

Anexo 4: Productos de la consultoría

Anexo 4.1: Instructivo para el curso: Estadística básica e inferencial con Microsoft Excel (v2013)

Fomento a la construcción de capacidades en relación a las políticas públicas apoyadas por Contratos de Reforma Sectorial (CRS) en Bolivia (DCI-ALA/2014/025-058)

Fortalecimiento de la capacidad institucional en los sectores de desarrollo integral con coca, tráfico ilícito de drogas y seguridad alimentaria para una eficiente gestión del apoyo presupuestario sectorial en Bolivia (DITISA)

Instructivo para el curso: Estadística básica e inferencial con Microsoft Excel (v2013)

Informe final de misión de Corta Duración

Contrato n° DCI/LA/2017/392-699
N° de identificación: EuropeAid/138320/IH/SER/BO



Proyecto Financiado por la
Unión Europea



Implementada por el consorcio:

AGRER — AECOM — TRANSTEC

La Paz / Bruselas, enero 2019

Disclaimer:

Este informe ha sido elaborado por el consorcio AGRER/AECOM/TRANSTEC con financiamiento de la Unión Europea. Las opiniones aquí expresadas son del consultor y no expresan necesariamente las de la Comisión Europea.



Fortalecimiento de la capacidad institucional en los sectores de desarrollo integral con coca, tráfico ilícito de drogas y seguridad alimentaria para una eficiente gestión del apoyo presupuestario sectorial – Contrato n° DCI/LA/2017/392-699



AGRADECIMIENTOS

Ministerio de Desarrollo Rural y Tierras

[Dirección General de Planificación y Observatorio Agropecuario de Productivo]

Programa Financiado por Unión Europea.

Fortalecimiento de la capacidad institucional en los sectores de desarrollo integral con coca, tráfico ilícito de drogas y seguridad alimentaria para una eficiente gestión del apoyo presupuestario sectorial.

DITISA-UE. LA/2017/392-699



CONTENIDO

Lista de las tablas	6
Lista de las figuras	6
1. Conceptos.....	8
1.1. Variables	8
1.2. Población	8
1.3. Muestra	8
1.4. Hipótesis.....	9
2. Organización base de datos	10
2.1. Variables en una matriz de doble entrada	10
3. Estadística descriptiva	12
3.1. Variables cuantitativas	12
3.2. Variables cualitativas.....	18
4. Estadística inferencial	22
4.1. Intervalo de confianza para la media poblacional " μ " siendo la varianza poblacional desconocido.	22
4.2. Intervalo de confianza para una proporción poblacional P.	24
4.3. Comparación de promedios en base la distribución de t de Student	25
4.4. Comparación de un promedio con un indicador.....	29
4.5. Prueba de diferencia de proporciones.....	30
4.6. Análisis de varianza de clasificación de una vía	31
4.7. Relación entre variables	35
5. Muestreo.....	44
5.1. Determinación del tamaño de la muestra	44
5.2. Muestreo aleatorio simple	45
5.3. Muestreo aleatorio sistemático	46
5.4. Algunas consideraciones para el diseño de un muestreo	47
Bibliografía	48



LISTA DE LAS TABLAS

Tabla 1. Prueba de t de Student contrastando promedios de rendimiento con el indicador rendimiento de papa 5.98 t/ha (30% adicional al rendimiento 4.6 t/ha del 2014)	29
Tabla 2. Datos de rendimiento de papa (t/ha) variedad Waycha de la campaña 2010-2011. (SEPA-PASA-MDRyT, 2011),	33
Tabla 3. Estadística descriptiva del rendimiento de papa (t/ha) variedad Waycha de la campaña 2010-2011 de Cochabamba, Provincia Arque, Municipio Tacopaya y nueve Comunidades. (SEPA-PASA-MDRyT, 2011),	33

LISTA DE LAS FIGURAS

Figura 1: Bolivia: Precio mensual al por mayor de la papa huaycha, Enero 2008 - Noviembre 2018 (Bolivianos por 1 arroba).....	10
Figura 2. Generación de una tabla en Excel	11
Figura 3. Ventana Complementos en Excel.....	13
Figura 4. Estadística descriptiva dentro Análisis de datos en Microsoft Excel.....	14
Figura 5. Demarcación de datos, opciones y salidas de Estadística descriptiva	15
Figura 6. Demarcación de datos, para obtener la Tabla dinámica	15
Figura 7. Tabla dinámica con cinco veces el precio de papa en La Paz	16
Figura 8. Estadística descriptiva del precio de papa en departamento de La Paz	16
Figura 9. Tabla dinámica del precio de papa en La Paz e incluida Año en la Fila.....	17
Figura 10. Presentación clásica de la tabla dinámica con estadística descriptiva del precio de papa de LA PAZ (Bs/arroba) por Mes, datos de 2008-01 a 2018-11 (OAP-MDRyT, 2018).....	17
Figura 11. Histograma del precio de papa de LA PAZ (Bs/arroba), datos de 2008-01 a 2018-11 (OAP-MDRyT, 2018).....	18
Figura 12. Tabla inicial para la distribución de frecuencias de una vía con la variable Variedad de papa.....	19
Figura 13. Distribución de frecuencias de una vía con conteo y expresado en porcentaje la variable Variedad de papa cultivada.....	20
Figura 14. Distribución de frecuencias de dos vías con conteo de la Variedad y P14_COMO_17 (14. ¿Cómo califica la cosecha del año agrícola 2016-2017?).	20
Figura 15. Distribución de frecuencias de dos vías con conteo y % de Variedad y P14_COMO_17 (14. ¿Cómo califica la cosecha del año agrícola 2016-2017?).	21
Figura 16. Resultados de Estadística descriptiva y límites de confianza.....	23
Figura 17. Límites de confianza para variable cualitativa con fórmulas y resultados.....	25
Figura 18. Prueba de F para determinar homogeneidad de varianzas.	28
Figura 19. Prueba de t para dos muestras suponiendo varianzas iguales.	29
Figura 20. Diferencia de dos proporciones en base a fórmulas.....	31
Figura 21. Ventana Análisis de datos, remarcada el Análisis de varianza de un factor y resultados... ..	35
Figura 22. Ventana de Coeficiente de correlación activada con las opciones.	36
Figura 23. Resultados de coeficientes de correlación simple (Pearson) de variables de superficie, semilla, plaga-1, enfermedad-1 y el rendimiento de papa.	37



Figura 24. Resultados de coeficientes de correlación simple (Pearson) y significancia de variables de superficie, semilla, plaga-1, enfermedad-1 y el rendimiento de papa (ns: no significativo a P: 0.05; *: Significativo a P: 0.05 y; ** significativo a P: 0.01).....	37
Figura 25. Resultados del análisis de regresión simple del rendimiento y la superficie sembrada de papa de la campaña 2017-2018	39
Figura 26. Ventana de análisis de regresión activadas las opciones de rótulo y Nivel de confianza....	40
Figura 27. Resultados del análisis de regresión del rendimiento y las variables superficie_17_18, semilla_toneladas, P20_PLAGA_1 y P20_ENFERMEDAD_1	41
Figura 28. Distribución de frecuencias observadas y esperadas de la variable variedad vs aplicación agroquímico.....	42
Figura 29. Ventana Argumento de función PRUEBA.CHICUAD con las tablas de distribución de frecuencias observadas y esperadas.	43
Figura 30. Distribución de frecuencias y en porcentaje por variedad.	43
Figura 31. Resultados del tamaño de muestra y estadística descriptiva del rendimiento de papa campaña 2017-2018 (OAP-MDRyT, 2018).	45
Figura 32. Vista de la opción Muestra para una selección aleatoria.	46
Figura 33. Vista de los elementos seleccionados en forma aleatoria y sistemática.	46



1. CONCEPTOS

1.1. Variables

Se refiere a características que no son constantes, sino que varían de un individuo a otro y que sirven para distinguir o describir

○ **Cuantitativas**

Es aquella para la cual las observaciones resultantes pueden medirse: por ejemplo, estaturas y pesos

- **Continuas:** Es aquella que puede presentar cualquier valor de cierto intervalo. La estatura y peso son ejemplos
- **Discretas:** Es aquella para la cual los valores posibles no se pueden observar en una escala continua debido a la existencia de espacios entre estos posibles valores. Generalmente son conteos con números enteros, por ejemplo el número de hijos de familias.

○ **Cualitativas**

Hay muchos casos en que no es posible hacer medidas numéricas, muchas de ellas son susceptibles solamente de clasificación

- **Ordinal:** Son categorías independientes y exclusivas, es decir, cada observación cae en una, y solamente una, categoría; pero hay un orden intrínseco entre las categorías, por ejemplo de menor a mayor.
- **Nominal:** Son categorías exclusivas e independientes, por ejemplo especies o sexos.

1.2. Población

Es el conjunto más grande de valores (de una variable), por el cual existe algún interés.

- **Parámetro** Son aquellas medidas como la media aritmética (μ), desviación estándar (σ), varianza (σ^2), número de elementos (N), pero cuando caracterizan a una población.

1.3. Muestra

Es una parte de la población. Por lo general, se trata de usar la información de la muestra para hacer inferencia a cerca de una población.



- **Estadísticos o estadígrafos:** Son aquellas medidas como la media aritmética (\bar{x}), desviación estándar (S), varianza (S^2), número de elementos (n), pero cuando caracterizan a una muestra.

1.4. Hipótesis

Indican lo que estamos buscando o tratando de probar y pueden definirse como explicaciones tentativas del fenómeno investigado formuladas a manera de proposiciones. Así misma, las hipótesis deben estar vinculadas con técnicas disponibles para probarlas.

Hipótesis estadística: Son afirmaciones sobre los parámetros (media, desviación estándar, etc.) de dos o más poblaciones, de tal forma que se pueden comprobar por medio de métodos estadísticos

- **Hipótesis nula:** Es una afirmación en la que se menciona que no hay ninguna diferencia entre dos parámetros poblacionales o entre el valor verdadero de algún parámetro y su valor hipotético

$$\mu_1 = \mu_2 = \mu_3 = \mu_4 = \mu \dots$$

- **Hipótesis alternativa:** Es la que menciona que los datos particulares de la muestra sí dan suficiente evidencia como para hacernos concluir que la hipótesis nula es falsa y por lo cual es rechazada.

$$\mu_1 \neq \mu_2 \neq \mu_3 \neq \mu_4 \neq \mu \dots$$



2. ORGANIZACIÓN BASE DE DATOS

2.1. Variables en una matriz de doble entrada

La organización de los datos levantados para el análisis con un programa estadístico suele realizarse en una matriz de dos entradas, una es filas y la segunda columnas. En las filas generalmente están las observaciones y en las columnas las variables (Figura 1). Estas variables son cualitativas y cuantitativas.

	A	B	C	D	E	F	G	H	I	J	K	L	
2	Bolivia: Precio mensual al por mayor de la papa huaycha, Enero 2008 - Noviembre 2018.												
3	(En Bolivianos por 1 arroba)												
4													
5	Fecha	Año	Mes	EI ALTO	LA PAZ	COCHA BAMBAMBA	SANTA CRUZ	ORURO	SUCRE	POTOSI	TARIJA	COBIJA	TRINIDAD
6	2008-01	2008	1		25.5	25.37							
7	2008-02	2008	2		25.27	24.68							
8	2008-03	2008	3		24.95	25.07							
9	2008-04	2008	4		26.41	26.22							
10	2008-05	2008	5		32.92	33.63	32.56						
11	2008-06	2008	6		31.82	37.47	36.27						
12	2008-07	2008	7		32.92	36.38	34.91	35.27					
13	2008-08	2008	8		31.78	35.02	28.53	35.2			48.89		
14	2008-09	2008	9		32.49	34.98	29.7	34.4			50		
15	2008-10	2008	10		37.72	39.03	33.77	39.01					
16	2008-11	2008	11		38.96	41.78	35.57	38.27			31.21		
17	2008-12	2008	12		40.47	40	34.81	40			39.38		
18	2009-01	2009	1		33.12	34.28	27.72	37.26		30.81	38.89		
19	2009-02	2009	2		29.48	27.22	26.87	31.11		28.94	38.89		
20	2009-03	2009	3		28.14	24.96	23.07	32.43		27.7	30		

Figura 1: Bolivia: Precio mensual al por mayor de la papa huaycha, Enero 2008 - Noviembre 2018 (Bolivianos por 1 arroba)

- Organización de una tabla en Microsoft Excel

En Microsoft Excel, una vez o antes de la introducción de datos, genere una tabla. Para ello remarcar el área donde están los datos o ubicar el cursor dentro del cuadro de datos. Luego del **Menú principal** seleccione **Insertar, Tabla y Aceptar** (Figura 2) y obtendrá la tabla, debe visualizarse los encabezados con filtro.

