

ANÁLISIS DE DATOS CON INFOSTAT

Introducción.

El material que se presenta a continuación contiene resultados del análisis estadístico de datos realizados con el software Infostat, empleando los procedimientos desarrollados en la asignatura.

Su objetivo es complementar las aplicaciones de las diferentes herramientas estadísticas desarrolladas en la asignatura, con la finalidad de familiarizar al alumno en la lectura e interpretación de salidas de software estadístico cuyo manejo es fundamental en esta disciplina.

Para facilitar la presentación del tema se ha tomado como referencia un problema ingenieril para el cual se plantean diferentes estrategias de análisis estadístico, se muestran las correspondientes salidas de Infostat y se realiza un breve análisis de los resultados obtenidos.

Problema Propuesto.

En una fábrica autopartista se debe controlar el proceso de producción de una pieza mecanizada por un torno, que últimamente ha registrado muchos reclamos por defectos de fabricación.

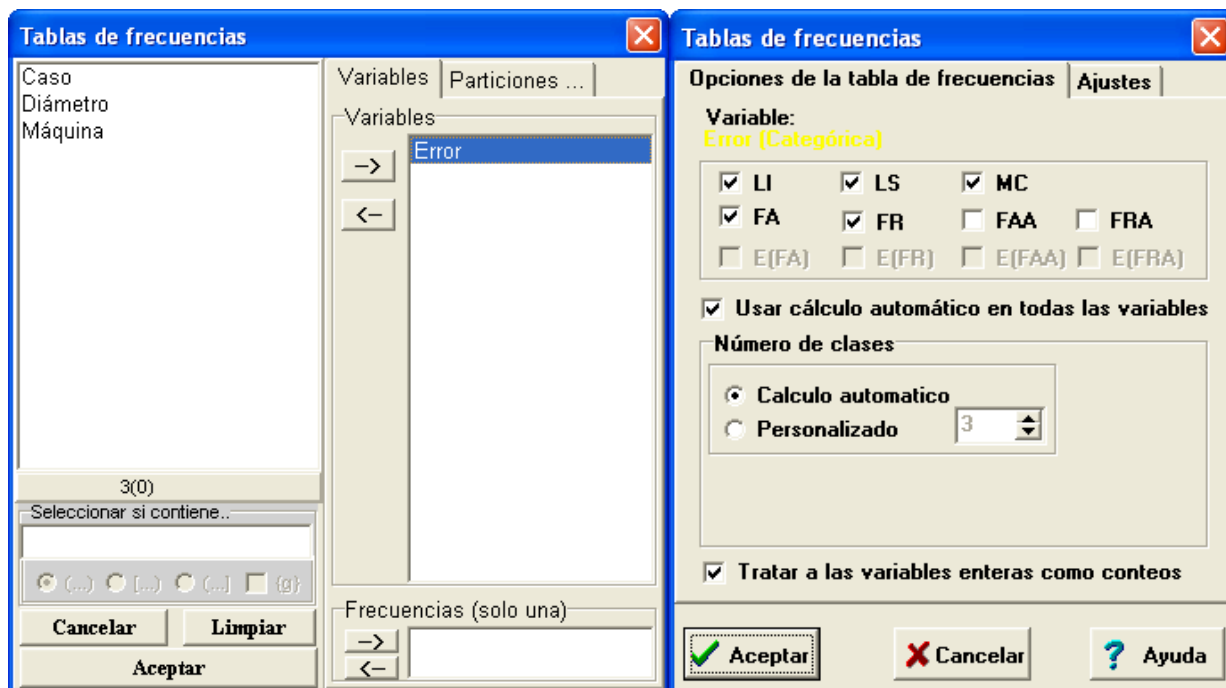
Se conforma un grupo de trabajo con la misión de resolver el problema. Este grupo debe reunir toda la información posible relacionada con el tema; se plantea como primera actividad explorar las causas de los reclamos. Buscando en la documentación del área de calidad, el grupo pudo identificar reclamos relacionados con defectos encontrados en:

Diámetro, Longitud, Rugosidad, Embalaje, Causas menores agrupadas como “otros”.

Realizado un relevamiento sobre una muestra de 50 entregas, se pudo identificar los siguientes defectos que ocasionaron reclamos:

Diámetro	Longitud	Diámetro	Longitud	Longitud
Diámetro	Longitud	Diámetro	Longitud	Rugosidad
Diámetro	Diámetro	Diámetro	Diámetro	Rugosidad
Rugosidad	Embalaje	Embalaje	Embalaje	Otros
Diámetro	Diámetro	Diámetro	Diámetro	Rugosidad
Longitud	Longitud	Longitud	Longitud	Diámetro
Diámetro	Diámetro	Diámetro	Diámetro	Diámetro
Longitud	Longitud	Rugosidad	Otros	Diámetro
Longitud	Longitud	Longitud	Diámetro	Diámetro
Diámetro	Diámetro	Diámetro	Diámetro	Diámetro

Con esos datos se elaboraron una tabla de frecuencias y un diagrama de Pareto.



Distribución de frecuencias

Variable	Clase	Categorías	FA	FR
Error	1	Diámetro	26	0,52
Error	2	Embalaje	3	0,06
Error	3	Longitud	14	0,28
Error	4	Otros	2	0,04
Error	5	Rugosidad	5	0,10

Ordenando por frecuencias decrecientes se obtiene

Error	FA	FR
Diámetro	26	0,52
Longitud	14	0,28
Rugosidad	5	0,10
Embalaje	3	0,06
Otros	2	0,04

Agrupando los reclamos de acuerdo al defecto que originó el mismo y teniendo en cuenta la información de la tabla precedente, se puede apreciar que los reclamos más frecuentes están relacionados con el diámetro de las piezas, ya que constituye el 52 % de los mismos.

Por otra parte considerando el diámetro y la longitud de las piezas, se observa que entre ambos suman el 80 % de los reclamos, es decir que estas dos causas son responsables de la mayor parte de los reclamos: **Se concluye que la mayoría de los problemas están relacionados con características dimensionales de las piezas.**

El análisis del Diagrama de Pareto, que se muestra a continuación, confirma las conclusiones obtenidas en base a la tabla, por lo que el Grupo de Trabajo decide continuar con el análisis particular de las características dimensionales de las piezas en una primera etapa, dejando el tratamiento de los problemas menos frecuentes para más adelante.

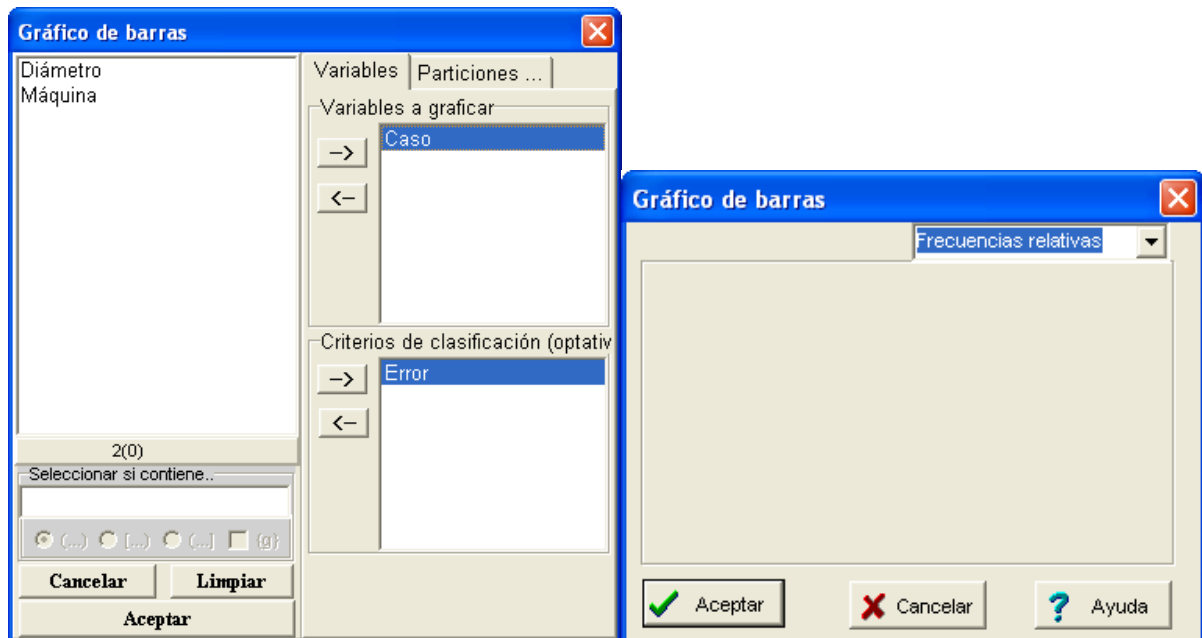
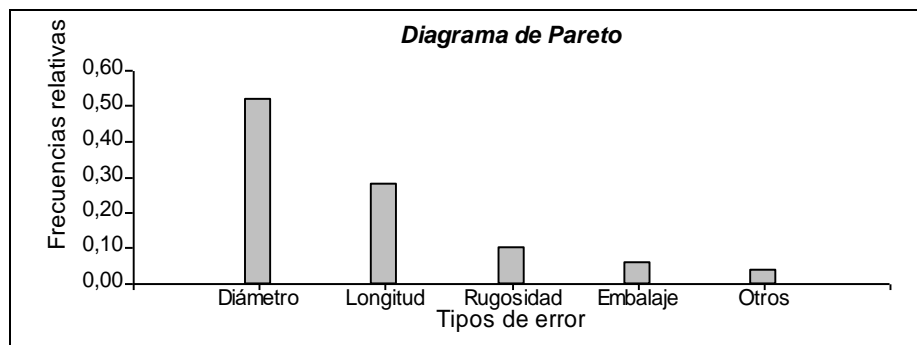


Diagrama de Pareto



En primer lugar se estudia el comportamiento de la siguiente variable:

X: diámetro de la pieza.

