Ejercicios de repaso, con respuesta, para el segundo parcial 2024 2c

1) Calcular la varianza y el desvío estándar de la variable x, cuyos valores observados fueron:

$$\boldsymbol{x_i} = \mathbf{60; 65; 75; 80}$$

Resp:

$varianza\ \boldsymbol{\sigma^2} = \mathbf{62,57} \quad desvío\ estándar\ \boldsymbol{\sigma} = \mathbf{7,91}$

2) A partir de una muestra de tamaño N de una variable numérica "x" se calcularon las medidas de posición para dicha muestra y se obtuvo que hay 50 observaciones x_i cuyos valores se encontraban ente el $P_{35} = 60$ y el $cuartil\ Q_3$=130

a) Entonces… ¿**Cuántas** observaciones tenía en total la muestra, o sea, ¿cuánto vale N ?

b) Supongamos que se dan cuenta de que a todos los valores de x_i hay que restarles 20 unidades, porque hubo un error en la toma de datos. ¿Cuántas observaciones tendrán ahora valores entre los nuevos P_{35} y el Q_3 calculados? ¿Cuánto valdrán ahora P_{35} y Q_3?

Resp:

a) Como hay un 35% de valores x_i menores que $P_{35} = 60$, y un 75% menores que el cuartil $Q_3 = 130$, entonces entre ambos hay $75\% - 35\% = 40\%$ de observaciones. Por lo tanto, si el 40% de N es 50, entonces N es igual a $\frac{100 \cdot 50}{40} = 125$ observaciones.

b) Dado que todos los valores disminuirán por igual, entonces entre las nuevas medidas de posición tendría que haber la misma cantidad de observaciones que antes, o sea 50. Los nuevos valores calculados serán $P_{35} = 40 \quad Q_3$=110).

3) Se midió la altura a los alumnos de un curso **A**, obteniéndose los siguientes valores:

23	34	21	21	30	16	16	29	27

En cambio, para otros alumnos de otro curso **B**, se obtuvo:

10	17	19	13	15	12	20	14

Calcular la **mediana (Me)** para cada uno de los dos turnos.

Resp: Curso A (N=9 es impar) $\boldsymbol{Me} = \mathbf{23}$ **Curso B** (N=8 es par) $\boldsymbol{Me} = \mathbf{14,5}$

4) Para analizar si hay una buena correlación lineal entre la edad de una planta y su altura se obtuvieron los siguientes datos en 5 momentos diferentes

X=Meses	3	5	7	9	11
Y=Altura en cm	25	80	130	150	170

$\boldsymbol{\sigma_x} = \mathbf{2,83} \qquad \boldsymbol{\sigma_y} = \mathbf{52,38}$

Calcule el coeficiente de correlación (ρ). ¿Le parece bueno el ajuste? ¿Por qué?

Resp:
a) $\Sigma x = 35$ $\quad \Sigma y = 555$ $\quad \Sigma xy = 4605$ $\quad cov_{xy} = 144$ $\quad \rho = \mathbf{0,97}$, es bueno, pues da cercano a 1

5) Calcular la moda (Mo), mediana (Me) y media ($\bar{x}$) de la variable x.

x_i	f_i
11	2
12	1
14	3
15	5
17	2

Resp: N=13 Moda= 15 Mediana=15 Media= $\frac{\mathbf{185}}{\mathbf{13}} = \mathbf{14,23}$

6) Completar correctamente las frases con solo una de las palabras optativas propuestas:

Palabras propuestas: **MEDIANA, VARIANZA, MUESTRA, CORRELACIÓN, MODA, MEDIA ($\bar{x}$).**

a) Para medir que tan dispersos están los valores de una variable cuantitativa hay que calcular la **……**

b) Cuando entre dos variables cuantitativas existe una buena relación entre ellas decimos que es buena la **……**

c) La medida de tendencia central utilizada para un conjunto de datos cualitativos es la ……

Resp: a) Varianza b) Correlación c) Moda

7) En una escuela se seleccionó al azar a 600 niños de primaria para preguntarles cual era el helado de su preferencia, dentro de tres gustos. complete la tabla de frecuencias a partir de la siguiente información:

a) El **30%** del total de los estudiante, sin distinción de sexo, prefiere limón.
b) Si nos restringimos a los que les gusta el chocolate, **un cuarto** de ellos son niñas.
c) El **40%** del total de los encuestados son niños que prefieren frutilla.

	Limón	Chocolate	Frutilla	**Total**
Niños				
Niñas				**165**
Total		**100**		**600**

Resp:

	Limón	Chocolate	Frutilla	**Total**
Niños	120	75	**c)** 240	435
Niñas	60	**b)** 25	80	**165**
Total	**a)** 180	**100**	320	**600**

8) La siguiente tabla exhibe la distribución de cargos docentes universitarios según renta y dedicación, por categoría en la Universidad de Buenos Aires en el año 2000.

