



**TECANA AMERICAN UNIVERSITY
POSTDOCTORAL PROGRAM IN BIOSTATISTICS**

**CURSO I:
B909 Fundamentos de Bioestadística**

**CURSANTE:
Oralia Nolasco Jáuregui**

“Por la presente juro que soy la única autora de este documento en el cual se integran opiniones personales en base a mi trabajo y fruto de la investigación bibliográfica”

**Jalisco México a 4 de Noviembre de 2019
Versión 1.0**

Índice

1.1 Introducción.....	13
1.1.1 Estadística.....	13
1.1.2 Estadística Descriptiva	14
1.1.3 Estadística Inferencial.....	14
1.1.3.1 Inferencia Estadística.....	14
2.1 Desarrollo	15
2.1.1 Significación Estadística.....	15
2.1.2 Variables.....	16
2.1.2.1 Tipos de Variables	16
2.1.2.1.1 Variables Cualitativas.....	16
2.1.2.1.2 Variables Cuantitativas.....	16
2.1.2.1.3 Escalas Nominales.....	16
2.1.2.1.4 Escalas Ordinales.....	16
2.1.3 Gráficas.....	16
2.1.4 Tipos de Eventos	16
2.1.4.1 Eventos Mutuamente Excluyentes.....	16
2.1.4.2 Eventos NO Mutuamente Excluyentes.....	17
2.1.5 Distribuciones de Probabilidad Discretas.....	17
2.1.5.1 Distribución Binomial	17
2.2 Estadísticos Descriptivos.....	17
2.2.1 Medias de Tendencia Central	17
2.2.1.1 Media Aritmética.....	17
2.2.1.2 La Mediana	18
2.2.1.3 La Moda.....	18
2.2.2 Medidas de Tendencia Central	18
2.2.3 Medidas de Dispersión	19
2.2.3.1 Diagramas de Dispersión.....	19
2.2.3.2 Rango Cuartílico.....	19
2.2.3.3 Medidas de Forma	20
2.2.4 Medidas de Posición	20
2.3 Distribución Normal.....	21

2.3.1	Características de la Distribución Normal.....	21
2.3.2	Test de la Distribución Normal	21
2.4	Muestreo Estadístico	22
2.4.1	Ventajas del Muestreo	22
2.5	Intervalos de Confianza	23
2.6	Teoría de Decisión Estadística	23
2.6.1	Hipótesis Estadística.....	23
2.6.2	Nivel de Significación	23
2.6.3	Ensayos de Una y Dos Colas.....	23
2.7	Teoría de Pequeñas Muestras	23
2.7.1	Distribución t de Student	23
2.8	Pruebas Paramétricas: Requisitos.....	24
2.9	Pruebas NO Paramétricas: Requisitos	24
2.9.1	Principales Pruebas NO Paramétricas	24
2.10	Regresión y Correlación	24
2.10.1	Regresión	24
2.11	Pruebas Robustas.....	25
2.11.1	Métodos robustos para el cálculo de medidas de posición: Medidas robustas de tendencia central	25
2.11.1.1	Media α -winsorizada muestral	25
2.11.1.2	Media α -recortada muestral.....	25
2.11.1.3	Mediana muestral	25
2.11.1.4	Estimador de Huber.....	25
2.11.2	Métodos robustos para el cálculo de medidas de posición: Medidas robustas de dispersión.....	25
2.11.2.1	Desviación absoluta mediana estandarizada.....	25
2.11.2.2	Cuasi desviación típica α -winsorizada muestral.....	25
2.11.3	Pruebas robustas de contraste de hipótesis	25
2.12	Prueba de Rachas.....	26
2.13	Estructura del proceso de Inferencia Estadística	26
2.13.1	Inferencia Estadística.....	26
2.13.2	Técnicas de Muestreo	26
2.13.3	Error Aleatorio.....	26

2.14	Proceso de selección de individuos para el estudio.....	26
2.14.1	Población Diana.....	26
2.14.2	Población de Estudio	26
2.14.3	Muestra	26
2.15	Concepto e importancia y utilidad práctica del error alfa (α) y error beta (β)	27
2.15.1	Determinación del tamaño muestral	27
2.15.2	Error tipo I (alfa).....	27
2.15.2.1	Recomendaciones para disminuir el error alfa (α) tipo I.....	27
2.15.3	Error tipo II (beta).....	27
2.15.3.1	Recomendaciones para disminuir el error beta (β) tipo II.....	28
2.15.3.2	Importancia del error alfa y error beta.....	28
2.16	Cálculo del tamaño muestral de un parámetro	28
2.16.1	Variabilidad del parámetro que se desea estimar	28
2.16.1.1	Variabilidad del parámetro	28
2.16.2	Precisión	28
2.16.3	Nivel de confianza	28
2.17	Nivel de confianza de 95%.....	28
2.17.1	Fórmulas para el cálculo del número de sujetos necesarios para la relación de un estudio objetivo es la estimación de un parámetro	28
2.17.2	Cálculo del tamaño muestral para contraste de hipótesis.....	29
2.17.3	Valor de significación estadística p.....	29
2.17.4	Evaluación de las pruebas diagnósticas.....	29
2.18	Sensibilidad	29
2.19	Especificidad	29
2.20	Exactitud.....	30
2.21	Valor Predictivo Positivo (VPP) de la prueba diagnóstica.....	30
2.22	Valor Predictivo Negativo (VPN) de la prueba diagnóstica.....	30
2.22.1	Falsos Positivos versus Falsos negativos.....	30
2.23	Prevalencia.....	31
2.24	Razón de Verosimilitud.....	31
2.25	Construcción de la curva ROC para cada prueba diagnóstica.....	31
2.25.1	Comparación del área de las curvas ROC en muestras independientes	31
2.25.2	Comparación del área de las curvas ROC en muestras apareadas	31

2.25.3 Ventajas de las curvas ROC en la evaluación de pruebas diagnósticas	31
2.25.4 Análisis Estadístico	32
2.25.5 Elección del punto de corte	32
2.26 Postulados de Galen y Gambino.....	32
2.26.1 Criterio de máxima Sensibilidad Diagnóstica	32
2.26.2 Criterio de máxima Especificidad Diagnóstica	32
2.26.3 Criterio de alto Valor Predictivo Positivo	32
2.26.4 Criterio de máxima Exactitud Diagnóstica.....	32
2.27 Datos Categóricos.....	32
2.27.1 Cualitativa versus cualitativa. Nominales	32
2.27.2 Cualitativa versus Cualitativa. Ordinales	33
2.28 Test de χ^2 de Pearson.....	33
2.29 Test de χ^2 de Yates.....	33
2.30 Fisher	33
2.31 Razón de Verosimilitud de χ^2	34
2.32 Test de McNemar para datos apareados	34
2.33 Q de Cochran.....	34
2.34 Test de χ^2 de independencia de Pearson	34
2.34.1 Contrastes a posteriori	34
2.35 Test de la Q de Yule	34
2.35.1 Coeficiente Gamma	34
2.32.2 Características de la Q de Yule.....	35
2.36 Coeficiente PHI Φ	35
2.37 Nominal x Nominal (Tablas 2 x 2).....	35
2.38 Limitaciones del test χ^2 de Pearson como medida de asociación	36
2.39 Coeficiente de Contingencia (C)	36
2.40 Tau de Goodman y Kruskal.....	36
2.41 Coeficiente Lambda λ	36
2.42 Características de la Odd Ratio (OR)	36
2.43 Tau-b de Kendall	37
2.44 Coeficiente Gamma γ	37
2.45 D de Somers	37

2.46	Cuantitativas versus cuantitativas.....	38
2.46.1	Regresión lineal simple	38
2.46.2	Correlación	38
2.46.3	Comparación de métodos analíticos	38
2.46.4	Relación directa e inversa.....	38
2.47	Covarianza de dos variables X e Y	38
2.48	Covarianza y coeficiente de correlación de Pearson (r)	39
2.48.1	Coeficiente de correlación lineal de Pearson.....	39
2.49	Propiedades de r	39
2.50	Coeficiente de determinación r^2	39
2.51	Otros coeficientes de correlación	39
2.52	Cualitativa dicotómica versus cuantitativas	40
2.53	Test de T de Student para dos muestras independientes	40
2.53.1	Condiciones de aplicación (1)	40
2.53.2	Condiciones de aplicación (2)	40
2.54	Test de U de Mann-Whitney	40
2.54.1	Condiciones de aplicación.....	40
2.54.2	Inconvenientes	40
2.54.3	Pasos	40
2.55	Test de Wilcoxon.....	41
2.55.1	Prueba No paramétrica o de libre distribución	41
2.55.2	Condiciones de aplicación.....	41
2.56	Análisis de Regresión Logística Simple.....	41
2.57	Modelo de Análisis de Regresión Simple.....	41
2.58	Cualitativa policotómica versus cuantitativa.....	42
2.59	Análisis de la varianza para un factor (ANOVA).....	42
2.59.1	Condiciones de aplicación.....	42
2.59.2	Análisis de la varianza para un factor.....	42
2.59.3	Test ANOVA: Observaciones	42
2.59.4	Reflexiones: existen varios incumplimientos para el ANOVA.....	42
2.59.5	Análisis de la Varianza (ANOVA) para muestras dependiente so apareadas (medidas repetidas).....	43
2.60	Pruebas Post Hoc Paramétrico.....	43

2.60.1 Pruebas Post Hoc No Paramétrico.....	43
2.61 La prueba de Kruskal-Wallis (comparar más de dos grupos)	43
2.61.1 Prueba de la Mediana	44
2.62 Test de Jonkheere-Terpstra.....	44
2.63 Comparación de > 2 grupos apareados.....	44
2.63.1 Comparación de > de 2 grupos apareados	44
2.64 Prueba de Esfericidad de Mauchly	44
2.65 Test de Friedman	45
2.66 Coeficiente de concordancia W de Kendall	45
2.67 Medidas de concordancia	45
2.68 Análisis de concordancia: Variables Cualitativas	46
2.69 Índice Kappa de Cohen.....	46
2.70 Índice Kappa Ponderado.....	46
2.70.1 Limitaciones del estadístico Kappa	46
2.70.2 Pruebas de homogeneidad de las kappas	46
2.70.3 Comparación de kappas: Distribución de χ^2 de Pearson.....	47
2.71 Medidas de concordancia	47
2.72 Evaluación de métodos diagnósticos cuantitativos	47
2.72.1 Evaluación metrológica	47
2.72.2 Especificaciones de la calidad para la imprecisión analítica.....	47
2.72.3 Fuentes de variación biológica	47
2.72.4 Cálculo de la variación total (C_v).....	47
2.72.5 Diagnósticos cuantitativos.....	48
2.72.6 Comparación de métodos diagnóstico cuantitativos	48
2.72.7 Análisis de concordancia: variables cuantitativas	48
2.72.7.1 Métodos diagnósticos	48
2.72.7.2 Métodos de regresión.....	48
2.72.7.3 Métodos gráficos	48
2.72.8 ¿Cuál es el método estadístico más adecuado para calcular el error que comete un método diagnóstico cuantitativo?	48
2.73 Regresión lineal por mínimos cuadrados.....	49
2.73.1 Condiciones de aplicación	49
2.74 Coeficiente de correlación de Pearson.....	49

2.75 Coeficiente de correlación intraclase.....	49
2.76 Valoración del coeficiente de correlación intraclase	50
2.76.1 Limitaciones del coeficiente de correlación intraclase.....	50
2.77 Modelos de efectos fijos	50
2.78 T de Student para datos apareados	50
2.79 Coeficiente de concordancia de Lin	50
2.80 Método de regresión ortogonal de Deming	50
2.81 Modelos de regresión de Passing-Bablok.....	51
2.82 Método de mínimos cuadrados ponderados	51
2.83 Método gráfico de Bland y Altman	51
2.83.1 Interpretación.....	51
2.83.2 Repetibilidad.....	51
2.84 Recomendaciones de Zaydman	52
2.84.1 Modelos de Regresión de Passing-Badblock.....	52
2.84.2 Método de mínimos cuadrados (método regresión lineal simple	52
2.84.3 Método de mínimos cuadrados ponderados	52
2.84.4 Método de regresión ortogonal de Deming	52
2.85 ANOVA de 2 factores	52
2.86 Regresión Logística Multivariable	52
2.86.1 Aplicación de los modelos de regresión Logística	52
2.86.2 Objetivos de los modelos de regresión logística	52
2.86.3 Ventajas de los modelos de regresión logística Multivariable	53
2.86.4 Funciones de los modelos de regresión logística multivariante	53
2.86.5 Variables explicativas o Predictoras en el modelo de Regresión Logística	53
2.86.5.1 Variables Cualitativas.....	53
2.86.5.2 Variables Cuantitativas.....	53
2.86.6 Regresión Logística	54
2.86.6.1 Estimación de coeficientes β_i	54
2.86.6.2 El signo de β_i tiene un significado importante.....	54
2.86.7 Definición de las hipótesis nula H_0 y alternativa H_a en la Regresión Logística	54
2.86.8 Procedimientos para evaluar el contraste de hipótesis en la Regresión Logística	54
2.86.8.1 Evaluación estadística del coeficiente β_i	54

2.86.8.2 Evaluación estadística de Wald.....	54
2.86.8.3 Evaluación estadística de la razón de verosimilitud (L_0).....	55
2.86.8.4 Evaluación estadística del cambio del estadístico $-2LL_0$	55
2.86.8.5 Evaluación del estadístico Z^2	55
2.86.10 Utilidades de la regresión lineal múltiple.....	55
2.86.11 Requisitos para el uso de la regresión lineal múltiple.....	55
2.86.12 Modelo de regresión lineal múltiple.....	55
2.86.13 Colinealidad en la Regresión Lineal Múltiple.....	56
2.86.13.1 Diagnóstico de colinealidad.....	56
2.86.13.2 Diagnóstico de Colinealidad.....	56
2.87 El test de Kaiser-Meyer-Olkin (KMO).....	56
2.88 Tipos de variable del estudio Horbar et al.....	57
2.89 Estudio de Marshall et al. (NEOCOSUR).....	57
2.90 Reflexiones Finales.....	57
3.1 Conclusiones.....	58
Anexos.....	65

Lista de Tablas

Tabla 1. Estadísticamente significativo a lo clínicamente relevante (tabla basada en Portillo, J. D., 2011, Pág. 32).	65
Tabla 2. Pruebas estadísticas de contraste de hipótesis más aplicables a cada situación (tabla basada en Portillo, J. D., 2011, Pág. 83).....	66
Tabla 3. Estadística descriptiva (tabla basada en Portillo, J. D., 2011, Pág. 146).....	66
Tabla 4. Tipos de error al contrastar hipótesis (tabla basada en Portillo, J. D., 2011, Pág. 179).....	66
Tabla 5. Fórmulas para el cálculo estadístico $(1 - \beta)$ en función del tipo de diseño del estudio (Portillo, J. D., 2011, Pág. 206).....	67
Tabla 6. Decisiones que se pueden tomar en un test de contraste de hipótesis después de conocer el valor de significación estadística p (tabla basada en Portillo, J. D., 2011, Pág. 212).....	67
Tabla 7. Análisis Estadístico (tabla basada en Portillo, J. D., 2011, Pág. 359).....	67
Tabla 8. Tipos de Test Analíticos Estadísticos Bivariantes de contraste de hipótesis más frecuentes (tabla basada en Portillo, J. D., 2011, Pág. 361).....	68
Tabla 9. Pruebas Estadísticas de contraste de hipótesis más aplicables a cada situación (tabla basada en Portillo, J. D., 2011, Pág. 361).....	68

Tabla 10. Coeficientes de Correlación (tabla basada en Portillo, J. D., 2011, Pág. 687).	69
Tabla 11. Índice Kappa de Cohen (medida dicotómica), (tabla basada en Portillo, J. D., 2011, Pág. 996).	69
Tabla 12. Valores del índice Kappa (tabla basada en Portillo, J. D., 2011, Pág. 999). ..	69
Tabla 13. Coeficiente de Correlación Intraclase (tabla basada en Portillo, J. D., 2011, Pág. 1059).	69
Tabla 14. Escala de McBrige para el coeficiente de concordancia de Lin (tabla basada en Portillo, J. D., 2011, Pág. 1088).	70
Tabla 15. Análisis de la Varianza (ANOVA) de dos factores (tabla basada en Portillo, J. D., 2011, Pág. 1167).	70
Tabla 16. Evaluación del modelo de regresión logística multivariante mediante el programa estadístico SPSS (tabla basada en Portillo, J. D., 2011, Pág. 1186).	70
Tabla 17. Evaluación del modelo de regresión logística multivariante, si se rechaza la hipótesis nula H_0 ($p < 0,05$) significa que el modelo de regresión es significativo (tabla basada en Portillo, J. D., 2011, Pág. 1187).	70
Tabla 18. Modelos de Regresión Múltiple de Cox (tabla basada en Portillo, J. D., 2011, Pág. 1330).	71

Lista de Figuras

Figura 1. Selección de muestral, imagen basada en Portillo, J. D., 2011, Pág. 164.	73
Figura 3. Exactitud, imagen basada en Portillo, J. D., 2011, Pág. 237.	74
Figura 5. Muestras, imagen basada en Portillo, J. D., 2011, Pág. 852.	75
Figura 7. Evaluación Metrológica, imagen basada en Portillo, J. D., 2011, Pág. 852. 76	
Figura 8. Cálculo de la Variación Total (CV_t), imagen basada en Portillo, J. D., 2011, Pág. 1033.	76

Resumen

El estudio de la bioestadística se encarga de las mediciones de la medicina y la biología. Es indispensable entender la estadística descriptiva inferencial y descriptiva para realizar una significación en los estudios en el campo de la bioestadística. Los primeros pasos dentro de un ambiente estadístico se describen en este documento, viene bien un campo introductorio como lo describe el capítulo 1 y un tratamiento profundo de las variables y utilización de técnicas de manera correcta descrito en el capítulo 2, finalmente el capítulo 3 tiene las conclusiones y opiniones personales que plasmo con respecto a la bibliografía leída con la que se construyó este documento.

Objetivo General

Las características de las muestras son esenciales para la selección de la técnica a implementar. Dentro de las generalidades de este documento es escoger las definiciones más acertadas de una serie de bibliografía, mostrar por ejemplos y una serie de metodologías las técnicas. Hay diferencias muy bastas entre las muestras tales como si son apareadas o independientes y las demostraciones que debe seguir el investigar para demostrar su equivalencia y/o diferencia según sea el caso.

Objetivos Específicos

Dentro de los objetivos particulares a tratar en este documento se encuentra la identificación del tipo de variables dentro de las muestras, el tipo de gráfico y sus características. Los tipos de eventos y las distribuciones. Además de las características de las muestras es necesario saber cómo es su comportamiento a lo largo de las distribuciones es decir los límites de confianza y a partir de estos estadísticos el investigador podrá tomar decisiones. Conocerá el nivel de significación de sus muestras, podrá cambiar de técnicas en caso de necesario ya que conocerá perfectamente sus datos y parámetros que los rigen. Dependiendo del número de muestras podrá trabajar con pruebas robustas y será consiente de los tipos de errores y los algoritmos de control individual o familiar de ellos. Además, de las pruebas de rachas, procesos de inferencia, desviación, dispersión, variabilidad, precisión, sensibilidad, especificidad y exactitud de los datos.

Justificación

Es imprescindible en la medicina y áreas biológicas tener exactitud en los datos, es inconcebible un mal estudio por un desconocimiento en los métodos estadísticos y no entender el impacto de un VPP (Valor Predictivo Positivo) ó de un VPN (Valor Predictivo Negativo) para un paciente en algún padecimiento particular ó general en la sociedad. Es verdad que los estudios de control y los individuos expuestos son estudios comparativos son procesos sujetos a mucha precisión, pero también es una realidad que hay errores en las técnicas y es obligación del investigador conocerlas.

Capítulo 1

1.1 Introducción

1.1.1 Estadística

La estadística es la ciencia que se ocupa de los métodos y procedimientos para recoger, clasificar, hallar regularidades y analizar los datos, siempre y cuando la variabilidad e incertidumbre sea una causa intrínseca de los mismos; así como de realizar inferencia a partir de ellos, con la finalidad de ayudar a la toma de decisiones y en su caso formular predicciones (Portillo, J. D., 2011, Pág. 11).

La palabra Estadística proviene del latín “status”. En la antigüedad chinos, egipcios, hebreos, griegos y romanos la practicaban en recuentos de población y riquezas. Con el tiempo se perfeccionó mediante el método matemático y probabilístico hasta generalizar su estudio y uso a cualquier actividad científica. La estadística se puede dividir en Descriptiva e Inferencial. La primera se refiere a la recolección de datos, agrupación, presentación, análisis e interpretación de datos obtenidos de una población, tales como: métodos estadísticos, de tendencia central, de posición, de dispersión y de forma. La segunda se refiere a hacer una estimación e inferencia para la toma de decisiones. La fuente de los datos es referente a la población para la estadística descriptiva y la muestra para la estadística inferencial (María Rosaría Ruggiero, YacelliBustamante, Claudia Mark, 1948, Pág. 6).

La estadística constituye un instrumento científico que requiere el cumplimiento de numerosas condiciones de numerosas condiciones para producir resultados válidos (Portillo, J. D., 2011, Pág. 54):

1. Es precioso que los experimentos tengan un diseño muy escrupuloso.
2. Es inexcusable aplicar en cada circunstancia aquellos métodos que en ella son idóneas y controlar rigurosamente su validez.
3. Nunca es legítimo deducir que un resultado significativo es clínicamente relevante.
4. Una correlación entre dos variables acabará siendo siempre significativa $p < 0,05$ si la muestra es lo suficientemente grande.
5. Por eso en algunos casos, se dice que: “La estadística es un instrumento, que si no se usa adecuadamente, puede demostrar cualquier mentira”.

Si los datos de un estudio se analizan con mucho enfoques distintos y con la suficiente intensidad hasta torturarlos, acaban dando los resultados que al investigador le hubiese gustado obtener, aunque tales resultados suelen ser científicamente nefastos (Portillo, J. D., 2011, Pág. 57).

- Estadística en Medicina

La medicina es una ciencia estadística, pues siempre se trabaja con la probabilidad de enfermar y con la oportunidad incierta de diagnóstico y del tratamiento. El médico debe adecuar el conocimiento científico y tecnológico a la situación clínica personal y social del paciente del que se ocupa en ese conocimiento, ya que no existen enfermedades sino enfermos. El buen juicio médico logra un equilibrio entre el riesgo que comporta toda intervención diagnóstica y/o terapéutica y el beneficio esperable de dicha intervención en el paciente concreto, lo que obliga a combinar ciencia y arte, pues no es fácil trabajar con la incertidumbre, y sólo la estadística nos proporciona el instrumento adecuado que nos

permite convivir con el azar y con las limitaciones del conocimiento médico (Portillo, J. D., 2011, Pág. 14).

1.1.2 Estadística Descriptiva

Describe, presenta y analiza un grupo de datos utilizando métodos numéricos y gráficos que resumen y presentan la información contenida en ellos (Portillo, J. D., 2011, Pág. 12).

1.1.3 Estadística Inferencial

Apoyándose en el cálculo de probabilidad y a partir de datos de una muestra (procedente de una población), efectúa estimaciones, decisiones, predicciones u otras generalizaciones sobre un conjunto mayor de datos. Permite generalizar los datos obtenidos a partir de una muestra a la población de la que proceden. Se basa en la teoría de las probabilidades y trabaja con los datos que le proporciona la estadística descriptiva (Portillo, J. D., 2011, Pág. 13).

1.1.3.1 Inferencia Estadística

La inferencia estadística constituye una parte de la estadística en la que se hacen estimaciones e inferencias para la toma de decisiones. Aquí utilizaremos técnicas de muestreo apropiadas a fin de estudiar determinadas características de la población que nos interesa analizar, tomando en cuenta que la muestra a estudiar, debe ser representativa de la población (María Rosaría Ruggiero, Yacelli Bustamante, Claudia Mark, 1948, Pág. 79).

Capítulo 2

2.1 Desarrollo

2.1.1 Significación Estadística

Un resultado estadísticamente significativo (ver Tabla 1) con $p < 0,005$ sin relevancia clínica no deja de ser una anécdota; un hallazgo clínicamente con $p > 0,1$ no puede ser asumido como concluyente, ya que no podemos atribuir con seguridad el hallazgo clínico a la intervención realizada (no se puede descartar la influencia del azar) (Portillo, J. D., 2011, Pág. 33).

Altman (Altman, D. G., & Bland, J. M., 1996, Pág. 311) y Bland (Argimon, J. M., 2002, Págs. 701-703) alertaron sobre un error común en la bibliografía biomédica: interpretar un resultado estadísticamente no significativo, cuando se comparan dos tratamientos, como sinónimos de su equivalencia terapéutica (Portillo, J. D., 2011, Pág. 37).

Un valor de significación estadística mayor $p > 0,05$ se etiqueta como no significativo. Estos estudios se les denominan negativos, un término que implica erróneamente que los tratamientos son iguales, cuando lo único que sucede es que no se ha podido demostrar una diferencia, ver Tabla 1. Ambos conceptos son distintos. Un resultado que no alcance significación estadística quiere decir que (si en realidad no existe diferencia), es probable que la que la observada en el estudio pueda haberse producido simplemente por el azar. No debe interpretarse como indicativo de que no existe una diferencia en la realidad sino tan sólo de que no puede descartarse esta posibilidad, en especial en los estudios que han incluido a pocos sujetos (Portillo, J. D., 2011, Pág. 38).

