

# Análisis Estadístico de Datos con R

Módulo I: Estadística Descriptiva y Probabilidad

## software descripción



Durante las últimas décadas se ha producido un desarrollo explosivo en las tecnologías de base de datos y la cantidad de datos que es recolectada. Esto ha creado una oportunidad sin precedentes para el uso de métodos estadísticos y de las ciencias de la computación en el análisis de datos; mediante el proceso de descubrimiento, ya sea supervisado o no, de información interesante y útil a partir de repositorios de datos disponibles.

Dentro de los distintos programas estadísticos disponibles, **R** proporciona una poderosa plataforma open source para la aplicación de las técnicas más importantes de la analítica de datos y es actualmente el programa más utilizado. R se caracteriza por incluir un mayor número de modelos estadísticos avanzados, aunque por lo general el manejo de código y la programación puede tener una curva de aprendizaje más lenta en comparación con otros programas y ser un reto para los analistas de datos que quieren utilizar esta herramienta. Ante ello, existen diversas herramientas como el entorno de desarrollo integrado (IDE) **RStudio** que adicionan una interfaz gráfica de usuario específicamente diseñada para facilitar la codificación y programación de los usuarios que no están acostumbrados al entorno de trabajo de este programa.

Este primer módulo del curso tiene como finalidad brindar los fundamentos sobre estadística descriptiva y probabilidad, necesarios en distintas áreas de estudio (como la Ciencia de Datos), usando el programa R a través del IDE RStudio. Se aprenderá a como acceder a una base de datos y utilizar una serie de métodos cualitativos y cuantitativos para resumirla y descubrir la información que se encuentra subyacente, así como la forma efectiva de comunicar estos hallazgos.

## publico objetivo

Analistas de datos. Profesionales relacionados a la inteligencia de negocios e investigación de mercados. Académicos e investigadores. Público en general que requiera aplicar técnicas de análisis de datos.

## contacto logros de aprendizaje



+51970227123



info@perustat.com



<http://www.perustat.com>



fb://perustat



Al finalizar este curso, el participante será capaz de pensar críticamente acerca de los datos, resumiéndolos mediante diferentes métodos descriptivos para extraer conclusiones a partir de estos análisis y facilitar la toma de decisiones con la ayuda del programa R y la interfaz RStudio. De manera específica el participante estará en capacidad de:

- Emplear R y RStudio como herramientas para el análisis de datos.
- Resumir la información contenida en bases de datos.
- Interpretar los resultados obtenidos luego de realizar un análisis descriptivo.
- Comprender y aplicar el concepto de probabilidad, así como sus principales propiedades.
- Identificar los principales modelos probabilísticos discretos y continuos.
- Presentar de manera efectiva los resultados obtenidos.

# contenidos

<b>Tema 1</b>	<b>Conceptos básicos</b> <ul style="list-style-type: none"><li>Definición de Estadística.</li><li>Población y muestra.</li><li>Tipos de datos.</li><li>Reconocimiento de Variables y su escala de Medición.</li><li>Medidas estadísticas: Parámetros y estadísticos.</li><li>Estudios experimentales y observacionales.</li><li>Introducción a R.</li><li>Uso de RStudio.</li><li><b>Taller 1:</b> Manejo de Datos con R</li></ul>	(8 horas)
<b>Tema 2</b>	<b>Exploración y Visualización de datos</b> <ul style="list-style-type: none"><li>Exploración de datos cualitativos.</li><li>Exploración de datos cuantitativos.</li><li>Medidas de resumen: Tendencia central, dispersión y asimetría.</li><li>Transformaciones: Puntuación Z.</li><li>Outliers y valores perdidos.</li><li><b>Taller 2:</b> Resumen y Visualización de datos con R (Parte 1)</li><li><b>Taller 3:</b> Resumen y Visualización de datos con R (Parte 2)</li></ul>	(16 horas)
<b>Tema 3</b>	<b>Probabilidad y Variables Aleatorias</b> <ul style="list-style-type: none"><li>Experimentos Aleatorios.</li><li>Probabilidad: Definiciones. Propiedades.</li><li>Probabilidad Condicional.</li><li>Teorema de Bayes.</li><li>Variables Aleatorias: Definición. Propiedades.</li><li>Modelos de probabilidad discretos. Distribución Binomial.</li><li>Modelos de probabilidad continuos. Distribución Normal</li><li><b>Taller 4:</b> Aplicaciones de cálculo de probabilidades.</li><li><b>Taller 5:</b> Modelos probabilísticos con R.</li></ul>	(16 horas)

## metodología

La metodología del curso se basa en una combinación de clases teóricas y análisis de casos prácticos en la computadora. Cada sección del curso está motivada por un conjunto de datos en particular, de tal forma que el participante gane experiencia trabajando con una amplia variedad de fuentes de datos similares a los que usa en la realidad. Los contenidos están estructurados en 10 sesiones con un total de 40 horas académicas.

## evaluación y asistencia

- Certificado de asistencia:** Para obtener este certificado debe de mantener un porcentaje mínimo de 75% de asistencia a las clases.
- Certificado de aprobación:** El 100% de la calificación final se obtiene sobre la base de pruebas objetivas o actividades colaborativas. Para recibir el certificado de aprobación el participante deben obtener al menos un 70% de los puntos posibles y contar con el porcentaje mínimo de asistencia a clases.

Serán otorgados certificados a nombre de PeruStat Analytics S.A.C. que acredita a los participantes del curso. La certificación que se otorga es excluyente.

## contacto



+51970227123



info@perustat.com



<http://www.perustat.com>



fb://perustat



## materiales

Material preparado por el equipo de capacitación con los contenidos del curso el cuál será entregado a los participantes en medios físicos o digitales.

## referencias

Anderson, D., Sweeney, D. and Williams, T. (2010). Estadística para administración y economía. Prentice Hall.

Dalgaard, P. (2008). Introductory Statistics with R. Springer.

De Veaux, R., Velleman, P. and Bock, D. (2013). Intro Stats. Pearson Education.

Heiberger, R. and Holland, B. (2015). Statistical Analysis and Data Display. An Intermediate Course with Examples in R. Springer.

Keen, K. J. (2010). Graphics for Statistics and Data Analysis with R. Chapman & Hall/CRC.

Teetor, P. (2011). R Cookbook (O'Reilly Cookbooks). O'Reilly Media

Zuur, A. F., Ieno, E. N., and Meesters, E. (2009). A beginner's guide to R. Springer

## contacto



+51970227123



info@perustat.com



<http://www.perustat.com>



fb://perustat



**PeruStat**

Analytics training, consulting  
& researching