



**USAC**  
TRICENTENARIA  
Universidad de San Carlos de Guatemala

**DC** Dirección General  
de Investigación  
Universidad de San Carlos de Guatemala

# Plan de análisis de datos para la investigación cuantitativa

Oscar Federico Nave Herrera  
Coordinador del Programa de  
Asesoría Estadística para  
Investigación



# Hipótesis

---

Explicación tentativa del fenómeno a investigar (problema de investigación), formulada en forma de proposición lógica.

Deben estar apoyadas en conocimientos organizados y estructurados relacionados con el problema de investigación.

# Tipos de hipótesis

---

1. Hipótesis de investigación

2. Hipótesis estadísticas

- Hipótesis nula
- Hipótesis alternativa

# Hipótesis empíricas

---

La respuesta a la pregunta de investigación se basa en la experiencia del investigador (no debe ser del todo subjetiva, sino más bien lógica y fundamentada en el conocimiento y estado del arte del tema).

“Al aplicar la nueva técnica para producción de etanol, el rendimiento será mayor al 90%”

# Hipótesis relacionales

Implican la mención de al menos dos variables vinculadas por medio de palabras que manifiesten alguna relación, como por ejemplo efecto, asociación, correlación o diferencia. En un estudio se espera que la variable independiente tenga algún efecto sobre la dependiente o que dos variables (supuestamente independientes) tengan un comportamiento que implique alguna relación entre ambas.

# Hipótesis relacionales

---

“El rendimiento del cultivo de cebolla con cepas genéticamente mejoradas será mayor que con cepas silvestres”

“El consumo de cigarrillos se asocia con el desarrollo de cáncer pulmonar”

“Al menos uno de los tratamientos ensayados tendrá efecto inhibitorio sobre leptospira”

# Hipótesis estadísticas

Escritura de las hipótesis nula y alternativa en términos simbólicos estadístico-matemáticos que implican la formulación de una ecuación o inecuación.

Ej:  $H_0: \mu_1 = \mu_2$

$H_a: \mu_1 \neq \mu_2$

$H_0: \mu_1 = \mu_2$

$H_a: \mu_1 > \mu_2$

# Plan de análisis

Exposición del plan que se deberá seguir para el tratamiento estadístico de los datos.

Está relacionado con el nivel de medición de las variables.



# Clases de estadística

---

- Descriptiva (deductiva). Es la fase de descripción, organización, síntesis y análisis de la información de interés, pero sin llegar a conclusiones fuertes o profundas sobre la misma.

No va más allá de lo que la muestra nos dice.

- Estadísticas descriptivas

Estadística numérica: Medidas de tendencia central, medidas de posición, medidas de dispersión.

Análisis de frecuencias.

Análisis de series de tiempo.

Estadística gráfica.

Análisis exploratorio de datos.

# Clases de estadística

- Inferencial (inductiva). Esta fase busca obtener conclusiones sólidas y más profundas que una simple descripción de la información, basados en el trabajo con muestras y su posterior generalización.
- Se obtiene información sobre una población de datos mediante el estudio de una muestra de los mismos.

Las muestras deben reflejar el verdadero comportamiento de las poblaciones.

- Estadísticas inferenciales

Estimación de valores poblacionales (parámetros).

Prueba de hipótesis.

Correlaciones y asociaciones.

Regresiones.

Técnicas de análisis multivariado.

Técnicas de análisis ráster (SIG)

# Estimación

Consiste en establecer o estimar el verdadero valor poblacional con base en una muestra, empleando la teoría de probabilidades para establecer un margen de precisión (intervalo de confianza).

Se pueden estimar valores para promedio, desviación estándar (varianza), parámetros de regresión, frecuencias, etc.

# Pruebas de hipótesis

---

Cuando se desarrolla un diseño experimental o cuasi experimental, las pruebas de hipótesis son generalmente la forma de análisis para determinar si el efecto fue significativo; el principio estadístico es el mismo que en las pruebas sobre poblaciones.

