

Apellido y Nombre: PANELLA, Agustina.

Primer Parcial - Estadística - 16/09/21 - IIB - INTECH

Guardá este archivo nombrándolo con tu apellido. Pegá las salidas de Infostat que justifiquen tus respuestas. Expresá los resultados con sus unidades y las conclusiones en términos del problema. Recordá grabar periódicamente tus avances. Al finalizar envíalo por mail a olguinsalinash@gmail.com. Buena suerte!



Argentina es uno de los tres principales productores de maní (*Arachis hypogae*) de alta calidad, siendo la Provincia de Córdoba donde se concentra su mayor producción. Las hormonas vegetales tienen una marcada influencia sobre el comportamiento del cultivo (desde la germinación hasta la producción de frutos), por lo que un adecuado equilibrio hormonal incidirá en un óptimo rendimiento y calidad del fruto. Con el objetivo de evaluar la efectividad de un nuevo producto (*Stimulate*®) que contiene tres hormonas del crecimiento (conkinetina, ácido giberélico y ácido3-indol-butírico) se diseñó el siguiente estudio:

A) 100 parcelas de 1m², todas ellas con características similares de suelo, fueron sembradas, cada una, con 20 semillas de maní. 50 de las parcelas, elegidas en forma aleatoria, fueron suplementadas con *Stimulate*® (**con sti**) mientras que a las restantes fueron dejadas como control (**sin sti**). A los 40 días se midió el rendimiento de la parcela, en términos de peso de los frutos producidos (gramos/m²). Los resultados se encuentran en el archivo [mani.IBD2](#).

1. Defina la unidad experimental e identifique y clasifique las variables de interés.
UNIDAD EXPERIMENTAL: parcela de 1m² de la Provincia de Córdoba.
VARIABLES DE INTERÉS: rendimiento de cada parcela en términos de peso de los frutos producidos (**Variable cualitativa continua**).
2. Identifique la(s) muestra(s) y su tamaño. ¿En esta investigación se tiene una o más poblaciones estadísticas?
MUESTRAS: 50 parcelas suplementadas con Stimulate y 50 parcelas control.
En esta investigación hay una sola población, la cual esta compuesta por las 100 parcelas de 1m².
3. Realice un análisis exploratorio de los datos y que incluya:
 - a) Tablas con medidas de tendencia central y de dispersión.

stimulate	Variable	Clase	LI	LS	MC	FA	FR	FAA	FRA
con sti	Peso de frutos	1	[0,38 0,43]		0,40	11	0,22	11	0,22
con sti	Peso de frutos	2	(0,43 0,48]		0,45	6	0,12	17	0,34
con sti	Peso de frutos	3	(0,48 0,53]		0,50	20	0,40	37	0,74
con sti	Peso de frutos	4	(0,53 0,58]		0,55	10	0,20	47	0,94
con sti	Peso de frutos	5	(0,58 0,63]		0,60	3	0,06	50	1,00

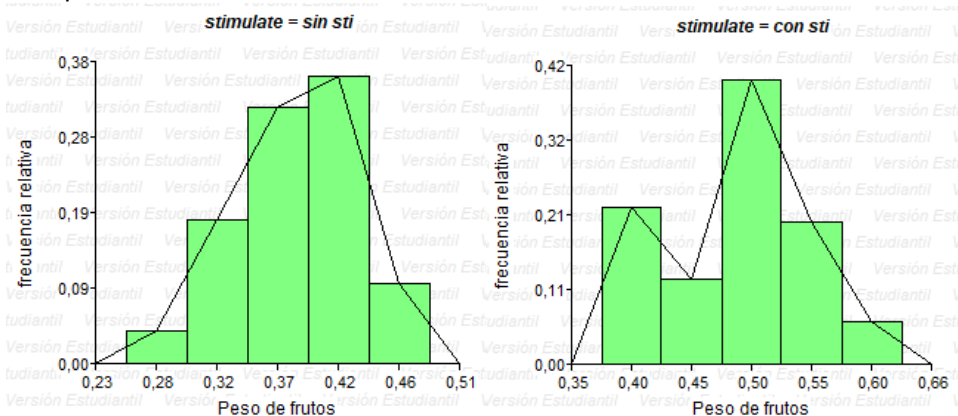
stimulate	Variable	Clase	LI	LS	MC	FA	FR	FAA	FRA
sin sti	Peso de frutos	1	[0,25 0,30]		0,28	2	0,04	2	0,04
sin sti	Peso de frutos	2	(0,30 0,35]		0,32	9	0,18	11	0,22
sin sti	Peso de frutos	3	(0,35 0,39]		0,37	16	0,32	27	0,54
sin sti	Peso de frutos	4	(0,39 0,44]		0,42	18	0,36	45	0,90
sin sti	Peso de frutos	5	(0,44 0,49]		0,46	5	0,10	50	1,00

