

IN3401: Estadística para Economía y Gestión
Semestre Otoño 2012

Temario General

El curso cubre conceptos esenciales de inferencia estadística y métodos de estimación. La primera parte del curso se enfoca a introducir conceptos de estadística tales como: (i) muestreo; (ii) estimadores y sus propiedades básicas; (iii) intervalos de confianza y distintas técnicas para formular estos; (iv) el marco conceptual de los test de hipótesis. La segunda parte del curso cubre test clásicos para analizar muestras, con especial énfasis en el análisis de encuestas y datos experimentales/cuasi-experimentales. La tercera parte del curso cubre regresión lineal simple y multivariada. La cuarta parte del curso cubre estimación paramétrica, analizando en detalle el método de Estimación de Máxima Verosimilitud (EMV). Se estudian aplicaciones de EMV, incluyendo regresión logística y modelos probabilísticos de clientes. La última parte del curso cubre una selección de métodos multivariados: análisis de componentes principales y análisis de factores.

Las cátedras del curso cubren aspectos teóricos de estadística, aspectos técnicos de metodologías de estimación y aplicaciones prácticas de estos conceptos y métodos usando datos de problemas reales. Las aplicaciones son cubiertas mediante ejemplos en cátedra, ejercicios en clase auxiliar, mini-tareas a ser evaluadas como nota de participación, y tareas computacionales más complejas. Se utilizará *Excel* y el software estadístico *Stata* a través de todo el curso.

Evaluación

La evaluación del curso está constituida por tres partes:

- **Pruebas escritas (45 %)**: un control a mitad de semestre (20 %) y un examen final (25 %)
- **Tareas computacionales (40 %)**: cuatro tareas donde los alumnos trabajan de manera individual.
- **Participación (15 %)**: se evaluará participación en clase y notas de varias mini-tareas a ser entregadas durante el semestre.

Los ítems 1 y 2 son reprobatorios – nota inferior en 4.0 en pruebas o tareas es causal de reprobación.

Acuerdo ético

El desarrollo de cursos con énfasis práctico y aplicado requiere un esfuerzo significativo en la preparación del material docente. En particular, en este curso hemos hecho el esfuerzo de utilizar ejemplos y tareas basados en datos reales en aplicaciones relevantes. El aprendizaje de estadística requiere que los alumnos “se ensucien las manos” trabajando con datos, por lo cual las tareas forman una parte fundamental del curso. Es por eso que exigimos que las tareas se hagan de forma individual, y la copia de tareas es considerada una falta gravísima y causal de reprobación – al igual que copia en pruebas escritas. Sin embargo, es difícil verificar la autenticidad de las tareas, por lo que resulta fundamental que los alumnos adquieran un compromiso ético de cumplir con las reglas establecidas en el curso. En cada evaluación y tarea, se incluirá un párrafo que ratifica este compromiso ético que cada alumno deberá firmar.

Temario en Detalle

Apunte 1: Introducción a Estadística

En estadística, típicamente queremos analizar ciertas características desconocidas de una *población* definida. La *inferencia estadística* se enfoca en *estimar* estas características utilizando una *muestra* de observaciones de la población de interés. Esta estimación tiene un error asociado, el cual nos interesa definir y medir, y tomarlo en cuenta a la hora de reportar nuestro análisis. Este capítulo define algunos conceptos generales relevantes en la inferencia estadística.

1. Muestras aleatorias
2. Simulación de variables aleatorias
3. Estimadores: Inferencia de la media de una población a partir de una muestra.
4. Error muestral
5. Propiedades de un estimador: insesgado, varianza y error estándar.
6. Intervalos de Confianza
7. Muestras aleatorias de una población con distribución Normal:
 - a) Derivación de la distribución t-student y chi-cuadrado.
 - b) Intervalos de confianza exactos para la media de la población.

Aplicaciones: (1) Obteniendo una muestra aleatorias de teléfonos de la guía; (2) Simulando la distribución del promedio muestral;

Apunte 2: Teoremas Limite y Aproximaciones Asintóticas

En el capítulo anterior vimos que el análisis exacto de los error muestrales e intervalos de confianza puede ser engorroso y en algunos casos imposible de realizar. En este capítulo estudiamos como usar los *teoremas límites* vistos en el curso de probabilidades para aproximar el error muestral y distribución de un estimador.

1. Tipos de convergencia
2. Estimadores consistentes
3. Consistencia del promedio muestral: Teorema de los Grandes Números
4. Distribución del promedio muestral: Teorema Central del Límite (TCL)
5. Aproximación Normal de la Distribución Binomial: aplicación a análisis de encuestas.
6. Otras herramientas útiles basadas en resultados asintóticos: Slutsky y el Método Delta.