En la bibliografía científica se está abusando del término significación estadística. Numerosos investigadores dan por supuesto que si en un estudio hallan una probabilidad (p) inferior a 0,005, ello certifica la existencia tanto de la validez científica de su observación como de la relevancia clínica de la misma. Y ello está lejos de ser cierto. Su fe ciega el valor de la p contrasta con los detractores de su significado (ateos de la estadística) (Portillo, J. D., 2011, Pág. 40).

Neyman et al (Shen, W., Neyman, J., Pearson, E. S., Bolch, G., Greiner, S., de Meer, H. & Trivedi, K. S., 2018, Pág. 285-289) lanzaron este tema exclusivamente bajo el punto de vista de la toma de decisión ante un problema que fuera trascendente y en el que, urgiera tal toma de decisión: rechazar o no la hipótesis nula con sus consecuencias derivadas. Lo que ocurre es que se ha universalizado demasiado y se lo ha sacado de su contexto, ya que se aplica a cualquier inferencia estadística, aunque el fin de la investigación no suponga obligación alguna de toma de una decisión tan drástica, basada en la ley del todo o nada (Portillo, J. D., 2011, Pág. 49).

La utilidad de la p es una medida de la influencia del azar en la posible asociación entre las variables analizadas, este azar debe ser siempre contemplado, evaluado y cuantificado mediante el valor de la p . La p NO es un indicador de fuerza de la asociación entre variables, ni de su importancia. Los valores de la p deben ser considerados solo como una guía y no como base de conclusiones definitivas e irrevocables (Portillo, J. D., 2011, Pág. 52).

2.1.2 Variables

Los tipos de variables utilizados por la estadística se dividen en: cualitativa y cuantitativa. Las primeras con aquellas que únicamente se pueden describir y se dividen en nominal y ordinal; en cambio las cuantitativas se pueden contar y medir. Las clases constituyen cada grupo de variantes. Los límites de una clase pueden ser el tipo aparente y del tipo real. La ventaja de utilizar límites reales es que no se rompe la continuidad, no existe la posibilidad de que un valor caiga en la frontera y no se altera la marca de la clase. La desventaja del uso de límites reales es que dan la impresión de continuidad, trabajan con decimales y no son recomendables para variables discretas (María Rosaría Ruggiero, YacelliBustamante, Claudia Mark, 1948, Pág. 8).

2.12.1 Tipos de Variables

2.12.1.1 Variables Cualitativas

Representan una cualidad o atributo que clasifica a cada individuo en una de varias categorías. Pueden ser de dos tipos (Portillo, J. D., 2011, Págs. 74-85):

1. Dicotómicas o binarias: La situación más sencilla es aquella en la que se clasifica cada individuo en uno de dos grupos (hombre/mujer, enfermo/sano, fumador/no fumador).
2. Policotómicas: Es necesario un mayor número de categorías (color de los ojos, grupo sanguíneo, profesión, etcétera).

2.12.1.2 Variables Cuantitativas

Son aquellas que pueden medirse, cuantificarse o expresarse numéricamente. Pueden ser de dos tipos:

1. Continuas: Son aquellas que si admiten tomar cualquier valor dentro de un rango numérico determinado, con uno o varios decimales (peso, talla, edad).
2. Discretas: Son aquellas que NO admiten todos los valores intermedios en un rango. Suelen tomar solamente valores enteros (número de hijos, abortos, partos, hermanos).

2.12.1.3 Escalas Nominales

Es una forma de observar o medir en la que los datos se ajustan por categorías que no mantienen una relación de orden entre sí.

2.12.1.4 Escalas Ordinales

En las escalas utilizadas para medirlas existe un cierto orden, grado o jerarquía entre las categorías.

2.1.3 Gráficas

La representación de datos numéricos es por medio de coordenadas o dibujos que hacen visible la relación o gradación que esos datos guardan entre sí, como: los histogramas, diagrama circular, pictogramas, diagrama de caja, etc. (María Rosaría Ruggiero, YacelliBustamante, Claudia Mark, 1948, Pág. 14).

2.1.4 Tipos de Eventos

2.1.4.1 Eventos Mutuamente Excluyentes

Sean A y B dos subconjuntos de S. Decimos que A y B son mutuamente excluyentes si $A \cap B = 0$, es decir, la aparición de uno de ellos impide la ocurrencia simultánea del otro.

$$P(A \cup B) = P(A) + P(B)$$

2.1.4.2 Eventos NO Mutuamente Excluyentes

En este caso los eventos parecen ser mutuamente excluyentes, pero existe una intersección en los eventos A y B, es decir, puede ocurrir que en el espacio muestral exista un evento que excluya a los eventos A y B, por lo tanto debemos restar dicha intersección para evitar contarla en las probabilidades de A y B.

$$P(A \cup B) = P(A) + P(B) - P(A \cap B)$$

Teorema de Bayes sea S un espacio muestral que contiene n eventos A_1, A_2, \dots, A_n . Sea B un evento de S tal que $P(B) > 0$. La probabilidad de cualquier evento A_i , $i = 1, 2, \dots, n$ el evento B es: $P(A_i / B) = P(A_i \cap B) / P(B)$, (María Rosaría Ruggiero, YacelliBustamante, Claudia Mark, 1948, Págs. 55-59).

2.1.5 Distribuciones de Probabilidad Discretas

2.1.5.1 Distribución Binomial

Los principios básicos de esta distribución teórica binomial los desarrolló el matemático suizo Jacob Bernoulli, en el siglo XVII. La distribución binomial proporciona la probabilidad de que un resultado específico ocurra de un número determinado de pruebas independientes. Bajo el supuesto de que durante n pruebas, la probabilidad de éxito en una sola prueba se mantenga fija; la determinación de la probabilidad de obtener un número dado de éxitos r, en las n pruebas, se simplifica utilizando la distribución binomial. Un experimento binomial es aquel cuyo experimento consta de n pruebas idénticas en donde cada respuesta tiene dos posibles resultados; éxito o fracaso.

Distribución de Poisson, la distribución teórica de Poisson se debe al matemático francés Simeón Poisson, es aplicable a fenómenos aleatorios que se caracterizan por el número de sucesos que ocurren en un determinado periodo de tiempo o en un determinado espacio. La hipótesis básica en este tipo de fenómenos es que los sucesos son independientes. Puede usarse para determinar la probabilidad de eventos poco frecuentes, es decir, proporciona la probabilidad de que un resultado suceda un número específico de veces cuando la cantidad de pruebas es grande y la probabilidad de ocurrencia es pequeña. Esta distribución tiene aplicación cuando estamos en presencia de "eventos raros", los cuales se caracterizan por tener una probabilidad de ocurrencia muy pequeña en una probabilidad de ocurrencia muy pequeña en una población muy grande, por lo que generalmente $\mu = n.p < 5$, de esta forma, la distribución Binomial tiende a la distribución de Poisson (María Rosaría Ruggiero, YacelliBustamante, Claudia Mark, 1948, Págs. 68-78).

2.2 Estadísticos Descriptivos

2.2.1 Medias de Tendencia Central

2.2.1.1 Media Aritmética

Es el promedio aritmético en una distribución de datos. Es el más usado de los promedios, siempre y cuando la serie no presente valores extremos, ya que esto distorsiona el valor de la media, en este caso sería aconsejable otra medida (la mediana por ejemplo). Es el valor típico representativo de un conjunto de datos y se caracteriza por depender de todas las medidas que forman la serie de datos.

Ventajas: Fácil de entender y calcular. Hace uso de todos los datos de una distribución. Es el más conocido. Es usada en la inferencia estadística. Se presenta a manipulación algebraica.

Desventajas: Puede ser influenciada por los valores extremos que hagan perder su medida central. En el caso de variables discretas, el valor no es exactamente alguno de los

datos, por lo que a veces se redondean. La media aritmética no puede ser calculada para una distribución con intervalos de clases abiertas, esto es, cuando los elementos están agrupados en intervalos de clase de tipo “por encima de” o “por debajo de”.

2.2.1.2 La Mediana

Es el valor de la variable que equidista de ambos extremos de la distribución cuando está ordenada de manera creciente, es decir, es el valor que deja por debajo de él el 50% de los datos, consecuentemente por encima de ella mediana se halla el 50% de los datos. El valor de la mediana puede coincidir o no con un valor de la serie de datos. No es un estadígrafo suficiente, ya que no considera a todos los datos.

Ventajas: Los valores extremos no la afectan ya que está determinada por el número de observaciones y no por el valor de las mismas. Se puede calcular aunque los valores extremos sean abiertos.

Desventajas: No se presenta a tratamientos algebraicos, es necesario ordenar las variantes antes de que se pueda calcular la mediana. Es poco conocida.

2.2.1.3 La Moda

Se define como el valor que tiene más frecuencia en una serie de datos. Puede que no exista o bien que existan varios valores candidatos a ser moda. Una distribución puede tener un solo valor modal se dice que es unimodal; si dos variantes se repiten con la misma frecuencia se dice bimodal. Si hay más de dos variantes con la misma frecuencia, la distribución es multimodal. Si todos los datos tienen la misma frecuencia, no existe moda. La moda corresponde al valor donde el histograma alcanza la máxima altura. No es un estadígrafo suficiente, ya que no toma en cuenta todos los datos y si algunos datos se alteran, es posible que la moda siga igual. Carece de significación en distribuciones que contengan pocos datos y no ofrezcan una marcada tendencia central. No es afectada por los valores extremos (María Rosaría Ruggiero, YacelliBustamante, Claudia Mark, 1948, Págs. 18-26).

2.2.2 Medidas de Tendencia Central

1. Media Aritmética (ver Ecuación 2.1):
 - Principal medida de tendencia central, pero menos robusta que la media geométrica o media armónica.
 - Se deja influir mucho por los valores extremos de la distribución.
 - Se suman todos los valores y se divide por el número total de individuos.
 - Expresión con pocos decimales.
2. Media Geométrica (ver Ecuación 2.2).
3. Media Armónica (ver Ecuación 2.3).
4. Media Ponderada (ver Ecuación 2.4):
 - Medida de tendencia central robusta, de cálculo fácil.
 - No se deja influir mucho por los valores extremos de la distribución.
 - Coincide con el valor central del conjunto ordenado de observaciones que deja por debajo y por encima a la mitad de las observaciones realizadas.
 - Se debe utilizar siempre en dos casos:
 - o Bajo tamaño muestral (para muestras pequeñas) y/o
 - o Cuando existen observaciones extremas (periféricos).
5. Mediana
6. Moda

2.2.3 Medidas de Dispersión

Las medidas de tendencia central o de localización dan una visión del grupo, pero la misma es incompleta. Ellas dan información acerca del centro de los datos pero no qué tan dispersos son ellos mismos.

Las medidas de variabilidad indican qué tan diseminados son los datos del grupo al cual se le calcula la medida. Si un grupo tiene una baja variabilidad esto indica que está compuesto por individuos aproximadamente iguales, los datos están poco esparcidos, están bastante agrupados. La mayoría de los puntajes estarán alrededor de la medida de tendencia utilizada. En este caso se dice que los individuos poseen características homogéneas.

Proporcionan la variación de datos respecto a la media aritmética. Su valor se encuentra en relación directa con la dispersión de los datos a mayor dispersión de ellos, mayor desviación típica; a menor dispersión menor desviación típica. Es la medida de dispersión adecuada para la medida de tendencia central es la media. Es susceptible a los valores extremos. La mayor utilidad de la desviación típica se presenta en una distribución normal, al encontrar que en los intervalos la media $\pm \sigma$ se concentra en el 68% aproximadamente de los datos, la media $\pm 2\sigma$ se concentra en el 95% de los datos y que la media $\pm 3\sigma$ se concentra aproximadamente en todos los datos.

1.

2.2.3.1 Diagramas de Dispersión

El diagrama de dispersión da una visualización y un medio más simple para estudiar la relación entre dos variables. En este diagrama, cada uno de ellos en pares de observaciones (x_i, y_i) se marca con un solo punto en la gráfica. Con la disposición de los puntos en la gráfica se detecta el patrón indicativo de la naturaleza de la forma funcional básica de los datos. Los puntos dentro del diagrama pueden sugerir una relación lineal, curvilínea o ninguna relación (María Rosaría Ruggiero, Yacelli Bustamante, Claudia Mark, 1948, Pág. 104).

2.2.3.2 Rango Cuartílico

Esta medida se basa en el cuartil 1 y 3, por lo que excluye al 25% inferior de los datos y el 25% superior de los mismos. Esto indica que el rango cuartílico mide la concentración de los datos en el 50% central de los mismos. El rango cuartílico expresa la distancia entre Q_1 y Q_3 . Dentro de sus características se encuentra que no es una medida segura de dispersión, ya que su valor se encuentra afectado por el 50% de los datos, 25% inferior y 25% superior. Igualmente obvia la distribución de datos entre Q_1 y Q_3 . Es posible que dos series de datos con diferentes distribuciones presenten igual rango cuartílico, por ser iguales en los valores de Q_1 y Q_3 . Una medida de dispersión derivada del rango cuartílico, es la desviación semicuartil, que es la semisuma de Q_1 y Q_3 . Es la medida de variabilidad adecuada cuando la mediana es la medida de tendencia central.

2. Varianza $[s^2]$ (ver Ecuación 2.5), se define como el cuadrado de la desviación típica.

3. Desviación Estándar $[s]$ (ver Ecuación 2.6), es la desviación que presenta los datos con respecto a la mediana o a la media aritmética. Se usa cuando las desviaciones extremas influyen en la desviación típica:

-Medida de dispersión que nos suministra mucha información sobre la naturaleza de una variable.

-Cuanto más separados están los valores de la media mayor será su dispersión, y mayor será su desviación estándar.

-Se debe representar como media (desviación estándar), y NO como media \pm desviación estándar.

-Su magnitud SI depende de las unidades de medida de la variable. NO permite la comparación directa de la dispersión de varias variables.

-Mide el grado de dispersión de todos los individuos que forman la muestra o la población (σ).

4. Coeficiente de variación [CV] es un índice de variabilidad con respecto a la media aritmética. Nos permite expresar el grado de homogeneidad del grupo de datos considerados en su conjunto (María Rosaría Ruggiero, YacelliBustamante, Claudia Mark, 1948, Págs. 33-38) (ver Ecuación 2.7):

-Medida de dispersión que relaciona que proporción porcentual de la media representa la desviación estándar.

-Representa la desviación estándar medida en unidades de magnitud relativas a la media.

-Permite la comparación directa de la dispersión de varias variables.

-Su magnitud NO depende de las unidades de medida de la variable.

5. Error Estándar de la Media [EEM] (ver Ecuación 2.8): Mide el grado de dispersión de todas las posibles muestras que pudieran extraerse de la población. Mide el grado de dispersión de las medias muestrales respecto a la media población μ (desconocida).

2.2.3.3 Medidas de Forma

Una distribución queda bien caracterizada mediante la tendencia central y la variabilidad, pero quedará mejor si éstas medidas con acompañadas con medidas que describan la asimetría y apuntamiento de la distribución.

Sesgo es un indicador que mide el grado de asimetría o falta de simetría de una distribución. Si el sesgo es igual a 0, hay simetría. Si el sesgo es mayor a cero, la cola derecha es más larga que la izquierda respecto al valor central. Se dice que la asimetría es positiva. Si el sesgo es menor a cero, la cola derecha es más corta que la izquierda con respecto al valor central. La asimetría es negativa.

Curtosis es el grado de apuntamiento de una distribución con respecto a una curva modelo o curva normal de Laplace- Gauss. Se divide en Leptocúrtica, Mesocúrtica y Platicúrtica, (María Rosaría Ruggiero, YacelliBustamante, Claudia Mark, 1948, Págs. 42-45).

1. Asimetría (coeficiente de asimetría γ_1 de Fisher)(Portillo, J. D., 2011):
 - Simétrica (coeficiente de asimetría $\gamma_1=0$), ver Ecuación 2.9.
 - Asimétrica (coeficiente de asimetría $\gamma_1 \neq 0$).
 - o Asimetría positiva (coeficiente de asimetría $\gamma_1 > 0$).
 - o Asimetría negativa (coeficiente de asimetría $\gamma_1 < 0$).
2. Curtosis o apuntamiento (coeficiente γ_2 curtosis de Fisher) mide el grado de aplastamiento de una curva, ver Ecuación 2.10.
 - o Mesocúrtica ($\gamma_2 = 0$).
 - o Platicúrtica ($\gamma_2 < 0$).
 - o Leptocúrtica ($\gamma_2 > 0$).

2.2.4 Medidas se Posición

Percentiles son valores que dividen la distribución en 100 partes iguales y nos dan la situación de los datos según el lugar que ocupan en tanto por ciento. Son 99 y se denotan

por P_1, P_2, \dots, P_{99} corresponde a aquel valor que supera $k\%$ de datos a lo más y a la vez es superado por el $(100 - k)\%$ de los datos a lo máximo.

Deciles son valores que dividen la distribución en 10 partes iguales, son 9 y se denotan por D_1, \dots, D_9 . Así D_2 corresponde a aquel valor que supera al 20% de datos a lo más y a la vez es superado por el 80% de los datos a lo máximo.

Cuartiles son los valores que dividen la distribución en 4 partes iguales, son 3 y se denotan por Q_1, \dots, Q_3 , (María Rosaría Ruggiero, YacelliBustamante, Claudia Mark, 1948, Pág. 27).

2.3 Distribución Normal

Distribución Normal, Gauss y Laplace estudiaron la distribución de errores de las observaciones, concluyendo que todas las distribuciones estadísticas se aproximan a una curva que llamaron Normal, cuando el número de observaciones es grande. La distribución normal es la distribución de probabilidad más famosa. Fue descubierta por primera vez sus trabajos en 1733. Sin embargo, dos astrónomos matemáticos, Pierre-Simon Laplace de Francia y Carl Friedrich Gauss de Alemania, se ocuparon de establecer los principios científicos de la distribución normal. La media μ y la desviación estándar σ son los parámetros de la distribución normal, esto es, μ y σ determinan completamente la ubicación de las cantidades y la forma de la curva. Para la curva normal tipificada, se tiene que la media es igual a 0 y la desviación típica es 1, (María Rosaría Ruggiero, YacelliBustamante, Claudia Mark, 1948, Págs. 68-78).

El 100% de los valores deben estar por encima y por debajo de 3 veces la desviación estándar (s) valores entre $X \pm 3s$. La distribución debe ser simétrica (Portillo, J. D., 2011, Pág. 124). La asimetría en valor absoluto γ_1 debe ser menor que 2 veces su error estándar, ver Ecuación 2.11. La distribución debe ser mesocúrtica. La curtosis en valor absoluto γ_2 debe ser menor que 2 veces su error estándar, ver Ecuación 2.12.

2.3.1 Características de la Distribución Normal

1. Tener campana.
2. Sera simétrica.
3. Mesocúrtica(no excesivamente plana ni muy picuda).
4. Coincidir la media, mediana y moda.
5. Tener aproximadamente el 95% de sus valores del intervalo $media \pm 2$ desviaciones estándar.

2.3.2 Test de la Distribución Normal

1. Test de Shapiro-Wilk (la prueba más potente para muestra < 30 casos) (Portillo, J. D., 2011, Pág. 140).
2. Test de D'Agostino.
3. Test de Anderson-Darling.
4. Test de chi-cuadrado de bondad de ajuste.
5. Test de Contraste de asimetría: Sirve para contrastar la hipótesis de que el coeficiente de asimetría de Fisher γ_1 es cero. Propiedad que verifica la distribución normal ($\gamma_1 = 0$).
6. Test de Curtosis: Contraste de apuntamiento, sirve para contrastar la hipótesis de que el coeficiente γ_2 curtosis de Fisher (coeficiente de apuntamiento) es cero. Propiedad que verifica la distribución normal ($\gamma_2 = 0$).
7. Test de kolmogoroff-Smirnoff-Lilliefors (test más potente para muestral > 30 casos):
 - Es el test de normalidad más difundido en los programas estadísticos.

- Otorga un menor peso a las observaciones extremas (periféricos o Outlier) por ello es menos tolerante a las desviaciones que normalmente se producen en estos tramos distales.
- Cuando se dispone de un número suficiente de datos (amplio volumen muestral), los demás tests de normalidad serán capaces de detectar diferencias pequeñas aun cuando estas no sean relevantes para la mayor parte de los objetos, por eso se recomienda el test de Kolmogorov-Smirnov (se considera uno de los tests más potentes para muestras mayores de 30 casos).
- Se basa en la idea de comparar la función de distribución acumulada de los datos observados con la de una distribución normal, midiendo la máxima distancia entre ambas curvas.
- Como en todos los tests de contraste de hipótesis, la hipótesis nula (H_0) se rechaza cuando el valor del estadístico calculado supera un cierto valor crítico (tabulado) que se obtiene de una tabla de cálculo de probabilidades.
 - o $\rho > 0,05$ se acepta H_0 (se cumple la normalidad), se pueden usar test paramétricos.
 - o $\rho < 0,05$ se acepta H_a (no se cumple con la normalidad), se deben usar test NO paramétricos, ver Tabla 1.

8. Test de Jarque-Bera (Portillo, J. D., 2011, Pág. 144):

- $H_0 : \gamma_1 = \gamma_2 = 0$, lo que significa simetría y curtosis exactas ($\gamma = 0$).

- $H_1 : \gamma_1 \neq \gamma_2$ ó $\gamma_2 \neq 0$ lo que significa que existe curtosis ($\gamma \neq 0$) o asimetría.

La consecuencia es que si existe asimetría o curtosis falla la normalidad:

- $H_0 : \rho > 0,05 \rightarrow$ simetría y curtosis exactas ($\gamma = 0$) \rightarrow test paramétricos.

- $H_1 : \rho < 0,05 \rightarrow$ curtosis ($\gamma \neq 0$) o asimetría \rightarrow test NO paramétricos.

Al contrastarlo conjuntamente (curtosis y asimetría) exige un fallo mayor de alguna de ellas o de ambas para rechazarse, ver Ecuación 2.13.

2.4 Muestreo Estadístico

El conjunto de técnicas que nos permiten diseñar la muestra más apropiada para un experimento, garantizado que esta sea representativa de la población de origen y controlar los errores cometidos, es lo que se conoce como Muestreo Estadístico. Dependiendo de la investigación a realizar, utilizaremos el método que mejor se adecúa al mismo (muestreo aleatorio simple, muestreo estadístico, entre otros), (María Rosaría Ruggiero, YacelliBustamante, Claudia Mark, 1948, Págs. 79-85).

2.4.1 Ventajas del Muestreo

Economía y rapidez en su realización. Más alcance en la investigación. Más entrenamiento, formación y control del personal. Mayor rapidez de procesamiento y presentación de resultados. Fácil verificación posterior. Mayor confiabilidad de los datos obtenidos.

2.4.2 Limitaciones del Muestreo

No permite hacer cálculos, tabulaciones o proyecciones con respecto a áreas o grupos pequeños. Presenta el error de muestreo. Se requiere de una preparación estadístico-matemática.

2.4.3 Distribuciones Muestrales

Si tenemos una población de tamaño N , de ella es posible extraer con o sin reemplazamiento una serie de muestras (n). Para cada muestra se puede calcular cualquier estadístico. Existe un llamado Teorema del Límite Central el cual en esencia dice que

cuando n aumenta, la distribución de las medias tiende a una ley normal con media y varianza específica.

2.4.4 Distribuciones Muestral de Proporciones (Población Finita)

Se considera todas las posibles muestras de tamaño n extraída de una población y para cada muestra se determina la proporción de p de éxito. Entonces se obtiene una distribución muestral de proporciones cuya media es μ_p y desviación típica σ_p y viene dada por una media.

2.5 Intervalos de Confianza

Si queremos estimar un parámetro de una población mediante una muestra de tamaño n , podemos obtener muchos valores distintos de ese parámetro muestral. Lo más recomendable es encontrar un intervalo alrededor del valor del estimador, acompañado de alguna medida que nos diga la confianza que se puede tener de que ese intervalo contenga el verdadero valor del parámetro (María Rosaría Ruggiero, YacelliBustaminate, Claudia Mark, 1948, Pág. 86).

2.6 Teoría de Decisión Estadística

Son decisiones sobre poblaciones, tomadas a partir de la información muestral de las mismas.

2.6.1 Hipótesis Estadística

Es una conjetura que se realiza respecto a una población, más concretamente, respecto a un parámetro de la población el cual cuantifica una característica de ella. Se formulan hipótesis con el solo propósito de rechazarla o aceptarla. Trabajemos con las llamadas hipótesis y la hipótesis alternativa.

1. Hipótesis Nula: Se denota por H_0 y es la hipótesis que se establece con el propósito de ver su posible rechazo.

2. Hipótesis Alternativa: se denota por H_a y es cualquier hipótesis que difiere de la hipótesis nula, referida la misma medida estadística, por lo tanto contradice a H_0 .

2.6.2 Nivel de Significación

La probabilidad máxima con la que en el ensayo de una hipótesis se puede cometer un error del tipo I, se llama nivel de significación del ensayo. El mismo es denotado por α . Comúnmente se usan los niveles 5% y 1%. Por ejemplo si es de 5%, se tiene que se está con un 95% de confianza de que se toma la decisión adecuada. En tal caso, se dice que la hipótesis ha sido rechazada al nivel de significación del 0.05, lo que significa que se puede cometer error con una probabilidad de 0.05.