Las especificaciones de ingeniería de la característica son las siguientes: 20 mm. \pm 1 mm. Las piezas que estén fuera de ese rango deben ser rechazadas (retrabajadas o descartadas). Para controlar el proceso de fabricación de la pieza, se extrajo una muestra de 125 piezas producidas consecutivamente, que arrojó los siguientes valores:

Diámetro	Máquina
19,81	1
21,87	1
20,78	1
20,50	1
21,06	1
20,83	1
20,98	1
21,67	1
21,47	1
19,88	1
21,23	1
20,94	1
20,00	1
20,68	1
21,14	1
20,58	1
20,24	1
19,78	1
20,78	1
20,34	1
19,91	1
20,61	1
19,98	1
19,27	1
20,54	1

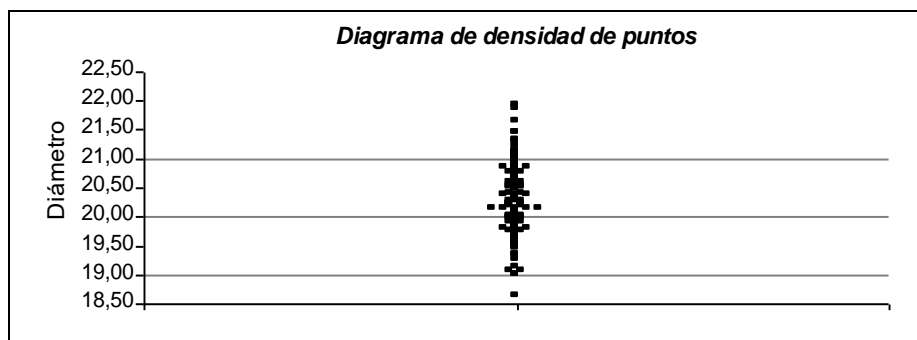
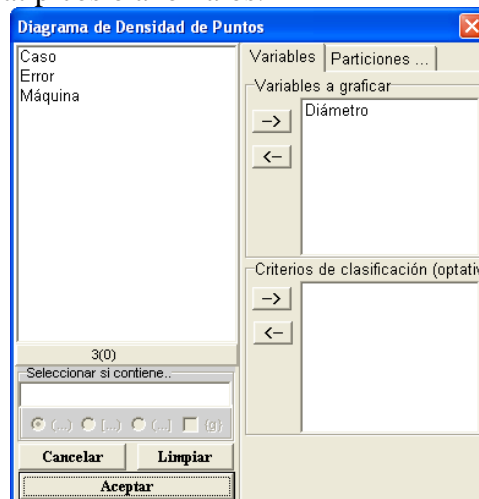
Diámetro	Máquina
19,35	1
20,73	1
20,15	1
20,63	1
20,01	1
20,19	1
21,05	1
19,60	1
20,28	1
20,88	1
20,04	1
20,10	1
20,29	1
19,01	1
20,25	1
20,81	1
20,15	1
19,47	1
19,85	1
19,83	1
19,70	1
20,86	1
19,74	1
19,75	1
20,31	1

Diámetro	Máquina
19,81	2
19,64	2
19,95	2
21,13	2
20,71	2
19,81	2
19,82	2
20,35	2
20,62	2
20,75	2
20,74	2
20,79	2
20,42	2
19,09	2
21,12	2
19,92	2
19,98	2
20,21	2
21,32	2
20,15	2
20,39	2
20,14	2
20,40	2
20,42	2
20,53	2

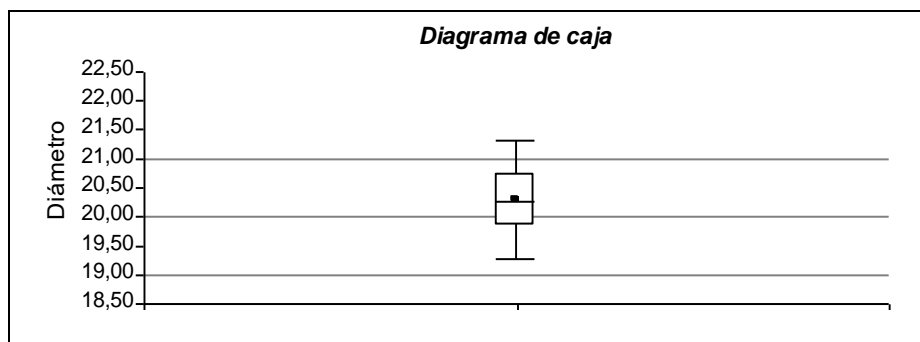
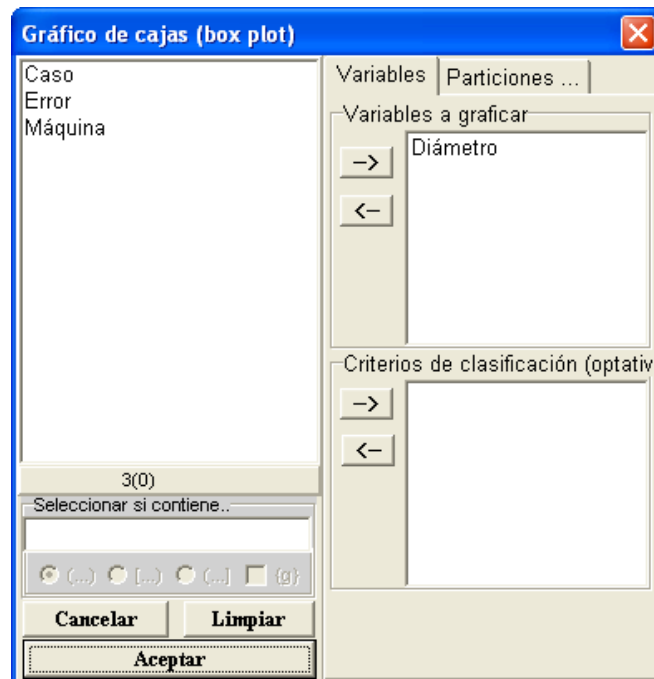
Diámetro	Máquina
20,64	2
19,53	2
20,87	2
20,53	2
20,21	2
20,05	2
18,66	2
20,25	2
19,76	2
20,96	2
20,01	2
20,29	2
19,09	2
20,15	2
20,54	2
19,86	2
19,76	2
21,08	2
20,38	2
20,11	2
21,94	2
19,59	2
21,46	2
19,61	2
20,61	2

Diámetro	Máquina
19,37	2
20,72	2
20,86	2
21,33	2
20,03	2
20,15	2
20,19	2
20,86	2
19,56	2
20,39	2
19,77	2
20,51	2
20,39	2
19,92	2
20,24	2
21,07	2
20,03	2
19,14	2
21,10	2
21,00	2
20,49	2
19,49	2
19,69	2
19,62	2
19,02	2

Como primera medida se realiza el Diagrama de Puntos y Diagrama de Caja, con el fin de detectar la presencia de datos atípicos o anómalos.

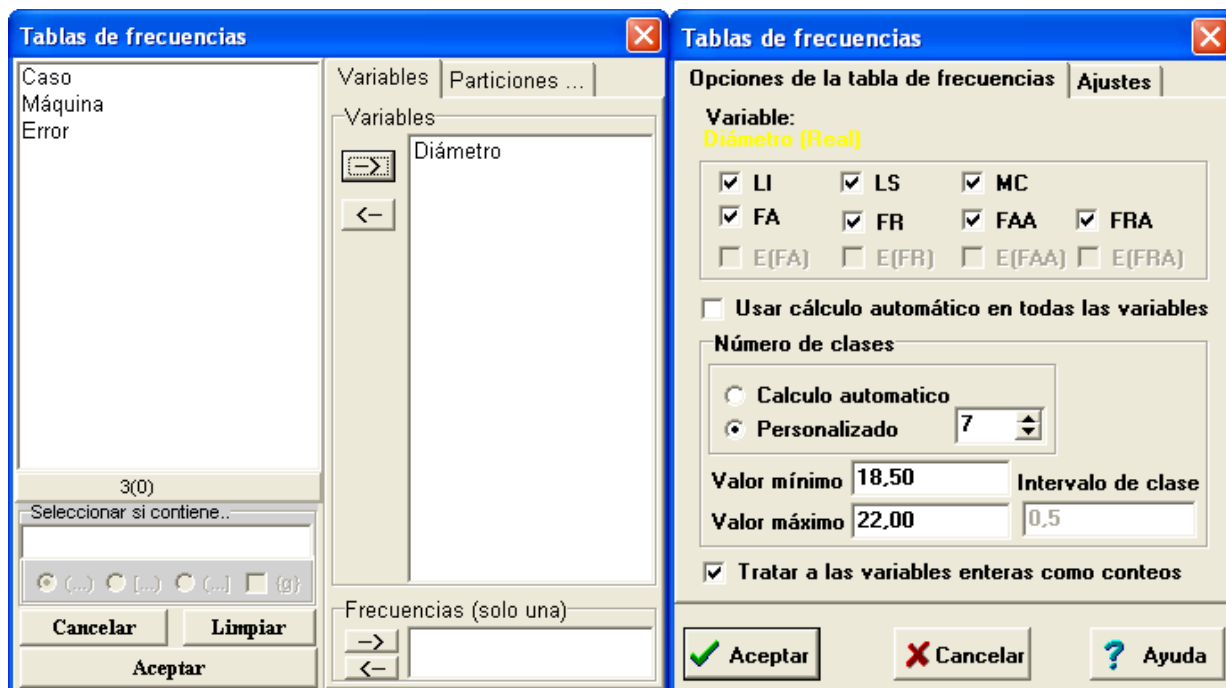


- El diagrama de puntos permite observar que la mayor concentración de datos se produce levemente por encima de los 20 mm.



- En el diagrama de caja se observa que la media aritmética prácticamente coincide con la mediana, y que los brazos del diagrama son bastante parecidos. Esto mostraría la simetría de la distribución.
- No se observan datos atípicos.

Luego, continuando con el análisis se elabora una tabla de frecuencias distribuyendo los datos en 7 intervalos de clase.



recordando que:

- *LI y LS representan respectivamente los Límites Inferior y Superior del Intervalo de clase.*
- *MC es la Marca de Clase o punto medio del intervalo.*
- *FA es la Frecuencia Absoluta, FR es la Frecuencia Relativa, FAA la Frecuencia Absoluta Acumulada y FRA la Frecuencia Relativa Acumulada.*

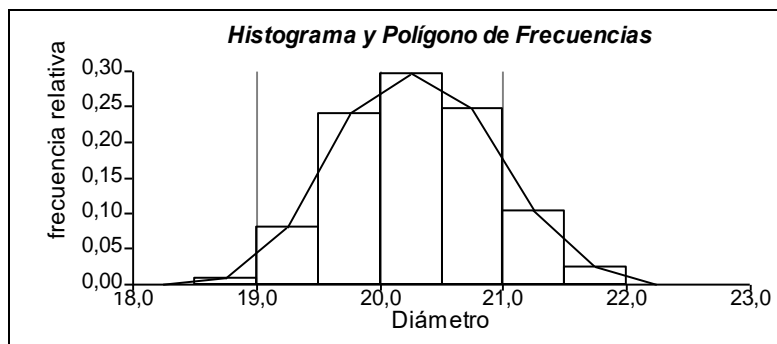
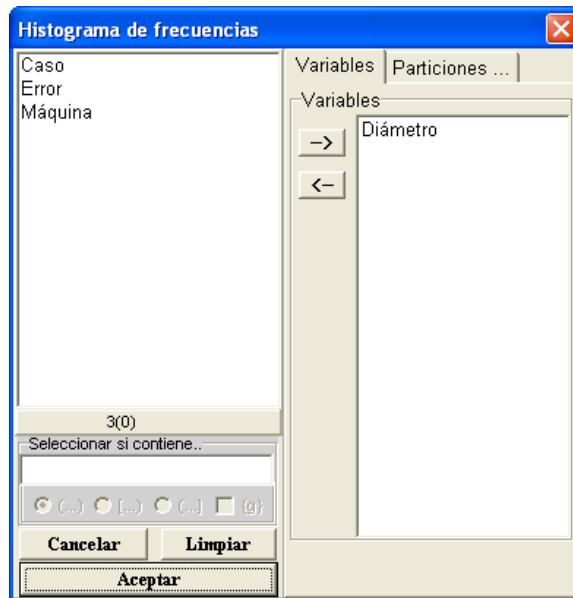
Se tiene

