



UNIVERSIDAD DE COSTA RICA
SEDE DE OCCIDENTE, RECINTO DE GRECIA
ESTADÍSTICA PARA INFORMÁTICOS (XS0105)
PROGRAMA DEL CURSO

Período lectivo	Del 6 de marzo al 14 de julio
Profesora	Andrea Collado Chaves, acollado@ccp.ucr.ac.cr
Horario de clase	Jueves de 8:30 a.m. a 12:15 a.m. Aula 213
Horario consulta	Jueves de 1:00 a 3:00 p.m. o por cita. Cubículo 02

1. DESCRIPCION DEL CURSO

Este curso pretende que el/la estudiante sea capaz de aplicar los conceptos y técnicas básicas de la Estadística en la resolución de problemas empíricos. En especial, se espera que comprenda las definiciones matemáticas e interprete los conceptos subyacentes y sus diferencias con los conceptos de la Informática.

El Excel se usará como herramienta para el análisis y servirá para aplicar los conceptos de la Estadística Descriptiva e Inferencial. Adicionalmente, con el trabajo final se pretende despertar un pensamiento crítico en el estudiante, que le permita crear su propio plan para el levantamiento, recolección, análisis y presentación de datos, confeccionar un instrumento de medición y una plataforma de entrada para el ingreso de datos.

El curso se impartirá en 15 días o el equivalente a 60 lecciones. Se divide en diez módulos de aprendizaje que cubren los elementos más importantes de la Estadística Descriptiva e Inferencial y sus aplicaciones en las distintas áreas del quehacer de un informático/a.

Los módulos de aprendizaje son:

1. Conceptos generales de Estadística
2. Distribución de frecuencias
3. Medidas descriptivas
4. Métodos de presentación de datos
5. Asociación estadística y regresión lineal
6. Elementos de probabilidad
7. Distribuciones de probabilidad
8. Elementos de muestreo y estimación
9. Pruebas de significancia, análisis de varianza y regresión lineal
10. Levantamiento y recolección de datos

Con estos módulos se cubren las técnicas cuantitativas básicas de la Estadística Descriptiva e Inferencial y se construyen nuevos conocimientos para el análisis de la información Estadística.

En síntesis, se espera que al final el curso los y las estudiantes estén en capacidad de realizar una investigación estadística completa.

1.1. Objetivo General

Desarrollar en el estudiantado la destreza y el criterio básico para el manejo de información estadística mediante la implementación de métodos y técnicas propias del análisis estadístico y su aplicación en la práctica del profesional de la informática.

1.2. Objetivos específicos de conocimiento

Al finalizar el curso, cada participante tendrá criterio y conocimiento básico para:

1. Identificar el problema estadístico y sus elementos.
2. Describir los diferentes conceptos que se aplican durante una investigación estadística.
3. Conocer las diferentes formas de recopilar, organizar y presentar grupos de datos.
4. Interpretar distribuciones de frecuencias, cuadros y gráficos estadísticos.
5. Establecer las diferencias entre la Estadística Descriptiva y la Estadística Inferencial.
6. Comprender los conceptos básicos sobre probabilidades, distribuciones de probabilidades discretas y continuas e inferencia estadística.

1.3. Objetivos específicos de desarrollo de habilidades

Al finalizar el curso se espera que el/la participante pueda:

1. Determinar el problema estadístico y sus elementos.
2. Manejar bases de datos provenientes de investigaciones cuantitativas.
3. Usar Excel para procesar datos, calcular indicadores y presentar información.
4. Efectuar los diferentes cálculos de las medidas de tendencia central y de variabilidad.
5. Construir e interpretar una distribución de frecuencias.
6. Realizar pruebas de hipótesis, asociación, correlación y regresión lineal.
7. Generar su propio plan para el levantamiento, recolección, análisis y presentación de datos estadísticos.
8. Confeccionar un instrumento de medición como es el cuestionario y una plataforma de entrada para el ingreso de datos.
9. Llevar a cabo una investigación estadística completa.

2. METODOLOGÍA

El enfoque pedagógico que inspira el curso se puede resumir en la frase “la mejor manera de aprender es haciendo”. Por tanto, cada lección pondrá especial énfasis en la aplicación práctica de los temas estudiados y se favorecerán las actividades participativas.