2.6.3 Ensayos de Una y Dos Colas

Si $H_1 = \mu = \mu_0$, estas pruebas conducen a una prueba de dos colas. Una prueba de dos colas es apropiada cuando el investigador no espera algo a priori respecto al valor a observar en la prueba. Solo desean saber si la muestra es diferente de la media de la población (María Rosaría Ruggiero, YacelliBustaminate, Claudia Mark, 1948, Págs. 88-90).

2.7 Teoría de Pequeñas Muestras

2.7.1 Distribución t de Student

Es una distribución de probabilidad continua y simétrica, pero más extendido que la normal y su amplitud depende del tamaño de la muestra; cuando ésta es muy grande coincide con la normal. El número de grados de libertad de un estadístico se define como el número de observaciones independientes de la muestra. Los valores del estadístico "t" vienen

expresados en función del nivel de confianza y grados de libertad de la prueba (María Rosaría Ruggiero, YacelliBustamante, Claudia Mark, 1948, Pág. 98).

2.8 Pruebas Paramétricas: Requisitos

Las variables cuantitativas tiene que ser continuas, y los datos siguen una distribución normal, además que los datos numéricos se expresan con la media y la desviación estándar o el EEM. Tiene homogeneidad entre las varianzas muestrales. Generalmente el tamaño muestral debe de ser menor a 30. Los datos presentan un mayor poder de contraste que los tests son no paramétricos, y pueden analizar interacciones entre variables predictoras. Cuando se dan las condiciones de aplicación, las pruebas paramétricas tienen más potencia que las no paramétricas, pero, cuando esto no es así, el riesgo alfa puede sr mayor que el especificado de antemano (Portillo, J. D., 2011, Pág. 147).

2.9 Pruebas NO Paramétricas: Requisitos

Las variables cuantitativas pueden ser ordinales o discretas. Los datos siguen una distribución libre (no necesariamente normal), uno de los grupos pueden tener una distribución normal mientras que el otro no. Los datos numéricos se expresan con la mediana y los intervalos intercuartílicos. Tiene homogeneidad sólo en uno o más de un grupo de las varianzas muestrales. Tamaño de las muestras menores a 30. Presentan un menor poder estadístico que los tests paramétricos. Trabajan con la simple ordenación y recuento (asignando ranking) a los valores de la variable sin importar el tipo de distribución (Portillo, J. D., 2011, Págs. 148-149).

2.9.1 Principales Pruebas NO Paramétricas

1. Chi cuadrado de Pearson.
2. Test exacto de Fisher.
3. Chi cuadrado de Mc Nemar.
4. Q de Cochran.
5. U de Mann Whitney.
6. Wilcoxon.
7. Kruskall Wallis.
8. Friedman.

2.10 Regresión y Correlación

De una población de tamaño N se pueden estudiar dos variables X e Y, los cuales los tendremos por pares de observaciones (x_i, y_i) los cuales los disponemos en forma de tablas. Dichas tablas expresan cómo se distribuyen las observaciones en función de los pares (x_i, y_i) por lo que reciben el nombre de distribución bidimensional de frecuencias. En estos casos se pretende estudiar de este conjunto, la relación existente entre las variables. Si $S_{xy} > 0$ entonces decimos que X e Y siguen el mismo comportamiento, si $S_{xy} < 0$ entonces decimos que X e Y se mueven en sentido contrario y si $S_{xy} = 0$ entonces no existe relación entre las variables. Más con la covarianza, existe otra magnitud que indica el grado de relación entre las dos variables, es llamada Coeficiente de Correlación r. Si r tiende a 1 las variables tienen el mismo sentido, si r tiende a -1 las variables tienen relación inversa y si r tiende a 0 no hay relación lineal entre las variables (María Rosaría Ruggiero, YacelliBustamante, Claudia Mark, 1948, Pág. 103).

2.10.1 Regresión

Consiste en obtener una ecuación que se pueda usar para predecir o calcular el valor de una variable correspondiente a un valor dado de la otra variable. Existen muchos modelos de regresión, a saber, lineal, cuadrático, logarítmico, logístico, exponencial, entre otros.

Regresión Lineal, el diagrama de dispersión sugiere la idea de intentar expresar la relación entre dos variables mediante una línea de regresión que sea recta. Si tenemos dos variables X e Y, decimos que están relacionadas según una línea recta cuando sus valores satisfacen la relación $Y = aX + b$, (María Rosarí Ruggiero, Yacelli Bustamante, Claudia Mark, 1948, Pág. 106).

2.11 Pruebas Robustas

Otra alternativa a los métodos paramétricos; la utilización de los llamados métodos robustos. Estos métodos son menos potentes que los paramétricos, pero se muestran superiores a los no paramétricos clásicos. La principal de sus ventajas es que no se afectan por la existencia de datos anómalos. Los métodos robustos pueden ser de utilidad para la realización de inferencias sin tener que depurar los datos extremos. Como parámetros de localización (posición central) se recomienda en primer lugar, el estimador de Huber y, en segundo lugar, la media α -recortada muestral (con $\alpha = 0,2$). Como medida de dispersión se utiliza la desviación absoluta mediana estandarizada. Métodos robustos para el contraste de hipótesis con los parámetros antes referidos, se pueden construir intervalos de confianza robustos y realizar contrastes de hipótesis (Portillo, J. D., 2011, Pág. 150).

2.11.1 Métodos robustos para el cálculo de medidas de posición: Medidas robustas de tendencia central

2.11.1.1 Media α -winsorizada muestral

Se sustituye un determinado porcentaje, α , (20% generalmente) de valores extremos a cada lado de la muestra por el valor más próximo no sustituido.

2.11.1.2 Media α -recortada muestral

Se eliminan las k observaciones extremas de cada lado en lugar de winsorizarlas, calculando la media aritmética de las observaciones restantes.

2.11.1.3 Mediana muestral

Divide la distribución de dos partes con el mismo número de elementos.

2.11.1.4 Estimador de Huber

Se encuentra dentro de los determinados M-estimadores, que generalizan al estimador de máxima verosimilitud con buenas propiedades de robustez y eficiencia. En este caso se descartan las observaciones que sean mayores (o menores) a una constante (Portillo, J. D., 2011, Pág. 151).

2.11.2 Métodos robustos para el cálculo de medidas de posición: Medidas robustas de dispersión

2.11.2.1 Desviación absoluta mediana estandarizada

Es la mediana de las desviaciones absolutas a la mediana.

2.11.2.2 Cuasi desviación típica α -winsorizada muestral

En la que sustituye un determinado porcentaje de valores extremos a cada lado de la muestra por el valor más próximo no sustituido (Portillo, J. D., 2011, Pág. 152).

2.11.3 Pruebas robustas de contraste de hipótesis

Métodos robustos para el contraste de hipótesis con los parámetros antes referidos, se pueden construir intervalos de confianza robustos y realizar contrastes de hipótesis. El test de Yuen es una alternativa a las pruebas de la t de Student y U de Mann-Whitney para tamaños muestrales pequeños y distribuciones no normales. Para comparar más de dos

medias, cuando no puede utilizarse el ANOVA clásico, deberíamos recurrir a los clásicos no paramétricos de Kruskal-Wallis o, mejor, a métodos robustos como la generalización robusta del test Welch, que utiliza medias a-recortadas muestrales. Se debería recurrir a métodos no paramétricos (Spearman) o, mejor, a métodos robustos como el coeficiente de porcentaje ajustado poblacional y el estimador robusto de regresión medio bponderado (Portillo, J. D., 2011, Págs. 153-154).

2.12 Prueba de Rachas

Sirve para determinar si una muestra de observaciones es o no aleatoria, es decir, si las observaciones de una determinada secuencia son independientes entre sí. En una serie temporal, las observaciones no son aleatorias: lo que ocurre con una observación cualquiera depende, de las características de una observación anterior. En una muestra aleatoria, lo que ocurre con una observación cualquiera sea independiente de las características de las observaciones anteriores. Esta prueba permite determinar si el número de rachas (R), observado en una determinada muestra de tamaño n es lo suficientemente grande o lo suficientemente pequeña como para poder rechazar la hipótesis de independencia-aleatoriedad de las observaciones. El concepto de Racha hace referencia a una secuencia de observaciones de un mismo tipo (Portillo, J. D., 2011, Págs. 155-158).

2.13 Estructura del proceso de Inferencia Estadística

2.13.1 Inferencia Estadística

Es el conjunto de procedimientos estadísticos que permiten pasar de lo particular (la muestra) a lo general (la población).

2.13.2 Técnicas de Muestreo

Es el conjunto de procedimientos que permiten elegir muestras que sean representativas de la población (de todos los individuos de la población).

2.13.3 Error Aleatorio

Es el error que se comete al estudiar un problema en base a una muestra elegida al zar (es siempre un error inevitable). La imposibilidad de estudiar a todos los individuos de la población obliga al investigador a trabajar con muestras, lo que supone siempre cometer un error aleatorio (Portillo, J. D., 2011, Pág. 165).

2.14 Proceso de selección de individuos para el estudio

2.14.1 Población Diana

Conjunto de individuos al que hace referencia la pregunta principal del objetivo de estudio. Es la población a la que se desearía generalizar los resultados obtenidos. Se define principalmente sobre la base de características demográficas, ver Figura 1.

2.14.2 Población de Estudio

Es un subconjunto diana definida por criterios de selección establecidos en un protocolo. Esta población presenta determinadas características geográficas y temporales que la hacen accesibles al estudio.

2.14.3 Muestra

Es un conjunto de individuos que pueden ser estudiados en el momento de realizar el estudio y que cumplen con los criterios definidos en el protocolo de estudio (Portillo, J. D., 2011, Pág. 167).

2.15 Concepto e importancia y utilidad práctica del error alfa (α) y error beta (β)

2.15.1 Determinación del tamaño muestral

-Estudios para determinar parámetros poblaciones (sólo error alfa) (Portillo, J. D., 2011, Pág. 174):

- Estimación de una proporción.
- Estimación de una media.

-Estudios para contraste de hipótesis (error alfa y beta):

- Comparación de dos proporciones.
- Comparación de dos medias.
- Correlación entre dos variables cuantitativas.

-Rentabilidad diagnóstico de una prueba-test:

- Exactitud.
 1. Sensibilidad.
 2. Especificidad.
- Seguridad:
 1. Valor predictivo positivo (VPP).
 2. Valor predictivo negativo (VPN).

2.15.2 Error tipo I (alfa)

El investigador rechaza la H_0 siendo esta verdadera en la población (el investigador concluye que existe diferencia cuando en realidad no la hay) (Portillo, J. D., 2011, Pág. 182). Se concluye que hay diferencias cuando realmente no las hay. Falsos Positivos. Se detecta significación estadística ($p < 0,05$) y se rechaza H_0 , cuando en realidad ésta es verdadera. La probabilidad de cometer este error es lo que mide la “ p ”, ver Tabla 4.

2.15.2.1 Recomendaciones para disminuir el error alfa (α) tipo I

Disponer de una teoría que guíe la investigación, evitando el “salir de pesca” con el computador buscando asociaciones entre variables (Sánchez-Villegas, A., 2014). Disminuir el número de test estadísticos llevados a cabo en el estudio. Depurar la base de datos para evitar errores de valores extremos que puedan producir hallazgos significativos (Portillo, J. D., 2011, Pág. 184). Utilizar valores de alfa más reducidos (0,01 ó 0,001). Reproducir el estudio. Si al reproducir el estudio se obtienen resultados similares, estaremos más seguros de no estar cometiendo el error de tipo I (Pita Fernández, S., &Pértega Díaz, S., 2000, Págs. 191-195).

2.15.3 Error tipo II (beta)

El investigador no rechaza la H_0 siendo esta falsa en la población (el investigador concluye que ha sido incapaz de encontrar una diferencia que existe en la realidad) (Portillo, J. D., 2011, Pág. 183). Cuando los resultados del estudio en turno, los resultados son NO significativos ($p > 0,05$) se concluye diciendo que no hay diferencias (“estudio negativo”) cuando realmente SI las hay, pero nosotros no hemos sido capaces de detectarlas, ver Tabla 4.

- La probabilidad de tomar esta decisión (y aceptar la H_0) siendo incorrecta se llama error beta.
- En realidad existen diferencia pero el estudio no ha incapaz de detectarlas. Son los Falsos Negativos.

2.15.3.1 Recomendaciones para disminuir el error beta (β) tipo II

Incrementar el tamaño de la muestra. Estimar el poder estadístico del estudio ($1 - \beta$) (Portillo, J. D., 2011, Pág. 185). Incrementar el tamaño del efecto a detectar. Incrementar el valor de alfa. Utilizar paramétricos (más potentes) en lugar de test no paramétricos (Pita Fernández, S., & Pértega Díaz, S., 2000, 191-195).

2.15.3.2 Importancia del error alfa y error beta

-Determinación del tamaño muestral: Estudios para determinar parámetros poblacionales (sólo error alfa) (Portillo, J. D., 2011, Pág. 186).

- Estimación de una proporción.
- Estimación de una media.

-Estudios para contraste de hipótesis (error alfa y beta).

- Comparación de dos proporciones.
- Comparación de dos medias.
- Correlación entre dos variables cuantitativas.

-Rentabilidad diagnóstico de una prueba-test

- Sensibilidad.
- Especificidad.
- Valor predictivo positivo (VPP).
- Valor predictivo negativo (VPN).

2.16 Cálculo del tamaño muestral de un parámetro

2.16.1 Variabilidad del parámetro que se desea estimar

Variabilidad del parámetro que se desea estimar (Portillo, J. D., 2011, Pág. 187):

1. Variables cuantitativas: varianza.
2. Variable cualitativas: 1- P.

2.16.1.1 Variabilidad del parámetro

- Cuando esta variabilidad es desconocida.
- Bibliografía.
- Datos propios previos.
- Estudio piloto.

2.16.2 Precisión

Precisión (i) con la que se desea obtener la estimación (amplitud deseada del IC). Se fija i previa al estudio. Cuanta más precisión más estrecha será el IC, y mayor deberá ser el volumen muestral.

2.16.3 Nivel de confianza

Nivel de confianza deseado (generalmente 95%), que corresponde a un valor de alfa 0,05. Este valor indica el grado de confianza de que el verdadero valor poblacional μ desconocido del parámetro en estudio se sitúe en el intervalo obtenido.

2.17 Nivel de confianza de 95%

2.17.1 Fórmulas para el cálculo del número de sujetos necesarios para la relación de un estudio objetivo es la estimación de un parámetro

1. Estimación de una variable Cualitativa (proporción – prevalencia)(Portillo, J. D., 2011, Pág. 188), ver Ecuación 2.14.
2. Estimación de una variable Cuantitativa (media), ver Ecuación 2.15.

Donde:

N = número de sujetos necesarios.

$Z^2\alpha$ = Valor del estadístico Z correspondiente al riesgo alfa fijado ($\alpha = 0,05$).

P = Valor de la proporción (prevalencia) que se supone que existe en la población.

S^2 = Varianza de la distribución de la variable cuantitativa que se supone que existe en la población.

I = Precisión con que se desea estimar el parámetro.

2.17.2 Cálculo del tamaño muestral para contraste de hipótesis

1. Magnitud de la diferencia a detectar que tenga interés clínico (Portillo, J. D., 2011, Pág. 196), ver Tabla 5.

- Comparación de dos medias.
- Comparación de dos proporciones: P1 y P2.

2. Aproximación de los parámetros de la variable que se estudia.

3. Nivel de confianza deseado o seguridad del estudio (generalmente 95%), que corresponde a un valor de α de 0,05 (riesgo de cometer un error alfa).

4. Poder estadístico ($1 - \beta$) (riesgo de cometer un error β).

5. Definir si la hipótesis va a ser unilateral o bilateral, ver Figura 2.

- Bilateral: cualquiera de los dos parámetros a comparar puede ser mayor o menor que el otro (no hay dirección) ($A \neq B$).
- Unilateral: cuando se considera que uno debe ser mayor que el otro (dirección de las diferencias) ($A < B$).

2.17.3 Valor de significación estadística p

Decisiones que se pueden tomar en un test de contraste de hipótesis después de conocer el valor de significación estadística p, ver Tabla 6.

2.17.4 Evaluación de las pruebas diagnósticas

1. Exactitud (validez, efectividad o rentabilidad) (Portillo, J. D., 2011, Pág. 216):

- Sensibilidad
- Especificidad

2. Seguridad: valores predictivos:

- Valor predictivo positivo (VPP).
- Valor predictivo negativo (VPN).

3. Razón de verosimilitud.

4. Curvas roc.

5. Reproducibilidad.

6. Invasividad.

7. Coste económico (URV).

2.18 Sensibilidad

- Sensibilidad = $VP / (VP + FN)$

-Es la probabilidad de clasificar correctamente a un individuo enfermo.

-Es la probabilidad de que para un sujeto enfermo se obtenga en la prueba un resultado positivo.

-Es la capacidad del test para detectar la enfermedad.

-Probabilidad de obtener un resultado positivo cuando el individuo tiene la enfermedad.

-Mide su capacidad para detectar la enfermedad cuando está presente.

-Su valor depende del error Beta (β) (Portillo, J. D., 2011, Pág. 226).

2.19 Especificidad

- Especificidad = $VN / (VN + FP)$

- Es la probabilidad de clasificar correctamente a un individuo sano.
- Es la probabilidad de que para un sujeto sano se obtenga un resultado negativo.
- Es la capacidad del test para detectar a los sanos.
- Es la probabilidad de obtener un resultado negativo cuando el individuo no tiene la enfermedad.
- Mide la capacidad para descartar la enfermedad cuando no está presente.
- Su valor depende del error ALFA (α)(Portillo, J. D., 2011, Pág. 228).

2.20 Exactitud

- Es la principal cualidad clínica de una prueba diagnóstica.
- Es la capacidad de una prueba diagnóstica para clasificar correctamente a los pacientes.
- Es la capacidad de una prueba diagnóstica para clasificar correctamente a los individuos en subgrupos clínicamente relevantes (enfermos y NO enfermos).
- Es su forma más simple es la capacidad para distinguir entre dos estados de salud (enfermos y NO enfermos)(Portillo, J. D., 2011, Pág. 230).
 - Exactitud = $(VP + VN) / (VP + FN + FP + VN)$
 - Exactitud = Aciertos / (Aciertos + Fallos)

2.21 Valor Predictivo Positivo (VPP) de la prueba diagnóstica

- $VPP = VP / \text{Total de Positivos}$
- $VPP = VP / (VP + FP)$
- Es la probabilidad de padecer la enfermedad si obtiene un resultado positivo en el test.
- Se puede definir el VPP como la proporción de resultados válidos (VP) entre todos los resultados positivos de la prueba (VP + FP).
- Su valor depende del error ALFA (α) y de la Potencia Estadística.
- Su valor depende de la Prevalencia (P) de la enfermedad; cuando menor sea la prevalencia menor será el VPP y cuanto mayor sea la prevalencia mayor será el VPP(Portillo, J. D., 2011, Pág. 232).
 - $VPN = VN / (VN + FN)$
 - $VPN = VN / \text{Total de Negativos}$

2.22 Valor Predictivo Negativo (VPN) de la prueba diagnóstica

- Es la probabilidad de que un sujeto con un resultado negativo en la prueba esté realmente sano.
- El VPN es la proporción de resultados válidos (VN) entre todos los resultados negativos (VN + FN).
- Su valor depende del error BETA (β).
- Su valor depende de la prevalencia (P) de la enfermedad en estudio. Cuanto menor sea la prevalencia, mayor será el VPN, y cuanto mayor sea la prevalencia, menor será el VPN(Portillo, J. D., 2011, Pág. 234).

2.22.1 Falsos Positivos versus Falsos negativos

1. Gravedad o importancia en salud pública. Si la enfermedad puede ser tratada y detectada evitando complicaciones posteriores han de utilizarse pruebas con alta Sensibilidad, es decir con baja tasa de Falsos Negativos.
2. Si el coste del proceso asistencial (posibles pruebas invasivas complementarias, exámenes y cuidados médicos y de enfermería requeridos) del enfermo es alto se deben elegir pruebas con alta Especificidad, para reducir los falsos positivos.

3. Si la enfermedad diagnosticaba conlleva consecuencias psicosociales graves, como la infección por VIH, han de minimizarse los falsos positivos, eligiendo pruebas de alta Especificidad.

4. Baja prevalencia de la enfermedad. En los programas de *screening* se eligen pruebas de alta Especificidad, para reducir la tasa de falsos positivos y así garantizar la eficiencia del programa (Portillo, J. D., 2011, Pág. 349).

2.23 Prevalencia

$$\circ \text{ Prevalencia} = (\text{VP} + \text{FN}) / (\text{VP} + \text{FP} + \text{FN} + \text{VN})$$

Los valores predictivos (VPP + VPN) dependen de la prevalencia de la enfermedad en estudio, ver Figura 3.

2.24 Razón de Verosimilitud

- $\text{RV} = \text{Sensibilidad} / (1 - \text{Especificidad})$
- $\text{RV} + = \text{Tasa de VP (\%)} / \text{Tasa de FP (\%)}$

2.25 Construcción de la curva ROC para cada prueba diagnóstica

-La curva ROC es un gráfico que muestra todos los pares posibles sensibilidad/especificidad resultantes de la variación continua de los puntos de corte de los puntos de corte en todo el rango de resultados observados.

-El eje x se sitúa la fracción de FP ó $1 - \text{especificidad}$, definida como $1 - E = \text{FP} / (\text{VN} + \text{FP})$ y calculada en los no enfermos (sanos).

-Las curvas ROC son los mejores índices de la efectividad, rentabilidad o exactitud diagnóstica y proporcionan un criterio unificador en el proceso de evaluación de una prueba, y de comparación simultáneas de varias pruebas diagnósticas.

- El área bajo la curva ROC es un índice de la exactitud o valor global de la prueba: la exactitud máxima corresponderían a un valor del área de 1 y la mínima de 0.5 (Portillo, J. D., 2011, Pág. 295).

2.25.1 Comparación del área de las curvas ROC en muestras independientes

A_1 = área bajo la curva ROC de la técnica 1.

A_2 = área bajo la curva ROC de la técnica 2.

EE_{A1} = error estándar del área bajo la curva ROC de la técnica 1.

EE_{A2} = error estándar bajo la curva ROC de la técnica 2, ver Ecuación 2.16.

Si el estudio se realiza en grupos Independientes, la comparación de áreas mediante métodos estadísticos es directa. Si se utilizan datos apareados o dependientes es necesario calcular previamente la correlación r entre las pruebas coeficientes de correlación entre ambas técnicas para muestras independientes vale cero ($r = 0$).

2.25.2 Comparación del área de las curvas ROC en muestras apareadas

r = coeficiente de correlación entre las áreas de ambas técnicas para muestras dependientes, ver Ecuación 2.17.

2.25.3 Ventajas de las curvas ROC en la evaluación de pruebas diagnósticas

1. Son una representación gráfica fácilmente comprensible de la capacidad de discriminación de la prueba diagnóstica en todo el rango de puntos de corte (Burgueño, M. J., García-Bastos, J. L., & González-Buitrago, J. M., 1995, Págs. 661-70).

2. Son simples, gráficas y fáciles de interpretar visualmente.

3. No requieren un nivel de decisión particular porque está incluido todo el espectro de puntos de corte (Zweig, M. H., & Campbell, G., 1993, Págs. 561-577).

4. Son independientes de la prevalencia, ya que ni la sensibilidad ni la especificidad dependen de ella.
5. Proporcionan una comparación visual directa entre distintas pruebas en una escala común, mientras que otro tipo de gráficos, como los diagramas de puntos o los histogramas de frecuencias requieren diferentes gráficos cuando difieren las escalas.
6. La especificidad y la sensibilidad son fácilmente accesibles en el gráfico, en contraste con los diagramas de puntos y los histogramas (Portillo, J. D., 2011, Pág. 354).

2.25.4 Análisis Estadístico

Tipos de test analíticos estadísticos de contraste de hipótesis según el tipo de variables implicadas en el estudio, ver Tabla 7.

2.25.5 Elección del punto de corte

Criterios basados en el equilibrio de riesgo y beneficios de las decisiones médicas. En un enfoque práctico se basa en el cociente entre los costes de un resultado falso (FP) frente a un falso negativo (FN), en función de la prevalencia (P) de la enfermedad (Zweig, M. H., & Campbell, G., 1993, Págs. 561-577). El valor de corte óptimo se determina hallando el punto de curva ROC cuya tangente a la curva tiene de pendiente el valor de m (McNeil, B. J., Keeler, E., & Adelstein, S. J., 1975, Págs. 211-215).

$$m = (\text{Coste de FP} / \text{Coste de FN}) * ((1-P) / P)$$

2.26 Postulados de Galen y Gambino

2.26.1 Criterio de máxima Sensibilidad Diagnóstica

- Se elige la mayor sensibilidad cuando la enfermedad sea grave y no pueda pasar inadvertida, la enfermedad sea tratable, y los resultados falsos positivos no supongan un traumatismo psicológico y económico en los individuos examinados.

2.26.2 Criterio de máxima Especificidad Diagnóstica

- Se elige la mayor especificidad cuando la enfermedad sea importante pero difícil de curar o incurable, y el hecho de conocer que no se padezca la enfermedad tiene una importante carga sanitaria y psicológica.

2.26.3 Criterio de alto Valor Predictivo Positivo

- Se elige un alto VPP cuando el tratamiento de los falsos positivos pueda tener consecuencias graves en personas que NO tienen la enfermedad.