Variable	Clase	LI	LS	MC	FA	FR	FAA	FRA
Diámetro	1	18,50	19,00	18,75	1	0,01	1	0,01
Diámetro	2	19,00	19,50	19,25	10	0,08	11	0,09
Diámetro	3	19,50	20,00	19,75	30	0,24	41	0,33
Diámetro	4	20,00	20,50	20,25	37	0,30	78	0,62
Diámetro	5	20,50	21,00	20,75	31	0,25	109	0,87
Diámetro	6	21,00	21,50	21,25	13	0,10	122	0,98
Diámetro	7	21,50	22,00	21,75	3	0,02	125	1,00

Observando la tabla de frecuencias, se tiene que:

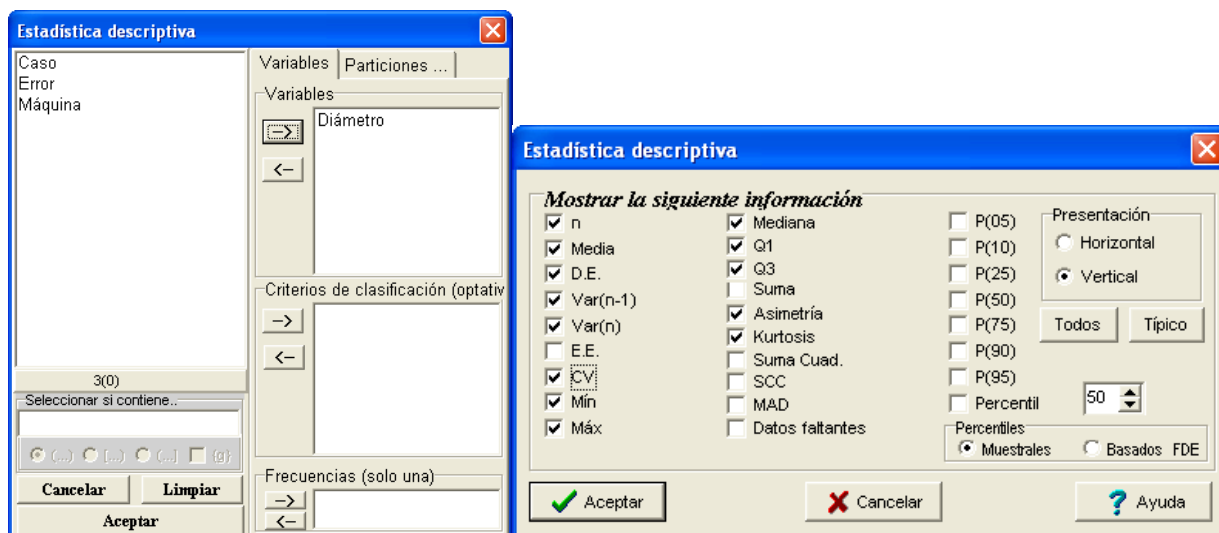
- *El intervalo modal (con mayor frecuencia) en este caso es el que va de 20 a 20,5 mm., e incluye al 30 % de los datos.*
- *Hay 17 piezas (un 13 %) con el diámetro fuera del intervalo de tolerancia de la característica (que va de 19 a 21 mm.).*

Además se dibujan algunos diagramas útiles para analizar la distribución de la variable: Histograma, Polígono de Frecuencias. Además, en el mismo gráfico se incluyen el valor nominal y las especificaciones de la característica controlada



- Tanto el histograma como el polígono de frecuencias muestran que la distribución de la variable es bastante simétrica.
- Se observa que la mayor concentración de observaciones se localiza en el intervalo de 20 a 20,5 mm.
- Teniendo en cuenta las especificaciones de la característica estudiada ($20 \text{ mm} \pm 1 \text{ mm}$) se puede apreciar que la distribución excede dichas especificación -marcados en el gráfico-, y no se encuentra correctamente centrada dentro de los mismos (esta desplazada hacia la derecha).

Finalmente se calculan las medidas descriptivas.



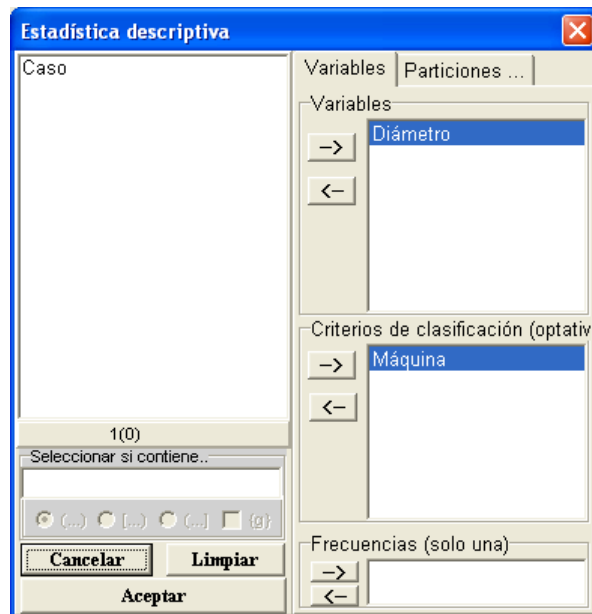
Estadística descriptiva

Resumen	Diámetro
n	125,000
Media	20,296
D.E.	0,624
Var (n-1)	0,389
Var (n)	0,386
CV	3,074
Mín	18,656
Máx	21,941
Mediana	20,255
Q1	19,852
Q3	20,744
Asimetría	0,075
Kurtosis	-0,046

- La media y la mediana son muy parecidas.
- El coeficiente de asimetría es muy bajo en valor absoluto (aunque su signo indica asimetría positiva con sesgo a la derecha) esto confirma lo señalado en los análisis presedentes, la distribución de la variable es prácticamente simétrica.
- El coeficiente de Kurtosis muy cercano a cero, esto indica que la distribución de la variable tiene un grado de apuntamiento similar a una Normal.
- La varianza común es prácticamente igual a la corregida, porque la muestra considerada es de tamaño considerable.
- El coeficiente de variación indica que el desvío es un 3 % de la media.

Uno de los integrantes del grupo de trabajo remarca que las piezas provienen de dos máquinas diferentes, por lo que se propone estratificar los datos por máquina para ver si existen diferencias en sus resultados.

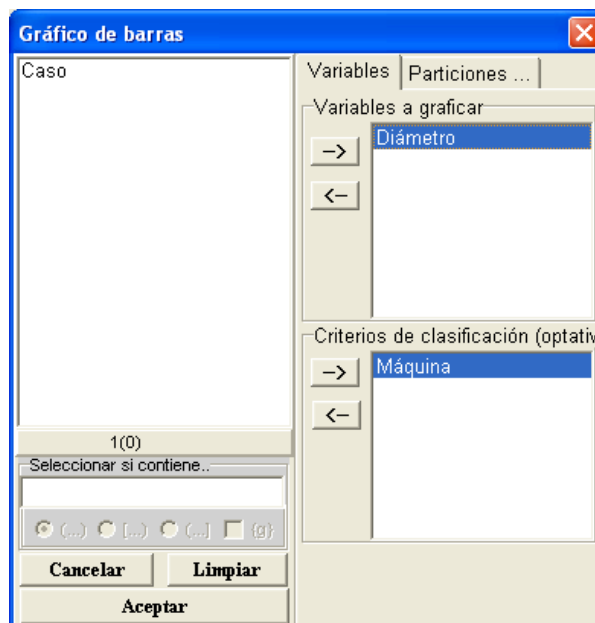
Para obtener la estadística descriptiva, estratificando por máquina se realiza

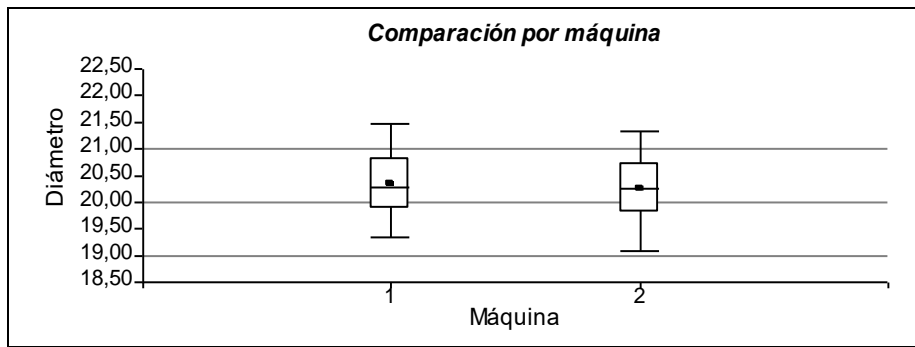


Estadística descriptiva

Máquina	Variable	n	Media	D.E.	CV	Mín	Máx	Asimetría
1	Diámetro	50	20,36	0,62	3,06	19,01	21,87	0,21
2	Diámetro	75	20,25	0,63	3,09	18,66	21,94	-4,6E-03

Asimismo, para realizar el Gráfico de Caja pero estratificando por Máquina, procedemos

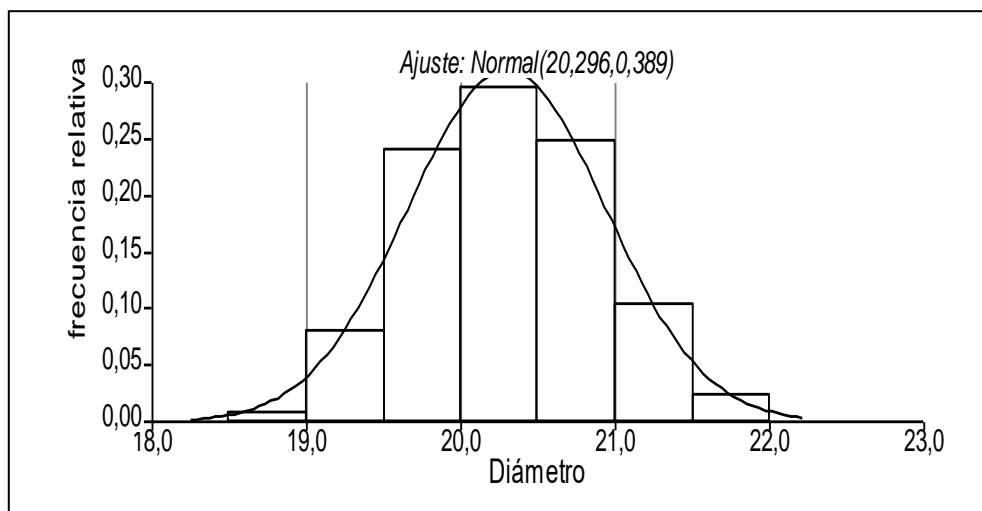




Como se puede apreciar tanto gráfica como analíticamente, las principales propiedades de la variable son muy parecidas en ambos casos (aunque la distribución de los datos de la Máquina 1 está ubicada levemente por encima de la 2), por lo que es razonable suponer que los resultados de ambas máquinas no difieren considerablemente en términos prácticos.

