

Guía de Trabajo

Estadística Inferencial

Ybnias Eli Grijalva Yauri



Contenido

Presentación	5
Primera Unidad	7
Análisis exploratorio de datos y <i>software</i> especializado	
Semana 1: Sesión 2	
Datos: Variables categóricas y numéricas	8
Semana 2: Sesión 2	
Variables aleatorias discretas: distribuciones de probabilidad	10
Semana 3: Sesión 2	
Variables aleatorias continuas: Distribuciones de probabilidad	13
Semana 4: Sesión 2	
Aplicación de <i>software</i> especializado (R y Python)	15
Segunda Unidad	17
Muestreo y distribuciones de muestrales	
Semana 5: Sesión 2	
Tipos de muestreo probabilístico	18
Semana 6: Sesión 2	
Estimación: Propiedades de los estimadores puntuales	20
Semana 7: Sesión 2	
Estimación de intervalos de confianza – media y proporción	23
Semana 8: Sesión 2	
Estimación de intervalos de confianza – varianza	27
Tercera Unidad	30
Pruebas de hipótesis	
Semana 9: Sesión 2	
Prueba de hipótesis: población única	31
Semana 10: Sesión 2	
Prueba de hipótesis: población múltiple	36

Semana 11: Sesión 2	
Prueba de hipótesis: población múltiple	41
Semana 12: Sesión 2	
Análisis de varianza	48
Cuarta Unidad	53
Correlación y análisis de regresión	
Semana 13: Sesión 2	
Análisis de datos categóricos	54
Semana 14: Sesión 2	
Relación entre correlación y causalidad	57
Semana 15: Sesión 2	
Introducción al análisis de regresión	60
Semana 16: Sesión 2	
Aplicaciones de regresión y series de tiempo	67
Referencias	71

Presentación

La presente guía de trabajo de la asignatura Estadística Inferencia, ha sido elaborada con el propósito de cimentar adecuadamente los aspectos teóricos y la correspondiente práctica que se irá orientando a lo largo del desarrollo de la asignatura a través de cada una de las sesiones.

La guía esta estructura correspondiente a las unidades que desarrollaremos durante el semestre académico:

Unidad 1: Muestreo y distribuciones de muestrales, son dos componentes fundamentales de la formación de estudiantes, equipándolos con habilidades analíticas esenciales para su futura carrera profesional.

Unidad 2: Medidas de resumen o descriptiva, son herramientas que facilitan la comprensión y el análisis de datos, apoyan la toma de decisiones basadas en evidencia y son esenciales para una comunicación efectiva. Además, proporcionan una base sólida para el aprendizaje de métodos estadísticos y análisis más avanzados, contribuyendo al desarrollo de habilidades analíticas esenciales.

Unidad 3: Pruebas de hipótesis, son una herramienta indispensable que proporcionan un marco metodológico sólido para evaluar datos, validar teorías y tomar decisiones basadas en evidencia.

Unidad 4: Correlación y análisis de regresión, son técnicas esenciales que proporcionan herramientas para analizar datos de manera cuantitativa, entender relaciones entre variables, hacer predicciones y tomar decisiones informadas.

El resultado de aprendizaje de la asignatura busca que el estudiante sea capaz de analizar información de carácter probabilístico para el planteamiento de pronósticos de naturaleza organizacional y empresarial.

Asimismo, en la primera Unidad, el estudiante será capaz de analizar las variables categóricas y numéricas, a través de un software especializado en situaciones estadísticas planteadas. En la segunda Unidad, el estudiante será capaz de analizar los tipos de muestreo, utilizando métodos estadísticos apropiados en contextos reales de investigación. En la tercera Unidad, el estudiante será capaz de aplicar pruebas de hipótesis para poblaciones únicas, múltiples y el análisis de varianza basadas en la evidencia estadística. Y en la cuarta Unidad el estudiante será capaz de analizar la correlación y causalidad entre variables y la regresión en problemas prácticos en diversos campos.

Finalmente, se desea promover el uso del aula virtual donde se comparte todo el material de aprendizaje de la asignatura y se plasma la estrategia de aprendizaje colaborativo entre los estudiantes a través del siguiente esquema de trabajo:

- Lee todo el material (lecturas, PPT, guías de aprendizaje, videos, *pódcast*, etc.) correspondiente al tema de la semana.
- Participación activa en clase para la mejora del aprendizaje.
- Pregunta cuando tenga dudas o algo no haya quedado claro, esto será beneficioso para toda la clase en general.
- Utiliza la guía para completar la clase desarrollada de manera teórica.

Ybnias Eli Grijalva Yauri

Primera **Unidad**

**Análisis exploratorio de datos y
software especializado**

Semana 1: Sesión 2

Datos: Variables categóricas y numéricas

Sección: Fecha:/...../..... Duración: 60 minutos

Docente: Unidad: 1

Nombres y apellidos:

Instrucciones:

- a) Ingresar a la página de microdatos del INEI, seleccione la Encuesta Nacional de Hogares – ENAHO Metodología actualizada – Condiciones de Vida y Pobreza – ENAHO. <https://proyectos.inei.gob.pe/microdatos/>
- b) Seleccione y descargue uno de los siguientes módulos:
 - Características de la vivienda y del hogar
 - Educación
 - Salud
 - Gastos en alimentos y bebidas (Módulo 601)
 - Programas sociales (Miembros del Hogar)
 - Gobernabilidad, democracia y transparencia
- c) Trabajo a desarrollar a nivel grupal, organizados por el docente en la primera clase.
- d) Para la presentación de sus respuestas debe utilizar un archivo digital MS Word (doc.) y para el procesamiento de sus respuestas debe usar un archivo digital de R y Python. Debiendo presentar ambos archivos.

I. Propósito: Al finalizar la sesión, el estudiante analiza datos de variables categóricas y cuantitativas utilizando técnicas estadísticas adecuadas,

desarrollando habilidades para la interpretación y comunicación eficaz de los resultados en contextos académicos y profesionales.

II. Descripción de la actividad por realizar

1. En el módulo seleccionado y descargado ubique el archivo digital: diccionario.pdf
2. Identifica las variables del módulo y agruparlos según los tipos de variables estudiadas:
 - a) Variables categóricas – cualitativas
 - b) Nominal – ningún atributo
 - c) Ordinal – un atributo
 - d) Variables numéricas – cuantitativas
 - Intervalo – dos atributos
 - Razón – tres atributos
3. Por cada variable identificada, debe presentar la siguiente información:

Nombre Variable	Etiqueta	Tamaño	Decimal	Formato

4. Presenta tu trabajo en un archivo digital MS Excel.

Semana 2: Sesión 2

Variables aleatorias discretas: distribuciones de probabilidad

Sección: Fecha:/...../..... Duración: 60 minutos

Docente: Unidad: 1

Nombres y apellidos:

Instrucciones:

- a) Lee cuidadosamente el enunciado del caso «Gestión de los servicios de *Ashland Multicomm*».
- b) Responde las preguntas, teniendo en consideración los conceptos de probabilidad y distribuciones de probabilidad desarrollados en clase.
- c) Los resultados deben ser presentados en archivo digital *MS Word* con el mayor detalle posible, acompañado del archivo de *software* especializado utilizado para el procesamiento de sus datos.
- d) El trabajo a desarrollar será por equipos, organizados por el docente en la primera clase.
- e) Para la presentación de sus respuestas deben utilizar un archivo digital *MS Word* (doc.) y para el procesamiento de tus respuestas debes utilizar un archivo digital de *R* y *Python*. Deberás presentar ambos archivos.

- I. **Propósito:** Al finalizar la sesión, el estudiante analiza las variables aleatorias discretas, las distribuciones y estadísticas relacionadas para la toma de decisiones informadas.

II. Descripción de la actividad por realizar

Caso: Gestión de los servicios de Ashland Multicomm

El departamento de *marketing* de Ashland MultiComm Services (AMS) está buscando aumentar las suscripciones a su paquete combinado de telefonía, cable e internet. Para lograrlo, han estado llevando a cabo una campaña de *marketing* directo intensiva, que incluye correos postales, electrónicos y llamadas telefónicas. Los comentarios recibidos sugieren que la inclusión de canales premium en este paquete es un factor clave tanto para los clientes actuales como para los potenciales. Tras varias sesiones de *brainstorming*, el equipo de marketing ha decidido ofrecer canales de cable premium como un incentivo sin costo adicional al suscribirse al servicio. La directora de investigación, Mona Fields, tiene la intención de realizar una encuesta entre posibles clientes para determinar cuántos canales premium deberían añadirse para atraer nuevas suscripciones. Basándose en campañas anteriores y en datos de la industria, ella estima lo siguiente:

Number of Free Premium Channels	Probability of Subscriptions
0	0.02
1	0.04
2	0.06
3	0.07
4	0.08
5	0.085

1. Selecciona una muestra de 50 clientes potenciales y se les ofrece una suscripción, pero no incluye los canales premium gratuitos en la oferta de servicios, dados los resultados anteriores, cuál es la probabilidad de que:

- a) ¿Menos de 3 clientes se suscribirán a la oferta de servicio?
- b) ¿0 clientes o 1 cliente se suscribirán a la oferta de servicio?
- c) ¿Más de 4 clientes se suscribirán a la oferta de servicio?
- d) Suponiendo que, en la encuesta real de los 50 clientes potenciales, 4 clientes se suscriben a la oferta de servicio. ¿Qué te dice esto sobre la estimación anterior de la proporción de clientes que se suscribirían a la oferta de servicio?
2. En lugar de no ofrecer canales gratuitos premium como en el problema 1, supón que se incluyen dos canales premium gratuitos en la oferta de servicio. Dados los resultados pasados, ¿cuál es la probabilidad de que:
- a) ¿Menos de 3 clientes se suscribirán a la oferta de servicio?
- b) ¿0 clientes o 1 cliente se suscribirán a la oferta de servicio?
- c) ¿Más de 4 clientes se suscribirán a la oferta de servicio?
- d) Compara los resultados de «a» hasta «c» con los del problema 1.
- e) Suponiendo que, en la encuesta real de 50 clientes potenciales, 6 clientes se suscriben a la oferta de servicio. ¿Qué te dice esto sobre la estimación anterior de la proporción de clientes que se suscribirían a la oferta de servicio 3 por todos?
3. Suponiendo que se realizaron encuestas adicionales a 50 clientes potenciales en las que varía el número de canales premium gratuitos. Los resultados fueron los siguientes:

Number of Free Premium Channels	Probability of Subscriptions
1	5
3	6
4	6
5	7

¿Cuántos canales premium gratuitos debería recomendar el director de investigación para su inclusión en el servicio? Explica.

