Fundamentos en Estadística

V. Trujillo

GRC-MERVEX (CO de Vigo. IEO)

Marzo 2021



Fundamentos en Estadística

V. Trujillo

GRC-MERVEX (CO de Vigo. IEO)

.- Intro

2.- Método científico y Método estadístico

Planteamiento del

l modelo

stadístico

a informac

DA. Análisis cploratorio de atos

Estimación d Parámetros

Parametros Contraste de

Crítica y A

del modelo

3.Diferenciación / Clasificación

descriptiva vs inferencial

> stadistica aramétrica v o-paramétric

Fundamentos en Estadística

7. Trujillo

GRC-MERVEX (CO de Vigo. IEO)

1.- Intro

1.- Intro

2.- Método científico y Método

Planteamiento del problema El modelo

estadístico

EDA. Análisis

cploratorio de atos

Parámetros
Contraste de
Simplificación

3.-Diferenciación

Diferenciación / Clasificación

Estadística descriptiva vs inferencial Estadística

paramétrica vs no-paramétrica ¿Cuales son las estrategias básicas de toda investigación?

Fundamentos en Estadística

V. Trujillo

RC-MERVEX (CO de Vigo. IEO)

Intro

.- Método ientífico y Iétodo stadístico

oblema modelo

stadístico a información

a información DA. Análisis

Datos
Estimación de Parámetros
Contraste de

3.Diferenciación

Clasificación Estadística Jescriptiva vs Inferencial

Fundamentos en Estadística

7. Trujillo

GRC-MERVEX (CO de Vigo. IEO)

.- Intro

2.- Metodo científico y Método estadístico

Planteamiento del

El modelo

La informa

EDA. Análisis Exploratorio de

xploratorio d atos

Estimación d Parámetros

Contraste de Simplificación

Simplificaci

Crítica y Anál del modelo

3.Diferenciación

Estadística descriptiva vs inferencial

> stadistica aramétrica vs o-paramétrics

► Explorar

► Identificar

Clasificar

► Relacionar

► Contrastar

Explicar

► Predecir

¿Históricamente, a qué necesidades atendía la Estadística?

Fundamentos en Estadística

V. Trujillo

RC-MERVEX
CO de Vigo.

Intro

Métode

adístico

oblema modelo tadístico

tadistico a información DA. Análisis

DA. Análisis xploratorio de atos

Estimación de Parámetros Contraste de Simplificación

ferenciación

ferenciación Clasificación stadística

nterencial Estadística paramétrica vs

Padrones de los estados: registro y descripción de los datos.

Cálculo de probabilidades: prácticas de los juegos de azar.

Fundamentos en Estadística

Contraste de

3 -

¿Qué problemas resuelve la Estadística?

Fundamentos en Estadística

V. Trujillo

GRC-MERVEX (CO de Vigo. IEO)

.- Intro

2.- Método científico y Método

problema El modelo

estadístico La información

EDA. Análisis

Estimación de Parámetros Contraste de Simplificación

3.-Diferenciación / Clasificación

Estadística descriptiva vs inferencial Estadística

Fundamentos en Estadística

V. Trujillo

GRC-MERVEX (CO de Vigo. IEO)

- 1.- Intro
- científico y
 Método
- Planteamiento del
- l modelo
- stadístico
- La informac
- EDA. Análisis Exploratorio de
- stimación (
- Parámetros Contraste de
- Simplificac
- Critica y Anál del modelo
- 3.Diferenciación

Estadística descriptiva v inferencial

> aramétrica v o-paramétric

- ▶ Descripción de datos
- ► Análisis de muestras
- ► Contraste de hipótesis
- ► Medición de relaciones¹
- ▶ Predicción y Proyección^{2,3}

2.- Método científico y Método estadístico

Fundamentos en Estadística

V. Trujillo

GRC-MERVEX (CO de Vigo. IEO)

.- Intro

2.- Método científico y Método

Planteamiento del problema

El modelo estadístico

EDA. Análisis

DA. Analisis Exploratorio d Oatos

Parámetros Contraste de

implificación Trítica y Anális el modelo

3.Diferenciación

Estadística descriptiva vs inferencial Estadística

aramétrica vs o-paramétrica

Método científico

Fundamentos en Estadística

V. Trujillo

RC-MERVEX (CO de Vigo. IEO)

