

# Actividad en Clase



Análisis Univariado: la Media

# Instrucciones

1. Crear un archivo de Word para esta Actividad en Clase, con el nombre

**202310 Actividad 2.1: Análisis univariado**

compartido con [nataliah@uninorte.edu.co](mailto:nataliah@uninorte.edu.co).

2. En este documento, debe copiarse el comando utilizado, seguido de la salida obtenida, y responder las preguntas de análisis.

**Recuerden que lo importante es el intento, no la perfección.**

# Cargar los datos en R

1. Recibirán una muestra en clase.
2. Copie los datos de su muestra en R como un vector.

Ejemplo:

Si mi muestra fuera de tres (3) notas 4.5, 5 y 2.4.

```
muestra <- c(4.5, 5, 2.4)
```

*Recuerden que los decimales se separan con puntos y los datos con comas*

# Primera Parte

# Distribución de los datos muestrales

3. Haga un histograma de las notas de su muestra.

Este histograma representa la **distribución de los datos muestrales**.

Modifique el título de su histograma por “Muestra \_” agregando lo siguiente al comando:

```
hist(muestra, main ="Muestra 11")
```

4. Exporte el histograma como una imagen, péguela en su documento y compártala en [este Padlet](#).

# Parámetros y Estadísticos

5. Complete esta tabla con el resumen estadístico de su muestra.

¿Hay algún parámetro que sea igual a su estadístico correspondiente?

**Para tener en cuenta:** Normalmente no conocemos el valor real de los parámetros. Si conociéramos estos valores poblacionales, no necesitaríamos trabajar con muestras.

	Parámetro	Estadístico
<b>Media</b>	3,54	
<b>Mediana</b>	3,8	
<b>Varianza</b>	0,9892	
<b>D.E.</b>	0,9946	
<b>Mínimo</b>	0,7	
<b>Q1</b>	2,8	
<b>Q2</b>	3,8	
<b>Q3</b>	4,3	
<b>Máximo</b>	5	

# Error muestral

A esta diferencia entre el parámetro y el estadístico se le llama **error muestral**.

Lo ideal sería que ambas medidas fueran iguales (es decir, que el error muestral sea igual a cero). Sin embargo, es poco común que eso ocurra.

# Distribución de los posibles valores la media muestral

6. Anoten en el tablero su media muestral de tal forma que todos puedan ver los posibles valores de la media por muestra.

7. Haga un histograma de estos posibles valores de la media.

Este histograma representa la distribución de los posibles valores de la media llamada **distribución muestral de la media**.

En inglés, se llama *sampling distribution*, una traducción más adecuada puede ser **distribución de muestreo**.

¿La forma de la distribución que se representa en el histograma es conocida?

Muestra	Media Muestral
1	
2	
3	
4	
...	
15	



# Segunda Parte

# ¿Qué tan probable es...?

El comando **pnorm** nos indica la probabilidad de que ocurra cierto valor de  $z$  o menos.

## Ejemplo:

“Probabilidad de que  $z$  sea menor a 1”:

```
> pnorm(1)  
[1] 0.8413447
```

Es decir, la probabilidad de que  $z$  sea menor a 1 es de 0.8413447, es decir, 84,1%.

8. Verifique el valor del ejemplo en la tabla normal (siguiente diapositiva).

9. Escoja cualquier valor de  $z$  en la tabla normal y verifique la probabilidad de que ocurra ese  $z$  o menor sea igual al propuesto en la tabla normal.

