



Información general

| | | | | | |
|-----------------------------------------------------------------|---------------------------------------------|-----------------------------------------------|----------------------------|-----------------------|-----|
| Asignatura | Fundamentos de Análisis de datos | | | | |
| Código | | | | | |
| Tipo de asignatura | Obligatoria X | | Electiva | | |
| Tipo de saber | Obligatoria básica o de fundamentación X | Obligatoria profesional | Obligatoria complementaria | | |
| Número de créditos | 3 | | | | |
| Tipo de crédito | A | | | | |
| Horas de trabajo con acompañamiento directo del profesor | 64 | Horas de trabajo independiente del estudiante | 192 | Total de horas | 256 |
| Prerrequisitos | | | | | |
| Correquisitos | | | | | |

| | | |
|------------------------------------|------------------------------------|--|
| Horario | | |
| Salón | | |
| Profesor | Nombre | |
| | Correo electrónico | |
| | Lugar y horario de atención | |
| | Página web | |
| Profesor auxiliar o monitor | Nombre | |
| | Correo electrónico | |
| | Lugar y horario de atención | |
| | Página web | |

Resumen y propósitos de formación del curso

La estadística es una disciplina aplicada en la toma de decisiones en condiciones de incertidumbre en todos los campos de la actividad humana. Actualmente, es evidente que la interpretación y avance de las investigaciones en las Ciencias Biológicas y de la Salud están basadas y dependen en gran parte de los métodos estadísticos. Por esta razón, es esencial que los estudiantes de las disciplinas mencionadas deben familiarizarse con los razonamientos, procesos y métodos estadísticos.

Además, el análisis e interpretación de datos es un elemento indispensable en la caracterización y entendimiento de los fenómenos biológicos, como e indico antes, a nivel general, igualmente desempeña un importante papel en todas las esferas del conocimiento. Por lo tanto, su conocimiento, entendimiento y adecuada aplicación le proporcionará al estudiante las técnicas y procedimientos necesarios para observar, ordenar, describir, graficar, tabular, correlacionar, analizar, inferir e interpretar datos relevantes a diferentes aplicaciones.

Propósitos del curso

Desarrollar capacidades y habilidades dirigidas al análisis e interpretación de los datos en el contexto de un problema de investigación con una mirada exploratoria, con la ayuda de un software estadístico y una hoja de calculo



UNIVERSIDAD DEL ROSARIO

Temas

- Tema 1. Métodos descriptivos y Análisis exploratorio de datos
- Tema 2. Teoría y aplicación de la probabilidad
- Tema 3: Variables aleatorias y distribuciones de probabilidad discretas y continuas
- Tema 4: Introducción al muestreo y concepto de distribuciones muestrales
- Tema 5: Pruebas de hipótesis paramétricas
- Tema 6: Distribución chi – cuadrado X^2 , datos categóricos
- Tema 7: Análisis de varianza
- Tema 8: Análisis de correlación y regresión

Resultados de aprendizaje esperados (RAE)

OBJETIVOS GENERALES

Desarrollar capacidades y habilidades dirigidas al análisis de los datos en el contexto de un problema de investigación con una mirada exploratoria

OBJETIVOS ESPECÍFICOS

- (1) Desarrollar una visión exploratoria para el análisis de datos
- (2) Describir un conjunto de datos partiendo de los objetivos propuestos para el análisis
- (3) Calcular probabilidades utilizando la definición clásica de probabilidad y construir distribuciones de probabilidad
- (4) Comprender la distribución muestral como una distribución de probabilidad que representa el comportamiento de los valores de un estadístico
- (5) Aplicar correctamente las pruebas de hipótesis y los intervalos de confianza en la realización de inferencias
- (6) Identificar las situaciones que pueden ser resueltas por los modelos de análisis de varianza
- (7) Analizar el tipo de distribución que siguen los datos
- (8) Utilizar la distribución ji-cuadrado en el análisis de hipótesis de independencia y homogeneidad en las tablas de contingencia
- (9) Construir e interpretar modelos de regresión**

Por tanto, al término del curso de fundamentos de análisis de datos el alumno estará en capacidad de:

- Resolver problemas donde se involucren eventos con incertidumbre, aplicando los modelos analíticos apropiados.