- Intro

2.- Método científico y Método

problema El modelo estadístico

La información

atos Stimación o

Estimación de Parámetros Contraste de Simplificación

3.Diferenciación
/ Clasificación

Estadística descriptiva vs inferencial Estadística



Fundamentos en Estadística

V. Trujillo

GRC-MERVEX (CO de Vigo. IEO)

.- Intro

2.- Método científico y Método

> Planteamiento de problema

modelo tadístico

La informac

EDA. Análisis Exploratorio de

Estimación d Parámetros

Parámetros Contraste de

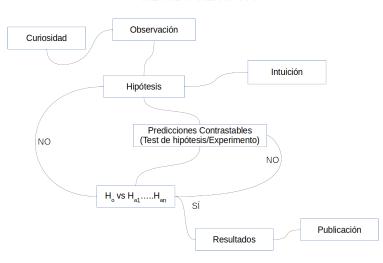
Simplificación

3.Diferenciación

Estadística descriptiva vs inferencial

> Estadística paramétrica v

MÉTODO CIENTÍFICO



Fundamentos en Estadística

V. Trujillo

GRC-MERVEX (CO de Vigo.

L.- Intro

2.- Método científico y Método

Planteamiento del

modelo

adístico

EDA Análisis

DA. Análisis Exploratorio de Datos

Estimación de Parámetros

Parámetros Contraste de

Crítica y Análisis lel modelo

3.Diferenciación

Estadística descriptiva vs

stadística aramétrica vs

Método estadístico

Fundamentos en Estadística

V. Trujillo

RC-MERVEX (CO de Vigo. IEO)

T... 4

2.- Método científico y Método

problema
El modelo

La información

xpioratorio d' atos stimación de

Estimación de Parámetros Contraste de Simplificación

3.-Diferenciación / Clasificación

Estadística descriptiva vs inferencial Estadística

MÉTODO ESTADÍSTICO Problema Real Planteamiento del Problema Objetivos y Medios

Modelo Estadístico Cálculo de Probabilidades

Recogida de Información Técnicas de Muestreo: Diseño.

Depuración de los Datos Análisis Exploratorio de Datos (EDA)

Estimación de los Parámetros Teoría del Estimación

Contraste de Simplificación Contraste de Hipótesis

Crítica y Diagnosis del Modelo Análisis de la Información

¿Es el Modelo Adecuado?

Conocimiento Científico

Prognosis

Decisiones

Fundamentos en Estadística

3 -

Planteamiento del problema

El primer paso consiste en definir los objetivos del estudio y relacionar estos con valores númericos de variables observables. También hay que tener bien definida la población sobre la que se va a actuar, las variables que mediremos y cómo hacerlo.

Fundamentos en Estadística

3 -

El modelo estadístico

El modelo estadístico más utilizado consiste en la descomposición de los valores de la variable en parte sistemática y parte aleatoria.

Se pueden clasificar como:

Fundamentos en Estadística

V. Trujillo

RC-MERVEX (CO de Vigo. IEO)

1.- Intro

2.- Método
científico y
Método
estadístico

Planteamiento del problema El modelo

información

DA. Análisis xploratorio de atos

Estimación de Parámetros

Contraste de

rítica y Análi el modelo

3.Diferenciación / Clasificación

descriptiva vs inferencial

> stadistica aramétrica vs o-paramétrics

Fundamentos en Estadística

Contraste de

Estáticos Dinámicos $\overline{Y} = \mu + \vartheta Y_{t-1} + \varepsilon_t$ Extrapolativos $Y = \mu + \varepsilon$ $Y = \mu + \beta X_t + \vartheta Y_{t-1} + \varepsilon_t$ Explicativos $\overline{Y} = \mu + \beta X + \varepsilon$

Dependiendo del <u>momento</u> en el cual se estudia la variabilidad tendremos dos tipos de modelos:

1. Estáticos

No tienen en cuenta el tiempo, son modelos aplicables o aplicados en un momento temporal dado, son los utilizados habitualmente en Biología y Química.

