

# Tarea Final: Análisis de Datos para la Toma de Decisiones

Christian Araya Muñoz

Magíster en Estadística, PUCV

## 1 Instrucciones

A continuación, desarrolle las siguientes preguntas de manera individual. Dispone de dos semanas de plazo para trabajar (fecha límite: martes 28 de abril, hasta las 23:59 hrs).

Escriba sus respuestas en un archivo Word o Latex, indicando su nombre y número de pregunta. Emplee las tablas, gráficos o resúmenes que considere pertinentes. Añada en una carpeta, que debe entregar comprimida, cualquier script o planilla Excel que utilice en sus cálculos.

### Question 1

La hoja *P3.xlsx* presenta el peso en Kg de 5 muestras de sacos de cemento envasados en 4 plantas diferentes (5 muestras por cada una de las 4 plantas).

A un nivel  $\alpha = 0,05$  desarrolle:

1. Use la prueba de Kruskal Wallis para determinar si existen diferencias significativas en los pesos de los sacos envasados en cada planta.
2. Usted sabe que el test de Kruskal Wallis es una alternativa al test de ANOVA, que compara las medias. El test de ANOVA exige como supuesto que las varianzas sean homogéneas (iguales) entre los grupos y que los datos provengan de poblaciones Normales. En este caso, ¿califica el set de datos para aplicar ANOVA? Utilice los tests revisados en clases y concluya.

### Question 2

La base en *P4.xlsx* contiene los resultados de una encuesta aplicada a 80 individuos que visitaron un restaurante de Valparaíso. Se midió a través de un puntaje continuo, entre 1 (peor puntuación) y 7 (mejor puntuación) el desempeño del local en 4 ámbitos (desde A1 a A4) y el nivel de satisfacción general del encuestado (*Satis*).

Para poder comprender cómo un individuo define su nivel de satisfacción, se buscará relacionar el puntaje de satisfacción general con el puntaje en los 4 ámbitos medidos.

A continuación responda:

1. ¿Con cuál de los puntajes presentados, la variable *Satis* (satisfacción general) evidencia el mayor nivel de correlación?
2. Observando los valores que usted obtiene, ¿qué puede comentar en términos generales?
3. Discuta y proporcione evidencia sobre: la correlación muestral entre *Satis* y cada uno de los puntajes obtenidos en los 4 ámbitos proporciona evidencia para asumir sin problemas que  $\rho$  es 0 en cada caso.

### Question 3

Existe una alternativa no paramétrica para estudiar tests de hipótesis para una población y dos muestras pareadas: *test de rango con signos de Wilcoxon*.

Sobre el test:

1. Investigue cómo opera y provea un ejemplo para una muestra y para el caso de dos muestras pareadas.
2. Estudie cómo aplicar estos tests en R y muestre la aplicación a su ejemplo.

### Question 4

La base de datos *P5.csv* contiene 440 registros del Departamento de Personas de una compañía. En ella se detalla el registro histórico de trabajadores que pertenecen a la organización (fueron contratados), en términos del desempeño en pruebas de ingreso. Dichos tests evalúan a los postulantes en una serie de ámbitos de interés para sus cargo.

Las variables del registro corresponden a:

1. ID: ID del trabajador.
2. AREA: con valores *N1* o *N2*. Representa el área de desempeño del trabajador al interior de la organización.

El siguiente conjunto de variables fue medido en una escala continua de 0 a 100, obtenida a partir de un cuestionario desarrollado por psicólogos laborales:

1. TM: capacidad de trabajo multitareas del postulante.
  2. TBP: capacidad de trabajo bajo presión del postulante.
  3. HLM: habilidad lógico matemática, medida a través de una serie de ejercicios lógicos.
  4. EE: desempeño en expresión escrita del postulante.
  5. DI: nivel de dominio de inglés del postulante.
  6. TE: capacidad de trabajo en equipo.
  7. DA: desempeño académico (pregrado) del postulante, evaluado a partir de sus antecedentes.
  8. AE: participación e involucramiento en actividades extraprogramáticas, artísticas o de servicio social.
  9. AR: autopercepción del nivel responsabilidad del trabajador.
  10. RUE: evaluación de las referencias del último empleo del postulante.
1. Escoja 5 de las 10 variables medidas en el cuestionario. Presente un resumen empleando tablas, gráficos o medidas que considere apropiadas y comente cómo cambian estos indicadores por grupo (Áreas *N1* y *N2*).
  2. Realice una comparación, en el contexto de estas 5 variables, que permita estudiar si existen diferencias significativas entre los grupos (*N1* y *N2*), cuidando de validar los supuestos necesarios y escogiendo el test de hipótesis que corresponda. Comente sus resultados y concluya.