



FESC

La Facultad de Estudios Superiores de Cuautla
en colaboración con el
Colegio Universitario Científico de Datos Invitan al

Diplomado

**Diseño de Investigación y
Análisis de Datos con R**

- ✓ Todo lo que un investigador necesita en aprendizaje avanzado y robusto en un solo programa
- ✓ Dominará el lenguaje R líder en ciencia de Datos aplicada a la investigación

MÓDULOS

- ✓ Diseño de investigación
- ✓ Técnicas clásicas
- ✓ Técnicas avanzadas y robustas
- ✓ Técnicas multivariante
- ✓ Big Data y ETL. Cubos OLAP y Tidyverse



Carr México Oaxaca 218, Plan de Ayala, 62743 Cuautla, Mor.



735 339 2795



FESC
UAEM | Facultad de
Estudios
Superiores de
Cuautla



Contenido

▶ **INFORMACIÓN GENERAL**

▶ **TEMARIO**

▶ **RESEÑA DE DOCENTE**



Carr México Oaxaca 218, Plan de Ayala, 62743 Cuautla, Mor.



735 339 2795



"Excelente estructura de los módulos del diplomado, empezando con la estadística descriptiva hasta llegar al análisis de clústeres y big data. Catedráticos altamente especializados con una amplia experiencia en el ramo. Lo recomiendo ampliamente."

Dr. Rubén Meraz.

"¡Me encantó el Diplomado! Formé parte de la séptima edición y me parece que todos los docentes e investigadores deberían acercarse a estos temas. Sin embargo, recomiendo este Diplomado con énfasis en el temario y metodología. El material que nos proporcionan es abundante y muy preciso. Las sesiones son fantásticas."

Dr. Horacio López.

"El diplomado es excelente muy completo. Recomendado 100%"

Dr. Edwin Mora.

"El curso ha sido un excelente guía para mis habilidades en estadística. Considero que es muy necesario estar al tanto de los avances en estadística como lo es el manejo de R. He publicado un artículo con los procedimientos que aprendí en el curso. Estoy agradecida con el Mtro. Tomás Salmerón"

Dra. Jessica Becerra.

"El curso desarrollado amerita una evaluación de excelente en función de los siguientes elementos:

- 1.- La pertinencia y actualidad de los temas tratados.
- 2.- El estudio profundo de los métodos estadísticos y su aplicación en la solución de problemas.
- 3.- La habilidad docente, la estructura metódica seguida en la impartición de la temática, la seriedad en el manejo de tiempo para las reuniones de trabajo y la disposición del instructor para aclarar dudas, profundidad de contenidos y aplicaciones prácticas.
- 4.- Los ejemplos y problemas propuestos permitieron entender conocimientos complejos asociados a la estadística.
- 5.- El material de apoyo proporcionado en el curso es de gran calidad, variedad y alto volumen de textos.

Todo en su conjunto hacen que el diplomado sea una gran experiencia de aprendizaje.

Muchas gracias Maestro Tomás."

Dr. Lamberto Vázquez.





Objetivo:

Convertirse en un científico de datos profesional con la herramienta del lenguaje R, aprendiendo su potencialidad para el procesamiento de datos, interpretando de forma correcta los resultados obtenidos y así innovar en sus proyectos de investigación.

A quién va dirigido:

- Licenciados, Maestros, Doctores e investigadores de Ciencias Experimentales (Biología, Psicología, química, matemáticas, etc.)
- A cualquier persona que quiera adentrarse en el mundo de la investigación y el análisis de datos
- Mínimo nivel de licenciatura
- Conocimientos básicos de estadística requeridos
- No se requieren conocimientos de programación en lenguaje R