Figura 1: Tabla de frecuencias.

stimulate	Variable	n	Media	D.E.	Var(n)	CV	Min	Máx	Mediana	Q1	Q3
con sti	Peso de frutos	50	0,49	0,06	3,8E-03	12,67	0,38	0,63	0,50	0,44	0,53
sin sti	Peso de frutos	50	0,38	0,05	2,1E-03	12,07	0,25	0,49	0,39	0,35	0,41

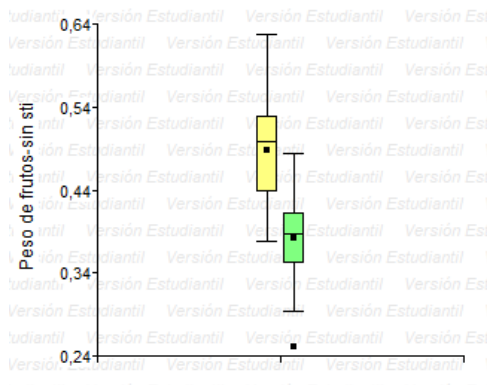
Figura 2: Medidas de resumen.

- b) Gráficos donde pueda describir la forma de la distribución de la variable de interés. Interpretélos.



A partir de estos gráficos podemos decir que aquellas parcelas que no contaron con la suplementación del *Stimulate*, presentan una distribución simétrica en la cual es más común que los frutos pesen entre 0.39 y 0.44 gramos. Por otro lado, en el gráfico de aquellas parcelas que fueron suplementadas con *Stimulate* se presenta una distribución más asimétrica. Podemos decir que en ese gráfico se observa un comportamiento bimodal debido a que posee dos picos: uno entre 0.37 y 0.43 gramos y otro entre 0.47 y 0.53 gramos.

- c) Gráficos que le permitan detectar la presencia de datos atípicos. Interpretélos



En este gráfico, la caja de color amarillo representa los datos obtenidos de las parcelas que recibieron suplementación con *Stimulate*, mientras que la de color verde son los valores correspondientes a las parcelas control.

Observando el box plot, podemos decir que aquellas parcelas que no recibieron estímulo presentan una distribución más simétrica debido a que la media se encuentra muy cercana a la mediana. Por otro lado, en la caja amarilla se observa que la media se encuentra un poco más alejada de la mediana y, por lo tanto, presenta una distribución más asimétrica.

Además, se puede observar uno de los valores registrados de las parcelas control es un dato atípico.

En cuanto a la variabilidad (bigotes), podemos decir que las parcelas estimuladas presentan mayor variabilidad que las parcelas control.

4. Compare ambos tratamientos en cuanto a tendencia central y variabilidad. ¿En cuál de los dos grupos (con *Stimulate*® y sin el producto) es más variable el rendimiento? Fundamente. Redacte una conclusión respecto al objetivo de la investigación. Podemos decir, al tener una media y mediana por encima de la media y mediana de las parcelas control, las parcelas estimuladas (amarillo) tuvieron un mayor rendimiento. En cuanto al objetivo de la investigación se puede concluir que, el uso de *Stimulate* es efectivo para el aumento del rendimiento en término del peso de los frutos.

B) Además, en este ensayo se registró en cada una de las parcelas si habían germinado todas las semillas sembradas o no, resultados que se encuentran en el archivo [mani.IBD2](#) (Germinación).