2.26.4 Criterio de máxima Exactitud Diagnóstica

- Se desea una elevada exactitud (efectividad, valor global o rentabilidad diagnóstica) cuando la enfermedad sea importante pero curable y tanto los falsos positivos (FP) como los falsos negativos (FN) supongan un traumatismo y conlleven consecuencias graves (Galen, R. S., & Gambino, S. R., 1975, Págs. 50-51).

2.27 Datos Categóricos

2.27.1 Cualitativa versus cualitativa. Nominales

Test de χ^2 de Pearson

- Relación entre dos variables cualitativas nominales o policotómicas.
- Comparación de proporciones histogramas (Portillo, J. D., 2011, Pág. 363).

Corrección de Yates

Test exacto de Fisher

- Valores esperados menores de 5, ver Tabla 8.

Razón de verosimilitud

- Evaluación de pruebas diagnósticas.

Test de χ^2 de McNemar

-Datos apareados o emparejados, ver Tabla 9.

Coeficiente de PHI (ϕ)

V de Cramer

Coeficiente de contingencia

Coeficiente de incertidumbre

Tau de Goodman y Kruskal

Lambda de Goodman y Kruskal (coeficiente de predictibilidad de Guttman)

2.27.2 Cualitativa versus Cualitativa. Ordinales

Test de χ^2 de tendencia lineal

Tau-B de Kendall o Stuart

Tau-C de Kendall

Coeficiente Gamma y de Goodman y Kruskal

D de Somers

Correlación de Sperman

Estos estadísticos NO se pueden utilizar cuando alguna de las variables sea nominal (Portillo, J. D., 2011, Pág. 364).

2.28 Test de χ^2 de Pearson

- $\chi^2 = n (ad - bc)^2 / ((a + b)(c + d)(b + d))$, ver Ecuación 2.18.
- Grados de libertad = $[K - 1][F - 1]$

Condiciones necesarias y suficientes de aplicación

1. Ambas variables deben ser cualitativas en escala Nominal.
2. Número de sujetos estudiados igual o superior a 40 ($n > 40$).
3. Ninguno de los valores esperados en cada casilla debe ser menor de 5.
4. No se debe aplicar cuando la variable dependiente o resultado es Ordinal.
5. Es un test no dirigido (por ejemplo, test de planteamiento bilateral), que nos indica si existe o no relación entre dos variables cualitativas pero no en qué sentido se produce tal, ni cuantifica la intensidad de dicha asociación (Portillo, J. D., 2011, Pág. 370).

Nota: Los estudios permiten comprobar si dos variables cualitativas están o no asociadas. Si al final del estudio concluimos que las variables son están relacionadas podremos decir que ambas variables son independientes, con un determinado nivel de confianza, generalmente 95% (Baldessarini, R. J., Finklestein, S., & Arana, G. W., 1983, Págs. 569-573).

2.29 Test de χ^2 de Yates

Corrección de continuidad

En la frecuencia esperada en una de las cuatro celdas es menor de 5, por lo que NO resulta adecuado aplicar el test de χ^2 de Pearson, aunque sí el test de Yates. Las variables sexo y tabaquismo son independientes, porque la probabilidad asociada a los datos del estudio ($p = 0,051$) es inferior a 0,05 (error alfa). Por tanto no podemos descartar la hipótesis nula (H_0) (Portillo, J. D., 2011, Pág. 399).

Regla de decisión

Menor a 3 → Fisher.

Mayor a 3 pero menor a 5 → Yates.

Mayor a 5 → Pearson.

2.30 Fisher

- $p = ((a + b)! * (c + d)! * (a + c)! * (b + d)!) / (n! * a! * b! * c! * d!)$

2.31 Razón de Verosimilitud de χ^2

1. Si la proporción o frecuencia observada f_0 es en realidad igual a la proporción esperada f_e según la hipótesis nula (H_0), las 2 probabilidades calculadas serán iguales y su razón será igual a 1.
2. Cuanto mayor sea la diferencia entre f_0 y f_e , mayor será la razón entre ambas. La razón de estas 2 probabilidades o verosimilitudes pueden utilizarse como estadístico para medir el grado de ajuste entre las frecuencias observadas y las esperadas (Portillo, J. D., 2011, Pág. 409). Esta prueba se calcula de la Ecuación 2.19.

2.32 Test de McNemar para datos apareados

Prueba de la significación de los cambios

$$\chi^2_{McNemar} = [(B - C) - 1]^2 / (B + C)$$

2.33 Q de Cochran

Para tres o más muestras dependientes

1. La prueba Q de Cochran es una extensión de la prueba de McNemar, que se utiliza en los modelos experimentales con 3 o más muestras dependientes o relacionadas entre sí, es decir, esta población sirve como su propio control, en el que existe un periodo previo y otro ulterior; además el tipo de escala debe ser nominal (Portillo, J. D., 2011, Pág. 446).
2. El valor calculado en la prueba Q de Cochran se distribuye igual que la ji cuadrada, por lo cual el símbolo utilizado será χ^2_Q , ver Ecuación 2.20.

Donde:

χ^2_Q = Estadístico ji cuadrado de la prueba de Q de Cochran.

K = Número de tratamientos.

Gn = número total de respuestas de cambio de cada tratamiento o columna.

Lc = número total de respuestas de cambio por individuo de la muestra o hileras.

2.34 Test de χ^2 de independencia de Pearson

2.34.1 Contrastes a posteriori

1. Test de la Q de Yule: No es un test de contraste de hipótesis (no calcula p) Tablas = 2 x 2.
2. Coeficiente PHI ϕ : Nominales, Tablas de 2 x 2.
3. V de Cramer: Nominales, Tablas de 2 x 2.
4. Coeficientes Lambda λ : Nominales, Tablas de 2 x 2.
5. Coeficiente de Incertidumbre: Nominales, Tablas de 2 x 2.
6. Coeficiente de Tau-b de Kendall: Ordinales, Tablas de 2 x 2.
7. Coeficiente Gamma γ : Ordinales, Tablas de 2 x 2.
8. D de Somers: Ordinales, Tablas de 2 x 2 (Portillo, J. D., 2011, Pág. 454).

2.35 Test de la Q de Yule

2.35.1 Coeficiente Gamma

1. Interpretación:

- Q = 0 → independencia.
- Q > 0 → asociación positiva.
- Q < 0 → asociación negativa.

2. Intervalo de Q: (-1, +1)

3. Información:

- Medida de la intensidad de la asociación.
- Sentido de la misma.

2.32.2 Características de la Q de Yule

1. Medida de asociación.
 - Valores mayores de 0 (asociación positiva).
 - Valores menores de 0 (asociación negativa).
 - Valor = 0 o próximos a (es probable que no exista asociación).
2. Válida para cualquier tipo de Estudios (diseño).
3. Alcanza valores extremos (+1, -1) bajo condiciones de asociación perfectas.
4. Es independiente de los cambios de ESCALA de filas y columnas.
5. Si se intercambian entre sí las filas o las columnas se mantiene la magnitud de la INTENSIDAD de la asociación, sólo cambia la dirección de la misma, es decir el signo de la Q.
6. Es más válida con un test previo de contraste de hipótesis que indique relación entre ambas variables (Portillo, J. D., 2011, Pág. 457).

2.36 Coeficiente PHI Φ

1. Condiciones: Tablas tetracóricas (2 x 2).
2. Interpretación:
 - $\Phi = 0 \rightarrow$ independencia.
 - $\Phi > 0 \rightarrow$ asociación positiva.
 - $\Phi < 0 \rightarrow$ asociación negativa.
3. Intervalo de Φ : entre -1 y +1.
4. Información: Es un test de contraste de hipótesis que ofrece una medida de la intensidad de la asociación. Cuando se acercan a cero, indican independencia o asociación muy débil entre las variables. Cuando se acercan a uno, indican una relación más fuerte o intensa entre las variables asociadas (Portillo, J. D., 2011, Pág. 468).

2.37 Nominal x Nominal (Tablas 2 x 2)

1. V de Cramer.
2. Coeficiente de contingencia.
3. Lambda λ .
4. Tau de Goodman y Kruskal.
5. Coeficiente de incertidumbre.

V de Cramer

1. Definición: V de Cramer es una manera de calcular la correlación en los cuadros que tienen más de 2x2 filas y columnas. Se utiliza como posterior a la prueba para determinar la intensidad de la asociación después de que el test de chi-cuadrado ha sido significativa. Chi-cuadrado, dice que existe una relación significativa entre las variables, pero no dice nada del grado de significancia ni de su importancia. V de Cramer es un post-test para dar esta información adicional (Cramer, H., & Cansado, E., 1953).
2. Condiciones: Tablas de más de 2 x 2 cuando la tabla es de 2 x 2 (tetracórica), utilice el coeficiente PHI.
3. Interpretación:
 - $V = 0 \rightarrow$ independencia.
 - $V > 0 \rightarrow$ asociación positiva.
4. Intervalo de V: entre 0 y 1.
5. Información: Es una medida de la intensidad de la asociación. Cuando se acercan a cero, indican independencia o asociación muy débil entre las variables. Cuando se acercan a uno, indican una relación más fuerte o intensa entre las variables estudiadas.

2.38 Limitaciones del test χ^2 de Pearson como medida de asociación

1. Una vez que tenemos hallado el valor del χ^2 podemos preguntarnos si existe o no asociación.
2. No se puede emplear el valor de χ^2 como una medida de asociación ya que este aumenta con el tamaño muestral. El valor absoluto de χ^2 no permite tener una idea clara de la fuerza de la asociación ya que sus valores no están acotados en un intervalo definido.
3. Un coeficiente que sí tiene la propiedad de que sus valores varíen dentro de un intervalo es el Coeficiente de Contingencia.

2.39 Coeficiente de Contingencia (C)

1. El valor de este coeficiente nunca será inferior a 0 ni superior a 1. El 0 correspondería a ausencia de asociación, con $\chi^2 = 0$. No en todos los casos la dependencia extrema necesariamente debe dar 1, dependerá del número de categoría k.
2. El coeficiente de contingencia se utiliza para saber la asociación de variables cualitativas Nominales, que tienen dos o más categorías. Este coeficiente requiere del cálculo previo del estadístico χ^2 de Pearson, ver Ecuación 2.21 y Ecuación 2.22.

Donde:

C = coeficiente de contingencia.

χ^2 = valor de chi-cuadrado de Pearson calculado previamente.

N = tamaño o número de la muestra.

Ejemplo:

C = 0 → Significa que NO hay asociación

C > 0,30 → Indica una buena asociación entre las variables.

C_{máx} No alcanza el valor 1 aunque las dos variables están totalmente relacionadas.

3. El coeficiente de contingencia se utiliza para saber la asociación de variables cualitativas Nominales, que tienen dos o más categorías. Este coeficiente requiere el cálculo previo del estadístico χ^2 de Pearson.

$$\circ C_{\text{ajustado}} = C / C_{\text{máx}} = \text{Ecuación 2.21} / \text{Ecuación 2.22}$$

2.40 Tau de Goodman y Kruskal

Al igual que el coeficiente de incertidumbre y lambda, el coeficiente Tau (de Goodman y Kruskal) es una medida de asociación basada en la reducción proporcional del error. Es una medida que expresa el grado de incertidumbre que conseguimos reducir cuando utilizamos una variable para efectuar pronósticos sobre otra (Portillo, J. D., 2011, Pág. 485).

2.41 Coeficiente Lambda λ

Al igual que el coeficiente de incertidumbre, lambda es una medida que expresa el grado de incertidumbre que conseguimos reducir cuando utilizamos una variable para efectuar pronósticos sobre otra. Lambda tiene tres versiones dos asimétricas (cuando conocemos la variable predictora y el resultado) y una simétrica cuando no sabemos cuál es la variable predictora y cuál es la variable resultado (Portillo, J. D., 2011, Pág. 494).

2.42 Características de la Odd Ratio (OR)

1. Medida de asociación.
 - Valores mayores de 1 (asociación positiva) (factor de riesgo).
 - Valores menores de 1 (asociación negativa) (factor de protección).
 - Valor = 1 (no existe asociación).
2. Estudios casos-controles.
3. OR:

-Es la razón entre la probabilidad de que un suceso ocurra y la probabilidad de que no ocurra. Cociente entre la tasa de exposición observada en el grupo de casos (A/C) y la tasa de exposición observada en el grupo control (B/D).

-Refleja la probabilidad de que los casos (cáncer de pulmón) están expuestos al factor de estudio (tabaco) comparada con la probabilidad de que lo estén los controles (Portillo, J. D., 2011, Pág. 505).

4. Riesgo Relativo (RR) = Tasa de Incidencia en expuestos / Tas de Incidencia de NO expuestos

2.43 Tau-b de Kendall

1. Condiciones de aplicación:

-Tablas cuadrada (3 x 3), (4 x 4), etc.

2. Interpretación:

-Tau = 0 independencia.

-Tau > 0 asociación positiva.

-Tau < 0 asociación negativa.

3. Intervalo de Tau: (-1, +1).

4. Información:

-Medida de la intensidad de la asociación de variables ordinales.

-Sentido de la misma (direccionalidad).

5. Condiciones de aplicación: número de filas distinto al número de columnas (Tablas NO cuadradas) (Portillo, J. D., 2011, Pág. 628).

○ Tau-b de Kendall → Ordinal x Ordinal → K = F

○ Tau-c de Kendall → Ordinal x Ordinal → K ≠ F

2.44 Coeficiente Gamma γ

1. Test de contrastes de hipótesis.

2. Test de asociación para variables ordinales.

3. Interpretación:

$\gamma = 0$ independencia (Variables no asociadas).

$\gamma > 0$ asociación positiva (A mayor nivel de instrucción Mayor nivel económico).

$\gamma < 0$ asociación negativa (A mayor nivel de instrucción Menor nivel económico).

4. Intervalo de γ : (-1, +1)

5. Información:

-Medida de la intensidad de la asociación de variables ordinales.

-Sentido de la misma (direccionalidad) (Portillo, J. D., 2011, Pág. 632).

2.45 D de Somers

1. Test de contraste de hipótesis.

2. Test de asociación para variables ordinales.

3. Interpretación:

D = 0 independencia (Variables NO asociadas).

D > 0 asociación positiva (A mayor nivel de instrucción Mayor nivel económico).

D < 0 asociación negativa (A mayor nivel de instrucción Menor nivel económico).

4. Intervalo de D de Somers: (-1, +1).

5. Información:

Medida de la intensidad de la asociación de variables ordinales.

Sentido de la misma (direccionalidad).

2.46 Cuantitativas versus cuantitativas

2.46.1 Regresión lineal simple

1. Describe como es la relación entre la variable predictora (X) y la variable resultado (Y).
2. Busca la línea que mejor se ajusta a los puntos (x, y), y por tanto busca la mejor predicción de Y a partir de X.

2.46.2 Correlación

Mide el grado o fuerza de asociación entre dos variables cuantitativas (X, Y).

1. Correlación paramétrica
 - Coeficiente de correlación de Pearson (r).
 - Coeficiente de determinación r^2 .
2. Correlación NO paramétrica
 - Coeficiente de correlación de Spearman (ρ).
3. Concordancia:
 - Coeficiente de correlación intraclase.

2.46.3 Comparación de métodos analíticos

1. Regresión paramétrica
 - Regresión lineal simple
 - Coeficiente de regresión β
 - Coeficiente de correlación intraclase
2. Regresiones NO paramétricas
 - Regresión de Theil
 - Regresión de Passing – Bablok
 - Regresión de Bland – Altman
 - Regresión de Tholen -Ludbrook

2.46.4 Relación directa e inversa

1. Incorrelación: Para valores de X por encima de la media tenemos valores de Y por encima y por debajo en proporciones similares. Incorrelación.
2. Fuerte relación directa: Para los valores de X mayores que la media le corresponden valores de Y mayores también. Para los valores de X menores que la media le corresponden valores de Y menores también. Esto se llama relación directa.
3. Cierta relación inversa: Para los valores de X mayores que la media le corresponden valores de Y menores. Esto es relación inversa o decreciente (Portillo, J. D., 2011, Pág. 657).

2.47 Covarianza de dos variables X e Y

1. La covarianza entre dos variables, S_{xy} , nos indica si la posible relación entre dos variables es directa o inversa.

Directa: $S_{xy} > 0$

Inversa: $S_{xy} < 0$

Incorreladas: $S_{xy} = 0$

2. El signo de la covarianza nos dice si el aspecto de la nube de puntos es creciente o no, pero no nos dice nada sobre el grado de relación entre las variables, ver Ecuación 2.22.

Los valores de la covarianza dependiente de la variación al unísono de x e y.

Covarianza Positiva > 0

Covariana Negativa < 0

Covarianza Nula $= 0$

3. El valor de la Covarianza es cero cuando no hay variación al unísono de x e y.

2.48 Covarianza y coeficiente de correlación de Pearson (r)

2.48.1 Coeficiente de correlación lineal de Pearson

1. El coeficiente de correlación lineal de Pearson de dos variables, r, nos indica si los puntos tienen una tendencia a disponerse alineadamente, ver Ecuación 2.23.

2. El coeficiente tiene el mismo signo que la covarianza S_{xy} por tanto su signo nos informa del tipo de relación, directa o inversa; r es útil para determinar si hay relación lineal entre dos variables, pero no servirá para otro tipo de relaciones (cuadráticas, logarítmica, etc.).

3. El coeficiente de correlación no se debe extapar más allá del rango de valores observado de las variables en estudio ya que la relación existente entre X y Y puede cambiar fuera de dicho rango.

4. La correlación no implica causalidad. La causalidad es un juicio de valor que requiere más información que un simple valor cuantitativo de un coeficiente de correlación (Pita, F. S., 1996, Págs. 59-260).

5. No debe utilizarse para comparar dos métodos que intentan medir el mismo evento (variable), como por ejemplo dos instrumentos que miden la tensión arterial. El coeficiente r mide el grado de asociación entre dos cantidades pero no mira el nivel de acuerdo o concordancia. Si los instrumentos de medida miden sistemáticamente cantidades diferentes uno del otro, la correlación puede ser $r \approx 1$ y su concordancia ser nula (Bland, J. M., & Altman, D., 1986, Págs. 307-310).

2.49 Propiedades de r

1. Es adimensional.

2. Sólo toma valores en $[-1, 1]$.

3. Las variables son incorreladas $\leftrightarrow r = 0$.

4. Relación lineal perfecta entre dos variables $\leftrightarrow r = +1$ ó $r = -1$.

-Excluimos los casos de puntos alineados horizontal o verticalmente.

5. Cuanto más cerca esté r de +1 ó -1 mejor será el grado de relación lineal (Horbar, J. D., Onstad, L. Y. N. N., & Wright, 1993, Págs. 12-18).

-Siempre que no existan observaciones anómalas.

$r > 0 \rightarrow$ relación directa perfecta, β constante (+).

$r = 0 \rightarrow$ variables incorreladas.

$r < 0 \rightarrow$ relación inversa perfecta, β constante (-).

2.50 Coeficiente de determinación r^2

-Indica el porcentaje de la variabilidad total de la variable dependiente o resultado (Y) que puede ser explicada por la variable independiente o predictora (X).

2.51 Otros coeficientes de correlación

1. Cuando las variables en vez de ser numéricas son ordinales, ver Ecuación 2.25.

2. Disponemos para estos casos de tres test estadísticos:

- ρ ("rho") de Spearman.

-T ("tau") de Kendall.

-Coeficiente de asociación ordinal de Goodman-Kruskal (coeficiente gamma).

3. Son estadísticos análogos al coeficiente de correlación lineal de Pearson r, que aparecen en estudios donde las variables cuantitativas no son continuas y/o no siguen una distribución normal, ver Tabla 10.

4. El coeficiente de correlación de Spearman es recomendable utilizarlo cuando los datos presentan valores externos ya que dichos valores afectan mucho el coeficiente de correlación de Pearson, o ante distribuciones NO normales (Portillo, J. D., 2011, Pág. 686).

2.52 Cualitativa dicotómica versus cuantitativas

Grupos Paramétricos y No paramétricos, ver Figura 4.

2.53 Test de T de Student para dos muestras independientes

2.53.1 Condiciones de aplicación (1)

- n^1 y $n^2 >= 30$.

-Normalidad (Requisitos necesarios) (Scaillet, O., 2005, Págs. 415-427).

-Comprobar que el máximo y el mínimo quedan dentro del intervalo definido por 3 DE por encima y por debajo de la media ($\text{media} \pm 3 \text{ DE}$) (Royston, P., 1992, Págs. 117-119).

-Que la asimetría (en valor absoluto) sea menor que dos veces el error estándar [asimetría] $< 2 \text{ EE de asimetría}$.

-Que la curtois (en valor absoluto) sea menor que 2 veces su EE [Curtois] $< 2 \text{ EE de curtois}$.

2.53.2 Condiciones de aplicación (2)

Homogeneidad de varianzas (homoscedasticidad):

-Prueba F de Snedecor $\rightarrow F = \text{Varianza Mayor} / \text{Varianza Menor}$.

-Test de Bartlett.

-Test de Levene.

H_0 = Las varianzas de ambos grupos son iguales \rightarrow Test T de Student.

H_a = Existen diferencias entre las varianzas \rightarrow Test de Welch.

$F_{\text{calculada}} > F_{\text{tabulada}} \text{ p} > 0,05$ (aceptamos la H_0).

$F_{\text{calculada}} < F_{\text{tabulada}} \text{ p} < 0,05$ (aceptamos la hipótesis nula).

2.54 Test de U de Mann-Whitney

2.54.1 Condiciones de aplicación

-Si alguna de las dos muestras contiene menos de 30 de observaciones y No se puede asumir la normalidad.

-Variable medida en escala Ordinal.

-Si la muestra es de muy pequeño tamaño (< 10 observaciones en alguno de los dos grupos).

2.54.2 Inconvenientes

-Se pierde potencia para detectar diferencias significativas (para una comparación que resulte significativa con 95 individuos en la t de Student, se necesitaría 100 individuos para que la misma comparación resulte significativa con el test de Mann-Whitney).

-No permite construir intervalos de confianza.

-Los valores de estadístico U y W calculados por el test se comparan con los tabulados para un contraste bilateral con un error alfa 0,05.

2.54.3 Pasos

Paso I

-Aplicación de la prueba estadística. De acuerdo con los pasos, las observaciones se deben ordenar en rangos del menor al mayor.

-Rangos de lectura en tablas.

Paso II

-Calculamos las U de Mann-Whitney (U_1 y U_2) con la Ecuación 2.26 y Ecuación 2.27.

Paso III

Teniendo en cuenta los pasos del test, hay que saber que cuando el tamaño muestral total es mayor de 25 (como ocurre en este caso), el estadístico del test de Mann-Whitney se distribuye normalmente, por tanto se calcula el estadístico Z, que se compara con el tabulado para conocer la probabilidad (p)(Portillo, J. D., 2011, Pág. 801). Esto se calcula con la Ecuación 2.28.

Paso IIIa

$$U = (n_1 * n_2) / 2$$

Donde:

U = cualquier valor de U calculado (ya sea U_1 o U_2).

U = valor promedio de U.

Z = valor estadístico de la curva normal.

σU = desviación estándar.

Paso IIIb

-La desviación estándar de U (σU) se determina de la forma en la Ecuación 2.29.

-Donde:

σU = desviación estándar de U.

n_1 y n_2 = tamaño de la muestra de los grupos 1 y 2.

N = tamaño total de la muestra (la suma de n_1 y n_2).

L_i = sumatoria de las ligas o empates.

Paso IIIc

-Una vez obtenida la sumatoria de L_i , se determina la desviación estándar de U (σU) mediante la expresión en la Ecuación 2.30.

Paso IIId

-Una vez calculados todos los parámetros necesarios, se obtiene el valor Z, conforme a la fórmula de la Ecuación 2.28.

Paso IV

-Obtener el valor de la p(Portillo, J. D., 2011, Pág. 807).

2.55 Test de Wilcoxon

2.55.1 Prueba No paramétrica o de libre distribución

-No se basa en la media, desviación estándar, etc.

2.55.2 Condiciones de aplicación

-Diseño emparejado o de medidas repetidas.

-Datos o compara son:

-Ordinales o, Cuantitativos con bajo tamaño muestral ($n < 30$) y No siguen una distribución normal, ver Figura 5.

2.56 Análisis de Regresión Logística Simple

-Variable Resultado: Dicotómica.

-Variable Predictora:

-Cuantitativa.

-Categorica.

2.57 Modelo de Análisis de Regresión Simple

Se describen en la Ecuación 2.31 y Ecuación 2.32 $\rightarrow OR = e^{\beta_1}$.

2.58 Cualitativa policotómica versus cuantitativa

Anova para muestras independientes y Análisis de la Varianza para muestras dependientes, ver Figura 6.

2.59 Análisis de la varianza para un factor (ANOVA)

2.59.1 Condiciones de aplicación

1. Tipos de variables:

-Variables resultado: Cuantitativa numérica.

-Variable predictora: Cualitativa policotómicas.

2. Normalidad: Variable resultado siga una distribución normal o de Gauss.

3. Homogeneidad de las varianzas (homocedasticidad): Las varianzas de cada grupo deben ser iguales.

-Prueba F de Snedecor $\rightarrow F = \text{Varianza Mayor} / \text{Varianza Menor}$.

-Test de Bartlett.

-Test de Levene.

- H_0 = Las varianzas de ambos grupos son iguales \rightarrow ANOVA.