Las personas que participan del curso tendrán la oportunidad de conocer y analizar diversos casos y aplicaciones reales extraídas del contexto nacional. En TODAS las sesiones prácticas se utilizará Excel para el análisis de estadístico y el CPro para el ingreso de datos.

Se asignarán lecturas que el/la estudiante deberá hacer ANTES de la clase. Periódicamente se harán quices y prácticas en el horario de clases y se asignarán tareas semanales para comprobar la lectura del material obligatorio y la aplicación y entendimiento de la materia. Además, se ha programado un examen parcial y un trabajo final.

El trabajo final consiste en el diseño y aplicación de una investigación estadística completa donde se elabore un instrumento de medición, una plataforma de ingresos de datos, se analice la información y se ejecuten pruebas de asociación. El trabajo final es individual. Las otras actividades podrían ser grupales.

Como material de apoyo, el curso contará con una página Web en la siguiente dirección: <http://fce.ucr.ac.cr/moodle>. Ahí los/las estudiantes podrán consultar el programa, referencias bibliográficas, resúmenes, prácticas, bases de datos, el cronograma y otros materiales complementarios.

3. EVALUACION

La evaluación se basa en un examen parcial, la nota promedio de quices y tareas, la realización de prácticas en clase y el trabajo final. El siguiente esquema presenta el desglose porcentual de la nota.

Un examen parcial	20%
Quices y tareas	20%
Prácticas en clase	10%
Trabajo final	50%

El examen parcial será el 7 de junio y cubrirá los módulos del 1 al 9. La clase antes del examen se hará una práctica. Dado que no se dispone de un laboratorio de computación el examen evaluará los conceptos y su aplicación pero no el uso de Excel.

El trabajo versa especialmente sobre el modulo 10 y se le dedicará casi el mes de junio. Se debe entregar impreso. Debe ser un trabajo escrito en un procesador de texto y deberá cumplir con los lineamientos de contenido y forma que la profesora especificará. Se comenzará a desarrollar desde inicio del curso y va guardar mucho paralelismo con los temas cubiertos. Habrá entregas parciales que serán componentes del documento final.

Los y las estudiantes tienen la posibilidad de rehacer, tantas veces como sea necesario los componentes del trabajo final, en caso de obtener una mala calificación en la primera entrega. La calificación del componte será el promedio de la nota inicial y la nota más alta lograda en las entregas posteriores.

El esquema de calificación para el trabajo final es el siguiente:

Definición del problema	5%
Plan para el levantamiento de la información	10%
Plan para el análisis de la información	5%
Diseño del instrumento	5%
Diseño de entrada de datos	5%
Aplicación del instrumento y entrada de datos	5%
Presentación de cuadros y gráficos	5%
Documento final	10%

El/la estudiante que según el esquema de evaluación presentado en la primera tabla, obtenga una nota de aprovechamiento superior o igual al 70%, aprobará el curso. Los/las estudiantes que, habiendo cumplido con todos los requisitos del curso, no alcanzan la nota mínima de aprobación tendrán derecho a una prueba general de reposición cuya nota máxima es un 7,0.

4. FALTA A EXÁMENES Y PRÁCTICAS

Se calificaran con cero las prácticas y tareas no entregadas en fecha fijada. Si por causa debidamente justificada (Artículo 24 del nuevo Reglamento de Régimen Académico Estudiantil) un estudiante falta a un quiz o un examen, tiene una segunda oportunidad, presentando la excusa correspondiente: si procede, la profesora le programará la reposición en los 10 días siguientes al reintegro a sus estudios.

Dado que los exámenes son en horas de clase de Estadística, no se acepta como justificación de ausencia la asistencia a actividades en otros cursos.

5. TEMARIO BÁSICO

Las referencias para cada modulo se muestran en itálica, estas se consideran complementarias al material visto en clase.

Módulo 1: Conceptos generales de Estadística

- Conceptos de estadística y estadísticas. *Hernández (2002). Capítulo 1*
- Estadística descriptiva e inferencial. *Hernández (2002). Capítulo 1*
- Información existente y no existente. Fuentes primarias y secundarias. *Gómez: 29-33*
- Definición de problema, objetivos y población. Definición de las características a investigar y de su naturaleza. *Hidalgo: 11-15*
- Selección de las unidades de estudio: Estudio por censo (enumeración total), estudio por muestreo: aleatorio y no aleatorio.
- Marco de muestra y su importancia durante una investigación.
- Tipos de muestreo: aleatorio simple al azar (tabla de números aleatorios), muestreo sistemático, muestreo por conglomerados.