Modalidad: presencial

200 hrs. de estudio



Fecha: Pendiente por definir



Distribución de tiempo semanal

- 3 hrs. presenciales
- 5 hrs. trabajo independiente
- 2 hrs. de actividades



Evaluación continua



**En permanente contacto
con el profesor especialista**



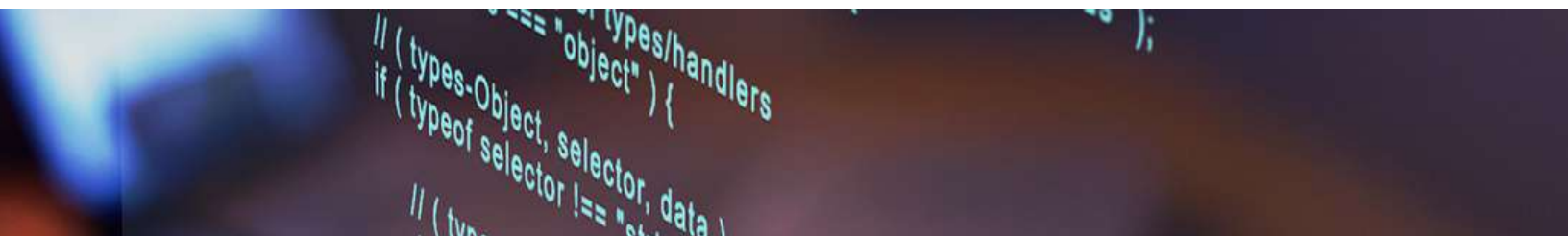
Inscripción





¿Por qué estudiar el Diplomado en Diseño de Investigación y Análisis de Datos con software R?

- 1) Comenzará de cero en programación.
- 2) Solo se le recomienda tener nociones básicas de estadística aplicadas a las ciencias experimentales.
- 3) Comprenderá y sabrá aplicar las nociones básicas para poder llevar a cabo una investigación, desde el póster de concepto, cálculos muestrales hasta los distintos tipos de diseños de investigación, entre otros.
- 4) Aprenderá las técnicas estadísticas más utilizadas en investigación.
- 5) Aplicará las técnicas clásicas paramétricas y no paramétricas, predictivas y de análisis multivariado.
- 6) Conocerá y sabrá en qué situaciones aplicar las técnicas robustas en las diferentes técnicas para el tratamiento de falta de supuestos y presencia de datos anómalos.
- 7) Aprenderá técnicas de machine learning.
- 8) Trabajará estrategias ETL y bases de datos (big data) con cubos OLAP.
- 9) Aprenderá a trabajar bases de datos con entorno tidyverse.
- 10) Al finalizar obtendrá un diploma avalado por la Facultad de Estudios Superiores de Cuautla de la UAEM





1

Módulo primero. Diseño de Investigación

1. Fundamentos y bases empíricas.

- 1.1. Evolución del método estadístico.
- 1.2. Ciencia en la sociedad.
- 1.3. Método científico.
- 1.4. Conocimiento.
- 1.5. Medición de datos.
- 1.6. Experimentación.
- 1.7. Clasificación.
- 1.8. Inicio de una investigación.

2. Elección del tema, objetivos e hipótesis.

- 2.1. Tema de investigación.
- 2.2. Objetivo de la investigación.
- 2.3. Hipótesis de la investigación.
- 2.4. Metas.

3. Muestreos, tipos, datos y controles.

- 3.1. Datos.
- 3.2. Muestreo.
- 3.3. Tipos de muestreos.
- 3.4. Con azarización.
- 3.5. Sin azarización.
- 3.6. Cálculo de tamaño muestral.
- 3.7. Grupo control y grupo/s experimental/es.
- 3.8. Principios éticos en la investigación.