- H_a = Existen diferencias entre las varianzas \rightarrow Kruskal-Wallis.

$F_{\text{calculada}} < F_{\text{tabulada}}$ $p > 0,05$ (aceptamos la H_0).

$F_{\text{calculada}} > F_{\text{tabulada}}$ $p < 0,05$ (rechazamos la H_0).

2.59.2 Análisis de la varianza para un factor

○ Hipótesis nula : $\mu_1 = \mu_2 = \mu_3 = \dots = \mu_n$

○ Hipótesis alternativa : al menos una de las medias μ_x es distinta.

○ $F = \text{Varianza entre grupos ("Efecto")} / \text{Varianza residual ("Error")}$.

2.59.3 Test ANOVA: Observaciones

1. Los resultados del contraste estadístico F en la prueba ANOVA son sustancialmente válidos aunque los datos no sean normales, en ese sentido se dice que es una técnica "Robusta" frente a desviaciones de la normalidad.

2. El efecto de desigualdad de las varianzas en los grupos sobre el contraste F y los contrastes de medidas dependen de que el número de observaciones en cada grupo sea igual o muy distinto. Si todos los grupos tienen el mismo número de observaciones, el contraste F es igualmente exacto aunque las varianzas sean distintas. Es decir, podemos despreocuparnos de las varianzas a efecto de contraste de medias, siempre que haya aproximadamente el mismo número de observaciones por grupo, en caso contrario, diferencias entre las varianzas pueden ser graves.

2.59.4 Reflexiones: existen varios incumplimientos para el ANOVA

1. Ausencia de normalidad.

2. Ausencia de homocedasticidad: el test de Levene para valorar la homogeneidad de las varianzas muestra que existen diferencias entre ellas.

3. Y existencia de valores anómalos (outliers) que impiden la utilización de métodos paramétricos (ANOVA clásicos).

Se tiene que recurrir a los clásicos no paramétricos de Kruskal-Wallis o, mejor, a métodos robustos como la generalización robusta del test Welch, que utiliza medias α -recortadas muestrales y es capaz de detectar diferencias significativas entre los grupos.

2.59.5 Análisis de la Varianza (ANOVA) para muestras dependiente so apareadas (medidas repetidas)

1. Análisis de la varianza (ANOVA) se utiliza para la obtención del coeficiente de correlación intraclase (CCI).
2. En los estudios que evalúan la concordancia, cada sujeto es examinado por más de un observador, siendo adecuado realizar un análisis de la varianza (ANOVA) para medidas repetidas (generalización de la prueba de la t de Student para datos apareados).
3. Este tipo de diseño permite aislar la variabilidad entre pacientes y concentrarnos en la variabilidad debida al factor de interés; el análisis de la varianza simple, la variabilidad entre pacientes está incluida en la variabilidad residual, (Portillo, J. D., 2011, Pág. 951).

2.60 Pruebas Post Hoc Paramétrico

Contrastes a posteriori : se comapan los grupos de 2 en 2.

- Procedimiento de Scheffe : se comapan los grupos de 2 en 2.
- Procedimiento de Bonferroni : se comapan los grupos de 2 en 2.
- Método de Dunnet : se comapan los grupos de 2 en 2.
- Método de Tukey : se comapan los grupos de 2 en 2.
- Método de Ducan : se comapan los grupos de 2 en 2.
- Método de Student- Mewman-Keuls : Se obtienen intervalos de medias.

2.60.1 Pruebas Post Hoc No Paramétrico

Contrastes a posteriori

- Procedimiento de Bonferroni.
- Procedimiento de Sidák.
- Método de Holm.
- Método de Holland-Copenhaver.
- Método de Hochberg.
- Método de Hommel.
- Método de Rom.

2.61 La prueba de Kruskal-Wallis (comparar más de dos grupos)

1. Para la comparación de una misma variable cuantitativa en más de tres grupos. Es la prueba NO paramétrica más potente par acomparar > 2 variables continuas independientes. Es el “similar” al test de ANOVA para un factor.
2. La prueba de kruskal-Wallis, es una alternativa a la prueba F de ANOVA (análisis de varianza) para diseños de clasificación simple donde No se cumple la normalidad d ella variable cuantitativa.
3. En este caso comparan varios grupos de la variable predictora (cualitativa policotómica) pero usando la mediana de cada uno de ellos, en lugar d ellas medias, donde, n es el total de datos, ver Ecuación 2.33.
 - H_0 = La mediana de las k poblaciones consideradas son iguales y
 - H_a = Al menos una de las poblaciones tiene mediana distintas a las otras.

Ventajas

- Permite datos ordinarles y cuantitativos discretos.
- No es necesaria la normalidad (test no paramétrico).
- El tamaño muestral puede ser pequeño.

Inconvenientes

- Menor potencia estadística.
- Menor sensibilidad para detectar diferencias entre grupos.

-No permite construir intervalos de confianza.

2.61.1 Prueba de la Mediana

1. Su objetivo es comparar las medianas de dos muestras y determinar si pertenecen a la misma población o no.
2. Se calcula la mediana de todos los datos conjuntamente. Después, se divide cada muestra en dos subgrupos: uno para aquellos datos que se sitúen por encima de la mediana y otro para los que se sitúen por debajo.
3. La prueba de chi-cuadrado determinará si las frecuencias observadas en cada grupo difieren de las esperadas con respecto a una distribución de frecuencias que combine ambas muestras.
4. Esta prueba está especialmente indicada cuando los datos sean extremos o estén sesgados.

2.62 Test de Jonkheere-Terpstra

1. Cuando no se cumplen las condiciones para el ANOVA, hay tres tests disponibles: H de Kruskal-Wallis, mediana y Jonkheere-Terpstra.
2. El test H de Kruskal-Wallis, una extensión del test U de Mann-Whitney es el análogo no-paramétrico del ANOVA.
3. El test de la mediana, que es una prueba más general pero no tan potente, detecta diferencias en la situación y forma de las distribuciones. Ambas pruebas asumen que no hay un ordenamiento a priori de las k poblaciones de las que se han extraído las muestras.
4. Cuando hay una ordenación natural a priori (ascendente o descendente) de las k poblaciones, es preferible usar el test de Jonckheere-Terpstra.

2.63 Comparación de > 2 grupos apareados

Test de Kolmogorov-Smirnov \rightarrow ($p > 0,05$; Normalidad) Prueba de la esfericidad de Mauchly

1. $\rightarrow p > 0,05$ Esfericidad \rightarrow Contrastes Univariados ($p < 0,05$) \rightarrow Post hoc: procedimiento de Bonferroni: comparación por pares.
2. $\rightarrow p < 0,05$ No Esfericidad \rightarrow Contrastes Multivariados ($p < 0,05$) \rightarrow Post hoc.

2.63.1 Comparación de > de 2 grupos apareados

Test de Kolmogorov-Smirnov \rightarrow ($p > 0,05$; Normalidad) Prueba de la esfericidad de Mauchly

1. \rightarrow Esfericidad ($p > 0,05$) Contrastes Univariados \rightarrow ($p < 0,05$) \rightarrow Post Hoc: Procedimiento de Bonferroni: Comparación de pares.
2. \rightarrow No Esfericidad ($p < 0,05$) Contrastes Multivariados \rightarrow ($p < 0,05$) \rightarrow Post Hoc: Procedimiento de Bonferroni: Comparación de pares.

2.64 Prueba de Esfericidad de Mauchly

1. Cuando el estadístico W de Mauchly nos lleve al rechazo de la esfericidad (No esfericidad), podemos basar nuestra decisión en los estadísticos F Multivariados que no les afecta la falta de esfericidad, o bien podemos utilizar el estadístico F Univariado aplicando el índice corrector Epsilon.
2. El índice corrector Epsilon expresa el grado en que la matriz de varianzas-covarianzas se aleja de la esfericidad. En condiciones de esfericidad perfecta, epsilon vale 1. En este caso epsilon es alto porque hay Esfericidad.
3. Existen 3 estimaciones del índice epsilon, Greenhouse e-Geisser es la más conservadora. El límite inferior expresa el valor que adoptaría epsilon en caso de incumplimiento extremo de la esfericidad.

4. El valor de ϵ se utiliza para corregir el valor de F Univariado. En este caso No es necesario utilizar el ϵ porque se cumple la Esfericidad ($p > 0,05$).
5. Si se cumple la Esfericidad es preferible utilizar la aproximación Univariada (versión esfericidad asumida), pues en condiciones de esfericidad el estadístico F Univariado es más potente que los estadísticos Multivariados, sobre todo con las muestras pequeñas.

2.65 Test de Friedman

1. Es la versión No paramétrica del ANOVA para muestras apareadas.
2. Utiliza la modificación estadística de Iman & Davenport (1980) que lo hace menos conservador.
3. El valor de z de cada hipótesis se puede utilizar para calcular la probabilidad de error (valor p) mediante la distribución normal, ver Ecuación 2.34.
4. ¿Se cumplen las condiciones impuestas por el test ANOVA?
SI: Utilizar ANOVA
NO: Utilizar el test de Friedman

2.66 Coeficiente de concordancia W de Kendall

1. Mide el grado de concordancia entre un grupo de elemento y un grupo de características.
2. Si la concordancia es máxima $W = 1$, si la concordancia es mínima $W = 0$, Ecuación 2.35.

Donde:

w = coeficiente de concordancia de Kendall.

S = suma de los cuadrados de las diferencias con respecto a un promedio.

N = Tamaño de la muestra en función del número de tripletes, tetrapletes, quintupletes, etc.

K = número de variables incluidas.

L_i = sumatoria de las ligas o empates entre los rangos.

3. El coeficiente W de Kendall y el estadístico χ^2 de Friedman son aplicables al mismo tipo de situaciones.
4. La hipótesis nula de que los tratamientos aplicados son iguales sean su efectividad es exactamente lo mismo que mantener la hipótesis nula de ausencia de concordancia en los resultados de los tratamientos ensayados.

2.67 Medidas de concordancia

1. Fiabilidad de los procedimientos de medida empleados.
2. Fleis: “ni el más elegante de los diseños sería capaz de paliar el daño causado por un sistema de medida poco fiable”.
3. Error de medida: Variabilidad entre observadores.
4. Estudios de fiabilidad.
 - Sesgo entre observadores.
 - Concordancia entre observadores.
5. Indicadores del grado de variabilidad.
 - Coeficiente de correlación intraclase (variables cuantitativas).
 - Índice de Kappa (variables cualitativas tipo categóricas).
6. Valoración de la fiabilidad externa de ellas publicaciones de dos revistas científicas durante el último quinquenio.
7. En 5 (4 de JAMA y uno de MC) de los 12 artículos se han llevado a cabo tratamientos estadísticos no idóneos para medir la concordancia, mediante las siguientes pruebas de

contraste de hipótesis (Basterra-Gortari, F. J., Bes-Rastrollo, M., Seguí-Gómez, M., Forga, L., Martínez, J. A., & Martínez-González, M. Á., 2007, Págs. 405-408):

- Comparación de medias con la t de Student
- Test exacto de Fisher,
- Análisis de la varianza (ANOVA)
- Test de Wilcoxon
- Coeficiente de Spearman

2.68 Análisis de concordancia: Variables Cualitativas

1. Índice kappa de Cohen (2 observadores).
2. Índice kappa de Fleiss (> 2 observadores), (Fleiss, J. L., Levin, B., & Paik, M. C., 2013).
3. Índice kappa ponderado.
4. Prueba de homogeneidad de kappas, (Portillo, J. D., 2011, Pág. 995).

2.69 Índice Kappa de Cohen

1. $k = 0$ cuando la concordancia observada p_0 es la que se espera a causa del azar p_e .
2. $k > 0$ cuando la concordancia observada p_0 es mayor de la esperada sólo por el azar p_e , ver Ecuación 2.36 y Ecuación 2.37, Tabla 12.

2.70 Índice Kappa Ponderado

1. Si las categorías de clasificación son más de dos y están definidas en una escala ordinal, como por ejemplo “sano”, “posiblemente enfermo” y “claramente enfermo”. A la hora de valorar el grado de discrepancia entre dos observadores, no es lo mismo que uno clasifique a un sujeto como “posiblemente enfermo” y otro lo declare “sano” a que uno lo clasifique como “sano” y el otro como “claramente enfermo”. La “distancia” entre ambas discrepancias no es la misma.

2. Cuando ciertos tipos de desacuerdos son más acusados que otros, al investigador le puede interesar tenerlo en cuenta en la construcción del índice kappa. Por ello, se pondera las diferentes discrepancias, usando una matriz de pesos que pueden variar según el criterio del investigador en función de lo que esté analizando, aunque siempre cumpliendo ciertas restricciones:

1. La matriz debe ser simétrica.
2. La diagonal de acuerdos se le asigna el máximo peso, que es 1 y el resto de pesos debenser inferiores, aunque siempre positivos o iguales a cero.

2.70.1 Limitaciones del estadístico Kappa

1. El valor de kappa se ve afectado por la prevalencia del rasgo estudiado. Por tanto, es necesario ser cuidadoso a la hora de generalizar los resultados de comparación de observadores en situaciones con prevalencias deferentes; esto quiere decir que kappa es un estadístico descriptivo útil, pero es inadecuado con fines que predicción o inferencia.

2. Kappa es dependiente del número de categorías. Cuantas más categorías se estén considerando, más difícil será clasificar correctamente los sujetos de observación, lo que implica valores de kappa más bajos. Por tanto, debe tenerse en cuenta el número de categorías a la hora de interpretar kappa.

3. Para datos ordinales derivados de categorizar variables continuas, kappa depende fuertemente las definiciones arbitrarias que se hacen de las categorías.

2.70.2 Pruebas de homogeneidad de las kappas

Hipótesis nula: $k_1 = k_2 = \dots = k_n$.

Hipótesis alternativa: al menos una de las kappas k_x es distinta.

2.70.3 Comparación de kappas: Distribución de χ^2 de Pearson

Hipótesis nula H_0 = Igualdad de las kappas comparadas. El grado de aprendizaje no varía entre los distintos residentes.

Hipótesis nula H_a = Desigualdad de las kappas comparadas. El grado de aprendizaje varía entre los distintos residentes. Existe al menos una kappa diferente al resto.

2.71 Medidas de concordancia

1. Fiabilidad de los procedimientos de medida empleados.
2. Fleiss: “ni el más elegante de los diseños sería capaz de paliar el daño causado por un sistema de medida poco fiable”.
3. Error de medida: Variabilidad entre observadores.
4. Estudios de fiabilidad.
 - Sesgo entre observadores.
 - Concordancia entre observadores.
5. Indicadores del grado de variabilidad.
 - Coeficiente de correlación intraclase (variables cuantitativas).
 - Índice de kappa (variables cualitativas tipo categóricas).

2.72 Evaluación de métodos diagnósticos cuantitativos

1. Variación metrológica:
 - Errores aleatorios.
 - Errores sistemáticos.
2. Comparación de métodos.

2.72.1 Evaluación metrológica

La evaluación metrológica de una prueba diagnóstica Cuantitativa persigue detectar los errores aleatorios y sistemáticos que origina, para cuantificarlos mediante cálculos estadísticos y determinar si son inferiores a los límites permisibles que no comprometen a la utilización clínica de los resultados, ver Figura 7.

Esta evaluación nos permite conocer el error que cometen en nuestras condiciones de trabajo, y si la prueba es adecuada para nuestros fines o es necesario cambiarlo por otros métodos, (Portillo, J. D., 2011, Pág. 1024).

2.72.2 Especificaciones de la calidad para la imprecisión analítica

1. Presentación óptima $\rightarrow CV_a < 0.25 CV_{BW}$
2. Presentación deseable $\rightarrow CV_a < 0.5 CV_{BW}$
3. Presentación mínima $\rightarrow CV_a < 0.75 CV_{BW}$

2.72.3 Fuentes de variación biológica

1. Variación biológica NO aleatoria:
 - A lo largo de la vida.
 - Ritmo biológico diario.
 - Ciclos mensuales.
 - Ritmos estacionales.
2. Variación biológica aleatoria:
 - Variación biológica intra-individual (CV_{BW}).
 - Variación biológica inter-individual (CV_{Bb}).

2.72.4 Cálculo de la variación total (Cv_t)

Cuando se obtiene una sola muestra y esta es analizada una sola vez, ver Figura 8 y Ecuación 2.38. Si minimizamos la variabilidad preanalítica mediante protocolización,

entrenamiento, buenas prácticas de laboratorio y seguimiento de los procedimientos operativos descritos y estandarizados, ver Ecuación 2.39 resultante.

2.72.5 Diagnósticos cuantitativos

Se utilizan para demostrar que el nuevo método (Y) presenta una buena concordancia con el método clásico de referencia (X). Demostrar que una misma medida tiene equivalencia con diferentes equipos. Detectar la presencia de SESGOS (errores sistemáticos):

1. Error (sesgo) constante: La diferencia entre ambos métodos se mantiene constante a lo largo del rango de medidas.
2. Error (sesgo) proporcional: La diferencia entre ambos métodos es lineal a lo largo de todo el rango de medición.

1. Métodos analíticos
2. Métodos de regresión
3. Métodos gráficos

2.72.6 Comparación de métodos diagnóstico cuantitativos

Si los errores sistemáticos detectados y medidos están en un límite aceptable, la evaluación se debe completar con un experimento de comparación de métodos frente a los resultados obtenidos con un método de referencia. La utilización de la estadística en los experimentos de comparación de métodos es la parte más difícil de análisis estadístico en los estudios de evaluación de métodos.

2.72.7 Análisis de concordancia: variables cuantitativas

2.72.7.1 Métodos diagnósticos

1. Coeficientes de correlación intraclass (ANOVA).
 - Modelo de efectos Aleatorios.
 - Modelo de efectos Fijos.
2. Coeficiente de correlación-concordancia de Lin.
3. Concordancia de Krippendorff.

2.72.7.2 Métodos de regresión

1. Método de regresión lineal por mínimos cuadrados.
2. Método de regresión ortogonal de Deming.
3. Modelo de Passing-Bablok.
4. Métodos de mínimos cuadrados ponderados.
5. Método de regresión de Bland y Altman.
6. ANOVA de medidas repetidas.
7. Modelos de ecuaciones estructurales.

2.72.7.3 Métodos gráficos

1. Análisis de las diferencias individuales método de Bland y Altman(que calcula un intervalo de concordancia).
2. Gráfico de distribución acumulativa empírica- “gráfico montaña” (mountainplot)-
3. Gráfico de la concordancia -supervivencia (survival-agreementplot).

2.72.8 ¿Cuál es el método estadístico más adecuado para calcular el error que comete un método diagnóstico cuantitativo?

La utilización de la estadística en los experimentos de comparación de métodos cuantitativos es la parte más difícil del análisis estadístico en los estudios de evaluación de método diagnósticos.

2.73 Regresión lineal por mínimos cuadrados

2.73.1 Condiciones de aplicación

1. Linealidad de datos.
2. Pequeños errores aleatorios.
3. Intervalo de valores normales suficiente amplio de concentraciones.
4. Ausencia de valores aberrantes (“outliers”).
5. Que la variable predictora “X” no tenga error, es decir que el método de referencia utilizado sea exacto, pero en la mayoría de ocasiones no es posible, por lo que los resultados del método de comparación (X) también presenta errores (Westgard, J. O., & Hunt, M. R., 1973, Págs. 49-57).

2.74 Coeficiente de correlación de Pearson

1. Mide la fuerza de asociación entre ambas variable (X-Y), es decir la intensidad de la relación lineal entre los dos métodos de diagnósticos cuantitativos.
2. No ofrece información sobre el grado de Concordancia, por tanto no puede detectar sesgo constante o proporcional.
3. Es posible encontrar un alto coeficiente de correlación con una baja intensidad de concordancia entre métodos.

Correlación \neq Concordancia

2.75 Coeficiente de correlación intraclass

1. Los valores del CCI pueden oscilar entre 0 y 1, de modo que la máxima concordancia posible corresponde a un valor de CCI = 1 (Bland, J. M., & Altman, D., 1986, Págs. 307-310).
2. Cuando el valor CCI = 0 se obtiene cuando la concordancia observada es igual a la que se esperaría que ocurriera sólo por azar. En este caso la variabilidad observada se explicaría por las discrepancias entre los métodos de medición o de los diferentes criterios utilizados por los observadores en el experimento (Pita Fernández, S., Pértega Díaz, S., & Rodríguez Maseda, E., 2003).

Correlación \neq Concordancia

3. La obtención del CCI que permite evaluar la concordancia general entre dos o más métodos u observaciones se basa en un modelo de análisis de la variancia (ANOVA) con medidas repetidas, pero su obtención directa no es posible con los paquetes estadísticos habituales.
4. La variabilidad total observada en el CCI puede dividirse en tres componentes:
 - La variabilidad debida a las diferencias entre los pacientes (σ^2P);
 - La variabilidad debida a las diferencias entre los observadores (médicos A y B) (σ^2O), yLa variabilidad (residual), inexplicable (aleatoria), asociada al error inherente a toda medición (σ^2R).
5. El CCI se define como la proporción de variabilidad total debida a la variabilidad de los pacientes, ver Ecuación 2.40.
6. Los valores del CCI pueden oscilar entre 0 y 1: donde el 0 indica ausencia de concordancia y el 1, la concordancia o fiabilidad absoluta de los resultados obtenidos.
7. Si bien el valor del CCI que define una fiabilidad satisfactoria es arbitraria (criterios de Fleiss, J. L., 2011):
 - CCI < 0,4 representan BAJA fiabilidad.
 - CCI entre 0,4 – 0,75 representan una fiabilidad entre Regular y Buena.
 - CCI > 0,75 representan una fiabilidad Excelente.

8. Los resultados ofrecidos por la r de Pearson para estos propósitos pueden resultar equívocos. Se recomienda el uso del CCI para cuantificar la fiabilidad de las mediciones clínicas.

2.76 Valoración del coeficiente de correlación intraclase

Valoración de la concordancia según los valores de Coeficiente de Correlación Intraclase (CCI), ver Tabla 13.

2.76.1 Limitaciones del coeficiente de correlación intraclase

A pesar de ser la medida de concordancia más adecuada para el caso de variables numéricas, el CCI presenta ciertas limitaciones:

1. A la hora de interpretar los valores del CCI.
2. Dificultad inherente a su cálculo.
3. Se trata de una prueba Paramétrica, por lo que limita su uso a:
 - variable distribuidas según una normal.
 - igualdad de varianzas.
 - independencia entre los errores de cada observador.
4. El CCI depende en gran medida de la variabilidad de los valores observados: cuanto más homogénea sea la muestra estudiada, más bajo tenderá a ser el valor del CCI.
5. Lo que más ha limitado la difusión del uso del CCI en la literatura médica es la carencia de interpretación clínica completamente objetiva (PRIETO, L., 1998, Págs. 142-145).

2.77 Modelos de efectos fijos

1. Este modelo es apropiado cuando los observadores que intervienen en el estudio son los únicos que participarán en el mismo.
2. Se utiliza cuando el objetivo es conocer la magnitud de las discrepancias entre dos médicos y determinar si el médico A puede ser sustituido por el médico B en esa tarea.
3. Solo estamos interesados en la comparación de los resultados obtenidos por estos 2 médicos y no deseamos generalizar los resultados al conjunto de médicos que miden la frecuencia cardíaca en el hospital.

2.78 T de Student para datos apareados

1. Mide la diferencia entre las medias de las mediciones realizadas por los dos métodos comparados.
2. Sólo puede detectar el sesgo o error sistemático constante.
3. No puede detectar el error sistemático Proporcional (Gardner, P. L., 1995, Págs. 283-289).

2.79 Coeficiente de concordancia de Lin

1. El coeficiente de correlación de las medidas de concordancia de LIN mide el grado de acuerdo entre dos variables cuantitativas continuas.
2. Es útil para evaluar la reproducibilidad o fiabilidad entre dos observadores o evaluadores. La Escala de McBrige para el coeficiente de concordancia de Lin se puede observar en la Tabla 14.

2.80 Método de regresión ortogonal de Deming

1. En el caso de que presenten errores aleatorios en el método de referencia (X) ha sido propuesto como más aceptable el método de regresión ortogonal establecido por Deming en 1943 (Wakkers, P. M., Hellendoorn, H. B. A., DeWeegh, G. O., & Heerspink, W., 1975, Págs.173-184.).

2. Es adecuado incluso si existen valores aberrantes (Cornbleet, P. J., & Gochman, N., 1979, Págs. 432-438).

3. Si el error que presenta esta variable es proporcional, cuando la varianza de los resultados aumenta con la concentración, ha sido considerado adecuado el cálculo de la correlación lineal mediante el cálculo de mínimos cuadrados ponderado.

4. Mientras que consideran mejor el método de Deming si los errores en “x” son significativos frente al intervalo ensayado, cuando el intervalo de concentraciones es estrecho (v.g. en Potasio) (Lasky, F. D., 1992, Págs. 1260-1267).

2.81 Modelos de regresión de Passing-Bablok

En los años 80-90, para evitar los inconvenientes del método simple de mínimos cuadrados, Passing y Bablock desarrollan el cálculo del método de regresión no paramétrica. Fue defendido como “método de elección debido a su robustez” y a permitir “obtener una estimación no sesgada de la recta de regresión” (Fuentes Arderiu, X., 1998).

1. La razón de varianzas en “Y” y “X” igual de la cuadrado de la pendiente.

2. Ausencia de error en la variable Predictora (independiente) “X”.

3. Que el error en la variable Resultado “Y” debe ser homogéneo en todo el intervalo de concentraciones (Lasky, F. D., 1992, Págs. 1260-1267).

