

Tarea 1. Estadística

Katherin A. Murcia S. y Juan Carlos Rojas V.

```
In [1]: import numpy as np
import matplotlib.pyplot as plt
import pandas as pd
```

1. Mean, standard deviation and standard error

An experiment was conducted to determine the concentration of a sodium hydroxide solution. The eight repeat measurements of the volume of hydrochloric acid titrated (al in mL) are: 25.8, 26.2, 26.0, 26.5, 25.8, 26.1, 25.8 and 26.3. Calculate **(i)** the mean, **(ii)** the standard deviation using the rough-and-ready approach; **(iii)** the standard deviation using equation (2.3); **(iv)** the standard error of the volume.

Solución:

(i) Para hallar el valor promedio de la medición de concentración de ácido clorhídrico en mL se debe usar la fórmula

$$\bar{x} = \frac{1}{N} \sum_{i=1}^N x_i$$

donde N es el número de volúmenes medidos y x_i son las medidas tomadas.

```
In [31]: # Generación de la lista de datos
V = np.array([25.8, 26.2, 26.0, 26.5, 25.8, 26.1, 25.8, 26.3]) # en mL

# Número de datos tomados
N = len(V)

"""
Usando la función np.sum() se hará la sumatoria de los datos sobre la lista
de datos V y luego dividir por el número de datos N. Así:
"""

mean_value = np.sum(V)/N
mean_value
```

```
Out[31]: 26.0625
```

Por lo tanto el valor medio de los volúmenes de ácido clorhídrico es de 26.1 mL.

(ii) Para encontrar la desviación estándar usando el *método tosco* se debe utilizar el spread entre el valor medio y valor máximo o mínimo de la lista de datos y luego multiplicar este valor por 2/3. Por tanto, se calcula la spread máximo $d_{\text{máx}} = \bar{x} - x_{\text{mín}}$ o $d_{\text{máx}} = x_{\text{máx}} - \bar{x}$

```
In [3]: #Cálculo del valor mínimo y máximo de la lista de datos
x_max, x_min = np.max(V), np.min(V)
x_max, x_min

(26.5, 25.8)
```

Out[3]:

```
In [4]: #Cálculos del spread máximo por con el valor mínimo y máximo de la lista.
d1 = mean_value - x_min
d2 = x_max - mean_value
d1,d2
```

Out[4]: (0.26249999999999993, 0.4375)

Dado que spread máximo más grande es 0.4 mL este se tomará para calcular la desviación estandar así

$$\sigma = \frac{2}{3}d_{\text{máx.}}$$

```
In [33]: std_dev = (2/3)*d2
std_dev
```

Out[33]: 0.29166666666666663

Por lo que la desviación estándar es de 0.3 mL.

(iii) Para encontrar la desviación se utiliza la fórmula

$$\sigma_{N-1} = \sqrt{\frac{1}{N-1} \sum_{i=1}^N d_i^2}$$

donde $d_i = x_i - \bar{x}$.

```
In [6]: # Generación de la lista de los spread para cada uno de los datos
d_list = V - mean_value
d_list
print("Lista de spreads para cada dato:%s"%( d_list))

#Generación de lista de los spread elevados al cuadrado

d_sqrd_list = d_list**2
print("Lista de spreads al cuadrado para cada dato: %s"%(d_sqrd_list))
```

```
Lista de spreads para cada dato:[-0.2625  0.1375 -0.0625  0.4375 -0.2625  0.0375 -0.2625
 0.2375]
Lista de spreads al cuadrado para cada dato: [0.06890625 0.01890625 0.00390625 0.1914062
5 0.06890625 0.00140625
 0.06890625 0.05640625]
```

```
In [7]: # Calculo de la desviación estándar por medio de la fórmula
std_dev_real = np.sqrt((1/(N-1))*np.sum(d_sqrd_list))
std_dev_real
```

Out[7]: 0.26152028055746845

Por lo tanto la desviación estándar es $\sigma = 0.3$ mL.

