# **ID 033521 Análisis Multivariado**

Créditos: 3

Prerrequisitos: Inferencia Estadística (ID 033519)

Modalidad: Presencial, 4 horas semanales.

Componente: Teórico

Descripción de la asignatura

El análisis multivariado es el área de la estadística que tiene como objetivos resumir, representar e interpretar los datos muestrales provenientes de poblaciones en las cuales las unidades experimentales son evaluadas en diferentes variables. La justificación del investigador para medir diversas variables en cada unidad experimental es que probablemente, ninguna de las variables individualmente puede caracterizar de forma adecuada la unidad o puede discriminar los individuos con relación a cualquier criterio que sea empleado. Esta asignatura extiende los conceptos de la inferencia estadística al caso de múltiples variables, abordando métodos estadísticos multivariados relacionados con la estructura de covarianza y presenta los conceptos básicos del análisis discriminante.  
  
Este curso busca resolver la pregunta: ¿Cómo analizar estadísticamente datos considerando varias variables simultáneamente?

**Objetivos de formación:**

1. Proporcionar a los alumnos los elementos necesarios para la adecuada utilización de métodos de análisis multivariado en datos de investigación.
2. Presentar metodologías multivariadas de análisis estadístico de datos.
3. Ilustrar el análisis de bases de datos y tomar de decisiones basadas en criterios estadísticos.
4. Presentar herramientas de software que permiten el uso de estadísticas multivariadas básicas y avanzadas en R.
5. Ilustrar la comunicación de los resultados y su interpretación.

**Resultados de aprendizaje esperado (RAE):**

Al finalizar el curso el estudiante estará en capacidad de:

1. Identificar los fundamentos de la inferencia estadística bajo el modelo normal multivariado.
2. Juzgar si se cumplen los supuestos para un método estadístico multivariado.
3. Estimar e interpretar los parámetros de un modelo multivariado utilizando software estadístico.
4. Construir, evaluar y comparar pruebas de hipótesis para coeficientes de correlación, vectores de medias, matrices de covarianzas.
5. Usar adecuadamente los métodos multivariados relacionados con la estructura de covarianza.
6. Organizar información utilizando algunos métodos de clasificación o discriminación.

**Contenidos temáticos**

1. Introducción. ¿Cómo describir datos multivariados? 1.1 Representación gráfica de datos multivariados. 1.2 Vectores aleatorios. 1.3 Algunos parámetros y estadísticas. 1.4 Distribuciones conjuntas, marginales y condicionales. 1.5 Panorama de los métodos multivariados.

2. Distribución normal multivariada y sus propiedades. ¿En qué se basa el análisis multivariado? 2.1 Función de densidad. 2.2 Distribuciones marginales y condicionales. 2.3 Distribución normal bivariada. 2.4 Pruebas de bondad y ajuste a la normal multivariada, transformaciones y detección de datos atípicos. 2.5 Correlación parcial y múltiple. 2.6 Inferencia sobre coeficientes de correlación.

3. Inferencia sobre vectores de medias. ¿Las afirmaciones sobre el vector de medias poblacional son verdaderas? 3.1 Estimación y propiedades del estimador. Distribución T2 de Hotelling. 3.2 Pruebas de hipótesis y regiones de confianza sobre el vector de medias. 3.3 Función de potencia y tamaños de muestra. 3.4 Distribuciones no centrales asociadas a la normal.

4. Inferencia sobre matrices de covarianzas y de correlaciones. ¿Las afirmaciones sobre las matrices de covarianzas y correlaciones son verdaderas? 4.1 Distribución de Wishart y sus propiedades. 4.2 Pruebas de hipótesis sobre matrices de covarianzas. 4.3 Pruebas de hipótesis sobre patrones específicos de matrices de correlaciones.

5. Análisis en componentes principales. ¿Cómo describir un conjunto de datos multivariado en términos de nuevas variables (componentes) no correlacionadas? 5.1 Objetivos del análisis en componentes principales. 5.2 Generación de componentes principales. 5.3 Criterios para la selección del número de componentes principales. 5.4 Interpretación geométrica del análisis en componentes principales.

6. Análisis de correlación canónica. ¿Cómo analizar relaciones multidimensionales entre múltiples variables independientes y múltiples variables dependientes? 6.1 Objetivos de la correlación canónica. 6.2 Modelo poblacional y su estimación muestral. 6.3 Interpretación geométrica del análisis de correlación canónica.

7. Análisis factorial. ¿Cómo explicar las correlaciones entre las variables observadas en términos de un número menor de variables no observadas (factores)? 7.1 Objetivos del análisis factorial. 7.2 Modelo y métodos de estimación. 7.3 Criterios para la selección del número de factores. 7.4 Rotación de factores.

8. Análisis discriminante. ¿Cómo establecer relaciones entre las variables, que clasifiquen individuos/objetos a grupos distintos? 8.1 Objetivos del análisis discriminante. 8.2 Reglas para discriminación de grupos. 8.3 Estimación de tasas de error de discriminación. 8.4 Discriminación bajo la distribución normal.