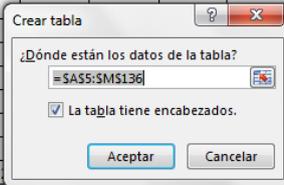


Fortalecimiento de la capacidad institucional en los sectores de desarrollo integral con coca, tráfico ilícito de drogas y seguridad alimentaria para una eficiente gestión del apoyo presupuestario sectorial – Contrato n° DCI/LA/2017/392-699



MENÚ PRINCIPAL

Fecha	Año	Mes	EL ALTO	LA PAZ	COCHA BAMBAMBA	SANTA CRUZ	ORURO	SUCRE	POTOSI	TARIJA	COBUJA	TRINIDAD
2008-01	2008	1		25.5	25.37							
2008-02	2008	2		25.27	24.68							
2008-03	2008	3		24.95	25.07							
2008-04	2008	4		26.41	26.22							
2008-05	2008	5		32.92	33.63	32.56						
2008-06	2008	6		31.82	37.47	36.27						
2008-07	2008	7		32.92	36.38	34.91	35.2				48.89	
2008-08	2008	8		31.78	35.02	28.53	35.2				50	
2008-09	2008	9		32.49	34.98	29.7	34.4					
2008-10	2008	10		37.72	39.03	33.77	39.01					
2008-11	2008	11		38.96	41.78	35.57	38.27				31.21	
2008-12	2008	12		40.47	40	34.81	40				39.38	
2009-01	2009	1		33.12	34.28	27.72	37.26		30.81		38.89	
2009-02	2009	2		29.48	27.22	26.87	31.11		28.94		38.89	
2009-03	2009	3		28.14	24.96	23.07	32.43		27.7		30	



Fecha	Año	Mes	EL ALTO	LA PAZ	COCHABAMBAMBA	SANTA CRUZ	ORURO	SUCRE	POTOSI	TARIJA	COBUJA	TRINIDAD
2008-01	2008	1		25.5	25.37							
2008-02	2008	2		25.27	24.68							
2008-03	2008	3		24.95	25.07							
2008-04	2008	4		26.41	26.22							
2008-05	2008	5		32.92	33.63	32.56						
2008-06	2008	6		31.82	37.47	36.27						
2008-07	2008	7		32.92	36.38	34.91	35.27					
2008-08	2008	8		31.78	35.02	28.53	35.2				48.89	
2008-09	2008	9		32.49	34.98	29.7	34.4				50	
2008-10	2008	10		37.72	39.03	33.77	39.01					
2008-11	2008	11		38.96	41.78	35.57	38.27				31.21	
2008-12	2008	12		40.47	40	34.81	40				39.38	
2009-01	2009	1		33.12	34.28	27.72	37.26		30.81		38.89	
2009-02	2009	2		29.48	27.22	26.87	31.11		28.94		38.89	
2009-03	2009	3		28.14	24.96	23.07	32.43		27.7		30	

Figura 2. Generación de una tabla en Excel

NOTA: Si una variable numérica y cuantitativa posee dentro las observaciones con texto u otro carácter diferente a números, por favor bórralas.

• Evaluación de los datos

Al realizar un clic izquierdo en el filtro (triangulo invertido) le generará una lista donde observará y valorará los datos numéricos o textos.



3. ESTADÍSTICA DESCRIPTIVA

La estadística descriptiva tiene el objetivo de resumir y presentar datos en forma fácilmente interpretable tanto para variables cuantitativas y cualitativas.

3.1. Variables cuantitativas

Promedio

Es la medida de tendencia central más utilizada y puede definirse como el promedio aritmético de una distribución. Es la suma de todos los valores dividida por el número de casos.

$$\bar{x} = \frac{\sum_{i=1}^n x_i}{n} \qquad \bar{x}_p = \frac{\sum_{i=1}^n n_i \bar{x}_i}{\sum_{i=1}^n n_i}$$

\bar{x} = Promedio

\bar{x}_p = Promedio ponderado

x_i = Un valor de una variable de la i-esima observación

\bar{x}_i = Promedio de la i-esima submuestra

n = Tamaño de la muestra

n = Tamaño de la muestra

n_i = Tamaño de la i-esima submuestra

Desviación estándar

Es el promedio de desviación de las puntuaciones con respecto a la media. Esta medida es expresada en las unidades originales de la distribución. Cuanto mayor es la dispersión de los datos alrededor de la media, mayor es la desviación estándar.

$$S = \sqrt{\frac{\sum_{i=1}^n (x_i - \bar{x})^2}{n - 1}} = \sqrt{\frac{\sum_{i=1}^n x_i^2 - \frac{\left(\sum_{i=1}^n x_i\right)^2}{n}}{n - 1}}$$

Varianza

Es la desviación estándar elevada al cuadrado y se simboliza como: S^2 . Es un concepto estadístico sumamente importante, ya que muchas de las pruebas cuantitativas se fundamentan en él.

Coefficiente de variación

Es la variación de las observaciones respecto al promedio, se determina:

$$CV = \left(\frac{S}{\bar{x}}\right) * 100$$

Rango

Es la diferencia entre la puntuación mayor y la puntuación menor, indica el número de unidades en la escala de medición necesaria para incluir los valores máximo y mínimo



$$Rango = x_{m\acute{a}ximo} - x_{m\acute{i}nimo}$$

En Microsoft Excel

Existen varias alternativas para el análisis estadístico: “Fórmulas”, “Análisis de datos” y “Tablas dinámicas”.

- **Opción Análisis de datos**

La opción de **Análisis de datos** en la mayoría de las computadoras dentro Microsoft Excel no está activado. Observe en el Menú principal “Datos” y ubique “Análisis de datos” si está presente pase al siguiente inciso.

Activación de Análisis de datos

Para la activar el “Análisis de datos” en la versión de Excel 2013, entre en el **Menú principal**, seleccionar **ARCHIVO**, luego a **Opciones de Excel, Complementos**, seleccione en **Administrar, Complementos de Excel** y hacer clic en **Ir**, activar **Herramienta para análisis** y **Aceptar**. Si el programa no fue instalado en forma completa le pedirá el CD o DVD instalador.

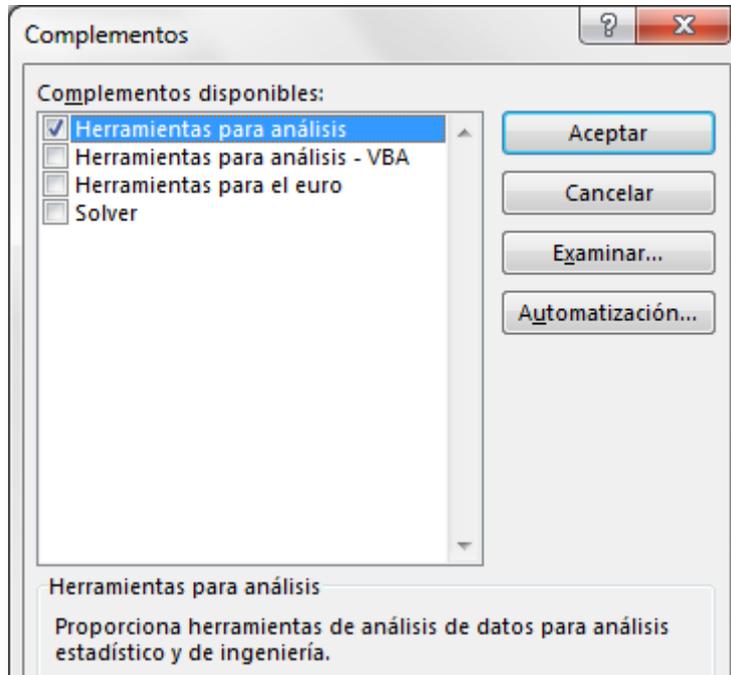


Figura 3. Ventana Complementos en Excel

- **Empleo de la opción Análisis de datos:**



Utilizar el archivo “precios para cbba.xlsx”, dicha información corresponde a **Precios de productos en Cochabamba** (OAP-MDRyT, 2018). Para obtener una estadística descriptiva. Ejecute los siguientes pasos; del **MENÚ PRINCIPAL**, luego a “**Datos**”, **Análisis de datos**, “**Estadística descriptiva**” y **Aceptar**:

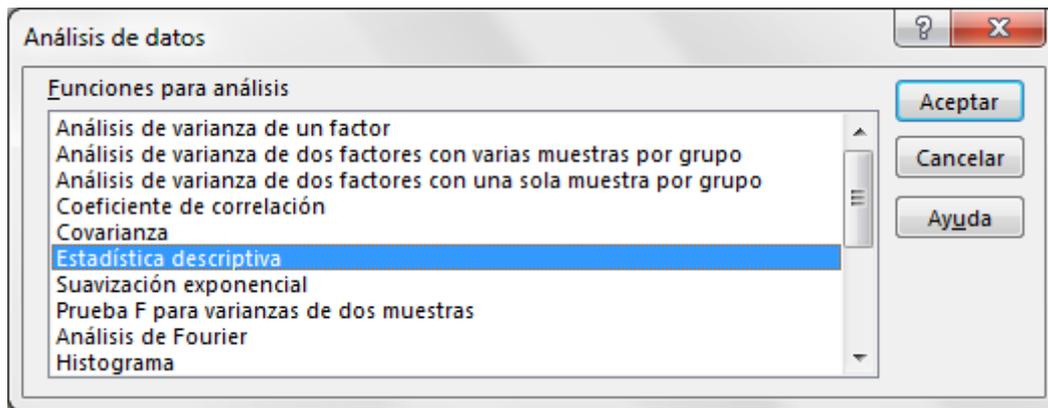


Figura 4. Estadística descriptiva dentro **Análisis de datos** en Microsoft Excel.

Luego demarcar los datos más el encabezado de **El Alto, La Paz, Cochabamba, Santa Cruz y Oruro** de la Tabla en **Rango de entrada**, activar en Agrupado por: **Columnas**, activar **Rótulo en la primera fila**, obsérvese la primera fila contiene el nombre de la variable. También active **Resumen de estadísticas**, **Nivel de confianza para la media** y **Aceptar**, la salida de los resultados por defecto saldrá en otra Hoja Excel (Figura 5).

	A	B	C	D	E	F	G	H	I	J	K	L	M
	Fecha	Año	Mes	EL ALTO	LA PAZ	COCHABAMBA	SANTA CRUZ	ORURO	SUCRE	POTOSI	TARIJA	COBIJA	TRINIDAD
5													
6	2008-01	2008	1		25.5	25.37							
7	2008-02	2008	2		25.27	24.68							
8	2008-03	2008	3		24.95	25.07							
9	2008-04	2008	4		26.41	26.22							
10	2008-05	2008	5		32.92	33.63							
11	2008-06	2008	6		31.82	37.47							
12	2008-07	2008	7		32.92	36.38							
13	2008-08	2008	8		31.78	35.02							
14	2008-09	2008	9		32.49	34.98							
15	2008-10	2008	10		37.72	39.03							
16	2008-11	2008	11		38.96	41.78							
17	2008-12	2008	12		40.47	40							
18	2009-01	2009	1		33.12	34.28							
19	2009-02	2009	2		29.48	27.22							
20	2009-03	2009	3		28.14	24.96							
21	2009-04	2009	4		23.96	21.82							
22	2009-05	2009	5		24.66	20.36							
23	2009-06	2009	6		23.86	19.79							
24	2009-07	2009	7		24.17	23.74	20.63	25.52		21.98			
25	2009-08	2009	8		25.91	26.78	19.57	23.84		21.43			



	A	B	C	D	E	F	G	H	I	J
1	EL ALTO		LA PAZ		COCHABAMBA		SANTA CRUZ		ORURO	
2										
3	Media	37.5790476	Media	34.4024806	Media	33.2950382	Media	32.7481746	Media	34.08
4	Error típico	1.39398141	Error típico	0.7790455	Error típico	0.84865276	Error típico	0.80251571	Error típico	0.786
5	Mediana	34.305	Mediana	32.92	Mediana	32.27	Mediana	33.08	Mediana	33
6	Moda	34.91	Moda	25	Moda	30.51	Moda	34	Moda	
7	Desviación estándar	9.03403208	Desviación	8.84825601	Desviación	9.71327485	Desviación	9.00821655	Desviación	8.79488
8	Varianza de la muestra	81.6137357	Varianza d	78.2916344	Varianza d	94.3477083	Varianza d	81.1479654	Varianza d	77.349
9	Curtosis	2.1433713	Curtosis	2.041771	Curtosis	1.48448002	Curtosis	1.17160545	Curtosis	2.44401
10	Coefficiente de asimetría	1.56641194	Coefficient	1.14609914	Coefficient	0.90353266	Coefficient	0.73544485	Coefficient	1.30731
11	Rango	38.9	Rango	46.02	Rango	52.22	Rango	44.82	Rango	46
12	Mínimo	24.52	Mínimo	19.36	Mínimo	16.69	Mínimo	17.1	Mínimo	18
13	Máximo	63.42	Máximo	65.38	Máximo	68.91	Máximo	61.92	Máximo	65
14	Suma	1578.32	Suma	4437.92	Suma	4361.65	Suma	4126.27	Suma	4260
15	Cuenta	42	Cuenta	129	Cuenta	131	Cuenta	126	Cuenta	
16	Nivel de confianza(95.0%)	2.81520258	Nivel de c	1.54147463	Nivel de c	1.678958	Nivel de c	1.5882782	Nivel de c	1.55697
17										

Figura 5. Demarcación de datos, opciones y salidas de Estadística descriptiva

- **Opción Tablas dinámicas:**

Utilizar el archivo “precios para cba.xlsx”, para obtener una estadística descriptiva ejecute los siguientes pasos; ubicar el cursor en la **Tabla** y del **Menú principal** entrar a “**Insertar**”, luego “**Tabla dinámica**”, obtendrá una ventana “**Crear tabla dinámica**” y nuevamente un clic en **Aceptar**. Por defecto esta remarcada la **Tabla** y generará en otra Hoja Excel las salida de **Tabla dinámica**.

