



# PEP 02

## Instrucciones:

- Los ejercicios a continuación corresponden a la parte práctica de la segunda prueba asociada al ramo de análisis estadístico para ingeniería (en horario vespertino).
- El tiempo con el que cuentan para resolver la prueba es de 1 hora y 30 minutos.
- Al finalizar, deberán enviar sus respuestas al correo de la profesora.
- Usted debe subir un archivo **comprimido** el cual contendrá en primer lugar su **código en R ordenado, estructurado** y debidamente **comentado**. Y también, un documento **PDF** que contendrá de manera ordenada el desarrollo de cada pregunta con las imágenes asociadas a sus desarrollos en R y las respectivas respuestas.
- Cada ejercicio debe tener sus respectivas justificaciones, desarrollos y respuestas **completas** asociadas al problema realizado. El desarrollo de cada una de las preguntas debe ser de forma **clara y ordenada**.
- La prueba es de carácter individual, por ende, cualquier evidencia de copia será evaluada con la nota mínima 1.0 sin derecho a revisión.

## Primera pregunta (35 puntos)

Para su ramo de análisis de algoritmos y estructura de datos le solicitaron crear un algoritmo que sea capaz de analizar bases de datos de una empresa con el objetivo de clasificar en percentiles los sueldos de los empleados de manera gráfica y a través de un listado con cada nombre, sueldo y cuartil del empleado. El día que presentaron los resultados, el profesor los organizó en parejas, les entregó los computadores de la Universidad (todos con misma marca, modelo y año) y diversas bases de datos para analizar con sus algoritmos elaborados. Los resultados de rendimiento de los algoritmos (suyo y de su pareja) fueron los siguientes:

Algoritmo 1 (Creado por usted)		Algoritmo 2 (Creado por su pareja designada)	
Prueba	Tiempo (S)	Prueba	Tiempo (S)
1	3.2	1	2.2
2	3.0	2	2.7
3	2.9	3	2.6
4	3.5	4	2.5
5	2.7	5	2.3
6	3.1	6	2.0
7	3.2	7	2.5
8	2.9	8	2.2

9	3.0	9	2.1
10	2.8	10	2.7
11	3.1	11	2.6
12	3.1	12	2.1
13	3.5	13	2.5
14	3.0	14	2.0
15	2.7	15	2.3
16	2.9	16	2.2
17	3.0	17	2.2
18	3.5	18	2.1
19	3.0	19	2.6
20	2.7	20	2.4
21	2.9	21	2.0
22	3.2	22	2.6
23	3.4	23	2.8
25	3.1	25	2.5

- Genere en R una representación para los datos entregados, calcule los parámetros básicos (media, desviación estándar y coeficientes de variación) y explique a que conclusión puede llegar con los resultados obtenidos. (10 puntos).
- Evalué la normalidad de los datos entregados para cada algoritmo, con los diversos test revisados en clases, determine que test puede utilizar para comparar ambos algoritmos. Fundamente su respuesta en detalle. (10 puntos)
- Aplice un test correspondiente a este ejercicio que le permita determinar y comparar los rendimientos de ambos algoritmos, explique los resultados obtenidos y compare con los resultados obtenidos en la parte a (10 puntos).
- Dentro de los dos grupos presentados, determine mediante algún test si los datos presentan o no homocedasticidad y explique los resultados y conclusiones obtenidas. (5 puntos)

### Segunda pregunta (25 puntos)

En este ejercicio, usted analizará un conjunto de datos que reflejan la relación que existe entre la expectativa de vida en años de las personas en algunos países del mundo y el gasto per cápita en salud (Datos correspondientes al año 2015 recolectados desde → <https://ourworldindata.org/grapher/life-expectancy-vs-health-expenditure> ).

