



## SÍLABO

### SEMESTRE 2020-2

#### I. INFORMACIÓN GENERAL

Nombre del curso	:	Estadística para el Análisis Político 2
Código del curso	:	POL 304
Carácter	:	Obligatorio
Créditos	:	4
Número de horas de teoría	:	3
Número de horas de práctica	:	2
Profesor del curso	:	Dr. José Manuel MAGALLANES, PhD. Profesor de la Sección de Ciencia Política y Gobierno -Departamento de Ciencias Sociales Oficina 105 – Departamento de Economía/CCSS
Correo electrónico	:	<a href="mailto:jmagallanes@pucp.edu.pe">jmagallanes@pucp.edu.pe</a>
Horario de clases	:	Jueves de 9:00 a 12:00 (0689) / Lunes de 15:00 a 18:00 (0690)
Jefes de Práctica	:	Lic. Cristhian Arturo JARAMILLO <a href="mailto:cristhian.jaramillo@pucp.pe">cristhian.jaramillo@pucp.pe</a> Lic. Juan Carlos GONZALES <a href="mailto:jcgonzalez@pucp.edu.pe">jcgonzalez@pucp.edu.pe</a> Lic. Jair Almen ALVA <a href="mailto:jair.alva@pucp.edu.pe">jair.alva@pucp.edu.pe</a> Lic Christian Omar CHIROQUE <a href="mailto:christian.chiroque@pucp.edu.pe">christian.chiroque@pucp.edu.pe</a>
Horario prácticas:	:	Viernes de 8:00 a 10:00

#### II. SUMILLA

Se analizan las técnicas cuantitativas de análisis multivariable aplicadas a la investigación en ciencia política. Construcción de modelos cuantitativos para representar relaciones entre fenómenos políticos empíricos. Análisis de la varianza y correlación parcial. Análisis de regresión múltiple: interpretación de coeficientes, regresión con variables categóricas, diagnóstico de regresión, especificación de modelos. Introducción a la regresión logística. Técnicas para identificar categorías de análisis más o menos complejas: análisis discriminante, análisis factorial y análisis de conglomerados.

#### III. PRESENTACION

El curso completa la formación cuantitativa y computacional del alumno de pre grado, con énfasis en el desarrollo de proyectos concretos de análisis político cuantitativo. El entrenamiento que recibe le permitirá que se desempeñe como organizador, analista y comunicador político basado en datos. El aprendizaje de software como el R le permitirá llevar sus capacidades a cualquier entidad pública o privada, pues su trabajo no requerirá la compra de software licenciado.

#### IV. RESULTADOS DEL APRENDIZAJE

1. Recolección de información desde repositorios, APIs o páginas web.
2. Pre procesamiento y organización de datos.
3. Análisis estadístico multivariado descriptivo e inferencial

#### 4. Presentación de resultados.

### V. CONTENIDO DEL CURSO

#### Unidad 1: Colección y pre procesamiento de Datos

- Carga de datos desde la nube
- Pre procesamiento de datos
- Preparación de tablas de datos

#### Unidad 2: Descripción Multivariada de Indicadores complejos

- Conglomeración
  - Particionante
  - Aglomerativa y divisiva
  - Basada en densidad
- Factorización
  - Explorativa
  - Confirmativa

#### Unidad 3: Modelos Estadístico Inferenciales

- Regresión Lineal y sus diagnósticos
- Regresión Logística
- Alternativas a la regresión lineal

#### Unidad 4: Presentación de reportes estadísticos

- Estructura
- Gráficas
- Publicación

### VI. METODOLOGÍA

- Las clases teóricas discuten el uso y limitaciones de las técnicas presentadas.
- Se presentan las técnicas ya implementadas en funciones en R.
- Cada clase práctica desarrolla ejercicios.
- La evaluación es constante.
- Tanto en teoría como en práctica se les pedirá a los alumnos participar de manera activa.
- Durante clase se prohíbe el uso de teléfono celular, así como de acceso a redes sociales.
- La clase no se puede grabar, a menos que el profesor lo autorice.
- Los alumnos deben tener una laptop en cada sesión.
- Se recomienda que los alumnos creen una cuenta en <https://rstudio.cloud>. Si desean instalarlos en su propia máquina, deben:
  - Instalar R: <https://cran.r-project.org/>
  - Luego Rstudio: <https://www.rstudio.com/products/rstudio/download/>
- Todos deben tener cuenta en <https://github.com/>, el mismo que debe aprenderse de manera independiente.