5. Arme una tabla de contingencia donde se indiquen las frecuencias absolutas del número de parcelas donde germinaron todas las semillas ([Si](#) en el archivo [mani.IBD2](#)) y de aquellas donde no germinó la totalidad de las semillas sembradas ([No](#) en el archivo [mani.IBD2](#)) para cada una de las dos condiciones ensayadas ([con sti](#) y [sin sti](#)).

<i>stimulate</i>	Variable	Clase	Categorías	FA	FR	FAA	FRA
con sti	Germinación	1	no	7	0,14	7	0,14
con sti	Germinación	2	si	43	0,86	50	1,00

<i>stimulate</i>	Variable	Clase	Categorías	FA	FR	FAA	FRA
sin sti	Germinación	1	no	20	0,40	20	0,40
sin sti	Germinación	2	si	30	0,60	50	1,00

	Si	No	TOTAL
Con sti	43	7	50
Sin sti	30	20	50
TOTAL	73	27	100

6. Si se elige al azar una de las parcelas y considerando que estas representan adecuadamente a la población, calcule la probabilidad de que:
 - a) en la parcela hayan germinado todas las semillas
 $P(\text{todas las semillas germinadas}) = \frac{73}{100} = 0.73$
 La probabilidad de que todas las semillas de la parcela hayan germinado es de **0.73** (Alta probabilidad de ocurrencia).

- b) si es una parcela que ha sido tratada con *Stimulate*, ¿cuál es la probabilidad de que no hayan germinado todas las semillas de la parcela?

$$P(\text{No germinadas/Con sti}) = \frac{7}{50} = 0.14$$

La probabilidad de que no hayan germinado todas las semillas de la parcela tratada con *Stimulate* es de **0.14** (Baja probabilidad de ocurrencia).

- c) ¿Es independiente el que hayan germinado o no todas las semillas en la parcela de la presencia de o no del producto hormonal? Justifique con alguna probabilidad. Estos sucesos no son independientes ya que NO se cumple la siguiente condición:

$$P(A/B) = P(A) \text{ o } P(B/A) = P(B)$$

En este caso:

$$P(\text{Si/Sin sti}) = \frac{30}{50} = 0.6$$

$$P(\text{No/Sin sti}) = \frac{20}{50} = 0.4$$

Probabilidades de ocurrencia distintas \Rightarrow Sucesos No independientes

C) Se sabe que el rendimiento de las parcelas en plantaciones de maní de la región de Río Cuarto se distribuye en normalmente con media de 4400 Kg/há y un desvío estándar de 650 kg/ha. Si se elige aleatoriamente una parcela de esta región:

ACLARACIÓN: EJERCICIOS REALIZADOS CON LA APLICACIÓN PROBABILITY DISTRIBUTIONS.

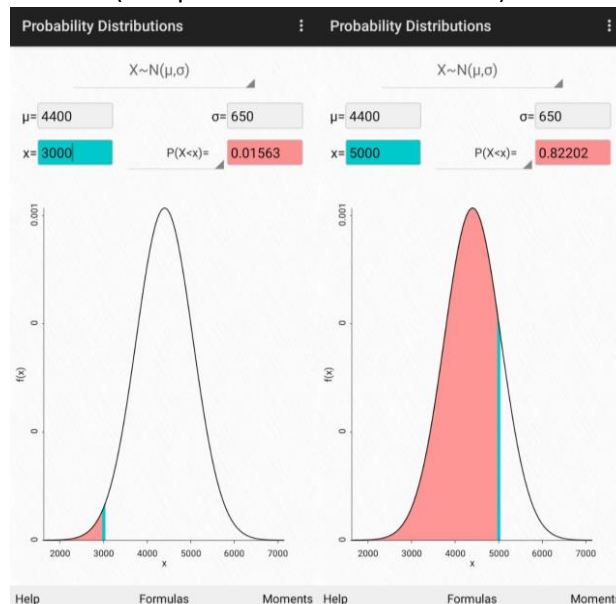
$\mu=4400$ kg/há

$\sigma=650$ kg/há

8. ¿Cuál será la probabilidad de que tenga un rendimiento entre 3000 y 5000 kg/ha?

$$P(3000 < x < 5000) = P(X < 5000) - P(X < 3000) = 0.82202 - 0.01563 = \mathbf{0.80639}$$

La probabilidad de que la parcela elegida al azar tenga un rendimiento entre 3000 y 5000 kg/há es de **0.80639** (Alta probabilidad de ocurrencia).