	A	B	C	D	E	F	G	H	I	J	K	L	M
5	Fecha	Año	Mes	EL ALTO	LA PAZ	COCHABAMBA	SANTA CRUZ	ORURO	SUCRE	POTOSI	TARIJA	COBIJA	TRINIDAD
6	2008-01	2008	1		25.5	25.37							
7	2008-02	2008	2		25.27	24.68							
8	2008-03	2008	3		24.95	25.07							
9	2008-04	2008	4		26.41	26.22							
10	2008-05	2008	5		32.92	33.63	32						
11	2008-06	2008	6		31.82	37.47	36						
12	2008-07	2008	7		32.92	36.38	34						
13	2008-08	2008	8		31.78	35.02	28						
14	2008-09	2008	9		32.49	34.98	29						
15	2008-10	2008	10		37.72	39.03	33						
16	2008-11	2008	11		38.96	41.78	35						
17	2008-12	2008	12		40.47	40	34						
18	2009-01	2009	1		33.12	34.28	27						
19	2009-02	2009	2		29.48	27.22	26						
20	2009-03	2009	3		28.14	24.96	23						
21	2009-04	2009	4		23.96	21.82	20						
22	2009-05	2009	5		24.66	20.36	17						
23	2009-06	2009	6		23.86	19.79	19						
24	2009-07	2009	7		24.17	23.74	20						
25	2009-08	2009	8		25.91	26.78	19						

Figura 6. Demarcación de datos, para obtener la **Tabla dinámica**

En la ventana **Campos de la tabla dinámica** seleccione una variable cuantitativa, por ejemplo el precio de papa (Bs/arroba) de **LA PAZ**, con un clic derecho del Mouse presione la opción **Agregar a valores**, esta operación realice cinco veces.



	A	B	C	D
1				
2				
3	Cuenta de LA PAZ	Cuenta de LA PAZ2	Cuenta de LA PAZ3	Cuenta de LA PAZ4
4	129	129	129	129
5				
6				
7				
8				
9				
10				
11				
12				
13				
14				
15				

Figura 7. Tabla dinámica con cinco veces el precio de papa en La Paz

Una vez obtenida la Tabla dinámica con la salida de valores, ubique el cursor en el primer valor (1ra columna) y realice un clic derecho, ubicar **Configuración de campo de valor..**, encontrar **“Contar números”** y presionar **Aceptar** (Figura 8). Luego realice este procedimiento para el promedio, desviación estándar, valores mínimo y máximo; al mismo tiempo puede configurar el número de decimales en **Formato de número** ubicada en la ventana **Configuración de campo de valor..**.

	A	B	C	D	E	F
1						
2						
3	Cuenta de LA PAZ	Promedio de LA PAZ2	Desvest de LA PAZ3	Mín. de LA PAZ4	Cuenta de LA PAZ5	
4	129	34.40	8.85	19.36	129	
5						
6						
7						
8						
9						
10						
11						
12						
13						
14						
15						
16						
17						
18						
19						
20						
21						

Figura 8. Estadística descriptiva del precio de papa en departamento de La Paz

La ventaja de la tabla dinámica, radica en que puede utilizar otra variable cualitativa o cuantitativa. Por ejemplo utilizar la fecha disgregada en Mes y Año (Figura 9); en la *Tabla dinámica* hacer un clic derecho y seleccione **Actualizar**. Posteriormente un clic derecho en la *Tabla dinámica*, seleccione **Mostrar lista de campos**. Nuevamente un clic derecho en la *Tabla dinámica* e incluya **Año** en **Filas**



Fortalecimiento de la capacidad institucional en los sectores de desarrollo integral con coca, tráfico ilícito de drogas y seguridad alimentaria para una eficiente gestión del apoyo presupuestario sectorial – Contrato n° DCI/LA/2017/392-699



	A	B	C	D	E	F
1						
2						
3	Etiquetas de fila	Cuenta de LA PAZ	Promedio de LA PAZ2	Desvest de LA PAZ3	Mín. de LA PAZ4	Máx. de LA PAZ5
4	2008	12	31.77	5.42	24.95	40.47
5	2009	12	27.29	3.15	23.86	33.12
6	2010	12	30.46	9.98	19.36	45.83
7	2011	12	36.36	3.89	29.14	42.01
8	2012	12	29.31	6.59	23.22	43.45
9	2013	10	38.19	13.47	25.28	64.21
10	2014	12	37.32	3.02	31.18	40.54
11	2015	12	35.75	5.93	29.97	45.75
12	2016	12	47.75	11.32	34.59	65.38
13	2017	12	36.05	3.50	32.05	45.13
14	2018	11	28.32	3.27	25.00	32.71
15	Total general	129	34.40	8.85	19.36	65.38
16						
17						
18						
19						
20						
21						
22						

Campos de tabla dinámica

Seleccionar campos para agregar al informe:

- Fecha
- Año
- Mes
- El ALTO
- LA PAZ
- COCHABAMBA
- SANTA CRUZ
- ORURO

Arrastrar campos entre las áreas siguientes:

FILTROS

FILAS: Año

COLUMNAS

Valores

Figura 9. Tabla dinámica del precio de papa en La Paz e incluida Año en la Fila.

NOTA: Para obtener una presentación clásica ubicar el cursor en una casilla del campo de valores y presionar el clic derecho del Mouse, seleccionar **Opciones de tabla dinámica**, pestaña **Mostrar**, activar **Diseño de tabla dinámica clásica** y presionar **Aceptar**. En Fila se cambió **Año** por **Mes**.

	A	B	C	D	E	F
1			Coloque los campos de filtro de informe aquí			
2						
3			Valores			
4	Mes	Cuenta de LA PAZ	Promedio de LA PAZ2	Desvest de LA PAZ3	Mín. de LA PAZ4	Máx. de LA PAZ5
5	1	11	35.87	5.88	25.50	45.13
6	2	11	33.55	5.97	25.27	42.46
7	3	11	30.55	6.15	19.38	39.38
8	4	11	27.91	4.53	19.36	34.59
9	5	11	29.22	5.67	20.43	38.02
10	6	11	30.13	6.83	21.28	42.93
11	7	10	31.86	7.20	24.17	46.58
12	8	10	33.31	8.09	24.94	51.77
13	9	11	36.88	10.88	24.66	60.32
14	10	11	40.86	12.03	28.66	64.21
15	11	11	41.21	9.71	30.13	65.38
16	12	10	41.82	7.45	30.68	58.26
17	Total general	129	34.40	8.85	19.36	65.38
18						

Figura 10. Presentación clásica de la tabla dinámica con estadística descriptiva del precio de papa de LA PAZ (Bs/arroba) por Mes, datos de 2008-01 a 2018-11 (OAP-MDRyT, 2018).

Comentario

Los resultados son interesantes, por ejemplo el promedio de ingreso es mayor a partir de septiembre a enero. También presenta mayor variación en septiembre y octubre.



Histograma

Con la información de estadística descriptiva del precio de papa de LA PAZ, presenta valor mínimo, máximo (19.36 y 65.38 de la Figura 9) y dicha amplitud dividida entre siete se tiene $6.57 \sim 7$ clases. La generación del histograma, en el **Menú principal** seleccionar **DATOS**, luego **Análisis de datos e Histograma**; posteriormente en la ventana **Histograma**, en **Rango de entrada** seleccionar los datos de **LA PAZ** sin el rótulo, en **Rango de clases**, seleccionar los datos de la columna O, activar **Crear gráfico** y **Aceptar**; los resultados se visualizará en otra Hoja. Las clases del precio de papa Bs/arroba son de 19-26, 26-33, de 33-40,..., y sus frecuencias 28, 38, 35,..., respectivamente.

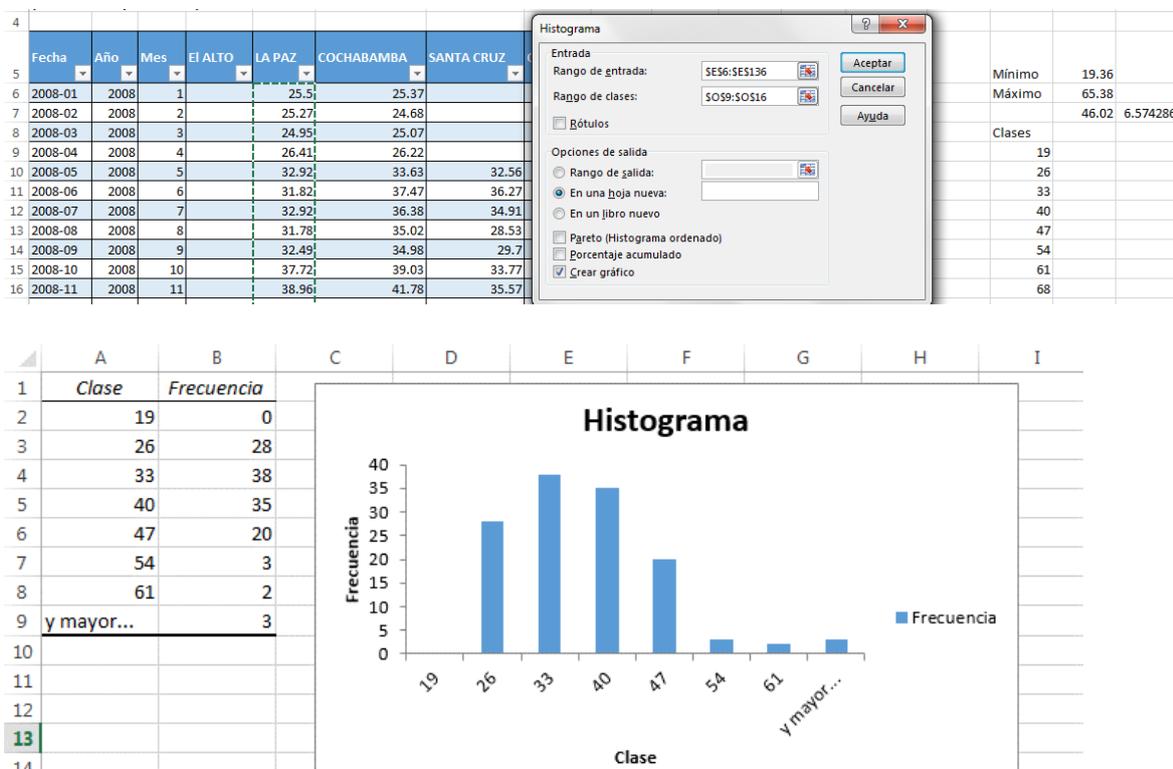


Figura 11. Histograma del precio de papa de LA PAZ (Bs/arroba), datos de 2008-01 a 2008-11 (OAP-MDRyT, 2018).

3.2. Variables cualitativas

Distribución de frecuencias

La distribución de frecuencias es la ocurrencia de observaciones con un carácter dentro una variable cualitativa. Estas distribuciones pueden ser de una o varias entradas (variables cualitativas).

- **Opción Tablas dinámicas:**

El archivo "Datos de papa evaluado 2017-2018.xlsx" contiene información de Datos de encuesta de papa (OAP-MDRyT, 2018) será empleado en el siguiente inciso.



Distribución de frecuencias de una vía

Ubicar el cursor en una casilla de la Tabla que contiene los datos. Posteriormente insertar una Tabla dinámica. Luego envíe una variable cualitativa, por ejemplo **Variedad**, al campo de **Agregar a valores**, en dos oportunidades. Repita esa operación, esta vez al campo de **Agregar a etiqueta de filas**. Por favor cambie a Tabla dinámica presentación clásica.

	A	B	C
3		Valores	
4	Variedad	Cuenta de Variedad	Cuenta de Variedad2
5	1(Waycha)	284	284
6	2(Imilla)	39	39
7	3(Sani)	41	41
8	4(Desire)	170	170
9	5(papa nativa)	13	13
10	6(otro)	49	49
11	Total general	596	596

Figura 12. Tabla inicial para la distribución de frecuencias de una vía con la variable Variedad de papa. La 3ra columna debe expresarse en % sobre total de la columna, en ese sentido en dicha columna (datos) realice un clic derecho, seleccione **Mostrar valores como** y **% del total de columnas**.

	A	B	C	D
3		Valores		
4	Variedad	Cuenta de Variedad	Cuenta de Variedad2	
5	1(Waycha)	284	47.65%	
6	2(Imilla)	39	6.54%	
7	3(Sani)	41	6.88%	
8	4(Desire)	170	28.52%	
9	5(papa nativa)	13	2.18%	
10	6(otro)	49	8.22%	
11	Total general	596	100.00%	



Figura 13. Distribución de frecuencias de una vía con conteo y expresado en porcentaje la variable Variedad de papa cultivada.

Otra forma de obtener la frecuencia en %, hacer un clic derecho en la 3ra columna (datos), ubicar **Configuración de campo del valor**, realizar un clic en la pestaña **Mostrar valores como**, en el filtro ubicar **% del total de columna** y **Aceptar**.

Distribución de frecuencias de dos vías

De la Tabla dinámica anterior, eliminar la 3ra columna, y hacer clic en la 2da columna (datos), de la ventana Campos de tabla dinámica seleccionar la variable P14_COMO_17 (14. ¿Cómo califica la cosecha del año agrícola 2016-2017?), hacer un clic derecho y utilizar **Agregar a etiquetas de columna**. Obtendrá los siguientes resultados.