CENSO DE DOCENTES 2000

Capítulo III - Cuadro 1: Cargos docentes universitarios según renta y dedicación, por categoría

Categoría	Rentado					Ad Honorem	Total
	Exclusiva	Semi Exclusiva	Parcial	Otra	Total		
Titular plenario	31	5	5	.	41	4	45
Titular	354	393	674	47	1468	160	1628
Asociado	228	126	290	8	652	267	919
Adjunto	597	637	2910	31	4175	505	4680
Consulto	35	30	72	5	142	32	174
Contratado, Invitado	5	20	108	97	230	30	260
Emérito	11	13	17	1	42	18	60
Honorario	8	2	7	2	19	380	399
Autorizado	7	15	57	9	88	268	356
Libre	.	2	20	7	29	100	129
Jefe Trabajos Prácticos	890	957	2827	32	4706	1142	5848
Ayudante 1°	404	689	4882	34	6009	2497	8506
Ayudante 2°	35	52	2404	53	2544	2659	5203
Otra	6	29	79	96	210	418	628
Total	2611	2970	14352	422	20355	8480	28835

a) ¿Qué porcentaje de los docentes tienen dedicación exclusiva?
b) ¿Qué porcentaje de los jefes de trabajos prácticos tienen dedicación parcial?
c) Si nos restringimos a los rentados ¿Qué porcentaje tienen dedicación parcial?
d) De los titulares plenarios ¿Cuántos no son rentados?
e) ¿Qué porcentaje de los docentes son Ayudantes, ya sea de 1ra o 2da?
f) De los ayudantes ¿Qué porcentaje no son rentados?

Resp:
a) 9,05 %

b) 48,34 %

c) $\frac{14352}{20355}100 = 70{,}51\%$
d) 4

e) $\frac{8506+5203}{28835}100 = 47{,}54\ \%$

f) $\frac{2497+2659}{8506+5203}100 = 37{,}61\%$

9) A lo largo de una semana se midió, durante cada día, la cantidad de lluvia (en mm) en 4 ciudades distintas (A, B, C, D) Los valores de las medias y desvíos obtenidos de los registros de esos siete días fueron los siguientes:

Ciudad:	$\bar{x}$ (mm)	σ (mm)
A	8,4	0,2
B	9,6	0,8
C	10	1
D	7	1,5

a) Supongamos que un día, en una ciudad se registró una lluvia de 8,9 mm ¿A cuál ciudad es más probable que pertenezca ese registro?
b) A la semana siguiente, en la ciudad "D" llovió durante cada día exactamente 3 mm más que en cada uno de los siete días correspondientes de la semana anterior. ¿Cuánto valdrían ahora entonces la media y el desvío?

Resp:

a) Si se calculan los intervalos $(\bar{x} - \sigma; \bar{x} + \sigma)$ el único que contiene al 8,9 es el de la ciudad **B.**

b) La media se incrementará en 3 mm y valdrá **10 mm**, el desvío se mantendrá igual: **1,5** mm.

10) ¿Cuál de las tres afirmaciones es correcta en cada caso?

I) Un estadístico es:
a) Un valor que se calcula siempre a partir de los datos de la población para estimar algo en la muestra.
b) Un valor de la muestra que sirve para estimar un parámetro de la población.
c) Un parámetro desconocido de la población.

II) De la media o promedio ($\bar{x}$) podemos decir que:
a) Es recomendable para utilizar cuando el gráfico de las frecuencias absolutas es muy asimétrico.
b) Es un valor que me asegura que el 50% de mis datos serán menores que él.
c) Se ve seriamente afectada por valores extremadamente grandes o chicos.

III) De la mediana (Me) podemos decir que:
a) Es un valor que me asegura que el 50% de mis datos serán menores que él.
b) Si hay valores extremos, es menos representativa del valor central de mis datos que la media $\bar{x}$
c) Es recomendable para utilizar en variables cualitativas.

IV) Dado un intervalo de dispersión del tipo $(\bar{x}-\sigma\ ;\ \bar{x}+\sigma)$

a) Aproximadamente el 68% de los valores de mi muestra o población se hallan dentro de los extremos de ese intervalo.
b) Aproximadamente el 68% de los valores de mi muestra o población se hallan fuera de los extremos de ese intervalo.
c) Cuánto más grande es su longitud, más representativa será la media del valor central de mis datos.

Resp:

Acá van las respuestas correctas (en negrita) y justificaciones para cada caso:

I)

a) No, no se calcula a partir de los datos de la población, pues justamente se utiliza cuando carecemos de todos los datos de ella.
b) En efecto, la idea es que el estadístico es un valor que se obtiene a partir de una muestra y que sirve para tener una idea, una estimación lo más precisa posible, de un determinado parámetro de la población.
c) Si bien me sirve para estimar el parámetro que no conocemos, no es “el” parámetro.