2.82 Método de mínimos cuadrados ponderados

En 1991 Zaydman et al., en su estudio sobre modelos de regresión lineal para comparación de métodos, rechazan el método de Passing-Block porque no cumple las condiciones anteriores. Obtiene los mejores resultados con el método de mínimos cuadrados ponderados cuando los errores en la variable Predictora “X” son pequeños frente al intervalo de concentraciones como el medio (v.g. Glucosa).

2.83 Método gráfico de Bland y Altman

Bland y Altman proponen su método de evaluación de las diferencias, con el atractivo de su sencillez de cálculo y su utilización frente al uso de los métodos de regresión anteriores.

-Representan gráficamente la diferencia entre los dos métodos ($D = Y - X$) frente a la media de la suma de los dos resultados ($A = X + Y / 2$) (Altman, D. G., & Bland, J. M., 1983, Págs. 307-317).

2.83.1 Interpretación

1. En el caso de que no hay error sistemático los puntos se distribuirán de forma aleatoria a uno y otro lado de la recta correspondiente a la diferencia 0 entre medidas (Altman, D. G., & Bland, J. M., 1983, Págs. 307-317).

2. La media de las diferencias (error sistemático del segundo método respecto al primero).

3. Límites de confianza definidos mediante ± 2 DE de la media de la diferencia, y se denominan límites de concordancia (Bland, J. M., & Altman, D., 1986, Págs. 307-310).

2.83.2 Repetibilidad

1. La repetibilidad es relevante para el estudio de comparación de métodos porque la repetibilidad de los métodos de medida limita el grado de concordancia.

2. Si un método presenta una repetibilidad pobre - es decir, existe una considerable variación en las mediciones repetidas en la misma muestra - el grado de concordancia entre ambos métodos es pobre también.

3. La mejor manera de examinar la repetibilidad de un instrumento es tomar medidas repetidas en una serie de muestras.

4. Se calcula la media y la desviación estándar de las diferencias medidas. La diferencia media de ser cero, ya que se trata del mismo método. Si la diferencia de medias es

significativamente diferente de cero, no podemos utilizar los datos para evaluar la repetibilidad, ya sea porque el conocimiento de la primera medida está afectando a la segunda o el de la medida es modificar la cantidad).

2.84 Recomendaciones de Zaydman

2.84.1 Modelos de Regresión de Passing-Badblock

Debe utilizarse cuando cumple la condición:

1. La razón de varianzas en “Y” y “X” iguale el cuadro de la pendiente (Zaydman, M. A., Silva, J. R., Delaloye, K., Li, Y., Liang, H., Larsson, H. P., ...& Cui, J., 2013, Págs. 13180-13185).

$$[\beta_1]^2 = S_y^2 / S_x^2$$

2.84.2 Método de mínimos cuadrados (método regresión lineal simple)

Debe de utilizarse cuando cumple con la siguiente condición:

1. Ausencia de error en la variable Predictora (independiente) “X”.
2. Que el error en la variable Resultado “Y” debe ser homogéneo en todo el intervalo de concentraciones.

2.84.3 Método de mínimos cuadrados ponderados

Debe utilizarse cuando:

- Los errores en la variable predictora (independiente) “X” son pequeños frente al intervalo de concentraciones medido.

2.84.4 Método de regresión ortogonal de Deming

Debe utilizarse cuando:

- Cuando los errores en la variable predictora “X” son significativos frente al intervalo ensayado, cuando el intervalo de concentraciones es estrecho.

2.85 ANOVA de 2 factores

Variable Predictora 1 → Causa o Factor 1: Cualitativa ↔ Interacción ↔ Variable Predictora 2 (causa o Factor 2: Cualitativa) → Variable Resultado (Desenlace o efecto: Cuantitativa), ver Tabla 15.

2.86 Regresión Logística Multivariable

Es un modelo de regresión que permite estudiar si la variable resultado de tipo cualitativa nominal depende, o no, de otras variables predictoras. La naturaleza de las variables que pueden ser:

- Categorías Dicotómica, es decir, sólo puede admitir dos posibles resultados.
- Cuantitativa (discreta o continua).

2.86.1 Aplicación de los modelos de regresión Logística

1. En ensayos clínicos en los que se evalúa un evento o resultado en relación a uno o más factores etiológicos, predictores o asociados (RR).
2. En estudios de Cohortes en los que se evalúa un evento o resultado en relación a uno o más factores de riesgo (etiológicos, predictores o asociados) (RR).
3. En estudios de Casos y Controles en los que se evalúa diferentes hechos en casos en cada grupo, para encontrar asociaciones entre ellos y la enfermedad o casos (OR).
4. En estudios Transversales en los que se evalúan diferentes variables predictoras (factores de riesgo), para encontrar asociaciones entre ellas y la enfermedad de estudio (Variable resultado).

2.86.2 Objetivos de los modelos de regresión logística

1. Estimar la relación entre una o más variables predictoras con la variable resultado.

- Estudios etiológicos
- Confusión e interacción

2. Predecir lo mejor posible la variable Resultado o respuesta, usando la combinación más eficiente de las variables predictoras incluidas en el estudio, ver Ecuación 2.42.

2.86.3 Ventajas de los modelos de regresión logística Multivariable

1. Es un modelo útil en muchas circunstancias de investigación en medicina.

-En las que se quiere analizar o predecir una respuesta dicotómica de la variable resultado:

- enfermo / sano
- vive / muere
- presente / ausente
- positivo / negativo

-En función del perfil de variables predictoras o factores de riesgo, que puede ser

- cualitativas: género, sexo, grupo cultural, estado civil, nivel de estudios, nivel socioeconómico.
- cuantitativas: edad, número de cigarrillos, IMC, colesterol, TAS.

2. Es la mejor aproximación a la realidad biológica, ya que la influencia que una variable predictora o factor de riesgo sobre la probabilidad de que aparezca una enfermedad (variable resultado) suele asemejarse más a una curva sigmoidea (función exponencial) que a una recta. La curva sigmoidea es la representación habitual de una respuesta a diferentes dosis de un estímulo.

3. El modelo matemático permite obtener directamente la Odds Ratio (OR) o el Riesgo Relativo (RR) para cada factor de riesgo o variable Predictora incluida en el modelo. El coeficiente β_i de cada variable predictora obtenido por regresión logística es el logaritmo neperiano de la OR o de la RR del factor de riesgo.

2.86.4 Funciones de los modelos de regresión logística multivariante

1. Cuantificar: la relación existente entre las distintas variables predictoras incluidas en el modelo y la variable resultado (respuesta).

2. Controlar y Ajusta: aquellas variables predictoras que actúan como potenciales Factores Modificadores de Efecto, especialmente los Factores de Confusión.

3. Predecir la Respuesta: de la variable resultado en función de las variables predictoras o factores de riesgo, clasificado a los individuos en una de las dos posibles respuestas en función de la probabilidad que arroje su perfil de variables predictoras incluidas en el modelo de regresión logística multivariable.

2.86.5 Variables explicativas o Predictoras en el modelo de Regresión Logística

2.86.5.1 Variables Cualitativas

-El modelo asume estimaciones basadas en variables continuas.

-Las variables cualitativas deben ser modeladas en 2 categorías (0,1) o transformadas

En tantas variables dicotómicas como categorías (variables Dummy).

2.86.5.2 Variables Cuantitativas

-La OR y el RR son medidas que cuantifican el cambio en el riesgo cuando se pasa de un valor del factor a otro.

-Este aumento o disminución del riesgo es lineal y proporcional al cambio.

2.86.6 Regresión Logística

2.86.6.1 Estimación de coeficientes β_i

1. $\beta_1, \beta_2, \beta_3, \dots, \beta_n$ son los coeficientes de regresión logística del modelo.
 - β_0 es el término independiente (o constante en SPSS).
 - El resto de las β_i son los coeficientes de regresión logística para las diferentes variables predictoras incluidas en el modelo.
 - La evidencia de relación entre una variable predictora y la variable resultado es la existencia de un coeficiente de regresión logística significativamente diferente a cero $\beta_i \neq 0$.

2.86.6.2 El signo de β_i tiene un significado importante

1. Si es positivo ($\beta_i > 0$) significa que esa variable predictora aumenta la probabilidad del suceso (variable resultado o respuesta) que estamos analizando.
2. Si es negativo ($\beta_i < 0$) su cálculo se lleva a cabo mediante la Función de Máxima Verosimilitud.

2.86.7 Definición de las hipótesis nula H_0 y alternativa H_a en la Regresión Logística

1. La hipótesis conceptual es que “una o varias variables predictoras influyen significativamente en la probabilidad del suceso en estudio (que la variable predictora tome un valor u otro)”.
2. Las hipótesis operativas se reducen al contraste de ellos diferentes coeficientes β_i obtenidos en el análisis de los datos como estimadores de las variables predictoras de interés.
3. Así cuando sólo hay una variable predictora X_1 :
 - $H_0 \rightarrow \beta_1 = 0$.
 - $H_a \rightarrow \beta_1 \neq 0$.
4. Hay varios procedimientos para llevar a cabo este contraste.

2.86.8 Procedimientos para evaluar el contraste de hipótesis en la Regresión Logística

2.86.8.1 Evaluación estadística del coeficiente β_i

-El coeficiente de regresión logística β_i de la variable predictora i debe ser estadísticamente distintos de cero $\beta_i \neq 0$ para que esta tenga influencia significativa sobre la variable resultado.

- La hipótesis H_0 será que $\beta_i = 0$.
- La hipótesis H_a será que $\beta_i \neq 0$.

-Para tener en cuenta el error aleatorio ($E\beta_i$) con que el coeficiente β_i (obtenido en la regresión logística) estima al coeficiente de regresión poblacional β_n se lleva a cabo un contraste a través del estadístico t de Student con un número de grados de libertad igual al número de individuos de la muestra N menos el número de variables predictoras (n) introducidas en el modelo.

2.86.8.2 Evaluación estadística de Wald.

- El estadístico de Wald es el resultado de dividir el cuadrado del coeficiente de regresión logística β_i por el cuadrado de su error estándar $E\beta_i$.
- El estadístico de Wald (W) es también muy utilizado para evaluar la significación estadística del modelo de regresión logística multivariante.
- El estadístico W se obtiene dividiendo el cuadrado de β_i por el cuadrado del error estándar de β_i .

$$W = [\beta_i]^2 / [EE\beta_i]^2$$

- El estadístico de Wald se distribuye según una Chi Cuadrado.
- La hipótesis a contrastar son las mismas que en el caso anterior.
- La hipótesis H_0 será que $W = 0$.
- La hipótesis H_a será que $W \neq 0$.

2.86.8.3 Evaluación estadística de la razón de verosimilitud (L_0).

2.86.8.4 Evaluación estadística del cambio del estadístico $-2LL_0$.

- Este estadístico resulta ser menor de dos veces el logaritmo neperiano de la razón de verosimilitud (L_0).

2.86.8.5 Evaluación del estadístico Z^2 , (Goodness of Fit) , ver Tabla 16.

2.86.9 Interpretación del coeficiente de regresión logística β_i

-Su signo es positivo ($\beta_i > 0$), la probabilidad de que se dé o aparezca la característica de la variable predictora aumentará según aumente el valor de x_n (la variable resultado a quien afecta), ver Tabla 17.

-Si $\beta_i \neq 0$ es significativamente distinta de cero (test de contraste de hipótesis $p < 0,05$), la variable predictora i influye significativamente en la probabilidad de que se dé o aparezca la característica de la variable resultado, y el modelo logístico tiene sentido (Portillo, J. D., 2011, Pág. 1184).

2.86.10 Utilidades de la regresión lineal múltiple

1. Predicción de respuestas (variables resultado) a partir de un conjunto de variables explicativas (predictoras).
2. Identificación de variables predictoras y construcción del modelo de regresión lineal múltiple.
3. Detección de interacciones entre las distintas variables predictoras que afectan a la variable resultado.
4. Identificación (y ajuste) de variables predictoras que actúan como Factores de Confusión.

2.86.11 Requisitos para el uso de la regresión lineal múltiple

1. Linealidad.

- La variable Resultado depende linealmente de las variables predictoras.
- Si encontramos respuestas no-lineales, esas variables predictoras se pueden transformar en lineales.

2. Normalidad de las variables cuantitativas.

2.86.12 Modelo de regresión lineal múltiple

$$Y = \beta_0 + \beta_1 X_1 + \dots + \beta_n X_n$$

1. "Y" es la variable resultado o dependiente (debe ser siempre numérica).
2. Los términos X_i son las variables predictoras o independientes o explicativas, las cuales pueden ser:
 - Cuantitativas (discretas o continuas).
 - Cualitativas dicotómicas; en este caso deben transformarse en cuantitativas (0,1).
 - Cualitativas policotómicas; en este caso deben transformarse en cuantitativas (0,1), son las variables dummy.
3. Los términos β_i son los coeficientes de regresión del modelo y los calcula el programa estadístico.

2.86.13 Colinealidad en la Regresión Lineal Múltiple

1. Es uno de los problemas más importantes que se puede encontrar en un análisis de regresión Si en un modelo de Regresión Lineal Múltiple alguna variable predictora es combinación lineal de otras variables predictoras, el modelo es irresoluble.
2. La colinealidad aparece cuando alguna variable predictora es “casi” combinación lineal de otra u otras variables predictoras, es decir sus coeficientes de correlación simple o múltiple entre las variables predictoras son cercanos a 1, aunque no llegan a dicho valor.
3. Como medida de la misma hay varios estadísticos propuestos, los más sencillos son los coeficientes de determinación (r^2) de cada variable predictora con todas las demás.

2.86.13.1 Diagnóstico de colinealidad

-Factor de inflación de la varianza (FIV)

FIV > 10: Colinealidad (*)

$$FIV = 1 / (1 - R^2_j)$$

-Tolerancia (T)

T < 0,1: Colinealidad (*)

$$T = 1 / FIV = 1 - R^2_j$$

(*) Regla empírica de Kleinbaum

2.86.13.2 Diagnóstico de Colinealidad

Índice de Condición: Regla empírica Belsley):

-Índices de Condición <5 : NO existe colinealidad.

-Índices de Condición entre 5 y 10 : colinealidad débil.

-Índices de Condición entre 10 y 30 : colinealidad moderada.

-Índices de Condición >30 : colinealidad fuerte.

2.87 El test de Kaiser-Meyer-Olkin (KMO)

Mide la idoneidad de los datos para realizar un análisis factorial comparando los valores de los coeficientes de correlación observados con los coeficientes de correlación parcial (grupos comparados). Si la suma de los cuadrados de ellos coeficientes de correlación parcial entre todos los pares de variables es pequeña en comparación con la suma de los coeficientes de correlación al cuadrado, esta medida tiende a uno. Para Kaiser los resultados de modelo factorial serán excelente si el índice KMO está comprendido entre 0,9 y 1; buenos, si está comprendido entre 0,8 y 0,9; aceptables, si está entre 0,7 y 0,8; mediocres o regulares, cuando resulte entre 0,6 y 0,7; malos, si está entre 0,5 y 0,6; e inaceptable o muy malos cuando sea menor que 0,5.

$$KMO = (r_{global})^2 / ((r_{global})^2 + (r_{hombres})^2)$$

El coeficiente de correlación parcial es un indicador de la fuerza de las relaciones entre dos variables eliminando la influencia del resto. Si las variables comparten factores comunes, el coeficiente de correlación parcial entre pares de variables deberá ser bajo, puesto que se eliminan los efectos lineales de las otras variables.

Una forma de evaluar este hecho es mediante la medida de adecuación de la muestra KMO propuesta por Kaiser, Meyer y Olkin. Dicha medida viene dada por la ecuación KMO donde $r_{ij}(p)$ es el coeficiente de correlación parcial entre las variables X_i y X_j eliminando la influencia del resto de las variables.

KMO es un índice que toma valores entre 0 y 1 y que se utiliza para comparar las magnitudes de los coeficientes de correlación observados con las magnitudes de los coeficientes de correlación parcial de forma que, cuando más pequeño sea su valor, mayor es el valor de los coeficientes de correlación parciales $r_{ij}(p)$ y, por lo tanto, menos deseable

es realizar un Análisis Factorial. Kaise, Meyer y Olkin aconsejan que si $KMO \geq 0,75$ la idea de realizar un análisis factorial es buena, si $0,75 > KMO \geq 0,5$ la idea es aceptable y si $KMO < 0,5$ es inaceptable, ver Tabla 18.

2.88 Tipos de variable del estudio Horbar et al.

Variable resultado (cualitativa dicotómica) (Horbar, J. D., Onstad, L. Y. N. N., & Wright, Elizabeth, 1993, Págs. 12-18):

- Muerte = 1.
- Sobrevino = 0.

1. Variable predictora 1: Peso del RN (gramos/100).

2. Variable predictora 2: Bajo peso según la edad gestacional (BPSEG).

- Peso inferior al P10 para su edad gestacional = 1.
- Peso igual o superior al P10 para su edad gestacional = 0.

3. Variable predictora 3: Raza.

- Raza negra = 1.
- Raza blanca = 0.

4. Variable predictora 4: Sexo.

- Hombre = 1.
- Mujer = 0.

5. Variable predictora 5: Índice de APGAR.

- Apgar ≤ 3 = 1.
- Apgar > 3 = 0.

2.89 Estudio de Marshall et al. (NEOCOSUR)

-Objetivo: Desarrollar una ecuación que permita predecir el riesgo de mortalidad neonatal en RN de muy bajo peso de nacimiento basado en variable predictoras conocidas en el momento del nacimiento, antes del ingreso a la unidad de cuidado intensivo neonatal (Marshall, G., Tapia, J. L., D'Apremont, I., Grandi, C., Barros, C., Alegria, A., ... & Bancalari, A., 2005, Pág. 577).

-Métodos: Fueron incluidos en este estudio 1801 RN con peso al nacimiento entre 500 a 1500 gramos.

-Variables Predictoras: Edad materna, peso de nacimiento, edad gestacional, Apgar al 1 minuto, malformaciones congénitas mayores, sexo, gestación múltiple, uso de esteroides prenatales y peso pequeño para su edad gestacional (BPSEG) (definida como más bajo de Percentil 10 de peso para edad gestacional).

2.90 Reflexiones Finales

1. Los distintos tipos de sesgos, la influencia del azar y la presencia de variables predictoras de Confusión deben finalmente siempre, tenerse en cuenta, en todos los estudios como explicación posible de cualquier asociación estadística ya sea positiva, negativa o no existente.

2. Y como señalaba M. Susser en sus reflexiones sobre causalidad. “Cuando hay minas por todas partes uno no debe aventurarse sin un detector de minas (Análisis Multivariante)”.

Susser, M. (1991). Conceptos y estrategias en epidemiología: el pensamiento causal en las ciencias de la salud. In Conceptos y estrategias en epidemiología: el pensamiento causal en las ciencias de la salud.

Capítulo 3

3.1 Conclusiones

En el campo de las ciencias de la salud se aplican las leyes de la probabilidad y conceptos relacionados en la evaluación de pruebas de detección y criterios de diagnóstico. En nuestro caso, nos interesa tener mayor capacidad de predecir correctamente la presencia o ausencia de enfermedad a partir del conocimiento de los resultados positivos o negativos y del estado de los síntomas (presentes o ausentes).

En pruebas de detección pueden ocurrir los siguientes resultados: valores falsos positivos (una prueba da positiva cuando debería dar negativa) y falsos negativos (una prueba da negativa cuando debería dar positiva). Por lo tanto, las pruebas de detección no siempre son pruebas infalibles y se debe evaluar la utilidad de los resultados de la prueba y los síntomas del paciente para determinar si el individuo tiene o no alguna enfermedad.

La sensibilidad de una prueba o síntoma es la probabilidad de un resultado positivo de la prueba (presencia del síntoma) dada la presencia de la enfermedad.

Las muestras son el elemento primordial al momento de iniciar un método, dentro de las características a distinguir en un primer paso es: las muestras dependientes son mediciones pareadas de un conjunto de elementos. Las muestras independientes son mediciones realizadas en dos conjuntos de elementos distintos.

Cuando el investigador realiza una prueba de hipótesis utilizando dos muestras aleatorias, debe escoger el tipo de prueba dependiendo de si las muestras son dependientes o independientes. Por lo tanto, es importante saber si las muestras son dependientes o independientes: en el caso que los valores de una muestra afecten los valores de la otra muestra, entonces las muestras son *dependientes*; cuando los valores de una muestra no revelan información sobre los valores de la otra muestra, entonces las muestras son *independientes*.

Dentro de las características que nosotros como usuarios de los métodos estadísticos debemos tener presentes es que las pruebas de hipótesis de una y dos colas, no necesariamente deben tener los mismos valores de Z en ambos lados, cuando se construyen intervalos con iguales niveles de confianza. Por ejemplo, se podría tener un valor de $Z = 1.65$ para un intervalo unilateral del 95% y tener una $Z = 1.96$ para un intervalo bilateral al mismo porcentaje.

Tener el valor de σ conocido genera ventajas, para las suposiciones de los intervalos de confianza para μ ya que son las mismas que fundamentan la prueba Z de una media. Alguna de las violaciones de las suposiciones anteriores generaría un problema en el resultado de los cálculos en los intervalos de confianza.

Otra de las situaciones comúnmente encontradas es el hecho de no conocer la σ , entonces el intervalo de confianza para μ , se utiliza la prueba t de una media en lugar de la prueba Z de una media.

Para el caso donde σ no es conocida se utiliza un s (desviación estándar de la muestra) en lugar de σ y por consecuencia la distribución más pertinente es t (con $n-1$ grados de libertad) en lugar de Z .

Un método aproximado que define los intervalos de confianza establecer sus límites L y U para una población (π). Sin embargo, suelen ser L y U inexactos cuando el tamaño de la

muestra (n) es muy grande o cuando \hat{p} está muy cerca de 0 y 1. Sea \hat{p} el tamaño de la proporción de éxitos y \hat{q} los fracasos, definiendo así que $\hat{q} = 1 - \hat{p}$.

Para el método exacto, puede utilizarse la distribución Binomial cuyos cálculos no son sencillos incluye a S para definir los éxitos y n como el número de observaciones de la muestra, para este tipo de método se encuentran F_U y F_L , que marcan los límites de la distribución F .

En la situación cuando el investigador quiere demostrar que no hay ninguna diferencia entre procesos o que sus datos conservan la misma media antes y después de un tratamiento, es posible utilizar la prueba t para muestras apareadas y establecer equivalencias; la hipótesis nula plantea una diferencia entre medias que no reside en el intervalo de equivalencia y la hipótesis alternativa afirma que si hay una diferencia y que está en el intervalo para una prueba de equivalencias de dos colas basada en la prueba t .

Uno de los consejos más utilizados en un conjunto de muestras independientes es aquél que dice que no es posible utilizar el no rechazo de la hipótesis nula como evidencia de que la hipótesis nula es cierta. Si se decide ese camino se debe de conocer el error tipo II (beta), ya que no se tendrá suficiente evidente evidencia de que haya un efecto diferencial de las pruebas para llegar a una conclusión así. La prueba t de muestras independientes determina si existe diferencia entre las medias de dos poblaciones incluso podría encontrarse qué tan grande es dicha diferencia, esta diferencia puede calcularse con un intervalo de confianza entre L y U ; así, simplemente hay que calcular el valor planteado por la hipótesis nula y si está contenido en el intervalo. Si el valor nulo está entre L y U , la hipótesis nula no se rechaza, caso contrario se desecha.

Una manera común de plantear una situación en la cual una persona está expuesta, es conocida como eventos de razón de probabilidad u oportunidad relativa o razón de oportunidad definida como OR. Cuando el cociente de OR es menor a 1 se dice que son eventos son del tipo protectores ya que la enfermedad implicada tiene menos probabilidad a la exposición que la ausencia de la enfermedad.

Para un conjunto de muestras apareadas, la relación OR se obtiene de convertir los límites de los intervalos L y U en intervalos para el cálculo de la razón de probabilidad de una población. Los diseños de este tipo de sujetos expuestos y no expuestos son estudiados se denominan cohorte potencial. Se basan en comparativas con sujetos expuestos o también llamados caso control.

La normalidad es una suposición sólida que se define como las muestras que provienen de poblaciones con distribuciones normalmente distribuidas y funcionan en tamaños iguales o menores a 30 muestras. La homogeneidad de varianza comprueba que dos muestras provienen de la misma población cuando tienen la misma varianza, siempre y cuando la heterogeneidad no tenga un valor muy extremo. La independencia se basa en que la normalidad y la homogeneidad no se cumplan.

Una de las cuestiones que se presenta con regularidad al investigador es cómo establecer la equivalencia por medio de una prueba Z de muestras independientes para la diferencia de proporciones, se debe demostrar que la hipótesis nula la cual evidencia una equivalencia entre las muestras es verdadera. La prueba Z de muestras independientes para una diferencia entre proporciones intenta determinar si existe una diferencia entre dos proporciones de población. En caso de querer conocer el valor de tal diferencia se podría saber con tan solo calcular el intervalo de confianza y se expresa en un resultado dicotómico.

El método ANOVA se considera una extensión de la prueba t de muestras independientes, el cual consiste en un análisis de varianza y es recomendable en comparaciones múltiples. Cuando se trabaja en análisis de comparación múltiples se refiere a situaciones en las que se comparan múltiples grupos sobre una variable resultante, y F factor de ANOVA es imprescindible para saber si todas las medias de la población son iguales.