4. Tipos de diseños de investigación.

- 4.1. Diseño de una investigación.
- 4.2. Tipo de diseño. Motivos de elección.
- 4.3. Propiedades de una investigación.
- 4.4. Clasificación de los diseños de investigación estadística.
- 4.5. Según la intención de los objetivos que originan la investigación.
- 4.6. Según la direccionalidad de la investigación.
- 4.7. Según el número de mediciones.
- 4.8. Según el grado de control.
- 4.9. Diseños observacionales.
- 4.9.1 Ensayos de prevalencia.
- 4.9.2. Estudios de casos y controles.
- 4.9.3. Estudios de cohortes.
- 4.10. Diseños Cuasiexperimentales.
- 4.10.1. Diseños intersujetos.
- 4.10.2. Diseños intrasujetos.
- 4.10.3. Pre-post test de grupo único.
- 4.10.4. Pre-post test con dos grupos.
- 4.11. Diseños experimentales.
- 4.11.1. Diseños de laboratorio.
- 4.11.2. Ensayos clínicos controlados.
- 4.11.3. Estudios de campo.

5. Comunicación de resultados.

- 5.1. Autoevaluación del trabajo.
- 5.2. Presentación oral de las investigaciones.
- 5.2.1. Métodos audiovisuales.
- 5.2.2. Exposición verbal.
- 5.2.3. Presentación en cartel.
- 5.3. Manuscrito. Publicación.
- 5.3.1. Título.
- 5.3.2. Autores.
- 5.3.3. Redacción del texto.
- 5.3.4. Resumen.
- 5.3.5. Introducción.
- 5.3.6. Material y métodos.
- 5.3.7. Discusión.
- 5.3.8. Agradecimientos.
- 5.3.9. Referencias.
- 5.4. Elección de la revista.

6. Técnicas estadísticas cuantitativas.

- 6.1. Estadística descriptiva.
- 6.2. Estadística predictiva.
- 6.3. Estadística multivariada.
- 6.4. Machine learning.
- 6.5. Técnicas robustas.





2 Módulo segundo. Técnicas clásicas.

1. R.

- 1.1. Instalación.
- 1.2. Primeros pasos.
- 1.3. Tipos de introducción de datos.
- 1.4. Importado y exportado de bases de datos.
- 1.5. Cambio de directorio.
- 1.6. Guardado de scripts.

2. Estadística descriptiva numérica y gráfica.

- 2.1. Medidas de posición y dispersión.
 - 2.1.1. Media.
 - 2.1.2. Mediana.
 - 2.1.3. Cuasivarianza.
 - 2.1.4. Cuasidesviación típica.
 - 2.1.5. Cuantiles.
 - 2.1.6. Resumen.
 - 2.1.7. describeBy.
 - 2.1.8. Descriptiva robusta. Huber. Alpha winsorizadas. Alpha recortadas.
 - 2.1.9. Test de Grubbs para detección de outliers.
- 2.2. La distribución normal.
- 2.3. Funciones útiles para investigación. Tratamiento de datos faltantes.
- 2.4. Gráficos de barras.
- 2.5. Gráficos de sectores.
- 2.6. Histograma.
- 2.7. Nubes de puntos.
- 2.8. Gráficos de cajas.
- 2.9. Gráficos para tablas de contingencia.
- 2.10.- Conclusión.

3. Pruebas no paramétricas.

- 3.1. Test de Wilcoxon.
- 3.2. Test de Wilcoxon – Mann-Whitney.
- 3.3. Test de Kolmogorov – Smirnov.
- 3.4. Test de Shapiro-Wilks.
- 3.5. Test de Kruskal – Wallis.
- 3.6. Test de Chi cuadrado.

4. T de Student.

- 4.1. Distribución t-Student.
- 4.2. Función t.test.
- 4.3. Función var.test.
- 4.4. Resolución. Supuestos.
- 4.5. Media de una población normal.
- 4.6. Media de una población no necesariamente normal. Muestras grandes.
- 4.7. Cociente de dos poblaciones independientes.

5. Análisis de varianza. Diseño completamente aleatorizado.

- 5.1. Un factor. HSD Tukey.
- 5.2. Generalizaciones robusta del test de Welch y del test de Box para un factor. Comparaciones robustas lincon.
- 5.3. Dos factores. HSD Tukey.
- 5.4. Generalización robusta del test de Welch para dos factores. Comparaciones robustas lincon.

