

Introducción al Diseño de Experimentos

Clase 2

Nicolás Mejía M.
n.mejia10@uniandes.edu.co

Probabilidad y Estadística II
Departamento de Ingeniería Industrial
Universidad de Los Andes, Bogotá, Colombia

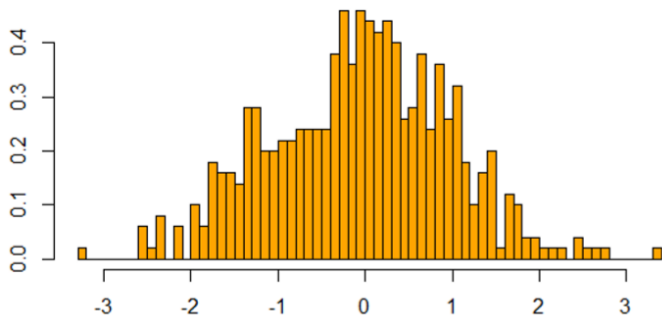
2020-19

Outline

- 1 La Variación en los Datos
 - Fuentes de Variación
 - Tipos de Datos
- 2 Diseño de Experimentos
 - Contexto
 - Objetivos
 - Notación
 - Prueba ANOVA
- 3 ANOVA de 1 Factor

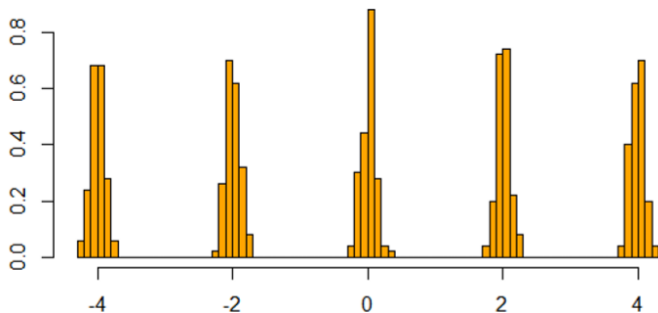
Fuentes de Variación

Cuando observamos realizaciones de un fenómeno aleatorio, usualmente atribuimos la **variación** en los valores que se dan a únicamente ese componente aleatorio.



Fuentes de Variación

Pero esas variaciones pueden ser producto de otros factores, y no exclusivamente del ruido aleatorio. Un caso puede ser una estructura de grupos definida, donde se tiene que el comportamiento cambia según la categoría correspondiente.



Fuentes de Variación

Ejemplo

Considere la estatura de una persona seleccionada de forma aleatoria. La estatura promedio es diferente si hacemos distinción entre hombres y mujeres. También sería diferente si hacemos distinción entre grupos de edades, etc.

De esta forma, se tiene que la variación que observamos en los datos realmente se puede **descomponer** en varias fuentes:

Variación Total = Var. por efecto de grupos + Var. por efecto aleatorio

Fuentes de Variación

En el ejemplo que vimos, sabíamos de antemano que la variable cambia su comportamiento según la categoría del grupo, caso que no es interesante. Lo que si es interesante, es determinar si dada una estructura de grupos para nuestra variable de interés, es necesario hacer una distinción entre las categorías.

Ejemplo

Se cree que la estatura promedio difiere según la estación del año en que se nace (invierno, primavera, verano, otoño).

Aquí las respuestas dejan de ser directas. Nuestro objetivo es contestar esas preguntas utilizando procedimientos de **inferencia estadística**.

Outline

1 La Variación en los Datos

- Fuentes de Variación
- Tipos de Datos

2 Diseño de Experimentos

- Contexto
- Objetivos
- Notación
- Prueba ANOVA

3 ANOVA de 1 Factor

Tipos de Datos

A la hora de realizar uno de estos análisis, es importante recordar con que tipo de datos se está trabajando:

- **Observacionales:** En los estudios observacionales no se manipulan las variables para tratar de producir resultados. En su lugar, simplemente se mira lo que sucede naturalmente.
- **Experimentales:** Se controlan tantas variables como sea posible con el objetivo de determinar el efecto que tiene una variable. Se evita la influencia variable no controladas.

Dependiendo del tipo de datos, se podrá concluir con certeza que el cambio en el comportamiento de la variable de interés es producto de la estructura de grupos que se analiza.