Teniendo en cuenta las propiedades de la variable analizada anteriormente, se decide emplear la distribución Normal para modelar el comportamiento de la variable. (OBSERVACIÓN: Mas adelante analizaremos con herramientas de inferencia estadística si la decisión fue la adecuada)

En base a los resultados de la muestra se ajusta el modelo Normal que describe la distribución de probabilidades de la variable en estudio.



Empleando la distribución Normal, con los parámetros estimados en base a la muestra anterior, se calcula con Infostat la probabilidad de que una pieza esté fuera de especificación.

$$P(X < 19) = 0,01885801534$$

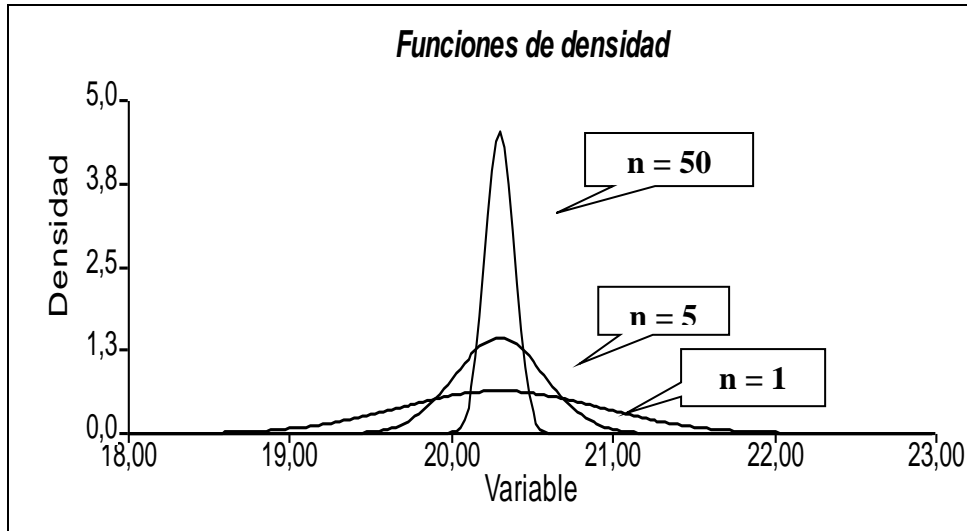
$$P(X > 20) = 0,1295018017$$

TOTAL = 0,14835981704 Esta probabilidad es bastante parecida a la proporción de piezas observadas fuera de tolerancia (que era igual a 0,13).

Conociendo la probabilidad de que una pieza sea rechazada, y sabiendo que los despachos de esta pieza se realizan en cajones que contienen 16 piezas, el grupo desea calcular la probabilidad de que un cajón cualquiera contenga al menos una pieza defectuosa. Empleando el modelo Binomial con $n=16$ y $p=0,1484$, el Infostat permite calcular la siguiente probabilidad:

$$P(X \geq 1) = 0,9234807959$$

Para monitorear el funcionamiento del proceso, se puede controlar el diámetro de una pieza tomada al azar de la producción con una determinada frecuencia, o se puede tomar una muestra de varias piezas y controlar el diámetro medio. A fin de evaluar la precisión de cada alternativa, se compara la distribución de probabilidades de la variable original con la distribución de la media muestral para muestras de tamaño 5 y 50.



Como se puede apreciar, la precisión de la estimación es más grande cuando se trabaja con promedios que cuando se trabaja con observaciones individuales, y a su vez es mayor a medida que aumenta el tamaño de la muestra (la distribución tiene menor dispersión en torno a su media).

Para realizar inferencias sobre el diámetro, el grupo debe comprobar en primer lugar si es razonable suponer que la variable tiene distribución Normal. Para esto se implementan las pruebas de bondad de ajuste Chi-Cuadrado y de Kolmogorov, empleando los datos de la muestra inicial.

A continuación se ilustrará el modo de realizar diversas pruebas de hipótesis utilizando el software. En ellas, no se han planteado las Hipótesis Nula y Alternativa ya que el objetivo en esta parte de la Guía es mostrar cómo se realizan estas pruebas con Infostat. En cada caso, usted deberá plantear las hipótesis correspondientes a cada prueba. Por lo tanto:

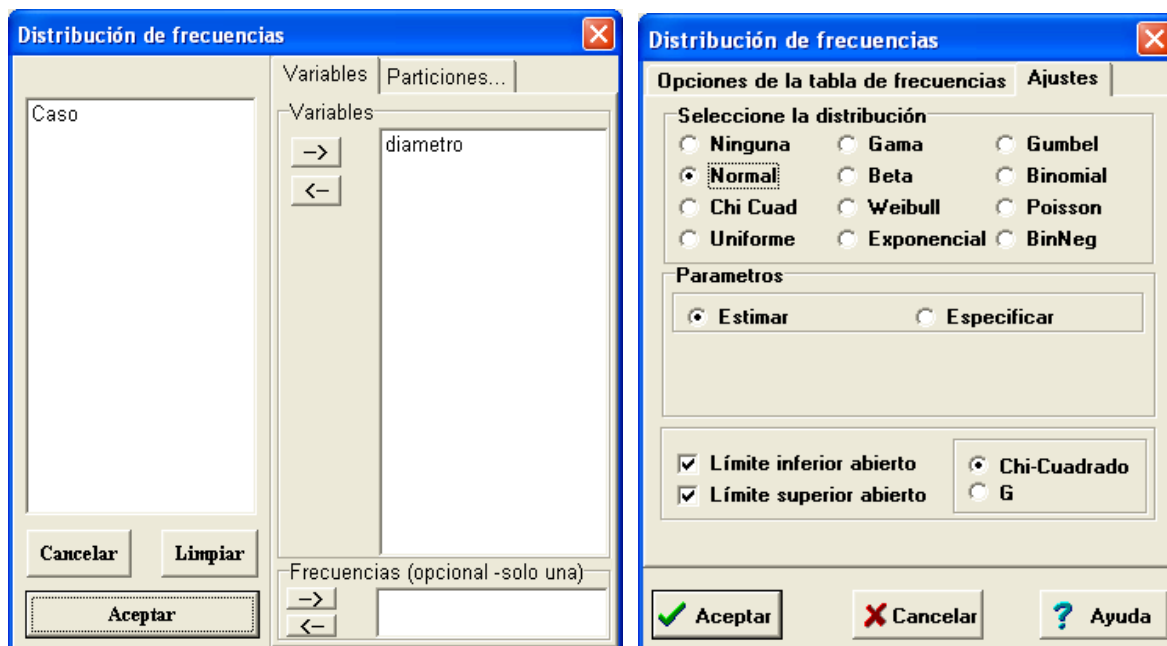
Recuerde plantear SIEMPRE la Hipótesis Nula y la Alternativa, en función de la variable estudiada y del problema planeado!

Volviendo al problema en estudio:

X: diámetro de la pieza

Utilizando Infostat en este ejemplo, se tiene:

1. Si se elige la Prueba de Bondad de Ajuste



Ajuste: Normal con estimación de parámetros: Media= 20,29574 y varianza= 0,38933

Variable	Clase	LI	LS	MC	FA	FR	E(FA)	E(FR)	Chi-Cuadrado	p
Diámetro	1	18,65	19,20	18,93	6	0,05	4,94	0,04	0,23	
Diámetro	2	19,20	19,75	19,48	15	0,12	18,92	0,15	1,04	
Diámetro	3	19,75	20,30	20,03	45	0,36	38,98	0,31	1,97	
Diámetro	4	20,30	20,85	20,58	34	0,27	38,76	0,31	2,55	
Diámetro	5	20,85	21,40	21,13	20	0,16	18,60	0,15	2,66	
Diámetro	6	21,40	21,95	21,68	5	0,04	4,80	0,04	2,67	0,4459

Teniendo en cuenta esta salida tenemos que: $\mu = 20,29574$ y $\sigma^2 = 0,38933$ y estamos planteando

H_0 : X tiene distribución Normal con media 20,29574 y varianza 0,38933.

H_1 : X no tiene distribución Normal con media 20,29574 y varianza 0,38933.

En la salida se vislumbra que:

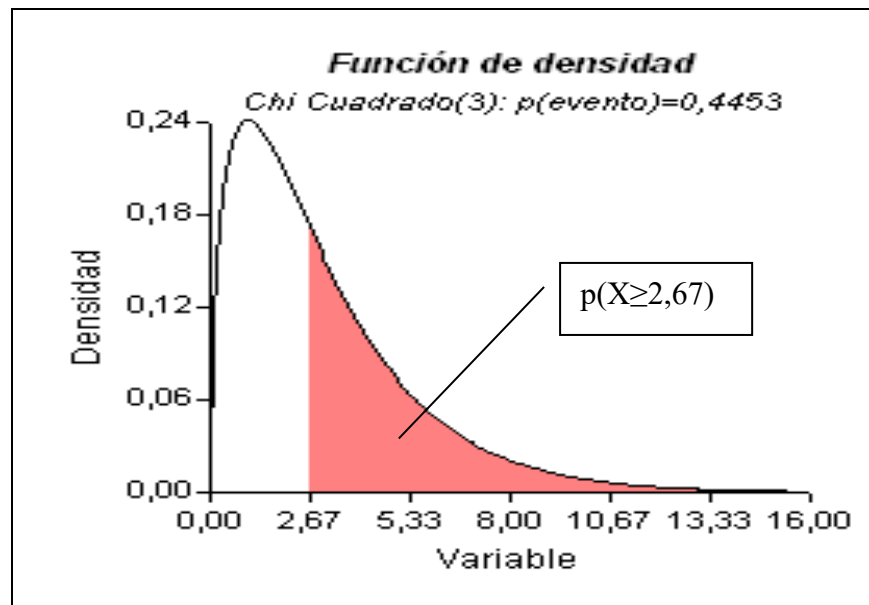
- LI y LS son los Límites Inferior y Superior de los intervalos de clase en que se agruparon las observaciones individuales, y MC es la Marca de Clase de cada intervalo.
- FA y FR son las Frecuencias Absolutas y Relativas “observadas”.
- E(FA) y E(FR) son las Frecuencias Absolutas y Relativas “esperadas”, calculadas en base al modelo propuesto (suponiendo H_0 verdadera).