Semana 3: Sesión 2

Variables aleatorias continuas: Distribuciones de probabilidad

Sección: Fecha:/...../..... Duración: 60 minutos

Docente: Unidad: 1

Nombres y apellidos:

Instrucciones:

- a) Lee cuidadosamente el enunciado del caso.
- b) Responde las preguntas, teniendo en consideración los conceptos de probabilidad y distribuciones de probabilidad desarrollados en clase.
- c) Los resultados deben ser presentados en archivo digital *MS Word* con el mayor detalle posible, acompañar el archivo de *software* especializado utilizado para el procesamiento de sus datos.
- d) El trabajo a desarrollar será por equipos, organizados por el docente en la primera clase.
- e) Para la presentación de sus respuestas debe utilizar un archivo digital *MS Word* (*doc.*) y para el procesamiento de tus respuestas debes utilizar un archivo digital en *R* y *Python*. Debiendo presentar ambos archivos.

I. Propósito: Al finalizar la sesión, el estudiante maneja las variables aleatorias continuas, las distribuciones de probabilidad para la toma de decisiones informadas en la interpretación y comunicación de los resultados.

II. Descripción de la actividad por realizar

Enunciado del caso de evaluación

Un investigador quiere evaluar el peso promedio de una especie de peces en un lago. Para ello, toma una muestra aleatoria de 50 peces y registra sus pesos en kilogramos.

Pesos de la muestra (kg)				
2.7	3.1	2.9	2.9	2.8
3.1	2.7	2.8	2.8	2.9
2.8	3	2.9	3	3.1
2.9	2.8	3.1	2.7	2.7
3	2.9	2.7	2.9	3
2.6	3.2	3	3.2	2.9
2.8	2.6	2.8	2.6	2.8
3.2	2.8	2.6	2.8	2.7
2.7	2.7	2.7	2.7	3.1
2.9	3	3.1	3	2.8

Se sabe que los pesos de esta especie de peces siguen una distribución normal

- Calcular el peso promedio de los peces.
- Determinar la varianza y la desviación estándar del peso.
- Calcular la probabilidad de que el peso de un pez sea superior al promedio.
- Calcular la probabilidad de que el peso de 5 peces sea superior al promedio.
- Calcular la probabilidad de que el peso de un pez sea inferior al promedio.
- Determinar la probabilidad de que el peso de 8 peces es menor al

promedio.

Semana 4: Sesión 2

Aplicación de software especializado (R y Python)

Sección: Fecha:/...../..... Duración: 60 minutos

Docente: Unidad: 1

Nombres y apellidos:

Instrucciones

- a) Ingresa a la página de *Kaggle*, www.kaggle.com. Si ingresas por primera vez, regístrate con el *email* institucional.
- b) En la sección *Datasets*, deberás seleccionar una base de datos.
- c) En la base de datos selecciona: dos variables cualitativas y dos variables cuantitativas.
- d) El trabajo a desarrollar será por equipos utilizando el *software* especializado (R y *Python*).
- e) Para la presentación de tus respuestas debes utilizar un archivo digital MS *Word* (doc.) y para el procesamiento de las respuestas debe usar un archivo digital de R y *Python*. Debes presentar ambos archivos.

I. Propósito: Al finalizar la sesión, el estudiante utiliza R y Python para el análisis estadístico para la interpretación de los resultados y la comunicación eficaz de los hallazgos.

II. Descripción de la actividad por realizar

1. Utiliza las variables cualitativas seleccionadas y realiza las siguientes actividades:
 - Calcula los estadísticos de tendencial central, de dispersión, y de forma. (de estar permitido).
 - Realiza cálculos de probabilidades (mayor y menor) relacionados con el promedio.

2. Utiliza las variables cuantitativas seleccionadas y realiza:
 - Calcula los estadísticos de tendencial central, de dispersión, y de forma.
 - Realiza cálculos de probabilidades (mayor y menor) relacionados con el promedio.

3. El desarrollo del caso debe ser presentado en un archivo digital *Word*, acompañado con el archivo digital del *software* especializado para el desarrollo de los cálculos.

Segunda **Unidad**

**Muestreo y distribuciones de
muestrales**

Semana 5: Sesión 2

Tipos de muestreo probabilístico

Sección: Fecha:/...../..... Duración: 60 minutos

Docente: Unidad: 2

Nombres y apellidos:

Instrucciones

- a) Ingresa a la página de Kaggle, www.kaggle.com.
- b) En la sección *Datasets*, deberás seleccionar una base de datos, la misma que será utilizada para determinar los tipos de muestra.
- c) El trabajo a desarrollar será por equipos utilizando el *software* especializado (R y *Python*).
- d) Para la presentación de tus respuestas debes utilizar un archivo digital MS *Word* (doc.) y para el procesamiento de sus respuestas debe utilizar un archivo digital de R y *Python*. Debes de presentar ambos archivos.

I. **Propósito:** Al finalizar la sesión, el estudiante identifica el muestreo de poblaciones, distribuciones muestrales y el teorema del límite central para el diseño de estudios de muestreo, interpretando las distribuciones

y aplicando las inferencias sobre poblaciones.

II. Descripción de la actividad por realizar

1. Con la base de datos seleccionar una base de datos y determinando una variable de referencia para cada tipo de muestro, calcular:
 - Muestreo aleatorio simple sin reposición
 - Muestreo aleatorio simple con reposición
 - Muestreo sistemático
 - Muestreo estratificado
 - Muestreo por conglomerados
 - Muestreo polietápico

2. Cada equipo debe presentar un archivo digital en *MS Excel*, la cual debe contener los integrantes de los diversos tipos de muestras. (utilizar una hoja del *Excel* para cada tipo) y acompañada con el archivo digital del *software* especializado (*R* y *Python*).

Semana 6: Sesión 2

Estimación: Propiedades de los estimadores puntuales

Sección: Fecha:/...../..... Duración: 60 minutos

Docente: Unidad: 2

Nombres y apellidos:

Instrucciones:

- a) Lee cuidadosamente cada uno de los enunciados y responde las preguntas.
- b) Para la presentación de tus respuestas debes utilizar un archivo digital MS Word (doc.) y para el procesamiento de tus respuestas debes utilizar un archivo digital de R y Python. Debes presentar ambos archivos.
- c) Ingresar a la página de microdatos del INEI, <https://proyectos.inei.gob.pe/microdatos/>
- d) Ingresa a la página de Kaggle, www.kaggle.com.

- I. Propósito:** Al finalizar la sesión, el estudiante identifica las propiedades de los estimadores puntuales en la inferencia estadística, como el sesgo, la consistencia, la eficiencia y la suficiencia, y aplicar estos conceptos en la selección y evaluación de estimadores.

II. Descripción de la actividad por realizar

1. Es preocupante el incremento de la velocidad de los vehículos que transitan por una determinada parte de una carretera. El radar señala las siguientes velocidades, en kilómetros por hora, de una muestra aleatoria de 28 vehículos:

59 63 68 57 56 71 59 69 53 58 60 66 51 59

54 64 58 57 66 61 65 70 63 65 57 56 61 59

- a) Investiga las evidencias de la falta de normalidad.
 - b) Encuentra una estimación puntual de la media poblacional que sea insesgada y eficiente.
 - c) Emplea un método de estimación insesgada para encontrar una estimación puntual de la varianza de la media muestral.
2. Se coge una muestra aleatoria de ocho casas en un barrio, cuyos costos de venta fueron los siguientes (en miles de dólares):

192 183 312 227 309 396 402 390

- a) Investiga prueba de la falta de normalidad.
- b) Halla una estimación puntual de la media poblacional que sea imparcial y eficiente.
- c) Usa un método de estimación imparcial para localizar una estimación puntual de la varianza de la media muestral (sugerencia: aplica la desviación estándar muestral para estimar la desviación estándar poblacional).

d) Usa un estimador imparcial para medir la proporción de casas en este barrio que se venden por debajo de 250.500 \$.

3. Una muestra aleatoria de diez economistas ha realizado las siguientes valoraciones del crecimiento porcentual del Producto Interno Bruto para el siguiente año:

2,2 2,8 3,0 2,5 2,4 2,6 2,5 2,4 2,7 2,6

Usa métodos de estimación imparcial para alcanzar estimaciones puntuales de:

- a) La media poblacional.
 - b) La varianza poblacional.
 - c) La varianza de la media muestral.
 - d) La proporción de economistas en la población que ha precedido un crecimiento del Producto Interno Bruto de al menos un 2,5 %.
4. Se seleccionó una muestra aleatoria de 12 trabajadores de una gran industria y se consiguió las siguientes cifras sobre el número de horas extras realizadas en el anterior mes:

22 16 28 12 18 36 23 11 41 29 26 31

Usa las técnicas de estimación imparcial para calcular estimaciones puntuales de:

- a) La media poblacional.
- b) La varianza poblacional.
- c) La varianza de la media muestral.
- d) La proporción de trabajadores en la población que laboraron más de 30 horas extras en esta empresa el mes anterior.

Semana 7: Sesión 2

Estimación de intervalos de confianza – media y proporción

Sección: Fecha:/...../..... Duración: 60 minutos

Docente: Unidad: 2

Nombres y apellidos:

Instrucciones:

- a) Lee cuidadosamente cada uno de los enunciados y responde las preguntas.
- b) Para la presentación de tus respuestas debes utilizar un archivo digital MS Word (doc.) y para el procesamiento de tus respuestas debes utilizar un archivo digital de R y Python. Debes presentar ambos archivos.
- c) Ingresar a la página de microdatos del INEI, <https://proyectos.inei.gob.pe/microdatos/>
- d) Ingresa a la página de Kaggle, www.kaggle.com.

- I. **Propósito:** Al finalizar la sesión, el estudiante calcula intervalos de confianza para la media y proporción para la toma de decisiones basadas y en la comunicación de sus hallazgos de manera efectiva.