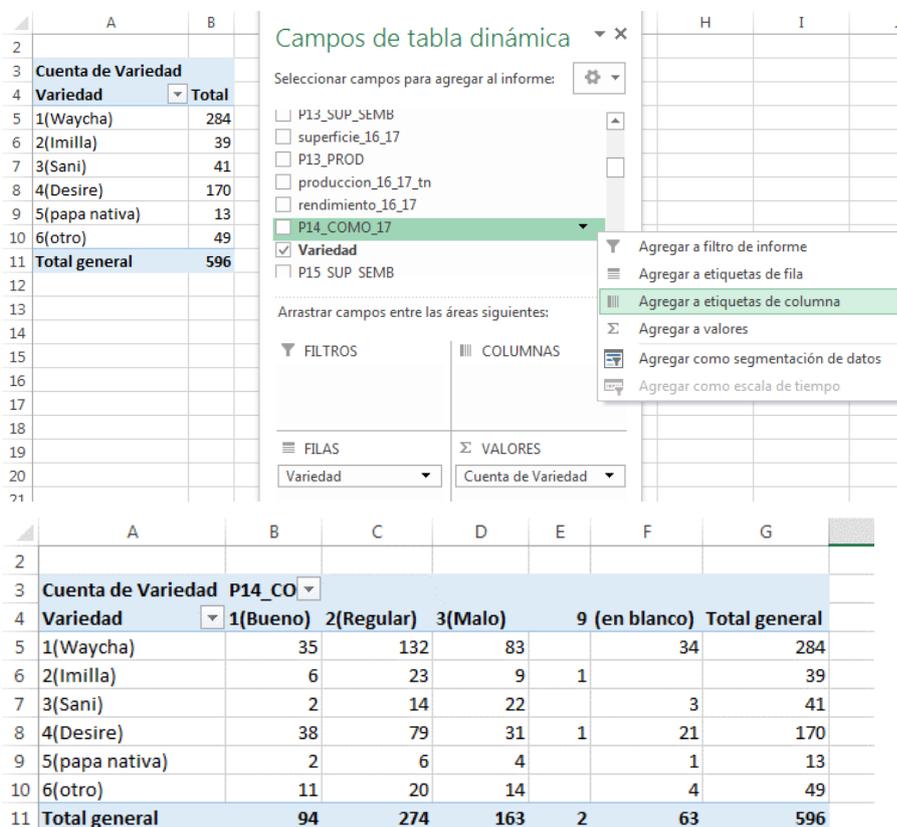


Figura 14. Distribución de frecuencias de dos vías con conteo de la Variedad y P14_COMO_17 (14. ¿Cómo califica la cosecha del año agrícola 2016-2017?).

Para expresar las frecuencias en porcentaje, copiar la Tabla dinámica de la Figura anterior y pegar debajo de ella. Realizar un clic derecho en la Tabla (datos), seleccione **Mostrar valores como** y **% del total de filas**.



	A	B	C	D	E	F	G	H
2								
3	Cuenta de Variedad P14_CO							
4	Variedad	1(Buena)	2(Regular)	3(Mala)		9 (en blanco)	Total general	
5	1(Waycha)	35	132	83		34	284	
6	2(Imilla)	6	23	9	1		39	
7	3(Sani)	2	14	22		3	41	
8	4(Desire)	38	79	31	1	21	170	
9	5(papa nativa)	2	6	4		1	13	
10	6(otro)	11	20	14		4	49	
11	Total general	94	274	163	2	63	596	
12								
13								
14	Cuenta de Variedad P14_CO							
15	Variedad	1(Buena)	2(Regular)	3(Mala)		9 (en blanco)	Total general	
16	1(Waycha)	12.32%	46.48%	29.23%	0.00%	11.97%	100.00%	
17	2(Imilla)	15.38%	58.97%	23.08%	2.56%	0.00%	100.00%	
18	3(Sani)	4.88%	34.15%	53.66%	0.00%	7.32%	100.00%	
19	4(Desire)	22.35%	46.47%	18.24%	0.59%	12.35%	100.00%	
20	5(papa nativa)	15.38%	46.15%	30.77%	0.00%	7.69%	100.00%	
21	6(otro)	22.45%	40.82%	28.57%	0.00%	8.16%	100.00%	
22	Total general	15.77%	45.97%	27.35%	0.34%	10.57%	100.00%	

Figura 15. Distribución de frecuencias de dos vías con conteo y % de Variedad y P14_COMO_17 (14. ¿Cómo califica la cosecha del año agrícola 2016-2017?).

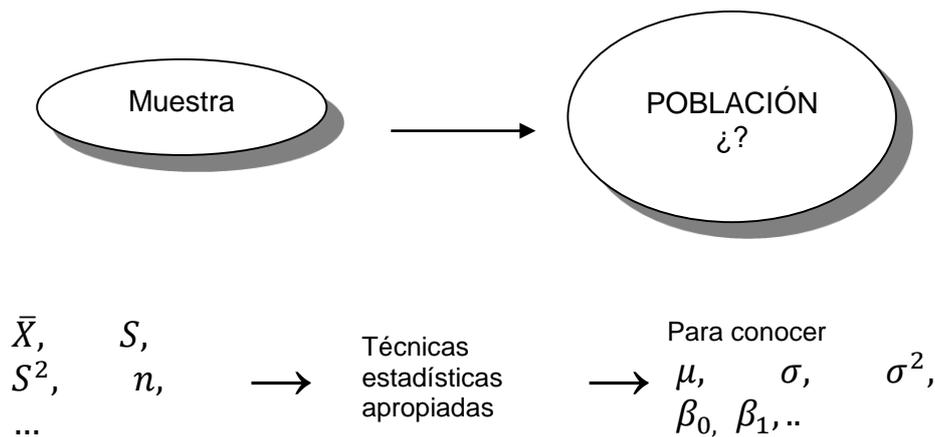
Comentario

La variedad Desire (1) presenta mayor valor (frecuencia y %) en la calificación de la cosecha como buena, le continúa otras variedades (6) en relación a los valores menores de Sani, Waycha, Imilla y papa nativa. La variedad Sani presenta mayor proporción la producción calificada como “Mala”.



4. ESTADÍSTICA INFERENCIAL

Se pretende generalizar los resultados obtenidos en la muestra a la población o universo (método inductivo), debido a que todos los elementos que forman una población no son conocidos o porque su número es demasiado grande. Las conclusiones obtenidas sobre la población basándose en la muestra se denominan inferencia estadística.



4.1. Intervalo de confianza para la media poblacional “ μ ” siendo la varianza poblacional desconocido.

$$\mu = \bar{x} \pm \frac{tS}{\sqrt{n}}$$

para población infinita

$$\mu = \bar{x} \pm \frac{tS}{\sqrt{n}} \sqrt{\frac{N-n}{N-1}}$$

para población finita

donde:

- \bar{x} : Media de la muestra
 - t : Valor de distribución t de Student a una probabilidad.
 - S : Desviación estándar de la muestra.
 - N : Número de elementos de la población.
 - n : Número de elementos de la muestra.
- Nota:** Si n es mayor a 30 puede utilizar el valor de Z de la distribución normal.

En Excel.



Con los datos del archivo “Datos de papa evaluado 2017-2018.xlsx” contiene información de Datos de encuesta de papa (OAP-MDRyT, 2018), de ella se plantea la interrogante ¿Cuánto es el rendimiento a nivel poblacional? Del Menú principal, entrar a **DATOS**, luego a **Análisis de datos** y **seleccionar Estadística descriptiva**. En esta ventana remarcar los datos de *rendimiento_16_17* en **Rango de entrada**, activar **Rótulo en la primera fila**, **Resumen de estadísticas**, **Nivel de confianza para la media** y presionar **Aceptar**. Luego realizar la operación de resta y suma con la **Media** y **Nivel de confianza (95%)** para la estimación de los límites de confianza.

	U	V	W	X	Y	Z	AA	AB	AC
2	p13prod	pro1617t	rend1617	p14ccalf	variedd	p15supsm	sup1718	p15prod	pro1718t
3									
	P13_PROD	produccion_16_17 tn	rendimien to_16_17	P14_COM O_17	Variedad	P15_SUP_S EMB	superficie_17 18	P15_PROD	produccion 17 18 tn
4									
5	4900.00	56.35	4.70						
6	500.00	25.00	2.50						
7	5000.00	57.50	7.19						
8	1200.00	13.80	1.73						
9	5660.00	65.09	9.30						
10	1100.00	37.95	5.42						
11	924.00	10.63	1.77						
12	100.00	4.60	0.77						
13									
14	30.00	30.00	5.00						
15	4000.00	46.00	9.20						
16	230.00	26.45	5.29						
17	5000.00	57.50	11.50						
18	150.00	16.50	3.30						

Entrada	Rango de entrada:	\$WS4:\$WS600	Aceptar
Agrupado por:	<input checked="" type="radio"/> Columnas	<input type="radio"/> Filas	Cancelar
<input checked="" type="checkbox"/> Rótulos en la primera fila			Ayuda
Opciones de salida			
<input type="radio"/> Rango de salida:			
<input checked="" type="radio"/> En una hoja nueva:			
<input type="radio"/> En un libro nuevo			
<input checked="" type="checkbox"/> Resumen de estadísticas			
<input checked="" type="checkbox"/> Nivel de confianza para la media	95	%	
<input type="checkbox"/> K-ésimo mayor:	1		
<input type="checkbox"/> K-ésimo menor:	1		

	A	B	C	D
1	<i>rendimiento_16_17</i>			
2				
3	Media	6.76403685		
4	Error típico	0.24957061		
5	Mediana	5.29		
6	Moda	4.6		
7	Desviación estándar	5.71838102		
8	Varianza de la muestra	32.6998815		
9	Curtosis	1.83345288		
10	Coficiente de asimetría	1.27043643		
11	Rango	32.1795		
12	Mínimo	0.008		
13	Máximo	32.1875		
14	Suma	3551.11935		
15	Cuenta	525	Inferior	Superior
16	Nivel de confianza(95.0%)	0.49028184	6.27	7.25
17				

Figura 16. Resultados de Estadística descriptiva y límites de confianza.

Comentario

El intervalo de confianza del promedio poblacional del rendimiento a nivel poblacional estaría entre 6.27 a 7.25 t/ha a una probabilidad de P: 0.05. Esta amplitud se debe a la variación elevada de los datos, obsérvese valores mínimos, máximos y la desviación estándar.



4.2. Intervalo de confianza para una proporción poblacional P.

$$P = p \pm Z \sqrt{\frac{p(1-p)}{n}}$$

para población infinita

$$P = p \pm Z \sqrt{\frac{p(1-p)}{n} \frac{(N-n)}{(n(N-1))}}$$

para población finita

donde:

p : Proporción de elementos que componen una muestra con una característica de interés.

Z : Valor de la distribución normal a una probabilidad.

n : Número de elementos de la muestra.

Nota: Si el número de elementos es menor a 30 puede utilizar el valor de t de Student.

Ejemplo.

Con los datos del archivo “Datos de papa evaluado 2017-2018.xlsx” contiene información de Datos de encuesta de papa (OAP-MDRyT, 2018), de ella se plantea la interrogante 14. ¿Cómo califica la cosecha del año agrícola 2016-2017?. Del Menú principal, entrar a **INSERTAR**, luego **Tabla dinámica** y seleccionar P14_COMO_17 tanto para Filas y Valores (Tabla 1).

Tabla 1. Distribución de frecuencias de P14_COMO_17 (14. ¿Cómo califica la cosecha del año agrícola 2016-2017?)

P14_COMO_17	Total
1 Bueno	94
2 Regular	274
3 Malo	163
9 ¿?	2
Total general	533

Posteriormente acceda al archivo “**Limites confianza.xlsx**”, esta contiene una plantilla con las fórmulas descritas anteriormente. Definiendo como caso positivo a 1 Bueno introduzca el valor 94 en **Número de casos del carácter de interés**, en el **Total de casos** el valor 533 y obtendrá los resultados. Se estima entre el 14.40 a 20.87 % de los entrevistados que expresaron su cosecha como Bueno a P: 0.05.



	A	B	C	D	E		A	B	C	D	E
1						1					
2						2					
3		Limite de confianza				3		Limite de confianza			
4		Número de casos del carácter de interés:	94			4		Número de casos del carácter de interés:	94		
5		Total de casos n =	533			5		Total de casos n =	533		
6		p =	=C4/C5			6		p =	0.176360225		
7		(1-p)=	=1-C6			7		(1-p)=	0.823639775		
8		Z _{0.05} =	1.96			8		Z _{0.05} =	1.96		
9		$Z \sqrt{\frac{p(1-p)}{n}} = =C8*RAIZ(C6*C7/C5)$				9		$Z \sqrt{\frac{p(1-p)}{n}} =$			
10							10				
11					%	11					%
12		Limite de confianza inferior =	=C6-C10		=C12*100	12		Limite de confianza inferior =	0.144003732		14.40
13		Limite de confianza superior =	=C6+C10		=C13*100	13		Limite de confianza superior =	0.208716718		20.87
14						14					

Figura 17. Límites de confianza para variable cualitativa con fórmulas y resultados.

4.3. Comparación de promedios en base la distribución de t de Student

Anterior a la prueba de comparación con la distribución de t de Student deberá verificarse el supuesto de homogeneidad de varianzas.

