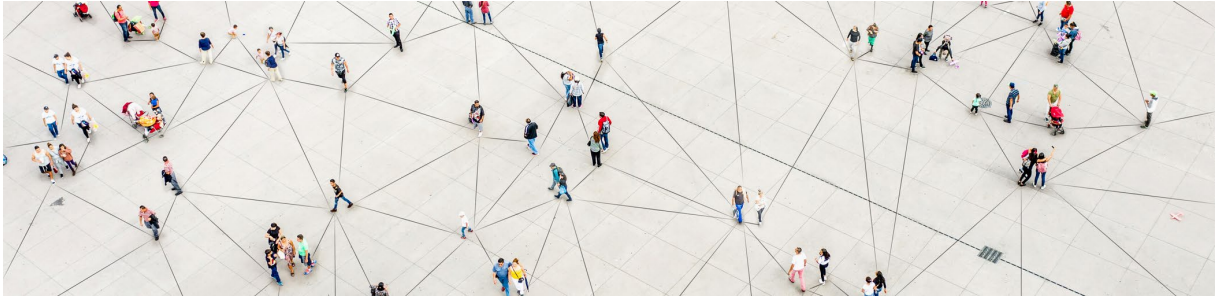




Bloque 3.

Tema 1:
Estadística
inferencial para
data science



Para entender con mayor facilidad cómo se utiliza la estadística inferencial, así como su relevancia en data science, enfatizaremos los siguientes conceptos:

Data science

Es la disciplina en la cual el científico de datos analiza los datos recolectados utilizando algoritmos estadísticos y potentes herramientas de cálculo para ejecutarlos (RedaccionMCP, 2019).

Estadística

De acuerdo con Roldán (s.f.), la estadística es una rama de las matemáticas que se ocupa de la obtención, orden y análisis de un conjunto de datos para obtener explicaciones y predicciones sobre fenómenos observados. Asimismo, consiste en métodos, procedimientos y fórmulas que permiten recolectar información para luego analizarla y extraer conclusiones. Su principal objetivo es mejorar la comprensión de los hechos a partir de la información disponible. Se divide en dos ramas: descriptiva e inferencial.

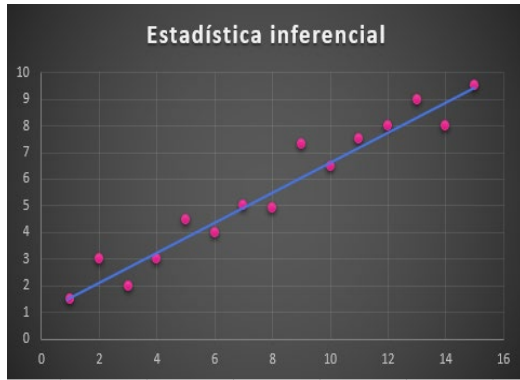
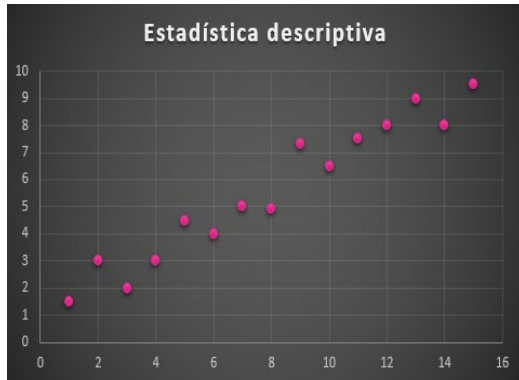
Estadística descriptiva

Es la encargada de obtener, organizar, mostrar y describir datos facilitando su comprensión a través del uso de tablas, gráficos, cálculo de dispersión y cálculo de tendencias. Algunos conceptos dentro de la estadística descriptiva son los siguientes: media, mediana, moda, rango, varianza, desviación estándar, visualización de datos, etcétera (Decena, 2019).

Estadística inferencial

Decena (2019) afirma que es la encargada de crear modelos predictivos basados en la información numérica de la muestra con el fin de llegar a una conclusión inferencial de la totalidad de la misma. Por lo que gracias a ella puedes llegar a formular predicciones basándote en la recolección de datos que obtuviste de la estadística descriptiva.

