con la variabilidad para que puedan ser solucionados y tratados con la importancia que amerita para asegurar un buen punto de partida para el estudio de la estadística en toda su magnitud.

Otro estudio importante, tiene como objetivo observar y describir la manera en que 6 profesores de secundaria modifican sus nociones de variabilidad estadística después de realizar actividades dentro de un contexto de distribuciones empíricas apoyadas con un software educativo de estadística. El modelo de clasificación utilizado en esta investigación, para explorar y analizar las nociones sobre variabilidad estadística, se basa en el modelo multimodal y la taxonomía SOLO desarrollados por Biggs y Collins (1991). El procedimiento consistió en 4 etapas: 1) los profesores contestaron preguntas relacionadas a distribuciones de datos generados por situaciones al azar en las que subyacen distribuciones básicas (uniforme y binomial), con la finalidad de obtener información acerca de sus nociones sobre variabilidad estadística. 2) los profesores desarrollaron una actividad guiada, cuyo objetivo era utilizar el software para resolver las tareas de las preguntas de la primera etapa mediante simulación estadística. 3) los profesores contestaron un cuestionario posterior, cuyo propósito era contrastar las respuestas obtenidas en las dos etapas anteriores, se esperaba crear un conflicto cognitivo que hiciera reflexionar al profesor sobre su comprensión de las nociones involucradas. Debían buscar justificaciones a las diferentes respuestas (antes y después del uso del software) y evaluar la variabilidad estadística. 4) por último, se trabajó el protocolo de la entrevista que se diseñó con el objetivo de estudiar el nivel de percepción de la variabilidad estadística que tienen los alumnos-docentes participantes en estudio. Luego de realizar una descripción detallada del proceso para cada uno de los profesores se llegó a la conclusión de que el uso de las herramientas tecnológicas crea la posibilidad de que los estudiantes entiendan de manera más profunda el sentido de la variabilidad y de la

predicción en estadística. Comprendieron que no se puede predecir eventos singulares sino que se deben dar intervalos en los que pueden caer los resultados; es decir, los estudiantes lograron relacionar la estructura con la aleatoriedad de los datos.

CAPITULO III

Marco teórico

3.1 Variabilidad:

El Pensamiento Estadístico se define como:

“Proceso del pensamiento, el cual reconoce que la Variación está alrededor de nosotros y está presente en todo lo que hacemos, todo el trabajo es una serie de procesos interconectados los que van identificando, caracterizando, cuantificando, controlando y reduciendo la variación para ofrecer oportunidades de mejora” Snee (1990, p.118)

La atención se centra entonces en la enseñanza de la variabilidad por ser una pieza fundamental del pensamiento estadístico y base para el buen desarrollo de todos los fundamentos estadísticos.

“La Variabilidad es... la esencia de la estadística como una disciplina y no basta la teoría para entenderlo. Tiene que ser experimentado” (Cobb, 1992)

“El término variabilidad se entenderá como la característica [que varía] de la entidad que es observable, y el término variación en el sentido de la descripción o la medición de esta característica. Por consiguiente, el "razonamiento acerca de la variación", se ocupará de los procesos cognitivos implicados en la descripción del fenómeno observado en

*situaciones que exponen la variabilidad, o la propensión para el cambio”
Shaughnessy (2004) (p. 202)*

La variación, es a menudo considerado como una medida de la cantidad de datos que se apartan de una medida de centro, como en la desviación estándar por ejemplo. La variación comprende más que una medida, ya que su conceptualización es un componente importante en el análisis de datos. En la conceptualización de la variación, se debe tomar en cuenta lo “que es” (su definición o fórmula), o “cómo” lo utilizan como una herramienta (procedimientos relacionados), pero también “para qué” es útil dentro de un contexto (el propósito).

"La incertidumbre y la variabilidad están estrechamente relacionados: porque existe una gran variabilidad, es que vivimos en incertidumbre, y porque no todo está determinado o cierto, hay variabilidad "(Bakker, 2004, p. 14).

Si bien nadie espera que todos los niños de cinco años de edad tengan la misma estatura, a menudo hay dificultad para comprender en qué medida las estaturas de los niños varían, y que la variabilidad de éstas es un mezcla de factores que lo explican (por ejemplo, las estaturas de los padres, nutrición) y del azar o factores inexplicables.

La variación presenta tres mensajes importantes:

3.1.1 La variación es omnipresente:

Es una realidad observable, está presente en todo lugar y en todas las cosas. La variabilidad afecta todos los aspectos de la vida y todas las cosas que observamos. Dos objetos fabricados no son idénticos, dos organismos no reaccionan de la misma forma. Lo antes mencionado se refiere solo a la variación real inherente en el sistema, a esto se debe agregar la variación inducida que se produce en el proceso de medición, en el muestreo o por accidente en la recolección de información.

3.1.2 Impacto en la práctica:

Una vez establecido que la variación está en todas partes, tenemos entonces que demostrar el importante impacto práctico de esta variación en la vida de la gente y en la forma de hacer negocios. Es la variación que hace que los resultados de las acciones sean impredecibles, que las preguntas de causa y efecto sean difíciles de resolver, y que también sean difíciles de descubrir los mecanismos.

“La variación es la razón por la cual la gente tiene que desarrollar métodos estadísticos sofisticados para filtrar los mensajes del ruido circundante” (Wild & Pfannkuch, 1999, pp. 235-236).

3.1.3 Solución al problema:

Las estadísticas nos dan una forma de entender un mundo acosado por la variación. Existen estudios que se refieren a cómo las estadísticas trata de plantear soluciones al respecto.

3.2 Visualización de la Variabilidad en los gráficos

Wild y Pfannkuch (1999) plantean que una de las formas básicas del razonamiento estadístico es el lenguaje gráfico. Como sostienen P. Arteaga, C. Batanero, C. Díaz, J. M. Contreras, (2009) que: el lenguaje gráfico tiene un papel esencial en la organización, descripción y análisis de datos, al ser un instrumento de transnumeración, que consiste en obtener una nueva información, al cambiar de un sistema de representación a otro. Por ejemplo, al pasar de una lista de datos desordenada a un histograma, se visualiza la moda y se percibe la simetría o asimetría de la distribución. La construcción e interpretación de gráficos estadísticos es también parte importante de la cultura estadística que Gal (2002, pg. 2) define como la unión de dos competencias relacionadas: a) Interpretar y evaluar críticamente la información estadística, los argumentos apoyados en datos o los fenómenos estocásticos que las personas pueden encontrar en diversos contextos, incluyendo los medios de comunicación, pero no limitándose a ellos, y b) discutir o comunicar sus opiniones respecto a tales

informaciones estadísticas cuando sea relevante (Gal, 2002, pp. 2-3). Según este autor, una persona culta debiera poder leer críticamente los gráficos estadísticos que encuentra en la prensa, Internet, medios de comunicación, y trabajo profesional. Esto supone no sólo la lectura literal del gráfico, sino identificar las tendencias y variabilidad de los datos, así como detectar los posibles errores conscientes o inconscientes que puedan distorsionar la información representada (Schield, 2006). Asimismo debiera conocer los convenios de construcción de los diferentes tipos de gráficos y ser capaz de construir correctamente un gráfico sencillo.

Según Cooper L. L., Shore F. S. (2010), se desarrolla la percepción visual de la variabilidad a través de la comparación de los gráficos y que esta variabilidad es percibida de manera diferente dependiendo del tipo de gráfico.

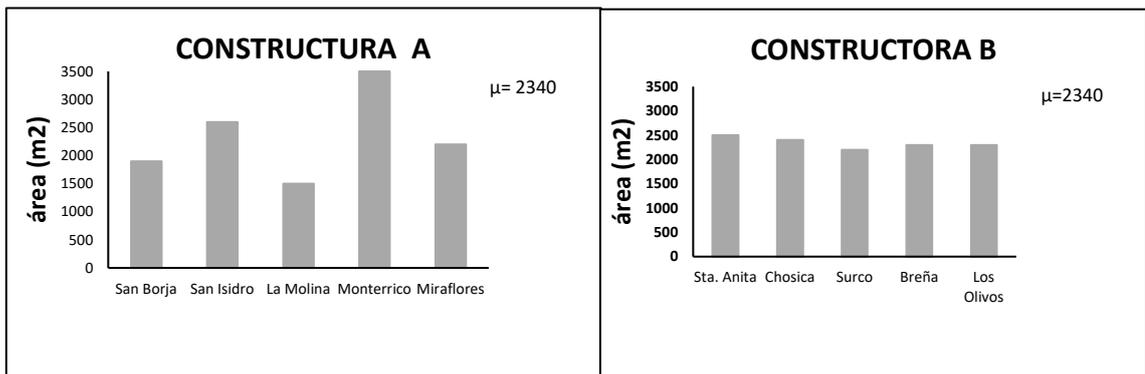
Se han elegido tres tipos de gráficos: valor de barras, histogramas, y diagrama de barras (para datos categóricos), específicamente porque: 1) comparten la similitud superficial del uso de barras, y sin embargo el papel de las barras en la representación de los datos es diferente en el tipo de gráfico, 2) se usan para datos cualitativos y cuantitativos; 3) son muestra de la distribución y la no distribución de los datos, y en consecuencia 4) los métodos de variabilidad para cada uno se juzga diferente.

3.2.1 Gráfico valor de barras

El gráfico valor de barras proporciona una ilustración de los datos numéricos en los que la altura de una barra corresponde a la magnitud de su valor correspondiente. La importancia del orden de las barras depende de si los valores a lo largo del eje horizontal son nominales u ordinales. Sólo, recientemente, se ha dado más atención en la distinción de estos gráficos, haciendo hincapié en su naturaleza pictórica, además que su estructura amerita una interpretación totalmente diferente del centro y la difusión. En los medios de comunicación populares e incluso en trabajos de investigación, los gráficos valor de barras se usan a menudo para comparar los datos de resumen, como medias o proporciones, para varios grupos.

En el gráfico 3.1, se ilustran los proyectos (áreas en metros cuadrados de los planos de construcción) de dos compañías constructoras de casas A y B. Ambas constructoras ofrecen cinco proyectos con áreas distintas. Las alturas de las barras son las áreas en metros cuadrados de cada proyecto. La media o promedio se puede interpretar como los metros cuadrados resultante si la superficie total se distribuye uniformemente entre los proyectos, por tanto, el metraje cuadrado medio, la altura a la que todas las barras estarían niveladas, es 2340 metros cuadrados para ambas constructoras.

Gráfico 3.1
ÁREAS DE LOS PROYECTOS DE DOS CONSTRUCTORAS DE
VIVIENDAS A Y B (En m²)



FUENTE: Journal of Statistics Education, Volume 18, Number 2 (2010)

La comparación de la variabilidad de los proyectos para las dos constructoras se hace al observar la magnitud de las diferencias entre las alturas de las barras y la línea horizontal de la media que debe trazarse sobre el gráfico, la mayor de estas diferencias muestra la mayor variación en los datos respecto a la media. Los cinco proyectos de la Constructora B, representado por las barras, todos rondan muy cerca de la línea horizontal media, mientras que las alturas de las barras, para la constructora A, los proyectos se desvían mucho de la línea media. Así, la constructora A ofrece los proyectos cuyas áreas son más variables que la constructora B.

3.2.2 Histogramas

El histograma muestra la distribución de los valores de una variable cuantitativa. Su uso de barras para mostrar la distribución requiere de un eje de frecuencias. Con el fin de interpretar los gráficos que muestran la distribución de una variable, el sentido gráfico debe ser lo suficientemente sofisticado como para comprender la interacción entre los valores de los datos y sus frecuencias absolutas o frecuencias relativas. Las dificultades se han observado en la investigación de Friel y Bright (1995) quienes postularon que la interpretación de gráficos que utilizan dos ejes (por ejemplo, histogramas, graficas de dispersión) puede presentar dificultades, "si las funciones de los ejes x e y a través de estos gráficos no están explícitamente reconocidos ". En su estudio sobre la escuela secundaria y estudiantes universitarios, delMas, et al., (2005) también especuló que los estudiantes no entienden realmente lo que representan los ejes.

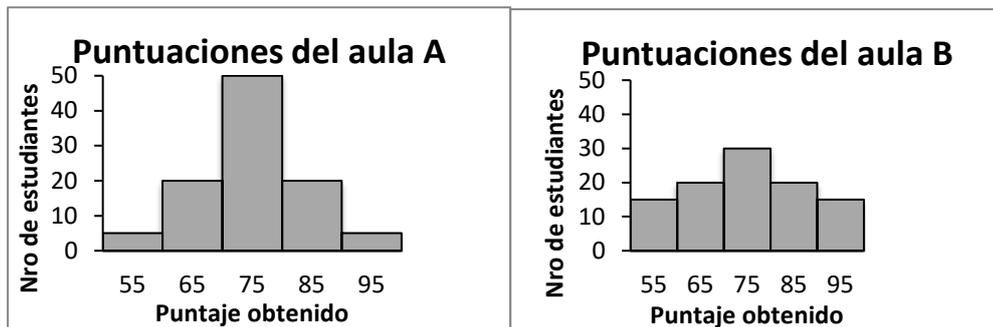
Los histogramas se clasifican por la forma, el centro, y la propagación. Las formas de distribución clásicas que se identifican son en forma de campana, simétrica, sesgada positiva y sesgada negativa. Si la forma es aproximadamente simétrica, la media es la medida preferida de centro, que puede ser visualizada como el punto de equilibrio a lo largo del eje horizontal del histograma, teniendo en cuenta el tamaño y la ubicación de las barras por encima de ella. Al igual que en la definición de centro, a juzgar por la variabilidad de un histograma, depende en gran medida de la forma. Se centra el análisis en cómo se percibe la variabilidad dentro de un grupo de histogramas a través de la comparación de las distribuciones simétricas y asimétricas aproximadamente, donde la variabilidad de los datos se juzgan por la cantidad de los datos que se concentran sobre, o hacia fuera de la media. Cuanto más próximos son los datos hacia la media, menor es la variabilidad, cuanto más alejados son los datos, mayor es la variabilidad. Esta idea de la magnitud de la desviación de los valores de los datos hacia la media es análoga al concepto de la variabilidad

en el caso del gráfico de valor de barras, aunque la interpretación visual es diferente para los dos.