- **Homogeneidad de varianzas**

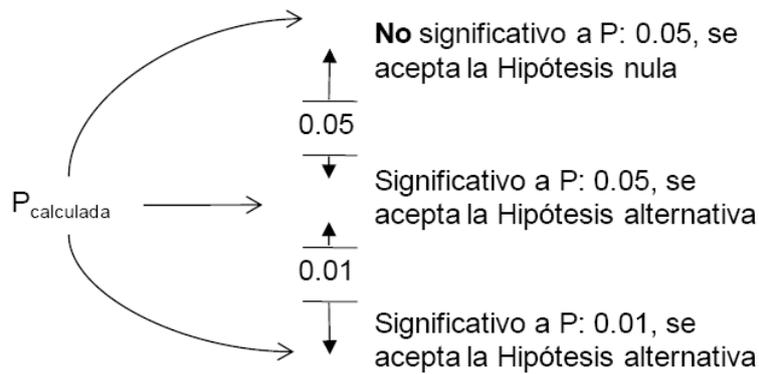
Supóngase que tenemos dos poblaciones con medias μ_1 y μ_2 . Se extrae una muestra aleatoria de cada población para probar la hipótesis de que μ_1 y μ_2 estén separados por una cantidad específica, usualmente se toma como cero.

Primero debe conocerse la homogeneidad de las varianzas muestrales (S^2) mediante la razón de F de Fisher.

$$F = \frac{S_{\text{mayor}}^2}{S_{\text{menor}}^2} \quad \text{con F Tab: y gl (mayor: menor)}$$

Donde la hipótesis nula es: $H_0: \sigma_1^2 = \sigma_2^2$

Para aceptar o rechazar la hipótesis nula es contrastar con la probabilidad calculada en base a la Razón de F calculada y grados de libertad, con las probabilidades de 0.05 y 0.01, así:



- **Comparación de dos promedios**

La hipótesis de la diferencia de dos promedios es

$$H_0: \mu_1 = \mu_2$$

$$H_1: \mu_1 \neq \mu_2$$

La misma se probaría con:

$$t = \frac{(\bar{Y}_1 - \bar{Y}_2) - (\mu_1 - \mu_2)}{S_{Y_1 - Y_2}}$$

donde:

\bar{Y}_1, \bar{Y}_2 : Son promedios muestrales de las poblaciones 1 y 2.

μ_1, μ_2 : Promedios poblacionales generalmente la diferencia es cero.

$S_{Y_1 - Y_2}$: Desviación estándar de las diferencias apropiada de dos promedios.

$$S_{Y_1 - Y_2} = \sqrt{S^2 * \left(\frac{1}{n_1} + \frac{1}{n_2} \right)}$$

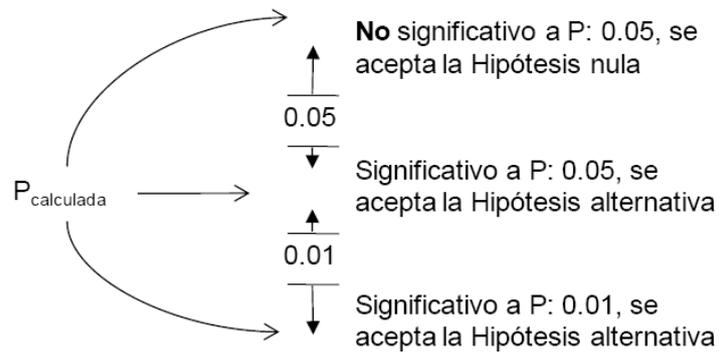
pero
$$S^2 = \frac{((n_1 - 1) * S_1^2) + ((n_2 - 1) * S_2^2)}{n_1 + n_2 - 2}$$

Las reglas de decisión son:

Se acepta H_0 si $t_{cal} < t_{tab}$ entonces los promedios son similares.

Se rechaza H_0 si $t_{cal} > t_{tab}$ entonces los promedios difieren

Otra alternativa es contrastar con la probabilidad calculada de la prueba de t, así:



NOTA:

Si el tamaño de muestra es mayor a 30, puede reemplazarse por comparación de promedio en base la distribución normal (Z).

Por otra parte, existen varias circunstancias en la comparación de dos promedios, para ello existen fórmulas en el caso de la distribución de t de Student.

En Excel

Con los datos del archivo “Datos de papa evaluado 2017-2018.xlsx” contiene información de Datos de encuesta de papa (OAP-MDRyT, 2018), de ella se plantea la interrogante ¿existe diferencias de rendimiento de papa entre la campaña 2016-2017 vs 2017-2018?

- **Prueba de homogeneidad de varianzas.**

En Excel, del **Menú principal** ubicar **Datos, Análisis de datos y Prueba F para varianzas de dos muestras** y presionar **Aceptar**. Seleccionar el grupo de valores de variedad 1 (Waycha) en **Rango para la variable 1**, variedad 4 (Desiré) en **Rango para la variable 2**, activar **Rótulos** y presionar **Aceptar**.

	V	W	X	Y	Z	AA	AB	AC	AD	
	produccion_16_17_tn	rendimien_to_16_17	P14_COMO_17	Variedad	P15_SUP_S EMB	superficie_17_18	P15_PROD	produccion_17_18_tn	rendimien_to_17_18	P16 O_1
4										
5	56.35	4.70	2(Regular)	4(Desire)	12.00	12.00	5000.00	57.50	4.79	
6	25.00	2.50	2(Regular)	1(Waycha)	10.00	10.00	1000.00	50.00	5.00	
7	57.50	7.19	2						7.19	
8	13.80	1.73	2						5.75	
9	65.09	9.30	2						8.05	
10	37.95	5.42	2						5.69	
11	10.63	1.77	2						4.43	
12	4.60	0.77	2						0.77	
13										
14	30.00	5.00	2						0.60	
15	46.00	9.20	1						9.20	
16	26.45	5.29	2						8.05	
17	57.50	11.50	1						1.15	
18	16.50	3.30	1						1.32	

Entrada		Aceptar
Rango para la variable 1:	SWS4:SW\$600	Cancelar
Rango para la variable 2:	SAD\$4:SAD\$600	Ayuda
<input checked="" type="checkbox"/> Rótulos		
Alfa:	0.05	
Opciones de salida		
<input type="radio"/> Rango de salida:		
<input checked="" type="radio"/> En una hoja nueva:		
<input type="radio"/> En un libro nuevo		



	A	B	C
1	Prueba F para varianzas de dos muestras		
2			
3		rendimiento _16_17	rendimiento _17_18
4	Media	6.76403685	7.339877388
5	Varianza	32.69988145	31.98866206
6	Observaciones	525	578
7	Grados de libertad	524	577
8	F	1.022233484	
9	P(F<=f) una cola	0.397829849	
10	Valor crítico para F (una cola)	1.150525796	
11			

Figura 18. Prueba de F para determinar homogeneidad de varianzas.

Comentario

La decisión de hipótesis va a favor que ambas varianzas son homogéneas a P: 0.05, debido a que el valor P(F<=f) es mayor a la probabilidad de 0.05.

- **Comparación de dos promedios.**

En Excel, del Menú principal ubicar **Datos, Análisis de datos y Prueba t para dos muestras suponiendo varianza iguales** y presionar **Aceptar**. Seleccionar el grupo de valores de la **rendimiento_16_17** en **Rango para la variable 1**, el grupo de datos del **rendimiento_17_18** en **Rango para la variable 2**, active **Rótulos** y presione **Aceptar**.

	V	W	X	Y	Z	AA	AB	AC	AD	E
	produccion _16_17_tn	rendimien to_16_17	P14_COM O_17	Variedad	P15_SUP_S EMB	superficie 17_18	P15_PROD	produccion _17_18_tn	rendimien to_17_18	P16_C O_18
4										
5	56.35	4.70	2(Regular)	4(Desire)	12.00	12.00	5000.00	57.50	4.79	
6	25.00	2.50							5.00	
7	57.50	7.19							7.19	
8	13.80	1.73							5.75	
9	65.09	9.30							8.05	
10	37.95	5.42							5.69	
11	10.63	1.77							4.43	
12	4.60	0.77							0.77	
13										
14	30.00	5.00							0.60	
15	46.00	9.20							9.20	
16	26.45	5.29							8.05	
17	57.50	11.50							1.15	
18	16.50	3.30							1.32	
19	79.10	15.82							17.63	
20	2.76	0.55	2(Regular)	1(waycha)	5.00	5.00	600.00	2.76	0.55	
21	26.57	8.86	3(Malo)	4(Desire)	4.50	4.50	3080.00	35.42	7.87	
22	14.38	9.58	2(Regular)	4(Desire)	4.50	4.50	607.00	69.81	15.51	



	A	B	C
1	Prueba t para dos muestras suponiendo varianzas iguales		
2			
3		rendimiento 16 17	rendimiento 17 18
4	Media	6.76403685	7.339877388
5	Varianza	32.69988145	31.98866206
6	Observaciones	525	578
7	Varianza agrupada	32.3271534	
8	Diferencia hipotética de las medias	0	
9	Grados de libertad	1101	
10	Estadístico t	-1.679864753	
11	P(T<=t) una cola	0.046633664	
12	Valor crítico de t (una cola)	1.646238786	
13	P(T<=t) dos colas	0.093267327	
14	Valor crítico de t (dos colas)	1.962120966	
15			

Figura 19. Prueba de t para dos muestras suponiendo varianzas iguales.

Comentario

El rendimiento de papa presento diferencias significativa entre la campaña 2016-2017 frente a 2017-2018 a una probabilidad de P: 0.05 (t prueba de una cola), por las magnitudes observadas en los promedios existe un incremento en la campaña 2017-2018.

4.4. Comparación de un promedio con un indicador.

La siguiente formula, en base a la distribución de t de Student, se refiere a la determinación que si la media muestral es similar a la media poblacional. La media poblacional puede ser definida o mayormente es considerada igual a cero. Esto permite utilizar para contrastar con el valor de un indicador correspondiente a una variable.

$$t = \frac{(\bar{Y} - \mu)}{S} \sqrt{n-1}$$

\bar{Y} Media de la muestra
 μ Constante o media poblacional
 n Tamaño de la muestra
 S Desviación estándar

En Excel para obtener Pr > t, emplearse la función =DISTR.T.CD(), considera el valor de t calculado y grados de libertad (n-1)

Tabla 1. Prueba de t de Student contrastando promedios de rendimiento con el indicador rendimiento de papa 5.98 t/ha (30% adicional al rendimiento 4.6 t/ha del 2014)

Variedad	Cuenta	Promedio	Desv. Est.	Indicador (t/ha)	t	t0.05	t0.01	Pr > t
1(Waycha)	278	6.76	5.79	5.98	2.25 *	1.65	2.34	0.012693137
2(Imilla)	39	7.37	6.21	5.98	1.40 ns	1.69	2.43	0.084974202
3(Sani)	40	6.30	4.49	5.98	0.46 ns	1.68	2.43	0.325701009
4(Desire)	161	8.66	5.40	5.98	6.28 **	1.65	2.35	0.000000001
5(Papa nativa)	12	7.83	6.25	5.98	1.03 ns	1.80	2.72	0.163648344



6(Otro)	48	7.00	5.46	5.98	1.29 ns	1.68	2.41	0.101308481
Total	578	7.34	5.66	5.98	5.78 **	1.65	2.33	0.000000006

Comentario

La variedad Waycha, Desire y en forma general el rendimiento (Total), presentó significación estadística a P: 0.05, es decir existe diferencias a favor de la media de la muestra sobre el indicador. En las variedades de Imilla, Sani, Papa nativa y Otro son no significativas, es decir no superan al indicador.

4.5. Prueba de diferencia de proporciones

Quando la base de datos posee variables de respuesta categóricas binomiales puede emplearse la siguiente fórmula para diferenciar ambas muestras.

$$Z = \frac{p_1 - p_2}{S_p}$$

donde:

$$p_1 \text{ y } p_2: \frac{\text{casos_positivos_con_característica_estudiada}}{\text{total_de_elementos_de_la_muestra}}$$

$$S_p = \sqrt{p(1-p)\left(\frac{1}{n_1} + \frac{1}{n_2}\right)} \quad p = \frac{(n_1 * p_1) + (n_2 * p_2)}{n_1 + n_2}$$

n_1 y n_2 : número total de elementos que contiene la muestra 1 y 2

Las reglas de decisión son:

Se acepta H_0 si $Z_{cal} < Z_{tab}$ las proporciones son similares

Se rechaza H_0 si $Z_{cal} > Z_{tab}$ los proporciones difieren

En Excel:

Con los datos del archivo "Datos de papa evaluado 2017-2018.xlsx" (OAP-MDRyT, 2018), de ella se plantea si existe diferencias en la opinión 14. ¿Cómo califica la cosecha del año agrícola 2016-2017? Respecto a la variedad 1 (Waycha) vs variedad 4 (Desiré). Del Menú principal, entrar a **INSERTAR**, luego **Tabla dinámica** y seleccionar P14_COMO_17 tanto para en **Agregar a etiquetas de Filas** y en **Agregar a Valores**; además de incluir **Variedad** en **Agregar a etiquetas de columna**.

Tabla 2. Distribución de frecuencias de P14_COMO_17 (14. ¿Cómo califica la cosecha del año agrícola 2016-2017?) por variedad



P14_COMO_17	Variedad de papa	
	1 (Waycha)	4 (Desiré)
1 Bueno	35	38
2 Regular	132	79
3 Malo	83	31
9 ¿?		1
Total	250	149

En el archivo “Diferencia proporciones.xlsx” esta la plantilla con las fórmulas de **Diferencia de dos proporciones**, en ese sentido introducir los valores de **1 Bueno** y el **Total**. Obsérvese que el valor Z calculada es mayor a las $Z_{0.05}$ y $Z_{0.01}$ por lo cual existe diferencias de proporciones a favor de Encuestados que afirman buena producción de papa en la variedad 4 (Desiré) a $P: 0.01$.