Se pueden clasificar en:

- 1.1 Estáticos extrapolativos: definidos por una variable.
- 1.2 Estáticos explicativos: definido por dos o más variables.

Fundamentos en Estadística

V. Trujillo

GRC-MERVEX (CO de Vigo. IEO)

- Intro

2.- Método científico y Método estadístico

Planteamiento del problema

stadístico

a información

A. Análisis ploratorio de tos

stimación de arámetros

Contraste de

Crítica y Anális del modelo

3.-Diferenciación

Clasificación

Estadística paramétrica vs

2. Dinámicos

Se aplican a situaciones que cambian a lo largo del tiempo, donde este aparece explícitamente.

Se pueden clasificar en:

- 2.1 Dinámicos extrapolativos: como el modelo de Box-Jenkins (B-J), que separa la parte estocástica de la variabilidad. Se usa por ej.: en macroeconomía.
- 2.2 Dinámicos explicativos: igual que B-J pero con dos variables y permite ver su relación, por ej.: Modelos de Regresión Dinámica

Fundamentos en Estadística

V. Trujillo

GRC-MERVEX (CO de Vigo. IEO)

l.- Intro

2.- Método científico y Método estadístico

Planteamiento del problema

estadísti

La información

EDA. Análisis Exploratorio de

stimación de arámetros

arámetros ontraste de

ontraste de implificación

Crítica y Análisis el modelo

3.Diferenciación
/ Clasificación

Estadística lescriptiva vs nferencial

stadistica aramétrica vs o-paramétrica

V. Trujillo

GRC-MERVEX (CO de Vigo. IEO)

.- Intro

2.- Método científico y Método estadístico

Planteamiento del problema

estadístic

La información

DA. Análisis Exploratorio de Datos

Estimación d Parámetros

'arámetros Contraste de

Contraste de Simplificación

Crítica y Análi: lel modelo

3.-Diferenciación

Estadística descriptiva vs inferencial

> stadística aramétrica vs o-paramétrics

Es decir, que dependiendo de la <u>variabilidad</u> que tenga el modelo, diremos que son modelos:

Extrapolativos:

Aquellos que explican la variabilidad tomando únicamente como información sus valores pasados. La parte sistemática del modelo es función de los valores pasados observados.

Explicativos:

Aquellos que tienen en cuenta el efecto de otras variables. La parte sistemática será la suma de la parte sistemática extrapolativa y del efecto de las variables explicativas.

Otra posible clasificación podría ser:

► Modelos Descriptivos:

Se ocupan de ordenar, clasificar, hacer visibles los datos y también medir relaciones simples entre variables.

► Modelos Inferenciales:

Se hace contraste de hipótesis, que puede ser paramétrica o no-paramétrica.

► Modelos Multivariados:

Se explican simultáneamente el comportamiento de varias variables, haciendo uso intensivo del cálculo matricial. Necesita obligadamente un ordenador debido a su complejidad.

Fundamentos en Estadística

V. Trujillo

GRC-MERVEX (CO de Vigo. IEO)

l.- Intro

2.- Método científico y Método estadístico

Planteamiento del problema

estadístic

EDA. Análisis

atos stimación de

Parámetros

Contraste de

Crítica y Análisi del modelo

3.-Diferenciación

> Estadística lescriptiva vs nferencial

stadística aramétrica vs

La información

La recogida de información y la forma que la Estadística se aproxima al método científico puede ser de dos formas:

Por procedimiento muestral (muestreo):
 El muestreo consiste en observar pasivamente una muestra de las variables y anotar sus valores.
 No se fija ninguna condición. Se utiliza principalmente en modelos extrapolativos. Ej. Modelo de efectos aleatorios

▶ Por diseño experimental:

(correlación).

El diseño de experimentos.

El diseño de experimentos consiste en fijar o controlar determinadas variables y observar la respuesta de las otras. El diseño experimental debe utilizarse cuando se quiera construir un modelo explicativo, ya que solo tendremos una base sólida para juzgar relaciones de "causalidad" entre variables, cuando los datos se obtengan mediante diseño experimental. Ej. Modelo de efectos fijos (regresión).