Para los MCP (Procedimientos de Comparación Múltiple), los análisis múltiples o de múltiples pruebas al cometer algún error, se presentan errores del tipo error familiar. Se han desarrollado muchos métodos para el control de los errores familiares. Algunas de estas pruebas están diseñadas para emplearse en contextos de investigación específicos, los más conocidos son el método de Bonferroni, la prueba de HSD de Tukey y la de Holm. La técnica de Bonferroni es aplicable a una variedad de situaciones en las cuales se tienen extremos múltiples y comparaciones múltiples. La prueba de Tukey está restringida sólo a los análisis de comparación múltiples. En 1979, Holm propuso una modificación al procedimiento de Bonferroni, él propone ser menos riguroso en la obtención de valores del α_{FWE} cercanos a 0.01 y no estrictamente 0.01 como plantea Bonferroni.

En los MCP, a veces sucede que una prueba como ANOVA de un factor o una prueba cuadrada como de 2 por k demuestra que no todos los valores paramétricos son iguales, sin que todas las comparaciones múltiples subsiguientes son no significativas. Entre otras razones, esto puede ser el resultado de una falta de potencia al nivel de comparación individual. También es imprescindible tener el conocimiento que el uso de MCP es un tema controversial entre los especialistas, algunos creen que estos métodos no deberían utilizarse. Dentro de las sugerencias al utilizar los MCP se encuentra no rechazar una o más hipótesis falsas en el conjunto de pruebas, si es de interés primordial entonces, probablemente debería de hacer los ajustes necesarios para controlar el α_{FWE} . Pero si no es primordial el rechazar esas hipótesis nulas entonces no se debería hacer ajustes. También se puede escoger una postura comprometida si decide controlar el α_{FWE} a 0.01 en vez de la forma tradicional a 0.05, ya que este eleva la probabilidad de un error familiar. Es necesario saber que se puede controlar el error familiar, pero no eliminar, siempre estará presente, ya sea por los propios cálculos y/o la precisión del sistema de cómputo.

En la bibliografía se encuentran documentos que exponen los métodos clásicos de control de error familiar y donde se discuten dichas técnicas y autores como Kirk y Hochberg y Tamhane proponen sus ajustes; queda por implícito en dichas prácticas que las técnicas dependen de sus datos y aplicaciones.

Para la estimación de relaciones se estudiaron los métodos de ANOVA y las pruebas de Chi-cuadrada de 2 por k. La prueba de Chi-cuadrada tiene una peculiaridad, puede ser utilizada para demostrar independencia; debido a que la prueba es una extensión directa de la prueba Chi-cuadrada de 2 por k.

También el coeficiente de correlación de Pearson del producto-momento (P-M) está diseñado para estimar una forma específica de relación entre dos variables continuas; en particular, el grado en el cual están relacionadas linealmente. P-M toma valores entre -1 y 1. La base de P-M son los conceptos de naturaleza y fortaleza de la relación entre dos variables continuas representados por la variable r y n es el número de observaciones apareadas, el cociente será la covarianza entre las dos variables (x y y). La covarianza es la medida de relación lineal, y depende de la escala sobre la cual estén medidas x y y. Es importante hacer notar que los valores de la covarianza no necesariamente están definida entre los valores -1 y 1 como el P-M.

Dos variables pueden estar correlacionadas cuando no existe una relación causal, es decir, el investigador debe ser muy cauteloso al afirmar la existencia de una relación causal entre x y y .

Es común que el investigador desee conocer si la correlación de una población es diferente de cero, para esa situación se puede iniciar con la afirmación de que hipótesis nula no es cero. Esto implicaría que la distribución muestral de r es simétrica cuando $\rho = 0$, así que la prueba t quizá produzca resultados engañosos debido al fracaso de la suposición de normalidad. Como resultado, los cálculos generalmente no serán apropiados para la prueba; entonces es momento de utilizar la prueba de Fisher. Fisher propuso un método de pruebas que supera en gran medida este problema. Este algoritmo emplea el logaritmo natural y sigue una distribución normal, así que puede realizarse mediante referencia a la curva normal.

Los algoritmos de regresión lineal resaltan por ser útiles para predecir el valor de una variable (y) a partir de otra (x). Estos modelos se manejan para calcular la proporción de variación de una variable explicada por alguna otra; comúnmente, se recaba tanto x como y para cada sujeto y después se utilizan para calcular el modelo. Incluso se recurre a él en situaciones en las que no hay ningún valor disponible de y .

La distribución muestral de muchas pruebas de permutaciones se basan en permutaciones de combinaciones de enteros y otros números. Estas pruebas por lo general se caracterizan por ser no paramétricas. Es decir, la distribución no depende de las especificaciones de los parámetros de la población. Recordando, la distribución muestral de la prueba Z de una medida requiere que se especifique σ (así como otros parámetros); en la prueba t de muestras independientes es necesario que las poblaciones estén distribuidas normalmente, lo que a su vez implica valores de ciertos parámetros como sesgo y curtosis. Como resultado de este requerimiento, estas pruebas se caracterizan por ser paramétricas. Al contrario, la prueba de Wilcoxon para muestras independientes no requiere que se especifique ningún parámetro de población, por lo que el sesgo de la población puede ser cero o cualquier otro parámetro de valor limitado, por consecuencia, esta es una prueba no paramétrica y distribución libre. Sin embargo, para construir los intervalos de confianza mediante métodos de permutación no es una tarea fácil; ya que, que es un problema de complejidad en los cálculos; Hay varios métodos de permutación basados en el principio de correlación de datos independientes tales como la prueba de correlación de Pitman, prueba de Hotelling y Pabst para correlación de rangos.

La prueba de Kruskal-Wallis es similar a la prueba F de permutación para un ANOVA de un factor, excepto que las puntuaciones se convierten en rangos antes de calcular la F obtenida y generar la distribución de permutación. Por ejemplo, para realizar la prueba F de permutación para una ANOVA de un factor con tres observaciones en cada uno de los 4 grupos, previamente encontramos que tendría que formar 369,600 disposiciones de datos calculando un estadístico F para cada una. El número de cálculos aumenta con rapidez a medida que aumenta el tamaño de la muestra y el número de grupos. Esta enorme tarea, y otras consideraciones animaron a W. H. Kruskal y W. A. Wallis a idear una prueba de permutación de muestras independientes múltiples basadas en rangos. Una ventaja de una estrategia así es que una vez que se genera la distribución de permutación para el estadístico de prueba para los tamaños de muestras dados se puede usar la distribución para cualquier muestra de los mismos tamaños sin importar los valores de los datos originales. Esto se deduce del hecho de que el dato original se reemplaza con rangos para los cuales ya se ha construido la distribución.

No todas las pruebas no paramétricas para los métodos basados en permutación están basadas en rangos que se pueden expresar como sus equivalentes paramétricas. Algunas pruebas no paramétricas como las pruebas de rangos con signos y de muestras independientes de Wilcoxon, pueden tener ventajas poderosas muy grandes sobre sus contrapartes paramétricas, como la prueba t de muestras apareadas e independientes. En la práctica del análisis de datos, los métodos paramétricos pueden tener ventajas poderosas en algunas situaciones, ninguno de los conjuntos de pruebas mantiene un dominio absoluto.

La mayoría de los métodos de permutación que utilizaban puntuaciones originales fueron de gran importancia teórica en vez de práctica ya que hasta la llegada del computador es que se facilitó el uso de estos algoritmos, paquetes como StatXact de Cytel software corporation, SC (StatisticalCalculator) de Mole software, entre otros.

La iteración es muy importante en el efecto de una variable dependiente de su asociación con otras variables. Un efecto importante de los métodos de permutación basados en rangos es su incapacidad para detectar las iteraciones de manera significativa.

Bibliografía

- Altman, D. G., & Bland, J. M. (1983). *Measurement in medicine: the analysis of method comparison studies*. Journal of the Royal Statistical Society: Series D (The Statistician), 32(3), 307-317.
- Altman, D. G., & Bland, J. M. (1996). *Absence of evidence is not evidence of absence*. AustVet J, 74(4), 311.
- Argimon, J. M. (2002). *La ausencia de significación estadística en un ensayo clínico no significa equivalencia terapéutica*. Medicina Clínica, 118(18), 701-703.
- Baldessarini, R. J., Finklestein, S., & Arana, G. W. (1983). *The predictive power of diagnostic tests and the effect of prevalence of illness*. Archives of general psychiatry, 40(5), 569-573.
- Basterra-Gortari, F. J., Bes-Rastrollo, M., Seguí-Gómez, M., Forga, L., Martínez, J. A., & Martínez-González, M. Á. (2007). *Tendencias de la obesidad, diabetes mellitus, hipertensión e hipercolesterolemia en España (1997-2003)*. Medicina Clínica, 129(11), 405-408.
- Blair, R. C., & Taylor, R. A. (2008). *Biostatistics for the health sciences*. Pearson Prentice Hall.
- Bland, J. M., & Altman, D. (1986). *Statistical methods for assessing agreement between two methods of clinical measurement*. The Lancet, 327(8476), 307-310.
- Burgueño, M. J., García-Bastos, J. L., & González-Buitrago, J. M. (1995). *Las curvas ROC en la evaluación de las pruebas diagnósticas*. Med Clin (Barc), 104(17), 661-70.
- Cornbleet, P. J., & Gochman, N. (1979). *Incorrect least-squares regression coefficients in method-comparison analysis*. Clinical chemistry, 25(3), 432-438.
- Cramer, H., & Cansado, E. (1953). *Métodos matemáticos de estadística (No. QA276. C72 1970.)*. Aguilar.
- Fleiss, J. L. (2011). *Design and analysis of clinical experiments*. (Vol. 73). John Wiley & Sons.
- Fleiss, J. L., Levin, B., & Paik, M. C. (2013). *Statistical methods for rates and proportions*. John Wiley & Sons.
- Fuentes Arderiu, X. (1998). *Bioquímica clínica y patología molecular*. (Vol. 2). Reverté. Condiciones y limitaciones de la aplicación.
- Galen, R. S., & Gambino, S. R. (1975). *Beyond normality: the predictive value and efficiency of medical diagnoses*. (pp. 50-51). New York: Wiley.
- Gardner, P. L. (1995). *Measuring attitudes to science: Unidimensionality and internal consistency revisited*. Research in science education, 25(3), 283-289.
- Hochberg, J., & TAMHANE, A. C. (1987). *Multiple comparison procedures*. (No. 519.535 H655m Ej. 1). John Wiley & Sons.
- Horbar, J. D., Onstad, L. Y. N. N., & Wright, E. L. I. Z. A. B. E. T. H. (1993). *Predicting mortality risk for infants weighing 501 to 1500 grams at birth: a National Institutes of Health Neonatal Research Network report*. Critical care medicine, 21(1), 12-18.
- Kirk, R. E. (1974). *Experimental design: Procedures for the behavioral sciences*. Pacific Grove, CA: Brooks/Cole.
- Kruskal, W. H., & Wallis, W. A. (1952). *Use of ranks in one-criterion variance analysis*. Journal of the American statistical Association, 47(260), 583-621.

- Lasky, F. D. (1992). *Proficiency testing linked to the national reference system for the clinical laboratory: a proposal for achieving accuracy*. *Clinical chemistry*, 38(7), 1260-1267.
- MariaRosariaRuggiero, YacelliBustamante, Claudia Mark. (1948). *Bioestadística I*. Universidad Central de Caracas Venezuela: Bioanálisis.
- Marshall, G., Tapia, J. L., D'Aprémont, I., Grandi, C., Barros, C., Alegria, A., ...&Bancalari, A. (2005). *A new score for predicting neonatal very low birth weight mortality risk in the NEOCOSUR South American Network*. *Journal of perinatology*, 25(9), 577.
- McNeil, B. J., Keeler, E., &Adelstein, S. J. (1975). *Primer on certain elements of medical decision making*. *New England Journal of Medicine*, 293(5), 211-215.
- Pita Fernández, S., &PértegaDíaz, S. (2000). *Significancia estadística y relevancia clínica*. *Cad Aten Primaria*, 8, 191-195.
- Pita Fernández, S., Pértega Díaz, S., & Rodríguez Maseda, E. (2003). *La fiabilidad de las mediciones clínicas: el análisis de concordancia para variables numéricas*. *Cad Aten Primaria*, 10(4), 290-6.
- Pita, F. S. (1996). *Correlación frente a la Causalidad*. *Jano*, 4, 59-260.
- Portillo, J. D. (2011). *Guía práctica del curso de bioestadística aplicada a las ciencias de la salud*. Instituto Nacional de Gestión Sanitaria, Servicio de Recursos Documentales y Apoyo Institucional.
- PRIETO, L. (1998). *La evaluación de la fiabilidad en las observaciones clínicas: el coeficiente de correlación intraclase*. *Med Clin (Barc)*, 110, 142-145.
- Royston, P. (1992). *Approximating the Shapiro-Wilk W-Test for non-normality*. *Statistics and computing*, 2(3), 117-119.
- Sánchez-Villegas, A. (2014). *Bioestadística amigable*. M. Á. Martínez-González, & F. J. Faulín (Eds.). Barcelona: Elsevier.
- Scaillet, O. (2005). *A Kolmogorov-Smirnov type test for positive quadrant dependence*. *Canadian Journal of Statistics*, 33(3), 415-427.
- Shen, W., Neyman, J., Pearson, E. S., Bolch, G., Greiner, S., de Meer, H., ...& Trivedi, K. S. (2018). *On the problem of the most efficient tests of statistical hypotheses*. *Interfaces*, 48(3), 285-289.
- Wakkers, P. M., Hellendoorn, H. B. A., De Weegh, G. O., &Heerspink, W. (1975). *Applications of statistics in clinical chemistry: A critical evaluation of regression lines*. *ClinicaChimicaActa*, 64(2), 173-184.
- Westgard, J. O., & Hunt, M. R. (1973). *Use and interpretation of common statistical tests in method-comparison studies*. *Clinical Chemistry*, 19(1), 49-57.
- Zaydman, M. A., Silva, J. R., Delaloye, K., Li, Y., Liang, H., Larsson, H. P., ... & Cui, J. (2013). *Kv7. 1 ion channels require a lipid to couple voltage sensing to pore opening*. *Proceedings of the National Academy of Sciences*, 110(32), 13180-13185.
- Zweig, M. H., & Campbell, G. (1993). *Receiver-operating characteristic (ROC) plots: a fundamental evaluation tool in clinical medicine*. *Clinical chemistry*, 39(4), 561-577.

Anexos

Tablas

¿Es estadísticamente significativo?	¿Es clínicamente relevante?	
*Si ($\rho < 0,05$)	*Si (éxito) *Clínicamente relevante *Estadísticamente significativo	*No (fracaso) *Clínicamente irrelevante ($\rho < 0,05$) *Estadísticamente significativo *¿Error Alfa?
*Si ($\rho \geq 0,1$)	*Clínicamente relevante *Estadísticamente NO significativo *¿Error Beta?	*Clínicamente irrelevante *Estadísticamente NO significativo

Tabla 1. Estadísticamente significativo a lo clínicamente relevante (tabla basada en Portillo, J. D., 2011, Pág. 32).

Variable Independiente o Predictora	Variable Dependiente o Resultado	Prueba de contraste de hipótesis	Observaciones y/o limitaciones del test
*Cualitativa	*Cualitativa	*Ji cuadrado *Prueba exacta de Fisher *Test de McNemar *Q de Cochran *Regresión Logística	*Si el tamaño muestral es grande *Si el tamaño muestral es pequeño *Dos grupos de muestras repetidas o apareadas * Más de dos grupos de muestras repetidas o apareadas *Varias variables predictoras cualitativas
*Cualitativa	*Cuantitativa	*t Student *Análisis de la Varianza *Test de Mann-Whitney *Test de Wilcoxon *Test de Kruskal-Wallis *Test de Fiedman	*Compara medias entre dos grupos independientes *Compara medias entre más de dos grupos independientes *Dos grupos independientes No paramétricos *Dos grupos medidas repetidas No paramétricos *Más de dos grupos No paramétrico, grupos independientes *Más de dos grupos No paramétrico medidas repetidas
*Cuantitativa	*Cualitativa	*Regresión logística simple *Regresión logística multivariable	* Una sola variable predictora cuantitativa * Varias variable predictoras cuantitativa o cualitativa
*Cuantitativa	*Cuantitativa	*Regresión lineal simple *Correlación de Pearson *Correlación de Sperman *Regresión Múltiple	* Paramétrico *Paramétrico * No paramétrico *Varias variables predictoras cuantitativas
*Cualitativa	*Supervivencia	*Kaplan-Meier *Long-Rank	*Curvas de supervivencia *Compara curvas de supervivencia

		*Regresión de COX	*Análisis Multivariante
--	--	-------------------	-------------------------

Tabla 2. Pruebas estadísticas de contraste de hipótesis más aplicables a cada situación (tabla basada en Portillo, J. D., 2011, Pág. 83).

Tipos de distribución	Simétrica	Asimétrica
Índice de tendencia central	Media	Mediana
Índice de dispersión	EEM	Rango intercuartil
“Peso” de valores extremos (outlier o periféricos)	Importante	Escaso o nulo
Fiabilidad del estadístico	Mucha	Poca

Tabla 3. Estadística descriptiva (tabla basada en Portillo, J. D., 2011, Pág. 146).

		Realidad
	H ₀ Falsa	H ₀ Cierta
	A ≠ B	A = B
Rechaza H ₀	Correcto	Error de tipo I
Acepta H ₁ Encuentra pruebas (+) ρ < 0,05	Se rechaza la hipótesis nula (H ₀) (se dice que los resultados son estadísticamente significativos y es verdad).	Se rechaza la hipótesis nula (H ₀) (se dice que los resultados son estadísticamente significativos y es verdad) pero nos equivocamos. Probabilidad α ERROR ALFA
	VP	FP
No Rechaza H ₀	Error de tipo II	Correcto
Acepta H ₀ A = B	No se rechaza la hipótesis nula (H ₀)	No se rechaza la hipótesis nula (H ₀)
No encuentra pruebas (-) ρ > 0,05	Se dice que no hay diferencias significativas pero no es así. Probabilidades β. ERROR BETA.	El azar puede explicar las diferencias observadas en los datos, y es verdad.
	FN	VN

Tabla 4. Tipos de error al contrastar hipótesis (tabla basada en Portillo, J. D., 2011, Pág. 179).

	Test Unilateral	Test Bilateral
Comparación de dos proporciones	$Z_{1-\beta} = \frac{ p_1 - p_2 \sqrt{n} - Z_{1-\alpha} \sqrt{2p(1-p)}}{\sqrt{p_1(1-p_1) + p_2(1-p_2)}}$	$Z_{1-\beta} = \frac{ p_1 - p_2 \sqrt{n} - Z_{1-\alpha/2} \sqrt{2p(1-p)}}{\sqrt{p_1(1-p_1) + p_2(1-p_2)}}$
Comparación de dos medias	$Z_{1-\beta} = \sqrt{\frac{\bar{n} d}{2S}} - Z_{1-\alpha}$	$Z_{1-\beta} = \sqrt{\frac{\bar{n} d}{2S}} - Z_{1-\alpha/2}$
Estimación de un OR en estudios de casos y controles	$p_1 = \frac{OR \cdot p_2}{(1-p_2) + OR \cdot p_2}$ $Z_{1-\beta} = \frac{ p_1 - p_2 \sqrt{n} - Z_{1-\alpha} \sqrt{(c+1)p(1-p)}}{\sqrt{c \cdot p_1(1-p_1) + p_2(1-p_2)}}$ $m = c \cdot n$	$p_1 = \frac{OR \cdot p_2}{(1-p_2) + OR \cdot p_2}$ $Z_{1-\beta} = \frac{ p_1 - p_2 \sqrt{n} - Z_{1-\alpha/2} \sqrt{(c+1)p(1-p)}}{\sqrt{c \cdot p_1(1-p_1) + p_2(1-p_2)}}$ $m = c \cdot n$
Estimación de un RR	$p_1 = RR \cdot p_2$ $Z_{1-\beta} = \frac{ p_1 - p_2 \sqrt{n} - Z_{1-\alpha} \sqrt{2p(1-p)}}{\sqrt{p_1(1-p_1) + p_2(1-p_2)}}$	$p_1 = RR \cdot p_2$ $Z_{1-\beta} = \frac{ p_1 - p_2 \sqrt{n} - Z_{1-\alpha} \sqrt{2p(1-p)}}{\sqrt{p_1(1-p_1) + p_2(1-p_2)}}$
Estimación de un coeficiente	$Z_{1-\beta} = \sqrt{n} - 3 \frac{1}{2} \ln \left(\frac{1+r}{1-r} \right) - Z_{1-\alpha}$	$Z_{1-\beta} = \sqrt{n} - 3 \frac{1}{2} \ln \left(\frac{1+r}{1-r} \right) - Z_{1-\alpha/2}$

de correlación lineal		
<ul style="list-style-type: none"> • n = Tamaño muestra. En un estudio de casos y controles, n es el número de casos. • p_1 = En un estudio transversal o de cohortes, proporción de expuestos que desarrollan la enfermedad. En un estudio de casos y controles, proporción de casos expuestos. • p_2 = En un estudio transversal o de cohortes, proporción de no expuestos que desarrollan la enfermedad. En un estudio de casos y controles, proporción de controles expuestos. $p = \frac{p_1 + p_2}{2}$ <ul style="list-style-type: none"> • d = Valor mínimo de la diferencia a detectar entre dos medias. • S^2 = Varianza en el grupo control o de referencia. 		

Tabla 5. Fórmulas para el cálculo estadístico $(1 - \beta)$ en función del tipo de diseño del estudio (Portillo, J. D., 2011, Pág. 206).

<p>$\rho < 0,05$ $A \neq B$ Se rechaza la hipótesis nula No parece que el azar pueda explicarlo todo</p> <p>El efecto observado es mayor que el error Hay diferencias estadísticamente significativas</p> <p>Existen evidencias a favor de la hipótesis alternativa H_a, y por tanto rechazamos la hipótesis nula H_0</p>	<p>$\rho \geq 0,10$ $A = B$ No se puede rechazar la hipótesis nula No se puede descartar que el azar lo explique todo</p> <p>El efecto observado es similar que el error NO hay diferencias estadísticamente significativas NO existen evidencias a favor de la hipótesis alternativa H_a, y por tanto aceptamos la hipótesis nula H_0</p>
------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------	------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------

Tabla 6. Decisiones que se pueden tomar en un test de contraste de hipótesis después de conocer el valor de significación estadística p (tabla basada en Portillo, J. D., 2011, Pág. 212).

		Variable Resultado	
ANÁLISIS ESTADÍSTICO		Dicotómica	Continua
Variable Predictora	Dicotómica	Test de χ^2	Regresión logística t de Student
Variable Independiente	Continua	Regresión Logística	Regresión Lineal

Tabla 7. Análisis Estadístico (tabla basada en Portillo, J. D., 2011, Pág. 359).

	Variable Resultado o Dependiente (criterio de evaluación)			
Variable Predictora o independiente (factor de estudio)	Cualitativa nominal dos categorías	Cualitativa nominal >2 categorías	Cualitativa ordinal o Cuantitativa No paramétrica	Cuantitativa paramétrica
Cualitativa dos grupos independientes	*Chi Cuadrado *Z comparación de proporciones *Test exacto de Fisher	*Chi cuadrado	*U de Mann-Whitney	*T de Student (prueba de Welch)

Cualitativa dos grupos apareados	*Test de McNemar	*Q de Cochran	*Prueba de los rangos de Wilcoxon	*T de Student para datos emparejados
Cualitativa > 2 grupos independientes	*Chi Cuadrado	*Chi cuadrado	*Prueba de Kruskal-Wallis	*Análisis de la Varianza (ANOVA)
Cualitativa > 2 grupos apareados	*Q de Cochran	*Q de Cochran	*Prueba de Friedman	*Análisis de la Varianza de dos vías
Cuantitativa	*Regresión logística	*Regresión logística	*Correlación de Spearman *Tau de Kendall	*Regresión lineal *Correlación de Pearson

Tabla 8. Tipos de Test Analíticos Estadísticos Bivariantes de contraste de hipótesis más frecuentes (tabla basada en Portillo, J. D., 2011, Pág. 361).

Variable Independiente o Predictora	Variable Dependiente o Resultado	Pruebas de contraste de hipótesis	Observaciones y/o limitaciones del test
Cualitativa	Cualitativa	Ji Cuadrado	Si el tamaño muestral es grande
		Prueba exacta de Fisher	Si el tamaño muestral es pequeño
		Test de McNemar	Dos grupos de muestras repetidas o apareadas
		Q de Cochran	Más de dos grupos de muestras repetidas o apareadas
		Regresión Logística	Varias variables predictoras cualitativas
Cualitativa	Cuantitativa	T de Student	Compara medias entre dos grupos independientes
		Análisis de la Varianza	Compara medias entre más de dos grupos independientes
		Test de Mann-Whitney	Dos grupos independientes NO Paramétricos
		Test de Wilcoxon	Dos grupos medias repetidas No Paramétricos
		Test de Kruskal-Wallis	Más de dos grupos No Paramétricos medidas repetidas
		Test de Friedman	Más de dos grupos No Paramétricos medidas repetidas
Cuantitativa	Cualitativa	Regresión Logística simple	Una sola variable predictora cuantitativa
		Regresión Logística Multivariante	Varias variables predictoras cuantitativa o cualitativas
Cuantitativa	Cuantitativa	Regresión lineal simple	Paramétrico
		Correlación de Pearson	Paramétrico
		Correlación de Spearman	No paramétrico
		Regresión Múltiple	Varias variables predictoras cuantitativas
Cualitativa	Supervivencia	Kaplan-Meier	Curvas de supervivencia
		Long-Rank	Compara curvas de supervivencia
		Regresión de COX	Análisis Multivariante

Tabla 9. Pruebas Estadísticas de contraste de hipótesis más aplicables a cada situación (tabla basada en Portillo, J. D., 2011, Pág. 361).