En el gráfico 3.2 se muestran dos histogramas en forma de campana (puntuaciones obtenidas en un examen para las aulas A y B). La diferencia notable entre las gráficas es que un histograma tiene un pico pronunciado con colas estrechas, mientras que el otro tiene más barras de alturas similares. En histogramas en forma de campana, las colas más gruesas indican una mayor variabilidad, mientras que la agrupación hacia el centro indica poca variabilidad. En el caso del gráfico 3.2, las puntuaciones del examen del aula B son más diseminadas que los del aula A, por lo tanto es más variable.

Gráfico 3.2

DISTRIBUCIÓN DE LAS PUNTUACIONES EN UN EXAMEN EN LAS AULAS A Y B



FUENTE: Journal of Statistics Education, Volume 18, Number 2 (2010)

Cuando Cooper y Shore (2008) presentan estos histogramas a 186 estudiantes, 50% de los estudiantes indicaron que el histograma con el pico más alto y colas más estrechas fue más variable.

En estudios posteriores, se explicó por qué los estudiantes pensaron que el aula B tuvo menor variabilidad, esto fue porque las alturas de las barras de la clase B eran básicamente planas, mientras que había un pico y colas pequeñas en la clase A. Así, los estudiantes confundieron la variabilidad de las alturas de las barras para ser una indicación de la variabilidad en los datos. Esto es consistente con la idea errónea que fue señalado por Friel y Bright (1995) en que no

tuvieron en cuenta la interacción entre el eje de las frecuencias y los valores de datos en el eje horizontal. En pocas palabras, los estudiantes interpretaron los histogramas como si fueran gráficos de valor de barras. Este simple proceso de juzgar la variabilidad por la nivelación de la barras conduce a una correcta interpretación de gráficos de valor de barras y una interpretación incorrecta de histogramas, esto apunta a la necesidad de ayudar a los estudiantes a entender estos gráficos y luego ingresar a los gráficos más sofisticados mediante la yuxtaposición de los tipos de gráficos y la toma de las principales características explícitas.

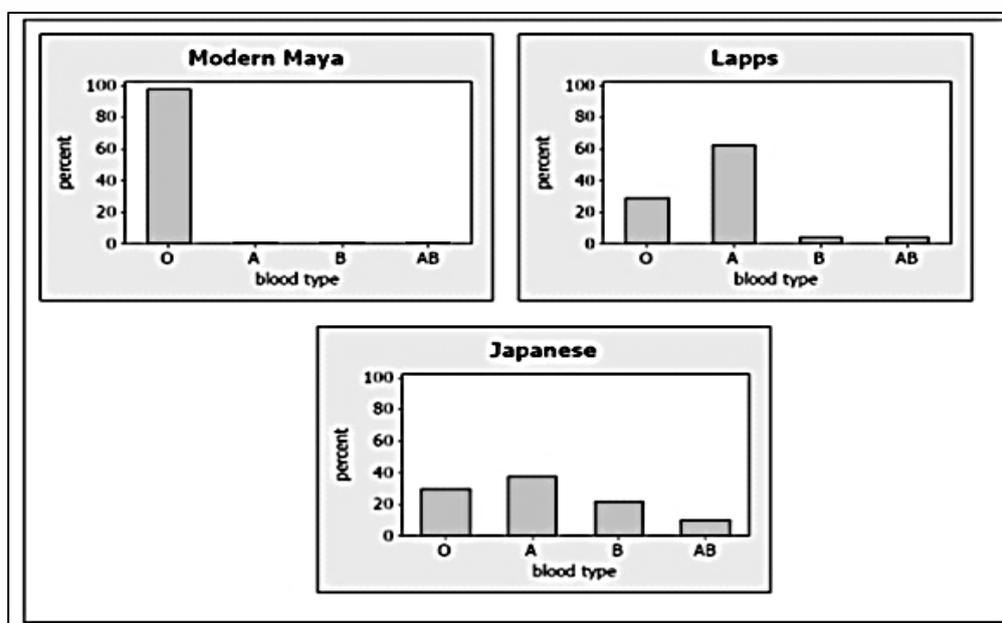
3.2.3 Diagrama de barras (variable cualitativa)

Los diagramas de barras se utilizan para presentar la distribución de los valores de una variable cualitativa y por lo tanto puede ser acertadamente denominado distribución de gráfico de barras, distinguiéndose de los gráficos de valor de barras. Las categorías de la variable suelen designarse en el eje horizontal. A diferencia de histogramas, el orden de las categorías en el eje horizontal no es importante, a menos que los datos sean ordinales. Se elaboran barras de la misma anchura para cada categoría de tal manera que su altura representa la frecuencia o frecuencia relativa de esa categoría. A diferencia del histograma, las barras no están juntas.

El gráfico 3.3 muestra las distribuciones de los tipos de sangre para tres grupos étnicos: los mayas modernos, los lapones escandinavos, y los japoneses. Casi no hay variabilidad en el tipo de sangre entre los mayas pues aproximadamente el 98% tienen sangre tipo "O". Hay más variabilidad en el tipo de sangre entre los lapones: aproximadamente el 63% tienen sangre tipo A, mientras que el 29% tienen sangre tipo "O". Los tipos de sangre entre los japoneses se distribuyen de manera más uniforme y por lo tanto son los más variables. Además, este es un caso de independencia en el orden de las barras.

Gráfico 3.3

DISTRIBUCIÓN DE BARRAS DEL TIPO DE SANGRE DE TRES GRUPOS ÉTNICOS, DISPUESTO DE MENOS A MÁS VARIABLES.



FUENTE: BloodBook.com 2006

El ejemplo anterior muestra cómo el concepto de variabilidad difiere para variables cualitativas y cuantitativas. En contraste con la variabilidad de los histogramas y los gráficos de valor de barras de los gráficos 3.1 y 3.2, la variabilidad en los datos cualitativos es juzgada por un concepto intuitivo de la variabilidad llamado “*unlikeability*” que plantea, ¿con qué frecuencia las observaciones difieren unos de otros? (Kader y Perry, 2007), siendo el punto principal simplemente si cada par de observaciones es diferente, en lugar de lo mucho que difieren en los datos cuantitativos. Por lo tanto, en la distribución del tipo de sangre de los Maya, el más pronunciado es el tipo "O" indica que prácticamente las observaciones no difieren unos de otros, es decir, hay poca variabilidad entre los tipos de sangre de los Maya. En menor medida, los lapones tienen más pronunciado el tipo A. Por el contrario, la distribución del tipo de sangre para los japoneses no tiene valor típico. Desde una perspectiva visual, cuanto más similares son

las alturas de las barras, los valores no se parecen y por lo tanto hay más variabilidad en los datos.

¿Cómo se compara esto con el ajuste cuantitativo en el que se enmarca la variabilidad como desviaciones de los valores de datos respecto del centro o media? La única medida adecuada del centro de datos cualitativos es el valor típico. La Moda de un conjunto de datos se identifica como la categoría que tiene la barra de mayor altura. La medida en que la frecuencia de la categoría modal se distingue de las frecuencias de otras categorías determina la idoneidad para referirse a la Moda como un valor típico. Para los conjuntos de datos que tienen un valor típico, si mayor es la frecuencia de la categoría modal en comparación con las frecuencias de otras categorías, significa que los datos son más parecidos, y por lo tanto hay menos variabilidad en los datos. Cuanto más difieren los datos de la categoría modal, hasta el punto extremo de que ya no hay una concentración de valores, hay más variabilidad en los datos.

El punto importante a destacar aquí es que hay conceptos diferentes de la variabilidad en función del tipo de datos y, posteriormente, diferentes métodos para evaluar la variabilidad.

3.3 Variabilidad en los problemas experimentales

El propósito de los problemas experimentales es explorar la noción intuitiva de la variabilidad estadística en situaciones de azar dentro de un contexto de distribuciones empíricas (distribución uniforme y binomial). La siguiente pregunta se tomó del cuestionario de exploración elaborado por Hernández (2005):

Imagina que lanzas un dado 60 veces. Llena la siguiente tabla escribiendo ¿cuántas veces crees que saldrá cada número?

Cara del dado	Nº de veces
1	
2	
3	
4	
5	
6	
Total	60

¿Por qué consideras este resultado?

Los resultados de este experimento deben presentar variabilidad en la respuesta. Si bien la distribución que subyace a este experimento es el Uniforme, esto no debe concluir en un resultado determinista. Pensar en los resultados de este experimento es verificar si el alumno reconoce la omnipresencia de la variabilidad.

“La base conceptual de la potencia pedagógica que obtenemos al hacer que los estudiantes realicen simulaciones es la conexión que se puede hacer entre la variación observada en los datos de la repetición de pruebas de un experimento, y los resultados que se esperan, basada en el conocimiento del espacio muestral subyacente o distribución de probabilidad.” (Shaughnessy & Ciancetta, 2002)

En el razonamiento informal de la variabilidad se pueden plantear hacer mediciones repetidas para descubrir dos tipos de variabilidad: 1) la variabilidad como un indicador de la diversidad (ejemplo A) y 2) la variabilidad como un error de medición (ejemplo B) (J.B Garfield, D. Ben-Zvi.).

Ejemplo A:

Se te proporciona una cinta métrica y te piden medir el contorno de las cabezas de 10 compañeros de tu salón de clase, escribe tus posibles resultados:

¿A qué se debe estos resultados?

Ejemplo B:

Se proporciona una cinta métrica a 10 de tus compañeros de tu salón de clase y se les piden medir el contorno de tu cabeza, escribe los posibles resultados:

¿A qué se debe estos resultados?

3.4 Medidas de dispersión

Una de las características de la distribución de los datos son las medidas de dispersión que proporcionan una medida del mayor o menor agrupamiento de los datos respecto a los valores de tendencia central.

3.4 1 Desviación Media Absoluta (DM):

Es la media aritmética de los valores absolutos de las desviaciones de las observaciones respecto de la media aritmética de éstas.

$$DM = \frac{\sum_{i=1}^n |x_i - \bar{x}|}{n}, \text{ si los datos no estan tabulados} \tag{3.1}$$

$$DM = \frac{\sum_{i=1}^m n_i |y_i - \bar{y}|}{n}, \text{ para datos tabulados} \tag{3.2}$$

Donde:

m = número de clases.

n_i = frecuencia absoluta de la clase i .

y_i = marca de clase o punto medio de la clase i .

3.4.2 Desviación Mediana Absoluta (DMe):

Es el promedio de los valores absolutos de las diferencias entre cada observación y la mediana de estas.

$$DMe = \frac{\sum_{i=1}^n |x_i - Me|}{n}, \text{ si los datos no estan tabulados} \tag{3.3}$$

$$DMe = \frac{\sum_{i=1}^m n_i |y_i - Me|}{n}, \text{ para datos tabulados} \quad (3.4)$$

3.4.3 Varianza Poblacional (σ^2)

Se define como la media aritmética del cuadrado de las desviaciones de las observaciones de la población respecto a su media μ .

$$\sigma^2 = \frac{\sum_{i=1}^N (x_i - \mu)^2}{N} \quad (3.5)$$

3.4.4 Varianza muestral (S^2)

Se define como la media aritmética del cuadrado de las desviaciones de las observaciones de la muestra con respecto de la media aritmética de los datos de dicha muestra.

$$S^2 = \frac{\sum_{i=1}^n (x_i - \bar{x})^2}{n-1}, \text{ para datos no tabulados} \quad (3.6)$$

$$S^2 = \frac{\sum_{i=1}^m n_i (y_i - \bar{y})^2}{n-1}, \text{ para datos no tabulados} \quad (3.7)$$

Un inconveniente al utilizar la varianza como medida de dispersión respecto de la media, es que no viene expresada en la misma unidad de medida que ésta, por tanto, no es posible la interpretación de esta cantidad. Por ello suele utilizarse en su lugar la Desviación Estándar.

3.4.5 Desviación Estándar (S)

Se define como la raíz cuadrada positiva de la varianza, en consecuencia esta medida sí permite la interpretación de la dispersión en la dimensión en que se miden los datos.

$$S = \sqrt{S^2} \quad (3.8)$$

La desviación estándar se puede utilizar para comparar la variabilidad de dos o más conjuntos de datos cuando las unidades en que están medidos y los promedios sean las mismas.

“... Yo no trato de eliminar la varianza de todo el campo de las estadísticas, pues tiene un papel muy importante y justificado de la estadística teórica y matemática con toda seguridad. Sin embargo para un curso introductorio yo apostaría, que, gustosamente los maestros tomarán acuerdos respecto a la fórmula de la varianza de la muestra y algún día se centrarán en la desviación estándar.”
(Deborah J. Rumsey)

Un ejemplo planteado por Carmen Batanero en su libro: “Didáctica de la Estadística”, dice:

Se presentan dos conjuntos de bloques de madera cuyas longitudes son:

A: 10; 20; 30; 40; 50; 60 cm.