6. Análisis de varianza. Medidas repetidas.

- 6.1. Análisis de Varianza. Medidas Repetidas.

7. Análisis de varianza multivariante (MANOVA).

- 7.1. Proceso de realización de MANOVA.

8. Regresión y correlación lineal simples.

- 8.1. Regresión lineal simple.
 - 8.1.1. Introducción.
 - 8.1.2. Modelo.
 - 8.1.3. Contraste.
 - 8.1.4. Tabla de análisis de varianza.
 - 8.1.5. Resolución e interpretación de resultados.
 - 8.1.6. Predicciones.
- 8.2. Correlación simple.

8.2.1. Introducción.

- 8.2.2. Coeficiente de correlación de Pearson. Conocimiento de otros métodos Kendall y Spearman.
- 8.2.3.- Resolución de correlación.

9. Regresión múltiple.

- 9.1. Modelo.
- 9.2. Contraste de la regresión lineal múltiple.
- 9.3. Tabla de análisis de la varianza para la regresión lineal múltiple.
- 9.4. Estimación de la varianza común.
- 9.5. Contraste de hipótesis sobre los coeficientes de regresión.
- 9.6. Regresión múltiple con R.
- 9.7. Selección secuencial de variables por pasos. Criterio de Akaike (AIC)

10. Correlación múltiple.

- 10.1.- Introducción.
- 10.2.- Resolución.
- 10.3.- Representaciones gráficas.





3

Módulo tercero. Técnicas avanzadas.

1. Técnicas actuales en regresión. Tratamiento de datos anómalos.

- 1.1. Introducción.
- 1.2. Bibliotecas.
- 1.3. Formulación.
- 1.4. Regresión lineal por mínimos cuadrados.
- 1.5. Técnicas actuales en Regresión (tratamiento de datos anómalos).
- 1.5.1. Recta de Huber.
- 1.5.2. LTS-Least Trimmed Squares.
- 1.5.3. LMS-Least Median of Squares.
- 1.5.4. Recta M-estimador (MM).

2. Regresión poisson.

- 2.1. Introducción.
- 2.2. Resolución del modelo.
- 2.3. Interpretación y predicciones.

3. Modelos de regresión logística. Logit y probit.

- 3.1. Introducción.
- 3.2. Modelos logit y probit.
- 3.3. Resolución del modelo.
- 3.4. Interpretación y predicciones.
- 3.5. Conclusión.

4. Regresión suavizada.

- 4.1. Introducción.
- 4.2. Formulación.
- 4.3. Resolución y toma de decisiones.
- 4.4. Conclusión.

5. Regresión múltiple robusta.

- 5.1. Introducción.
- 5.2. Regresión de Huber.
- 5.3. Regresión múltiple robusta de Huber.
- 5.4. MM-estimadores.
- 5.5. Regresión múltiple M-estimadores.
- 5.6. Interpretación de resultados, predicciones y toma de decisiones.
- 5.7. Conclusión.

6. Correlación simple y múltiple robustas.

- 6.1. Introducción.
- 6.2.- Correlación robusta para dos variables.
- 6.2.1. Correlación de porcentaje ajustado.
- 6.2.2. Correlación winsorizada.
- 6.2.3. M-estimadores de Goldberg e Iglewicz.
- 6.3. Correlación robusta para p variables.
- 6.3.1. Correlación de porcentaje ajustado entre p variables.
- 6.3.2. Correlación winsorizada entre p variables.
- 6.4. Interpretación de resultados, predicciones y toma de decisiones.
- 6.5. Conclusión.

7. Modelos lineales generalizados univariantes.

- 7.1. Desarrollo.
- 7.2. Conclusión.

8. Modelos lineales mixtos generalizados. Diseño por bloques aleatorios.

- 8.1. Introducción.
- 8.2. Realización del modelo lineal.
- 8.3. Comprobación de efectos fijos.
- 8.4. Comprobación de efectos aleatorios.
- 8.5. Creación del modelo.
- 8.6. Comprobación de supuestos del modelo.
- 8.7. Transformaciones.
- 8.8. Interpretación de resultados y toma de decisiones.
- 8.9.- Conclusión.