II. Descripción de la actividad por realizar

1. Se conoce que la desviación típica de los volúmenes de los envases de 591 mililitros de agua mineral embotellada por una fábrica es de cinco mililitros. Se ha tomado una muestra aleatoria de 100 envases y se han medido.
 - a) Evalúa el error típico de la media.
 - b) Halla el margen de error de un intervalo de confianza al 90 % de la media poblacional de los volúmenes.
 - c) Evalúa la amplitud de un intervalo de confianza al 98 % de la media poblacional de los volúmenes.

2. La secretaría de admisiones de un programa de maestría en administración de empresas se ha dado cuenta de que, históricamente, las notas medias de los interesados en sus estudios de grado siguen una distribución normal con una desviación estándar de 0,45. Este año, se sacó una muestra aleatoria de 25 interesados, y la calificación media de esta muestra fue de 2,90.
 - a) Halla un intervalo de confianza del 95 % para la media poblacional.
 - b) Usando estos resultados muestrales, un estadístico precisa un intervalo de confianza para la media poblacional que va de 2,81 a 2,99. Halla el nivel de confianza asociado a este intervalo.

3. Un director de personal se ha percatado que, históricamente, las

puntuaciones de los test de aptitud realizados por los interesados en el empleo para niveles de entrada siguen una distribución normal con una desviación estándar de 32,4 puntos. Una muestra aleatoria de nueve puntuaciones del grupo actual de solicitantes arrojó una media de 187,9 puntos.

- a) Señalar un intervalo de confianza del 80 % para la media poblacional de las puntuaciones del grupo actual de interesados.
 - b) Partiendo de estos resultados muestrales, un estadístico ha calculado un intervalo de confianza para la media poblacional que va de 165,8 a 210,0 puntos. Determina el nivel de confianza de este intervalo.
4. Define el factor de fiabilidad, $z_{(\alpha/2)}$, para conocer la media, k , de una población que sigue una distribución normal con varianza poblacional conocida en cada uno de los casos siguientes:
- a) Un nivel de confianza del 93 %
 - b) Un nivel de confianza del 96 %
 - c) Un nivel de confianza del 80 %
5. Encuentra en cada uno de los casos presentados el factor de fiabilidad, $z_{\frac{\alpha}{2}}$, para estimar la media, k , de una población que sigue una distribución normal que tiene una varianza poblacional conocida:
- a) $\alpha = 0.08$.
 - b) $\frac{\alpha}{2} = 0.02$.
6. Partiendo de una distribución normal que tiene una varianza poblacional conocida. Halla la amplitud para calcular la media poblacional, k , en los siguientes casos:
- a) Un nivel de confianza del 90 %, $n = 100, \sigma^2 = 169$.
 - b) Un nivel de confianza del 95 %, $n = 120, \sigma = 25$.

7. Considerando que se encuestó a una muestra aleatoria de 142 encargados de admisiones en programas de posgrado sobre la importancia de las puntuaciones en exámenes normalizados en la evaluación de un postulante. 87 respondieron que estas calificaciones desempeñan «un papel primordial». Halla un intervalo de confianza del 95 % para la proporción poblacional de responsables que comparten esta opinión.
8. En una muestra aleatoria de 95 empresas manufactureras, 67 han señalado que lograron la certificación ISO en los últimos dos años. Halla un intervalo de confianza del 99 % para la proporción de empresas en la población que han recibido esta certificación en los dos últimos años.
9. Considerando que las autoridades locales en una zona residencial densamente poblada de Hong Kong están evaluando la posibilidad de construir una nueva alberca municipal y un centro de esparcimiento. Dado que el costo de la construcción sería alto, primero se debe determinar si los residentes creen que el gasto público en estas instalaciones está justificado. Si 243 de los 360 residentes en una muestra aleatoria opinan que se debe construir la alberca y el centro de esparcimiento, halla la proporción de todos los vecinos que apoyarían la propuesta con un nivel de confianza del 95 %
10. Es crucial que las aerolíneas cumplan con los horarios de salida programados para los vuelos. Suponiendo que una aerolínea ha revisado recientemente las horas de partida de una muestra aleatoria de 246 vuelos y ha encontrado que diez vuelos se demoraron por el

mal clima, cuatro por razones de mantenimiento, y el restante partió sin demora.

- a) Halla el porcentaje de vuelos que salieron a tiempo con un nivel de confianza del 98 %
- b) Halla el porcentaje de vuelos que demoraron por un mal clima con un nivel de confianza del 98 %

Semana 8: Sesión 2

Estimación de intervalos de confianza – varianza

Sección: Fecha:/...../..... Duración: 60 minutos

Docente: Unidad: 2

Nombres y apellidos:

Instrucciones:

- a) Lee cuidadosamente cada uno de los enunciados y responde las preguntas.
- b) Para la presentación de tus respuestas debes utilizar un archivo digital MS Word (doc.) y para el procesamiento de tus respuestas debes utilizar un archivo digital de R y Python. Debes presentar ambos archivos.
- c) Ingresar a la página de microdatos del INEI,
<https://proyectos.inei.gob.pe/microdatos/>

d) Ingresa a la página de Kaggle, www.kaggle.com.

I. Propósito: Al finalizar la sesión, el estudiante calcula intervalos de confianza para la varianza mediante la aplicación de los conceptos para la toma de decisiones basadas en datos y en la comunicación de sus hallazgos de manera efectiva.

II. Descripción de la actividad por realizar

1. LDS desea asegurarse de que la tasa de incidencia de fugas (en centímetros cúbicos por segundo) de los enfriadores de aceite de la transmisión (TOC) que cumpla con los límites establecidos. Se estudia una muestra aleatoria de 50 TOC y se registran las tasas de casos de fugas en el archivo de datos llamado TOC. Halla la varianza de la tasa de incidencia de fugas con un nivel de confianza del 95 % (verifica la normalidad).
2. Un hospital presenta un programa de adelgazamiento. Según sus registros, una muestra aleatoria de diez pacientes presentó las siguientes reducciones de peso al concluir el programa:

18,2 25,9 6,3 11,8 15,4 20,3 16,8 19,5 12,3 17,2

Halla un intervalo de confianza del 90 % para la varianza poblacional de adelgazamiento de los usuarios de este programa de disminución de peso.

3. El director de control de calidad de una empresa química ha recogido una muestra aleatoria de veinte sacos de fertilizante de 100 kilos para estimar la varianza en los kilos de contaminantes. La varianza muestral observada es de 6,62. Halla un intervalo de confianza del 95 % para la

varianza poblacional de los kilos de contaminantes.

4. Un psicólogo quiere valorar la varianza de las puntuaciones de los trabajadores en un test. Se ha tomado una muestra aleatoria de 18 puntuaciones, y la desviación estándar muestral es de 10,4. Halla un intervalo de confianza del 90 % para la varianza poblacional. ¿Qué supuestos, si los hay, se deben cumplir para hallar este intervalo?

5. Un empresario está preocupado por la variabilidad en los niveles de contaminantes presentes en los mandados de una materia prima de un distribuidor. Se ha cogido una muestra aleatoria de 15 mandados y la desviación estándar en la concentración de impurezas observada es de 2,36. Considera que la población sigue una distribución normal.
 - a) Calcula un intervalo de confianza del 95 % para la varianza poblacional.
 - b) ¿Sería un intervalo de confianza del 99 % para esta varianza más amplia o más estrecha que el logrado en el apartado (a)?

6. Un empresario se especializa en revestir superficies metálicas con plástico. Se ha cogido una prueba aleatoria de nueve mediciones del espesor del revestimiento plástico de la producción de una semana; las mediciones del espesor (en milímetros) son las siguientes:

19,8 21,2 18,6 20,4 21,6 19,8 19,9 20,3 20,8

Estimar un intervalo de confianza del 90 % para la varianza poblacional, asumiendo que la población sigue una distribución normal.

Tercera **Unidad**

Pruebas de hipótesis

Semana 9: Sesión 2

Prueba de hipótesis: población única

Sección: Fecha:/...../..... Duración: 60 minutos

Docente: Unidad: 3

Nombres y apellidos:

Instrucciones:

- a) Lee cuidadosamente cada uno de los enunciados y responde las preguntas.
- b) Para la presentación de tus respuestas debes utilizar un archivo digital MS Word (doc.) y para el procesamiento de tus respuestas debes utilizar un archivo digital de R y Python. Debes presentar ambos archivos.
- c) Ingresar a la página de microdatos del INEI, <https://proyectos.inei.gob.pe/microdatos/>
- d) Ingresar a la página de Kaggle, www.kaggle.com.

I. Propósito: Al finalizar la sesión, el estudiante interpreta pruebas de hipótesis basadas en los resultados en una población única para la toma de decisiones y en la comunicación de sus hallazgos de manera efectiva a audiencias técnicas y no técnicas.

II. Descripción de la actividad por realizar

1. Se obtiene una muestra aleatoria de una población que tiene una varianza $\sigma^2 = 625$ y se calcula la media muestral. Contrasta la hipótesis nula $H_0: \mu = 100$ frente a la hipótesis alternativa $H_1: \mu > 100$ con $\alpha = 0.05$. Calcula el valor crítico \underline{x}_c y formula tu regla de decisión para las siguientes opciones. (Tomado de Newbold, Carlson & Thorne, 2022, cap. 9)
 - a) Tamaño de la muestra $n = 25$.
 - b) Tamaño de la muestra $n = 16$.
 - c) Tamaño de la muestra $n = 44$.
 - d) Tamaño de la muestra $n = 32$.
2. Un fabricante de detergente afirma que el peso promedio de las cajas que vende es de al menos 160 gramos. Se sabe que el peso sigue una distribución normal con una desviación estándar de 40 gramos. Una muestra aleatoria de 16 cajas mostró un peso promedio muestral de 158,4 gramos. Realiza una prueba de hipótesis con un nivel de significación del 10 % para evaluar la hipótesis nula de que la media poblacional del peso es al menos 160 gramos. (Tomado de Newbold, Carlson & Thorne, 2022, cap. 9)
3. Una compañía que recibe envíos de pilas examina una muestra aleatoria de nueve antes de aceptar el envío. La compañía requiere

que la durabilidad media real de todas las pilas en el envío sea de al menos 50 horas. Según su práctica, la durabilidad de las pilas sigue una distribución normal con una desviación estándar de tres horas. La durabilidad media de una muestra de nueve pilas del envío es de 48,2 horas. Desarrolla una prueba de hipótesis con un nivel de significación del 10 % para evaluar la hipótesis nula de que la media poblacional de la durabilidad es al menos 50 horas. (Tomado de Newbold, Carlson & Thorne, 2022, cap. 9)

4. Una compañía farmacéutica quiere que la concentración de impurezas en sus píldoras no supere el 3 %. Se conoce que la concentración de impurezas en un lote sigue una distribución normal con una desviación estándar de 0,4 %. Se recoge una muestra aleatoria de 64 píldoras del lote y se encuentra que la media muestral de la concentración de impurezas es de 3,07 %. (Tomado de Newbold, Carlson & Thorne, 2022, cap. 9)
 - a) Desarrolla una prueba de hipótesis con un nivel de significación del 5 % para evaluar la hipótesis nula de que la media poblacional de la concentración de impurezas es del 3 %, en contraste con la hipótesis alternativa de que es superior al 3 %.
 - b) Halla el valor p para esta prueba de hipótesis.
 - c) Considerando que la hipótesis alternativa hubiera sido bilateral en lugar de unilateral, con la hipótesis nula $H_0: \mu=3$. Señala sin realizar cálculos si el valor p del contraste sería mayor, menor o igual al obtenido en el apartado (b). Gráfica tu razonamiento.
 - d) Justifica el porqué, en este problema, es más apropiada una hipótesis alternativa unilateral en lugar de una bilateral.