normal	0,00	0,01	0,02	0,03	0,04	0,05	0,06	0,07	0,08	0,09
0,0	0,50000	0,50399	0,50798	0,51197	0,51595	0,51994	0,52392	0,52790	0,53188	0,53586
0,1	0,53983	0,54380	0,54776	0,55172	0,55567	0,55962	0,56356	0,56749	0,57142	0,57535
0,2	0,57926	0,58317	0,58706	0,59095	0,59483	0,59871	0,60257	0,60642	0,61026	0,61409
0,3	0,61791	0,62172	0,62552	0,62930	0,63307	0,63683	0,64058	0,64431	0,64803	0,65173
0,4	0,65542	0,65910	0,66276	0,66640	0,67003	0,67364	0,67724	0,68082	0,68439	0,68793
0,5	0,69146	0,69497	0,69847	0,70194	0,70540	0,70884	0,71226	0,71566	0,71904	0,72240
0,6	0,72575	0,72907	0,73237	0,73565	0,73891	0,74215	0,74537	0,74857	0,75175	0,75490
0,7	0,75804	0,76115	0,76424	0,76730	0,77035	0,77337	0,77637	0,77935	0,78230	0,78524
0,8	0,78814	0,79103	0,79389	0,79673	0,79955	0,80234	0,80511	0,80785	0,81057	0,81327
0,9	0,81594	0,81859	0,82121	0,82381	0,82639	0,82894	0,83147	0,83398	0,83646	0,83891
1,0	0,84134	0,84375	0,84614	0,84849	0,85083	0,85314	0,85543	0,85769	0,85993	0,86214
1,1	0,86433	0,86650	0,86864	0,87076	0,87286	0,87493	0,87698	0,87900	0,88100	0,88298
1,2	0,88493	0,88686	0,88877	0,89065	0,89251	0,89435	0,89617	0,89796	0,89973	0,90147
1,3	0,90320	0,90490	0,90658	0,90824	0,90988	0,91149	0,91308	0,91466	0,91621	0,91774
1,4	0,91924	0,92073	0,92220	0,92364	0,92507	0,92647	0,92785	0,92922	0,93056	0,93189
1,5	0,93319	0,93448	0,93574	0,93699	0,93822	0,93943	0,94062	0,94179	0,94295	0,94408
1,6	0,94520	0,94630	0,94738	0,94845	0,94950	0,95053	0,95154	0,95254	0,95352	0,95449
1,7	0,95543	0,95637	0,95728	0,95818	0,95907	0,95994	0,96080	0,96164	0,96246	0,96327
1,8	0,96407	0,96485	0,96562	0,96638	0,96712	0,96784	0,96856	0,96926	0,96995	0,97062
1,9	0,97128	0,97193	0,97257	0,97320	0,97381	0,97441	0,97500	0,97558	0,97615	0,97670
2,0	0,97725	0,97778	0,97831	0,97882	0,97932	0,97982	0,98030	0,98077	0,98124	0,98169
2,1	0,98214	0,98257	0,98300	0,98341	0,98382	0,98422	0,98461	0,98500	0,98537	0,98574
2,2	0,98610	0,98645	0,98679	0,98713	0,98745	0,98778	0,98809	0,98840	0,98870	0,98899
2,3	0,98928	0,98956	0,98983	0,99010	0,99036	0,99061	0,99086	0,99111	0,99134	0,99158
2,4	0,99180	0,99202	0,99224	0,99245	0,99266	0,99286	0,99305	0,99324	0,99343	0,99361

# ¿Qué tan probable es...?

El comando **pnorm** nos indica también la probabilidad de que ocurra cierto valor de  $x$  o menos de una distribución normal cualquiera con cierta media (mean) y cierta desviación estándar (sd).

## Ejemplo:

“Probabilidad de que  $x$  sea menor a 11 de una distribución normal con media 8 y desviación estándar 4”:

```
> pnorm(11,mean=8,sd=4)
[1] 0.7733726
```

Es decir, la probabilidad de que  $x$  sea menor a 11 con esos parámetros es de 0.7733726, es decir, 77,3%.

10. Verifique este valor de manera manual en la tabla normal.

**Pista:** Para esto, debe estandarizar el valor de  $x$  teniendo en cuenta los parámetros del ejemplo.

11. Revise el comando `pnorm` con la ayuda `?pnorm` e identifique cómo calcular “la probabilidad de que  $x$  sea **mayor** a 11 de una distribución normal con media 8 y desviación estándar 4”. Calcule la probabilidad de este evento.

**Pista:** Si lo hizo bien, la suma de la probabilidad del punto 10 y el punto 11 debe dar 1 ó 100%. **¿A qué se debe esto?**

# Volviendo al caso de las notas...

12. Se sabe que en ciertas condiciones:

La distribución muestral de la media sigue una **distribución normal** con

- media = media poblacional
- varianza = varianza poblacional / tamaño de muestra

Calcule con ayuda de R los siguientes eventos con la población de notas resumida en la tabla:

- Probabilidad de que salga una muestra aleatoria de tamaño 20 con una media muestral menor a 3.
- Probabilidad de que salga una muestra aleatoria de tamaño 20 con una media muestral mayor a 3.
- Probabilidad de que salga una muestra aleatoria de tamaño 20 con una media muestral entre 3 y 4.