Por consiguiente, ambas ramas de la estadística se complementan, ya que la estadística descriptiva se encarga de recolectar los datos y graficarlos para un mejor entendimiento inicial, mientras que la estadística inferencial (utilizando diversos modelos) nos ayuda a comprender dichos datos con mayor exactitud.



La estadística inferencial, también llamada inferencia estadística, se relaciona con el cálculo de un parámetro basado en la estadística de una muestra, por ejemplo, si una muestra de 10 calculadoras solares TI-36X revela que dos estaban defectuosas, se puede inferir que el 20% de la producción es defectuosa (Lind, Marchal y Wathen, 2015).

De esta forma, la estadística inferencial nos muestra cómo podría ser el futuro basado en el comportamiento de datos anteriores, mientras que la estadística descriptiva nos presenta el comportamiento pasado de la muestra o de los datos extraídos.

Asimismo, se pueden emplear diversos modelos para que dichas inferencias o predicciones tengan mayor exactitud:

Regresión lineal

Técnica de pronóstico que supone que los datos del pasado y las proyecciones del futuro se ubican en torno a una línea recta. Es utilizada para pronósticos de series de tiempo y de relaciones causales. Por lo tanto, cuando la variable dependiente cambia como resultado del tiempo se trata de un análisis temporal, mientras que si esta cambia debido al cambio en otra, se trata de una relación causal (Chase, 2018). Existen los siguientes tipos de regresiones lineales:

Regresión lineal simple

Examina la relación lineal entre dos variables continuas: una respuesta (Y) y un predictor (X). Cuando las dos variables están relacionadas es posible predecir un valor de respuesta a partir de un valor predictor con una exactitud mayor que la asociada únicamente a las probabilidades.

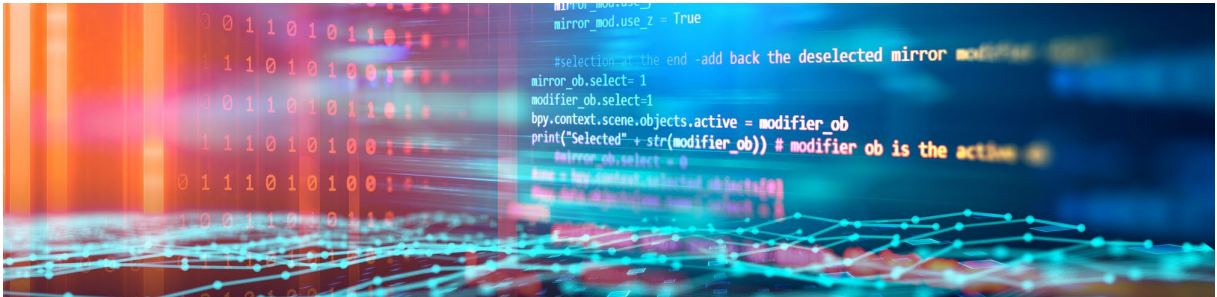
La regresión proporciona una línea que se ajusta mejor a los datos y que se puede utilizar después para lo siguiente:

- Examinar cómo cambia la variable de respuesta a medida que cambia la variable predictora.
- Predecir el valor de una variable de respuesta (Y) para cualquier variable predictora (X).

Regresión lineal múltiple

La regresión lineal múltiple examina las relaciones lineales entre una respuesta continua y dos o más predictores.

Si el número de predictores es grande (antes de ajustar un modelo de regresión con todos los predictores), se deberían utilizar las técnicas de selección de modelo paso a paso o de los mejores subconjuntos para excluir los predictores que no estén asociados con las respuestas.



Para saber más sobre la **estadística inferencial**, te invitamos a ver el siguiente video:

Los siguientes enlaces son externos a la Universidad Tecmilenio, al acceder a estos considera que debes apegarte a sus términos y condiciones.



Udacity. (2015, 23 de febrero). *Report Results - Intro to Inferential Statistics* [Archivo de video]. Recuperado de <https://www.youtube.com/watch?v=xNbzHOnSx5Q>

Para saber más sobre la **regresión lineal**, te invitamos a ver el siguiente video:



Brandon Foltz. (2013, 23 de noviembre). *Statistics 101: Linear Regression, The Very Basics* [Archivo de video]. Recuperado de <https://www.youtube.com/watch?v=ZkjP5RJLQF4>

Para saber más sobre la **regresión lineal múltiple**, te invitamos a ver el siguiente video:



