



**XS-2130: MODELOS DE REGRESIÓN APLICADOS – Grupo 02**

**PROGRAMA  
II SEMESTRE 2017**

**Docente:** Gilbert Brenes Camacho

**Correo electrónico:**

[gilbert.brenes.camacho@gmail.com](mailto:gilbert.brenes.camacho@gmail.com)

**Oficina:** 18 Estadística.

**Teléfono:** 2511-6529

**Horario de clases:** L: 9:00-11:50 (340 CE)  
J: 9:00-10:50 (340 CE)

**Horas de consulta:** L: 15:30-17:30  
.....M: 17:00-18:00

**1. Descripción**

Los modelos de regresión son un tema central en la práctica de la estadística y forman la base de un amplio rango de métodos estadísticos. En este curso se presentarán las bases de los modelos de regresión múltiple y los procedimientos desarrollados para asegurar una aplicación correcta de los mismos.

Además de los conocimientos teóricos se brindará al estudiante la posibilidad de aplicar los métodos mediante el uso del lenguaje de programación R.

- **Requisitos:** XS-2110 Métodos Estadísticos, XS-2210 Estadística Computacional I y MA-1004.
- **Correquisitos:** XS-2330 Estadística Computacional II.
- **Horas:** 5 horas semanales (2 de teoría y 3 de práctica).
- **Créditos:** 4.

**2. Objetivo General**

Proveer a los estudiantes el conocimiento conceptual y los instrumentos estadísticos para la aplicación de los modelos de regresión lineal.

**3. Objetivos Específicos**

Al finalizar el curso el estudiante tendrá criterio y conocimiento básico para:

- Reconocer las situaciones donde se pueden aplicar las técnicas de regresión lineal.
- Utilizar un modelo con fines de predicción (un valor estimado individual, un valor estimado promedio).
- Usar diferentes técnicas de selección de variables para formar un modelo apropiado a partir de un conjunto de predictores disponibles.
- Llevar a cabo los diagnósticos del modelo de regresión ajustado.
- Aplicar las medidas apropiadas en los casos de no conformidad con los supuestos del modelo de regresión.
- Analizar casos con respuestas de dos categorías y conteos.





#### 4. Contenidos

<b>I. El modelo de regresión lineal.</b>	
1.1	El modelo y los supuestos.
1.2	Estimación de los coeficientes (notación matricial).
1.3	Valores ajustados y residuales.
1.4	Coefficientes estandarizados.
1.5	Inferencias. El cuadrado de medio de error. Varianzas de los estimadores. Intervalos de confianza para los coeficientes.
1.6	Intervalos de confianza para la respuesta media y predicción de nuevas observaciones.
1.7	Coefficientes de determinación (múltiple y parcial).
1.8	Comparación de modelos anidados con la distribución F. Uso de la distribución t.
1.9	Predictores cualitativos. Modelos con interacciones.
<b>II. Selección de predictores.</b>	
2.1	Procedimiento con todos los posibles subconjuntos: $R^2$ ajustado, suma de cuadrados de predicción (PRESS) y estadístico de Mallows ( $C_p$ ).
2.2	Selección por pasos: criterios de información (Akaike y Bayes).
<b>III. Diagnósticos.</b>	
3.1	Verificación de los supuestos: a. No autocorrelación: prueba de Durbin-Watson. b. Normalidad: QQ plot. c. Homoscedasticidad: prueba de Breush Pagan. d. No multicolinealidad: factor de inflación de la variancia (VIF). e. Linealidad entre predictor y respuesta: gráficos de regresión parcial.
3.2	Medidas para detectar valores extremos: residuales estudentizados, "leverage".
3.3	Medidas para detectar casos de influencia: DFFITs, distancia de Cook, DFBETAs.
<b>IV. Construcción del modelo.</b>	
4.1	Transformaciones de Box-Cox.
4.2	Mínimos cuadrados ponderados.
4.3	Modelo heteroscedástico por máxima verosimilitud
4.4	Componentes principales.
4.5	Regresión robusta.
4.6	Regresión polinomial.
4.7	Evaluación de la precisión en casos no estándar: bootstrap.
4.8	Validación del modelo.
<b>V. Modelos lineales generalizados.</b>	
5.1	Regresión de Poisson.
5.2	Regresión logística.





## 5. Metodología

- Presentaciones teóricas: se impartirán lecciones magistrales por parte del docente donde se explicarán los conceptos y sus aplicaciones.
- Prácticas: se usará el laboratorio de computadoras de la escuela de estadística para el desarrollo de lecciones prácticas durante el semestre. Durante las sesiones de laboratorio se utilizará el lenguaje de programación R para realizar ejercicios de la materia vista en clase. Se realizarán ejercicios y sus respectivas prácticas.
- Tareas y quizzes: se asignarán tareas y ejercicios para asegurar la “puesta en práctica” de los conceptos estudiados. Las tareas incluirán aplicaciones con datos para ser analizados con el programa elegido así como interpretaciones de los resultados. La realización de las tareas queda como responsabilidad de los estudiantes.
- Trabajo con datos reales:
  - Los estudiantes realizarán un trabajo de análisis de datos reales utilizando lo aprendido en el curso.
  - El trabajo será en grupos de 3 (máximo).
  - Las aplicaciones podrán ser tomadas de investigaciones en áreas como la economía, medicina, psicología, biología, población, etc.
  - Los estudiantes presentarán un anteproyecto donde expondrán los objetivos del estudio, el cual será criticado por el profesor y devuelto para que se realicen las mejoras sugeridas.
  - Al final de curso se espera que los estudiantes escriban un documento donde se describa detalladamente el procedimiento de análisis de los datos seleccionados.
  - El trabajo debe contar con una introducción donde el estudiante contextualiza los datos, lo cual le permite profundizar en el área de aplicación donde fue realizado el estudio.
  - Los resultados deberán ser expuestos en forma oral.
- Se usará como texto el libro **Applied Linear Statistical Models** de **Neter et al**, y el libro **Econometría**, de **Gujarati**.