9. Árboles de regresión y clasificación CARTS.

- 9.1. Introducción.
- 9.2. Árboles de regresión.
- 9.3. Pruned o podado del Árbol.
- 9.4. Interpretación de resultados y toma de decisiones.
- 9.5. Árboles de clasificación.
- 9.6. Interpretación de resultados y toma de decisiones.
- 9.7. Conclusión.

10. Modelos aditivos generalizados GAM.

- 10.1. Introducción.
- 10.2. Modelos GAM para validación de modelos CARTs.
- 10.3. Interpretación de resultados y toma de decisiones.
- 10.4. Conclusión.

11. Análisis de supervivencia.

- 11.1. Introducción.
- 11.2. Análisis de supervivencia sin covariables.
- 11.3. Estimador de Kaplan Meier.
- 11.4. Comparación de curvas de supervivencia.
- 11.5. Análisis de supervivencia con covariables.
- 11.6. Uno y k grupos.

12. Modelos SARIMA de series temporales.

- 12.1. Introducción.
- 12.2. Objetivos y usos en las diferentes ciencias.
- 12.3. Paquetes a instalar en R.
- 12.4. Bibliotecas a abrir en R.
- 12.5. Cómo introducir los datos en R.
- 12.6. Resolución, interpretación de resultados, predicciones y toma de decisiones.
- 12.6.1. Filtrado lineal.
- 12.6.2. Modelos SARIMA
- 12.6.2.1. Identificación del modelo.
- 12.6.2.2. Estimación de los parámetros.
- 12.6.2.3. Diagnóstico.
- 12.6.2.4. Predicción.
- 12.6.2.5. Test de serie estacionaria.
- 12.6.2. Cointegración de Series.
- 12.7. Conclusión.

13. Parámetros gráficos.

- 13.1. Gráfico plot. Función legend.
- 13.2. Recta de regresión.
- 13.3. Histograma.
- 13.4. Gráfico de cajas.
- 13.5. Gráficos de barras con barras de error.
- 13.6. Gráficos para regresión y correlación con ggplot2.





4

Módulo cuarto. Técnicas de análisis multivariante.

1. Análisis de correspondencias.

- 1.1. Introducción.
- 1.2. Análisis de correspondencias bidimensional.
- 1.3. Análisis de correspondencias múltiple.
- 1.4. Interpretación de resultados y toma de decisiones.
- 1.5. Conclusión.

2. Análisis factorial. Validación de instrumentos tipo likert

- 2.1. Introducción.
- 2.2. Confección de un instrumento cuantitativo.
- 2.3. Proceso de validación de un instrumento.
- 2.4. Validación interna. Alpha de Cronbach.
- 2.5. Validación externa. Análisis factorial exploratorio y confirmatorio.
- 2.6. Interpretación de resultados y toma de decisiones.

3. Análisis de componentes principales. Reducción de dimensionalidad.

- 3.1. Introducción.
- 3.2. Formulación.
- 3.3. Análisis de componentes principales. Proporción de varianza y gráfico de sedimentación.
- 3.4. Representaciones gráficas
- 3.5. Interpretación de resultados y toma de decisiones.

4. Análisis discriminante.

- 4.1. Introducción.
- 4.2. Formulación.
- 4.3. Análisis discriminante lineal. Clasificación conocida.
- 4.4. Cross-validation.
- 4.5. Método de los k vecinos más próximos (knn).
- 4.6. Método de los k vecinos más próximos (knn) con cross-validation.
- 4.7. Análisis discriminante. K grupos y clasificación desconocida.
- 4.8. Interpretación de resultados y toma de decisiones.
- 4.9. Conclusión.