### VII. EVALUACIÓN

Nº	Tipo de Evaluación	Ponderación sobre la nota final
1	PROYECTO APLICATIVO	20%
2	EXAMEN DE LABORATORIO	20%
3	EXAMEN INTEGRAL	40%
4	EVALUACION CONTINUA	20%

El proyecto aplicativo tiene 10 puntos a observar:

	Entrega	Semana Limite	Puntaje maximo a obtener
Pr1	Inscripción de Grupos Inscripción oportuna del grupo las dos primeras semanas. Se pierde este punto si el grupo se altera en el tiempo o si se conforma luego de fecha.	2	1
Pr2	Presentación de Datos por alumno en Repositorio en GitHub Cada alumno debe crear su propio Repositorio en Github, Los datos deben consistir en variables simples que representen en conjunto algun 'concepto macro' (o latente) de interés político, tecnologico, social, o similar. Se debe preparar un reporte de metadata, explicando tipo de escala y significado de cada variable simple, y justificación de porqué debe ser parte de la latente.	4	1
Pr3	Exploración de datos Los datos deben explorarse con mucho detalle, usando todos los conceptos de estadística 1 (univariado y bivariado)	7	1
Pr3	Presentación de Hipotesis y su operacionalización La Hipotesis es multivariada, y debe redactarse en términos de los conceptos latentes, debe además informarse cómo se contruyó cada latente.	11	1
Pr5	Análisis con Técnica Multivariada Descriptiva Usar y justificar las técnicas multivariadas descriptivas en el contexto de los objetivos e hipótesis del trabajo. Debe presentarse todo los temas del curso.	12	2
Pr6	Análisis con Modelo Inferencial Usar y justificar las técnicas multivariadas inferenciales en el contexto de los objetivos e hipótesis del trabajo. Debe presentarse todo los temas del curso.	13	2
Pr7	Poster En una sola infografía explicar lo planteado, desarrollado y concluido.	15	3
Pr8	Redacción Final Usar formato de artículo científico similar a (Mejia Guinand et al., 2008). Debe manejar lenguaje adecuado, poner lo más importante de lo antes entregado, y organizar adecuadamente los anexos con lo no incluido en el cuerpo principal del documento.	16	4
Pr9	Código reproducible Implica que el código debe estar ordenado, comentado, explicado en cada paso, describe las variables utilizadas, y poder ejecutarse sin problemas. Si el código NO se ejecuta, no se revisa y se obtienen la nota mínima (0).	16	5

Tome en cuenta:

- Para cada ítem, se dará los puntos según la calidad de la entrega (ello se precisa en cada entrega).
- Entrega fuera de fecha es CERO puntos (0)
- Alumno que abandone un grupo, pierde los puntos acumulados.
- **Los puntos se contabilizan al final del curso, si un proyecto no es expuesto ni entregado, la nota es cero, así haya acumulado puntos previamente.**

El proyecto es grupal y el tema se decide de manera independiente. Los grupos son de 4 alumnos. Cada alumno debe proponer y justificar una base de datos para construir una variable. Los datos para las variables deben provenir de algún sector del gobierno (subnacional, nacional o internacional), partido/movimiento político, u organismo de la sociedad civil interesada en temas públicos. Si se usan datos de gestión o políticas públicas, éstos deben ser mínimo a nivel provincial, para todas las provincias del Perú. Si se usan datos de encuestas de opinión, estos procurar ser de preguntas hechas en varios países.

Los Jefes de Práctica aclararán cualquier duda. TODOS ellos manejarán los mismo criterios.