B: 10; 10; 10; 60; 60; 60 cm.

¿Cuál de los dos conjuntos es más homogéneo?

3.4.6 Coeficiente de Variación (C.V)

Es una medida de la dispersión relativa de los datos e indica la fracción de la desviación estándar respecto de la media. Se aplica cuando se desea comparar la variabilidad de dos o más conjuntos de datos que tienen promedios y/o unidades de medidas diferentes. Se puede expresar en porcentajes para una mejor interpretación.

$$C.V = \frac{s}{\bar{x}} * 100\% \quad (3.9)$$

Ejemplo:

Se realiza un estudio sobre dos compañías que producen zapatillas Marca A y Marca B, con el objeto de determinar cuál de ellas tiene un mejor programa de control de calidad sobre el desgaste de su producto. Se tomó dos muestras por cada marca y se controló el número de horas de uso hasta que se detecta un desgaste significativo, se obtuvieron los siguientes resultados:

Marca A: 97, 83, 75, 82, 98, 65, 75 hrs.

Marca B: 78, 56, 87, 54, 89, 65 hrs.

¿Cuál de las dos marcas de zapatilla tiene un mejor programa de control de calidad sobre el desgaste?

3.5 Técnicas ideales para la investigación

Los modelos de análisis de varianza (ANOVA) con medidas repetidas (MR) sirven para estudiar el efecto de uno o más factores cuando al menos uno de ellos es un factor intra-sujeto. En los factores inter-sujetos o completamente aleatorizados a cada nivel del factor se le asigna o le corresponde un grupo diferente de sujetos. Por el contrario, un factor intra-sujetos o con medidas repetidas se caracteriza porque todos los niveles del factor se aplican a los mismos sujetos.

El diseño más simple de medidas repetidas consiste en medir dos variables en una misma muestra de sujetos. Los datos de este diseño se analizan con la prueba T para muestras relacionadas. Pero los diseños de medidas repetidas pueden tener más de dos medidas y más de un factor. Por ejemplo se considera una investigación diseñada para conocer la opinión de los consumidores sobre cinco productos alternativos o rivales. Se puede optar por seleccionar tantos grupos de sujetos como productos disponibles (cinco) y hacer que cada grupo opine sobre un solo producto. De esta manera, se tiene un diseño con un factor (tipo de producto, con cinco niveles) y tantos grupos de sujetos como niveles tienen el factor (cinco). Para analizar los datos de este diseño, se usa un ANOVA de un factor completamente aleatorizado.

En lugar de esto, se selecciona un único grupo de sujetos y pedirles que expresen su preferencia por cada uno de los cinco productos rivales. En ese caso, se sigue teniendo un diseño de un factor (el tipo de producto, con cinco niveles), pero un sólo grupo de sujetos que se hace pasar por las cinco condiciones definidas por los niveles del factor (se tiene a todos los sujetos opinando sobre todos los productos). Para analizar los datos de este diseño, se puede utilizar un ANOVA de un factor con medidas repetidas.

Las ventajas de los diseños de medidas repetidas son evidentes: requieren menos sujetos que un diseño completamente aleatorizado y permiten eliminar la variación residual debida a las diferencias entre los sujetos (pues se utilizan los mismos). Como contrapartida, es necesario vigilar algunos efectos atribuibles precisamente a la utilización de los mismos sujetos, tales como el efecto de arrastre, que ocurre cuando se administra una condición antes de que haya finalizado el efecto de otra administrada previamente; o el efecto del aprendizaje por la práctica que ocurre cuando las respuestas de los sujetos pueden mejorar con la repetición y, como consecuencia de ello, los tratamientos administrados en último lugar parecen más efectivos que los administrados en primer lugar, sin que haya diferencias reales entre ellos (en estos casos es importante controlar el orden de presentación de las condiciones). Obviamente, conviene conocer las ventajas e inconvenientes de estos diseños para decidir correctamente cuándo es apropiado utilizarlos.

3.6 Conceptos generales

3.6.1 Experimento

Se circunscribe a investigación es que establecen un conjunto particular de circunstancias, bajo un protocolo específico para observar y evaluar las implicaciones de las observaciones resultantes. El investigador determina y controla los protocolos de un experimento para evaluar y probar algo que en su mayor parte no se conoce hasta ese momento.

3.6.2 Experimento comparativo

Es el tipo de experimento que utilizan los investigadores en áreas como biología, medicina, agricultura, ingeniería, sicología y otras ciencias experimentales. El adjetivo comparativo implica que se establezca más de un conjunto de circunstancias en el experimento y que se comparen entre sí las respuestas a las diferentes circunstancias.

3.6.3 Tratamiento

Son el conjunto de circunstancias creadas para el experimento en respuesta a la hipótesis de investigación y son el centro de la misma. Entre los ejemplos de tratamientos se encuentran dietas de animales, producción de variedades de cultivos, temperaturas, tipos de suelo y cantidades de nutrientes.

En un estudio comparativo se usan dos o más tratamientos y se comparan sus efectos en el sujeto de estudio

3.6.4 Unidad experimental

Es la entidad física o el sujeto expuesto al tratamiento independientemente de otras unidades. La unidad experimental, una vez expuesta al tratamiento, constituye una sola réplica del tratamiento.

3.6.5 Error experimental

Describe la variación entre las unidades experimentales tratadas de manera idéntica e independiente. Los distintos orígenes del error experimental son:

3.6.5.1 La variación natural entre unidades experimentales.

3.6.5.2 La variabilidad en la medición de la respuesta.

3.6.5.3 La imposibilidad de reproducir las condiciones del tratamiento con exactitud de una unidad a otra.

3.6.5.4 La interacción de los tratamientos con las unidades experimentales.

3.6.5.5 Cualquier otro factor externo que influya en las características medidas.

3.6.6 Estudios por observación comparativos

Son aquellos para los que se desearía hacer un experimento, pero no es posible por razones prácticas o éticas.

El investigador tiene en mente condiciones o tratamientos que tienen efectos causales en sujetos para los que no es posible efectuar

experimentos hasta obtener respuestas. Los investigadores de las ciencias sociales, ecología, vida silvestre, vida marina y otros recursos naturales, a menudo han de llevar a cabo estudios por observación en vez de la experimentación directa. Tal vez la unidad básica de estudio en la investigación sean sujetos humanos, animales individuales, hábitats u otros microcosmos, que tienen el papel de unidad experimental en los experimentos diseñados.

Los sujetos se autoseleccionan en grupos identificables, o sencillamente existen en sus circunstancias particulares; estos grupos o circunstancias se usan como clasificaciones del tratamiento en el estudio por observación. Por el contrario, el investigador asigna los tratamientos para las unidades experimentales del experimento diseñado.

3.6.7 La hipótesis de investigación genera el diseño de los tratamientos

La hipótesis de investigación establece un conjunto de circunstancias y sus consecuencias.

Los tratamientos son una creación de las circunstancias para el experimento. Así, es importante identificar los tratamientos con el papel que cada uno tiene en la evaluación de la hipótesis de investigación. Si no se logra delinear con claridad esta hipótesis y el objetivo del estudio, puede haber dificultades en la selección de los tratamientos y experimentos sin éxito.

3.6.8 Relación entre los experimentos y la hipótesis

Cuando se eligen los tratamientos adecuados en respuesta a una hipótesis de investigación, es posible comprender mejor los mecanismos fundamentales, ya sean físicos, químicos, biológicos o sociales. En algunos casos, el objetivo puede ser "escoger al ganador" para encontrar un tratamiento que proporcione la respuesta deseada; en otros casos, el experimento se usa para aclarar los mecanismos básicos asociados con los tratamientos, en términos de

su influencia sobre las variables de respuesta medidas. En este último caso, una hipótesis sólida motiva la selección de los tratamientos.

Es forzoso que el investigador se asegure de que los tratamientos elegidos concuerden con la hipótesis de investigación. Puede ser suficiente y por consiguiente más sencillo diseñar un estudio sólo para descubrir el mejor tratamiento. Sin embargo, con poco esfuerzo adicional puede obtenerse aún más información fundamental del experimento en respuesta a la hipótesis de investigación.

3.6.9 Los tratamientos de control son un punto de referencia

El tratamiento de control es un punto necesario para evaluar el efecto de los tratamientos experimentales; existen diversas circunstancias en las que el tratamiento de control es útil y necesario. Las condiciones del experimento pueden ser un obstáculo para la efectividad de tratamientos experimentales que, en general, han sido efectivos. Un control al que no se da tratamiento revelará las condiciones en las que se efectuó el experimento.

Por ejemplo, los fertilizantes con nitrógeno suelen ser efectivos, pero no producirán respuestas en campos con alta fertilidad. Un control de fertilizante sin nitrógeno señalará las condiciones básicas de fertilidad del experimento.

En ocasiones, los tratamientos requieren la manipulación de las unidades experimentales o sujetos cuando esa sola manipulación puede producir una respuesta.

En estos casos, los controles placebo establecen una base para la efectividad del tratamiento. La unidad o sujeto placebo se procesa de la misma manera que las unidades en tratamiento, pero sin incluir en su protocolo el tratamiento activo.

Uno de los experimentos de salud más famosos, las pruebas de campo en 1954 de la vacuna Salk contra la poliomielitis, usó controles placebo o inocuos en cerca de la mitad de las áreas de prueba en Estados Unidos. El placebo se preparó con una apariencia igual a la de la vacuna, pero sin la actividad antigénica. Los sujetos placebo se

inocularon de la misma forma que los sujetos que recibieron la vacuna (Tanur et al., 1978).

Por último, el control puede representar una práctica normal con la que se puede comparar el método experimental. En algunas situaciones es necesario incluir dos tipos de control distintos. Por ejemplo, el no tratamiento y el tratamiento placebo pueden indicar el efecto de manipular la unidad experimental en ausencia de tratamiento.

3.6.10 El diseño de tratamientos con múltiples factores amplía las inferencias

En el artículo "The Arrangement of Field Experiments" (Arreglo de los experimentos de campo), Fisher (1926) observó que el proverbio más repetido respecto a los experimentos de campo era el que decía:

"Debemos hacer algunas preguntas a la naturaleza, y de preferencia, una a la vez."

Fisher estaba convencido de que se trataba de un punto de vista equivocado. A este respecto escribió: *"la naturaleza... responderá mejor a un cuestionario concebido lógicamente y cuidadosamente; en realidad, si le planteamos una sola pregunta, se rehusará a responder hasta que se analice algún otro tema"*.

Él entendía que en los sistemas naturales no se sabe si la influencia de un tratamiento es independiente de otra o si su influencia se relaciona con la variación de otros tratamientos. En consecuencia, las condiciones en las que se comparan los tratamientos pueden ser aspectos importantes del diseño.

3.6.11 Control local de errores experimentales

Los principales objetivos de la mayoría de los experimentos son las comparaciones claras y exactas entre tratamientos a través de un conjunto apropiado de condiciones.

Estos objetivos requieren estimaciones precisas de las medias y poderosas pruebas estadísticas; al reducir la varianza del error experimental aumenta la posibilidad de lograrlos. El control local

describe las acciones que emplea un investigador para reducir o controlar el error experimental, incrementar la exactitud de las observaciones y establecer la base de la inferencia de un estudio.

El investigador controla: 1) técnica, 2) selección de unidades experimentales, 3) definición de bloques o aseguramiento de la uniformidad de información en todos los tratamientos, 4) selección del diseño experimental y 5) medición de covarianzas.

3.6.12 La técnica afecta la variación y el sesgo

Si las tareas experimentales se realizan sin una planeación cuidadosa, las observaciones mostrarán un incremento en la variación. Las técnicas incluyen tareas sencillas como una medición exacta, la preparación de medios, obtención de soluciones o calibración de instrumentos. En un nivel más complejo, el investigador puede contar con varios métodos o instrumentos de laboratorio para medir propiedades químicas o físicas. Los métodos pueden variar en exactitud, precisión y alcance de aplicación. El investigador debe elegir el método o instrumento que le proporcione las observaciones más precisas dentro de los recursos presupuestados.

Cuando la técnica tiene un efecto adverso en la precisión, las varianzas estimadas del error experimental se incrementan de manera innecesaria. Las observaciones lejanas de la media, causadas por el registro de errores o por condiciones ambientales extremas, pueden aumentar la variación. Cualquiera que sea la causa, el investigador debe decidir si incluir o no estas observaciones en el análisis.

Por otro lado, tal vez no exista un patrón discernible, en las observaciones, distinto a un incremento general en su variabilidad. Este tipo de aumento en la variación puede señalar una técnica equivocada en el curso del experimento. El investigador deberá verificar la selección de las unidades experimentales para el estudio, el protocolo del tratamiento, las técnicas de medición y el personal, en busca de fuentes de error y luego intentar hacer los ajustes necesarios.

Las técnicas pobres pueden afectar la exactitud de las observaciones y sesgar los resultados. Más aún, la variación introducida en las observaciones por las técnicas pobres, no necesariamente es aleatoria y por lo tanto, no está sujeta a las mismas leyes de probabilidad que se asocian con la inferencia estadística.

La aplicación uniforme de los tratamientos durante el experimento aumenta la probabilidad de mediciones no sesgadas de sus efectos. Por ejemplo, se requiere la ingestión de cantidades uniformes de alimento para medir con exactitud las diferencias entre las dietas de los animales. Es necesario aplicar cantidades uniformes de fertilizante a las parcelas para medir con precisión las diferencias en la cosecha debidas a éste.