Módulo 2: Distribuciones de frecuencia

- Distribuciones de frecuencias simples y cruzadas. *Hernández: 94-103. Pérez: 35-44*
- Las frecuencias acumuladas y su interpretación. *Hernández: 83-93*
- Bases de datos en Excel: ventajas y desventajas. Auto filtro, caracteres alfanuméricos
- Tabla dinámica para distribuciones de frecuencia simple. *Pérez: 44-49*

Módulo 3: Medidas de posición central y dispersión

- Medidas de tendencia central: moda, mediana y promedio aritmético para datos desagregados. Interpretación, usos y limitaciones.
- Efectos de los valores extremos. Distribuciones simétricas y asimétricas.
- El fenómeno de la variabilidad y su importancia. Necesidad de las medidas de dispersión de un conjunto o población.
- El recorrido o amplitud general y el intervalo intercuartil. Ventajas y desventajas.
- La variancia y la desviación estándar. Cálculo e interpretación de la desviación estándar.
- El coeficiente de variación: definición, utilidad, cálculo e interpretación.
- Medidas de posición y dispersión con tabla dinámica. *Pérez: 99-118*
- Medidas de posición y dispersión con el complemento de análisis estadístico en Excel. *Pérez: 78-99.*
- Referencia: *Hernández, O. (2004): Capítulo 4.*

Módulo 4: Métodos de presentación de la información estadística

- Series estadísticas. *Pérez: 26-34*
- Detalles esenciales en la presentación de datos numéricos: formas de presentar datos.
- Formas de texto y semitabular. *Hernández: cap 3*
- Cuadros estadísticos. *Hernández: cap 5*
- Análisis gráfico: partes indispensables de un gráfico estadístico. Errores frecuentes. *Hernández: cap 6. Pérez: 50-60*

Módulo 5: Asociación, correlación y regresión

- Introducción: Asociación y Causalidad. *Hernández (2004): Capítulo 7*
- Pruebas estadísticas de asociación: para características cuantitativas (el modelo de correlación lineal y su interpretación, diagrama de dispersión) *Pérez (2005): Capítulo 4* y para características cualitativas: prueba de independencia. *Pérez (2005): Capítulo 6*
- Modelo de regresión simple y regresión múltiple: Cálculo e interpretación de la ecuación. Coeficiente de regresión y coeficiente de determinación. *Pérez (2005): Capítulo 5*
- Evaluación y uso de la ecuación de regresión. Algunas precauciones. *Pérez (2005): Capítulo 5*

Módulo 6: Elementos de probabilidad

- Concepto y definiciones de probabilidad: subjetiva, estadística y clásica.
- Propiedades básicas de la probabilidad. Tabla empírica de probabilidad. Probabilidad complementaria. Propiedad básica de multiplicación. Permutaciones.
- Eventos no excluyentes. Probabilidades conjuntas y marginales.
- Tabla de contingencia. Probabilidad condicional. Independencia
- Tabla dinámica para distribuciones de frecuencia combinada. *Pérez: 44-49*
- Referencias: *Gómez (1998)*

Módulo 7: Distribuciones de probabilidad

- Concepto de distribución de probabilidad.
- Tipos de distribuciones: discretas y continuas.
- La esperanza matemática (valor esperado) y la variancia.
- Distribución Binomial y de Poisson.
- Distribución Normal y normal estándar.
- Aproximación normal a la binomial.
- Referencias: *Gómez (1998)*

Módulo 8: Elementos de muestreo y estimación

- Inferencia estadística. Población y muestra.
- El censo y los estudios por muestreo, justificación.
- Diferencias entre un censo y un estudio por muestreo. Errores de muestreo y no de muestreo. La representatividad de la muestra.
- Censo y Encuestas INEC
- Conceptos teóricos de: parámetro, estimador, error de muestreo y sesgo de selección.
- Parámetros, estimadores y estimaciones. Los estimadores como variables.
- Estimación por intervalo de un promedio y de una proporción.
- El teorema del límite central.
- Intervalos de confianza para comparar poblaciones normales.
- El muestreo simple al azar sin reemplazo en poblaciones infinitas.
- Referencias: *Pérez (2005), Capítulos 10 y 13*