1. El examen de laboratorio (EXLA) está basado sólo en ejercicios prácticos. Se permite uso de materiales.
2. El examen integral (EXIN) es principalmente teórico, por lo que no se permite uso de materiales. El profesor puede decidir incluir ejercicios, y de ser el caso, dividirá la hora del examen tal se pueda usar materiales en la parte de ejercicios.
3. En la evaluación continua, se toma al menos una práctica calificada (PC) por cada Unidad. Puede ser durante o fuera de horas de clase. Tienen control de tiempo.

**Por reglamento de la universidad, tener menos del 70% de asistencia a las prácticas genera desaprobación del curso.**

Tenga en cuenta que:

- Las asesorías se brindan por el profesor del curso y/o el jefe de practica.
- Las asesoría deben ser a al menos tres alumnos a la vez.
- Las asesorías individuales sólo se dan via ZOOM o MEET.
- El curso no contempla exámenes sustitorios ni de aplazados. Estos no se exigen bajo ningun criterio fuera de este silabo.

## VIII. BIBLIOGRAFIA

Chumney, F. L. (2014). *Principal Components Analysis , Exploratory Factor Analysis , and Confirmatory Factor Analysis*.

Kaufman, L., & Rousseeuw, P. J. (Eds.). (1990). Partitioning Around Medoids (Program PAM). En *Wiley Series in Probability and Statistics* (pp. 68-125). John Wiley & Sons, Inc. <https://doi.org/10.1002/9780470316801.ch2>

Lutins, E. (2017, septiembre 15). *DBSCAN: What is it? When to Use it? How to use it*. Medium.

<https://medium.com/@elutins/dbscan-what-is-it-when-to-use-it-how-to-use-it-8bd506293818>

Magallanes Reyes, J. M. (2017). *Introduction to data science for social and policy research: Collecting to organizing data with R and Python*. Cambridge University Press.

Mejia Guinand, L. B., Botero, F., & Rodriguez Raga, J. C. (2008). ¿Pavimentando con votos? Apropiación presupuestal para proyectos de infraestructura vial en Colombia, 2002-2006. *Colombia Internacional*, 68, 14-42. <http://www.redalyc.org/articulo.oa?id=81211204002>

Muñoz, P. (2009). *¿La Política Importa? Los determinantes políticos de la eficiencia del gasto municipal* (p. 65) [Informe]. Asociación Servicios Educativos Rurales.

Perez Liñan, A., & Altman, D. (2017). Explaining the Erosion of Democracy: Can Economic Growth Hinder Democracy? *SSRN Electronic Journal*. <https://doi.org/10.2139/ssrn.2929501>

Phillips, N. (2016, junio 11). *R Pubs. Introduction to Sweave (R and LaTeX)*. <https://rpubs.com/YaRrr/SweaveIntro>

Reddy, C. (2019, mayo 9). *Understanding the concept of Hierarchical clustering Technique*. Medium.

<https://towardsdatascience.com/understanding-the-concept-of-hierarchical-clustering-technique-c6e8243758ec>

Swaminathan, S. (2019, enero 18). *Logistic Regression—Detailed Overview*. Medium.

<https://towardsdatascience.com/logistic-regression-detailed-overview-46c4da4303bc>

Wooldridge, J. M. (2013). Limited Dependent Variable Models and Sample Selection Corrections. En *Introductory econometrics: A modern approach* (5th ed). South-Western Cengage Learning.

Xie, Y., Allaire, J. J., & Grolemund, G. (2018). *R Markdown: The definitive guide*. Taylor & Francis, CRC Press.