3.6.13 Selección de unidades experimentales uniformes

Las unidades experimentales heterogéneas producen valores grandes en la varianza del error experimental. La comparación precisa entre los tratamientos requiere la elección de unidades experimentales uniformes para reducir el error experimental.

Sin embargo, una selección demasiado restringida puede producir condiciones de uniformidad artificiales. Un conjunto estrecho de condiciones restringe la base de inferencia del estudio. Entonces, para asegurar una confiabilidad razonable de las conclusiones del experimento, es deseable que las unidades representen una variedad suficiente de condiciones, sin que aumente innecesariamente la heterogeneidad de las unidades experimentales.

La naturaleza del experimento señala el equilibrio entre la variedad de las condiciones y la uniformidad de las unidades. Por ejemplo, la selección de plantas en un estudio de su crecimiento se puede examinar en el conjunto de condiciones en el cual se espera cultivar las variedades; así, las condiciones pueden ser muy diversas. Si las selecciones se examinan en varios sitios lejanos entre sí, la uniformidad en cada lugar cobra importancia. La uniformidad de las unidades en un experimento con vacas lecheras requiere elegir vacas

de la misma cría, en la misma etapa de lactancia y con un número similar de lactancias.

3.6.14 La segregación en bloques para reducir la variación del error experimental

Fisher (1926) aseguró que no había que desistir de alguna ventaja para obtener una estimación válida del error, pero que eran necesarias dos cosas:

3.6.14.1 Hacer una distinción clara entre las componentes del error que debían eliminarse en el campo y las que no debían eliminarse.

3.6.14.2 Modificar el proceso estadístico de la estimación del error de manera que los componentes del error que se eliminaron en el campo sean igualmente eliminados en el laboratorio estadístico.

El bloqueo proporciona control local del ambiente para reducir el error experimental. Las unidades experimentales se agrupan de manera que su variabilidad dentro de los grupos sea menor que entre las unidades antes de agruparlas. La práctica de definir bloques o agrupar las unidades experimentales en conjuntos homogéneos, va de la mano con la selección de las unidades experimentales para tener uniformidad. Los tratamientos se comparan entre sí dentro de los grupos de unidades en un entorno más uniforme y las diferencias entre ellos no se confunden con las grandes discrepancias entre las unidades experimentales. En el análisis estadístico es posible separar el error experimental de la variabilidad asociada con las diferencias del entorno entre los grupos de unidades.

Las unidades experimentales se definen en bloques como grupos de unidades similares, con base en un factor o factores que se espera o se sabe que tienen alguna relación con la variable de respuesta o con la medición que se supone responde de manera diferente a los diversos tratamientos.

3.6.15 Cuatro importantes criterios para definir bloques

Los cuatro criterios que se usan con más frecuencia para definir bloques de unidades experimentales son: 1) proximidad (parcelas vecinas), 2) características físicas (edad o peso), 3) tiempo, 4) administración de tareas en el experimento.

La práctica de definir bloques clásicos se originó en los experimentos agrícolas; en los que las parcelas contiguas pertenecían a un grupo y cada uno de los tratamientos se asignaba a una parcela de ese grupo. Después, se usaba el segundo grupo de parcelas contiguas de la misma manera, y así sucesivamente, hasta tener un diseño de bloques completos.

El motivo principal para este tipo de bloqueo es que las parcelas cercanas en campos de cultivo se parecen más que las separadas por mayores distancias.

Algunos patrones de variabilidad pueden requerir arreglos diferentes para reducir el error experimental.

La definición de bloques se usa para dividir el experimento en unidades de tamaño razonable y administrar de manera uniforme el tiempo o las tareas.

3.6.16 Una demostración de la reducción de la varianza mediante la definición de bloques

Una prueba de uniformidad muestra la posible efectividad que pueden tener los bloques en la reducción de la varianza dentro de un estudio de investigación.

En esencia, la prueba de uniformidad es un experimento en el que las unidades experimentales se miden sin someterlas a un tratamiento. Por ejemplo, una prueba de uniformidad clásica en agricultura es un campo de trigo de la misma variedad, dividido en parcelas con las mismas dimensiones. Se mide la cosecha de trigo en cada parcela. Como la variación en los campos agrícolas normalmente ocurre en gradientes, se determina qué grupos de parcelas adyacentes tienen la menor varianza. De manera similar, en la medida de interés o alguna variable conocida que se sabe que tiene una relación estrecha con

esa medida, la observación de la línea base previa a la aplicación de tratamientos es equivalente a una prueba de uniformidad para propósitos de bloqueo.

3.6.17 Estrategias de asignación por compatibilidad

El agrupamiento de las unidades a menudo usa estrategias de compatibilidad para equiparar unidades similares. Los sujetos o unidades se eligen para cada tratamiento por su similitud en todos los factores que pueden influir. Cada variable que pueda influir en el valor de las características a observar en los sujetos es un candidato como variable a controlar. Los sujetos asignados tienen valores comunes para las variables a controlar, con la excepción del tratamiento que se administra. Durante el diseño del estudio, las estrategias para asignar intentan lograr que los sujetos sean similares en todos los factores que pueden sesgar en exceso la comparación de los tratamientos. La asignación por compatibilidad uno a uno y compatibilidad por clase son dos estrategias generales empleadas para la elección de unidades o sujetos.

Con la estrategia por compatibilidad uno a uno (por parejas), para cada sujeto de cada tratamiento se identifica parejas con sujetos que tienen los mismos valores de las variables a controlar y se asignan a cada uno de los otros tratamientos. Los valores de las variables a controlar pueden elegirse como:

1) Valores exactos o 2) valores calibrados. Es posible encontrar valores exactos con sujetos humanos en variables como género, profesión, uso del cinturón de seguridad y nivel de educación. Los valores exactos en un estudio del desempeño del tránsito en tramos de carretera son factibles en variables como número de carriles, ancho de los mismos, presencia o ausencia de carril central y límites de velocidad.

Los valores calibrados permiten cierta tolerancia en los valores de las variables comparadas. Es posible que un estudio sobre la manipulación de un ecosistema boscoso requiera que los sitios sean compatibles en composición de especies, pendiente y aspecto. Es

difícil lograr la compatibilidad exacta entre estos factores. Sin embargo, quizá se consigan sitios con valores similares para algunas o todas las variables de control comparadas, al grado de que la variación no tenga efectos serios en la comparación de los tratamientos.

Una compatibilidad por clase se puede lograr a través de una estrategia de: 1) frecuencias o de 2) medias. El método de frecuencia estratifica las unidades en grupos con base en las variables de control. Supongamos que la edad es una variable de influencia potencial en un estudio con seres humanos. Estos sujetos se pueden estratificar a partir de una distribución de frecuencias para sus edades, de manera que haya un número suficiente de sujetos en cada estrato de edad para asignarlos a todos los tratamientos.

La estrategia de compatibilidad por clase basada en las medias agrupa a los sujetos o unidades de manera que tengan los mismos valores promedio de las variables de control en cada grupo de tratamiento.

La naturaleza de las investigaciones señala la estrategia de compatibilidad más efectiva y si ésta es el protocolo deseable. Los detalles de los métodos de compatibilidad, sus ventajas y desventajas para estudios por observación comparativos, se pueden encontrar en Cochran (1983) y Fleiss (1981).

3.6.18 El diseño experimental asigna los tratamientos

El diseño experimental es el arreglo de las unidades experimentales utilizado para controlar el error experimental, a la vez que acomoda los tratamientos. Existe en la literatura una amplia variedad de arreglos diseñados para controlar el error experimental y se observa una tendencia natural a diseñar los experimentos de acuerdo con diseños ya existentes. Pero desarrollar un diseño de experimento que satisfaga la demanda del experimento que se está realizando es una actitud más adecuada.

El logro de la máxima información, precisión y exactitud en los resultados, junto con el uso más eficiente de los recursos existentes,

es un principio a seguir en la elección del diseño adecuado del experimento.

3.6.19 Replicar para obtener experimentos válidos

La comunidad científica considera la réplica de un experimento como el primer requisito para obtener resultados experimentales válidos. La réplica implica una repetición independiente del experimento básico. Dicho de manera más específica, cada tratamiento se aplica de manera independiente a dos o más unidades experimentales.

Existen diversas razones para hacer réplicas de un experimento, las más notables son:

- Demuestra que se pueden reproducir los resultados, al menos bajo las condiciones experimentales actuales.
- Proporciona cierto grado de seguridad contra resultados anormales en el experimento, debidos a accidentes no previstos.
- Proporciona las medias para estimar la varianza del error experimental. Aun cuando la experimentación previa proporcione estimaciones de la varianza, la estimación a partir del experimento en curso puede ser más exacta porque refleja el comportamiento actual de las observaciones.
- Proporciona la posibilidad de aumentar la precisión en la estimación de las medias de los tratamientos. Al incrementar las réplicas r , disminuye $s_{\bar{y}}^2 = \frac{s^2}{r}$ lo que aumenta la precisión de \bar{y} .

3.6.20 Las unidades de observación y las experimentales pueden ser claramente distintas

La unidad de observación puede no ser equivalente a la unidad experimental. La primera puede ser una muestra de la última, como muestras individuales de plantas de una parcela o muestras del plasma de un sujeto.

La varianza en las observaciones de las unidades experimentales es la varianza del error experimental; es una medida válida de la

variación entre las unidades experimentales que recibieron tratamiento de manera independiente.

Muchas veces la varianza entre las observaciones múltiples de la misma unidad experimental se usa equivocadamente como medida del error experimental para comparar los tratamientos.

3.6.21 ¿Cuántas réplicas?

El número de réplicas en un estudio de investigación afecta la precisión de las estimaciones de las medias de los tratamientos y la potencia de las pruebas estadísticas para detectar las diferencias entre las medias de los grupos en tratamiento.

Sin embargo, el costo de conducir estudios de investigación restringe las réplicas a un número razonable. Entonces, el número de réplicas está determinado por las restricciones prácticas que se pueden asignar al problema.

3.6.22 Número de réplicas para pruebas de hipótesis

El método para determinar el número de réplicas con frecuencia se basa en un examen de la hipótesis sobre las diferencias entre las medias de los grupos en tratamiento. Aquí se utiliza un método elemental para los experimentos con dos muestras independientes, para ilustrar algunos atributos del problema del número de réplicas.

Este método se basa en una prueba de hipótesis acerca de las diferencias entre las medias de dos grupos de tratamientos:

$d = m_i - m_j$, con una varianza de error experimental conocida s^2 , utilizando la prueba estadística de distribución normal.

Este método determina el número de réplicas necesario para probar la diferencia entre dos medias muestrales con errores especificados tipo I y tipo II.

El número de réplicas necesario está influenciado primordialmente por cuatro factores que se requieren para los cálculos:

- La varianza poblacional (σ^2).

- El tamaño de la diferencia (que tiene un significado físico) entre las dos medias δ .
- El nivel de significancia de la prueba (α), o la probabilidad del error tipo I.
- La potencia de la prueba $1 - \beta$, o la probabilidad de detectar δ , donde β es la probabilidad del error tipo II.

El número de réplicas necesario para cada grupo en tratamiento r , para alternativas bilaterales se estima mediante:

$$r \geq 2 \left[z_{\alpha/2} + z_{\beta} \right]^2 \left(\frac{\sigma}{\delta} \right)^2 \quad (3.10)$$

Donde $z_{\alpha/2}$ es una variable normal estándar excedida con probabilidad $\alpha/2$ y z_{β} está excedida con probabilidad β .

3.6.23 Aleatorizar para tener inferencias válidas

“Al reconciliar de esta manera las dos necesidades de reducción del error y de una estimación válida del error,... ningún principio se encuentra comprometido en el menor grado. Un experimento admite una estimación válida del error, o no lo hace; lo haga o no, no depende del arreglo real de las parcelas, sino sólo del modo en que se llegó a ese arreglo” (Fisher, 1926).

La réplica de un experimento proporciona los datos para estimar la varianza del error experimental. La definición de bloques proporciona un medio para reducir el error experimental. Sin embargo, las réplicas y los bloques por sí solos no garantizan estimaciones válidas de la varianza del error experimental o de las comparaciones de tratamientos.

Fisher (1926) señaló que la sola aleatorización proporciona estimaciones válidas de la varianza del error para los métodos de inferencia estadística justificados para la estimación y pruebas de hipótesis en el experimento. La aleatorización es la asignación aleatoria de tratamientos a las unidades experimentales.

3.6.24 Razonamiento para la aleatorización

El análisis de datos a partir de un experimento supone que las observaciones constituyen una muestra aleatoria de una población con distribución normal. Esta suposición es aceptable para los estudios comparativos por observación que usan muestras aleatorias de las unidades de observación disponibles de las distintas poblaciones en tratamiento. Sin embargo, cuando se hizo una selección cuidadosa, controlada y supervisada durante el experimento, es cuestionable que las unidades experimentales se puedan considerar una muestra aleatoria.

3.6.25 Las pruebas de la teoría normal aproximan a las pruebas de aleatorización

Fisher (1935) demostró primero que las pruebas de teoría normal son una buena aproximación a las pruebas de aleatorización, siempre que se haya hecho una asignación aleatoria de los tratamientos a las unidades experimentales y los tamaños de las muestras sean razonablemente grandes. Las aproximaciones a las pruebas de aleatorización mediante las de teoría normal mejoran al incrementar el tamaño de la muestra.