5. Escalado multidimensional.

- 5.1. Introducción.
- 5.2. Formulación.
- 5.3. Escalado multidimensional métrico.
- 5.4. Escalado multidimensional no métrico.
- 5.5. Interpretación de resultados y toma de decisiones.
- 5.6. Trabajo de matrices.
- 5.7. Conclusión.

6. Análisis clusters.

- 6.1. Introducción.
- 6.2. Formulación.
- 6.3. Elección del número de clusters. Matriz negativa de distancias de cuadrados.
- 6.4. Análisis clusters. Método jerárquico y no jerárquico.
- 6.5. Método de Ward.
- 6.6. Interpretación de resultados y toma de decisiones.
- 6.7. Conclusión.





5

Módulo Quinto: Big Data y ETL. Cubos OLAP y Tidyverse.

1. Trabajo de bases de datos.
 - 1.1. Importado de bases de datos.
 - 1.2. Creación de bases de datos con == y &.
 - 1.3. Guardado de bases de datos creadas.
 - 1.4. Estudio de variables cuantitativas y cualitativas.
2. Big data, ETL y cubos OLAP.
 - 2.1. OLAP (On-Line Analytical Processing – Procesamiento analítico en línea)
 - 2.1.1. Los Cubos OLAP
 - 2.1.2. Gestores de bases de datos con OLAP
 - 2.1.3. Hechos y dimensiones
 - 2.1.4. Operaciones OLAP
 - 2.1.4.1. Rebanada
 - 2.1.4.2. Dados
 - 2.1.4.3. Enrollar
 - 2.1.4.4. Profundizar
 - 2.1.4.5. Pivote
3. Tidyverse.
 - 3.1. Uso del pipe %>%
 - 3.2. Selección de columnas.
 - 3.3. Exclusión de columnas.
 - 3.4. Filtrado.
 - 3.5. Realizar operaciones entre columnas.
 - 3.6. Agrupar y desagrupar por categorías.
 - 3.7. Organizar tablas.
 - 3.8. Combinar variables.
 - 3.9. Separar variables.
 - 3.10. Joins para concatenar tablas.
4. Bootstrap.
 - 4.1. Simulación de variables en base a la media.
 - 4.2. Simulación de variables en base a la sd.
 - 4.3. Simulación de variables en base a medidas robustas.
 - 4.4. Integración de variables simuladas en nuevas bases de datos.
 - 4.5. Simulación y técnicas de regresión y anova.
5. Visualización de datos con ggplot y plotrix.





DOCENTE TITULAR



Mtro. Tomás Alberto Salmerón Enciso

Maestro en Técnicas Actuales de Estadística Aplicada, con una Especialización en Técnicas Robustas y Avanzadas de Estadística Aplicada y Licenciado en Psicopedagogía por la Universidad Nacional de Educación a Distancia (UNED – Madrid, España –). Trabajó para el Dpto. de Protección de Cultivos del Instituto de Formación Agraria y Pesquera de Andalucía (IFAPA – Gobierno de Andalucía, España) desarrollando la tarea de asesor en estadística e investigación. Realizó estudios en colaboración con el Comisionado de Energía de la Unión Europea y sus políticas Medioambientales en el año 2014. Especialista en software R, ha realizado publicaciones y desarrollado varios libros de estadística avanzada con software R y temarios para universidades. Actualmente, radica en México, donde es Director de Colegio Científico de Datos, imparte formación de posgrado en estadística avanzada para universidades, es asistente de investigación y estadística para proyectos conahcyt, consultor externo para el Fondo de Población de las Naciones Unidas capítulo México y expone en congresos nacionales e internacionales.

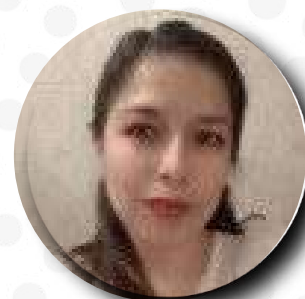
Docentes Invitados



Mtro. Enrique Díaz Ocampo



Dra. Violeta Larios Serrato



Dra. Patricia Alejandra
Cuevas Chávez