- La columna Chi-Cuadrado va calculando y acumulando $\frac{(f_{obs} - f_{est})^2}{f_{est}}$, valores necesarios

para obtener una estimación del estadístico de prueba $\chi^2 = \sum_i \frac{(f_{obs_i} - f_{est_i})^2}{f_{est_i}}$, que en este caso

vale 2,67 (valor reportado en la última fila). En este caso los grados de libertad son 3 (pues $k = 6$, $r = 2$ cantidad de parámetros estimados).

- la última columna (p) reporta el *valor p*: que es la probabilidad de observar un valor mayor o igual al valor del estadístico de prueba estimado dado que la hipótesis nula es verdadera. Gráficamente



- Al comparar el valor p con el nivel de significación elegido permite concluir acerca del resultado de la prueba (rechazar o no la hipótesis nula).
 - Si el $p\text{-valor} < \alpha$ significa que la estimación del estadístico de prueba cae en zona de rechazo, por lo que se rechaza la hipótesis nula.
 - En cambio, si el $p\text{-valor} > \alpha$ no se rechaza la hipótesis nula.

En el ejemplo planteado, si suponemos que se había elegido un nivel de significación $\alpha = 0.05$ entonces para elaborar su conclusión puede

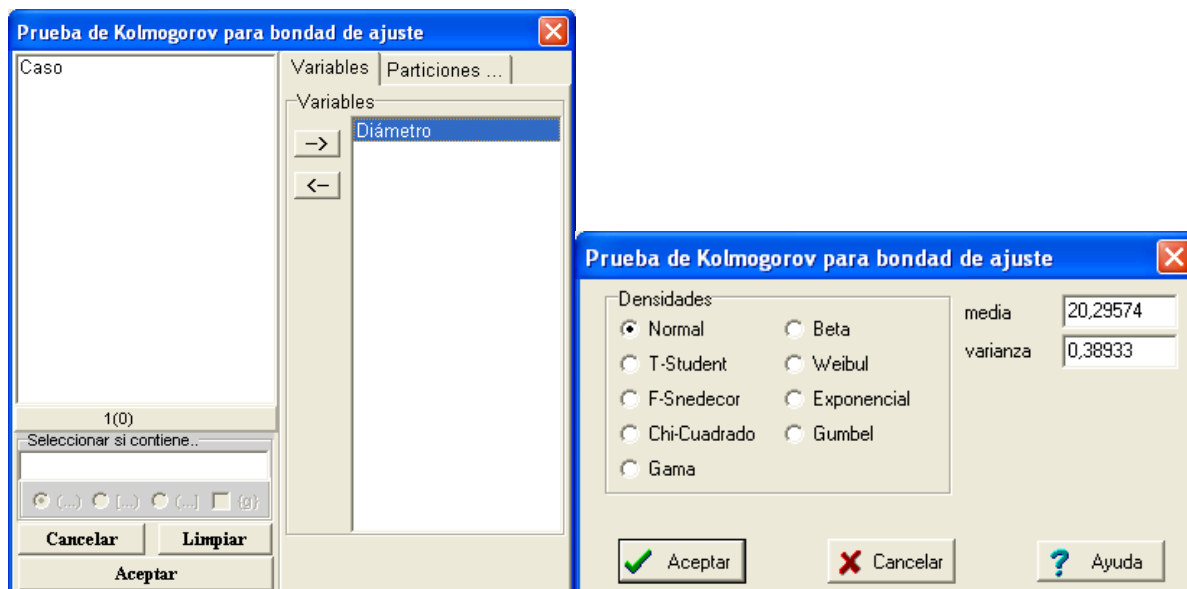
- Tener presente todas las columnas de la salida a excepción de la columna donde figura el valor p y buscar en la tabla de la distribución Chi-cuadrado el punto crítico $\chi^2_{\alpha, k-r-1}$ que en este caso sería $Y_c = \chi^2_{0,05;3} = 12,837$. Luego como $\chi^2_{obs} = 2,67$ es menor que Y_c no se rechaza H_0 . En conclusión la distribución del *diámetro de la pieza* puede modelarse con la distribución Normal con los parámetros especificados en el encabezamiento de la tabla.
- Tener presente todas las columnas de la salida. Entonces como $p(X \geq 2,67)$ es mayor al nivel elegido 0,05 se puede decir que no se rechaza H_0 y por ende que la distribución del *diámetro de la pieza* puede modelarse con la distribución Normal con los parámetros especificados en el encabezamiento de la tabla (los parámetros son estimados automáticamente a partir de la muestra en estudio).

2. Si se elige la **Prueba de Bondad de Kolmogorov-Smirnov**

Teniendo en cuenta que: $\mu = 20,29574$ y $\sigma^2 = 0,38933$ estamos planteando

H_0 : X tiene distribución Normal con media 20,29574 y varianza 0,38933.

H_1 : X no tiene distribución Normal con media 20,29574 y varianza 0,38933.



Prueba de Kolmogorov para bondad de ajuste

Variable	Ajuste	media	varianza	n	Estadístico D	p-valor
Diámetro	Normal (20,296,0,389)	20,30	0,39	125	0,03	0,9993

En la salida se vislumbra que:

- En la columna “Ajuste” se indica el modelo propuesto y sus parámetros.
- El “Estadístico D” es la mayor diferencia entre la Distribución de Frecuencias Relativas Acumuladas “observada” y “esperada” (bajo el modelo supuesto).
- El p-valor es la probabilidad asociada a dicho estadístico.

En el ejemplo planteado, si suponemos que se había elegido un nivel de significación $\alpha = 0.05$ entonces para elaborar su conclusión puede

- Tener presente todas las columnas de la salida a excepción de la columna donde figura el valor p y buscar en la tabla de Kolmogorov-Smirnov el punto crítico D_c que en este caso sería $D_c = 1,36 / \sqrt{125} = 0,1216$. Luego como $D = 0,03$ es menor que Y_c no se rechaza H_0 . En conclusión la distribución del *diámetro de la pieza* puede modelarse con la distribución Normal con los parámetros especificados en el encabezamiento de la tabla.
- Tener presente todas las columnas de la salida. Entonces como $p = 0,9993$ es mayor al nivel elegido 0,05 se puede decir que no se rechaza H_0 y por ende que la distribución del *diámetro de la pieza* puede modelarse con la distribución Normal con los parámetros especificados en el encabezamiento de la tabla.

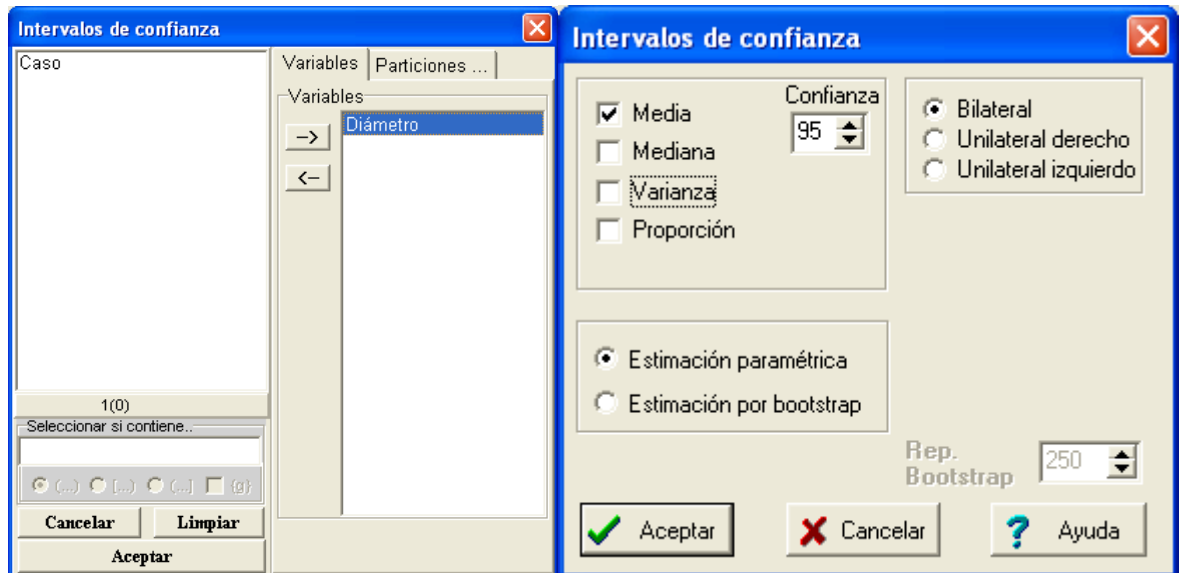
Comprobada la normalidad de la variable, ahora el grupo desea verificar si el diámetro cumple o no con los requerimientos de diseño. En primer lugar, recordemos que con los datos de la muestra inicial se calcularon estimaciones insesgadas de la verdadera media y varianza.

Resumen	Diámetro
n	125,000
Media	20,296
D.E.	0,624
Var (n-1)	0,389

Se calcularon la media aritmética y la varianza muestral corregida porque son estimaciones insesgadas de la verdadera media y varianza poblacionales (significa que el valor esperado de esos estimadores coincide con los parámetros poblacionales).

Esto es: $\mu = 20,29574$ y $\sigma^2 = 0,38933$

El diámetro tiene un valor nominal de 20 mm. Empleando un intervalo de confianza del 95 % para la media, el grupo va a explorar si la media de la variable cumple con esa especificación.