II)

a) No, al contrario, conviene utilizarlo cuando es lo más simétrico posible.
b) No, es la mediana la que cumple esta propiedad. La media no.
c) Tal cual, su valor tiende a agrandarse o achicarse mucho si en la muestra hay un valor mucho más grande o mucho más chico que el resto.

III)

a) En efecto, surge de la definición misma de mediana esa propiedad.
b) No, al contrario, es la media $\bar{x}$ la que suele ser menos representativa en el caso de haber valores extremadamente grandes o chicos.
c) No, es imposible utilizarla en variables cualitativas, solo está definida para las cuantitativas.

IV)

a) En efecto, si bien es una cifra aproximada, es lo que normalmente ocurre.
b) No, afuera quedarán aproximadamente el 32% de los valores de la muestra.
c) Al revés, cuánto más grande, menos representativa será la media del valor central de los datos.

11) La variable “x” representa las notas de 0 a 100 de 120 alumnos de un de curso universitario. Para dicho curso se obtuvo el intervalo $(\bar{x}-\sigma;\bar{x}+\sigma)=(56;78)$

Elija de entre estas tres afirmaciones la que le parezca más probable:

a) “Más de 100 alumnos obtuvieron una nota entre 56 y 78”.
b) “Menos de 40 alumnos obtuvieron una nota que no estuvo entre 56 y 78”.
c) “Un tercio de los alumnos obtuvieron una nota de entre 56 y 78”.

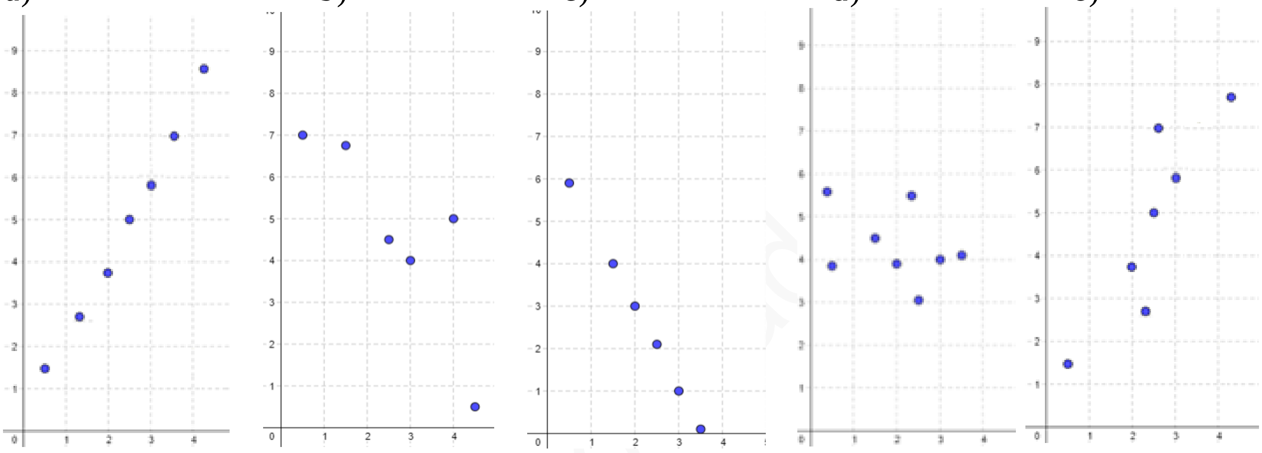
Resp:

La **b)**, pues dado que el intervalo $(\bar{x}-\sigma;\bar{x}+\sigma)$ suele contener al 68% de las observaciones de la variable x, quiere decir que uno esperaría que aproximadamente el 68% de 120, o sea, unos 82 alumnos hayan sacado esas notas. Por lo tanto, la opción a) parece ser demasiado exagerada en cuanto a la cantidad de alumnos, la c) está lejos de ese 68%, pues dice que solo el 33% de los alumnos obtuvieron esas notas, en cambio en la **b),** al afirmar que 40 alumnos (la tercera parte) no obtuvo esas notas, quiere decir que 80 si la obtuvieron (66%), que es el porcentaje más cercano a 68% de los tres casos.

12) Para cada uno de los cinco gráficos de nubes de puntos se calcularon los coeficientes de correlación lineal (ρ), y aunque no sabemos qué valor se corresponde con cada gráfico, sí sabemos que valores se obtuvieron. ¿Qué gráfico asignaría a cada uno de los coeficientes?