(iv) Para hallar el error en la medición de los volúmenes está dado por

$$\alpha = \frac{\sigma_{N-1}}{\sqrt{N}}$$

por lo que el error está dado por:

```
In [8]: # Calculo del error estándar de los volúmenes
alpha = std_dev_real/np.sqrt(N)
```

alpha

Out[8]: 0.09246138189999717

Por tanto el error estándar está dado por $\alpha = 0.09$

2. Reduction of the standard error

In a magnetometry experiment, after a minute of collecting data the statistical noise was reduced to 1 picotesla. For how much longer should data be collected in order to reduce the random error by a factor of 10?

Solución:

Suponiendo que la rata con la que se toman datos es constante y recordando la fórmula para la incertidumbre de la medida está dada por

$$\alpha = \frac{\sigma_{N-1}}{\sqrt{N}}$$

Por lo que si se desea reducir el error por 10, entonces se tiene que, si t es el tiempo que se hace las medidas y M es la rata en que se miden los datos, se tiene que

$$\begin{aligned}\alpha' &= \alpha \times 10^{-1} \\ &= \frac{1}{10} \frac{\sigma_{N-1}}{\sqrt{N}} \\ &= \frac{\sigma_{N-1}}{\sqrt{100N}} \\ &= \frac{\sigma_{N-1}}{\sqrt{100t \cdot M}}\end{aligned}$$

Por lo que se debería hacer el experimento 100 veces más largo. Esto es, en vez de 1 min, se debería tomar por 100 min. Esto es una 1 hora y 40 minutos.

3. Error in the error

Consider a set of measurements with the standard error calculated to be $\alpha = 0.987\,654\,321$. Here we address the question of how many significant figures should be quoted. Construct a spreadsheet with four columns. The first column should be N , the number of measurements on which α is based. In the second column write α to the nine significant figures quoted above. The third and fourth columns should be $\alpha \times \left(1 - \frac{1}{\sqrt{2N-2}}\right)$ and $\alpha \times \left(1 + \frac{1}{\sqrt{2N-2}}\right)$, respectively. As we are interested in the variation over a large dynamic range, choose values for N such as 2, 3, 5, 10, 20, 30, etc. Verify the statement from Section 2.7.1 that the number of data points, N , needs to approach a few tens of thousands before the second significant figure in the error can be quoted, i.e. when the values in the three columns become equal to the second significant figure. Repeat the analysis for the case where $\alpha = 0.123\,456\,789$, i.e. the first significant digit of the error is 1. How many data points must be collected before the third significant figure can be quoted?

Solución:

```
In [29]: # Generación de lista de números y error en el error.
N_ = np.array([5,10,20,50,100,500,1000,5250,7500,10000,12000,15000,30000,30000,50000,69500])
alpha_ = 0.987654321
a_minus = np.round(alpha_*(1 - 1/(np.sqrt(2*N_-2))),2)
a_plus = np.round(alpha_*(1 + 1/(np.sqrt(2*N_-2))),2)

data_frame_1 = {r"$N$":N_,r"$\alpha$":np.ones_like(N_)*0.123456789,r"$\alpha_{-}$":a_minus,
r"$\alpha_{+}$":a_plus}
data_F_1 = pd.DataFrame(data_frame_1)
data_F_1
```

```
Out[29]:
```

	N	α	α_-	α_+
0	5	0.123457	0.64	1.34
1	10	0.123457	0.75	1.22
2	20	0.123457	0.83	1.15
3	50	0.123457	0.89	1.09
4	100	0.123457	0.92	1.06
5	500	0.123457	0.96	1.02
6	1000	0.123457	0.97	1.01
7	5250	0.123457	0.98	1.00
8	7500	0.123457	0.98	1.00
9	10000	0.123457	0.98	0.99
10	12000	0.123457	0.98	0.99
11	15000	0.123457	0.98	0.99
12	30000	0.123457	0.98	0.99
13	30000	0.123457	0.98	0.99
14	50000	0.123457	0.98	0.99
15	69500	0.123457	0.99	0.99
16	700000	0.123457	0.99	0.99

El efecto mencionado en la sección se muestra cuando $N \sim 70000$.