UNIVERSIDAD DEL ROSARIO

- Obtener, analizar y representar gráficamente conjuntos de datos tomados de una situación real, haciendo síntesis de ellos, mediante descripciones numéricas.
- Aplicar los fundamentos de la teoría de la probabilidad en el cálculo de probabilidades de diferentes tipos de sucesos.
- Particularizar en el estudio de las distribuciones Binomial, Poisson. Determinar el modelo matemático apropiado que deba aplicarse.
- Aplicar los conceptos de variable aleatoria continua, con base a situaciones reales o simuladas y establecer la correspondiente distribución de probabilidad continua, particularizar en el estudio de las distribuciones Uniforme, Exponencial y Normal.

Actividades de aprendizaje

- El estudiante deberá tener en cuenta y aplicar los siguientes aspectos durante todo el curso: reflexión, análisis, pensamiento analítico y método científico.
- Deberá leer el tema de la sesión con anterioridad a la clase. Cada sesión se iniciará resolviendo las dudas que traigan los estudiantes del tema anterior o ejercicios donde hayan tenido dificultad en su desarrollo.
- El profesor expondrá la parte conceptual complementada con ejercicios y dará espacio a la participación de los estudiantes.
- El estudiante desarrollará durante la clase ejercicios con apoyo del profesor.
- El estudiante debe realizar un trabajo continuo y sistemático fuera de clase. Para lo cual debe realizar los ejercicios propuestos en cada tema de este programa, los cuales responden a problemáticas de biología y ciencias de la salud y al mismo tiempo los entrenará para la realización del trabajo final.

Usar un software computacional como R, SPSS y Excel. Utilizar correctamente calculadoras manuales que permiten reducir el tiempo de cálculo a fin de hacer énfasis en el análisis e interpretación de los resultados logrados.



UNIVERSIDAD DEL ROSARIO

Actividades de evaluación

| Tema | Actividad de evaluación | Porcentaje |
|----------------------|------------------------------------------------------------------------|------------|
| 1-2 | Parcial 1 | 20% |
| 3-4-5 | Parcial 2 | 20% |
| | Proyecto final | 20% |
| Labor en clases | Labor durante el curso Trabajo en clases (10%) Talleres - Quices | 15% |
| Resumen del semestre | Examen final – | 25% |

Programación de actividades por sesión

| Fecha | | Tema | Contenidos | Trabajo independiente del estudiante | Recursos que apoyan la actividad |
|--------|-----------|------|-----------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|--------------------------------------|----------------------------------|
| Semana | Actividad | | | | |
| 1 | 1 | 1 | Importancia de la estadística y probabilidad en la biología y ciencias de la salud. Elementos, variables y observaciones. Las escalas de medición. Datos cualitativos y cuantitativos. -Datos de sección transversal y de series de tiempo | Ejercicios 1.1, 1.2 | texto |
| | 2 | | Presentaciones tabulares y gráficas: resumen de datos cuantitativos y cualitativos. Tabulaciones cruzadas y diagramas de dispersión | Ejercicios 1.3 | Sala de Computo |
| 2 | 3 | 1 | Distribuciones de frecuencias, histogramas, reglas agrupar datos en categorías o clases. Medidas de localización. Medidas de variabilidad | Ejercicios 1.4, 1.4 | texto |
| | 4 | | Medidas de la forma de la | Ejercicios | Sala de |



UNIVERSIDAD DEL ROSARIO

| | | | | | |
|---|----|-----|----------------------------------------------------------------------------------------------------------------------------------------|--------------------------|----------------------------------------|
| | | | distribución, de la posición relativa y de la detección de observaciones atípicas. Cajas de bigotes. Análisis exploratorio de datos | 1.5, 1.6 | Cómpu o |
| 3 | 5 | 1 | Medidas de la asociación entre dos variables La media ponderada y el empleo de datos agrupados | Ejercicios 1.7 | texto |
| | 6 | | Interpretación de las probabilidades. Experimentos, reglas de conteo y asignación de probabilidades | | Sala de Cómput o |
| 4 | 7 | 2 | Eventos y sus probabilidades Algunas relaciones básicas de probabilidad | Ejercicios 2.1 – 2.4 | salón de clases. Guía de ejercicios |
| | 8 | | Diagramas de árbol – Repaso temas – Primer entrega proyecto de estudio | | salón de clases. Guía de ejercicios |
| 5 | 9 | | PRIMER PARCIAL | | salón de clases |
| | 10 | | Retroalimentación condicional Probabilidad Teorema de bayes | Ejercicios 3.1 – 3.5 | Sala de Cómput o |
| 6 | 11 | 2-3 | Variables discretas, Funciones de densidad discreta, esperanza matemática. La función de distribución acumulada | Ejercicios 4.1, 4.4, 4.3 | texto |
| | 12 | | La distribución binomial, Poisson. | Ejercicios 4.4, 4.5 | Sala de Cómput o |