Intervalos de confianza

Bilateral

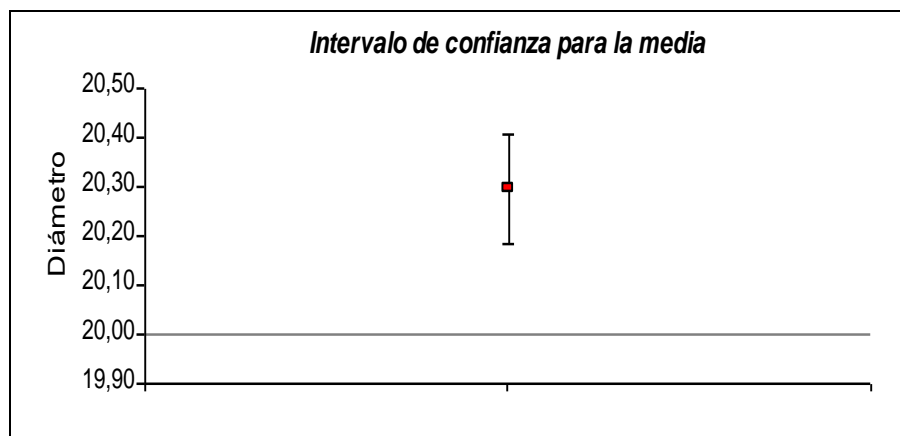
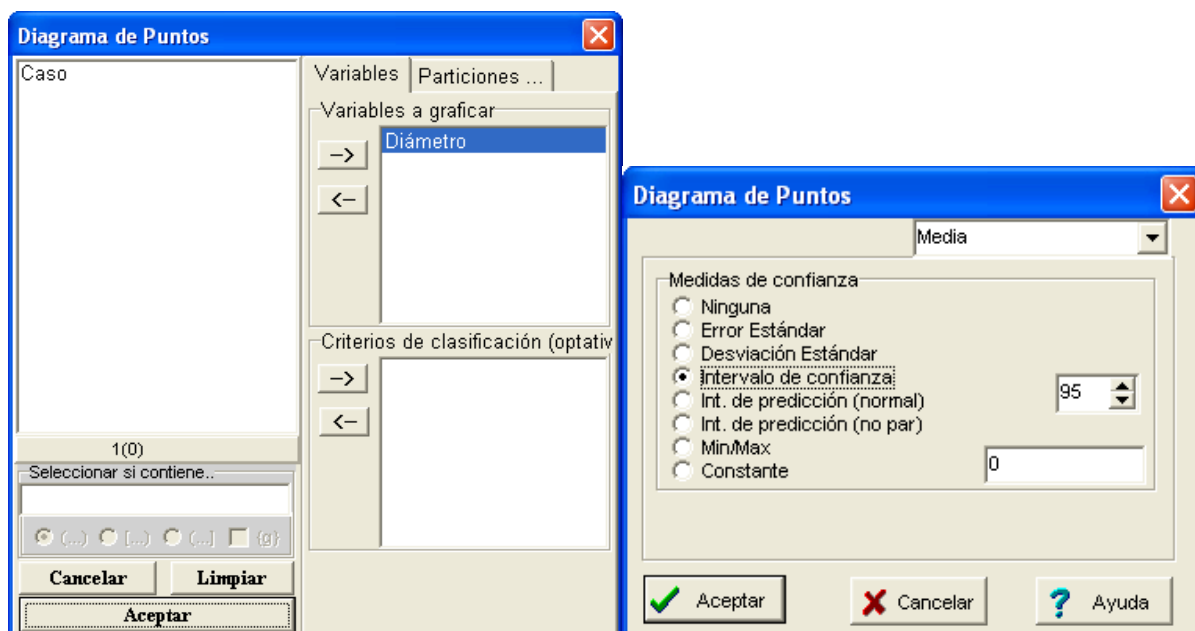
Estimación paramétrica

Variable	Parámetro	Estimación	E.E.	n	LI (95%)	LS (95%)
Diámetro	Media	20,30	0,06	125	20,19	20,41

La tabla muestra

- la estimación puntual de la media,
- el error estándar de la media,
- el tamaño de la muestra y
- un intervalo de confianza del 95 % para la verdadera media calculado en base a la distribución t.

Graficando



El intervalo obtenido [20,19; 20,41] tiene un 95% de confianza de incluir a la verdadera media poblacional del diámetro (es decir la media de toda la producción de piezas de ese tipo), pero no incluye al valor nominal y está íntegramente por encima de ese valor. Por lo tanto existen fuertes evidencias estadísticas que permiten suponer que la verdadera media poblacional es significativamente mayor al valor nominal, por lo que el proceso no está correctamente centrado en el valor objetivo.

Otro procedimiento que el grupo emplea para comprobar si el diámetro medio difiere significativamente del valor nominal (20 mm.) es la prueba de hipótesis. En este caso, como las evidencias parecen sugerir que el diámetro obtenido es mayor, se decide utilizar una prueba de unilateral derecha. Es decir, se plantea que:

$$H_0: \mu = 20 \text{ mm.}$$

$$H_1: \mu > 20 \text{ mm.}$$

Valor del
parámetro bajo H_0

Prueba T para un parámetro

Valor del parámetro probado: 20

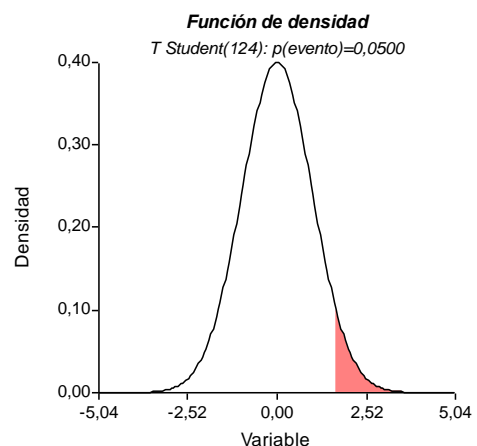
Variable	n	Media	DE	LI (95)	LS (95)	T	p (Bilateral)
Diámetro	125	20,30	0,62	20,19	20,41	5,30	<0,0001

La tabla proporciona

- el tamaño de la muestra,
- estimaciones puntuales de la media y el desvío estándar de la variable,
- un intervalo de confianza del 95 % para la verdadera media,
- la estimación del estadístico de prueba (que tiene distribución t), y
- el p-valor asociado a dicha estimación.

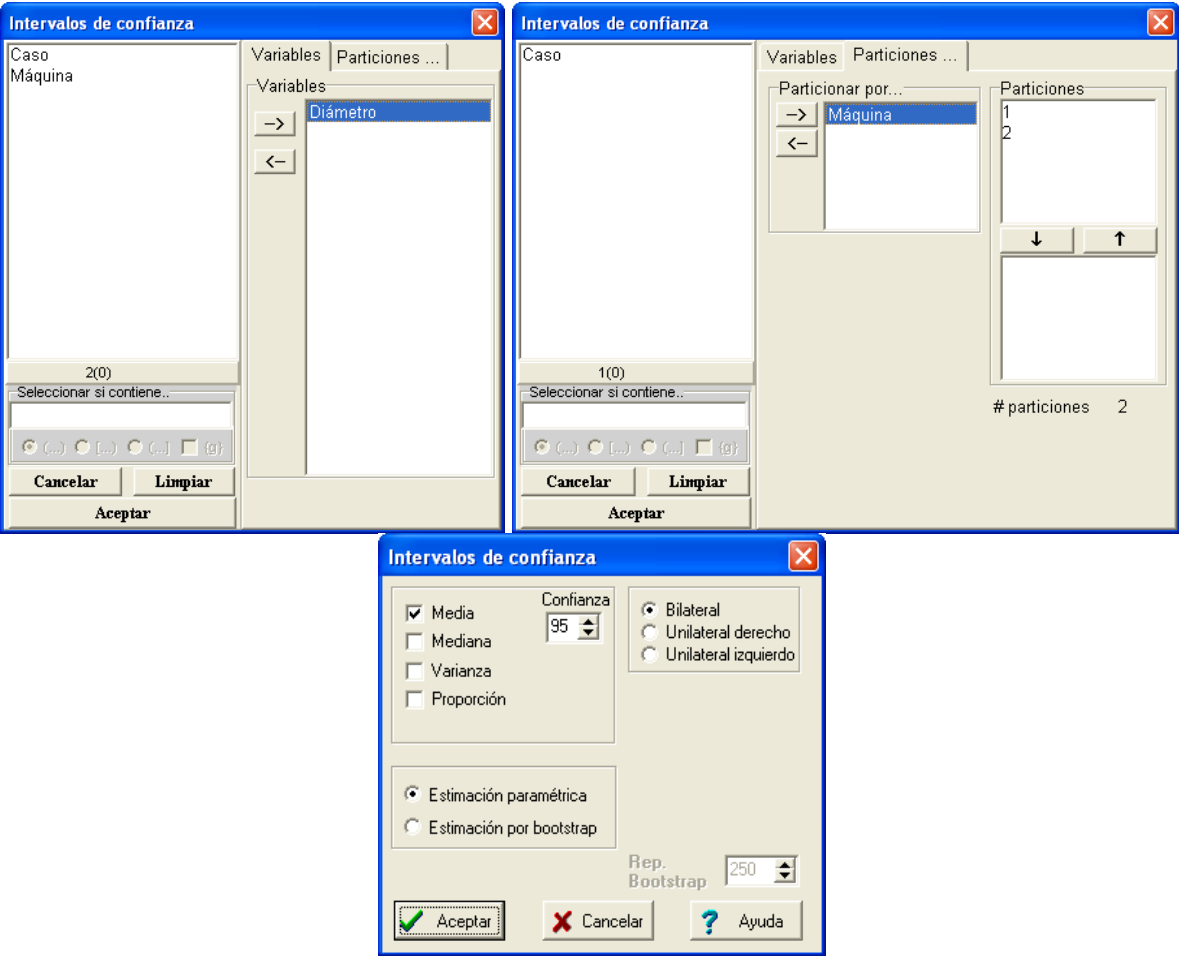
En este caso el valor p es menor a cualquiera de los niveles de significación usados, lo que indica que el estadístico de prueba (T) cae en zona de rechazo. Por lo tanto se rechaza la hipótesis nula de que la media poblacional es igual a 20, comprobándose la hipótesis alternativa que postula que es significativamente mayor a ese valor.

Observación: si no se desea considerar la información acerca del p-valor, se puede calcular el punto crítico $t_c = t_{\alpha, n-1}$ que en este caso sería $t_c = t_{0,05;124} = 1,657$.



Luego como $T = 5,30$ es mayor que t_c , se rechaza H_0 . En conclusión *media poblacional del diámetro de la pieza es significativamente mayor a 20.*

Por otra parte el grupo desea comparar el funcionamiento de las dos máquinas que producen la pieza en estudio. Para comparar el centrado de las dos máquinas, el grupo compara los intervalos de confianza del 95 % para las verdaderas medias.



