

## Prueba EAE 253 B

Profesor: Carlos Alvarado  
Ayudante: Pablo González

31 de octubre de 2018

### INSTRUCCIONES

1. Lea todas las instrucciones con calma.
2. Escriba su número de alumno en todas las hojas.
3. Escriba sólo por el anverso de la hoja.
4. Si requiere más espacio, al final tiene 2 hojas extras.
5. NO extraiga hojas de la prueba. Debe entregarlas todas al final.
6. Puntaje máximo: 100 puntos (110 puntos totales)

### CÓDIGO DE HONOR

En el desarrollo personal del alumno y futuro egresado de la Pontificia Universidad Católica de Chile la integridad académica es fundamental. Ella se basa en cinco valores: honestidad, confianza, imparcialidad, respeto y responsabilidad. Para obtener el mejor provecho de este curso, el profesor y los estudiantes deben trabajar juntos en un ambiente que propicie estos valores.

Acciones que violen estos valores son incompatibles con los objetivos de esta Facultad en el proceso de formación de sus alumnos y no serán aceptados. Se espera que cada estudiante promueva siempre el espíritu de integridad académica, reforzándolo entre sus compañeros.

**Declaro conocer el Código de Honor de la Facultad de Ciencias Económicas y Administrativas de la Pontificia Universidad Católica de Chile.**

**Declaro estar en condiciones de salud adecuadas para asistir a la evaluación y que es mi responsabilidad decidir rendirla. En caso de sentirme mal o tener alguna complicación, deberé informarlo inmediatamente al ayudante en sala.**

Nombre Alumno: \_\_\_\_\_

Firma Alumno: \_\_\_\_\_

## 1. Preguntas Cortas (25 puntos)

1. [5 puntos] Escriba una función o procedimiento que le permita saber si un valor  $X$  determinado se encuentra dentro de una lista de números llamada  $L$ . Suponiendo que  $n = \text{len}(L)$ , ¿cuántas operaciones realizará su procedimiento en términos de  $n$ ?

2. [5 puntos] Escriba el *output* del siguiente código:

```
c = True
d = False
c = c and d
c = not c or d

print(c)
```

3. [15 puntos] Al realizar análisis de datos, en ocasiones se recurre a un proceso llamado *winsorizing*, cuyo objetivo es reducir el impacto de valores extremos (outliers) en las estimaciones. El proceso consiste en reemplazar el valor de los outliers por el valor correspondiente a algún percentil determinado. Por ejemplo, al “winsorear” una variable al 90 %, se reemplaza el valor de todas las observaciones en el 5 % inferior y 5 % superior de la variable por el valor de los percentiles 5 % y 95 % respectivamente. Escriba una función o procedimiento que le ayude a implementar este proceso sobre una lista de números llamada  $L$ , usando un nivel de “winsor” al  $X$  %. Su función o procedimiento debe retornar una nueva lista “winsoreada” al  $X$  %.

## 2. Votaciones I (Python - 40 puntos)

En su universidad se están realizando votaciones para elegir al próximo Presidente del Centro de Alumnos. Usted es parte del equipo organizador de las votaciones, y le pidieron ayuda con el procesamiento y análisis de los datos.

El primer código que usted escribió es el siguiente:

```
In [1]: f = open('datos_mesas_votacion.txt', encoding='utf-8-sig')

        data = []
        for line in f:
            data.append(tuple([int(x) for x in line.strip().split()]))

        print(data)

[(1, 20, 25), (4, 30, 12), (5, 52, 32), (18, 17, 8), (32, 35, 40),
(40, 6, 8), (1, 22, 22), (18, 8, 9), (123, 26, 28)]
```

El archivo **datos\_mesas\_votación.txt** es una muestra de los datos de votaciones finales que usted recibirá. Para procesar dicho archivo, usted guardó los datos en un objeto **data**, que corresponde a una lista de tuplas, donde cada tupla contiene información sobre el resultado parcial observado en una mesa de votación dada. En particular, el primer elemento de la tupla es el código de la mesa de votación, el segundo elemento es el número de votos que recibió el candidato A en el correspondiente conteo parcial, y el tercer elemento es el número de votos por el candidato B del correspondiente conteo parcial.

Cabe destacar que, en la medida que se realicen nuevos cálculos parciales, los nuevos archivos de texto incluirán más filas para reflejar dichos nuevos cálculos (**note, por ejemplo, que las mesas 1 y 18 ya tienen 2 cálculos parciales cada una**). Además usted sabe que, en total, **hay 350 mesas de votación que eventualmente enviarán su información**.