5. Se solicita a una muestra aleatoria de 172 estudiantes de marketing

que califiquen en una escala del 1 (nada importante) al 5 (muy importante) la importancia de los servicios sanitarios complementarios como característica del puesto. La media muestral de las calificaciones es 3,31 y la desviación estándar muestral es 0,70. Realiza una prueba de hipótesis con un nivel de significación del 1 % para evaluar la hipótesis nula de que la media poblacional de las valoraciones es como máximo 3,0, en contraste con la hipótesis alternativa de que es superior a 3,0. (Tomado de Newbold, Carlson & Thorne, 2022, cap. 9)

6. Se presenta a una muestra aleatoria de 170 individuos un problema de predicción. Cada participante debe prever el próximo valor de una variable relacionada con las ventas al por menor de dos maneras: viendo los 20 valores anteriores tanto en formato numérico como en gráfico. Luego, se les solicita que pronostiquen el siguiente valor y se registran los errores absolutos de sus predicciones. La muestra consiste en 170 diferencias entre los errores absolutos de las predicciones basadas en números y las basadas en gráficos. La media muestral de estas diferencias es -2,91 y la desviación estándar muestral es 11,33. Calcula e interpreta el valor p para una prueba de hipótesis donde la hipótesis nula es que la media poblacional de las diferencias es 0, frente a la hipótesis alternativa de que es negativa (lo que sugiere que, en general, las personas hacen mejores predicciones cuando utilizan gráficos históricos en lugar de valores numéricos). (Tomado de Newbold, Carlson & Thorne, 2022, cap. 9)

7. En una muestra aleatoria de 361 dueños de pequeñas compañías que se habían declarado en bancarrota, 105 indicaron que no realizaron ningún estudio de mercado antes de iniciar la transacción. Realiza

una prueba de hipótesis con un nivel de significación del 5 % para evaluar la hipótesis nula de que el porcentaje máximo de todos los miembros de esta población que no realizaron estudios de mercado es del 25 %. (Tomado de Newbold, Carlson & Thorne, 2022, cap. 9)

8. En una muestra aleatoria de 360 directivos de exportación de Gran Bretaña, 69 manifestaron cierta disconformidad con la afirmación: «En 10 años, el mercado de exportaciones más importante para los empresarios británicos será el continente asiático». Realiza una prueba de hipótesis con un nivel de significación del 5 % para evaluar la hipótesis nula de que al menos el 25 % de todos los integrantes de esta población estaría en contra con dicha aseveración. (Tomado de Newbold, Carlson & Thorne, 2022, cap. 9)

9. En una muestra aleatoria de 160 estudiantes de administración de empresas, 72 indicaron cierto acuerdo con la afirmación: «Las notas de un examen de selección son intrascendentes para el éxito académico de un estudiante que las notas obtenidas en el bachillerato». Hacer una prueba de hipótesis para evaluar la hipótesis nula de que la mitad de todos los estudiantes de administración de empresas concuerda con esta afirmación, frente a una hipótesis alternativa bilateral. Halla e interpreta el valor p del contraste. (Tomado de Newbold, Carlson & Thorne, 2022, cap. 9)

Semana 10: Sesión 2

Prueba de hipótesis: población múltiple

Sección: Fecha:/...../..... Duración: 60 minutos
Docente: Unidad: 3
Nombres y apellidos:

Instrucciones:

- a) Lee cuidadosamente cada uno de los enunciados y responde las preguntas.
- b) Para la presentación de tus respuestas debes utilizar un archivo digital MS Word (doc.) y para el procesamiento de tus respuestas debes utilizar un archivo digital de R y Python. Debes presentar ambos archivos.
- c) Ingresar a la página de microdatos del INEI, <https://proyectos.inei.gob.pe/microdatos/>
- d) Ingresa a la página de Kaggle, www.kaggle.com.

I. Propósito: Al finalizar la sesión, el estudiante interpreta pruebas de

hipótesis para comparar las medias de varias poblaciones, aplicando los conceptos en la toma de decisiones basadas en datos y en la comunicación de los hallazgos de manera efectiva.

II. Descripción de la actividad por realizar

1. En un estudio comparativo entre bancos de Alemania y Gran Bretaña, se extrajo una muestra de 145 pares de bancos, donde cada par consistía en un banco de Alemania y uno de Gran Bretaña. Los bancos se emparejaron para que fueran lo más similares posible en términos de tamaño y antigüedad. Se calculó el cociente entre los préstamos totales pendientes y los activos totales de cada banco. La diferencia entre las medias muestrales de este cociente (alemán/británico) fue de 0,0518 y la desviación estándar muestral de las diferencias fue de 0,3055. Realiza una prueba de hipótesis para evaluar si las dos medias poblacionales son iguales frente a una hipótesis alternativa bilateral. (Tomado de Newbold, Carlson & Thorne, 2022, cap. 10)
2. Le han pedido que realice un estudio nacional de los precios de venta de las viviendas urbanas para averiguar si han subido con el paso del tiempo. Se sospecha que los precios de las viviendas de las grandes zonas urbanas no han subido al mismo ritmo que la inflación con el paso del tiempo. En su estudio utilizará datos de Atlanta, Chicago, Dallas y Oakland, que se encuentran en el fichero de datos *House Selling Price*. Formula un contraste de hipótesis adecuado y, utilice su paquete estadístico para calcular los estadísticos adecuados para el análisis. Realice el contraste de hipótesis e indique sus conclusiones. Repite el análisis utilizando datos únicamente de la ciudad de Atlanta. (Adaptado de Newbold, Carlson & Thorne, 2022, cap. 10)

3. Una academia ofrece materias de preparación para el examen de admisión a un programa de posgrado. Para evaluar la efectividad del programa, se seleccionaron 12 estudiantes y se dividieron en seis pares, asegurando que los miembros de cada par tuvieran expedientes académicos similares. Se asignó aleatoriamente a un miembro de cada par para que tomara el programa de entrenamiento, mientras que el otro no lo hizo. Las notas obtenidas en la evaluación están registradas en el archivo de datos «*Student Pain*». Considera las diferencias entre las notas siguen una distribución normal, realiza una prueba de hipótesis al 5 % para evaluar si las medias poblacionales de las notas son iguales, frente a la hipótesis alternativa de que la media es mayor para los estudiantes que asistieron al curso de preparación. (Tomado de Newbold, Carlson & Thorne, 2022, cap. 10)

4. Un docente de ciencia política desea contrastar las características de los estudiantes que sufragaron en las elecciones generales con los que no lo hacen. En una muestra aleatoria de 114 estudiantes que afirman haber sufragado en las últimas elecciones generales, la media de las calificaciones es 2,71, con una desviación típica de 0,64. Por otro lado, en una muestra independiente de 123 estudiantes que no sufragaron, la media de las calificaciones es 2,79 y la desviación típica es 0,56. Realiza una prueba de hipótesis para determinar si las medias poblacionales de las calificaciones son iguales, frente a una hipótesis alternativa bilateral. (Tomado de Newbold, Carlson & Thorne, 2022, cap. 10)

5. Dado el aumento reciente de quiebras de grandes compañías, los

auditores están cada vez más intranquilos por la posibilidad de desfalcos. Para evaluar esta posibilidad, los auditores podrían revisar minuciosamente el flujo de caja. En un estudio, se les pide a auditores de nivel medio en compañías de auditoría que evalúen la probabilidad de desfalco significativo basándose en la información sobre el flujo de caja, utilizando una escala de 0 a 100. En una muestra aleatoria de 36 auditores que utilizan esta información, la media de sus evaluaciones es 36,21, con una desviación típica de 22,93. En comparación, una muestra independiente de 36 auditores que no utilizan la información sobre el flujo de caja tiene una media de 47,56 y una desviación típica de 27,56. Suponiendo que las distribuciones poblacionales son normales y tienen la misma varianza, realiza una prueba de hipótesis para determinar si las medias poblacionales son iguales frente a una hipótesis alternativa bilateral. (Tomado de Newbold, Carlson & Thorne, 2022, cap. 10)

6. Los dependientes de una cadena de venta de materiales de construcción que está a punto de ser liquidada fueron encuestados para conocer su punto de vista sobre una estrategia de compra de la empresa. Algunos se comprometieron a aportar 10.000 dólares al plan, entregando de inmediato 800 dólares, mientras que otros manifestaron que no tenían intención de hacer ninguna aportación. En una muestra aleatoria de 175 dependientes que se comprometieron a dar dinero, 78 ya habían sido separados. En contraste, en una muestra aleatoria de 604 empleados que no se comprometieron a dar nada, 208 ya habían sido separados. Constata, al nivel del 5 %, la hipótesis nula de que las proporciones poblacionales de dependientes separados son iguales en ambos grupos frente a una hipótesis alternativa bilateral. (Tomado de Newbold, Carlson & Thorne,

2022, cap. 10)

7. En una muestra aleatoria de 381 acciones de alta calidad, 191 presentaban un déficit inferior al 30 %. En otra muestra aleatoria, independiente de 166 acciones de alto riesgo, 145 tenían una deuda de menos del 30 %. Compara la hipótesis nula de que las proporciones poblacionales de deuda inferior al 30 % son iguales en ambos tipos de acciones frente a una hipótesis alternativa bilateral. (Tomado de Newbold, Carlson & Thorne, 2022, cap. 10)

8. Se consultó a dos muestras aleatorias independientes de usuarios sobre su satisfacción con su plataforma informática, utilizando enfoques ligeramente diferentes en cada caso. Las opciones de respuesta variaban entre los dos métodos. En el primer grupo de 240 personas, 138 respondieron «muy satisfecho» cuando se les preguntó su nivel de satisfacción. En el segundo grupo, también de 240 personas, 128 respondieron «muy satisfecho» cuando se les preguntó sobre su insatisfacción. Contrasta al nivel de significación del 5 % la hipótesis nula de que las proporciones poblacionales de satisfacción son similares en ambos grupos frente a una hipótesis alternativa unilateral. (Tomado de Newbold, Carlson & Thorne, 2022, cap. 10)

9. En una muestra aleatoria de 1.200 daneses, 480 mostraron una actitud positiva hacia los vendedores de vehículos. En una muestra aleatoria independiente de 1.000 franceses, 790 tuvieron una actitud positiva hacia los vendedores de vehículos. Verifica el nivel del 1 % la hipótesis nula de que las proporciones poblacionales de actitud positiva hacia los vendedores de automóviles son iguales en ambos países frente a la hipótesis alternativa de que la proporción de

franceses con una actitud positiva es mayor que la de los daneses.