<https://bookdown.org/yihui/rmarkdown/>

## IX. CRONOGRAMA

Semana	Unidad	Tema	EVALUACION			Lectura
			Proyecto	Examen	Evaluación Continua	
1	I	Carga de datos desde la nube				(Magallanes Reyes, 2017)
2		Pre procesamiento de datos	Pr1			
3		Preparación de tablas de datos				
4	II	Conglomeración Particionante	Pr2			(Kaufman & Rousseeuw, 1990)
5		Conglomeración Aglomerativa y divisiva				(Reddy, 2019)
6		Conglomeración Basada en densidad				(Lutins, 2017)
7		Factorización Explorativa y Confirmativa	Pr3			(Chumney, 2014)
8						
9		Regresión Lineal y sus diagnósticos		EXLA		
10		Regresión Logística y sus diagnósticos				(Swaminathan, 2019)
11		Alternativas a la regresión lineal	Pr4			(Wooldridge, 2013)
12	IV	Estructura	Pr5			(Perez Liñan & Altman, 2017) (Muñoz, 2009)(Mejía Guinand et al., 2008)
13		Gráficas y Publicación	Pr6			(Phillips, 2016) (Xie et al., 2018)
14	Cierre de Curso			EXI		
15		Presentación de Poster	Pr7			
16		Presentación de Reporte	Pr8 – Pr9			

## **X. Del Profesor del Curso**

### **Dr. José Manuel Magallanes, Ph.D.**

Es PhD en Ciencias Sociales Computacionales por la George Mason University; Doctor en Psicología por la Universidad Nacional Mayor de San Marcos; Magister en Ciencia Política con mención en Gestión Pública por la Pontificia Universidad Católica del Perú; Maestría en Gestión Tecnológica en la Universidad Nacional de Ingeniería. Bachiller y Licenciado en Computación por la Universidad Nacional Mayor de San Marcos. Otros estudios diversos de postgrado en la Università degli Studi di Trento; National University of Singapore; Carnegie Mellon University; University of Chicago-Argonne National Lab; Kennedy School of Government-Harvard University; Universita Oberta de Catalunya.

Es Profesor de la Sección de Ciencia Política y Gobierno, así como de la Escuela de Gobierno y Políticas Públicas de la Pontificia Universidad Católica del Perú; Visiting Professor de la Evans School of Public Policy and Governance, así como Senior Data Scientist del eScience Institute, en la Universidad de Washington, Seattle. Fellow Catalyst de la U.C Berkeley Initiative for Transparency in the Social Sciences; Affiliate Researcher del Center for Social Complexity en George Mason University.

En el plano de investigación, el Dr. Magallanes trabaja en análisis de datos políticos para Perú (financiado por el eScience Institute de la Universidad de Washington) utilizando técnicas de análisis geoespaciales y de extracción de información desde usuarios de redes virtuales. Por otro lado, investiga las dinámicas de transfugismo legislativo (financiado por el Departamento de Computacional Social Science de la George Mason University) utilizando técnicas de análisis y modelamiento de grafos dinámicos a partir de las relaciones de co auspicio de iniciativas legislativas. Así mismo, viene trabajando el análisis de la dinámica decisional en sistemas de gobernanza local utilizando técnicas de simulación social (financiado por el Evans School of Government de la Universidad de Washington). Investiga además, con un enfoque prospectivo computacional, los efectos sociales y políticos del cambio climático en los Andes centrales (financiado por el Center for Social Complexity de la George Mason University y la National Science Foundation de EEUU).

Ha recibido en tres oportunidades el Premio a la Investigación en la PUCP. Recientemente, con financiamiento de la UC Berkeley, viene investigando y desarrollando material educativo para el uso de herramientas computacionales para mejorar la transparencia en la inversión en políticas públicas; trabajo que ha venido testeando en la PUCP, la Universidad de los Andes y la PUC de Chile.

Ha sido Director del Sistema Nacional de Bibliotecas de la Biblioteca Nacional del Perú; Gerente General de la Empresa Municipal de Servicios Informáticos de la Municipalidad de Miraflores; Responsable científico del proyecto INFOGOB del Jurado Nacional de Elecciones; así como consultor para diversas entidades públicas. En el 2020, ha sido miembro del equipo de modelamiento y análisis de datos de la pandemia COVID-19 ("Grupo Prospectiva") para el Gobierno del Perú.