Fundamentos en Estadística

V. Truiillo

RC-MERVEX (CO de Vigo. IEO)

- Intro

2.- Método científico y Método estadístico

Planteamiento del problema

El modelo stadístico

EDA. Análisis Exploratorio de Datos

stimación de arámetros

Contraste de implificación

rítica y Anális el modelo

3.-Diferenciación

stadística escriptiva vs ferencial

stadística aramétrica v

Por ejemplo:

Regresión / Correlación

	Diseño Experimental	Muestreo	
Pros	Influencia clara entre variables	Puedo hacer predicciones de comportamiento de variables, siempre <u>dentro</u> del rango analizado.	
Cons	Puedo introducir artefactos.	No me asegura relación causa-efecto entre variables.	
	No puedo predecir nada a otras condiciones.		

Fundamentos en Estadística

V. Trujillo

GRC-MERVEX (CO de Vigo. IEO)

l.- Intro

2.- Método científico y Método estadístico

lanteamiento del roblema

El modelo estadístico

> EDA. Análisis Exploratorio de Datos

Estimación de Parámetros

Contraste de Simplificación

Crítica y Anál del modelo

3.Diferenciación / Clasificación

Estadística descriptiva vs inferencial Estadística

EDA. Análisis Exploratorio de Datos

En la fase de depuración de la muestra, se espera que alrededor de un 5 % de las observaciones tengan errores de medición, por tanto conviene utilizar técnicas estadísticas simples como las técnicas de análisis exploratorio de los datos (EDA), para así poder identificar los posibles errores, valores faltantes ("missing values") o los valores anómalos ("outliers").

Fundamentos en Estadística

V. Trujillo

GRC-MERVEX (CO de Vigo. IEO)

1.- Intro

2.- Método científico y Método estadístico

Planteamiento del

l modelo

La inforn

Estimación de Parámetros Contraste de

Crítica y Anál del modelo

3.Diferenciación / Clasificación

escriptiva vs iferencial

Estimación de Parámetros

Los modelos estadísticos dependen de constantes conocidas como **parámetros**⁴. En la fase de estimación se utiliza la información disponible para decidir el valor concreto de estos y cuantificar el posible error de la estimación.

Fundamentos en Estadística

V. Trujillo

GRC-MERVEX (CO de Vigo. IEO)

l.- Intro

2.- Método científico y Método estadístico

Planteamiento del

El modelo

estadístico

DA. Análisis

Contraste de Simplificación Crítica y Análisis

3.Diferenciación

Estadística descriptiva vs inferencial

Contraste de Simplificación

Una vez estimados los valores de los parámetros, estudiaremos si el modelo puede simplificarse.

El objetivo es conseguir un modelo lo más simple posible. sin más parámetros de los necesarios. "Entre dos modelos que expliquen unos datos igualmente, es mejor el más sencillo" (navaja de Ockham)⁵. Esta fase es especialmente importante en los modelos explicativos.

Las hipótesis tienen que ser contrastables favorablemente => leves (ecuaciones matemáticas).

Fundamentos en Estadística

Crítica y Análisis del modelo

Los resultados de las dos etapas anteriores se obtienen suponiendo que el modelo es correcto. Esta fase investiga la compatibilidad entre los valores empíricos y el modelo. Es de especial interés comprobar, que la parte aleatoria no contiene ninguna estructura sistemática.

Si después de esta fase se acepta el modelo como correcto, se usará para tomar decisiones o predecir valores de la variable. En caso contrario, se volverá a la segunda fase y se reformulará el modelo, repitiendo el proceso hasta encontrar un modelo más adecuado⁶.