3.6.26 Construcción del diseño de investigación

La hipótesis de investigación, el diseño del tratamiento y el diseño del estudio experimental o por observación, constituyen el diseño de investigación para el estudio. Los tratamientos se diseñan para resolver preguntas e hipótesis específicas que surgen en los programas de investigación. Por ejemplo, si un microbiólogo plantea la hipótesis de que la actividad de los microbios del suelo depende de las condiciones de humedad, se establecen tratamientos con distintos niveles de humedad para medir la actividad de los microbios y evaluar la hipótesis. Si un ingeniero de tránsito plantea la hipótesis de que la velocidad del tránsito se relaciona con el ancho de los carriles en las calles, para evaluar la hipótesis se seleccionan carriles con diferentes anchuras y se mide la velocidad de los automóviles en cada uno.

El diseño del tratamiento debe encontrarse dentro del diseño del experimento. El investigador debe decidir qué constituye una unidad experimental, cuántas réplicas de unidades experimentales exige cada tratamiento y qué tratamiento asignar a cada una de ellas. El investigador también debe determinar si agrupará por bloques las unidades experimentales en grupos homogéneos para controlar el error experimental. Los pormenores de cálculo para el análisis estadístico de los distintos diseños pueden variar de un diseño a otro, aunque muchos procedimientos estadísticos usados en el análisis son comunes a la mayoría de los diseños existentes. Esto se debe a que los mismos procedimientos generalmente se relacionan con los diseños del tratamiento específico, cada uno de los cuales puede aparecer en varias configuraciones del diseño experimental.

3.6.27 Selección de las unidades experimentales para estudios comparativos por observación

El planteamiento del ingeniero de tránsito sobre el ancho de las calles es un estudio comparativo por observación. En éste el investigador no puede asignar al azar una unidad a un grupo de tratamiento. Según el tipo de programa de investigación, las unidades básicas se autoseleccionan, o bien, existen en sus grupos característicos. Debe seleccionarse una muestra probabilística de unidades entre los miembros disponibles de cada población en tratamiento. Las unidades se eligen de cada población de manera que cada unidad tenga la misma oportunidad de entrar a la muestra. Se observa que cada población representa una clasificación de tratamiento diferente y que el muestreo aleatorio se lleva a cabo sólo dentro de la población.

El primer paso requiere una identificación de las poblaciones que representan las condiciones o tratamientos de interés para el estudio por observación. Se construye una lista de todas las unidades disponibles en cada población.

3.6.28 Un modelo estadístico para la experimentación

El análisis estadístico se basa en un modelo estadístico formal subyacente. La interpretación adecuada del análisis requiere la comprensión del modelo. En los estudios comparativos, la característica de las unidades o sujetos medida en la observación es la variable de respuesta, identificada como la variable Y .

El modelo estadístico para los estudios comparativos supone que existe una población de referencia de los sujetos o unidades experimentales. En la mayoría de los casos la población es conceptual, aunque es posible imaginar una población de motores de automóvil, tiendas, parcelas, corrales o carne empacada. Cada unidad individual en la población tiene un valor para la variable de respuesta Y , y esta variable tiene una media μ y una varianza σ^2 .

Se supone una población de referencia para cada condición de tratamiento en el estudio, y también se supone que las unidades son sus representantes seleccionados al azar como resultado de una aleatorización. En los estudios por observación se infiere que las unidades se seleccionan al azar de las poblaciones de tratamiento.

3.6.29 Modelo lineal estadístico general

El modelo de medias de celdas es un caso especial del modelo lineal general. El modelo más general describe las relaciones entre dos tipos de variables como una función lineal en un conjunto de parámetros. Un tipo de variable es la de respuesta Y , considerada dependiente del segundo tipo de variable: las variables de diseño x_1, x_2, \dots, x_k . Estas x_i , pueden fijarse mediante el diseño del tratamiento, como la temperatura del tratamiento, o pueden ser medidas covariadas, como la edad de los sujetos. Las x_i , también pueden representar categorías del tratamiento.

El modelo estadístico relaciona Y con las x_i , a través de un conjunto de parámetros $\beta_1, \beta_2, \dots, \beta_k$, con una relación lineal, es decir:

$$Y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \dots + \beta_k x_k + \epsilon \quad (3.11)$$

3.7 Diseños con mediciones repetidas

Cuando los sujetos que se estudian presentan numerosas diferencias entre ellos ante el mismo tratamiento, entonces se introduce una mayor fuente de error experimental. No obstante, aumenta la media de cuadrados de los errores haciendo difícil distinguir las diferencias entre los tratamientos. Una manera de controlar esta variabilidad entre los sujetos, consiste en aplicar a cada uno de ellos los K tratamientos. Este diseño se denomina medidas repetidas.

Si se supone un experimento donde aparece un factor con K tratamientos, y que cada tratamiento se aplica exactamente sobre cada uno de los n individuos:

Cuadro 3.1

DATOS PARA EL DISEÑO UNIFACTORIAL CON MEDIDAS REPETIDAS

		Sujetos				
		1	2	...	N	Totales
Tratamientos	1	Y_{11}	Y_{12}	...	Y_{1n}	$Y_{1.}$
	2	Y_{21}	Y_{22}	...	Y_{2n}	$Y_{2.}$
	⋮	⋮	⋮	⋮	⋮	⋮
	K	Y_{K1}	Y_{K2}	...	Y_{Kn}	$Y_{K.}$
Totales		$Y_{.1}$	$Y_{.2}$...	$Y_{.n}$	$Y_{..}$

La observación Y_{ij} es la respuesta del sujeto j al tratamiento i y sólo se usan n sujetos. El modelo se escribe como:

$$Y_{ij} = \mu + \tau_i + \beta_j + \varepsilon_{ij} \quad (3.12)$$

Donde τ_i es el efecto del i -ésimo tratamiento y β_j el efecto del j -ésimo sujeto.

Se supone que:

$$\sum_{i=1}^K \tau_i = 0 \quad (3.13)$$

Y que los individuos son una muestra aleatoria de una población dada, de manera que cada uno de ellos actúan como un efecto aleatorio.

Esto es:

$$\beta_j \sim N(0, \sigma_\beta^2) \quad (3.14)$$

Dado que β_j es común a todos los K tratamientos medidos sobre el mismo sujeto j, la covarianza entre Y_{ij} e Y'_{ij} es, en general, diferente de 0, asumiéndose que es constante sobre los tratamientos y los sujetos.

La suma total de cuadrados se descompone en dos partes:

$$\sum_{i=1}^K \sum_{j=1}^n (Y_{ij} - \bar{Y}_{..})^2 = K \sum_{j=1}^n (\bar{Y}_{.j} - \bar{Y}_{..})^2 + \sum_{i=1}^K \sum_{j=1}^n (Y_{ij} - \bar{Y}_{.j})^2 \quad (3.15)$$

El primer término de la suma de cuadrados recoge la diferencia entre sujetos y el segundo término la diferencia dentro de sujetos, esto es,

$$SCT = SC_{entre} + SC_{dentro} \quad (3.16)$$

De modo que ambas sumas de cuadrados son independientes entre sí con

$Kn - 1 = (n - 1) + n(K - 1)$ Grados de libertad.

Las diferencias dentro de los sujetos dependen tanto de las diferencias en los efectos de los tratamientos como del ruido. Así, se descompone la suma de cuadrados dentro de sujetos de la siguiente forma:

$$\sum_{i=1}^K \sum_{j=1}^n (Y_{ij} - \bar{Y}_{.j})^2 = n \sum_{i=1}^K (\bar{Y}_{i.} - \bar{Y}_{..})^2 + \sum_{i=1}^K \sum_{j=1}^n (Y_{ij} - \bar{Y}_{i.} - \bar{Y}_{.j} + \bar{Y}_{..})^2 \quad (3.17)$$

El primer término mide la contribución de las diferencias entre las medias de los tratamientos a la suma de cuadrados dentro de los sujetos, y el

segundo término es la variación residual debido al error. Ambos términos son independientes. No obstante:

$$SC_{dentro} = SCT_{ra} + SCE \quad (3.18)$$

Con $n(K - 1) = (K - 1) + (K - 1)(n - 1)$ grados de libertad respectivamente.

Se contrasta:

$$H_0: \mu_i = \mu, \quad i = 1, \dots, K$$

$$H_1: \mu_i \neq \mu, \quad \text{para algún } i$$

Con el cociente:

$$F_0 = \frac{\frac{SCT_{ra}}{K-1}}{\frac{SCE}{(K-1)(n-1)}} = \frac{MCT_{ra}}{MCE} \quad (3.19)$$

Que, cuando H_0 es cierta, se distribuye como una F de Snedecor $F_{K-1, (K-1)(n-1)}$. Se rechaza H_0 con una significación α , si $F_0 > F_{K-1, (K-1)(n-1)}$.

Las observaciones repetidas de la misma unidad experimental con frecuencia son más eficientes que el uso de una unidad experimental diferente para cada observación. No sólo se requieren menos unidades, lo que reduce costos, también la estimación de las tendencias serán más precisas. El aumento en la precisión se debe a que las mediciones en la misma unidad tienden a ser menos variables que las mediciones en unidades distintas; por tanto el efecto de las mediciones repetidas es similar al efecto del uso de bloques.

3.7.1 Diseños entre y dentro de unidades experimentales

Los diseños de mediciones repetidas se pueden describir en términos del diseño entre sujetos y el diseño dentro sujetos. Los diseños entre sujetos se refieren a los diseños en que una unidad experimental se

asigna a un tratamiento. Los diseños dentro sujetos se refieren a las diferentes mediciones en cada unidad experimental.

3.7.2 Relaciones entre las mediciones repetidas

Las relaciones entre las observaciones gobiernan los métodos estadísticos necesarios para el diseño de investigación específico usado en un estudio. Aquí se explora la correspondencia de las relaciones con el método de análisis para diseños de medidas repetidas junto con algunas estrategias útiles para el análisis.

3.7.3 Suposiciones del análisis de varianza

Varianzas iguales para los grupos de tratamiento y observaciones independientes con distribución normal son las suposiciones usuales necesarias para un análisis de varianza válido de los datos.

3.7.4 La condición de Huynh-Feldt es menos estricta que la simetría completa

Huynh y Feldt (1979) mostraron que las condiciones necesarias para el análisis de varianza usual en los diseños de mediciones repetidas eran menos estrictas que la condición de simetría completa. Demostraron que la condición necesaria es tener la misma varianza de la diferencia para todos los pares posibles de observaciones tomadas en diferentes periodos, digamos, Y_i y Y_j es decir:

$$\sigma_{(y_i - y_j)}^2 = 2\lambda \text{ para } i \neq j \quad (3.20)$$

Para algún $\lambda > 0$. Esta condición también se puede establecer como:

$$\sigma_{ij} = \frac{1}{2}(\sigma_i^2 + \sigma_j^2) - \lambda \text{ para } i \neq j \quad (3.21)$$

La matriz de varianzas y covarianzas que satisface esta condición se conoce como matriz tipo H. Los cuadrados medios del análisis de

varianza se pueden usar para probar las hipótesis sobre los tratamientos dentro sujetos si se cumple la condición de Huynh-Feldt.

3.7.5 Si cada sujeto recibe todos los tratamientos

La realidad de muchos estudios de investigación desde el punto de vista económico y de control del error experimental exige la obtención de más de una observación de cada unidad experimental. Por ejemplo, el considerable costo de mantener animales grandes hace necesario obtener toda la información posible sobre los tratamientos con cada animal. Además, la variabilidad de las observaciones entre los animales tiende a ser mucho mayor que entre observaciones múltiples en el mismo animal, por lo que el bloqueo de los animales con los tratamientos administrados a cada animal aumenta la precisión de las comparaciones de tratamientos.

3.7.6 Transferencia de efectos

Los efectos de cierto tipo de tratamiento se transfieren al siguiente periodo de tratamiento cuando se administran en secuencia a los sujetos. Estas transferencias pueden sesgar severamente las estimaciones de las medias de tratamiento debido a que los tratamientos suministrados en periodos anteriores influyen en el efecto del tratamiento aplicado en periodos sucesivos.

Los efectos de transferencia son particularmente complicados en sujetos humanos y animales con un suministro sucesivo de tratamientos dietéticos o médicos que afectan su fisiología. Antes de suministrar un segundo tratamiento suele usarse un periodo de descanso, o "de lavado", entre dos tratamientos sucesivos para eliminar los efectos del tratamiento más reciente.

3.7.7 Una prueba para la suposición Huynh-Feldt

Se puede usar el análisis de varianza univariado con cualquiera de los tres conjuntos de suposiciones sobre las mediciones repetidas, éstos son: independencia, simetría compuesta o condición Huynh-Feldt. La condición de Huynh-Feldt con la matriz tipo H para varianzas y

covarianzas de las mediciones repetidas es la menos restrictiva de las tres suposiciones. Se pueden usar sencillos métodos univariados para el análisis si se puede suponer que la condición Huynh-Feldt se cumple para las mediciones repetidas. La suposición de una matriz tipo H se puede evaluar con la prueba atribuida a Mauchly (1940).

3.7.8 Análisis de varianza univariado para mediciones repetidas

Si se puede tener una seguridad razonable de que las suposiciones del análisis de varianza son válidas para las mediciones repetidas en cada uno de los grupos de tratamientos, se pueden usar los cuadrados medios del análisis de varianza de parcelas divididas para probar las hipótesis sobre las medias de tratamiento y sus interacciones temporales.

3.7.9 Prueba de esfericidad de Mauchly

La condición Huynh-Feldt para la matriz de varianzas y covarianzas de las K mediciones repetidas de los sujetos necesita (K-1) contrastes ortogonales normalizados para que las mediciones repetidas no se correlacionen y tengan varianzas iguales.