Pearson	Spearman
----------------	-----------------

-Ambas variables deben seguir la distribución Normal. -Sólo detecta asociaciones lineales. -Las observaciones han de ser independientes.	-No requiere supuestos previos de normalidad. -Detecta asociaciones distintas de la lineal. -Es un método no paramétrico pero permite la construcción de intervalos de confianza. -Estima el grado de asociación de cualquier tipo de independencia de la función matemática que relaciona ambas variables.
------------------------------------------------------------------------------------------------------------------------------------------------	----------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------

Tabla 10. Coeficientes de Correlación (tabla basada en Portillo, J. D., 2011, Pág. 687).

Observador A (o primera medición)	Observador B (o segunda medición)		Total
	Anormal (+)	Normal (-)	
Anormal (+)	a	b	a + b
Normal (-)	b	d	c + d
Total	a + c	b + d	N

Tabla 11. Índice Kappa de Cohen (medida dicotómica), (tabla basada en Portillo, J. D., 2011, Pág. 996).

Valor de k	Fuerza de Concordancia
< 0,20	pobre
0,21 – 0,40	débil
0,41 – 0,60	moderada
0,61 – 0,80	buena
0,81 – 1,00	muy buena

Tabla 12. Valores del índice Kappa (tabla basada en Portillo, J. D., 2011, Pág. 999).

CCI	Fuerza de la concordancia
>0,90	Muy buena
0,71 - 0,90	Buena
0,51 – 0,70	Moderada
0,31 – 0,50	Mediocre
< 0,30	Mala o nula

Tabla 13. Coeficiente de Correlación Intraclase (tabla basada en Portillo, J. D., 2011, Pág. 1059).

Valor de la σ	Fuerza de la Concordancia
>0,90	Pobre
0,90 - 0,95	Moderada
0,95 – 0,99	Sustanciales
> 0,99	Casi perfecto

Tabla 14. Escala de McBride para el coeficiente de concordancia de Lin (tabla basada en Portillo, J. D., 2011, Pág. 1088).

	Regresión Lineal Múltiple	Regresión Logística Multivariante
Variable Resultado	Cuantitativa	Cualitativa (dicotónica)
Variable Predictoras	Cualitativa - Cuantitativa	Cualitativa - Cuantitativa
Ecuación	Ecuación 2.41	Ecuación 2.42

Tabla 15. Análisis de la Varianza (ANOVA) de dos factores (tabla basada en Portillo, J. D., 2011, Pág. 1167).

Estadístico	Hipótesis Nula	Contraste	Significación
β_i	$\beta_i = 0$	t de Student	$p < 0,05$
Wald	$\beta_i = 0$	Chi Cuadrado	$p < 0,05$
$(\beta_i)^2 / (EE\beta_i)^2$			
-2Log Likelihood (-2LL)	$L = 1$ ó $(-2LL) = 0$ “ajuste perfecto”	Chi Cuadrado con n-k-1 grados de libertad	Si rechaza la hipótesis nula ($p < 0,05$) significa que el modelo de regresión no es significativo
Improvement o cambio en la razón de verosimilitud en cada paso	$(-2LLK-1) - (-2LLK)$	Chi Cuadrado con k grados de libertad	$p < 0,05$
Z^2 (Goodness of Fit)	$Z^2 \neq 0$	Chi Cuadrado	Si se rechaza la hipótesis nula ($p < 0,05$) significa que el modelo de regresión no es significativo
Model Chi-Square	$(-2LL^0) - (-2LL^K)$	Chi Cuadrado con k grados de libertad	$p < 0,05$

Tabla 16. Evaluación del modelo de regresión logística multivariante mediante el programa estadístico SPSS (tabla basada en Portillo, J. D., 2011, Pág. 1186).

Estadístico	Hipótesis Nula	Constante	Significación
β_i	$\beta_i = 0$	t de Student	$p < 0,05$
Wald	$W = 0$	Chi Cuadrado	$p < 0,05$
$(\beta_i)^2 / (EE\beta_i)^2$			

Tabla 17. Evaluación del modelo de regresión logística multivariante, si se rechaza la hipótesis nula H_0 ($p < 0,05$) significa que el modelo de regresión es significativo (tabla basada en Portillo, J. D., 2011, Pág. 1187).

	Regresión Múltiple de Cox
--	----------------------------------

Variable Resultado	Cuantitativas Desenlace: supervivencia a tiempo t
Variables Predictoras	Cualitativa Cuantitativa (factores de pronóstico)
Ecuación	$\ln(\lambda_t) = a + b_1x_1 + b_2x_2 + \dots + b_px_p$

Tabla 18. Modelos de Regresión Múltiple de Cox (tabla basada en Portillo, J. D., 2011, Pág. 1330).

Ecuaciones

$$\bar{X} = \sum_{i=1}^n \frac{x_i}{n} \quad \text{Ecuación 2.1}$$

$$\text{Media geométrica} = \sqrt[n]{x_1 x_2 \dots x_n} = \sqrt[n]{\prod_i^n x_i} \quad \text{Ecuación 2.2}$$

$$\text{Media armónica} = \frac{n}{\sum\left(\frac{1}{x_i}\right)} \quad \text{Ecuación 2.3}$$

$$\text{Media Ponderada} = \frac{\sum(w_i x_i)}{\sum(x_i)} = \frac{w_1 x_1 + w_2 x_2 + \dots + w_n x_n}{w_1 + w_2 + \dots + w_n} \quad \text{Ecuación 2.4}$$

$$s^2 = \frac{\sum(x_i - \bar{x})^2}{n-1} \quad \text{Ecuación 2.5}$$

$$s = \sqrt{\frac{\sum(x_i - \bar{x})^2}{n-1}} \quad \text{Ecuación 2.6}$$

$$\text{Coef. de variación} = \frac{\text{Desviación típica}}{\text{media}} = \frac{s}{\bar{x}} \times 100 \quad \text{Ecuación 2.7}$$

$$\text{EEM} = \frac{s}{\sqrt{n}} \quad \text{Ecuación 2.8}$$

$$\gamma_1 = \frac{\sum \varepsilon_i^3}{\sigma_R^3} \quad \text{Ecuación 2.9}$$

$$\gamma_2 = \frac{\sum \varepsilon_i^4}{\sigma_R^4} - 3 \quad \text{Ecuación 2.10}$$

$$[\text{Asimetría}] < 2EE_{\text{asimetría}} \quad \gamma_1 < 2EE\gamma_1 \quad \text{Ecuación 2.11}$$

$$[\text{curtois}] < 2EE_{\text{curtois}} \quad \gamma_2 < 2EE\gamma_2 \quad \text{Ecuación 2.12}$$

$$JB = t_1^2 + t_2^2 = T \left[\frac{\hat{\gamma}_1^2}{6} + \frac{\hat{\gamma}_2^2}{24} \right] \quad \text{Ecuación 2.13}$$

$$N = \frac{Z^2 \alpha * P * (1-P)}{i^2} \quad \text{Ecuación 2.14}$$

$$N = \frac{Z^2 \alpha * S^2}{i^2} \quad \text{Ecuación 2.15}$$

$$Z = \frac{(A_1 - A_2)}{\sqrt{(EE_{A1})^2 + (EE_{A2})^2}} \quad \text{Ecuación 2.16}$$

$$Z = \frac{(A_1 - A_2)}{\sqrt{(EE_{A1})^2 + (EE_{A2})^2 - 2r * EE_{A1} * EE_{A2}}} \quad \text{Ecuación 2.17}$$

$$\chi^2 = \sum \frac{(f_0 - f_e)^2}{f_0} \quad \text{Ecuación 2.18}$$

$$RV\chi^2 = 2 * \sum f_0 * \ln \frac{f_e}{f_0} \quad \text{Ecuación 2.19}$$

$$\chi^2_Q = \frac{(K-1)[K \sum Gn^2 - (\sum Gn)^2]}{K \sum Lc - \sum Lc^2} \quad \text{Ecuación 2.20}$$

$$C_{\text{máx}} = \sqrt{\frac{k-1}{k}} \quad \text{Ecuación 2.21}$$

$$C = \sqrt{\frac{\chi^2}{\chi^2+n}}$$

Ecuación 2.22

$$S_{xy} = \frac{1}{n} \sum (x_i - \bar{x}) (y_i - \bar{y})$$

Ecuación 2.23

$$r_{xy} = \frac{\sum_{i=1}^n (X_i - \bar{X})(Y_i - \bar{Y})}{n \cdot S_x \cdot S_y}$$

Ecuación 2.24

$$r_s = 1 - \frac{6 \cdot \sum_{i=1}^n d_i^2}{n(n^2-1)}$$

Ecuación 2.25

$$U_1 = n_1 n_2 + \frac{n_1(n_1+1)}{2} - S_1$$

Ecuaciones 2.26

$$U_2 = n_1 n_2 + \frac{n_2(n_2+1)}{2} - S_2$$

Ecuaciones 2.27

$$Z = \frac{U - \bar{U}}{\sigma_u}$$

Ecuación 2.28

$$Li = \sum_i^k \frac{t_i^3 - t_i}{12}$$

Ecuación 2.29

$$\sigma_u = \sqrt{\frac{n_1 n_2}{n(n-1)} \left(\frac{n^3 - n}{12} - \sum_i^k \frac{t_i^3 - t_i}{12} \right)}$$

Ecuación 2.30

$$\log \frac{P}{1-P} = \beta_0 + \beta_1 X$$

Ecuación 2.31

$$P = \frac{e^{\beta_1 X_1}}{1 + e^{\beta_1 X_1}}$$

Ecuación 2.32

$$H = \frac{12}{n(n+1)} \sum_{i=1}^k \frac{R_i^2}{n_i} - 3(n+1)$$

Ecuación 2.33

$$z = (R_i - R_j) \sqrt{\frac{M(M+1)}{6N}}$$

Ecuación 2.34

$$W = \frac{S}{1/12 K^2 (N^3 - N) - K \sum Li}$$

Ecuación 2.35

$$k = \frac{\sum (x_n/n) \sum (x_i/n)(x_i/n)}{1 - \sum (x_i/n)(x_i/n)}$$

Ecuación 2.36

$$\hat{k} = \frac{p_0 - p_e}{1 - p_e}$$

Ecuación 2.37

$$CV_t = \sqrt{CV_p^2 + CV_a^2 + CV_{BW}^2}$$

Ecuación 2.38

$$CV_t = \sqrt{CV_a^2 + CV_{BW}^2}$$

Ecuación 2.39

$$CCI = \frac{\sigma_p^2}{\sigma_p^2 + \sigma_0^2 + \sigma_R^2}$$

Ecuación 2.40

$$E(Y) = \beta_0 + \beta_1 X_1 + \dots + \beta_n X_n$$

Ecuación 2.41

$$\log \frac{P}{1-P} = \beta_0 + \beta_1 X_1 + \dots + \beta_p X_p$$

Ecuación 2.42

Figuras

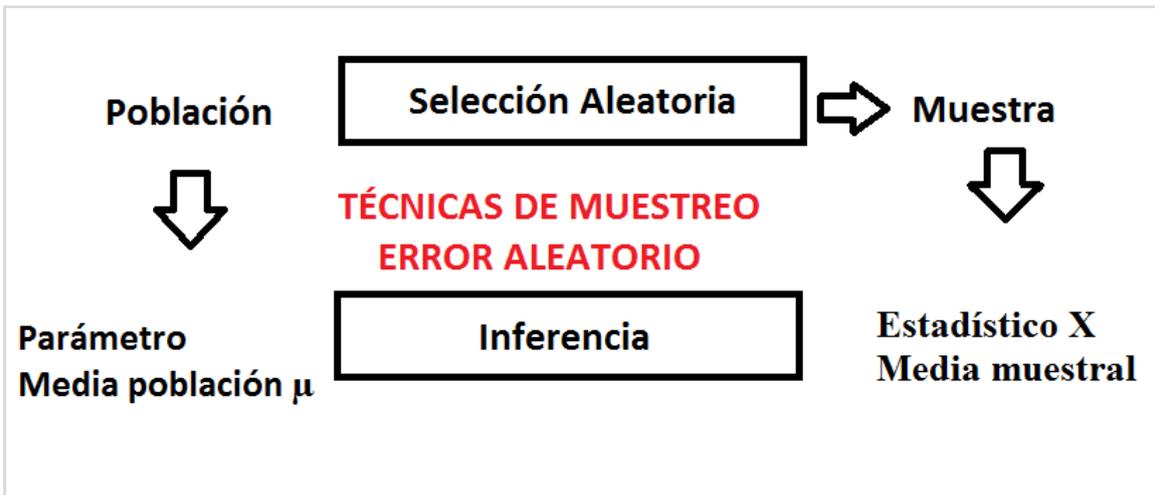


Figura 1. Selección de muestral, imagen basada en Portillo, J. D., 2011, Pág. 164.

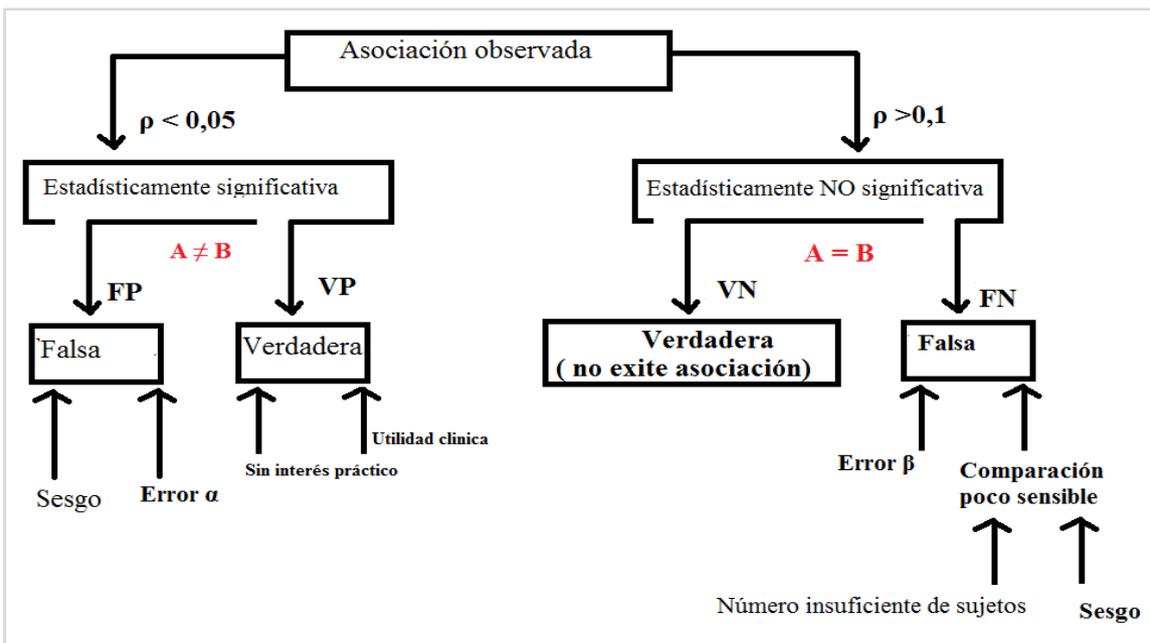


Figura 2. Interpretación de una asociación (o diferencia) observada en un estudio, imagen basada en Portillo, J. D., 2011, Pág. 210.

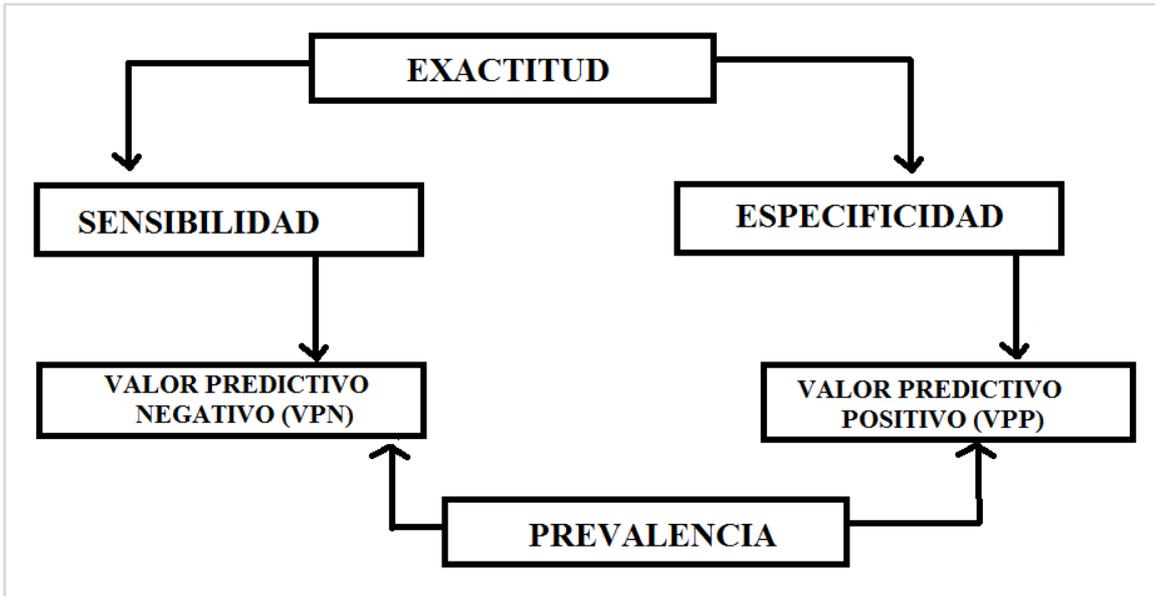


Figura 3. Exactitud, imagen basada en Portillo, J. D., 2011, Pág. 237.

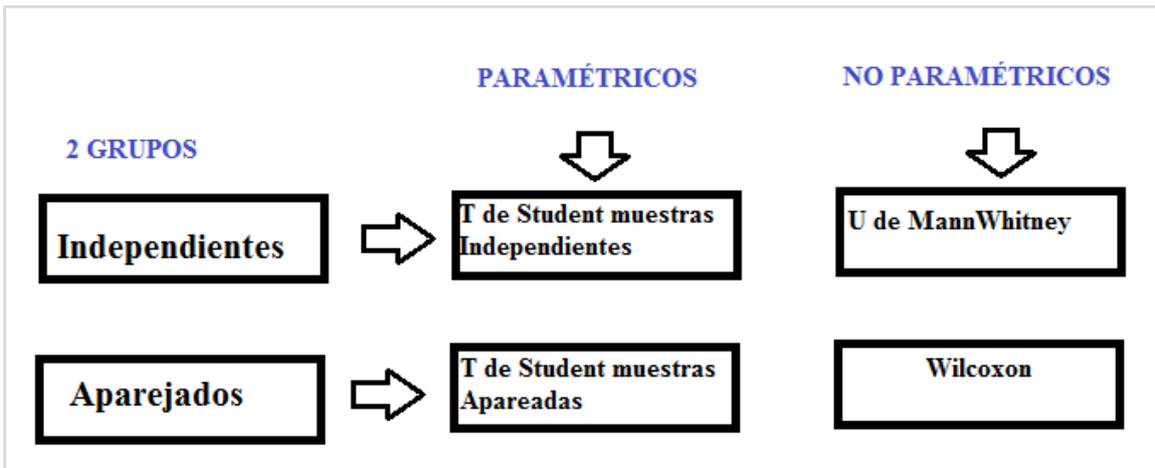


Figura 4. Cualitativa dicotómica versus cuantitativa, imagen basada en Portillo, J. D., 2011, Pág. 736.

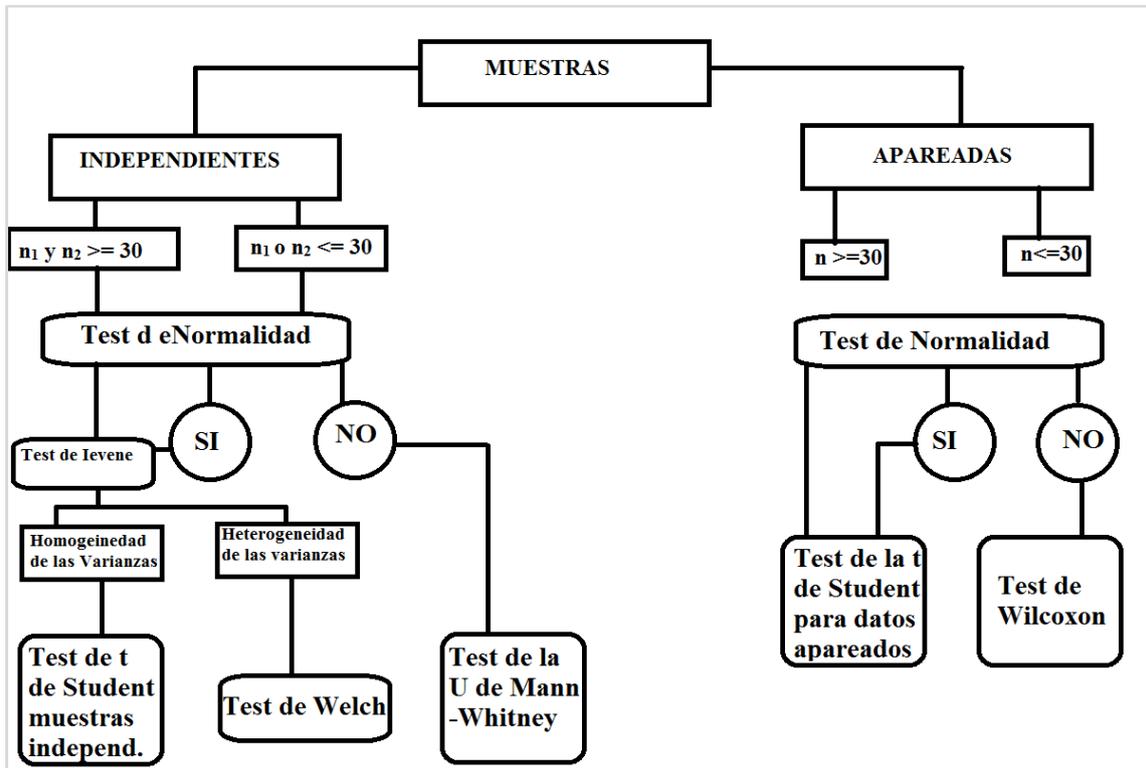


Figura 5. Muestras, imagen basada en Portillo, J. D., 2011, Pág. 852.



Figura 6. Cualitativa policotómica versus cuantitativa, imagen basada en Portillo, J. D., 2011, Pág. 880.

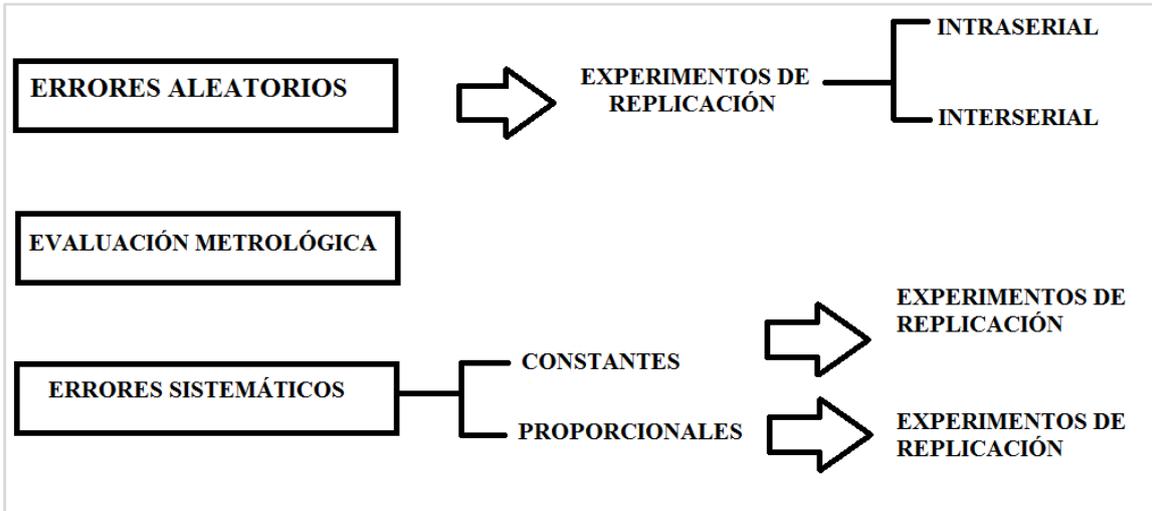


Figura 7. Evaluación Metrológica, imagen basada en Portillo, J. D., 2011, Pág. 852.

$$CV_t = \sqrt{CV_p^2 + CV_a^2 + CV_{BW}^2}$$

↓
↓
↓

Variación Preanalítica
Variación Analítica
Variación Biológica Intraindividual

Figura 8. Cálculo de la Variación Total (CV_t), imagen basada en Portillo, J. D., 2011, Pág. 1033.