Intervalos de confianza

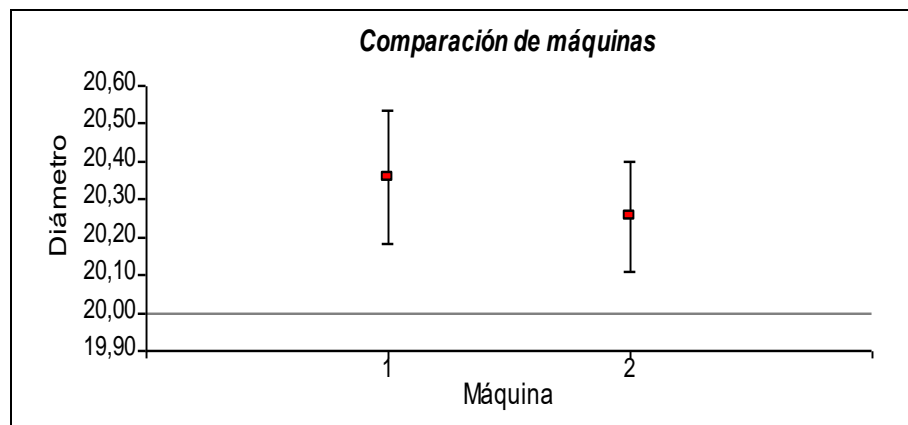
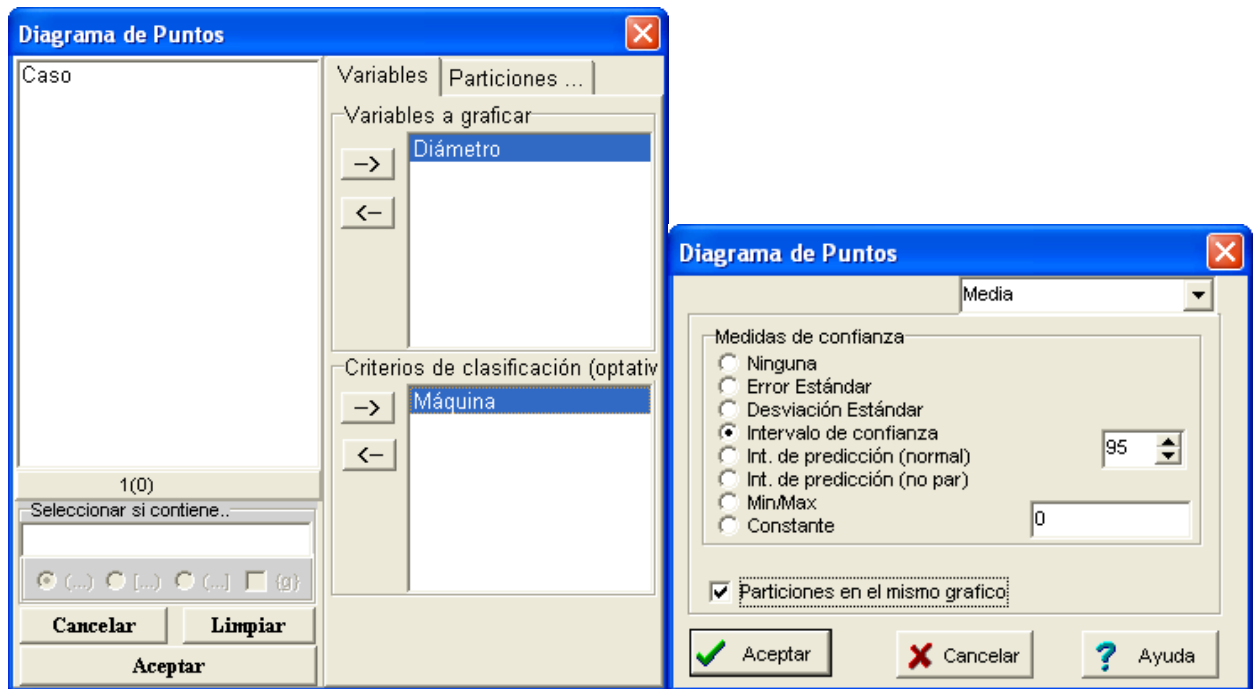
Bilateral
Estimación paramétrica

Máquina	Variable	Parámetro	Estimación	E.E.	n	LI (95%)	LS (95%)
1	Diámetro	Media	20,36	0,09	50	20,18	20,53
2	Diámetro	Media	20,25	0,07	75	20,11	20,40

La tabla contiene

- estimaciones puntuales de la media para ambos grupos,
- los errores estándar de la media,
- los tamaños muestrales y
- los intervalos de confianza del 95 % para la verdadera media de cada población.

Graficando

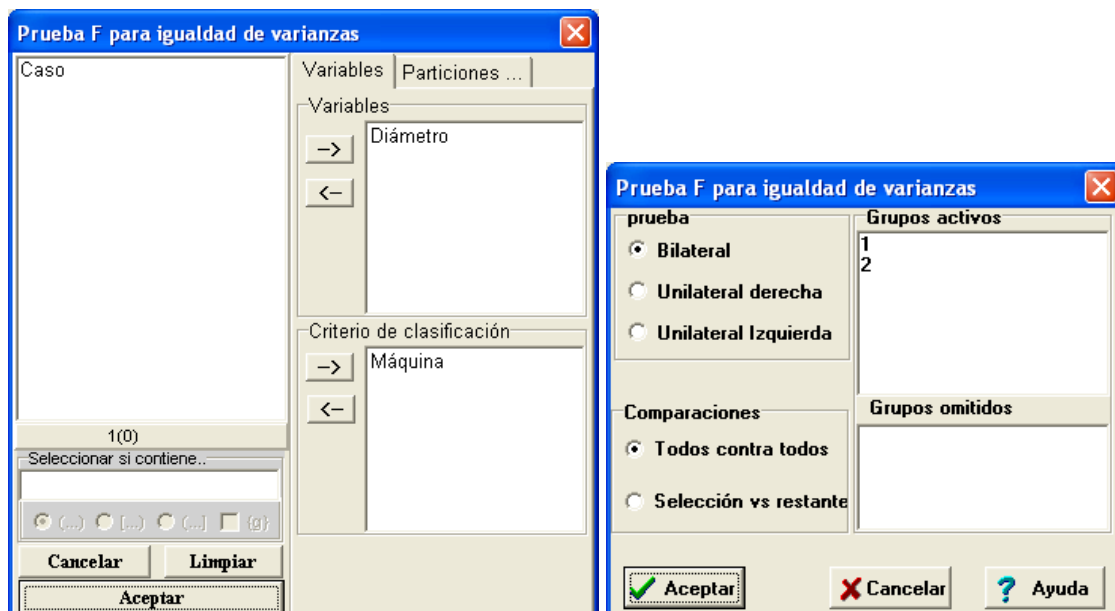


Los intervalos de confianza para la media se solapan, lo que indica que es razonable suponer que las verdaderas medias no difieren significativamente para ambas máquinas.

Ahora, para comprobar si la dispersión de las dos máquinas es similar, el grupo usa la prueba de comparación de varianzas poblacionales. Planteamos

$$H_0: \sigma^2_{\text{maq1}} = \sigma^2_{\text{maq2}}$$

$$H_1: \sigma^2_{\text{maq1}} \neq \sigma^2_{\text{maq2}}$$



Prueba F para igualdad de varianzas

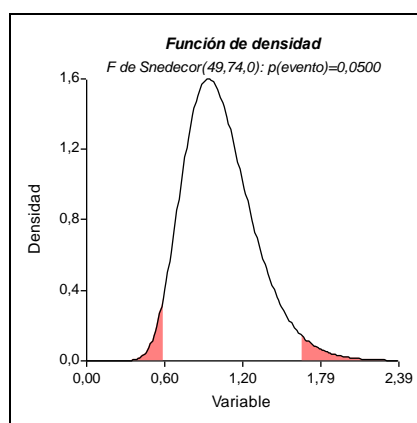
Variable	Grupo (1)	Grupo (2)	n (1)	n (2)	Var (1)	Var (2)	F	p	prueba
Diámetro	{1}	{2}	50	75	0,39	0,39	0,99	0,9546	Bilateral

La tabla proporciona

- los tamaños de cada muestra,
- estimaciones puntuales de las varianzas de ambas muestras,
- una estimación del estadístico de prueba (que tiene distribución F),
- y el valor p asociado a dicha estimación. Se explicita que es una prueba bilateral.

Como el valor p es mayor a cualquiera de los niveles de significación usados, significa que el estadístico de prueba (F) cae en zona de aceptación, por lo que no se rechaza la hipótesis de igualdad de varianzas poblacionales. Entonces la dispersión de las dos máquinas es similar.

Observación: recuerde que si no se desea considerar la información acerca del p-valor, se pueden calcular los puntos críticos $F_{c1} = F_{1-\alpha/2, (n1-1), (n2-1)}$ y $F_{c2} = F_{\alpha/2, (n1-1), (n2-1)}$ para luego analizar si el valor obtenido F, cae o no en la región de rechazo.



$$F_{c1} = F_{1-\alpha/2, (n1-1), (n2-1)} = F_{0,0975, 49, 74} = 0,589$$

$$F_{c2} = F_{\alpha/2, (n1-1), (n2-1)} = F_{0,025, 49, 74} = 1,652$$

Calculador de Probabilidades y cuantiles

Seleccione la distribución

- ☐ Uniforme (a,b)
- ☐ Normal (media,varianza)
- ☐ T Student (v)
- ☐ Chi Cuadrado (v,lambda)
- ☒ F no central (u,v,lambda)
- ☐ Exponencial (lambda)
- ☐ Gamma (lambda,r)
- ☐ Beta (a,b)
- ☐ Weibull (a,b)
- ☐ Doble Exponencial (a,b)
- ☐ Logística (a,b)
- ☐ Pareto (Theta,a)
- ☐ Gumbel (a,b)
- ☐ Rangos estudent (k,v)
- ☐ Poisson (lambda)
- ☐ Binomial (n,p)
- ☐ Geométrica (p)
- ☐ Hipergeométrica (m,k,n)
- ☐ Binomial Negativa (m,k)
- ☐ Beta-Binomial(p,Rho,N)

49 **u**

74 **v**

0 **lambda**

Valor de x
0,5890319590

Prob. [X<=x]
0,025

Prob. [X>x]
0,975

Prob. [X=x]

Calcular Ayuda

Calculador de Probabilidades y cuantiles

Seleccione la distribución

- ☐ Uniforme (a,b)
- ☐ Normal (media,varianza)
- ☐ T Student (v)
- ☐ Chi Cuadrado (v,lambda)
- ☒ F no central (u,v,lambda)
- ☐ Exponencial (lambda)
- ☐ Gamma (lambda,r)
- ☐ Beta (a,b)
- ☐ Weibull (a,b)
- ☐ Doble Exponencial (a,b)
- ☐ Logística (a,b)
- ☐ Pareto (Theta,a)
- ☐ Gumbel (a,b)
- ☐ Rangos estudent (k,v)
- ☐ Poisson (lambda)
- ☐ Binomial (n,p)
- ☐ Geométrica (p)
- ☐ Hipergeométrica (m,k,n)
- ☐ Binomial Negativa (m,k)
- ☐ Beta-Binomial(p,Rho,N)

49 **u**

74 **v**

0 **lambda**

Valor de x
1,6519853752

Prob. [X<=x]
0,975

Prob. [X>x]
0,025

Prob. [X=x]

Calcular Ayuda

Luego, como $0,589 \leq F (= 0,99) \leq 1,652$ no se rechaza hipótesis nula.

Si queremos comparar el centrado de las dos máquinas, entonces realizamos la prueba de igualdad de medias poblacionales. Planteamos

$$H_0: \mu_{\text{maq1}} = \mu_{\text{maq2}}$$

$$H_1: \mu_{\text{maq1}} \neq \mu_{\text{maq2}}$$

Prueba T para muestras Independientes

Caso

Variables

Particiones ...

Variables

→ Diámetro

←

Criterio de clasificación

→ Máquina

←

1(0)

Seleccionar si contiene...

☐ () ☐ [] ☐ (] ☐ [)

Cancelar Limpiar

Aceptar

Prueba T para muestras Independientes

prueba

☒ Bilateral

☐ Unilateral derecha

☐ Unilateral Izquierda

Comparaciones

☒ Todos contra todos

☐ Selección vs restantes

☒ N obs. ☒ Medias

☐ Varianzas ☒ p

☐ gl ☒ T

☐ Inter conf.