Fundamentos en Estadística

V. Trujillo

GRC-MERVEX (CO de Vigo. IEO)

..- Intro

2.- Método científico y Método estadístico

Planteamiento del problema

l modelo stadístico

La informació

DA. Análisis Exploratorio de Datos

stimación de trámetros

ontraste de

3.-Diferenciación

stadística escriptiva vs iferencial

Fundamentos en Estadística

El modelo

Contraste de

3.-

Estadística

3.- Diferenciación / Clasificación

Fundamentos en Estadística

V. Trujillo

GRC-MERVEX (CO de Vigo. IEO)

.- Intro

2.- Método científico y Método

Planteamiento del problema

El modelo estadístico

La información

EDA. Análisis Exploratorio de

Exploratorio o Datos Estimación de

Parámetros Contraste de

Crítica y Aná del modelo

3.-

Diferenciación / Clasificación

Estadística descriptiva vs inferencial

Estadística paramétrica vs no-paramétrica La estadística se puede clasificar en relación a varios conceptos:

Estadística descriptiva vs inferencial

La estadística **descriptiva** realiza una exploración de los datos sin tomar decisiones respecto a los valores de los parámetros en la población, se refiere a la construcción de tablas, gráficos, índices etc. En cambio, la estadística **inferencial** aplica los resultados de la estadística descriptiva sobre la población, estimando los valores de los parámetros poblacionales.

Fundamentos en Estadística

V. Trujillo

GRC-MERVEX (CO de Vigo. IEO)

.- Intro

2.- Método científico y Método

Planteamiento del

l modelo

La información

EDA. Análisis Exploratorio de Datos

Estimación de Parámetros

Parámetros Contraste de

implificación Crítica y Aná

del modelo

3.Diferenciación
/ Clasificación

Estadística paramétrica vs no-paramétrica

Estadística paramétrica vs no-paramétrica

La estadística paramétrica tiene en cuenta la distribución de los datos, estimando parámetros como la media (μ) y la varianza paramétrica (σ^2) .

La estadística **no-paramétrica** no asume ningún tipo de distribución de los datos, es de distribución libre.

Podemos tener una estadística inferencial tanto paramétrica como no-paramétrica, ya que ambas hacen contraste de hipótesis.

Fundamentos en Estadística

3 -

Métodos no-paramétricos o de distribución libre.

Condiciones para aplicar este tipo de métodos:

- Cuando los datos son enumerativos o representan el número de observaciones en una categoría. Ej.: frecuencias en escala de medida débil.
- ▶ Nivel de medida ordinal. Comparación de orden.
- ► Cuando la escala de medida es fuerte, pero **no** nos interesan los parámetros de la distribución poblacional
- ▶ No se asume que los parámetros de la población sean por ej.: normales. No nos planteamos que la distribución pueda ser normal o de otro tipo de distribución.

Fundamentos en Estadística

V. Trujillo

GRC-MERVEX (CO de Vigo. IEO)

1.- Intro

2.- Método científico y Método estadístico

Planteamiento del problema

modelo tadístico

La información

DA. Análisis Exploratorio de Oatos

Estimación d Parámetros

Parámetros Contraste de

implificación Crítica y Anális

del modelo

3.Diferenciación / Clasificación

descriptiva vs inferencial

Estadística no-paramétrica vs paramétrica:

- Aparato matemático simple. Los no-paramétricos son más simples.
- ► Facilidad de aplicación. Son más fáciles.
- Rápidos de aplicar. Igual de rápidos a pesar de que a veces necesitan más tiempo para la ordenación.
- Eficiencia estadística. Son <u>menos eficientes</u>, en principio, que los paramétricos.
- Niveles de significación. Menor significación, en principio, que los paramétricos.
- Aplicación a problemas reales. Es aplicable siempre, mayor aplicabilidad que paramétricos.
- ▶ Tamaño de la muestra. Se pueden usar en muestras reducidas.
- Robustez de la prueba. Mayor que los paramétricos. Un dato que sea dispar, en una técnica no-paramétrica es muchas veces un dato más, pero en cambio en una paramétrica puede verse muy afectada, ej.: los "outliers" o valores anómalos⁷.