Tipos de Datos

- **Observacionales:** Se desea saber si el desempeño en matemáticas cambia según el género. Para esto se utilizaron las notas de estudiantes de colegios puramente masculinos y femeninos.
- **Experimentales:** Se desea evaluar la diferencia en calidad de combustibles de motos. Para esto se mide el kilometraje recorrido con un galón de gasolina por motos bajo las mismas condiciones, difiriendo solo en el combustible.

Si en el primer ejemplo existen diferencias, no podemos atribuírselas directamente al género: podría pasar que en un colegio se haga más énfasis en matemáticas que en el otro, luego lo que afecta allí es la pedagogía y no el género. En el segundo ejemplo, si podemos atribuir las diferencias al factor dadas las condiciones de control.

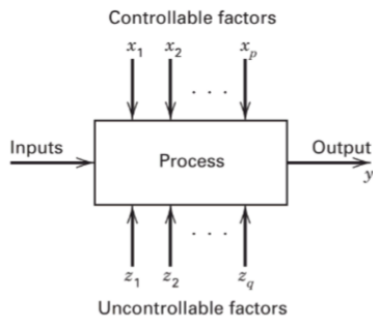
Durante el curso nos enfocarnos en **datos experimentales**, que es el caso donde hay control sobre lo que está afectando.

Outline

- 1 La Variación en los Datos
 - Fuentes de Variación
 - Tipos de Datos
- 2 Diseño de Experimentos
 - Contexto
 - Objetivos
 - Notación
 - Prueba ANOVA
- 3 ANOVA de 1 Factor

Contexto

Observar un sistema o un proceso en operación es una parte fundamental para entender su funcionamiento. Un proceso lo podemos entender como una combinación de operaciones que transforman insumos en productos, siendo estos últimos nuestras variables de interés.



Outline

- 1 La Variación en los Datos
 - Fuentes de Variación
 - Tipos de Datos
- 2 Diseño de Experimentos
 - Contexto
 - **Objetivos**
 - Notación
 - Prueba ANOVA
- 3 ANOVA de 1 Factor

Objetivos

En el caso en el que se quiere entender que sucede con el sistema cuando algunos factores de control cambian, es necesario hacer el cambio directamente en el sistema. En otras palabras, experimentar.

Experimento

Un experimento es un serie de pruebas en las que se hacen cambios deliberados a las variables *insumo* de un sistema para así identificar cambios en sus variables *producto*.

Outline

- 1 La Variación en los Datos
 - Fuentes de Variación
 - Tipos de Datos
- 2 Diseño de Experimentos
 - Contexto
 - Objetivos
 - **Notación**
 - Prueba ANOVA
- 3 ANOVA de 1 Factor

Notación

Trabajaremos con los siguientes conceptos:

- **Factor:** Es la variable que se supone afecta a nuestra Y (variable respuesta). En nuestro caso sería la estructura de grupos.
- **Niveles:** Corresponden a los diferentes valores que puede tomar el factor. En nuestro caso sería las distintas categorías de la agrupación, denotaremos a la cantidad de categorías como a .
- **Tratamientos:** Cuando hay varios factores en el estudio, los tratamientos corresponden a todas las posibles combinaciones de niveles entre los factores.
- **Réplicas:** Corresponde a la cantidad de veces que se repite el experimento bajo un tratamiento dado. Este número lo denotaremos como n_i y de manera informal es el número de datos por celda.

Objetivos

En un experimento estadístico existen dos preguntas fundamentales:

- 1 **¿Influye el factor sobre la variable de respuesta?** Generalmente es la pregunta científica de interés y puede ser entendida como la causalidad del cambio en la respuesta dado el cambio en el factor.
- 2 **¿Cuál es el tratamiento óptimo?** Si el factor si influye, una pregunta práctica es cuál debe ser seleccionado dado un interés particular.

En la primera parte del curso, nos dedicaremos a construir procedimientos estadísticamente correctos para responder la primera pregunta. En particular, desarrollaremos la prueba ANOVA.

Diseño y Análisis de Experimentos Estadísticos

Cuando se tiene **un solo factor** Y_{ij} representa la variable de interés para la j -ésima observación del i -ésimo nivel. Se asume que $Y_{ij} \sim \text{Normal}(\mu_i, \sigma^2)$.