UNIVERSIDAD DEL ROSARIO

| | | | | | |
|----|----|---|-------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|-------------------------|----------------------|
| 7 | 13 | 4 | Variables continuas, Funciones de densidad continua, esperanza matemática. La función de distribución acumulada. Distribución normal, reglas de la normal, tipificación | Ejercicios de 5.1 – 5.4 | texto |
| | 14 | | Muestreo, tipos de muestreo | Ejercicios 6.1 | Sala de Computo |
| 8 | 15 | 5 | Estimación puntual de la media e introducción a la estimación por intervalo: teorema central del límite | Ejercicios 6.2 | Texto |
| | 16 | | Estimación por intervalo / Teorema central del límite | Ejercicios 6.2 | Sala de Computo |
| 9 | 17 | 5 | Tamaño de la muestra: estimación / Tamaño de la muestra. Introducción a las pruebas de hipótesis. Valores del nivel de significancia | Ejercicios 6.3 – 6.6 | texto |
| | 18 | | Prueba de hipótesis para una muestra: media, varianza y proporción | | Sala de Computo |
| 10 | 19 | 5 | Inferencias sobre medias, varianzas y proporción. Intervalos de confianza | Ejercicios 7.1 – 7.2 | texto |
| | 20 | | Contraste T para datos dependientes o emparejados – | | Ejercicios 8.5 – 8.6 |



UNIVERSIDAD DEL ROSARIO

| | | | | | |
|---------------------|----|---|--------------------------------------------------------------------------|-------------------------------|-----------------|
| | | | Repaso temas tratados – Avances proyecto de estudio | | o |
| 11 | 21 | 6 | SEGUNDO PARCIAL | | texto |
| | 22 | | Distribución chi cuadrado, bondad de ajuste, independencia, homogeneidad | Ejercicios 9.1 – 9.3 | Sala de Computo |
| 12 | 23 | 7 | Aplicaciones del análisis de varianza a una vía | Ejercicios 10.1 – 10.2 - 10.3 | texto |
| | 24 | 7 | Aplicaciones del análisis de varianza a dos vías | | Sala de Computo |
| 13 | 25 | 7 | Comparaciones múltiples y por parejas. Contraste T de Bonferroni | Ejercicios 10.3 | texto |
| | 26 | | Taller Análisis de varianza a una y dos vías | | Sala de Computo |
| 14 | 27 | 8 | Introducción a la regresión lineal simple | Ejercicios 11.1 – 11.2 | texto |
| | 28 | 8 | Método de los mínimos cuadrados | | Sala de Computo |
| 15 | 29 | | Sustentación y entrega Proyecto Final | | Guía ejercicios |
| | 30 | 8 | Coeficientes de correlación y determinación | Ejercicios 11.3 | Sala de Computo |
| 16 | 31 | 8 | Estimaciones por intervalos de confianza, estimaciones | Ejercicios 11.4 – 11.5 | salón de clases |
| | 32 | 8 | Regresión múltiple – Repaso Final | Ejercicios 11.6 | Sala de Computo |
| EXAMEN FINAL | | | | | |



UNIVERSIDAD DEL ROSARIO

Bibliografía

Milton, Susan. Estadística para Biología y Ciencias de la Salud. Mc Graw Hill, 2001
Jay Kerns, G, Introduction to probability and statistics Using R. Third Edition, 2018. En <https://cran.r-project.org/web/packages/IPSUR/vignettes/IPSUR.pdf>

Bibliografía complementaria

Ayala, G., Estadística básica, 2015. En <https://www.uv.es/ayala/docencia/nmr/nmr13.pdf>
Diez D, Barr C, Çetinka-Rundel M. OpenIntro Statistics, 2nd Edition. En <https://www.openintro.org/stat/textbook.php>. [Open Intro 2014](https://www.openintro.org/stat/textbook.php)
Faraway, J. Practical Regression and Anova Usisn R, 2002
Pagano M, Gauvreau K. Principles of Biostatistics. Duxbury. Segunda edición, 2000