Fundamentos en Estadística

V. Trujillo

GRC-MERVEX (CO de Vigo. IEO)

..- Intro

2.- Método científico y Método

Planteamiento del

modelo tadístico

La información

EDA. Análisis Exploratorio d Datos

stimación de

Parámetros

Simplificación

Crítica y Análisis del modelo

3.Diferenciación

Estadística descriptiva vs inferencial

Ejemplo de eficacia entre métodos para y no-paramétricos

No-paramétricos	Paramétricos	ERA (Eficacia Relativa Asintótica)
Prueba de los signos	z o t de Student	0.637
Prueba de Wilcoxon	z o t de Student	0.955 (no detectan el 4% de las diferencias \cong a paramétricos)
Prueba de Friedman χ^2	F de Snedecor	0.955
U de Mann-Whitney Kruskal-Wallis	z o t de Student	

Fundamentos en Estadística

7. Trujillo

GRC-MERVEX (CO de Vigo. IEO)

.- Intro

- Metodo entífico y létodo stadístico clanteamiento del roblema

El modelo estadístico

> DA. Análisis xploratorio de

> > imación de ámetros

arámetros ontraste de

Contraste de Simplificación

del modelo

3.Diferenciación / Clasificación

descriptiva vs inferencial

Estadística univariante vs multivariante

Dependiendo del número de variables que se estudian en conjunto, la estadística se divide en **univariante**, si estudia la relación entre los casos de una o dos variables. La estadística **multivariante** entra en juego cuando hay tres o más variables, teniendo en cuenta la interacción entre ellas y los casos.

Fundamentos en Estadística

V. Trujillo

GRC-MERVEX (CO de Vigo. IEO)

1.- Intro

- 2.- Método científico y Método estadístico
- Planteamiento del
 - El modelo
 - tadistico información
 - DA. Análisis
 - estimación de
 - Parámetros
 - Contraste de
- Crítica y Análi del modelo
- 3.Diferenciación

Estadística descriptiva vs inferencial

paramétrica vs

	Una Muestra	Dos o más Muestras
Un conjunto de variables	ı	п
	Análisis de Componentes Principales (ACP)	Análisis discriminante/clasificación
		MANOVA
		Análisis canónico (poblaciones)
	Análisis Factorial	Análisis factorial de correspondencias
		Métodos Bi-Plot
		MSD (escalamiento multidimensional)
Dos o más conjuntos de variables	III	IV
	Regresión múltiple	
	Correlación múltiple	ANCOVA multivariado
	Correlación canónica	
	Correlación parcial y múltiple	

Fundamentos en Estadística

V. Trujillo

RC-MERVEX (CO de Vigo. IEO)

- Intro

- Método entífico y létodo stadístico

problema
El modelo

adístico

DA A-41:

DA. Análisis xploratorio d atos

Estimación de Parámetros Contraste de

> plificación tica y Análi

3.-Diferenciación

Estadística descriptiva vs inferencial Estadística En el segundo cuadrante (II) el Análisis Discriminante se usa en taxonomía numérica, que nos da una línea de corte entre especies, o también en medicina para la toma de decisión entre parto o cesárea. El MANOVA es semejante al ANOVA, o sea analiza la varianza multivariante (basándose en la distancia de Mahalanobis). El Análisis Canónico se usa para diferenciar poblaciones, el Análisis

En el cuadrante III por ej.: la Correlación Canónica se usa en meteorología, donde los conjuntos de variables son por ej. los días, las zonas etc.

Factorial de Correspondencias mejora al ACP.

Fundamentos en Estadística

V. Trujillo

GRC-MERVEX (CO de Vigo. IEO)

.- Intro

2.- Método científico y Método estadístico

Planteamiento del problema

El modelo

La informac

DA. Análisis xploratorio d

stimación de rámetros

arámetros

Contraste de Implificació

Crítica y Aná lel modelo

3.-Diferenciación

> stadística escriptiva vs ferencial

Estadística paramétrica vs



Fundamentos en Estadística

V. Trujillo

GRC-MERVEX (CO de Vigo. IEO)

1.- Intr

2.- Método científico y Método estadístico

Planteamiento del problema

El modelo estadístico

estadistico

EDA Análisis

EDA. Análisis Exploratorio d

Estimación Parámetros

Contraste de

Simplificación Crítica y Anál del modelo

3.-Diferenciación

Estadística descriptiva vs inferencial

Estadística paramétrica y