Así mismo, se utilizará la siguiente notación para los **promedios**:

$$N = \sum_{i=1}^a n_i$$

$$\bar{Y}_{..} = \frac{1}{N} \sum_{i=1}^a \sum_{j=1}^{n_i} Y_{ij}$$

$$\bar{Y}_{i.} = \hat{\mu}_i = \frac{1}{n_i} \sum_{j=1}^{n_i} Y_{ij}, \quad i \in \{1, \dots, a\}$$

Diseño y Análisis de Experimentos Estadísticos

Si ubicamos los datos en una tabla, se tendría lo siguiente:

	Factor A				
Obs (j)	Nivel 1	...	Nivel i	...	Nivel a
1	Y_{11}	...	Y_{i1}	...	Y_{a1}
2	Y_{12}	...	Y_{i2}	...	Y_{a2}
\vdots	\vdots	\vdots	\vdots	\vdots	\vdots
j	Y_{1j}	...	Y_{ij}	...	Y_{aj}
\vdots	\vdots	\vdots	\vdots	\vdots	\vdots
n	Y_{1n_1}	...	Y_{in_i}	...	Y_{an_a}
Promedios columna	$\bar{Y}_{1.}$...	$\bar{Y}_{i.}$...	$\bar{Y}_{a.}$
Desv. Est.	S_1	...	S_i	...	S_a

Outline

- 1 La Variación en los Datos
 - Fuentes de Variación
 - Tipos de Datos
- 2 Diseño de Experimentos
 - Contexto
 - Objetivos
 - Notación
 - Prueba ANOVA
- 3 ANOVA de 1 Factor

Prueba ANOVA (*Analysis of Variance*)

La pregunta de si el factor influye o no sobre a variable de respuesta puede ser puesta como una prueba estadística con la siguiente prueba de hipótesis:

H_0 : El Factor NO influye sobre $Y \Leftrightarrow \mu_1 = \mu_2 = \dots = \mu_a$

H_1 : El Factor SI influye sobre $Y \Leftrightarrow$ Algún par $\mu_i \neq \mu_j$

En el caso en que $a = 2$, y dado los supuestos de normalidad y varianzas iguales que se han hecho, la prueba se puede resolver con el procedimiento para **diferencia de medias** (prueba t).

Sin embargo, cuando se tienen más de dos tratamientos (niveles), se debe desarrollar un nuevo procedimiento.

Outline

- 1 La Variación en los Datos
 - Fuentes de Variación
 - Tipos de Datos
- 2 Diseño de Experimentos
 - Contexto
 - Objetivos
 - Notación
 - Prueba ANOVA
- 3 ANOVA de 1 Factor

Caso Particular de 2 Niveles

Veamos un caso particular. Supongamos que $a = 2$, es decir, el factor tiene 2 niveles. En ese caso nuestra tabla queda como:

	Factor A	
Obs (j)	Nivel 1	Nivel 2
1	Y_{11}	Y_{21}
\vdots	\vdots	\vdots
j	Y_{1j}	Y_{2j}
\vdots	\vdots	\vdots
n	Y_{1n_1}	Y_{2n_2}
Promedios columna	$\bar{Y}_{1.}$	$\bar{Y}_{2.}$
Desv. Est.	S_1	S_2

La hipótesis asociada a la influencia del factor queda como:

$$H_0 : \mu_1 = \mu_2 \text{ vs } H_1 : \mu_1 \neq \mu_2$$

Ejemplo

Se realizó un experimento para determinar la influencia del fármaco sobre la concentración andrógenos en la sangre. Se obtuvieron muestras de sangre de venados inmediatamente antes de recibir una inyección intramuscular. Treinta minutos después se obtuvo una segunda muestra de sangre y después los venados fueron liberados.

Antes	Después
2.76	7.02
1.18	4.12
2.68	5.44
8.1	5.21
7.05	9.26
6.6	8.91
8.79	9.53
8.79	9.53
7.3	6.85

2 Niveles Pareados

En la prueba anterior asumimos que las muestras eran independientes, y es un supuesto que haremos a menos que sea claro lo contrario. En el caso en que exista dependencia, podemos crear una variable auxiliar asociada a la diferencia:

$$d_j = Y_{1j} - Y_{2j}$$

De esta forma la prueba de hipótesis se reduce a probar:

$$H_0 : \mu_d = \mu_1 - \mu_2 = 0 \text{ vs } H_1 : \mu_d \neq 0$$

Este caso, la prueba de hipótesis corresponde a probar una media de una población normal.

$$t = \frac{\bar{d}}{\sqrt{\frac{\hat{\sigma}_d^2}{n}}} \sim t_{n-1}$$

Reconsidere el ejemplo visto. ¿Considera que la prueba pareada es más recomendable?