10. En una muestra aleatoria de 1.556 personas del país A, el 38,4 % estaba de acuerdo con la confirmación: «El incremento del comercio mundial puede acrecentar nuestro auge per cápita». En una muestra aleatoria independiente de 1.108 personas del país B, el 52,0 % estuvo conforme con la misma afirmación. Compara la hipótesis nula de que las proporciones de personas que están de acuerdo con esta afirmación son iguales en ambos países frente a la hipótesis alternativa de que la proporción en el país B es mayor. (Tomado de Newbold, Carlson & Thorne, 2022, cap. 10)

Semana 11: Sesión 2

Prueba de hipótesis: población múltiple

Sección: Fecha:/...../..... Duración: 60 minutos

Docente: Unidad: 3

Nombres y apellidos:

Instrucciones:

- Lee cuidadosamente cada uno de los enunciados y responde las preguntas.
- Para la presentación de tus respuestas debes utilizar un archivo digital MS Word (doc.) y para el procesamiento de tus respuestas debes utilizar un archivo digital de R y Python. Debes presentar ambos archivos.
- Ingresa a la página de microdatos del INEI, <https://proyectos.inei.gob.pe/microdatos/>
- Ingresa a la página de Kaggle, www.kaggle.com.

I. Propósito: Al finalizar la sesión, el estudiante interpreta pruebas de hipótesis para comparar proporciones de múltiples poblaciones, aplicando los conceptos para la toma de decisiones basadas en datos y en la comunicación de sus hallazgos de manera efectiva.

II. Descripción de la actividad por realizar

1. Se parte de la suposición de que a medida que aumenta la experiencia de un grupo de personas en la evaluación de declaraciones de impuestos, también aumenta la variabilidad en sus opiniones sobre la exactitud de estas declaraciones. Se seleccionaron muestras aleatorias independientes de 30 personas cada una, representando grupos con diferentes niveles de experiencia. El grupo con poca experiencia estaba compuesto por individuos que habían terminado recientemente su primer curso de contabilidad intermedia. El grupo con mucha experiencia estaba formado por profesionales con estudios universitarios completos y experiencia en firmas de auditoría reconocidas. A los miembros de ambas muestras se les pidió que evaluaran la exactitud de las declaraciones de impuestos. La varianza muestral del grupo con poca experiencia fue de 451,770, mientras que la del grupo con mucha experiencia fue de 1.614,208. Contrasta la hipótesis nula de que las varianzas poblacionales de ambos grupos son iguales frente a la hipótesis alternativa de que la varianza es mayor en el grupo con más experiencia. (Tomado de Newbold, Carlson & Thorne, 2022, cap. 10)
2. Se solicita a muestras aleatorias independientes de ejecutivos de compañías y de docentes universitarios de economía que puntúen

en una escala del 1 (totalmente en desacuerdo) al 7 (totalmente de acuerdo) la aseveración: «Los puntajes obtenidos en los cursos de economía avanzada son buenos indicadores de la capacidad analítica de los estudiantes». En una muestra de 70 ejecutivos de compañías, la media de las respuestas es 4,4 con una desviación estándar muestral de 1,3. En una muestra de 106 profesores de economía, la media es 5,3 con una desviación estándar muestral de 1,4. (Tomado de Newbold, Carlson & Thorne, 2022, cap. 10)

- a) Realiza una prueba al nivel del 5 % para contrastar la hipótesis nula de que la media poblacional de las respuestas de los directivos de empresas no supera 4,0.
 - b) Realiza una prueba al nivel del 5 % para contrastar la hipótesis nula de que las medias poblacionales de las respuestas son iguales, frente a la hipótesis alternativa de que la media poblacional de las respuestas es mayor para los docentes de economía en comparación con los ejecutivos de compañías.
3. En un estudio, se tomaron muestras aleatorias independientes de titulados medios y titulados superiores en estadística que iniciaron laborando en una gran compañía actuarial y luego se trasladaron a una compañía de seguros. En una muestra de 44 titulados medios, el promedio de meses que tardaron en cambiar de empleo fue de 35,02, con una desviación típica muestral de 18,20. En una muestra de 68 titulados superiores, el promedio de meses fue de 36,34, con una desviación típica muestral de 18,94. Contrasta, al nivel del 10 %, la hipótesis nula de que las medias poblacionales del número de meses para cambiar de empleo en ambos grupos son iguales, frente a una hipótesis alternativa bilateral. (Tomado de Newbold, Carlson & Thorne, 2022, cap. 10)

4. Se encuesta a muestras aleatorias independientes de pacientes que han obtenido prótesis de rodilla y de cadera sobre la calidad del servicio en una escala de 1 (baja) a 7 (alta). En una muestra de 83 pacientes con prótesis de rodilla, la valoración media es de 6,543, con una desviación típica muestral de 0,649. En una muestra de 54 pacientes con prótesis de cadera, la valoración media es de 6,733, con una desviación típica muestral de 0,425. Compara la hipótesis nula de que las medias poblacionales de las valoraciones de estos dos grupos de pacientes son semejantes frente a una hipótesis alternativa bilateral. (Tomado de Newbold, Carlson & Thorne, 2022, cap. 10)

5. En una muestra aleatoria de 148 estudiantes de contabilidad, 75 piensan que tener sentido del humor es una habilidad muy importante para su profesión. En una muestra aleatoria independiente de 178 estudiantes de economía financiera, 81 comparten la misma opinión. (Tomado de Newbold, Carlson & Thorne, 2022, cap. 10)
 - a) Realiza una prueba al nivel del 5 % para evaluar la hipótesis nula de que al menos el 50 % de todos los estudiantes de economía financiera que consideran que el sentido del humor es primordial.
 - b) Realice una prueba al nivel del 5 % para evaluar la hipótesis nula de que las proporciones de estudiantes de contabilidad y de economía financiera que consideran que el sentido del humor es primordial, son iguales frente a una hipótesis alternativa bilateral.

6. En una muestra aleatoria de 69 compañías de seguros médicos, 47 contaban con su propio departamento de relaciones públicas. De manera similar, 40 de una muestra aleatoria independiente de 69 empresas de seguros de accidentes también tenían un departamento de relaciones públicas. Calcula e interpreta el p-valor para una prueba que contraste la igualdad de las proporciones poblacionales, considerando una hipótesis alternativa bilateral. (Tomado de Newbold, Carlson & Thorne, 2022, cap. 10)

7. En una investigación, se seleccionaron muestras aleatorias independientes de varones y damas clientes del Centro de Iniciativa Empresarial que estaban evaluando la posibilidad de iniciar un negocio. De los 94 varones en la muestra, 53 efectivamente fundaron una empresa, mientras que 47 de 68 damas en la muestra hicieron lo mismo. Calcula e interpreta el p-valor para una prueba que contraste la igualdad de las proporciones poblacionales, con la hipótesis alternativa de que la proporción de damas clientes que realmente fundaron una empresa es mayor que la de los varones. (Tomado de Newbold, Carlson & Thorne, 2022, cap. 10)

8. Una persona encargada del desarrollo económico rural en un país en rápido crecimiento, que está empleando petróleo recientemente descubierto para impulsar su desarrollo, debe investigar si las nuevas técnicas de cultivo de arroz han acrecentado la producción por hectárea. Se tomó una muestra aleatoria de 27 arrozales que se cultivaron con el método antiguo, obteniendo una media muestral de producción de 60 por hectárea y una varianza muestral de 100. En siguiente año, se aplicó la técnica en los mismos arrozales, resultando en una media muestral de producción de 64

por hectárea y una varianza muestral de 150. La correlación muestral entre las dos técnicas fue de 0,38. Se asume que las varianzas poblacionales son iguales y el análisis debe realizarse bajo esta suposición. (Tomado de Newbold, Carlson & Thorne, 2022, cap. 10)

- a) Emplea una prueba de hipótesis con un nivel de significación del 5 % para determinar si existen evidencias suficientes que demuestren que la nueva técnica aumenta la producción por hectárea, e interpreta los resultados obtenidos.
- b) Dado que se asume que las varianzas poblacionales son semejantes, elabora un intervalo de confianza del 95 % para el cociente entre las varianzas muestrales. ¿Permiten las varianzas muestrales observadas llegar a la conclusión de que las varianzas poblacionales son semejantes? Explica tu respuesta.

9. Susana Perales, presidenta de Comercios Planetarios Reunidos (CPR), ha solicitado asistencia para evaluar la inserción del nuevo teléfono móvil de la compañía en dos mercados y determinar si la diferencia en los pagos de mercado se mantiene constante. Históricamente, CPR ha tenido un pago del 30 % en el mercado 1, ubicado en el oeste de Polonia, y del 35 % en el mercado 2, en el sur de Austria. Se ha obtenido una muestra aleatoria de compradores potenciales en cada zona. En el mercado 1, de una muestra de 800 personas, 258 han indicado que comprarán el teléfono de CPR. En el mercado 2, de una muestra de 700 personas, 260 han expresado la misma intención. (Tomado de Newbold, Carlson & Thorne, 2022, cap. 10)

- a) Usa un nivel de significación del 3 %, realiza una prueba para comparar la hipótesis de que los pagos de mercado son semejantes con la hipótesis de que son diferentes (diferencia

entre el mercado 2 y el mercado 1).