# Pruebas de hipótesis o de significación

---

Con base en datos muestrales que representan a la (s) población (es), se procede a establecer la probabilidad de que los datos sean iguales a lo que indica la  $H_0$  simplemente por causa del azar. Si dicha probabilidad es más grande que un nivel de error pre-establecido ( $\alpha$ ), la  $H_0$  no se puede rechazar: por el contrario, si la probabilidad del azar es menor o igual a dicho nivel de error, la  $H_0$  se rechaza.

# Las pruebas de significación

---

La “significación estadística” es simplemente la base para el rechazo de una hipótesis nula.

Un estudio debe diseñarse de forma tal que tenga una alta probabilidad de detectar efectos de la magnitud que el investigador considera importantes, si esos efectos realmente existen. Esa probabilidad es lo que se conoce como “Potencia estadística”.

**Clave para el cálculo de muestras o réplicas**

“Se evitará la dependencia exclusiva de las pruebas estadísticas de verificación de hipótesis, tal como el uso de los valores  $p$ , que no aportan ninguna información cuantitativa importante.”

(Comité Internacional de Directores de Revistas Médicas. Octubre, 2001. Requisitos uniformes de los manuscritos enviados a revistas biomédicas).

# Alternativas a las pruebas de significación

---

- En el diseño se debe considerar la Potencia.
- Si no se hace, es incorrecto calcular la Potencia como parte del análisis estadístico.

## Alternativas:

- Cálculo del tamaño del efecto (diferencia de medias estandarizada, correlación biserial puntual,  $\eta^2$ ,  $\omega^2$ , etc.)
- Intervalos de confianza

# Asociación

---

Conjunto de procedimientos estadísticos que cuantifican y evalúan la relación entre variables, generalmente de tipo cualitativo, aunque también entre variables cuantitativas y cualitativas.

# Correlación

---

Procedimiento estadístico que cuantifica y evalúa si existe alguna relación entre dos variables.

Cuando las variables son cuantitativas de tipo continuo, el medio estadístico adecuado para medir la asociación y ser expresada como un índice, se conoce como correlación.

# Regresión

---

Proceso estadístico-matemático, por medio del cual se calcula una ecuación que mejor se ajuste a los datos, explicando con parámetros estadísticos el comportamiento de dos variables cuantitativas continuas, con el objeto de predecir o estimar valores.

Una variable debe depender de la otra (u otras).

# Tipos de pruebas estadísticas

- Estadística paramétrica

Aplicable a variables continuas con distribución normal y se basa en que la muestra representa adecuadamente a uno o más parámetros poblacionales

- Estadística no paramétrica

Para variables que no siguen una distribución normal, variables ordinales o discretas, frecuencias o proporciones y número de muestras pequeño.

# Estadística bayesiana

---

“Para el 2021 el adjetivo bayesiano para la estadística sería superfluo al ser bayesianas todas las aproximaciones a la estadística” (Dennis Lindley, 1976).

La metodología bayesiana está basada en la interpretación subjetiva de la probabilidad y tiene como punto central el Teorema de Bayes.

# El teorema de Bayes

---

La probabilidad a priori de una hipótesis,  $Pr(H)$ , se ve transformada en una probabilidad a posteriori,  $Pr(H|\text{datos})$ , una vez incorporada la evidencia que aportan los datos.

Entonces una vez observados los datos, el teorema "devuelve" una nueva distribución, que no es otra cosa que la percepción probabilística original actualizada por los datos.

# Modelos lineales generalizados

---

Constituyen una generalización de los modelos lineales clásicos (regresión lineal y regresión múltiple).

Incluye la regresión lineal, los modelos de análisis de varianza, modelos probit y logit para ensayos cuantales, modelos log-lineares (logística y Poisson), modelos de respuesta multinomial y modelos de análisis de sobrevivencia.

# Técnicas de análisis multivariado

---

Métodos estadísticos cuya finalidad es analizar simultáneamente conjuntos de datos en los que se tienen varias variables medidas para cada individuo u objeto estudiado (cualitativas o cuantitativas).