Sea Σ la matriz de covarianzas de las mediciones repetidas y C una matriz de $(K-1) \times K$, donde los renglones son contrastes ortogonales normalizados de las K mediciones repetidas. La condición Huynh-Feldt necesaria para la covarianza de los contrastes es $C\Sigma C' = \lambda I$, donde I es la matriz identidad y C' es la transpuesta de C . Si se satisface la condición, se dice que la matriz de covarianzas λI es esférica.

Sea S_{ij} el elemento en el i-ésimo renglón y la j-ésima columna de la matriz de covarianzas de $(K \times K)$ para los errores experimentales dentro sujetos S, con $\nu = (K - 1) \times (n - 1)$ grados de libertad. Se eligen (K-1) contrastes ortogonales normalizados en las K mediciones repetidas. Sea C la matriz de $(K-1) \times K$, donde los renglones son contrastes ortogonales normalizados en las K mediciones repetidas, se calcula la matriz de orden $(K - 1) \times (K - 1)$, definida por CSC' . El

estadístico de prueba (Mauchly, 1940) para contrastar la hipótesis nula $H_0: C\Sigma C' = \lambda I$ es:

$$W = \frac{(K-1)^{(K-1)} |CSC'|}{Tr(CSC')^{(K-1)}} \quad (3.22)$$

Donde $TrCSC'$ es la traza de la matriz. El estadístico de prueba se pone en la escala adecuada para mejorar la exactitud de sus aproximaciones por la distribución chi cuadrada. El factor de escala para la aproximación chi cuadrada con $f = \frac{1}{2}K(K-1) - 1$ grados de libertad es:

$$\gamma = v - \frac{2K^2 - 3K + 3}{6(K-1)} \quad (3.23)$$

La hipótesis nula se rechaza al nivel de significancia γ si $W > \chi_{\alpha, f}^2$

3.8 Comparaciones por parejas (Efectos principales)

Una vez determinada la significación de las diferencias se procede a la comparación de los tratamientos por parejas, es decir, a contrastar las hipótesis

$$\begin{cases} H_0: \mu_i = \mu_j \\ H_1: \mu_i \neq \mu_j \end{cases} \quad (3.24)$$

Para cada pareja de medias. El estadístico de contraste que se utiliza es:

$$t^{ij} = \frac{|\bar{X}_i - \bar{X}_j|}{S_R \sqrt{\frac{1}{n_i} + \frac{1}{n_j}}} \sim t_{n-K} \quad (3.25)$$

Que sigue una distribución t de Student con n-K grados de libertad, y la desviación común se define por:

$$S_R = \sqrt{\frac{SCE}{(K-1)(n-1)}} \quad (3.26)$$

Donde SCE está definido en la ecuación (3.18).

La realización directa de cada contraste para un nivel de significación α incrementa la probabilidad de cometer un error de tipo I para el contraste global de igualdad de todas las medias por lo que, en general, el nivel de significación de cada contraste particular tiene que ser corregido.

Si no importa el incremento en el riesgo tipo I puede realizarse directamente el contraste t de Student, asumiendo que probablemente encontraremos más diferencias de las que se encuentran realmente en los datos.

Se rechaza la hipótesis nula si el estadístico de contraste supera el valor crítico de la distribución t de Student al nivel α .

$$t^{ij} = \frac{|\bar{X}_i - \bar{X}_j|}{S_R \sqrt{\frac{1}{n_i} + \frac{1}{n_j}}} > t_{n-K, \alpha} \quad (3.27)$$

O lo que es lo mismo:

$$|\bar{X}_i - \bar{X}_j| > t_{n-K, \alpha} S_R \sqrt{\frac{1}{n_i} + \frac{1}{n_j}} \quad (3.28)$$

Es decir, cuando la diferencia de medias supera el valor $t_{n-K, \alpha} S_R \sqrt{\frac{1}{n_i} + \frac{1}{n_j}}$ que se denomina “*diferencia significativa mínima*” (least significant difference). Por esta razón, a veces al test se le conoce como LSD.

Si se está interesado en mantener el nivel de significación global por debajo del nivel predeterminado, es necesario corregir el nivel que se utilizará en cada contraste individual de forma que al aumentar el riesgo, permanezca por debajo del fijado.

La corrección consiste en utilizar para cada contraste un nivel α' más pequeño que α y que resulta de dividir éste por una constante de penalización Q.

$$\alpha' = \frac{\alpha}{Q} \quad (3.29)$$

Dependiendo de los valores que tomemos para “Q” obtendremos distintos procedimientos de contraste.

3.8.1 Test de Bonferroni

Q es el número de comparaciones posibles por parejas.

$$Q = \binom{K}{2} = \frac{K(K-1)}{2} \quad (3.30)$$

El test de Bonferroni es el que más penaliza de forma que, el nivel de significación para el contraste global queda, en realidad, por debajo del fijado. Se trata entonces de un test conservador que detecta menos diferencias de las que hay realmente en los datos.

3.9 ANOVA de un factor

El análisis de la varianza permite contrastar la hipótesis nula de que las medias de K poblaciones ($K > 2$) son iguales, frente a la hipótesis alternativa de que por lo menos una de las poblaciones difiere de las demás en cuanto a su valor esperado. Este contraste es fundamental en el análisis de resultados experimentales, en los que interesa comparar los resultados de K “tratamientos” con respecto a la variable dependiente o de interés.

3.9.1 Supuestos

El ANOVA requiere el cumplimiento de los siguientes supuestos:

- Las poblaciones (distribuciones de probabilidad de la variable dependiente correspondiente a cada factor) son normales.
- Las K muestras sobre las que se aplican los tratamientos son independientes.
- Las poblaciones tienen todas igual varianza (homocedasticidad).

3.9.2 Fundamentos del ANOVA de un factor

El ANOVA se basa en la descomposición de la variación total de los datos con respecto a la media global (SCT), que bajo el supuesto de que la hipótesis nula es cierta, es una estimación de σ^2 obtenida a partir de toda la información muestral, en dos partes:

- Variación dentro de las muestras (SCD) o Intra-grupos, cuantifica la dispersión de los valores de cada muestra con respecto a sus correspondientes medias.
- Variación entre muestras (SCE) o Inter-grupos, cuantifica la dispersión de las medias de las muestras con respecto a la media global.

Las expresiones para el cálculo de los elementos que intervienen en el ANOVA son las siguientes:

3.9.3 Media global

$$\bar{X} = \frac{\sum_{j=1}^K \sum_{i=1}^{n_j} x_{ij}}{n} \quad (3.31)$$

3.9.4 Variación total

$$SCT = \sum_{j=1}^K \sum_{i=1}^{n_j} (x_{ij} - \bar{X})^2 \quad (3.32)$$

3.9.5 Variación intra grupos

$$SCD = \sum_{j=1}^K \sum_{i=1}^{n_j} (x_{ij} - \bar{X}_j)^2 \quad (3.33)$$

3.9.6 Variación inter grupos

$$SCE = \sum_{j=1}^K (\bar{X}_j - \bar{X})^2 n_j \quad (3.34)$$

Siendo x_{ij} el i -ésimo valor de la muestra j -ésima; n_j el tamaño de dicha muestra y \bar{X}_j su media.

Cuando la hipótesis nula es cierta $SCE/K-1$ y $SCD/n-K$ son dos estimadores insesgados de la varianza poblacional y el cociente entre ambos se distribuye según una F de Snedecor con $K-1$ grados de libertad en el numerador y $n-K$ grados de libertad en el denominador. Por lo tanto, si la hipótesis nula es cierta es de esperar que el cociente entre ambas estimaciones será aproximadamente igual a 1, de forma que se rechazará dicha hipótesis nula si dicho cociente difiere significativamente de 1.

CAPITULO IV

Metodología

Para alcanzar los objetivos de la presente investigación, se han definido las siguientes bases sobre las cuales se fundamenta el estudio:

4.1 Unidad de análisis

El sujeto de estudio es el estudiante de la carrera profesional de ingeniería estadística-UNI, inscrito en el semestre 2011-I.

4.2 Población

Todos los estudiantes de la carrera profesional de ingeniería estadística UNI, inscritos en el semestre 2011-I, sumando un total de N=298 estudiantes regulares, en 10 estratos definido por los 10 ciclos académicos. A continuación se indica los totales poblacionales por cada estrato:

Cuadro 4.1

TAMAÑOS POBLACIONALES POR ESTRATO

N1	N2	N3	N4	N5	N6	N7	N8	N9	N10
37	26	32	29	20	40	23	23	24	44

Elaboración: Propia

Fuente: UNI

4.3 Muestra

En función a la naturaleza de un estudio exploratorio, se definió:

4.3.1 Tamaño de la muestra:

Se fijó un error máximo permisible del 0.5 ($E=0.5$), una desviación de 1.8 ($\sigma = 1.8$), y un valor critico de $Z_{1-\frac{\alpha}{2}} = 1.96$

con $\alpha = 0.05$, se utilizó la siguiente fórmula:

$$n = \frac{Z_{1-\frac{\alpha}{2}}^2 \sigma^2}{E^2} \quad (4.1)$$

Se obtuvo un valor de $n=50$ alumnos

4.3.2 Tamaño de cada estrato:

Se calcula la fracción de muestreo $f = \frac{n}{N} = 0.16778523$ y con la relación $n_i = N_i f$ con $i=1, \dots, 10$ se obtienen los tamaños de muestra en cada estrato.

Cuadro 4.2

TAMAÑOS DE MUESTRAS

Ciclo	N_i	n_i
1	37	6
2	26	4
3	32	5
4	29	5
5	20	3
6	40	7
7	23	4
8	23	4
9	24	4
10	44	7
Total	N=298	n=50

Elaboración: Propia

4.4 Instrumento de recolección de datos

Se utilizó un cuestionario de 8 preguntas, dispuestas de la siguiente manera (ver anexo 1): las preguntas 1, 2, y 3 miden la interiorización que tienen los estudiantes de comparar la variabilidad en los gráficos. Las preguntas 4, 5, y 6 miden la interiorización que tienen los estudiantes de resolver problemas experimentales, basados en la variabilidad. Las preguntas 7 y 8 miden la interiorización que tienen los estudiantes de aplicar conceptos e interpretar resultados sobre la variabilidad.

Todas estas preguntas son extraídas de los ejemplos planteados y sustentados en el marco teórico del presente informe y están organizados como se presenta en el anexo 1 para propósitos del estudio.

4.5 Diseño del estudio para la hipótesis general

4.5.1 Operacionalización de las variables para el estudio de la hipótesis general

Para este estudio se han definido las siguientes variables:

Cuadro 4.3

OPERACIONALIZACION DE LAS VARIABLES PARA LA HIPOTESIS GENERAL

Variable	Indicador	Ítem
Interiorización de la concepción de la variabilidad obtenido por el estudiante en cada tipo de prueba o test	Puntaje obtenido por el estudiante en la primera prueba que consiste en comparar la variabilidad en los gráficos.	Preg1 Preg2 Preg3 (Anexo I)
	Puntaje obtenido por el estudiante en la segunda prueba que consiste en resolver problemas experimentales que implícitamente se relacionan con la variabilidad.	Preg4 Preg5 Preg6 (Anexo I)
	Puntaje obtenido por el estudiante en la tercera prueba que consiste en aplicar conceptos, hacer cálculos e interpretar resultados sobre la variabilidad.	Preg7 Preg8 (Anexo I)

Elaboración: Propia

4.5.2 Hipótesis: El estudiante de la carrera profesional de ingeniería estadística – UNI no tiene interiorizado el concepto de la variabilidad a lo largo de sus estudios universitarios.

4.5.3 Diseño del tratamiento: Para evaluar la hipótesis se utilizó un factor “tipo de prueba o test” con tres niveles: problemas de comparación de gráficos, problemas experimentales y aplicación de conceptos, hacer cálculos e interpretación, que influyen en la interiorización de la concepción de la variabilidad.

Con el puntaje obtenido en cada una de estas pruebas o test se mide lo siguiente: el primer nivel del factor mide la interiorización de la concepción de la variabilidad que tienen los estudiantes al comparar la variabilidad en los gráficos; el segundo nivel del factor mide la interiorización de la concepción de la variabilidad que tienen los estudiantes de resolver problemas experimentales que implícitamente se relacionan con la variabilidad; y el tercer nivel del factor mide la interiorización de la concepción de la variabilidad que tienen los estudiantes de aplicar conceptos, hacer cálculos e interpretar resultados sobre la variabilidad.

4.5.4 Diseño del experimento: A los 50 estudiantes se les consideró como sujetos a los cuáles individualmente se les aplicó los 3 tipos de prueba o test en forma repetida, obteniendo un puntaje para cada prueba. Así se cumple con el criterio de aleatorización en cada sujeto que, para propósito del análisis, representa a los bloques.

4.5.5 Técnica estadística: Se utilizó el modelo de Medidas Repetidas. Esta técnica requiere del cumplimiento de los supuestos de normalidad, independencia y homocedasticidad.