- b) Con un nivel de significación del 3 %, realice una prueba para comparar la hipótesis de que el pago de mercado es igual con la hipótesis de que el pago del mercado 2 es mayor que la del mercado 1.

10. María Perlas es responsable de mezclar dos tipos de harina para crear pan exótico de alta calidad. Utiliza harina A y harina B, ambas con un peso medio poblacional de 80 gramos, pero con varianzas poblacionales de 0,04 y 0,06 respectivamente, y una correlación de 0,40 entre los pesos. Al mezclar las dos harinas para obtener un paquete de 160 gramos, cada 60 minutos se toma una muestra aleatoria de cuatro paquetes de harina exótica para calcular el peso medio de los paquetes. Para asegurar que el peso de los paquetes se mantenga dentro de los estándares, se debe preparar un intervalo de aceptación del 99 % para un gráfico de control de calidad. Esto implica calcular el rango dentro del cual se espera que caigan las medias muestrales de las muestras de cuatro paquetes, mostrando los pasos y el razonamiento para el cálculo. Este gráfico ayuda a garantizar que los pesos de los paquetes continúen aplicando las normas establecidas, indicando cuándo el proceso podría estar fuera de control. (Tomado de Newbold, Carlson & Thorne, 2022, cap. 10)

Semana 12: Sesión 2

Análisis de varianza

Sección: Fecha:/...../..... Duración: 60 minutos

Docente: Unidad: 3

Nombres y apellidos:

Instrucciones:

- a) Lee cuidadosamente cada uno de los enunciados y responde las preguntas.
- b) Para la presentación de tus respuestas debes utilizar un archivo digital MS *Word* (doc.) y para el procesamiento de tus respuestas debes utilizar un archivo digital de R y *Python*. Debes presentar ambos archivos.
- c) Ingresar a la página de microdatos del INEI, <https://proyectos.inei.gob.pe/microdatos/>
- d) Ingresa a la página de Kaggle, www.kaggle.com.

- I. Propósito:** Al finalizar la sesión, el estudiante ANOVA para comparar las medias de múltiples grupos y en la comunicación de sus hallazgos de manera efectiva a audiencias técnicas y no técnicas.

II. Descripción de la actividad por realizar

1. Utilizando la tabla proporcionada para el análisis de la varianza, determina las medias de los cuadrados entre los grupos y dentro de los grupos. Calcula el valor del cociente F y verifica si las medias de los grupos son iguales. (Tomado de Newbold, Carlson & Thorne, 2022, cap. 15)

Fuente de variación	Suma de los cuadrados	Grados de libertad
Entre los grupos	1.000	4
Dentro de los grupos	750	15
Total	1.750	19

2. Utilizando la tabla del análisis de la varianza proporcionada, determina las medias de los cuadrados entre los grupos y dentro de los grupos. Luego, calcula el cociente F y verifica si las medias de los grupos son iguales. (Tomado de Newbold, Carlson & Thorne, 2022, cap. 15)

Fuente de variación	Suma de los cuadrados	Grados de libertad
Entre los grupos	879	3
Dentro de los grupos	798	16
Total	1.677	19

3. Una empresa está considerando tres opciones de automóviles para su cuadrilla: nacionales, japoneses o europeos. Se adquieren cinco vehículos de cada tipo y, tras recorrer 10.000 kilómetros con ellos, se halla el costo de rendimiento por kilómetro para cada tipo. Los resultados obtenidos en centavos por kilómetro son los siguientes. (Tomado de Newbold, Carlson & Thorne, 2022, cap. 15)

Nacionales	Japoneses	Europeos
18.0	20.1	19.3
15.6	15.6	15.4
15.4	16.1	15.1
19.1	15.3	18.6
16.9	15.4	16.1

- a) Elabora una tabla de análisis de la varianza para estos datos.
- b) Verifica la hipótesis nula de que las medias poblacionales de los costos de explotación por kilómetro son iguales para los tres tipos de automóviles.
- c) Determina la diferencia mínima significativa e identifica cuáles subgrupos tienen medias distintas.
4. Se seleccionan muestras aleatorias de siete estudiantes de primer año, siete de segundo año y siete de tercer año que participan en una clase de estadística para negocios. La tabla adjunta presenta las calificaciones obtenidas en el examen final. (Tomado de Newbold, Carlson & Thorne, 2022, cap. 15)

Nacionales	Japoneses	Europeos
82	71	64
93	62	73
61	85	87
74	94	91
53	78	56
70	66	78
53	71	87

- a) Prepara la tabla para el análisis de la varianza.
 - b) Verifica la hipótesis nula de que las medias poblacionales de las puntuaciones de los tres grupos son similares. Calcula la diferencia mínima significativa e identifica qué subgrupos presentan medias distintas.
5. En un problema con tres subgrupos, donde las sumas de los puestos en cada subgrupo son 45, 98 y 88, y el tamaño de los subgrupos es 6, 6 y 7, realiza el contraste de Kruskal-Wallis. Verifica la hipótesis nula de que los puestos en los subgrupos son iguales. (Tomado de Newbold, Carlson & Thorne, 2022, cap. 15)
 6. En un problema con cuatro subgrupos, donde las sumas de los puestos en cada subgrupo son 49, 84, 76 y 81, y el tamaño de los subgrupos es 4, 6, 7 y 6, realiza el contraste de Kruskal-Wallis. Verifica la hipótesis nula de que los puestos en los subgrupos son iguales. (Tomado de Newbold, Carlson & Thorne, 2022, cap. 15)
 7. En un problema con cuatro subgrupos, donde las sumas de los puestos en cada subgrupo son 71, 88, 82 y 79, y los tamaños de los subgrupos

son 5, 6, 6 y 7, realiza el contraste de Kruskal-Wallis. Verifica la hipótesis nula de que los puestos en los subgrupos son iguales. (Tomado de Newbold, Carlson & Thorne, 2022, cap. 15)

8. Se toman muestras aleatorias independientes de 101 estudiantes de primer año, 112 de segundo año y 96 de tercer año, quienes evalúan en una escala de 1 a 7 la importancia que le dan a la marca al comprar un automóvil. El valor del estadístico de Kruskal-Wallis calculado es 0,15. (Tomado de Newbold, Carlson & Thorne, 2022, cap. 15)
- a) ¿Qué hipótesis nula puede contrastarse usando la información?
 - b) Realiza el contraste.

9. Considera un análisis de varianza bifactorial con una observación por celda y bloques aleatorizados, con los siguientes resultados:
(Tomado de Newbold, Carlson & Thorne, 2022, cap. 15)

Fuente de variación	Suma de los cuadrados	Grados de libertad
Entre los grupos	231	4
Dentro de los grupos	348	5
Error	550	20
Total	1.129	29

Determina las medias de los cuadrados y comprueba las hipótesis nulas de que las medias entre los grupos son iguales y de que las medias entre los bloques son iguales.

10. Realiza un análisis de varianza bifactorial con una observación por celda y bloques aleatorizados, utilizando los siguientes resultados:
(Tomado de Newbold, Carlson & Thorne, 2022, cap. 15)

Fuente de variación	Suma de los cuadrados	Grados de libertad
Entre los grupos	380	6
Dentro de los grupos	232	5
Error	387	30
Total	989	41

Calcula las medias de los cuadrados y contraste las hipótesis nulas de que las medias entre grupos son iguales y de que las medias entre bloques son iguales.

Cuarta Unidad

Correlación y análisis de regresión

Semana 13: Sesión 2

Análisis de datos categóricos

Sección: Fecha:/...../..... Duración: 60 minutos

Docente: Unidad: 4

Nombres y apellidos:

Instrucciones:

- a) Lee cuidadosamente cada uno de los enunciados y responde las preguntas.
- b) Para la presentación de tus respuestas debes utilizar un archivo digital MS Word (doc.) y para el procesamiento de tus respuestas debes utilizar un archivo digital de R y Python. Debes presentar ambos archivos.
- c) Ingresar a la página de microdatos del INEI,
<https://proyectos.inei.gob.pe/microdatos/>

d) Ingresa a la página de Kaggle, www.kaggle.com.

I. Propósito: Al finalizar la sesión, el estudiante analiza datos categóricos, interpretando los resultados y aplicando sus conocimientos en situaciones reales para desarrollar sus habilidades y en la comunicación de sus hallazgos de manera efectiva.

II. Descripción de la actividad por realizar

1. ¿Qué tipo de datos categóricos se miden en una escala que tiene un orden, pero no tiene intervalos definidos? Presente tres ejemplos. (Tomado de Newbold, Carlson & Thorne, 2022, cap. 14)
2. Presente tres casos de una variable categórica nominal identificados en el mundo de la economía y negocios.
3. ¿Qué técnica se utiliza comúnmente para analizar tablas de contingencia? Presente un caso aplicado al *marketing* y un caso adicional relacionado con finanzas y economía.
4. ¿Cuál es el objetivo principal de la prueba de independencia de chi-cuadrado? Presenta situaciones donde no se cumple los requisitos para la aplicación de una prueba de independencia.
5. Realice un listado y definición de los tipos de gráficos que se utilizan en estadística e indique ¿Qué gráfico es más adecuado para representar datos categóricos?
6. ¿Cuál es una medida adecuada para la tendencia central de datos categóricos nominales? Presente dos casos en los cuales no se aplicaron la medida identificada por usted y señale cuáles son las

consecuencias para la toma de decisiones.

7. Presenta un caso de estudio de mercado para el lanzamiento de una nueva carrera en una universidad privada, presenta un listado de las variables utilizadas y describe una variable que tiene más de dos categorías sin orden específico.
8. Utiliza un caso hipotético y construye una tabla de contingencia, ¿qué representan las celdas individuales?
9. En una prueba chi-cuadrado ¿Cuál es la interpretación de un valor p menor que 0.05? Utilizando un caso hipotético sobre gestión de personal, analiza las implicancias de tener un p menor que 0.05.
10. ¿Qué prueba se usa para comparar las proporciones de dos grupos categóricos? Presenta un caso de negocios internacionales que respalde tu respuesta.
11. ¿Qué indica un valor de phi cercano a 1 en una tabla de contingencia 2x2?
12. ¿Cuál es el término para una variable categórica con solo dos categorías?
13. ¿Qué medida se utiliza para evaluar la concordancia entre dos observadores categóricos?
14. ¿Cuál de las siguientes es una ventaja del análisis de correspondencias?
15. ¿Cuál es la función principal de una matriz de confusión en análisis

categorico?