Bases de datos sugeridas: <http://www.gapminder.org/>
http://estadisticas.cepal.org/cepalstat/WEB_CEPALSTAT/estadisticasIndicadores.asp?idioma=e

Consultar las disponibles en la Universidad

Acuerdos de funcionamiento (Reglas de juego)

- El curso se desarrollará en forma presencial
- Cualquier inquietud o reclamación por parte del estudiante se resolverá de acuerdo con lo contemplado en el reglamento.
- No se permite el uso de celulares en el aula de clase. No es calculadora. No fotos.
- Los exámenes parciales y quices se llevarán a cabo en hora de clase y en la fecha predeterminada. Tienen carácter individual.
- Los parciales y examen se deben presentar en hoja de examen
- El profesor no responde preguntas en los parciales y examen
- Se respetarán los porcentajes indicados en cada corte.
- El Programa del curso lo encuentra en MOODLE.
- El estudiante debe verificar fechas de evaluaciones
- En las evaluaciones puede llevar un resumen de las fórmulas, tablas (normal, t-student, Chi cuadrado, Fisher) sin definiciones, ni ejercicios resueltos, no se permite prestar la calculadora o las fórmulas, ni sacar cuaderno o libros
- El examen final no se reemplazará por trabajo final y será de los temas vistos durante el semestre. Esta evaluación será individual. No se eximirá del examen



final a ningún estudiante.

- Solo se harán supletorios de **parciales** autorizados por la Secretaria Académica.
- En la plataforma Moodle, estará el programa a fin de que el estudiante consulte las actividades extraclase que deben desarrollar. De igual manera, allí el estudiante podrá consultar las temáticas que se abordarán en cada sesión, fechas importantes o cualquier información que el profesor considere que es de interés para el estudiante.
- El trabajo final debe desarrollarse en parejas. Los criterios de presentación serán a manera de informe ejecutivo tal como lo propone Webster, en uno de sus apéndices y debe hacerse en computador utilizando Word, Excel o SPSS o R.
- Teniendo en cuenta el reglamento formativo-preventivo y disciplinario de la Universidad del Rosario, y la certeza de que las acciones fraudulentas van en contra de los procesos de enseñanza y aprendizaje, cualquier acto corrupto vinculado a esta asignatura será notificado a la secretaría académica correspondiente de manera que se inicie el debido proceso disciplinario. Se recomienda a los estudiantes leer dicho reglamento para conocer las razones, procedimientos y consecuencias que este tipo de acciones pueden ocasionar, así como sus derechos y deberes asociados a este tipo de procedimientos.
- La asignatura no tiene ningún tipo de Bono.

Observación de estricto cumplimiento en parciales y examen final:

No se permite el uso o consulta de libros, apuntes, compañeros o cualquier medio electrónico. Los celulares o dispositivos móviles deben estar apagados y guardados durante todo el parcial (Tener alguno de estos dispositivos a la mano será considerado intento de fraude). Cualquier incumplimiento de lo anterior conlleva la anulación del parcial. Las respuestas deben estar totalmente justificadas, de no estar sustentadas la respuesta del punto su valor será cero. El profesor no responderá ninguna inquietud sobre la prueba. La comprensión de los enunciados también son parte de la evaluación

Cualquier duda de tipo académico remítase a:

<http://www.urosario.edu.co/La-Universidad/Documentos-Institucionales/ur/Reglamentos/Reglamento-Academico-de-Pregrado/>

- El trabajo final debe desarrollarse en grupos. Los criterios de presentación serán a manera de informe ejecutivo tal como lo propone Webster, en uno de sus apéndices y debe hacerse en computador utilizando Word, Excel o SPSS o R.
- Teniendo en cuenta el reglamento formativo-preventivo y disciplinario de la Universidad del Rosario, y la certeza de que las acciones fraudulentas van en contra de los procesos de enseñanza y aprendizaje, cualquier acto corrupto vinculado a esta asignatura será notificado a la secretaría académica correspondiente de manera que se inicie el debido proceso disciplinario. Se recomienda a los estudiantes leer dicho reglamento para conocer las razones, procedimientos y consecuencias que este tipo de acciones pueden ocasionar, así como sus derechos y deberes asociados a este tipo de procedimientos.