## 6. Evaluación

Se realizarán tres exámenes parciales. Además los estudiantes presentarán un trabajo de análisis de datos, así como algunas tareas que deben entregar al asistente via correo electrónico.

Si un estudiante faltase a algún examen por **causa justificada**, debe solicitar por escrito la reposición del examen indicando las razones de la ausencia, acompañada de los documentos justificantes. La misma debe entregarse ante el profesor **a más tardar en cinco días hábiles del reintegro a lecciones**.

<b>Exámenes</b>	<b>70 %</b>
Primer examen parcial	20%
Segundo examen parcial	25%
Tercer examen parcial	25%
<b>Trabajos</b>	<b>30 %</b>
Trabajo de análisis de datos	20%
Trabajo s/ supuestos, posibles evaluaciones cortas	10%



## 7. Cronograma

	Módulo	L	J	Actividad	
AGOSTO		7		Presentación. Exposición temas 1.1 y 1.2	
	I. Modelo de regresión lineal	10		Exposición temas 1.1 -1.5.	
		14		Lab1	
		17		Exposición temas 1.5 - 1.7	
		21		Lab2.	
		24		Exposición temas 1.8 - 1.9	
		28		Lab3 y Lab 4	
			31	Exposición tema 2.1 y 2.2	
SETIEMBRE	II. Selección de predictores	4		Repaso examen	
		7		Exposición temas 2.1 y 2.2	
		Sab 9		<b>EXAMEN I: Capítulo I</b>	
			11		Lab5 y adicionales.
	III. Diagnósticos	14		Exposición temas 3.1: d,e	
		18		Práctica con R. Exposición temas 3.2 y 3.3	
21			Exposición temas 3.2 y 3.3		
		25		Lab 6	
			28	Exposición temas 4.1 y 4.2	
OCTUBRE	IV. Construcción del modelo	2		Repaso examen	
		5		Exposición temas 4.3 - 4-6	
		9		<b>EXAMEN II: capítulos II y III</b>	
		12		Exposición temas 4.6 - 4-7	
			16		<b>Feriado: Reemplazo por día de las Culturas (12 de oct)</b>
	V. Modelos lineales generalizados	19		Exposición tema 5.1	
23			Lab7		
26			Exposición tema 5.2		
30			Práctica con R		
			2		<b>Posible ausencia del profesor</b>
NOVIEMBRE		6		Lab8	
		9		Conclusión capítulo V.	
		13		Práctica con R	
		16		Entrega de trabajo final	
		20		<b>Práctica de examen</b>	
DIC		23		<b>PRESENTACION DE TRABAJOS</b>	
		27		<b>EXAMEN III: capítulos IV y V</b>	
		30		<b>PRESENTACION DE TRABAJOS</b>	
		7		Ampliación	

## 8. Referencias bibliográficas

- Chatterjee, S. y Hadi, A.S. (1988). Sensitivity Analysis in Linear Regression. New York, N.Y.: Wiley.  
BIBLIOTECA LUIS DEMETRIO TINOCO 519.536 Ch495s c.1
- Faraway, J. (2002). Practical Regression and Anova using R.  
Documento en internet: <http://www.biostat.jhsph.edu/~iruczins/teaching/jf/faraway.html>
- Faraway, J. (2004). Linear Models with R. Chapman & Hall/CRC.  
BIBLIOTECA LUIS DEMETRIO TINOCO 519.536 F219L
- Faraway, J. (2005). Extending the Linear Model with R: Generalized Linear, Mixed Effects and Nonparametric Regression Models. Chapman & Hall/CRC.  
BIBLIOTECA LUIS DEMETRIO TINOCO 519.536 F219e
- Fox, J. (2002). An R and S-Plus Companion to Applied Regression. Thousand Oaks, Cal.: Sage Publications, Inc. Disponible la 2a. edición (2011) : Fox, J y Weisberg, S: An R companion to applied regression. BIBLIOTECA LUIS DEMETRIO TINOCO 519.536.028.551.33 F792a2
- Gujarati, D. (1990). Econometría básica. McGraw-Hill.**



**BIBLIOTECA LUIS DEMETRIO TINOCO 330.015.195 G969e2**  
**Neter, J et al. (1996). Applied Linear Statistical Models. 4a. ed. WCB, McGraw-Hill. Disponible la**  
**5a. edición (2005): Kutner, M, Nachtsheim, C, Neter, J, Li, William: Applied Linear Statistical**  
**Models. BIBLIOTECA LUIS DEMETRIO TINOCO 519.536 K97a5**