Aplicaciones: (1) Simulando la convergencia del TCL; (2) Análisis de encuestas usando la aproximación normal; (3)

Apunte 3: Metodos de Bootstrapping

En este capítulo veremos como construir intervalos de confianza para estimar el error muestral de un estimador a través de un método de simulación de muestras.

1. Bootstrap paramétrico
2. Bootstrap no paramétrico
3. Ejemplos en Excel y Stata.

Aplicacion: Intervalo de confianza para la duracion promedio de una cirugia.

Apunte 4: Test de Hipótesis

Un aspecto fundamental del metodo científico se basa en la formulacion de hipotesis y la validacion empirica de estas hipotesis. En esta parte del curso estudiamos el marco conceptual de los *test de hipotesis* basados en el analisis estadístico de una muestra de datos.

1. Definicion de los errores de Tipo I y Tipo II en un test.
2. Conceptos Generales de un Test de Hipótesis: hipotesis nula y alternativa, estadístico del test, distribución nula, región de rechazo, nivel de significancia y p-valor.
3. Dualidad entre Intervalo de Confianza y Test de Hipótesis

Aplicacion: (1) Control estadístico de procesos; (2) Testeando poderes extra-sensoriales.

Apunte 5: Comparando Muestras

En el capítulo anterior aprendimos el marco conceptual de los test de hipótesis. En esta seccion estudiamos tests usados en la practica para comparar muestras de distintas poblaciones.

1. Test de medias y test de proporciones
2. Test de una cola vs Test de dos colas
3. Test de varianza
4. Tests no-paramétricos

Aplicacion: Usando encuestas para testear hipotesis.

Apunte 6: Análisis de Varianza (ANOVA)

En el capítulo anterior vimos cómo comparar los promedios de dos muestras para validar si las medias de dos poblaciones son iguales o distintas. El analisis de varianza (ANOVA) es una metodologia que permite comparar las *medias* de varias poblaciones (mas de dos).

1. Descomposición de varianza entre grupos e intra grupos.
2. Test-F de igualdad de varianza en ANOVA con un factor.

3. ANOVA en grupos no-balanceados
4. Analizando dos o mas factores.

Minitarea: Aplicacion de ANOVA en experimentos de campo en retail.

Apunte 7: La Función de Esperanza Condicional (FEC)

ANOVA permite comparar las medias entre un numero discreto de poblaciones. La regresión lineal extiende esta idea definiendo una población mediante un conjunto finito de variables observables, denominadas variables independientes o explicativas (denotadas por X). Estas variables explicativas pueden ser utiles para *predecir* el valor de una variable aleatoria de interes, la cual denominamos la variable dependiente (denotada por y). Este capitulo presenta el marco conceptual que fundamenta el uso de la “esperanza condicional de y dado X ”, $E(y|X)$ – la cual llamamos la Funcion de Esperanza Condicional – como predictor de la variable dependiente y .

1. Esperanza Condicional y la Ley de Esperanzas Iteradas
2. La descomposición de una variable aleatoria en su Funcion de Esperanza Condicional (FEC)
3. Propiedades de la FEC: El Teorema de ANOVA
4. FEC y Predicción
5. Aproximacion lineal de la FEC.

Apunte 8: Regresión Lineal

En este capítulo profundizaremos en el método de regresión lineal, el cual modela la relación que existe entre una variable aleatoria dependiente y otras variables dependientes. Primero, presentamos el metodo de Minimos Cuadrados Ordinarios (MCO) y sus propiedades basicas. Luego estudiamos la relacion entre MCO, el modelo de regresion lineal y la FEC, los fundamentos que validan el modelo y cómo interpretar los resultados. Además aprenderemos a realizar analisis de regresion lineal usando Excel y Stata con ejemplos practicos.

1. El modelo de Regresión Lineal Simple
2. La Mecánica de Mínimos Cuadrados Ordinarios y sus propiedad algebraicas
3. Coeficiente de Determinación R^2
4. Propiedades Estadísticas de MCO bajo distintos supuestos.
5. El efecto “Regresión a la Media”
6. MCO con multiples (mas de una) variables explicativas: Representación matricial de la regresion lineal multivariada.
7. Comparacion entre regresion lineal simple y multivariada.
8. Errores estandar e Intervalos de confianza: metodos exactos, asintoticos y bootstrapping.
9. Test de hipotesis en regresion lineal: test-t de coeficientes, test-F de restricciones sobre multiples coeficientes.

Aplicaciones:(1) Encuesta de alturas entre padres e hijos; (2) La Regresion de Bordeaux

Apunte 9: Temas adicionales sobre Regresión

En esta seccion nos centramos en varios temas relacionados a la implementacion de regresion lineal en la practica tales como:

1. Reportando e interpretando resultados de la regresión
2. ISignificancia estadística vs significancia económica de los coeficientes
3. Definiendo la forma funcional de la regression
4. Interacciones entre variables explicativas
5. Selección del modelos
6. Usando la regresion para hacer predicción
7. Validando los supuestos de la regresion lineal: e.g. Multicolinealidad, Heterocedasticidad
8. Identificando Outliers y Puntos de Influencia

Aplicacion: Prediciendo el precio de casas.