Brandon Foltz. (2014, 1 de diciembre). *Statistics 101: Multiple Regression, The Very Basics* [Archivo de video]. Recuperado de <https://www.youtube.com/watch?v=dQNpSa-bq4M&t=12s>

Referencias

- Chase, R. (2018). *Administración de operaciones: producción y cadena de suministros* (15a. ed.). México: McGraw-Hill Interamericana. Libro disponible en Biblioteca Digital.
- Decena, E. (2019). *Estadística para Data Scientists*. Recuperado de <https://medium.com/@eddydecena/estadistica-para-data-scientists-574df6adb13a>
- Roldán, P. (s.f.). *Estadística*. Recuperado de <https://economipedia.com/definiciones/estadistica.html>
- Lind, D., Marchal, W., y Wathen, S. (2015). *Estadística aplicada a los negocios y a la economía* (16a. ed.). México: McGraw-Hill Interamericana. Libro disponible en Biblioteca Digital.
- RedaccionMCP. (2019). *Métodos estadísticos en Data Science: Saca el máximo valor a tus datos*. Recuperado de <https://www.muycmputerpro.com/2019/07/01/metodos-estadisticos-en-data-science-saca-el-maximo-valor-a-tus-datos>

Para expandir tu conocimiento, te recomendamos los siguientes recursos adicionales:

Los siguientes enlaces son externos a la Universidad Tecmilenio, al acceder a estos considera que debes apegarte a sus términos y condiciones.

Podcasts

Para saber más sobre **estadística y regresión lineal**, te recomendamos escuchar los siguientes podcasts:

- Michigan Tech. (2014). *FW5411 - Applied Regression Analysis* [Audio podcast]. Recuperado de <https://podcasts.apple.com/us/podcast/fw5411-applied-regression-analysis/id802041068>
- Renelle, T. (2017). *005 Linear Regression* [Audio podcast]. Recuperado de <https://player.fm/series/machine-learning-guide-1457335/5-linear-regression>

MOOC

Para aprender más sobre **estadística descriptiva e inferencial**, te recomendamos cursar los siguientes MOOC:

- Sistema Universitario de Georgia. (s.f.). *Six Sigma Advanced Defined and Measure Phases*. Recuperado de <https://es.coursera.org/lecture/six-sigma-define-measure-advanced/descriptive-vs-inferential-statistics-QBIS2>
- Duke University. (s.f.). *Inferential Statistics*. Recuperado de <https://www.coursera.org/learn/inferential-statistics-intro>
- Seoul National University. (s.f.). *Introductory Statistics: Sample Survey and Instruments for Statistical Inference*. Recuperado de <https://www.edx.org/es/course/introductory-statistics-sample-survey-and-instruments-for-statistical-inference>

- Berkeley. (s.f.). *Foundations of Data Science: Inferential Thinking by Resampling*. Recuperado de <https://www.edx.org/es/course/foundations-of-data-science-inferential-thinking-by-resampling-0>
- Udacity. (s.f.). *Intro to Inferential Statistics*. Recuperado de <https://www.udacity.com/course/intro-to-inferential-statistics--ud201>

La obra presentada es propiedad de ENSEÑANZA E INVESTIGACIÓN SUPERIOR A.C. (UNIVERSIDAD TECMILENIO), protegida por la Ley Federal de Derecho de Autor; la alteración o deformación de una obra, así como su reproducción, exhibición o ejecución pública sin el consentimiento de su autor y titular de los derechos correspondientes es constitutivo de un delito tipificado en la Ley Federal de Derechos de Autor, así como en las Leyes Internacionales de Derecho de Autor.

El uso de imágenes, fragmentos de videos, fragmentos de eventos culturales, programas y demás material que sea objeto de protección de los derechos de autor, es exclusivamente para fines educativos e informativos, y cualquier uso distinto como el lucro, reproducción, edición o modificación, será perseguido y sancionado por UNIVERSIDAD TECMILENIO.

Queda prohibido copiar, reproducir, distribuir, publicar, transmitir, difundir, o en cualquier modo explotar cualquier parte de esta obra sin la autorización previa por escrito de UNIVERSIDAD TECMILENIO. Sin embargo, usted podrá bajar material a su computadora personal para uso exclusivamente personal o educacional y no comercial limitado a una copia por página. No se podrá remover o alterar de la copia ninguna leyenda de Derechos de Autor o la que manifieste la autoría del material.