

# Modelos no paramétricos y de regresión

## Tarea 1

Fecha de entrega: 11 de marzo

*Instrucciones.* La tarea se puede realizar en equipos de máximo 3 integrantes. Se entregará física (o digital) a la hora de la clase y no habrá prórroga. Las actividades de DataCamp son individuales y se podrán terminar a lo largo del día de la entrega de la tarea. Estas actividades contarán como un punto extra sobre la tarea y de no realizarlos habrá una penalización de un punto sobre la misma.

1. (1 punto) Sea  $R$  el número de rachas en una secuencia de  $n$  objetos con  $n_1$  objetos del tipo I y  $n_2$  objetos del tipo II. Considere a  $R_1$  el número de rachas del tipo I y realice lo siguiente:

- (a) Demuestra que la distribución marginal de  $R_1$  está dada por

$$f_{R_1} = \frac{\binom{n_1-1}{r_1-1} \binom{n_2+1}{r_1}}{\binom{n_1+n_2}{r_1}}.$$

$$\text{Hint: } \binom{n}{k} = \binom{n-1}{k-1} + \binom{n-1}{k}$$

- (b) Demuestra que la esperanza de  $R_1$  está dada por:

$$\mathbb{E}(R_1) = \frac{(n_2 + 1)n_1}{n_1 + n_2}.$$

$$\text{Hint: } \sum_{i=1}^x \binom{x-1}{i-1} \binom{y}{i-1} = \binom{x+y-1}{x-1}$$

- (c) ¿Qué puedes decir sobre la distribución marginal y la esperanza de  $R_2$ ?
- (d) Demuestra que la esperanza de  $R$  está dada por:

$$\mathbb{E}(R) = 1 + \frac{2n_1n_2}{n_1 + n_2}$$

2. (1 punto) Para la prueba de suma de rangos de Wilcoxon se define el estadístico  $W$  como

$$W = \sum_{i=1}^N iZ_i,$$

donde  $Z_i = 1$  si la  $i$ -ésima variable es  $x_l$  para alguna  $l$ . Demuestra lo siguiente

- (a)  $\mathbb{E}(Z_i) = \frac{m}{N}$ .
- (b)  $\text{Var}(Z_i) = \frac{mn}{N^2}$ .
- (c)  $\text{Cov}(Z_i, Z_j) = \frac{-mn}{N^2(N-1)}$ .

Con estos resultados encuentra la esperanza y la varianza de  $W$ .

3. (1 punto) Para la prueba de Kruskal-Wallis prueba que el estadístico  $H$  se puede escribir como:

$$H = \frac{12}{N(N+1)} \sum_{i=1}^k \frac{R_i^2}{n_i} - 3(N+1),$$

donde  $N = n_1 + \dots + n_k$ .

4. (1 punto) Para la prueba de Friedman prueba que el estadístico  $S$  se puede reescribir como:

$$S = \sum_{j=1}^k R_j^2 - \frac{n^2 k(k+1)^2}{4}.$$

5. (0.5 puntos) De 20 actuarios graduados de la Facultad de Ciencias se tiene que 15 de ellos pasaron el examen P de la SOA. ¿Esto prueba que la probabilidad de que un actuario graduado de la Facultad de Ciencias pase el examen P de la SOA sea mayor al promedio de la Ciudad de México, que es del 70%?
6. (1 punto) En un estudio sobre brucelosis, se midió el marcador CD4 de la sangre en un grupo de control y en un grupo de enfermos de brucelosis obteniendo los resultados que se pueden encontrar en la Cuadro 3. ¿Hay alguna diferencia en estas dos poblaciones?

Cuadro 1: Marcador CD4.

Control	59	66	45	62	51	50	49	58	53	42	50	47	51	62	48
Enfermos	72	70	69	82	68	59	76	61	59	73	49	77			

7. (1 punto) El consejo técnico de la Facultad de Ciencias, realizó un estudio sobre la ansiedad en los estudiantes al inicio y al final del semestre. ¿Qué puedes concluir de los niveles de ansiedad con los datos de la Cuadro 1, donde valores más altos indican mayor ansiedad?

Cuadro 2: Ansiedad de los estudiantes.

Alumno	Inicio	Final	Alumno	Inicio	Final
Lorien	20	20	Nicolás	19	34
Josué	18	21	José Luis	13	28
Martha	10	23	Luis Angel	21	20
José Manuel	16	26	Eduardo	12	29
Cristián	35	32	Ben	23	22
Emmanuel	20	27	Carol	14	30
Carlos Alberto	20	38	Karen	17	25

8. (1 punto) Un grupo farmacéutico está intentando conocer si hay diferencia en 4 medicinas utilizadas para reducir el dolor de personas que sufren migrañas. Para este estudio se le dio a 7 pacientes la medicina por un mes y midió en una escala del 1 al 10 (siendo 10 el dolor más severo) el dolor sufrido por las migrañas en ese mes. Los resultados se pueden apreciar en la Cuadro 2. ¿Qué le dirías a este experto?

Cuadro 3: Medición del dolor de migraña para 4 medicamentos diferentes.

Medicina/Paciente	Angel	Diego	César	Israel	Yahir	Celeste	Aaron
Ibuprofeno	7	10	7	9	8	8	8
Ácido acetilsalicílico	7	6	5	8	7	5	7
Naproxeno	3	7	3	5	4	6	3
Sumatriptán	4	3	2	1	0	1	0

9. (1 punto) Considera los 30 lapsos en minutos entre las erupciones del géiser Old Faithful del Parque Nacional de Yellowstone (Cuadro 4) ¿consideras que estos datos son aleatorios?

Cuadro 4: Tiempo entre erupciones del geyser Old Faithful.

68	63	66	63	61	44	60	62	71	62	62	55	62	67	73
72	55	67	68	65	60	61	71	60	68	67	72	69	65	66

10. (1 punto) En un concurso de comida, a 3 chefs se les evaluó por la preparación de un salmón a la naranja con arroz salvaje. Dentro de la calificación se consideró el aroma, el sabor y la textura del platillo por críticos internacionales. Los resultados del concurso se pueden apreciar en la Cuadro 5. ¿Encuentras alguna diferencia entre las calificaciones recibidas?

Cuadro 5: Calificaciones del concurso de comida.

Chef Alan	Chef Zaid	Chef Andrés Manuel
4.05	4.35	2.24
4.96	3.88	3.93
3.45	3.02	3.37
3.57	4.56	3.21
4.23	4.37	2.35
4.18	3.31	2.59

## Actividades de DataCamp

1. Sampling in R
2. Hypothesis Testing in R