Para el otro error, se tiene que

```
In [20]: # Generación de lista de números y error en el error.
N_2 = np.array([5,10,20,50,100,500,1000,5250,7500,10000,12000,15000,30000,1000000,1100000])
alpha_2 = 0.123456789
a_minus = np.round(alpha_2*(1 - 1/(np.sqrt(2*N_-2))),3)
a_plus = np.round(alpha_2*(1 + 1/(np.sqrt(2*N_-2))),3)

data_frame = {r"$N$":N_2,r"$\alpha$":np.ones_like(N_2)*0.123456789,r"$\alpha_{-}$":a_minus,
r"$\alpha_{+}$":a_plus}
data_F = pd.DataFrame(data_frame)
data_F
```

```
Out[20]:
```

	N	α	α_-	α_+
0	5	0.123457	0.080	0.167
1	10	0.123457	0.094	0.153

2	20	0.123457	0.103	0.143
3	50	0.123457	0.111	0.136
4	100	0.123457	0.115	0.132
5	500	0.123457	0.120	0.127
6	1000	0.123457	0.121	0.126
7	5250	0.123457	0.122	0.125
8	7500	0.123457	0.122	0.124
9	10000	0.123457	0.123	0.124
10	12000	0.123457	0.123	0.124
11	15000	0.123457	0.123	0.124
12	30000	0.123457	0.123	0.124
13	1000000	0.123457	0.123	0.124
14	11000000	0.123457	0.123	0.123

Para este caso, el efecto se vio cuando $N \sim 11\,000\,000$.

4. Reporting results

Analysis of a Rydberg spectrum yields a quantum defect, δ , for each line. How would you report the results if you obtain (i) $\bar{\delta} = 3.273\,46$, with $\alpha_{\delta} = 0.019\,13$ from five measurements; (ii) $\bar{\delta} = 3.265\,13$, with $\alpha_{\delta} = 0.002\,506$ from 50 measurements; or (iii) $\bar{\delta} = 3.266\,81$, with $\alpha_{\delta} = 0.000\,270$ from 100 measurements?

Solución:

Para decidir cuántas cifras significativas se necesitan, se deben calcular los errores en los errores igual que en el ejercicio anterior.

(i) Para $\bar{\delta} = 3.27346$ y $\alpha_{\delta} = 0.01913$ con $N = 5$.

Dado que la muestra no es representativa, se debe utilizar únicamente la primera cifra significativa de la incertidumbre, así:

$$\delta = 3.27 \pm 0.02$$

(ii) Para $\bar{\delta} = 3.26513$ y $\alpha_{\delta} = 0.002506$ y $N = 50$.

Para ver si se necesita la segunda cifra significativa, se calculan los errores en los errores así:

```
In [11]: #Errores en el error
alpha2 = 0.002506
N2=50
alpha_plus2 = 0.002506*(1 + 1/np.sqrt(2*N2 -2))
alpha_minus2 = 0.002506*(1 - 1/np.sqrt(2*N2 -2))
alpha_minus2, alpha_plus2
```

```
Out[11]: (0.002252855772335216, 0.0027591442276647836)
```

Dado que las segundas cifras significativas con diferentes, entonces se tiene se debe reportar como

$$\delta = 3.265 \pm 0.002$$

.

(iii) Para $\bar{\delta} = 3.26681$ y $\alpha_{\delta} = 0.00027$ con $N = 100$.

```
In [12]: #Errores en el error
alpha3 = 0.000270
N3=100
alpha_plus3 = alpha3*(1 + 1/np.sqrt(2*N2 -2))
alpha_minus3 = alpha3*(1 - 1/np.sqrt(2*N2 -2))
alpha_minus3,alpha_plus3

Out[12]: (0.00024272588129709032, 0.0002972741187029097)
```

Dado que las segundas cifras significativas en el error en el error son las mismas, entonces se debe reportar como

$$\delta = 3.2668 \pm 0.0003$$

5. Scientific notation

Rewrite the then numbers from Exercise (2.7) in scientific notation.

1. $602.20 = 6.022 \times 10^2$
2. $0.001\ 380\ 6 = 1.3806 \times 10^{-3}$
3. $0.022\ 413\ 83 = 2.241383 \times 10^{-2}$
4. $1.602\ 19 = 1.602\ 19 \times 10^0$
5. $91.095 = 9.1095 \times 10^1$
6. $0.1660 = 1.660 \times 10^{-1}$
7. $299\ 790\ 000 = 2.9979 \times 10^8$
8. $66.2617 = 6.62617 \times 10^1$
9. $0.000\ 006\ 672 = 6.672 \times 10^{-6}$
10. $3.141\ 593 = 3.141593 \times 10^0$