Apunte 10: Introducción a la Estimación Paramétrica

La regresion lineal permite analizar factores que influyen la media de una poblacion y generalmente funciona bien prediciendo valores promedios dentro del rango observado en la muestra. Sin embargo, muchas veces nos interesa predecir otros estadísticos de la poblacion mas alla de la media (e.g. percentiles), o extrapolar predicciones a rangos no observados en la muestra. En algunos de estos casos la regresion lineal no es suficiente. En esta seccion analizaremos *modelos parametricos*, en donde se haremos supuestos mas fuertes, asumiendo la distribucion parametrica especifica que caracteriza completamente a una poblacion (no solo su media). Esta mayor estructura permite, en algunos casos, ajustar mejor los datos mas alla de la media y extrapolar fuera de la muestra.

1. Evaluando un modelo estadístico: predicción dentro y fuera de la muestra
2. Introducción a los modelos paramétricos
3. Método de los momentos
4. Intervalos de Confianza en Estimación Paramétrica: distribución exacta, distribución asintótica, métodos de bootstrap

Aplicacion: Estimando la tasa de retencion de clientes.

Apunte 11: Estimación de Máxima Verosimilitud

Maxima verosimilitud es el metodo mas utilizado para estimar modelos parametricos. Analizaremos la implementacion de estos metodos y sus propiedades.

1. Definición del Estimador de Máxima Verosimilitud (EMV)
2. La logica detras de EMV
3. Distribución asintótica de EMV
4. Varianza y Errores Estandar de EMV

Aplicacion: Analizando datos de visitas a un sitio web.

Apunte 12: Temas Adicionales de Máxima Verosimilitud

En este capítulo analizamos una serie de resultados teóricos que muestran las ventajas del EMV. También estudiamos cómo hacer test de hipótesis con el EMV para seleccionar entre distintos modelos.

1. Error Cuadrático Medio
2. Trade-off entre dispersión y sesgo
3. Eficiencia del EMV: la cota inferior de Cramer-Rao
4. Test de Hipótesis con Máxima verosimilitud: test de ratios de verosimilitud
5. Propiedades de los tests de ratios de verosimilitud: Lema de Neyman-Pearson
6. Test de ratios de verosimilitud generalizados
7. Test Chi-cuadrado de Bondad de Ajuste
8. Otros tests para validar distribuciones paramétricas

Apunte 13: Variables Dependientes Binarias

En este capítulo estudiamos cómo hacer análisis similares a la regresión lineal cuando la variable dependiente es binaria.

1. Modelo Lineal de Probabilidad
2. El modelo Logit: estimación vía Máxima Verosimilitud.
3. Test de Hipótesis
4. Interpretando la estimación del Logit
5. Evaluando la bondad de ajuste del modelo

Aplicación: Analizando factores que influyen en la tasa de atención de clientes en una tienda de retail.

Apunte 14: Análisis Multivariado – Componentes Principales

Tipicamente los resultados de un experimento o encuesta contienen muchas variables de medición que presentan una alta colinealidad entre ellas, lo cual dificulta el análisis y la interpretación de resultados. En esta sección estudiamos métodos de análisis multivariado que permiten reducir la alta dimensionalidad de estos datos a un grupo selecto de nuevas variables que resumen de mejor manera la información contenida en los datos originales.

1. Generalidades de Análisis Multivariado
2. Repaso Álgebra Lineal
3. Análisis de Componentes Principales (ACP)
4. Interpretando soluciones de ACP

Aplicaciones: (1) Mapas poblacionales genéticos; (2) Análisis del producto bruto de regiones.

Apunte 15: Análisis Factorial Exploratorio

Al igual que ACP, el Análisis Factorial Exploratorio (AFE) es una técnica de modelamiento que intenta explicar correlaciones entre un conjunto de variables observadas y reducir la complejidad de estas a un número reducido de *factores*. Mientras que ACP busca reducir al máximo la varianza, AFE busca factores *comunes* que expliquen la varianza en los datos originales que puedan tener una interpretación práctica y intuitiva. AFE tiene múltiples aplicaciones, dentro de las cuales destaca el análisis de encuestas de percepción de clientes y la identificación de distintas habilidades cognitivas en las respuestas de un test.

1. Análisis Factorial Exploratorio (AFE)
2. Modelos con uno o más factores
3. Metodología analítica del AFE
4. Interpretación del AFE: métodos de rotación

Aplicaciones: (1) Modelando la inteligencia humana mediante tests; (2) Preferencias de consumidores por analgésicos.