UNIVERSIDAD DEL ROSARIO

Observación de estricto cumplimiento en parciales y examen final:

No se permite el uso o consulta de libros, apuntes, compañeros o cualquier medio electrónico. Los celulares o dispositivos móviles deben estar apagados y guardados durante todo el parcial (Tener alguno de estos dispositivos a la mano será considerado intento de fraude). Cualquier incumplimiento de lo anterior conlleva la anulación del parcial. Las respuestas deben estar totalmente justificadas, de no estar sustentadas la respuesta del punto su valor será cero. El profesor no responderá ninguna inquietud sobre la prueba. La comprensión de los enunciados también son parte de la evaluación.

Si el estudiante se presenta 20 minutos luego de dar inicio a alguna evaluación parcial o final, no podrá presentarla y deberá solicitar supletorio siguiendo la reglamentación institucional.

Cualquier duda de tipo académico remítase a:

- <http://www.urosario.edu.co/La-Universidad/Documentos->



UNIVERSIDAD DEL ROSARIO

Trabajo aplicado fundamentos de análisis de datos

El objetivo de este trabajo es que al finalizar el semestre los estudiantes den respuesta a la pregunta de investigación del trabajo presentado; haciendo uso de los temas de formación del curso y de las diferentes fases de una investigación estadística: Planteamiento del problema, objetivos (general y específicos), planteamiento de las hipótesis de investigación y pregunta de investigación. Definir cuál es la población, como selecciono la muestra y como calculo su tamaño, indicar tipo de variables. Como obtuvo la base de datos (fuente) entre otros aspectos a tener en cuenta.

La pregunta de trabajo debe estar en el contexto de la biología y ciencias de la salud y puede ser impuesta por el docente o propuesta por el estudiante.

Para dar respuesta a la pregunta de investigación deberán realizar los siguientes pasos:

1. Recoger datos que permitan dar respuesta a la pregunta de trabajo. Para esto pueden hacer uso de encuestas o base de datos que les permita obtener al menos 5 variables (3 cuantitativas donde al menos dos sigan una distribución normal y 2 cualitativas). Esta información debe presentarse en el primer corte. En cada caso encuesta o base de datos, mínimo se requieren 60 observaciones.
2. Realizar una revisión bibliográfica con mínimo tres referentes teóricos asociados a la pregunta de investigación. La revisión bibliográfica debe ser presentada en mínimo una (1) página con fuente Times New Roman o Arial de 12 puntos e interlineado de 1.5 centímetros con normas APA. La revisión debe estar incluida en la entrega que realizarán en el segundo corte.
3. Dar respuesta a la pregunta de investigación por medio de los análisis solicitados en los tres cortes y con base en la revisión bibliográfica. Esto deben presentarlo en el tercer corte y dicho análisis debe tener una extensión mínima de 2 páginas con fuente Times New Roman o Arial de 12 puntos e interlineado de 1.5 cm.

Los temas que desarrollan en el trabajo aplicado por corte deben contener como mínimo los contenidos que se presentan a continuación.

Primer corte:

1. Título de la investigación, objetivos (general y específicos), preguntas de investigación.
2. Definir población y muestra.
3. Para cada una de las 5 variables del estudio indique tipo de variable y escala de medición.
4. Realice las distribuciones de frecuencia de cada una de las variables.
5. Calcule y analice la distancia de la que se encuentra la estimación puntual del valor del parámetro poblacional.
6. Plantee hipótesis sobre el valor del parámetro poblacional y realice el procedimiento necesario para ser o no rechazado.



UNIVERSIDAD DEL ROSARIO

Segundo corte:

1. Inferencia estadística con medias y proporciones con dos muestras.
2. Inferencia acerca de varianzas poblacionales.
3. Prueba de bondad de ajuste.
4. Análisis de varianza
5. Construcción de un modelo de regresión lineal simple.

Tercer corte

Informe final

Evaluación

El trabajo aplicado tiene un valor de 20% dentro de la nota del curso. Esta nota aplica para la nota del tercer corte. La ejecución del trabajo aplicado se realizará a partir de la semana cuatro. Esto al tener en cuenta que si son los estudiantes son quienes proponen el tema del trabajo aplicado requieren tener conocimientos básicos en estadística inferencial para realizar una propuesta viable y optimizar el tiempo.

El 20% será asignado de la siguiente manera:

10% Presentación de trabajo escrito

10% Sustentación oral del trabajo

Se aclara que si un estudiante no presenta el trabajo escrito no podrá realizar la sustentación. La nota del trabajo es individual.