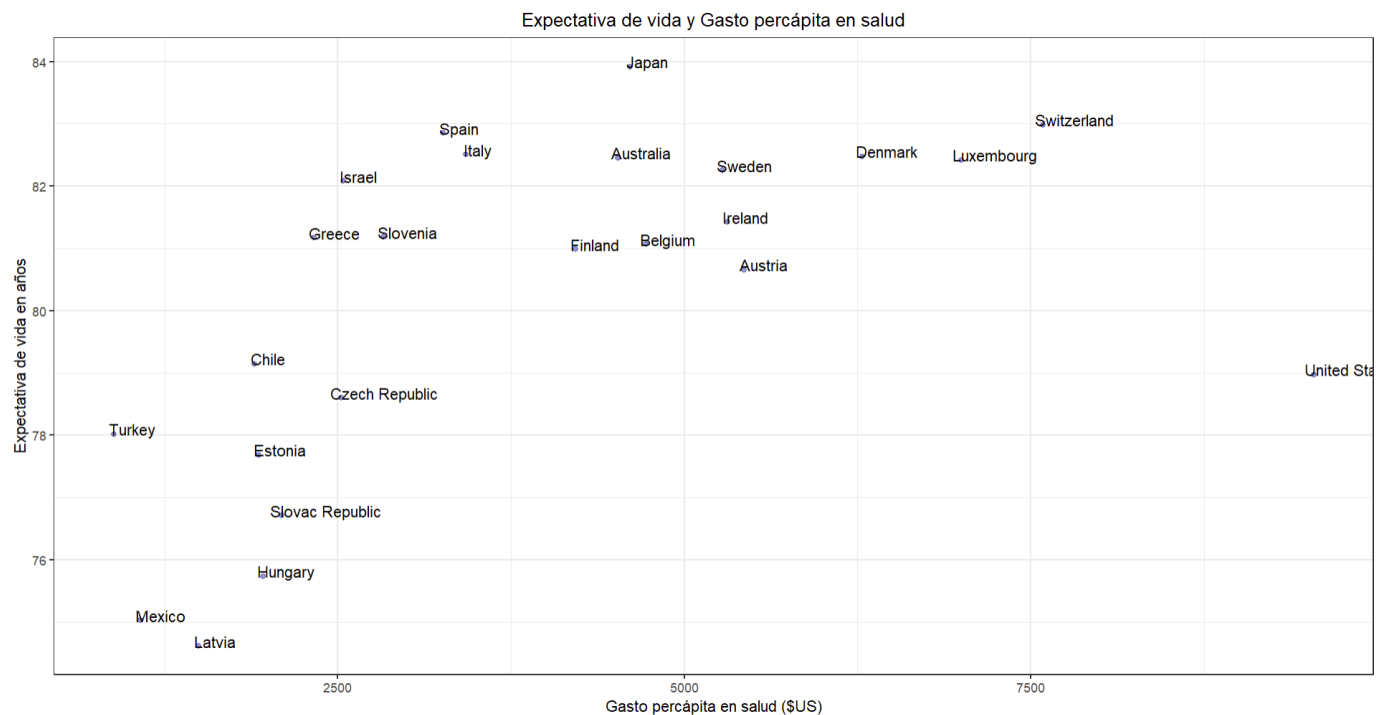
Los datos se han organizado y se encuentran en el archivo “.csv” disponible junto al enunciado.

Los datos se pueden graficar de la siguiente manera:

```
datos=read.csv("Datos_pregunta2.csv",sep=";",header=T)

grafico=ggplot(datos,aes(Gasto_salud_dolares ,expectativa_anual,,label=Pais)) +
  geom_point(aes(Gasto_salud_dolares,expectativa_anual),datos,color="darkblue",alpha=0.4) +
  theme_bw() + xlab("Gasto per cápita en salud ($US)") + ylab("Expectativa de vida en años") +
  ggtitle("Expectativa de vida y Gasto per cápita en salud") +
  theme(plot.title = element_text(hjust = 0.5)) + geom_text(hjust=0.1, vjust=0.1)
```

plot(grafico)



- Con los datos presentados previamente, proponga un modelo de regresión lineal que relacione las variables presentadas (la expectativa de vida en años (variable dependiente) con el gasto en salud per cápita (variable independiente)). Evalúe el modelo creado en base a los criterios que fueron revisados en clases. (10 puntos)
- Con los datos obtenidos previamente, grafique el intervalo de confianza asociado a los datos y el modelo de regresión, analice los resultados a lo largo del ejercicio y explique si el modelo de regresión se ajusta a los datos. Según lo visto en clases, ¿Qué otro tipo de regresión lineal se adaptaría de mejor manera a los datos? ¿Por qué?. (15 puntos)