Módulo 9: Pruebas de significancia, análisis de varianza y regresión lineal

- Esquema general de las pruebas de significancia
- Posibilidad de error en la decisión. Nivel de significancia.
- Pruebas bilaterales respecto a un promedio o a una proporción.
- Comparación de una característica cuantitativa en dos poblaciones independientes.
- Comparación de una característica cuantitativa en dos poblaciones dependientes.
- Comparación de una característica cuantitativa en dos o más poblaciones independientes.
- Análisis de varianza en una vía,
- Análisis de varianza en dos factores
- Comparación de una característica cualitativa en dos o más poblaciones: prueba de homogeneidad.
- Referencias: *Pérez (2005), Capítulos 10, 11, 12 y 13*

Módulo 10: Levantamiento y recolección de datos

- Fuentes primarias y secundarias. *Gómez (1998): 29-33*
- Definición de problema, objetivos y población. Poblaciones finitas e infinitas. Estudio parcial de la población. Definición de las características a investigar y de sus categorías. Naturaleza de las características. *Hidalgo (2004): 11-15*
- Diseño de la investigación. Planeamiento de la recolección de los datos. *Hidalgo (2004): 1-6.*
- Modelaje de datos, Elaboración del cuestionario, Recolección y procesamiento
- Propuesta de análisis y divulgación

6. CRONOGRAMA

Las fechas propuestas para concluir cada módulo son las siguientes:

Módulo	Tema	Fecha de finalización
1	Conceptos generales de Estadística	15 de marzo
2	Distribución de frecuencias	22 de marzo
3	Medidas descriptivas	29 de marzo
4	Métodos de presentación de datos	12 de abril
5	Asociación estadística y regresión lineal	26 de abril
6	Elementos de probabilidad	3 de mayo
7	Distribuciones de probabilidad	10 de mayo
8	Elementos de muestreo y estimación	17 de mayo
9	Pruebas de significancia, análisis de varianza y regresión	31 de mayo
10	Levantamiento y recolección de datos	28 de junio

7. REFERENCIAS BIBLIOGRÁFICAS Y ENLACES

Cualquier libro de Estadística Descriptiva es útil para los módulos del 1 a 5 (primera parte del curso). Para la segunda parte del curso (Módulos 6-9) es apropiado cualquier libro introductorio a la inferencia estadística. Para la última parte, Módulo 10, usaremos una tesis de Maestría en Informática (Hidalgo, 2004), el manual del CPro y los materiales más recientes sobre el tema.

7.1. Libros y artículos

Berk, K. y P.Carey (2001). Análisis de datos con Microsoft EXCEL. Actualizado para Office 2000. Thomson Learning. Inc

Gómez, M. (1998). Elementos de Estadística Descriptiva. Editorial UNED, tercera edición.

Hidalgo Céspedes, J. (2004). *“Diseño de un modelo computacional y desarrollo de un prototipo para la recolección de datos de censos y encuestas con computador de mano.”* Tesis para optar por el grado de Masgíster Scientiae en Computación e Informática. Universidad de Costa Rica.

Hernández Rodríguez, Oscar. Estadística Elemental para Ciencias Sociales. Editorial Universidad de Costa Rica. 2002.

Pérez, C. (2005). Estadística Aplicada a través de Excel. Pearson Educación. Madrid, 2002. Última reimpresión.

7.2. Enlaces de cursos en la línea, material complementario (en revisión)

- Curso de Estadística para Informáticos, Profesor Carlomagno Araya, quien impartió el curso anteriormente www.geocities.ws/cmastat
- Curso de Estadística descriptiva. Este curso en español es gratuito, presenta en línea los conceptos básicos de la estadística descriptiva, probabilidades y distribuciones de probabilidad. <http://www.aulafacil.org/CursoEstadistica/CursoEstadistica.htm>
- Sitio en español (Universidad de Cádiz) que presenta problemas de Estadística Descriptiva Unidimensional y Bidimensional, Probabilidad. Variable aleatoria. Distribuciones de Probabilidad. Inferencia Estadística. Estimación y Contrastes de Hipótesis. Las respuestas están disponibles en línea. <http://thales.cica.es/rd/Recursos/rd98/Matematicas/19/matematicas-19.html>
- Enlace que presenta definiciones estadísticas básicas y ejemplos de aplicación <http://www.informatica.us.es/~calvo/alumnos/t02es2inf02/tema1.html>