16. ¿Qué tipo de gráfico es útil para comparar la distribución de una variable categórica en diferentes grupos?
17. ¿Cuál es la finalidad de realizar un análisis de frecuencia en datos categóricos?
18. ¿Qué mide el coeficiente de contingencia en tablas de contingencia?
19. ¿Qué término describe la variable categórica dependiente en un modelo de regresión logística?
20. ¿Cuál es un uso común de la prueba de McNemar?

Semana 14: Sesión 2

Relación entre correlación y causalidad

Sección: Fecha:/...../..... Duración: 60 minutos

Docente: Unidad: 4

Nombres y apellidos:

Instrucciones:

- a) Lee cuidadosamente cada uno de los enunciados y responde las preguntas.
- b) Para la presentación de tus respuestas debes utilizar un archivo digital MS Word (doc.) y para el procesamiento de tus respuestas debes utilizar un archivo digital de R y Python. Debes presentar ambos archivos.

- c) Ingresar a la página de microdatos del INEI, <https://proyectos.inei.gob.pe/microdatos/>
- d) Ingresa a la página de Kaggle, www.kaggle.com.

I. Propósito: Al finalizar la sesión, el estudiante diferencia entre correlación y causalidad para llegar a interpretaciones correlacionales como inferencias de causalidad en estudios estadísticos y comunicar eficazmente las limitaciones y alcances de sus hallazgos

II. Descripción de la actividad por realizar

1. Analiza la importancia de la temporalidad en la determinación de una relación causal entre dos variables. Proporciona ejemplos específicos donde la falta de control de la temporalidad podría llevar a conclusiones incorrectas sobre causalidad y discute cómo se puede evaluar adecuadamente la temporalidad en estudios empíricos.
2. Discute el concepto de endogeneidad y sus implicaciones en la inferencia causal. Proporciona ejemplos específicos de cómo la endogeneidad puede sesgar los resultados de un análisis estadístico y explica las técnicas econométricas que pueden utilizarse para abordar este problema.
3. Explica cómo los estudios de casos y controles pueden ser utilizados para inferir causalidad en situaciones donde los experimentos aleatorizados no son factibles. Discute las ventajas y limitaciones de este enfoque y cómo se pueden minimizar los sesgos potenciales.
4. Evalúa la utilidad del análisis de series temporales en la determinación de causalidad. Proporciona ejemplos específicos de técnicas de

series temporales que pueden ayudar a identificar relaciones causales y discute sus ventajas y limitaciones.

5. Critica la afirmación «correlación no implica causalidad». Discute en qué contextos esta afirmación puede ser malinterpretada o mal aplicada y cómo los investigadores pueden utilizar la correlación como un primer paso hacia la identificación de relaciones causales.
6. Analiza cómo los métodos de diferencias en diferencias (DiD) pueden ayudar a identificar efectos causales en estudios no experimentales. Proporciona un ejemplo detallado de un estudio que utilice esta metodología y discuta los supuestos críticos y cómo se pueden verificar.
7. Explica el papel de los experimentos naturales en la inferencia causal. Proporciona un ejemplo concreto de un experimento natural y discuta las características que deben tener los experimentos naturales para ser considerados válidos.
8. Discute el uso de la regresión discontinua (RDD) para establecer causalidad. Proporciona un ejemplo detallado de un estudio que utilice RDD y explique cómo se interpretan los resultados de esta técnica.
9. Analiza la importancia de los diseños de panel en la inferencia causal. Discute las ventajas de utilizar datos de panel sobre datos transversales y cómo los métodos de efectos fijos y aleatorios pueden mejorar la validez de los resultados.

10. Explica cómo los modelos de ecuaciones estructurales (SEM) pueden ser utilizados para establecer relaciones causales complejas en estudios económicos. Discute las ventajas y limitaciones de esta metodología y proporciona un ejemplo de cómo se puede aplicar en un estudio empírico.

Semana 15: Sesión 2

Introducción al análisis de regresión

Sección: Fecha:/...../..... Duración: 60 minutos

Docente: Unidad: 4

Nombres y apellidos:

Instrucciones:

- a) Lee cuidadosamente cada uno de los enunciados y responde las preguntas.
- b) Para la presentación de tus respuestas debes utilizar un archivo digital MS Word (doc.) y para el procesamiento de tus respuestas debes utilizar un

archivo digital de R y *Python*. Debes presentar ambos archivos.

- c) Ingresar a la página de microdatos del INEI,
<https://proyectos.inei.gob.pe/microdatos/>
- d) Ingresa a la página de *Kaggle*, www.kaggle.com.

I. Propósito: Al finalizar la sesión, el estudiante interpreta modelos de regresión y la aplicación de las técnicas en problemas reales.

II. Descripción de la actividad por realizar

1. Una gran empresa de bienes de consumo ha estado estudiando la influencia de la publicidad en los beneficios totales. Para realizar este estudio se han recogido datos sobre los gastos publicitarios y las ventas totales de un periodo de cinco meses:

(10, 100) (15, 200) (7, 80) (12, 120) (14, 150)

La primera cifra corresponde a los gastos publicitarios y la segunda a las ventas totales. (Tomado de Newbold, Carlson & Thorne, 2022, cap. 11)

- a) Representa los datos.
 - b) ¿Demuestra el gráfico que la publicidad influye positivamente en las ventas?
 - c) Calcula los coeficientes de regresión, b_0 y b_1 .
2. Abdul Hassan, presidente de Floor Coverings Unlimited, le ha solicitado que investigue la relación entre el precio de mercado y las toneladas de alfombras ofrecidas por su competidor, Best Floor, Inc. Le proporciona las siguientes observaciones sobre el precio por tonelada y el número de toneladas, provenientes de sus archivos confidenciales:

(2, 5) (4, 10) (3, 8) (6, 18) (3, 6) (5, 15) (6, 20) (2, 4)

En cada observación, el primer número representa el precio y el segundo la cantidad. (Tomado de Newbold, Carlson & Thorne, 2022, cap. 11)

- a) Dibuja un diagrama de dispersión.
 - b) Precisa los coeficientes de regresión, b_0 y b_1 .
 - c) Describe brevemente por escrito la ecuación de regresión y cómo Abdul puede usarla para analizar su competencia. Incluye una explicación del rango en el que la ecuación es aplicable.
3. Supongamos que hemos estimado una ecuación de la regresión para las ventas semanales de agendas electrónicas en función del precio durante la semana. Explica el significado de la constante b_0 para el director de la marca. (Tomado de Newbold, Carlson & Thorne, 2022, cap. 11)
4. Se ha estimado un modelo de regresión que vincula las ventas totales de productos alimenticios con la renta disponible, basado en datos de pequeñas ciudades aisladas en el oeste de Estados Unidos. Elabora una lista de factores que podrían afectar el término de error aleatorio. (Tomado de Newbold, Carlson & Thorne, 2022, cap. 11)
5. *Transportation Research Inc.* le ha solicitado que desarrolle ecuaciones de regresión múltiple para estimar el impacto de varias variables en el ahorro de combustible. Encuentra una base de datos adecuada para este análisis, y la variable dependiente será las millas por galón (*milpgal*), de acuerdo con la certificación del Departamento de Transporte. (Tomado de Newbold, Carlson & Thorne, 2022, cap. 11)

- a) Desarrolla una ecuación de regresión que utilice la potencia de los vehículos (*horsepower*) y el peso de los vehículos (*weight*) como variables independientes. Interpreta los coeficientes obtenidos.
 - b) Desarrolla una segunda ecuación de regresión que incluya el número de cilindros (*cylinder*) como una variable independiente adicional a la ecuación del apartado (a). Interpreta los coeficientes resultantes.
 - c) Elabora una ecuación de regresión que emplee el número de cilindros y el peso de los vehículos como variables independientes. Interpreta los coeficientes y compara los resultados con los obtenidos en los apartados (a) y (b).
 - d) Desarrolla una ecuación de regresión que incluya la potencia de los vehículos, el peso de los vehículos y el precio como variables predictoras. Interpreta los coeficientes obtenidos.
 - e) Elabora un breve informe sobre los resultados.
6. *Transportation Research Inc.* le ha solicitado que desarrolle varias ecuaciones de regresión múltiple para evaluar el impacto de diferentes variables en la potencia de los vehículos. Encuentra una base de datos adecuada para este análisis, y la variable dependiente será la potencia de los vehículos (*horsepower*), según la certificación del Departamento de Transporte. (Tomado de Newbold, Carlson & Thorne, 2022, cap. 11)
- a) Desarrolla una ecuación de regresión que emplee el peso de los vehículos (*weight*) y el volumen de desplazamiento de los cilindros (*displacement*) como variables predictoras. Interpreta los coeficientes resultantes.
 - b) Desarrolla una ecuación de regresión que emplee el peso de los

vehículos, el desplazamiento de los cilindros y el número de cilindros (*cylinder*) como variables predictoras. Interpreta los coeficientes y compara los resultados con los obtenidos en el apartado (a).

- c) Desarrolla una ecuación de regresión que utilice el peso de los vehículos, el desplazamiento de los cilindros y las millas por galón (*milpgal*) como variables predictoras. Interpreta los coeficientes y compara los resultados con los obtenidos en el apartado (a).
- d) Desarrolla una ecuación de regresión que incluya el peso de los vehículos, el desplazamiento de los cilindros, las millas por galón y el precio como variables predictoras. Interpreta los coeficientes y compara los resultados con los del apartado (c).
- e) Redacte un breve informe sobre los resultados de su análisis de este problema.

