



MÉTODOS ESTADÍSTICOS PARA ANALIZAR DATOS MULTIVARIADOS EN CIENCIAS DE LA TIERRA

DOCENTE: Dra. Mabel Mena.

MODALIDAD: Virtual sincrónica.

FUNDAMENTACIÓN

Dentro de la rutina de trabajo profesional resulta normal el manejo de información, expresada en forma de datos y por lo general, este manejo requiere del empleo de procedimientos estadísticos. De lo anterior surge la necesidad de conocer los fundamentos básicos de las técnicas y métodos estadísticos que puedan emplearse en diferentes estudios, sus alcances y sus limitaciones. Esto es así, tanto si el profesional se encarga de implementar métodos estadísticos, interpretando y evaluando los resultados de los mismos como cuando se interactúa con estadísticos que se encargan de esta tarea.

OBJETIVO GENERAL

- Repasar conceptos básicos de estadística previamente adquiridos y comprender los fundamentos de los diferentes métodos estadísticos que se verán en el curso.
- Aprender a extraer la mayor cantidad posible de información de los muestreos disponibles.
- Adquirir conocimientos sobre técnicas y herramientas estadísticas que pueden emplearse sobre datos uni-, bi- o multivariantes para resolver problemas en ciencias naturales.
- Ser capaz de emplear técnicas y métodos estadísticos teniendo en cuenta el cumplimiento de sus supuestos y sus limitaciones.

OBJETIVOS PARTICULARES

- Diferenciar los distintos tipos de variables con las que pueden representarse los datos.
- Poder describir un conjunto de datos a partir de medidas estadísticas que resuman las probabilidades de los diferentes valores posibles. Poder elegir modelos que representen las distribuciones de probabilidades halladas empíricamente.
- Probar diferentes hipótesis sobre las distribuciones de frecuencia o de probabilidades de los datos.

CURSO DE POSTGRADO



- Buscar y caracterizar relaciones entre dos variables o entre varias variables y si es posible determinar funciones subyacentes a las que respondan estas relaciones, empleando métodos de correlación y de regresión.
- Analizar si las relaciones encontradas permiten agrupar los datos en diferentes conjuntos, con cierta afinidad interna, o bien si es posible separarlos o adjudicarlos a grupos previamente definidos.

CONTENIDOS MÍNIMOS

Descripción de conjuntos de datos mediante gráficos, tablas y medidas resúmenes. Distribución de frecuencias y modelos de distribuciones de probabilidad. Pruebas de hipótesis, confianza, significación, *valor-p*, errores. Métodos paramétricos y no paramétricos. Búsqueda de posibles relaciones entre dos o más variables, empleando métodos de correlación, regresión, agrupamiento, análisis discriminante: su implementación para variables continuas, discretas, nominales y composicionales.

PROGRAMA ANALÍTICO

1) *Breve revisión de conceptos básicos de estadística.*

Población, muestra, observaciones. Tipos de variables y escalas de medición. Distribución de frecuencias, histogramas, polígonos de frecuencias. Medidas de tendencia central, dispersión y forma. Frecuencia y probabilidad. Modelos de distribución de probabilidad para variables continuas. Pruebas de hipótesis, confianza, significación, *valor-p*, errores.

2) *Correlación paramétrica y no paramétrica. Análisis de regresión.*

Diagramas de dispersión bivariados. Covarianza y Correlación. Coeficiente de correlación lineal de Pearson. Correlación no paramétrica. Coeficiente de correlación de rangos de Spearman. Autocorrelación. Correlogramas. Correlación cruzada. Pruebas de aleatoriedad, uniformidad, tendencias y patrones en perfiles. Regresión tipo I y tipo II, estimación de coeficientes. Análisis de Varianza de la regresión. Bondad de ajuste, R^2 . Intervalos de confianza para los parámetros de la regresión. Predicciones e intervalos de confianza para las predicciones. Regresión múltiple y multivariante.

3) *Datos multivariantes*

Matrices de desviaciones, de covarianza y de correlación. Vector de medias. Pruebas de hipótesis sobre vectores de medias. Distancia de Mahalanobis. Prueba T2 de Hotelling. Prueba de homogeneidad de matrices de covarianza. Aporte de una variable a las diferencias. Análisis discriminante. Funciones discriminantes. Eficiencia discriminante de una variable. Análisis de agrupamiento (clusters). Índices y coeficientes de similitud. Distancia Euclidiana, distancia de Manhattan, correlación entre objetos. Matriz de similitud. Clasificación jerárquica de a pares. Dendrogramas. Análisis de interrelaciones entre variables (Modo R), y entre objetos o entidades (Modo Q). Análisis de Componentes principales (ACP). Análisis de factores. Rotación de factores. Biplot. Representación vectorial de la

CURSO DE POSTGRADO



matriz de covarianza. Análisis de Componentes principales (ACP). Biplot. Análisis de factores. Rotación de factores: ortogonal (método Varimax) y oblicua.

4) *Datos* composicionales

Diagrama ternario. Simplex. Operador clausura. Subcomposiciones. Transformación de datos: alr- y clr-. Medidas de diferencias. Distancia de Mahalanobis. Medidas de tendencia central y de dispersión. Divergencias composicionales. Clasificación. El problema de los ceros.

5) *Variables categóricas.*

Tablas de contingencia. Perfiles Marginales y Condicionales. Dependencia e independencia en Tablas de Correspondencias. Pruebas de hipótesis. Análisis de Correspondencias simple y múltiple: Descomposición en valores singulares. Interpretación de resultados. Inercia. Contribuciones. Gráficos de puntuaciones. Distancias Chi cuadrado. Coordenadas baricéntricas. Biplots

DESTINATARIOS

Profesionales o estudiantes de postgrado en ciencias de la Tierra (preferentemente geólogos, geoquímicos, químicos, físicos, agrónomos, biólogos)

CUPO: 25 alumnos.

MODALIDAD TEÓRICA-PRÁCTICA: Todo el curso se desarrolla de manera Teórico-práctica, intercalándose dentro de las explicaciones problemas cortos a resolver durante el transcurso de la clase. El curso se dictará vía Google Meet, mediante links que se proveerán a los alumnos vía email.

EVALUACIÓN: Presentación de un trabajo final individual.

CARGA HORARIA: 40 hs totales. El curso se dictará durante 5 semanas, con dos días de clase semanales, de 4 hs cada una.

MATERIAL Y BIBLIOGRAFÍA: Se entregarán en archivos .pdf las diapositivas que se emplearán en el dictado de las clases y también guías de problemas que se podrán resolver individualmente durante el transcurso de mismo y fuera de las horas de clase.

CURSO DE POSTGRADO



BIBLIOGRAFÍA

- **La Estadística como Herramienta en Ciencias de la Tierra.** Mena M., 2016. Ed. Akadia. 625pp.
- **Statistics and Data Analysis in Geology** , 3rd Edition. Davis, J.C., 2003. Wiley, 656pp.
- **Statistics of Earth Science Data: the Distribution in Time, Space and Orientation.** Borradaile, G., 2003. Springer, 351p.