	A	B	C	D
3				
4			Grupo 1	Grupo 2
5		Total de observaciones n=	250	149
6		Número de casos positivos =	35	38
7		p_i	0.14	0.2550336
8				
9		p	0.182957393	
10		S_p	0.040014799	
11		Z calculada	-2.874775305	
12		$Z_{0.05}$	1.96	
13		$Z_{0.01}$	2.58	
14				

Figura 20. Diferencia de dos proporciones en base a fórmulas.

4.6. Análisis de varianza de clasificación de una vía

El objetivo del análisis de varianza de clasificación de una vía es determinar diferencias debido a una variable dependiente, esta técnica suele usarse en diseños muestrales (Cochran, 1974; Leon y Quiroz, 1999).

Modelo estadístico: Se tiene por consideración básica que cualquier observación puede ser representada por medio de un modelo lineal:

$$X_{ij} = \mu + \tau_i + \xi_{ij}$$

donde:

- i: 1, 2, 3 ... t de estrato o grupos
- j: 1, 2, 3 ... r de muestra



- X_{ij} : Un valor de una variable de respuesta de la unidad muestral que reside en el i-ésimo estrato y corresponde a la j-ésima muestra.
- μ : Media general.
- τ_i : Efecto fijo del i – ésimo estrato.
- ξ_{ij} : Efecto aleatorio del residuo, $\xi_{ij} \sim \text{DNII}(0, \sigma_e^2)$

La verificación del supuesto:

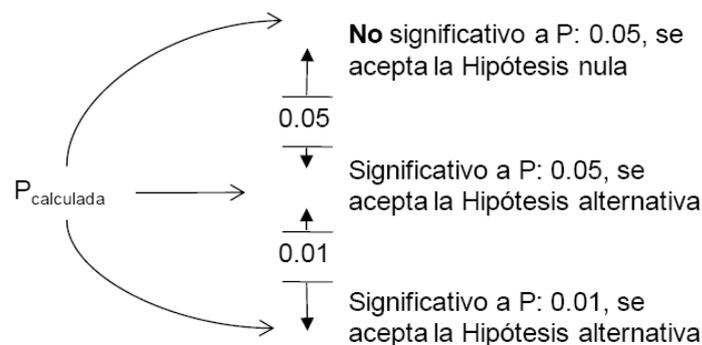
La normalidad de los residuos puede emplearse estadística descriptiva como el diagrama de caja, histograma, tallos y hojas. La homogeneidad de varianzas puede utilizar la prueba de O'Brien's (1979), Levene's (1960) y Bartlett's (1937) citados por SAS (2013).

Hipótesis y Regla de decisión

La hipótesis nula (H_0) se plantea a la similitud entre los estratos y la hipótesis alternativa (H_1) al menos dos estratos difieren

$$H_0 : \bar{X}_1 = \bar{X}_2 = \dots = \bar{X}_t \quad \text{y} \quad H_1 : \bar{X}_1 \neq \bar{X}_2 \neq \dots \neq \bar{X}_t$$

Para decidir cuál de las hipótesis se acepta o rechaza, considere la siguiente regla de decisión en base a probabilidades:



En Excel

El Proyecto Apoyo a la producción de papa de pequeños productores de los departamentos de La Paz, Cochabamba, Oruro, Potosí y Santa Cruz (SEPA-PASA-MDRyT, 2011), generó información sobre el rendimiento, de ella se extrajo datos del Cochabamba, Provincia Arque, Municipio Tacopaya de nueve Comunidades, los datos se exponen en el archivo "Papa 2010-2011.xlsx". Los datos se exponen en la Tabla 2 y la estadística descriptiva (obtenida con tabla dinámica) en la Tabla 3. Surge la interrogantes que los rendimientos poseen diferencias entre comunidades, en ese sentido se aplicará el Análisis de varianza con Excel.



Tabla 2. Datos de rendimiento de papa (t/ha) variedad Waycha de la campaña 2010-2011. (SEPA-PASA-MDRyT, 2011),

Nob	Comunidad								
	Chajwa Chajwa	Condori ri	Kjochi ni	K'otu Mayu	Murmunta ni Grande	Okioki Kotumayu	Siquipampa Chico	Siquipampa Grande	Viluyo
1	12.146	13.330	11.797	13.476	7.870	6.537	12.856	12.308	14.399
2	10.204	14.344	7.577	17.778	12.170	10.176	17.528	13.105	13.171
3	12.475	17.660	13.661	14.205	9.901	9.607	16.097	13.958	9.980
4	13.145	8.729	14.097	10.690	16.327	10.613	16.865	13.542	16.327
5			9.911	9.945		12.286	16.611	8.125	15.625
6			10.204	9.384		8.132	16.147	6.400	12.564
7			9.314	10.161		10.677		18.703	14.896
8			18.132	14.545		12.089		9.218	9.333
9				14.213		14.396		16.750	13.052
10				14.751				6.240	12.509
11				11.229				18.797	9.563
12				14.354				11.475	13.320
13									12.245
Mediana	11.993	13.516	11.837	12.894	11.567	10.501	16.017	12.385	12.845

Tabla 3. Estadística descriptiva del rendimiento de papa (t/ha) variedad Waycha de la campaña 2010-2011 de Cochabamba, Provincia Arque, Municipio Tacopaya y nueve Comunidades. (SEPA-PASA-MDRyT, 2011),

Comunidad	Valores				
	Cuenta	Promedio	Desv. Estand.	Mínimo	Máximo
Chajwa Chajwa	4	11.993	1.263	10.204	13.145
Condoriri	4	13.516	3.688	8.729	17.660
Kjochini	8	11.837	3.358	7.577	18.132
K'otu Mayu	12	12.894	2.560	9.384	17.778



Murmuntani Grande	4	11.567	3.627	7.870	16.327
Okioki Kotumayu	9	10.501	2.323	6.537	14.396
Siquipampa Chico	6	16.017	1.635	12.856	17.528
Siquipampa Grande	12	12.385	4.338	6.240	18.797
Viluyo	13	12.845	2.212	9.333	16.327
Total	72	12.555	3.104	6.240	18.797

Para procesar los rendimientos en Excel, del **Menú principal** seleccionar **Datos, Análisis de datos y Análisis de varianza de un factor**. En el **Rango de entrada** remarcar los datos con sus encabezados, activar **Agrupado por Columna**, activar **Rótulos en la primera fila** y **Aceptar**.

The screenshot shows an Excel spreadsheet with the following data (rows 4-17):

Nob	Chajwa	Condoriri	Kjochini	K'otu	Murmuntani	Okioki	Siquipampa	Siquipampa	Viluyo
	Chajwa		Mayu		Grande	Kotumayu	Chico	Grande	
1	12.146	13.330	11.797	13.476	7.870	6.537	12.856	12.308	14.399
2	10.204	14.344	7.577	17.778	12.170	10.176	17.528	13.105	13.171
3	12.475	17.660	13.661	14.205	9.901	9.607	16.097	13.958	9.980
4	13.145	8.729	14.097	10.690	16.327	10.613	16.865	13.542	16.327
5			9.911	9.945		12.286	16.611	8.125	15.625
6			10.204	9.384		8.132	16.147	6.400	12.564
7								18.703	14.896
8								9.218	9.333
9								16.750	13.052
10								6.240	12.509
11								18.797	9.563
12								11.475	13.320
13									12.245
Total general	11.993	13.516						12.385	12.845

The ANOVA dialog box is open with the following settings:

- Entrada: Rango de entrada: \$B\$4:\$J\$17
- Agrupado por: Columnas
- Rótulos en la primera fila
- Alfa: 0.05
- Opciones de salida: En una hoja nueva



	A	B	C	D	E	F	G
1	Análisis de varianza de un factor						
2							
3	RESUMEN						
4	<i>Grupos</i>	<i>Cuenta</i>	<i>Suma</i>	<i>Promedio</i>	<i>Varianza</i>		
5	Chajwa Chajwa	4	47.97	11.9925	1.59445233		
6	Condoriri	4	54.063	13.51575	13.6027616		
7	Kjochini	8	94.693	11.836625	11.2789277		
8	K'otu Mayu	12	154.731	12.89425	6.55203111		
9	Murmuntani Grande	4	46.268	11.567	13.154858		
10	Okioki Kotumayu	9	94.513	10.50144444	5.39410378		
11	Siquipampa Chico	6	96.104	16.01733333	2.67405627		
12	Siquipampa Grande	12	148.621	12.38508333	18.8189308		
13	Viluyo	13	166.984	12.84492308	4.89289508		
14							
15							
16	ANÁLISIS DE VARIANZA						
17	<i>Origen de las variaciones</i>	<i>Suma de cuadrados</i>	<i>Grados de libertad</i>	<i>Promedio de los cuadrados</i>	<i>F</i>	<i>Probabilidad</i>	<i>Valor crítico para F</i>
18	Entre grupos	125.691455	8	15.71143192	1.77283197	0.099344616	2.08918504
19	Dentro de los grupos	558.327143	63	8.862335607			
20							
21	Total	684.018599	71				
22							

Figura 21. Ventana Análisis de datos, remarcada el Análisis de varianza de un factor y resultados.

Comentario

Según el análisis de varianza es no significativo el efecto de Comunidades a P: 0.05, es decir los rendimientos de papa variedad Waycha son similares en estas Comunidades de Tacopaya (Figura 21).

4.7. Relación entre variables

- **Correlación simple**

El coeficiente de correlación (r) se utiliza para determinar la relación entre dos variables y la fórmula es:

$$r = \frac{\sum(X - \bar{X})(Y - \bar{Y}) / (n - 1)}{\sqrt{\sum(X - \bar{X})^2 / (n - 1)} \sqrt{\sum(Y - \bar{Y})^2 / (n - 1)}}$$

donde:

r : varía de -1 a +1

$\sum(X - \bar{X})(Y - \bar{Y}) / (n - 1)$: Covarianza de las variables Y y X,

$\sqrt{\sum(X - \bar{X})^2 / (n - 1)}$, $\sqrt{\sum(Y - \bar{Y})^2 / (n - 1)}$: Desviación estándar de la variable Y y X.



En Excel

Con los datos del archivo “Datos de papa evaluado 2017-2018 ejercicio.xlsx Hoja: Datos_papa” contiene información de Datos de encuesta de papa (OAP-MDRyT, 2018), de ella extraemos utilizando Tabla dinámica la información de las variables FOLIO, promedios de: rendimiento_17_18, superficie_17_18, semilla_toneladas, P20_PLAGA_1 y P20_ENFERMEDAD_1 en la Hoja: “Rendim y otras”. Se eliminan las observaciones que no tienen datos. Se desea determinar la relación o asociación entre ellas, el set de datos tiene n=442 y cinco variables.

Del Menú principal seleccionar **Datos, Análisis de datos y Coeficiente de correlación**. En la ventana **Coeficiente de correlación** remarcar las variables promedio de: rendimiento_17_18, superficie_17_18, semilla_toneladas, P20_PLAGA_1 y P20_ENFERMEDAD_1 incluyendo el rótulo o encabezado en **Rango de entrada**, activar **Rótulos en la primera fila** y **Aceptar**.

	A	B	C	D	E	F	G
3		Promedio de					
4	FOLIO	rendimiento_17_18	superficie_17_18	semilla_toneladas	P20_PLAGA_1	P20_ENFERMEDAD_1	
5	79	6.3	1	0.46	98	99	
6	21	13.6	0.05	2.07	12	99	
7	282	4.375	0.72	0.14375	99	99	
8	20	14.4	0.2	0.2875	12	99	
9	381	11.11111111	0.45	0.69	98	99	
10	253	4.6	0.5	1.15	98	98	
11	427	0.92	0.5	0.115	14	98	
12	340	0.525	2	0.94875	13	98	
13	7	4.189944134	0.179				
14	463	3.45	0.12				
15	8	4.326923077	0.52				
16	283	1.4375	0.72				
17	9	3.333333333	0.15				
18	403	5.75	0.16				
19	2	6.75	0.5				
20	443	0.75	2				
21	3	0.002130682	0.176				
22	22	13.6	0.066666667				
23	24	3.066666667	0.225				
24	327	3.85	1				
25	25	6.9	0.3	0.575	13	98	

Coeficiente de correlación

Entrada
Rango de entrada: SBS4:SF5446

Agrupado por:
 Columnas
 Filas

Rótulos en la primera fila

Opciones de salida
 Rango de salida:
 En una hoja nueva:
 En un libro nuevo

Aceptar Cancelar Ayuda

Figura 22. Ventana de Coeficiente de correlación activada con las opciones.

	A	B	C	D	E	F
1		rendimiento_17_18	superficie_17_18	semilla_toneladas	P20_PLAGA_1	P20_ENFERMEDAD_1
2	rendimiento_17_18	1				
3	superficie_17_18	-0.112765118	1			
4	semilla_toneladas	0.400936322	-0.076057665	1		
5	P20_PLAGA_1	0.056062859	0.098967003	0.087483176	1	
6	P20_ENFERMEDAD_1	-0.041332315	-0.001445086	-0.106595824	0.289371948	1



Figura 23. Resultados de coeficientes de correlación simple (Pearson) de variables de superficie, semilla, plaga-1, enfermedad-1 y el rendimiento de papa.