9. Análisis de conglomerados. ¿Cómo establecer grupos? 9.1 Objetivos del análisis de conglomerados. 9.2 Medidas de similitud y disimilitud. 9.3 Métodos gráficos útiles. 9.4 Métodos jerárquicos, métodos de uniones: vecino más cercano, más lejano, distancia promedio, varianza mínima de Ward. Métodos de divisiones. 9.5 Métodos no jerárquicos, K-medias.

|  |  |
| --- | --- |
| SEMANA | TEMAS |
| 1  15/07 – 20/07 | Introducción. Panorama de los métodos multivariados. Representación gráfica de datos multivariados. Vectores aleatorios. Distribuciones conjuntas, marginales y condicionales. Algunos parámetros y estadísticas. |
| 2  22/07 – 27/07 | Distribución normal multivariada y sus propiedades. Función de densidad. Distribuciones marginales y condicionales. Distribución normal bivariada. |
| 3  29/07 – 03/07 | Pruebas de bondad y ajuste a la normal multivariada, transformaciones y detección de datos atípicos. Correlación parcial y múltiple. Inferencia sobre coeficientes de correlación. |
| 4  05/08 – 10/08 | Inferencia sobre vectores de medias. Estimación y propiedades del estimador. Distribución T2 de Hotelling. Pruebas de hipótesis y regiones de confianza sobre el vector de medias. Función de potencia y tamaños de muestra. Distribuciones no centrales asociadas a la normal. |
| 5  12/08– 17/08 | Inferencia sobre matrices de covarianzas y de correlaciones. Distribución de Wishart y sus propiedades. Pruebas de hipótesis sobre matrices de covarianzas. Pruebas de hipótesis sobre patrones específicos de matrices de correlaciones. |
| 6  19/08 – 24/08 | PRIMER PARCIAL |
| 7  26/08 – 31/08 | Análisis en componentes principales. Objetivos del análisis en componentes principales. Generación de componentes principales. |
| 8  02/09 – 07/09 | Criterios para la selección del número de componentes principales. Interpretación geométrica del análisis en componentes principales. |
|  | SEMANA DE REFLEXIÓN |
| 9  16/09 – 21/09 | Análisis de correlación canónica. Objetivos de la correlación canónica. Modelo poblacional y su estimación muestral. Interpretación geométrica del análisis de correlación canónica. |
| 10  23/09 – 28/09 | Análisis factorial. Objetivos del análisis factorial. Modelo y métodos de estimación. |
| 11  30/09 – 05/10 | SEGUNDO PARCIAL |
| 12  07/10 – 12/10 | Criterios para la selección del número de factores. Rotación de factores. |
| 13  14/10 – 19/10 | Análisis discriminante. Objetivos del análisis discriminante. Reglas para discriminación de grupos. |
| 14  21/10 – 26/10 | Estimación de tasas de error de discriminación. Discriminación bajo la distribución normal. |
| 15  28/10 – 02/11 | Análisis de conglomerados. Objetivos del análisis de conglomerados. Medidas de similitud y disimilitud. Métodos gráficos útiles. |
| 16  04/11 – 09/11 | Métodos jerárquicos, métodos de uniones: vecino más cercano, más lejano, distancia promedio, varianza mínima de Ward. Métodos de divisiones. Métodos no jerárquicos, K-medias. |
| 17  11/11 – 16/11 | TERCER PARCIAL |

Estrategias pedagógicas

Durante el curso se utilizarán estrategias centradas en el aprendizaje activo del estudiante, entre otras están: clase magistral interactiva. El aprendizaje basado en problemas. El aprendizaje por proyectos. Sesiones de trabajo colaborativo. Uso del software R.

Evaluación

Las estrategias de evaluación están centradas en la valoración de los resultados de aprendizaje esperado de la asignatura; las cuales pueden ser formativas en las que se brindará retroalimentación por parte del profesor y los estudiantes, que suscitan la comprensión y construcción de conocimiento, y sumativas, las cuales incluyen porcentajes de evaluación con el fin de corroborar el logro de los aprendizajes y el desarrollo de las competencias en los estudiantes.

|  |  |  |
| --- | --- | --- |
| Componente | Semana | Porcentaje |
| Primer Parcial | 6 | 25% |
| Segundo Parcial | 11 | 25% |
| Tercer Parcial | 17 | 25% |
| Talleres, Tareas, Quices y Presentaciones, entre otras |  | 25% |

Observación: Las notas de Talleres, Trabajos y Quices se consolidan en una sola nota, es decir se promedian y el promedio obtenido es el que se reporta en el cuaderno de evaluación (con una ponderación de 25%) a más tardar en la semana 14.

Bibliografía

Requerido:

Johnson, R. and Wicher, D.W. Applied multivariate statistical analysis. 6th Edition. Prentice Hall.

Profundización:

* Mardia, K.V.; Kent, J.T. and Bibby, J.M. Multivariate analysis. Academic Press.
* Morrison, D.F. Multivariate statistical methods. Mc GRaw Hill.
* Peña, D. Análisis multivariado de datos. Mc Graw Hill.
* Rencher, A.C. Multivariate statistical inference and applications. John Wiley and Sons.
* Srivastava, M.S. Methods of multivariate statistics. John Wiley and Sons.