4.6 Diseño del estudio para la primera hipótesis específica

4.6.1 Operacionalización de las variables para el estudio de la primera hipótesis específica

Para este estudio se han definido las siguientes variables:

Cuadro 4.4

OPERACIONALIZACION DE LAS VARIABLES PARA LA PRIMERA HIPOTESIS ESPECÍFICA

Variable	Indicador	Ítem
Segmento de estudio	Estudiantes de 1er ciclo al 4to ciclo	Pregunta: ¿En qué ciclo te encuentras?
	Estudiantes de 5to ciclo al 7mo ciclo	Pregunta: ¿En qué ciclo te encuentras?
	Estudiantes de 8vo ciclo al 10mo ciclo	Pregunta: ¿En qué ciclo te encuentras?
Interiorización de la concepción de la variabilidad obtenido por el estudiante UNI en toda la prueba.	Puntaje obtenido por el estudiante en toda la prueba (la suma de los puntajes obtenidos en las tres pruebas).	Preg1 Preg2 Preg3 Preg4 Preg5 Preg6 Preg7 Preg8 Preg9 Preg10 (Anexo I)

Elaboración: Propia

4.6.2 Hipótesis: El avance de sus estudios universitarios influye significativamente en la interiorización del concepto de variabilidad en el estudiante de la carrera profesional de ingeniería estadística – UNI.

4.6.3 Diseño del experimento: Se ha definido el factor “segmento de estudio” con tres niveles de la siguiente manera: el segmento 1 está conformado por los alumnos que pertenecen del 1er ciclo al 4to ciclo; el segmento 2, por alumnos del 5to ciclo al 7mo ciclo; y del 8vo ciclo al 10mo ciclo segmento 3, y se ha medido el puntaje total en la prueba o test para cada uno de los alumnos en los tres segmentos, formándose así una muestra de puntajes por segmento.

4.6.4 Técnica estadística: La técnica elegida para ésta hipótesis es el ANOVA de un factor, que requiere el cumplimiento de los supuestos de normalidad, independencia, y homocedasticidad.

4.7 Diseño del estudio para la segunda hipótesis específica

4.7.1 Hipótesis: La importancia que se da a los cálculos influye significativamente en la interiorización del concepto de la variabilidad, en el estudiante de la carrera profesional de ingeniería estadística-UNI.

4.7.2 Diseño del estudio: Esta hipótesis se prueba con un análisis que es consecuencia de la hipótesis general, para la cual se usa el mismo diseño dado en 4.6.2, 4.6.3, y 4.6.4

4.7.3 Técnica estadística: La técnica elegida es Medidas Repetidas mediante Efectos Principales.

CAPITULO V

Análisis y resultados

5.1 Análisis de la hipótesis general

Se consideró la metodología planteada en 4.5.1, 4.5.2, 4.5.3, 4.5.4 y 4.5.5 para la cual se usó el siguiente modelo de medidas repetidas:

$$Y_{ij} = \mu + \tau_i + \beta_j + \varepsilon_{ij} \quad (5.1)$$

Donde: $i = 1,2,3$ $j = 1,2,3, \dots, 50$

Para la prueba de la hipótesis general se define el constructo conceptual que se fundamenta en el hecho de que si el estudiante de la carrera profesional de ingeniería estadística – UNI, responde en igual medida a los tres niveles del factor (a las tres pruebas o test) se concluye que efectivamente tiene interiorizado el concepto de la variabilidad a lo largo de sus estudios universitarios. En caso contrario, es decir, si no responde de manera homogénea a los tres niveles del factor se concluye que no tiene interiorizado el concepto de la variabilidad.

5.2 Supuestos del modelo de medidas repetidas:

5.2.1 Prueba de normalidad

Se utilizó el test no paramétrico de Kolmogorov Smirnov, que plantea las siguientes hipótesis para los puntajes de cada una de las pruebas o test:

Para la primera prueba o test:

H_{0Y1} : La población de los puntajes que obtienen los estudiantes al comparar la variabilidad en los gráficos, sigue una distribución normal.

H_{1Y1} : La población de los puntajes que obtienen los estudiantes al comparar la variabilidad en los gráficos, no sigue una distribución normal.

La segunda prueba o test:

H_{0Y2} : La población de los puntajes que obtienen los estudiantes al resolver problemas experimentales que implícitamente se relacionan con la variabilidad, sigue una distribución normal.

H_{1Y2} : La población de los puntajes que obtienen los estudiantes al resolver problemas experimentales que implícitamente se relacionan con la variabilidad, no sigue una distribución normal.

Para la tercera prueba o test:

H_{0Y3} : La población de los puntajes que obtienen los estudiantes al aplicar conceptos, hacer cálculos e interpretar resultados sobre la variabilidad, sigue una distribución normal.

H_{1Y3} : La población de los puntajes que obtienen los estudiantes al aplicar conceptos, hacer cálculos e interpretar resultados sobre la variabilidad, no sigue una distribución normal.

Los resultados se muestran en la siguiente tabla:

Cuadro 5.1

TEST DE KOLMOGOROV SMIRNOV PARA CADA FACTOR

	Y ₁	Y ₂	Y ₃
Z de Kolmogorov Smirnov	1.249	1.298	1.243
Sig. Asintótico (bilateral)	0.089	0.069	0.091

Fuente: Encuesta realizada a los alumnos de Ing. Estadística FIECS UNI

Elaboración: propia

Los resultados del cuadro 5.1 permiten concluir que no se puede rechazar las hipótesis nulas con un nivel de significancia de $\alpha = 0.05$ pues:

Y₁: primer nivel del factor: $0.089 > 0.05$

Y₂: segundo nivel del factor $0.069 > 0.05$

Y₃: tercer nivel factor $0.091 > 0.05$

Por lo tanto, los puntajes en cada una de las tres pruebas, provienen de poblaciones con distribución normal.

5.2.2 Prueba de independencia

Se utilizó el estadístico d de Durbin Watson, en las observaciones para los 3 niveles del factor:

$$d_i = \frac{\sum_{j=2}^{50} (e_j - e_{j-1})^2}{\sum_{j=1}^{50} e_i^2} \quad (5.2)$$

Donde $i = 1, 2, 3$

Para la primera prueba o test:

H₀: Las puntajes obtenidos en la primera prueba referida a la comparación de gráficos, son independientes.

H₁: Las puntajes obtenidos en la primera prueba referida a la comparación de gráficos, no son independientes.

Para la segunda prueba o test:

H₀: Los puntajes obtenidos en la segunda prueba referida a problemas experimentales, son independientes.

H₁: Los puntajes obtenidos en la segunda prueba referida a problemas experimentales, no son independientes.

Para la tercera prueba o test:

H₀: Las puntajes obtenidos en la tercera prueba referida a aplicar conceptos, hacer cálculos e interpretar resultados, son independientes.

H₁: Las puntajes obtenidos en la tercera prueba referida a aplicar conceptos, hacer cálculos e interpretar resultados, no son independientes.

Se obtuvo los siguientes resultados:

Cuadro 5.2

TEST DE DURBIN WATSON

	d₁	d₂	d₃
Durbin Watson	2.09	1.96	1.85

Elaboración: propia

En el cuadro 5.2 se confirma la independencia en cada uno de los tres niveles del factor dado que el estadístico **d** de Durbin Watson está comprendido entre 1.5 y 2.5. Por lo tanto se cumple el supuesto de independencia en las observaciones correspondientes a los puntajes en cada una de las tres pruebas.

5.2.3 Prueba de homocedasticidad

Se utilizó la prueba W de Mauchly, (1940) que sostiene la hipótesis nula de esfericidad ((H₀: igualdad de varianzas tomadas dos a dos) en las poblaciones de los 3 niveles del factor. Los resultados se muestran en la siguiente tabla:

Cuadro 5.3

TEST DE ESFERICIDAD

W de Mauchly	Chi cuadrado aproximado	Grados de libertad	Sig
0.938	3.088	2	0.214

Elaboración: propia

Con los resultados del cuadro 5.3, no se rechaza la hipótesis nula y se concluye en que se cumple el supuesto de homocedasticidad con un nivel de significancia del 5% ($\alpha = 0.05$) dado que $0.214 > 0.05$.

Por lo tanto se cumple los tres supuestos que el modelo de medidas repetidas requiere.

5.3 Análisis de varianza para la hipótesis general

Siguiendo el constructo conceptual especificado al inicio de éste capítulo, se plantea la siguiente hipótesis estadística:

$H_0: \mu_1 = \mu_2 = \mu_3 = \mu$ (El estudiante de la carrera profesional de ingeniería estadística – UNI tiene interiorizado el concepto de la variabilidad a lo largo de sus estudios universitarios).

$H_1: \exists \text{ algún } \mu_i \neq \mu; i = 1, 2, 3$ (El estudiante de la carrera profesional de ingeniería estadística – UNI no tiene interiorizado el concepto de la variabilidad a lo largo de sus estudios universitarios).

Se fija en 5% ($\alpha = 0.05$) el nivel de significación para todo el análisis. Para probar esta hipótesis se utilizó el ANOVA en medidas repetidas, cuyos resultados se muestran en el siguiente cuadro:

Cuadro 5.4

ANÁLISIS DE VARIANZA PARA LA HIPÓTESIS GENERAL

Fuente de variación	Suma de cuadrados	Grados de libertad	Media cuadrática	F	Sig.
Tratamientos	17036.444	2	8518.222	14.846	0
Error	56230.222	98	573.778		

Elaboración: propia

Con los resultados del cuadro 5.4, se rechaza la hipótesis nula pues $0 < 0.05$ y se acepta la hipótesis alterna.

Por lo tanto se concluye para un nivel de significancia del 5% ($\alpha = 0.05$) que el estudiante de la carrera profesional de ingeniería estadística – UNI no tiene interiorizado el concepto de la variabilidad a lo largo de sus estudios universitarios.

5.4 Análisis de la primera hipótesis específica

Se consideró la metodología planteada en 4.6.1, 4.6.2, 4.6.3 y 4.6.4, para el cual se usó el siguiente modelo de ANOVA de un factor:

$$Y_{ij} = \mu + \tau_i + \varepsilon_{ij} \quad (5.3)$$

Dónde: $i = 1,2,3$ $j = 1,2,3, \dots, 50$

Esta prueba se fundamenta en el hecho de que si los estudiantes de cada segmento de estudio de la escuela profesional de ingeniería estadística – UNI, responden en diferente medida al total de la prueba o test, se concluye que efectivamente el avance de sus estudios universitarios influye significativamente en la interiorización del concepto de la variabilidad.

5.5 Supuestos del ANOVA:

5.5.1 Prueba de normalidad:

Se plantearon las hipótesis nulas:

H_{0s1}: La población de los puntajes que obtienen los estudiantes en la prueba sobre la concepción de la variabilidad, correspondientes al segmento1, tiene una distribución normal.

H_{0s2}: La población de los puntajes que obtienen los estudiantes en la prueba sobre la concepción de la variabilidad, correspondientes al segmento2, tiene una distribución normal.

H_{0s3}: La población de los puntajes que obtienen los estudiantes en la prueba sobre la concepción de la variabilidad, correspondientes al segmento3, tiene una distribución normal.

Para probar éstas hipótesis se utilizó el test no paramétrico de Kolmogorov – Smirnov, con un nivel de significación del 5% ($\alpha = 0.05$) los resultados se muestran a continuación:

Cuadro 5.5
TEST DE KOLMOGOROV SMIRNOV

	Segmento1	Segmento2	Segmento3
Z de Kolmogorov Smirnov	0.964	1.065	0.655
Significación asintótica bilateral	0.311	0.207	0.783

Fuente: Encuesta realizada a los alumnos de Ing. Estadística FIECS UNI

Elaboración: Propia

Con los resultados del cuadro 5.5, se concluye que:

La población de los puntajes que obtienen los estudiantes en la prueba sobre la concepción de la variabilidad correspondientes al segmento1, se distribuye normalmente ($0.311 > 0.05$)

La población de los puntajes que obtienen los estudiantes en la prueba sobre la concepción de la variabilidad correspondientes al segmento2, se distribuye normalmente ($0.207 > 0.05$)

La población de los puntajes que obtienen los estudiantes en la prueba sobre la concepción de la variabilidad correspondientes al segmento3, se distribuye normalmente ($0.783 > 0.05$)

5.5.2 Prueba de independencia:

Se utilizó el estadístico d de Durbin Watson, planteado en la ecuación 5.2, en cada uno de los segmentos definidos, los resultados se muestran en la siguiente tabla:

Cuadro 5.6
TEST DE DURBIN WATSON

	Segmento1	Segmento2	Segmentos3
Durbin Watson	2.078	2.371	2.372

Elaboración: propia

Con los resultados del cuadro 5.6, el estadístico d de Durbin Watson, en los 3 segmentos, está entre 1.5 y 2.5, por tanto se cumple el supuesto de independencia en las observaciones (puntajes) correspondientes a cada segmento de estudio.

5.5.3 Prueba de Homocedasticidad:

Se plantearon las hipótesis:

$H_0: \sigma_1^2 = \sigma_2^2 = \sigma_3^2 = \sigma^2$ (Las varianzas de los puntajes de los estudiantes en la prueba sobre la concepción de la variabilidad en las poblaciones del segmento1, segmento2, y segmento3 son iguales).

$H_1: \sigma_i^2 \neq \sigma^2$ para algun $i = 1,2,3$ (Al menos una de las varianzas de los puntajes de los estudiantes en la prueba sobre la concepción de la variabilidad, es diferente).

Para ésta prueba se utilizó el test de Levene, cuyos resultados se muestran en el siguiente cuadro:

Cuadro 5.7
TEST DE LEVENE

Estadístico de Levene	Grados de libertad en el numerador	Grados de libertad en el denominador	Significación
0.984	2	47	0.381

Fuente: Encuesta realizada a los alumnos de Ing. Estadística FIECS UNI

Elaboración: Propia

Con los resultados del cuadro 5.7 se concluye que:

Las varianzas de los puntajes de los estudiantes en la prueba sobre la concepción de la variabilidad en las poblaciones del segmento1, segmento2, y segmento3 son iguales, ($0.381 > 0.05$). Por lo que se acepta el supuesto de homocedasticidad.