Los coeficientes de correlación pueden contrastar con valores tabulares de la Distribución R de correlación en el Texto de Steel y Torrie (1992) donde el valor de r a 0.05 con 441 grados de libertad es 0.0939 y para r a 0.01 es 0.12267

	A	B	C	D	E	F	G	H	I	J
1		rendimiento _17_18	superficie _17_18	semilla_toneladas		P20_PLAGA_1		P20_ENFER MEDAD_1		
2	rendimiento_17_18	1								
3	superficie_17_18	-0.11276512 *	1							
4	semilla_toneladas	0.400936322 **	-0.076058 ns	1						
5	P20_PLAGA_1	0.056062859 ns	0.098967 *	0.087483176 ns	1					
6	P20_ENFERMEDAD_1	-0.04133232 ns	-0.001445 ns	-0.106595824 *	0.289371948 **	1				

Figura 24. Resultados de coeficientes de correlación simple (Pearson) y significancia de variables de superficie, semilla, plaga-1, enfermedad-1 y el rendimiento de papa (ns: no significativo a P: 0.05; *: Significativo a P: 0.05 y; ** significativo a P: 0.01)

Comentario

El rendimiento está asociado con la superficie sembrada en forma negativa, es decir a medida que aumenta la superficie el rendimiento disminuye, este resultado puede ser a la deficiencia de mano de obra; así mismo el rendimiento también está asociado con la semilla empleada y en forma positiva.

La superficie sembrada con la ocurrencia de plagas es significativa y en forma positiva, es decir a medida que aumenta la superficie sembrada se tiene el incremento de la ocurrencia de la plaga; también se observa asociación de semilla empleada con la ocurrencia de enfermedad y en forma negativa, esta relación tiene contradicción con el sentido lógico de la naturaleza, pues el coeficiente de correlación daría a entender a medida que aumenta la cantidad de semilla la enfermedad disminuye.

La ocurrencia de plaga y enfermedad están asociadas y en forma positiva, deduciéndose a medida que aumenta la plaga la enfermedad aumenta.



- **Regresión lineal y múltiple**

El análisis de regresión lineal se usa cuando se quiere determinar si la variación de la variable independiente o explicativa (X) explica una variación de la variable dependiente o a explicar (Y). La ecuación es la siguiente (Gujarati, D. 2002):

$$y_i = \beta_0 + \beta_1 x_i + \xi_i$$

donde:

β_0 : Intersección de la recta con el eje vertical, es decir indica el valor esperado de Y cuando X es igual a 0. $\beta_0 = \bar{Y} - \beta_1 \bar{X}$

β_1 : Pendiente de la recta de regresión, es decir es la variación esperada de Y cuando X varía en una unidad.

$$\beta_1 = \frac{n \sum_{i=1}^n x_i y_i - \left(\sum_{i=1}^n x_i \right) \left(\sum_{i=1}^n y_i \right)}{n \sum_{i=1}^n x_i^2 - \left(\sum_{i=1}^n x_i \right)^2}$$

ξ_i : Valor residual \sim NID (0, σ^2)

En Excel

Continuando con los datos del archivo “Datos de papa evaluado 2017-2018.xlsx” (OAP-MDRyT, 2018), del ejercicio anterior utilizar los datos para determinar cómo es afectada el rendimiento_17_18 por la variable superficie_17_18.

Del **Menú principal** seleccionar **Datos, Análisis de datos y Regresión**. En la ventana **Regresión** remarcar la variable rendimiento_17_18 en **Rango Y de entrada** y las variables superficie_17_18 en **Rango X de entrada** incluyendo el rótulo o encabezado, activar **Rótulos en la primera fila**, activar **Nivel de confianza 95%** o **99%** y **Aceptar**.



	A	B	C	D	E	F	G
1	Resumen						
2							
3	<i>Estadísticas de la regresión</i>						
4	Coefficiente de correlación múltiple	0.11276512					
5	Coefficiente de determinación R^2	0.01271597					
6	R^2 ajustado	0.01047214					
7	Error típico	5.37415028					
8	Observaciones	442					
9							
10	ANÁLISIS DE VARIANZA						
11		<i>Grados de libertad</i>	<i>Suma de cuadrados</i>	<i>Promedio de los cuadrados</i>	<i>F</i>	<i>Valor crítico de F</i>	
12	Regresión	1	163.674015	163.674015	5.667090163	0.01771095	
13	Residuos	440	12707.8561	28.8814912			
14	Total	441	12871.5302				
15							
16		<i>Coefficientes</i>	<i>Error típico</i>	<i>Estadístico t</i>	<i>Probabilidad</i>	<i>Inferior 95%</i>	<i>Superior 95%</i>
17	Intercepción	7.53282869	0.32857829	22.9255218	4.14418E-77	6.88705074	8.178606639
18	superficie_17_18	-0.447253	0.18787682	-2.38056509	0.017710947	-0.81650048	-0.078005511

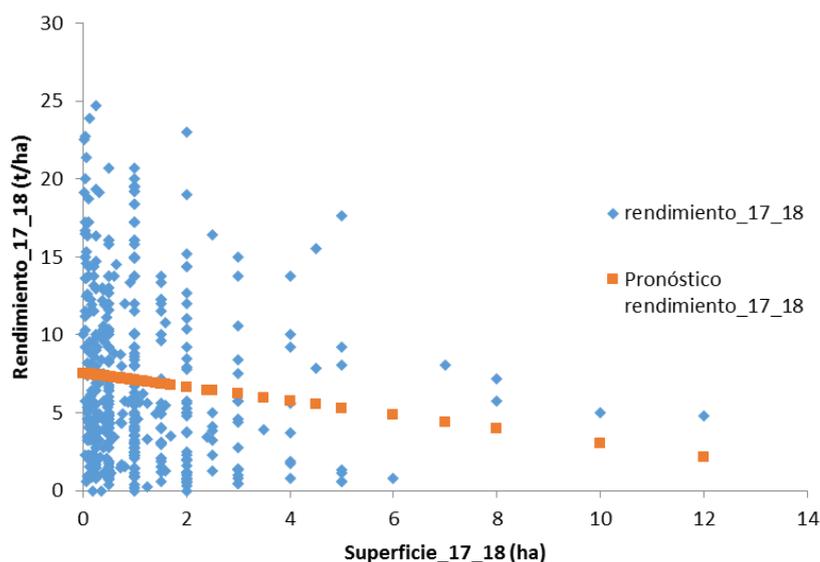


Figura 25. Resultados del análisis de regresión simple del rendimiento y la superficie sembrada de papa de la campaña 2017-2018

Comentario

El coeficiente de determinación presentó 1.27 % traduciéndose a la variación explicada debido a la superficie sobre la variación del rendimiento; así mismo el análisis de varianza presentó significación a la P: 0.05 debido a la regresión lineal; Por otra parte el coeficiente intercepto también es significativo deduciéndose que el intercepto es diferente de cero a nivel poblacional a P: 0.01 y el coeficiente de regresión también es significativo a P: 0.05, este coeficiente en la muestra implica por cada unidad de superficie a incrementar, disminuye en 0.447 t el rendimiento y a nivel poblacional el rendimiento disminuye según los valores del intervalos de confianza a P: 0.05.



- **Regresión múltiple**

Similar al de regresión lineal simple, en la múltiple intervienen dos o más variables independientes.

$$Y_i = b_0 + b_1X_{1i} + b_2X_{2i} + b_3X_{3i} + u_i$$

En Excel

Continuando con los datos del archivo “Datos de papa evaluado 2017-2018 ejercicio.xlsx Hoja: Rendim y otras” (OAP-MDRyT, 2018), utilizar los datos del rendimiento (rendimiento_17_18) y otras variables (superficie_17_18, semilla_toneladas, P20_PLAGA_1 y P20_ENFERMEDAD_1).

Del **Menú principal** seleccionar **Datos, Análisis de datos y Regresión**. En la ventana **Regresión** remarcar la variable rendimiento_17_18 en **Rango Y de entrada** y las variables superficie_17_18, semilla_toneladas, P20_PLAGA_1 y P20_ENFERMEDAD_1 en **Rango X de entrada** incluyendo el rótulo, activar **Rótulos en la primera fila**, activar **Nivel de confianza** 95% o 99% y **Aceptar**.

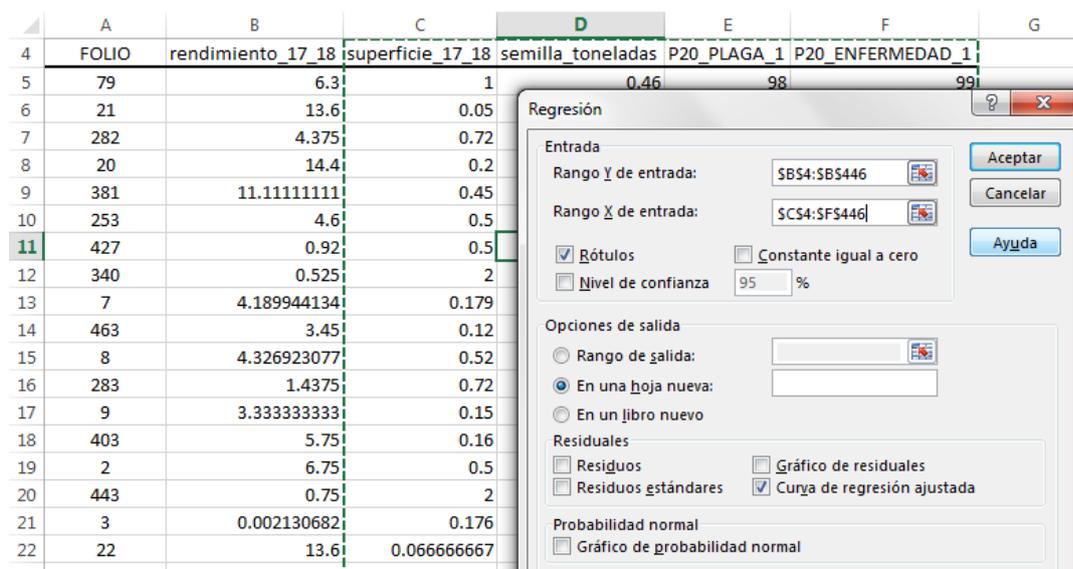


Figura 26. Ventana de análisis de regresión activadas las opciones de rótulo y Nivel de confianza.



	A	B	C	D	E	F	G
1	Resumen						
2							
3	<i>Estadísticas de la regresión</i>						
4	Coeficiente de correlación m	0.410534343					
5	Coeficiente de determinación	0.168538447					
6	R ² ajustado	0.160927815					
7	Error típico	4.948749341					
8	Observaciones	442					
9							
10	ANÁLISIS DE VARIANZA						
11		<i>Grados de libertad</i>	<i>Suma de cuadrados</i>	<i>Promedio de los cuadrados</i>	<i>F</i>	<i>Valor crítico de F</i>	
12	Regresión	4	2169.347704	542.3369261	22.14513139	1.15691E-16	
13	Residuos	437	10702.18246	24.49012004			
14	Total	441	12871.53016				
15							
16		<i>Coefficientes</i>	<i>Error típico</i>	<i>Estadístico t</i>	<i>Probabilidad</i>	<i>Inferior 95%</i>	<i>Superior 95%</i>
17	Intercepción	4.140900513	0.709199316	5.838838841	1.02658E-08	2.747034977	5.53476605
18	superficie_17_18	-0.342545434	0.174666061	-1.961144782	0.050496754	-0.685835391	0.000744523
19	semilla_toneladas	2.905959421	0.330411172	8.794979296	3.32872E-17	2.25656688	3.555351963
20	P20_PLAGA_1	0.004511964	0.006286025	0.717776993	0.473278325	-0.007842635	0.016866563
21	P20_ENFERMEDAD_1	-0.00147577	0.007201528	-0.204924611	0.83772644	-0.015629706	0.012678166

Figura 27. Resultados del análisis de regresión del rendimiento y las variables superficie_17_18, semilla_toneladas, P20_PLAGA_1 y P20_ENFERMEDAD_1

Comentario

Según la salida (Figura 27) obsérvese el coeficiente de determinación presenta un valor 0.1685 en porcentaje 16.85, la misma se refiere a la variación explicada de las variables superficie, semilla, plaga e incidencia de enfermedad sobre la variación del rendimiento.

El análisis de varianza, presento una probabilidad calculada menor a 0.01 por lo cual se concluye que el ajuste del modelo debido a las regresiones es significativa, es decir algunas variables tienen su incidencia positiva, negativa o algunas no la tienen.

Observando los parámetros (coeficientes de regresión) a nivel poblacional la variable semilla es significativa a P: 0.01, deduciéndose su coeficiente de regresión es diferente de cero y positivo; es decir por unidad de semilla se incrementa el rendimiento a nivel poblacional según los valores de los límites de confianza.



- **Tablas de contingencia**

La tabla de contingencia analiza la distribución de frecuencia conjunta de dos variables cualitativas y en base la distribución de Chi cuadrado se determina la relación de ambas variables y viene con la siguiente formula:

$$\chi^2 = \sum_{i \rightarrow 1}^f \sum_{j \rightarrow 1}^c \frac{(\theta_{ij} - \xi_{ij})^2}{\xi_{ij}}$$

donde:

θ_{ij} : Frecuencia observada en la i-ésima fila, j-ésima columna

ξ_{ij} : Frecuencia esperada en la i-ésima fila, j-ésima columna (total marginal de fila * total marginal de columna y dividida por el gran total)

f : número de filas

c : número de columnas

gl = (f - 1)(c - 1).