7. Se llevó a cabo un estudio para determinar si ciertas características podían explicar la variabilidad en los precios de los hornos. Se estimó una regresión para una muestra de 19 hornos:

(Tomado de Newbold, Carlson & Thorne, 2022, cap. 12)

$$y = -68,236 + 0,0023x_1 + 19,729x_2 + 7,653x_3$$

(0,005) (8,992) (3,082)

$$R^2 = 0,84$$

Donde:

y = precio en dólares

x_1 = potencia del horno en BTU por hora

x_2 = índice de eficiencia energética

x_3 = número de posiciones

Los números entre paréntesis debajo de los coeficientes representan los errores estándar de los coeficientes estimados.

- a) Determina un intervalo de confianza del 95 % para el incremento esperado en el precio debido a un aumento en las posiciones, manteniendo constantes los valores de potencia y el índice de eficiencia energética.
 - b) Verifica la hipótesis nula de que, manteniéndose constantes los demás factores, el índice de eficiencia energética de los hornos no influye en su precio, frente a la hipótesis alternativa de que, a mayor índice de eficiencia energética, mayor será el precio.
8. Se realiza una regresión de una variable dependiente con respecto a K variables independientes utilizando n conjuntos de observaciones muestrales. SCE es la suma de los cuadrados de los errores y R^2 es el coeficiente de determinación de esta regresión estimada. Queremos contrastar la hipótesis nula de que K_1 de estas variables independientes, consideradas en conjunto, no afectan linealmente a la variable dependiente, dado que las demás variables independientes ($K-K_1$) también se utilizan. Supongamos que se vuelve a estimar la regresión excluyendo las K_1 variables independientes de interés. Sea SCE^* la suma de los cuadrados de los errores y R^{*2} el coeficiente de determinación de esta regresión. Demuestre que el estadístico para contrastar nuestra hipótesis nula, introducido en el Apartado 12.5, puede expresarse de la forma siguiente. (Tomado de Newbold, Carlson & Thorne, 2022, cap. 12)
9. Se le ha solicitado que desarrolle una función de producción exponencial del tipo Cobb-Douglas para predecir el número de

microprocesadores producidos por un fabricante, Y , en función de las unidades de capital, X_1 ; las unidades de trabajo, X_2 ; y el número de informáticos que realizan investigación básica, X_3 . Especifique la forma del modelo y describa detalladamente cómo estimaría los coeficientes. Comience con un modelo sin restricciones y luego incluya la restricción de que la suma de los coeficientes de las tres variables debe ser igual a 1. (Tomado de Newbold, Carlson & Thorne, 2022, cap. 12)

10. En un estudio realizado con 27 estudiantes de la Universidad de Illinois, se recopilaron datos sobre la calificación media (y), el número de horas semanales dedicadas a estudiar (x_1), el promedio de horas dedicadas a estudiar para los exámenes (x_2), el número de horas semanales pasadas en los bares (x_3), si los estudiantes toman notas o subrayan al leer los libros de texto (x_4 %1 si lo hacen, 0 si no), y el número promedio de créditos cursados por semestres (x_5). Estime la regresión de la calificación media en función de estas cinco variables independientes y presente los resultados en un informe. Encuentra una base de datos adecuada para llevar a cabo este estudio, exponga en un informe sus resultados. Busca una base datos para realizar este estudio. (Tomado de Newbold, Carlson & Thorne, 2022, cap. 12)

Semana 16: Sesión 2

Aplicaciones de regresión y series de tiempo

Sección: Fecha:/...../..... Duración: 60 minutos

Docente: Unidad: 4

Nombres y apellidos:

Instrucciones:

- a) Lee cuidadosamente cada uno de los enunciados y responde las preguntas.
- b) Para la presentación de tus respuestas debes utilizar un archivo digital MS Word (doc.) y para el procesamiento de tus respuestas debes utilizar un

archivo digital de R y *Python*. Debes presentar ambos archivos.

- c) Ingresar a la página de microdatos del INEI, <https://proyectos.inei.gob.pe/microdatos/>
- d) Ingresa a la página de *Kaggle*, www.kaggle.com.

I. Propósito: Al finalizar la sesión, el estudiante aplica modelos de regresión y análisis de series de tiempo, interpretando resultados para realizar predicciones y en la toma de decisiones informadas, comunicando sus hallazgos.

II. Descripción de la actividad por realizar

1. Encuentra una base de datos que registre las ventas trimestrales realizadas por una empresa durante un período de seis años.
 - a) Elabora un gráfico temporal de esta serie y analiza sus características.
 - b) Aplica el método del índice estacional para eliminar la estacionalidad de esta serie. Luego, representa gráficamente la serie desestacionalizada y analiza sus características. (Tomado de Newbold, Carlson & Thorne, 2022, cap. 16)
2. Encuentra una base de datos que registre el importe del oro (en dólares) al término de cada año durante 14 años consecutivos. Halla una serie de medias móviles simples centradas de 3 puntos utilizando estos datos sobre el importe del oro. Gráfica la serie suavizada y analiza el gráfico final.
3. Encuentra una base de datos que muestre las casas iniciadas por cada mil residentes en Norteamérica a lo largo de un periodo de 24 años. Halla una serie de medias móviles simples centradas de 5 puntos

utilizando estos datos sobre la edificación de casas. Luego, realiza un gráfico temporal de la serie suavizada y analiza los resultados.

4. Encuentra una base de datos que registre la rentabilidad por acción de una empresa a lo largo de un periodo de 28 años. Calcula una serie de medias móviles simples centradas de 7 puntos utilizando estos datos sobre la rentabilidad de la empresa. A partir de un gráfico temporal de la serie suavizada, analiza los componentes regulares que puedan identificarse.
5. Encuentra una base de datos que incluya información sobre los beneficios por acción y estima un modelo autorregresivo de primer orden para estos beneficios. Utiliza el modelo ajustado para realizar predicciones sobre los beneficios para los próximos cuatro días.
6. Encuentra una base de datos que registre el volumen de transacciones (en cientos de miles) de acciones de una compañía durante un espacio de 12 meses. Estima un modelo autorregresivo de primer orden utilizando estos datos y emplea el modelo ajustado para predecir el volumen de transacciones para las tres semanas siguientes.
7. Encuentra una base de datos y estima modelos autorregresivos de órdenes 1 a 4. Aplica el método correspondiente para verificar la hipótesis de que el orden de la autorregresión es $p=1$ frente a la alternativa de que es p , utilizando un nivel de significación del 10 %. Selecciona uno de estos modelos y realiza predicciones sobre la edificación de casas para los siguientes cinco años. Traza un gráfico temporal que muestre tanto las observaciones originales como las predicciones. ¿Las predicciones serían distintas si se usa un nivel de

significación del 5 % en los contrastes del orden autorregresivo?

8. Encuentra una base de datos que contenga información sobre la rentabilidad por acción de una compañía y ajusta modelos autorregresivos de órdenes 1 a 4. Utiliza el método adecuado para contrastar la hipótesis de que el orden de la autorregresión es $p=1$ frente a la alternativa de que el orden verdadero es p , con un nivel de significación del 10 %. Selecciona uno de estos modelos y realiza predicciones sobre la rentabilidad por acción para los próximos cinco años. Traza un gráfico que muestre tanto los datos originales como las predicciones. ¿Cambiarían los resultados si se usara un nivel de significación del 5 % para los contrastes?
9. Identifica una base de datos, desarrolla un modelo autorregresivo de los gastos de consumo personal. Utiliza primero los datos del periodo comprendido entre el primer trimestre de 1980 y el cuarto de 2000 para hacer una predicción para los trimestres de los años 2001-2003. A continuación, utiliza los datos del periodo comprendido entre el primer trimestre de 1980 y el cuarto de 2007 para hacer una predicción para los trimestres de los años 2008 y 2009. Analiza las diferencias de precisión entre las predicciones y los resultados reales e indica las causas de estas diferencias.
10. Identifica una base de datos, desarrolla un modelo autorregresivo de la inversión fija. Utiliza primero los datos del periodo comprendido entre el primer trimestre de 1965 y el cuarto de 2000 para hacer una predicción para los trimestres de los años 2001-2003. A continuación, utiliza los datos del periodo comprendido entre el primer trimestre de 1965 y el cuarto de 2007 para hacer una predicción para los

trimestres de los años 2008 y 2009. Analiza las diferencias de precisión entre las predicciones y los resultados reales e indica las causas de estas diferencias.

Referencias

Díaz, A. (2019). *Estadística aplicada a la administración y la economía*.

McGraw-Hill & Interamericana

Newbold, P., Carlson, W. & Thorne, B. (2013). *Estadística para la administración*

y la economía. Pearson Educación

https://catalogo.continental.edu.pe/permalink/51UCCI_INST/1ud8d5s/alma990000003110107836

Anderson, D., Sweeney, D. & Williams, T. (2019). *Estadística para Negocios y*

Economía (13ª Edición). Cengage

Newbold, P., Carlson, W. & Thorne, B. (2023). *Statistics for Business and Economics*. (10ª edición). Pearson

Triola, M. (2019). *Estadística*. (12ª. ed.). Pearson Educación

Alberich, R., Gomila J. & Mir, A. (2022). *Curso completo de estadística inferencial con R y Python*. [ebook].

<https://joanby.github.io/estadistica-inferencial/>

Code, VS. (2024) *Visual Studio Code*. (versión 1.89.1) [software]

<https://code.visualstudio.com>

CRAN. (2024) *The Comprehensive R Archive Network*. (versión R.4.4.0) [software]. <https://cran.r-project.org/index.html>

José Antonio. (2022, 25 de agosto) *Curso de Estadística Inferencial Aplicada*. [video Youtube].

<https://www.youtube.com/playlist?list=PLC25JaspeZd1zGav1A2HRief9T8Gs3PbA>

Mesa, D. & Mereles, J. (noviembre de 2022). *Curso de Inferencia Estadística II. Aplicaciones con el programa estadístico R-project*. [ebook]. RPubS.

https://rpubs.com/diegomeza/inferencia_con_R

Posit. (2024) *RStudio IDE*. (Version: 2024.04.0+735) [software]

<https://posit.co/downloads/>

Python, S.F. (2024) *Python a programming language changes the world*. (version Python 3.12.3) [software] <https://www.python.org/downloads/>

Quintela del Rio, A. (2019 4 de setiembre). *Estadística Básica Edulcorada*.

[ebook]. Bookdown. <https://bookdown.org/aquintela/EBE/>

Romero. J. (2020, 12 de julio). *Inferencia Estadística*. [Video Youtube]

https://www.youtube.com/live/tGBEQg_tJEM?si=K9YQhjcpp70mZR94