Correlación:	$\mathbf{0,01}$	$\mathbf{-0,97}$	$\mathbf{-0,60}$	$\mathbf{0,62}$	**0,94**
Gráfico letra:					

a) b) c) d) e)



Resp:

Correlación:	$\mathbf{0,01}$	$\mathbf{-0,97}$	$\mathbf{-0,60}$	$\mathbf{0,72}$	**0,94**
Gráfico letra:	**d)**	**c)**	**b)**	**e)**	**a)**

13) La siguiente tabla resume la estructura familiar, en su hogar, de alumnos de una Universidad:

Edad:	Viven solos	Viven con sus padres y nadie más	Viven con sus padres y otros familiares	Viven con conocidos no familiares	**TOTAL**
18 a 21	100	400	300	300	**1100**
22 a 25	100	300	200	100	**700**
26 a 29	150	50	50	100	**350**
más de 30	200	35	25	90	**350**
TOTAL	**550**	**785**	**575**	**590**	**2500**

a) ¿Qué porcentaje de los alumnos de la Universidad tiene de 22 a 25 años?

b) ¿Qué porcentaje del total de alumnos viven "con sus padres y otros familiares" y además tienen menos de 30 años?

c) ¿Qué porcentaje de los que tienen menos de 26 años viven acompañados (o sea, **no** viven solos)?

d) ¿Qué porcentaje del total tiene menos 26 años y viven acompañados?

e) Si nos restringimos a los alumnos que contestaron que "viven solos" ¿Qué porcentaje de ellos tiene menos de 30 años?

Resp:

a) $\frac{700 \cdot 100}{2500} = \mathbf{28}\%$

b) $\frac{(300+200+50) \cdot 100}{2500} = \mathbf{22}\%$

c) Con menos de 26 hay $1100 + 700 = 1800$. De estos hay $400 + 300 + 300 + 300 + 200 + 100 = 1600$ que no viven solos, por lo tanto, la respuesta es $\frac{1600 \cdot 100}{1800} = \mathbf{88,89}\%$

d) Pregunta sobre el porcentaje del total: ya vimos en c) que 1600 no viven solos, la respuesta ahora sería $\frac{1600 \cdot 100}{2500} = \mathbf{64}\%$

e) Hay 550 que viven solos. De ellos hay $100 + 100 + 150 = 350$ que tienen menos de 30. Entonces la respuesta es $\frac{350 \cdot 100}{550} = \mathbf{63,36}\%$**.**

14. Con el objetivo de analizar la relación entre los gastos en publicidad de seis empresas de productos lácteos y las ventas realizadas durante un determinado período de tiempo disponemos de los siguientes datos:

Gastos en publicidad en millones de $	1	2	3	4	5	6
Ventas en millones de $	12	14	14	15	18	16

Sabiendo que $\boldsymbol{\sigma_x = 1,71}$ $\boldsymbol{\sigma_y = 1,86}$ calcule la covarianza y el coeficiente de correlación lineal.

Resp: $\boldsymbol{cov_{xy} = 2,75}$ $\boldsymbol{\rho = 0,86}$

Los siguientes dos ejercicios fueron desarrollados en la última clase, se los dejo para los que no vinieron o por si quieren reintentarlos

A) Se realizó una encuesta relacionada con preguntas sobre si padecían hipertensión a 1000 personas. Completar la tabla sabiendo que:

a. El 15% de la población padece de hipertensión

b. El 75% cree no tener hipertensión.

c. De los que tienen hipertensión, el 6% cree que no tiene esta enfermedad.

d. Completada la tabla, contestar: De los que **creen que no tienen** hipertensión, ¿qué porcentaje en realidad sí la tiene?

	Tiene	No tiene	Total
Cree que tiene			
Cree que NO tiene			
Total			1000

Resp

	Tiene	No tiene	
Cree que tiene	141	109	250
Cree que NO tiene	c) 9	741	b) 750
	a) 150	850	1000

La respuesta al punto d) es $\frac{9}{750}100 = \mathbf{1,2}\%$

a.

B) Durante 4 semanas un comerciante vendió un determinado artículo. En cada una de esas semanas ofreció el producto a un precio distinto para analizar en cuanto aumentaba o disminuía la cantidad de compradores. El resultado se resume en la tabla:

Algunos datos:

$\bar{\boldsymbol{x}} = \mathbf{16,75}$ $\quad \bar{\boldsymbol{y}} = \mathbf{7,5}$ $\quad \boldsymbol{\sigma_x} = \mathbf{4,7}$

Calcular la correlación lineal $\boldsymbol{\rho}$ ¿Le parece buena o mala?

Precio	Compradores	$x \cdot y$
10	15	150
15	7	105
20	5	100
22	3	66

Resp: $\Sigma xy = 421$ $\sigma_y{}^2 = 20{,}79$ $\sigma_y = 4{,}56$ covarianza$= -20{,}375$

$\boldsymbol{\rho} = -\mathbf{0,95}$ Muy buena