### Preguntas:

1. [5 puntos] Escriba un procedimiento o función que tome la lista **data** y calcule el total de votos para cada candidato. Su función debería retornar dos valores, el primero correspondiente a la suma de votos del candidato A, y el segundo, correspondiente a la suma de votos del candidato B.
2. [10 puntos] Preocupados por los votos de las mesas faltantes, le pidieron que analizara qué mesas aún no han mandado sus votaciones. En particular, se le pide a usted que escriba un procedimiento o función que tome el objeto **data** y retorne una lista, donde los elementos de la lista son los números de las mesas que aún no han enviado información (recuerde, existen 350 mesas, numeradas de la 1 a la 350).

3. [10 puntos] Escriba un procedimiento que tome la información en **data** y calcule el número de votos totales por mesa para cada candidato. En particular, su procedimiento debiera entregar un diccionario, donde la llave sea el número de mesa, y el “valor” asociado corresponda al total de votos por mesa.

4. [10 puntos] Otra persona de su equipo le entregó código, pero no le explicó para qué era. Describa qué hace el siguiente código y usando los datos de muestra del enunciado escriba cuál es el *output* de éste.

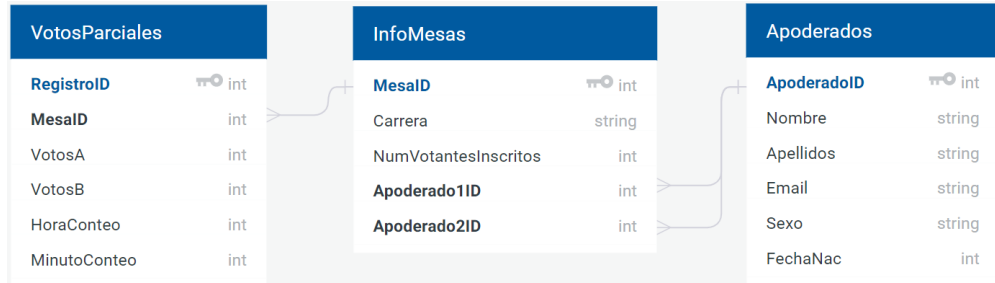
```
num1 = 0
num2 = 0
for row in data:
    if row[1]/row[2] > num2:
        num1 = row[0]
        num2 = row[1]/row[2]

print(num1, num2)
```

5. [5 puntos] Conceptualmente, ¿usted considera que el código anterior cumple su propósito? ¿por qué sí? ¿por qué no? (no se requiere codificar; sólo una respuesta conceptual).

### 3. Votaciones II (SQL - 45 puntos)

Después de las elecciones, su candidato (A) perdió, y le pidieron a Ud. analizar los datos de votaciones para tratar de entender qué pasó. Afortunadamente, consiguieron la información en una base de datos estructurada (SQL), con la siguiente descripción:



#### Preguntas:

1. [5 puntos] Construya una query que le permita saber el número de votos obtenidos por cada candidato, por carrera.



2. [10 puntos] Una de las hipótesis que maneja el equipo es que después de almuerzo (2pm) su candidato (A) recibió muchos menos votos, ya que a la misma hora se viralizó un video suyo en las redes sociales... ¿cómo testearía esta hipótesis? Construya una query o queries que le permita obtener los datos necesarios para testear dicha hipótesis.

3. [15 puntos] Construya una query que le permita determinar cuál fue la carrera donde más bajó el número de votos de su candidato (A) después de las 2pm. En concreto, se pide que compare el porcentaje de votos por carrera antes y después de las 2pm de su candidato (A), calculando la diferencia de porcentajes en términos absolutos (por ejemplo, si en una determinada carrera el porcentaje fue 70 % antes de las 2pm y 40 % después de esa hora, entonces la diferencia fue de -30 %).

# Alumno: \_\_\_\_\_

4. [10 puntos] Construya una query que le permita obtener la tasa de participación por carrera (Ayuda: la variable *NumVotantesInscritos* es el número de votantes totales inscritos **por mesa**. En una determinada carrera puede haber más de 1 mesa de votación.)
5. [5 puntos] ¿Qué variable o variables adicionales agregaría usted al análisis? ¿cómo conseguiría esos datos? (Pregunta abierta; no existe una respuesta correcta...)

# Alumno: \_\_\_\_\_

# Alumno: \_\_\_\_\_