En Excel

Del archivo “Datos de papa evaluado 2017-2018 ejercicio.xlsx Hoja: Datos_papa” generar una tabla dinámica con las variables Variedad vs Aplicación agroquímico (P23_AGROQ) (OAP-MDRyT, 2018). Así mismo, incluir las frecuencias esperadas.

	A	B	C	D
3	Cuenta de P23_AG P23_AGROQ			
4	Variedad	1(Si)	2(No)	Total
5	1(Waycha)	209	71	280
6	2(Imilla)	22	17	39
7	3(Sani)	21	20	41
8	4(Desire)	158	6	164
9	5(papa nativa)	8	5	13
10	6(otro)	38	11	49
11	Total	456	130	586
12				
13		P23_AGROQ		
14	Variedad	1(Si)	2(No)	
15	1(Waycha)	217.883959	=D5*\$C\$11/\$D\$11	
16	2(Imilla)	30.34812287	8.651877133	
17	3(Sani)	31.90443686	9.09556314	
18	4(Desire)	127.6177474	36.38225256	
19	5(papa nativa)	10.11604096	2.883959044	
20	6(otro)	38.12969283	10.87030717	

Figura 28. Distribución de frecuencias observadas y esperadas de la variable variedad vs aplicación agroquímico.

Del **Menú principal** seleccione **Fórmulas** e **Insertar función**. En la ventana **Insertar función** seleccionar la categoría **Estadísticas**, seleccionar la función **PRUEBA.CHICUAD** y presione **Aceptar**. Esta función



devuelve la probabilidad de la distribución de Chi cuadrado. En la ventana **Argumento de función**, remarcar las **frecuencias observadas** en el Campo **Rango_real**, remarcar las **frecuencias esperadas** en el campo **Rango_esperado** y presionar **Aceptar**.

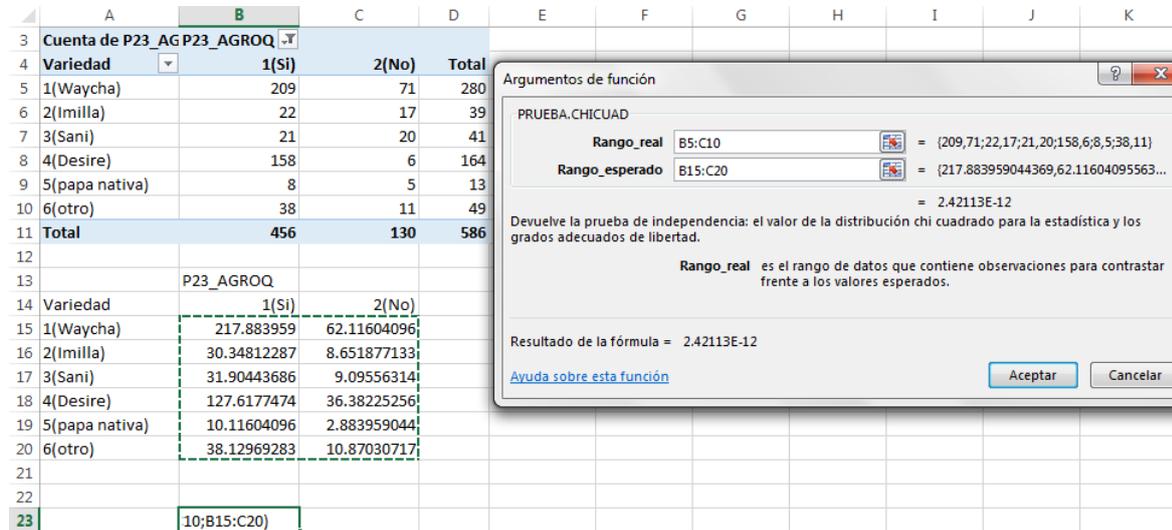


Figura 29. Ventana Argumento de función PRUEBA.CHICUAD con las tablas de distribución de frecuencias observadas y esperadas.

Variedad	1(Si)	2(No)	Total	Variedad	1(Si)	2(No)	Total
1(Waycha)	209	71	280	1(Waycha)	74.64%	25.36%	100.00%
2(Imilla)	22	17	39	2(Imilla)	56.41%	43.59%	100.00%
3(Sani)	21	20	41	3(Sani)	51.22%	48.78%	100.00%
4(Desire)	158	6	164	4(Desire)	96.34%	3.66%	100.00%
5(papa nativa)	8	5	13	5(papa nativ)	61.54%	38.46%	100.00%
6(otro)	38	11	49	6(otro)	77.55%	22.45%	100.00%
Total	456	130	586	Total	77.82%	22.18%	100.00%

Figura 30. Distribución de frecuencias y en porcentaje por variedad.

Comentario

Obsérvese el valor de la probabilidad de independencia es 2.42113E-12, este valor es menor a 0.01 por lo cual la aplicación del agroquímico tiene dependencia de las variedades de papa, es decir que la aplicación de agroquímicos en algunas variedades es menor y en otras es mayor, por ejemplo la Desire fue declarada con mayor frecuencia con aplicación de agroquímicos, seguida por Waycha y Otros frente a la Imilla, Sani y Papa nativa.



5. MUESTREO

5.1. Determinación del tamaño de la muestra

Dentro de un diseño muestral, ayuda bastante conocer información de trabajos anteriores de las variables con parámetros o estadísticos. Con esta información se puede estimar el tamaño de la muestra (n) en base a :

Población infinita	Población finita
$n = \frac{t^2 S^2}{d^2}$	$n = \frac{t^2 S^2}{d^2} \cdot \frac{1}{1 + \frac{1}{N} \left(\frac{t^2 S^2}{d^2} - 1 \right)}$

donde:

t: Valor de confiabilidad de distribución t de Student a la probabilidad 0.05 o 0.01 con grados de libertad n-1. Si n es mayor a 30 usar el valor de la distribución normal (Z).

S^2 : Varianza

d^2 : Diferencia entre la media poblacional y de la muestra

N: Tamaño de la población

Estadístico	Datos cualitativos	Datos cuantitativos
Media	$p = \frac{\text{Positivos}}{n}$	\bar{x}
Varianza	$p(1-p)$	$\frac{\sum (x_i - \bar{x})^2}{n-1}$

En Excel

Con los datos del archivo “Datos de papa evaluado 2017-2018 ejercicio.xlsx Hoja: Rendim y otras” (OAP-MDRyT, 2018), de ella extraemos utilizando Tabla dinámica la estadística descriptiva del rendimiento campaña 2017-2018. Posteriormente abrir el archivo “**tamaño muestra.xlsx**” e introducir los valores de **Desviación estándar** y el **Error permisible** o diferencia entre la media muestral y la media poblacional.



	A	B	C	D	E	F	G	H	I
	FOLIO	rendimiento_17_18	superficie_17_18	semilla_toneladas	P20_PLAGA_1	P20_ENFERMEDAD_1		N Lista	Folio Aleatorio
5	1	6.048387097	0.155	0.890322581	12	98		1	
6	2	6.75	0.5	0.483	12	98		2	
7	3	0.002130682	0.176	0.000196023	12	98		3	
8	4	8.928571429	0.105	2.957142857	99	98		4	
9	5	10.27930403	0.03775	2.422161172	12	98		5	
10	6	3.511235955	0.178	0.646067416	12	98		6	
11	7	4.189944134	0.179	1.156424581	99	98		7	
12	8	4.326923077	0.52	1.194230769	99	98		8	
13	9	3.333333333	0.15	1.38	99	98		9	
14	10	13.8	0.5	0.414	13	28		10	
15	11	13.8	0.5	0.414	13	30		11	
16	12	2.76	0.25	0.092	13	29		12	
17	13	2.07	1	0.184	13	27		13	
18	14	2.07	1.5	0.46	13	27		14	
19	15	8.28	0.5	0.138	13	27		15	
20	16	5	0.25	0.644	13	30		16	
21	17	4.83	0					17	

Muestra

Entrada

Rango de entrada:

Rótulos

Método de muestreo

Periódico

Período:

Aleatorio

Número de muestras:

Opciones de salida

Rango de salida:

En una hoja nueva:

En un libro nuevo

Figura 32. Vista de la opción Muestra para una selección aleatoria.

5.3. Muestreo aleatorio sistemático

A diferencia del Muestreo aleatorio simple, se selecciona el primer elemento en entre el 1 y el valor del factor de elevación. Considere 442 a la lista donde debe seleccionarse 147:

$$fe = \frac{N}{n} = \frac{442}{147} = 3.00680272 \approx 3$$

Seleccionar al azar un número entre 1 a 3 [en Excel la función es =ALEATORIO.ENTRE(inferior; superior)], este valor corresponde al primer elemento seleccionado, a éste se adiciona el factor de elevación (3) para el segundo elemento, para el tercer se vuelve adicionar el factor de elevación, este procedimiento se continua hasta completar el tamaño de la muestra.

	A	B	C	D	E	F	G	H	I	J
	FOLIO	rendimiento_17_18	superficie_17_18	semilla_toneladas	P20_PLAGA	P20_ENFERMEDAD_1		N Lista	Folio Aleatorio	Folio Sistemático
5	1	6.048387097	0.155	0.890322581	12	98		1	448	1
6	2	6.75	0.5	0.483	12	98		2	164	4
7	3	0.002130682	0.176	0.000196023	12	98		3	425	7
8	4	8.928571429	0.105	2.957142857	99	98		4	362	10
9	5	10.27930403	0.03775	2.422161172	12	98		5	16	13
10	6	3.511235955	0.178	0.646067416	12	98		6	438	16
11	7	4.189944134	0.179	1.156424581	99	98		7	276	19
12	8	4.326923077	0.52	1.194230769	99	98		8	375	22
13	9	3.333333333	0.15	1.38	99	98		9	215	25
14	10	13.8	0.5	0.414	13	28		10	133	28
15	11	13.8	0.5	0.414	13	30		11	420	31
16	12	2.76	0.25	0.092	13	29		12	488	34
17	13	2.07	1	0.184	13	27		13	292	37
18	14	2.07	1.5	0.46	13	27		14	162	40
19	15	8.28	0.5	0.138	13	27		15	10	43
20	16	5	0.25	0.644	13	30		16	2	46

Figura 33. Vista de los elementos seleccionados en forma aleatoria y sistemática.

La ventaja del muestreo sistemático, está en que los elementos de la población son seleccionados en forma bien distribuida.



5.4. Algunas consideraciones para el diseño de un muestreo

Las etapas del levantamiento de información con alguna técnica de muestreo según Cochran W. (1974) y Peña y Romo (1998) son:

- **Objetivo**
Una clara exposición de los objetivos es lo más útil (implica alcance del trabajo y variables)
- **Definición población**
Es un conjunto de elementos homogéneos respecto a una variable que se desea estudiar.
- **Nivel de precisión**
Medida del error o desvío con que se desea aproximarse con la muestra a la población
- **Métodos de medición**
Procedimiento a emplear para levantar los datos ya sea en aspecto social, económico, biológico (fitotecnia, zootecnia, etc) y otras.
- **Definición de unidad muestral**
Considerada como la unidad básica de muestreo y es un elemento de la población
- **Estimación del tamaño muestra**
Se requiere información preliminar de la variable como la media, varianza y número de observaciones. Esta información puede ser de trabajos anteriores o de un **muestreo piloto** a realizarse.
- **Técnica muestreo**
La técnica de muestreo puede ser definida en base a la dimensión del objetivo y las características de la población y se mencionan a:
 - Muestreo aleatorio simple,
 - Muestreo aleatorio sistemático,
 - Muestreo aleatorio estratificado,
 - Muestreo aleatorio conglomerado, etc
- **Trabajo de campo**
Se considera importante la planificación (detalles operativos) para ejecutar el levantamiento de los datos
- **Resumen y análisis de los datos**
La información levantada puede procesarse empleando estadísticas en base los requerimientos de los objetivos, para ello suele utilizarse programas estadístico o planillas electrónicas.



BIBLIOGRAFÍA

- COCHRAN W. 1974. Técnicas de Muestreo. Compañía Editorial Continental S. A. México. 507p.
- GUJARATI, D. 2002. Econometría. Trad. de tercera edición BASIC ECONOMETRICS. McGraw-Hill. Colombia pp 100-101.
- MICROSOFT OFFICE. 2013. Programa Microsoft Excel 2013.
- PEÑA Y ROMO. 1998. Introducción a la estadística para ciencias sociales. McGRAW-HILL. España. 428p.
- SEPA-PASA-MDRyT, 2011. Proyecto: Apoyo al fortalecimiento de la producción de papa de pequeños productores en los Departamentos de La Paz, Cochabamba, Potosí, Oruro y Santa Cruz. SEPA; Unidad de Producción de Semilla de Papa, PASA; Programa de Apoyo a la Seguridad Alimentaria, Ministerio de Desarrollo Rural y Tierras. Informe elaborado por Edwin Iquize, Rubén Campos, Miguel Osina, Pedro Pally, Nestor Yave, Modesto Felipe, Rodolfo Zambrana, Grover Iporre, Eddy Aracena, Herlan Sandoval, Elio Duran. Cochabamba Bolivia.
- SPIEGEL Y STEPHENS. 2009. Estadística. McGraw-Hill. Cuarta edición.
- STEEL, TORRIE AND DICKEY. 1992. Bioestadística: Principios y procedimientos. McGraw-Hill.