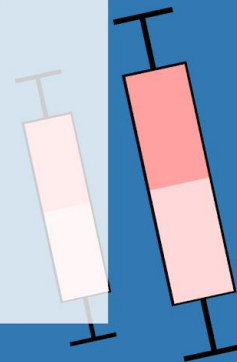


Herramientas Estadísticas y Forecast

Clase 7: Introducción a los modelos estadísticos

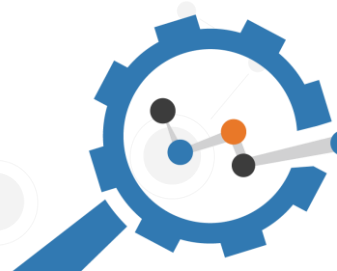


FACULTAD DE MATEMÁTICAS
PONTIFICIA UNIVERSIDAD
CATÓLICA DE CHILE



Contenido del curso

1. Introducción a la estadística y análisis descriptivo
2. Análisis descriptivo y gráfico
3. Probabilidad y Distribuciones
4. Muestreo
5. Inferencia estadística: Pruebas de hipótesis
6. Taller Práctico de inferencia
7. **Introducción a los modelos estadísticos**
8. Modelos predictivos I: Modelos de regresión lineal.
9. Modelos predictivos II: Regresión logística y otros modelos.
10. Modelos de Forecasting I.
11. Modelos de Forecasting II
12. Taller de Forecast



Modelos ANOVA

El análisis de varianza (ANOVA) se emplea cuando se desean comparar **k** muestras diferentes a través de un test de hipótesis.

En este caso, el estudio está enfocado que una gran población es subdividida en **k** grupos, a partir de un **factor** (o más) que los diferencia.

En general, se dice que un modelo ANOVA tendrá varios factores cuando una variable se pueda describir a través de una suma de promedios que representan a diversos grupos.

Modelos ANOVA

Un modelo **ANOVA con un factor** es una prueba de hipótesis donde se consulta si las medias de k grupos son iguales o distintas.

Se debe cumplir que:

- La variable posee distribución Normal,
- Las k muestras son independientes
- Los grupos poseen homocedasticidad

Modelos ANOVA

Un modelo ANOVA realiza el análisis a la variabilidad total de los datos, y la compara con la variabilidad de cada grupo, realizando las pruebas necesarias para determinar si el factor tiene algún tipo de efecto.

Se denomina **varianza intra-grupal (SCE)** a la dispersión de cada muestra con respecto a su propias medias.

Se denomina **varianza inter-grupal (SCR)** a la dispersión de las medias de cada muestra respecto a la media global.

Modelos ANOVA

Usualmente se reportará una tabla de la siguiente forma:

	Grados de Libertad	Suma de Cuadrados	Media Cuadrática	Prueba de Hipótesis F	Valor-p
Inter-grupos (Factor)	$k - 1$	SCR	$SCR / (k - 1)$	MCR/MCE	
Intra-grupos (Error)	$n - k$	SCE	$SCE / (n - k)$		
Total	$n - 1$	$SCT = SCR + SCE$	-	-	-

La función **aov(Y~Factor)** realiza el análisis de varianza. La función **summary()** de lo anterior entrega la tabla ANOVA.

Modelos ANOVA

Suponga que tres sucursales reportan las ventas en cientos de miles de sus ejecutivos (descargue los datos [aquí](#))

	Grados de Libertad	Suma de Cuadrados	Media Cuadrática	Prueba de Hipótesis F	Valor-p
Inter-grupos: Sucursal	2	111	55.45	3.314	0.0369
Intra-grupos: Error	747	12499	16.73		
Total	749		-	-	-

En este caso, con una significancia del 5% se rechaza la hipótesis nula, entonces si hay diferencia de medias en ventas entre las sucursales.

Actividad

Utilizando los datos de la base de datos PlayStore (Descargue los datos [aquí](#))

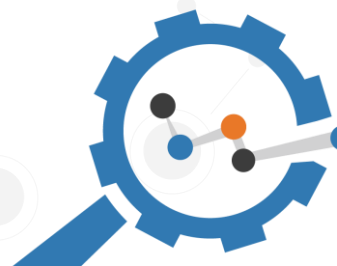
- I. Suponga que la variable TAMAÑO(M) sigue una distribución Normal. Muestre si existe una diferencia del promedio del TAMAÑO(M) según el tipo de la app (PAGADO vs GRATIS), Luego, pruebe muestre si existe diferencia de los promedio según la clasificación de las apps.
- II. Suponga que la valoración sigue una distribución Normal. Realice un test de hipótesis para determinar si existe efecto de la clasificación de las apps para determinar la valoración de estas.
- III. Proponga un gráfico para cada situación anterior y comente.

Modelos ANOVA

Un modelo **ANOVA con dos o más factores** se dan cuando, efectivamente, una variable se intenta proyectar a través de diferentes tratamientos.

En este caso, se debe considerar dos posibles métodos:

- Modelos aditivos
- Modelos con interacción



Modelos ANOVA

Un modelo **ANOVA con k factores aditivos** se dispone como un método de describir una variable en términos de las medias de los k factores. En el caso simple que fuera dos factores, la igualdad se da de la forma:

$$y_{ijk} = \mu + \alpha_i + \beta_j + \epsilon_{ijk}$$

Variable a
describir

Media global

Media del primer
factor

Media del segundo
factor

Error aleatorio

Modelos ANOVA

Un modelo **ANOVA con k factores con interacción** se dispone como un método de describir una variable en términos de las medias de los k factores, junto con la interacción entre ellos.

En el caso simple que fuera dos factores, la igualdad se da de la forma:

$$y_{ijk} = \mu + \alpha_i + \beta_j + \gamma_{ij} + \epsilon_{ijk}$$

Variable a describir

Media global

Media del primer factor

Media del segundo factor

Interacción de Factores

Error aleatorio

Modelos ANOVA

Suponga que tres sucursales, reportan las ventas de sus ejecutivos, y además se considera el sexo.

	Grados de Libertad	Suma de Cuadrados	Media Cuadrática	Prueba de Hipótesis F	Valor-p
Sucursal	2	111	55.45	3.330	0.0363
Sexo	1	13	13.14	0.789	0.3746
Sucursal:Sexo	2	97	48.48	2.912	0.0550
Errores	744	12389	16.65		
Total	49				

Notar que con una significancia del 5%, la interacción no es significativa, así como el sexo por si solo.

Modelos ANOVA

Desde la perspectiva anterior, un modelo ANOVA permite diseñar una descripción de la variable de interés en términos de otras variables de categoría o segmentación.

Actividad

Utilizando los datos de PlayStore

IV). Realice un modelo ANOVA para determinar si en la VALORACIÓN hay efecto conjunto entre CLASIFICACIÓN y PAGO.

V). ¿Podría considerar la CATEGORÍA como un nuevo factor?