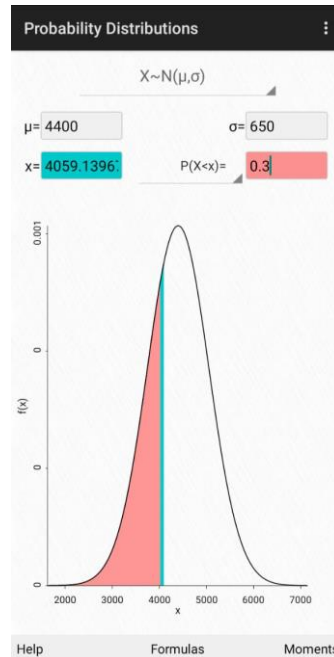


9. Del 70% de las parcelas más productivas ¿cuál será el rendimiento mínimo? ¿Cómo se denomina dicho estadístico?

Sabiendo que $P(X < x) = 0.3$, podemos determinar x , el cual es el valor del rendimiento mínimo que tiene el 70% de las parcelas más productivas.

$$x = 4059.13967.$$

Entonces, el rendimiento mínimo del 70% de las parcelas más productivas es de **4059.13967 kg/há**. Este estadístico se denomina percentil 30, el cual deja por debajo de él al 30% de la población y al 70% por encima.



10. Si se seleccionan aleatoriamente 5 de estas parcelas ¿cuál será la probabilidad de que **en promedio** tengan un rendimiento mayor a 5000 kg/ha?

La media del promedio muestral es igual a la media de la población (Teorema central del límite).

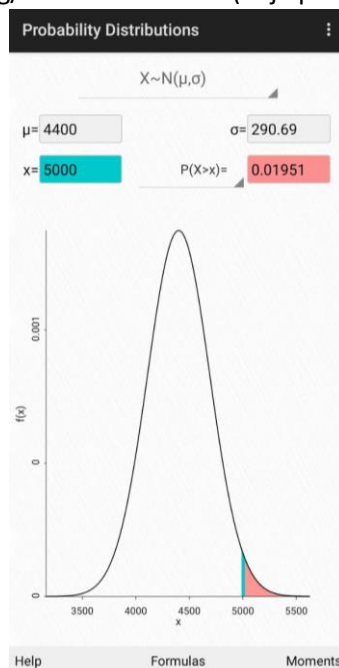
$$\mu = 4400 \text{ kg/há}$$

$$n = 5$$

$$E.E = \frac{\sigma}{\sqrt{n}} = \frac{650}{\sqrt{5}} = 290.6888371 \text{ kg/há}$$

$$P(X > 5000) = \mathbf{0.01951}$$

La probabilidad de que las 5 parcelas elegidas aleatoriamente tengan en promedio un rendimiento mayor a 5000 kg/há es de **0.01951** (Baja probabilidad de ocurrencia).



11. De 10 parcelas elegidas al azar ¿cuál será la probabilidad de que 5 de ellas tengan un rendimiento de entre 3000 y 5000 kg/ha? Indique los parámetros de la distribución de probabilidades empleada.

$$X \sim \text{Bin}(n, p)$$

$$n=10$$

$$x=5$$

$$p = \text{probabilidad de éxito} = P(3000 < X < 5000) = 0.80639 \text{ (Calculado en el ejercicio 8).}$$

$$X \sim \text{Bin}(10, 0.80639)$$

$$P(X=x) = \mathbf{0.02338}$$

La probabilidad de que 5 de las 10 parcelas elegidas al azar tengan un rendimiento de entre 3000 y 5000 kg/há es de **0.02338** (Baja probabilidad de ocurrencia).