	Parámetro
<b>Media</b>	3,54
<b>Mediana</b>	3,8
<b>Varianza</b>	0,9892
<b>D.E.</b>	0,9946
<b>Mínimo</b>	0,7
<b>Q1</b>	2,8
<b>Q2</b>	3,8
<b>Q3</b>	4,3
<b>Máximo</b>	5

# Tercera Parte

# ¿Entre qué valores se encuentra...?

El comando **qnorm** nos indica el valor de  $z$  que tiene cierta probabilidad acumulada.

## Ejemplo:

“El valor de  $z$ , de tal forma que la probabilidad de que pase él o algo menor es de 75%”:

```
> qnorm(.75)  
[1] 0.6744898
```

Es decir, el valor de  $z$  que tiene una probabilidad acumulada de 75% es 0,6744898, es decir, 0,67.

13. Verifique este valor en la tabla normal ([Diapositiva 11](#)).

14. Escoja cualquier valor de probabilidad en la tabla normal y verifique que el valor de  $z$  correspondiente sea igual al propuesto en la tabla normal.

# ¿Entre qué valores se encuentra...?

El comando **qnorm** nos indica el valor de  $x$  de una distribución normal cualquiera con cierta media (mean) y cierta desviación estándar (sd) que tiene cierta probabilidad acumulada.

## Ejemplo:

“El valor de  $x$  de una distribución normal con media 8 y desviación estándar 4, de tal forma que la probabilidad de que pase él o algo menor es de 75%”:

```
> qnorm(.75, mean = 8, sd = 4)
[1] 10.69796
```

Es decir, el valor de  $x$  de una distribución normal con media 8 y desviación estándar 4 que tiene una probabilidad acumulada de 75% es 10.69796, es decir, 10.7.

15. Verifique este valor de manera manual en la tabla normal ([Diapositiva 11](#)).

**Pista:** Para esto, debe desestandarizar el valor de  $z$  encontrado en la tabla teniendo en cuenta los parámetros del ejemplo.

16. Revise el comando `qnorm` con la ayuda de `?qnorm` e identifique cómo encontrar “el valor  $x$  de una distribución normal con media 8 y desviación estándar 4 de tal forma que la probabilidad de que pase  $x$  o algo mayor sea de 75%.

**Pista:** Si lo hizo bien, el valor de  $x$  obtenido en el punto 16 menos la media 8, es el mismo valor que lo que se obtiene al restar la media 8 menos el resultado del ejemplo (10.69796). **¿A qué se debe esto?**



# ¿Entre qué valores se encuentra...?

17. Calcule con ayuda de R los siguientes eventos:

- a) El valor de  $z$ , de tal forma que la probabilidad de que pase él o algo menor es de 5%.
- b) El valor de  $z$ , de tal forma que la probabilidad de que pase él o algo mayor es de 5%.
- c) Los dos valores de  $z$ , de tal forma que entre los dos se encuentren el 90% de los casos centrales.

18. Repita el punto 17 cambiando los valores 5%, 5% y 90% por 2,5%, 2,5% y 95%, respectivamente.

19. Repita el punto 17 cambiando los valores 5%, 5% y 90% por 0,5%, 0,5% y 99%, respectivamente.

# Volviendo al caso de las notas...

Recuerde que en ciertas ocasiones (como ésta):

La distribución muestral de la media sigue una **distribución normal** con

- media = media poblacional
- varianza = varianza poblacional / tamaño de muestra

20. Tome los valores de  $z$  del apartado 17.c) y desestandarícelos teniendo en cuenta que la muestra de notas es de tamaño 20 y los parámetros de la población son los de la tabla.

**Pista:** pueden apoyarse de lo hecho en la [diapositiva 15](#).

21. Repita el punto 20 con los valores de  $z$  del apartado 18.c)

22. Repita el punto 20 con los valores de  $z$  del apartado 19.c)

23. Describa en palabras estos rangos (puntos 20, 21 y 22) qué describen.

	Parámetro
<b>Media</b>	3,54
<b>Mediana</b>	3,8
<b>Varianza</b>	0,9892
<b>D.E.</b>	0,9946
<b>Mínimo</b>	0,7
<b>Q1</b>	2,8
<b>Q2</b>	3,8
<b>Q3</b>	4,3
<b>Máximo</b>	5