Su razón de ser radica en un mejor entendimiento del fenómeno objeto de estudio obteniendo información que los métodos estadísticos univariantes y bivariantes son incapaces de conseguir.

# Geoestadística

---

Es un conjunto de técnicas para el análisis y predicción de valores distribuidos en el espacio y/o en el tiempo, dichos valores se asumen correlacionados entre sí.

Evalúa interacciones espaciales, es decir que los valores de una variable están relacionados con su distribución en el espacio.

# Consideraciones finales

- Los aspectos relacionados con el número de muestra deben ser claros.
- El diseño de muestreo debe explicarse con detalle.
- El nivel de análisis debe corresponder con los objetivos, tipo de variables y muestra.
- No deben obviarse elementos indispensables para la correcta interpretación de los resultados (nivel de confiabilidad, nivel de significancia, comprobación de supuestos, etc.).
- El uso de software no es requisito.

# Investigación exploratoria

---

## Análisis:

- Basado en el criterio del investigador más que en métodos estadísticos
- Presentar los hallazgos de manera ordenada y lógica
- Presentar los datos numéricos necesarios, buscar tendencias, similitudes, diferencias, etc.
- No puede generalizarse

# Investigación descriptiva

---

## Análisis

- Resumen de los datos con estadística descriptiva: Medidas de tendencia central y dispersión, cuantiles y percentiles, frecuencias y porcentajes, tasas o índices (demográficos, económicos, epidemiológicos o de biodiversidad)

# Investigación descriptiva

---

## Análisis

- Uso de estadística gráfica: Barras, ciclogramas, barras de error, cajas, diagramas de tallos y hojas, histogramas.
- Estimación de parámetros por medio de intervalos de confianza

# Investigación correlacional (asociativa)

---

## Análisis

- Bivariado: Correlaciones bivariadas, Ji cuadrado, cálculo de Odds Ratio y riesgo relativo, análisis de concordancia (correlación de concordancia, Mc Nemar, Kappa de Cohen)
- Multivariado: Análisis de correspondencias, agrupamiento y discriminante, componentes principales, correlación canónica, Ji cuadrado de Mantel y Haenszel, regresión logística

# Investigación correlacional (asociativa)

---

## Análisis

- Estadística gráfica: Diagramas de dispersión, diagramas multivariados y dendrogramas
- En varios procedimientos se contempla el uso de intervalos de confianza o pruebas de hipótesis confirmativas para la correlación o asociación.

# Investigación explicativa: Estudios comparativos

---

## Análisis

- Pruebas de hipótesis sobre la media, mediana o distribución de una, dos o más poblaciones independientes (t de Student, análisis de varianza o covarianza, prueba del signo, U de Mann Whitney, Wilcoxon, Kruskal Wallis)
- Pruebas de hipótesis sobre la proporción de una, dos o más poblaciones independientes (prueba binomial, Z para proporciones, Ji cuadrado)

# Investigación explicativa: Estudios comparativos

---

## Análisis

- Análisis multivariados (Manova)
- Estadística gráfica: Barras de error, cajas, barras comparativas

# Investigación explicativa: Estudios experimentales y cuasi experimentales

## Análisis

- Un grupo: t de Student pareada, Ji cuadrado de Mc Nemar, Wilcoxon para muestras relacionadas
- Dos grupos: t de Student, prueba del signo, U de Mann Whitney, Wilcoxon para muestras independientes, Ji cuadrado, Z para proporciones

# Investigación explicativa: Estudios experimentales y cuasi experimentales

## Análisis

- Tres o más grupos: Análisis de varianza o covarianza (depende del diseño), pruebas no paramétricas (Kruskal Wallis, Friedman, Cochran), Ji cuadrado
- Análisis multivariado (Manova, factorial)
- Gráficas de caja, barras de error, líneas

# Investigación predictiva

## Análisis

- Análisis exploratorio de datos
- Regresión lineal simple o múltiple
- Regresión logística o regresión probit
- Análisis de varianza o covarianza
- Modelos lineales generalizados
- Modelos mixtos o jerárquicos (multinivel)
- Geoestadística