Por lo tanto se cumple con los supuestos de normalidad, independencia, y homocedasticidad, que el ANOVA de un factor exige.

5.6 Análisis de varianza para la primera hipótesis específica:

Para probar la primera hipótesis específica se utilizó el ANOVA de un factor, con las siguientes hipótesis estadísticas:

$H_0: \mu_1 = \mu_2 = \mu_3 = \mu$ (El avance de sus estudios universitarios no influye significativamente en la interiorización del concepto de

variabilidad en el estudiante de la carrera profesional de ingeniería estadística UNI).

$H_1: \mu_i \neq \mu, \text{ para } i = 1, 2, 3$ (El avance de sus estudios universitarios influye significativamente en la interiorización del concepto de variabilidad en el estudiante de la carrera profesional de ingeniería estadística - UNI).

Los resultados se muestran en la siguiente tabla:

Cuadro 5.8
ANOVA DE UN FACTOR

	Suma de cuadrados	Grados de libertad	Media Cuadrática	Estadístico F	Significación
Inter grupos	31800	2	15900	4.391	0.018
Intra grupos	170200	47	3621		
Total	202000	49			

Fuente: Encuesta realizada a los alumnos de Ing. Estadística FIECS UNI

Elaboración: Propia

Con los resultados del cuadro 5.8, se rechaza la hipótesis nula pues ($0.018 < 0.05$) y se acepta la hipótesis alternativa que sostiene que el avance de sus estudios universitarios influye significativamente en la interiorización del concepto de variabilidad en el estudiante de la carrera profesional de ingeniería estadística UNI,

5.7 Análisis de la segunda hipótesis específica

El siguiente análisis parte del hecho de haber aceptado que el estudiante de la carrera profesional de ingeniería estadística en la

UNI, no tiene interiorizado el concepto de la variabilidad a lo largo de sus estudios universitarios (Hipótesis general). Es decir:

$$\exists \text{ algún } \mu_i \neq \mu_j \quad \forall i \neq j \text{ para } i = 1, 2, 3 \quad j = 1, 2, 3 \quad (5.4)$$

Donde:

μ_1 → Puntaje promedio que tienen los estudiantes al comparar la variabilidad en los gráficos.

μ_2 → Puntaje promedio que tienen los estudiantes al resolver problemas experimentales que implícitamente se relacionan con la variabilidad.

μ_3 → Puntaje promedio que tienen los estudiantes al aplicar conceptos, hacer cálculos e interpretar resultados sobre la variabilidad.

Se plantean las siguientes hipótesis:

H₀: La importancia que se da a los cálculos influye significativamente en la interiorización del concepto de la variabilidad, en el estudiante de la carrera profesional de ingeniería estadística-UNI.

H₁: La importancia que se da a los cálculos no influye significativamente en la interiorización del concepto de la variabilidad, en el estudiante de la carrera profesional de ingeniería estadística-UNI.

Se comparó dos a dos los niveles de un factor de medidas repetidas, mediante la técnica: comparar efectos principales. Los niveles críticos se ajustan mediante la corrección de Bonferroni, como plantea Dunn, (1961) y los resultados se indican en la siguiente tabla:

Cuadro 5.9
COMPARACIONES POR PARES

Nivel i	Nivel j	Diferencia de medias (i-j)	Significación
1	2	-24.667	0
1	3	-4.933	0.857
2	1	24.667	0
2	3	19.733	0.002
3	1	4.933	0.857
3	2	-19.733	0.002

Fuente: Encuesta realizada a los alumnos de Ing. Estadística FIECS UNI

Elaboración: Propia

Con los resultados del cuadro 5.9, se determina que el puntaje promedio en los problemas experimentales (nivel 2) se diferencia en forma significativa de los demás (del nivel 1 y nivel 3) ($\mu_2 \neq \mu_1$ y $\mu_2 \neq \mu_3$), por tanto, es el razonamiento intuitivo, en los problemas experimentales, lo que influye en la interiorización del concepto de la variabilidad, mas no la importancia que se da a los cálculos.

CONCLUSIONES

El presente estudio de investigación basado en un margen de error del 5%, concurre en las siguientes conclusiones:

1. Con los resultados del estudio de la hipótesis general se concluye que el estudiante de la carrera profesional de ingeniería estadística de la UNI no tiene interiorizado el concepto de la variabilidad a lo largo de sus estudios universitarios.
2. Del análisis de la primera hipótesis específica se concluye que el avance de sus estudios universitarios es un factor de influencia en la interiorización del concepto de variabilidad en el estudiante de la carrera profesional de ingeniería estadística de la UNI.
3. Con referencia al análisis de la segunda hipótesis específica se concluye que, la capacidad que tienen los estudiantes de resolver problemas experimentales que implícitamente se relacionan con la variabilidad, es el factor que más influye en la interiorización del concepto de la variabilidad, en el estudiante de la carrera profesional de ingeniería estadística de la UNI y no son los conceptos y los cálculos, como empíricamente se suponía.

Esta conclusión evidencia que el alumno se caracteriza por un razonamiento implícito sobre la variabilidad al pensar en los posibles resultados de los experimentos y que a su vez, permite el razonamiento acerca de las causas que producen dichos resultados.

RECOMENDACIONES

Los resultados de la presente investigación, basados en una confianza del 95%, permiten plantear las siguientes recomendaciones:

Es necesario un rediseño del plan curricular, en el cuál el nuevo plan debe ponderar en gran medida la interiorización de la variabilidad en el estudiante de la carrera profesional de ingeniería estadística. Sin embargo, esta misión también se centra en el eje motriz de la ejecución de todo plan curricular, que es el docente universitario, a quien se recomienda incluir metodologías de la enseñanza mediante dinámicas de grupo y el planteamiento de experimentos con materiales e instrumentos con los cuales se puedan desarrollar trabajos que permitan el análisis de los resultados.

El uso del software es necesario en el desarrollo de simulaciones de los experimentos en gran escala y así fortalecer el concepto de la variabilidad en los alumnos. No obstante, ponderar la importancia al desarrollo de distintos tipos de gráficos estadísticos, que permita al alumno tener un mejor entendimiento de los análisis estadísticos y al mejor planteamiento de los mismos.

BIBLIOGRAFIA

- Arteaga, P., Batanero, C., Díaz, C., Contreras, J. M. (2009) “El Lenguaje de los gráficos estadísticos”, Revista Iberoamericana de Educación Matemática. Nro. 18.
- Batanero, C. (2001). “Didáctica de la Estadística” Grupo de Investigación en Educación Estadística. Departamento de Didáctica de la Matemática. Universidad de Granada.
- Biggs, J. B., Collis, K. F. (1991) Multimodal learning and the quality of intelligence behavior. H. A. Rowe (ed.). Intelligence reconceptualization and measurement. Hillsdale, New Jersey: Lawrence Erlbaum Associates Publishers
- Cochran, W. G. (1983). Planning and analysis of observational studies, L. E. Moses and F. Mosteller (eds.). New York: Wiley.
- Cooper L. L., Shore F. S. (2010). “The Effects of Data and Graph Type on Concepts and Visualizations of Variability” Journal of Statistics Education, Volume 18, Number 2.
- Del Puerto, S., Seminara, S., y Minnaard, C. (2007). “Identificación y análisis de los errores cometidos por los alumnos en estadística descriptiva”, Revista Iberoamericana de Educación, ISSN: 1681-5653

- Dunn, C. W. (1961). "Multiple comparisons among means." *Journal of the American Statistical Association*, 56, 52-64.
- Fisher, R. A. (1926). "The arrangement of field experiments." *Journal of the Ministry of Agriculture of Great Britain* 33, 503-5 13.
- Fisher, R. A. (1935). *The design of experiments*. Edinburgh: Oliver and Boyd.
- Fleiss, J. L. (1981). *Statistical methods for rates and proportions*, 2d ed. New York: Wiley.
- Hader, R. J. (1973). "An improper method of randomization in experimental design ." *The American Statistician* 27, 82-84.
- Huynh, H., Feldt, L. S. (1970). "Conditions under which mean square ratios in repeated measurements designs have exact F-distributions ." *Journal of the American Statistical Association* 65, 1582-1 589.
- Mauchly, J. W. (1940). "Significance test for sphericity of a normal n-variate distribution. " *Annals of Mathematical Statistics* 1 1,204-209.
- Mevarech, Z. R. (1983). A deep structure model of students statistical misconceptions. *Educational Studies in Mathematics*, 14, 415-429.
- Tanur, J. F., Mosteller, F., Kruskal, W., Link, R., Pietero, R., Rising, G., y Lehman, E. (1978). *Statistics: A pide to the unknown*. San Francisco: Holden-Day.
- Wild, C. J., Pfannkuch, M. (1999). "Statistical Thinking in Empirical Enquiry", *International Statistical Review*, 67, 3, 223-265

ANEXO I

Cuestionario

CUESTIONARIO DE ENCUESTA

A continuación se le tomará un test, sobre una investigación que se está realizando en la facultad de ingeniería económica y ciencias sociales de la UNI, agradeciendo desde ya su colaboración.

Marque con una X, su respuesta:

¿Cuál es tu escuela profesional?	1. Ing. Económica	2. Ing. Estadística	3. Otra
----------------------------------	-------------------	---------------------	---------

¿En qué ciclo te encuentras?	1. I	2. II	3. III	4. IV	5. V	6. VI	7. VII	8. VIII	9. IX	10. X
------------------------------	------	-------	--------	-------	------	-------	--------	---------	-------	-------

<p>Preg1</p> <p>Dos constructoras presentan cinco proyectos de construcción, con áreas distintas en algunos distritos de Lima. ¿Cuál de las dos constructoras tiene mayor riesgo en sobrevaluar sus proyectos?</p>			
1. Constructora A	2. Constructora B	3. Ambos tienen el mismo riesgo	4. Ninguno

<p>Preg2</p> <p>Se presentan los siguientes histogramas correspondientes a las puntuaciones obtenidas en un examen por los estudiantes en dos aulas distintas, considerando un mínimo de 0 y un máximo de 100 puntos. ¿En cuál de las dos aulas las puntuaciones son más homogéneas respecto a la otra aula?</p>			
1. Aula A	2. Aula B	3. Ambas son homogéneas	4. NA

<p>Preg3</p> <p>Los siguientes gráficos de barra representan la distribución de tipos de sangre de tres grupos étnicos. ¿Cuál de estos grupos étnicos es menos homogéneo?</p>			
1. Modern Maya	2. Lapps	3. Japanese	4. Ninguno

Preg4	<table border="1"> <tr> <td>Cara del dado</td> <td>Nº de veces</td> </tr> <tr> <td>1</td> <td></td> </tr> <tr> <td>2</td> <td></td> </tr> <tr> <td>3</td> <td></td> </tr> <tr> <td>4</td> <td></td> </tr> <tr> <td>5</td> <td></td> </tr> <tr> <td>6</td> <td></td> </tr> <tr> <td>total</td> <td>60</td> </tr> </table>	Cara del dado	Nº de veces	1		2		3		4		5		6		total	60
Cara del dado		Nº de veces															
1																	
2																	
3																	
4																	
5																	
6																	
total	60																
Imagina que lanzas un dado 60 veces. Llena la siguiente tabla escribiendo ¿cuántas veces crees que saldrá cada número?																	
¿Por qué consideras este resultado?.....																	

Preg5	<p>Se te proporciona una cinta métrica y te piden medir el contorno de las cabezas de 10 compañeros de tu salón de clase, escribe tus posibles resultados:</p> <p>_____</p> <p>_____</p>
¿A qué se debe estos resultados?.....	

Preg6	<p>Se proporciona una cinta métrica a 10 de tus compañeros de tu salón de clase y se les piden medir el contorno de tu cabeza, escribe los posibles resultados:</p> <p>_____</p> <p>_____</p>
¿A qué se debe estos resultados?.....	

Preg7	A: 10; 20; 30; 40; 50; 60 cm.	B: 10; 10; 10; 60; 60; 60 cm.
Se presentan dos conjuntos de bloques de madera cuyas longitudes son:		
¿Cuál de los dos conjuntos es más homogéneo?		
Conjunto A	2. Conjunto B	3. Tienen igual homogeneidad
		4. NA

Preg8	<table border="1"> <tr> <td>Marca A</td> <td>97</td> <td>83</td> <td>75</td> <td>82</td> <td>98</td> <td>65</td> <td>75</td> </tr> <tr> <td>Marca B</td> <td>78</td> <td>56</td> <td>87</td> <td>54</td> <td>89</td> <td>65</td> <td></td> </tr> </table>	Marca A	97	83	75	82	98	65	75	Marca B	78	56	87	54	89	65	
Marca A		97	83	75	82	98	65	75									
Marca B	78	56	87	54	89	65											
Se realiza un estudio sobre dos compañías que producen zapatillas Marca A y Marca B, con el objeto de determinar cuál de ellas tiene un mejor programa de control de calidad sobre el desgaste de su producto. Se tomo dos muestras por cada marca y se controló el número de horas de uso hasta que se detecta un desgaste significativo, se obtuvieron los siguientes resultados:																	
¿Cuál de las dos marcas de zapatilla tiene un mejor programa de control de calidad sobre el desgaste?																	
1. Marca A	2. Marca B	3. Tienen igual programa	4. No se puede afirmar														