☒ Si las varianzas no son homogéneas a nivel del 5 % hacer automáticamente una prueba T para varianzas heterogéneas

Grupos activos

1

2

Grupos omitidos

Aceptar Cancelar Ayuda

Prueba T para muestras Independientes

Clasific	Variable	Grupo (1)	Grupo (2)	n (1)	n (2)
Máquina	Diámetro	{1}	{2}	50	75

media (1)	media (2)	p (Var.Hom.)	T	p	prueba
20,36	20,25	0,9828	0,90	0,3677	Bilateral

La tabla proporciona

- los tamaños de cada muestra,
- estimaciones puntuales de la media para cada grupo,
- el valor p para la prueba de homogeneidad de varianzas (donde se hipotetiza que las varianzas poblacionales son iguales),
- el estadístico T para la prueba de igualdad de medias poblacionales (que tiene distribución t), y
- el valor p asociado a la estimación de dicho estadístico.

Nuevamente el valor p es mayor a cualquiera de los niveles de significación usados, lo que indica que la estimación del estadístico de prueba (T) cae en zona de aceptación, por lo que no se rechaza la hipótesis de igualdad de medias poblacionales.

Por lo tanto, para ambas máquinas, las verdaderas medias no difieren significativamente. Esta conclusión es coincidente con la obtenida a partir de la comparación de los intervalos de confianza para la media.

Observación: recuerde que si no se desea considerar la información acerca del p-valor, se pueden calcular los puntos críticos t_{c1} y t_{c2} para luego analizar si el valor obtenido T, cae o no en la región de rechazo.

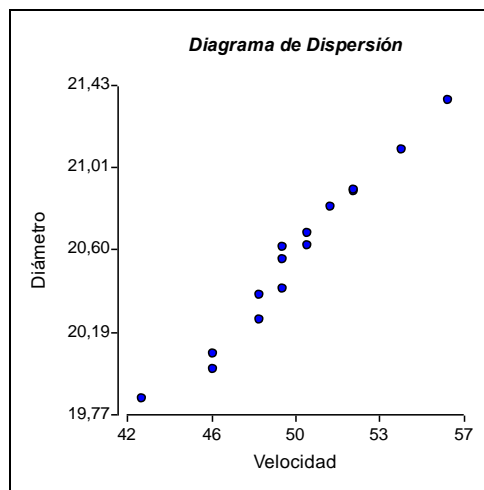
Finalmente, el grupo de trabajo sospecha que en el proceso de fabricación de la pieza, el diámetro obtenido depende significativamente de la velocidad de avance de la máquina que lo produce.

Para comprobar si la suposición es razonable, se extrae la siguiente muestra de observaciones conjuntas de velocidad vs. diámetro.

Diámetro	19,85	20,00	20,08	20,25	20,37	20,40	20,55	20,61	20,62	20,68	20,81	20,89	20,90	21,10	21,35
Velocidad	43	46	46	48	48	49	49	49	50	50	51	52	52	54	56

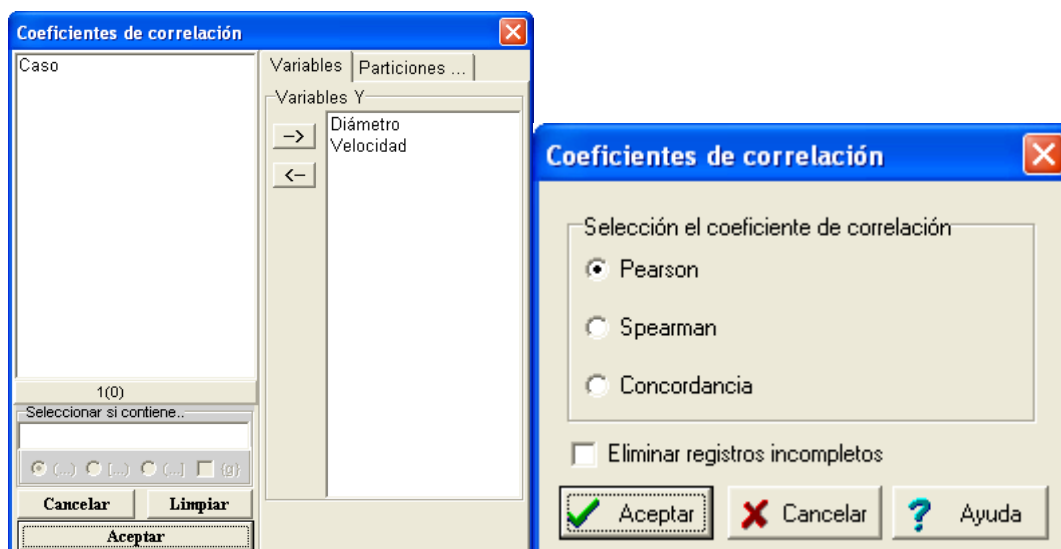
Para analizar la relación entre las dos variables, se elabora un diagrama de dispersión, se realiza el análisis de correlación, y se ajusta un modelo de regresión lineal simple.





En ordenadas se coloca la variable de respuesta y en abscisas la variable regresora o explicativa. Analizando el gráfico de dispersión, este nos muestra una clara relación directa entre ambas variables, de tipo lineal.

Coefficientes de correlación



Correlacion de Pearson: coeficientes\probabilidades

	Diámetro	Velocidad
Diámetro	1,00	3,8E-11
Velocidad	0,98	1,00

Notar que

- Por debajo de la diagonal principal se proporciona el coeficiente de correlación
- Por encima el valor p asociado a la prueba de hipótesis sobre el coeficiente de correlación poblacional, es decir $H_0: \rho = 0$ vs. $H_1: \rho \neq 0$.

Como el coeficiente de correlación es igual a 0,98 nos está indicando que existe un elevado grado de asociación lineal entre las dos variables. Por otro lado, el valor p es menor a cualquiera de los niveles de significación usados, indicándonos que se rechaza la hipótesis de que el coeficiente de correlación poblacional es igual a cero. Es decir que el coeficiente de correlación poblacional es significativamente diferente de cero, indicando que existe asociación lineal.

Se ajusta un modelo de regresión lineal simple.

Análisis de regresión lineal

Caso

Variables

Particiones ...

Variable dependiente

→ Diámetro

←

Regresoras

→ Velocidad

←

1(0)

Seleccionar si contiene...

☐ (..) ☐ (..) ☐ (..) ☐ (0)

Cancelar Limpiar

Aceptar

Pesos (solo una)

→

←

Análisis de regresión lineal

General Diagnóstico Polinomios Hipótesis Selección de modelo

Mostrar...

☒ Coeficientes de regresión y estadísticos asociados

☒ Tabla de análisis de la varianza ☐ SC Tipo I

☐ Tabla resumen criterios de diagnóstico

☐ Matriz de covarianzas de los coef. de regresión

☐ Solo mostrar los coeficientes de regresión

Opciones

☒ Ordenada al origen

☐ Regresoras centradas

☐ Prueba de Atkinson

☐ Error puro

☐ Todas las reg. simples

Método de Estimación

☒ Mínimos Cuadrados

Aceptar Cancelar Ayuda

Análisis de regresión lineal

General Diagnóstico Polinomios Hipótesis Selección de modelo

Guardar....

☒ Residuos ☒ Predichos ☐ Rparcial

☐ Residuos Jackknife ☐ DCook ☐ DFBetas

☒ Res. estudentizados

☐ Res. ext. student. ☐ Leverage

Típico

☒ Sobrescribir

Intervalos de ..

☐ Confianza ☐ Predicción 95

Graficar...

☒ REstud. vs predichos ☒ Ajuste

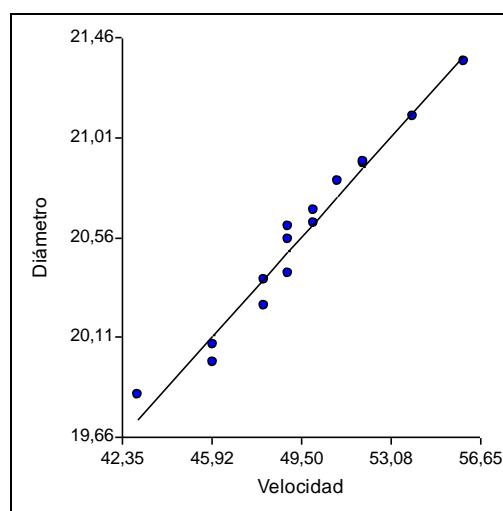
☐ REstud vs R ExEstud ☐ Bandas de confianza

☐ Leverage ☐ Bandas de predicción

☐ DCook

Aceptar Cancelar Ayuda

Gráficamente obtenemos



Análisis de regresión lineal

Variable	N	R^2	R^2 Aj	ECMP
Diámetro	15	0,97	0,97	0,01

La tabla muestra

- el coeficiente de determinación,
- su valor ajustado (por los grados de libertad), y
- el Error Cuadrático Medio de Predicción.

El coeficiente de determinación es igual a 0,97, entonces el 97 % de la variación de la variable dependiente (Diámetro), al variar la explicativa (Velocidad), es explicada por el modelo.

Coefficientes de regresión y estadísticos asociados

Coef	Est.	EE	LI (95%)	LS (95%)	T	p-valor
const	14,36	0,31	13,68	15,03	46,13	<0,0001
Velocidad	0,13	0,01	0,11	0,14	19,99	<0,0001

Esta tabla proporciona

- una estimación de cada uno de los coeficientes del modelo lineal simple,
- su error estándar,
- un intervalo de confianza del 95 % para el verdadero valor del parámetro,
- el estadístico T para la prueba de hipótesis sobre el coeficiente. En cada caso se hipotetiza $H_0: \alpha = 0$ vs. $H_1: \alpha \neq 0$ y $H_0: \beta = 0$ vs. $H_1: \beta \neq 0$ respectivamente.
- el p-valor asociado a cada una de dichas pruebas.

*Teniendo en cuenta la información que proporciona la **Tabla de coeficientes de regresión y estadísticos asociados** resulta que el modelo ajustado es el siguiente:*

$$\hat{y} = a + b * x$$

$$\hat{y} = 14,36 + 0,13 * x$$

siendo el diámetro la variable dependiente “y” y la velocidad la variable regresora “x”.

Si observamos los intervalos de confianza para el verdadero valor de cada uno de los parámetros, vemos que ninguno de ellos incluye el cero.

Los valores p asociados a ambos coeficientes son menores a cualquiera de los niveles de significación usuales, lo que permite rechazar la hipótesis nula de que los coeficientes no son significativos.

Observación: recuerde que si no se desea considerar la información acerca del p-valor, se pueden calcular para cada prueba, los puntos críticos t_{c1} y t_{c2} para luego analizar si el valor obtenido T, cae o no en la región de rechazo.

Cuadro de Análisis de la Varianza (SC tipo III)

F.V.	SC	gl	CM	F	p-valor
Modelo.	2,35	1	2,35	399,63	<0,0001
Velocidad	2,35	1	2,35	399,63	<0,0001
Error	0,08	13	0,01		
Total	2,43	14			

Esta tabla muestra los siguientes ítems:

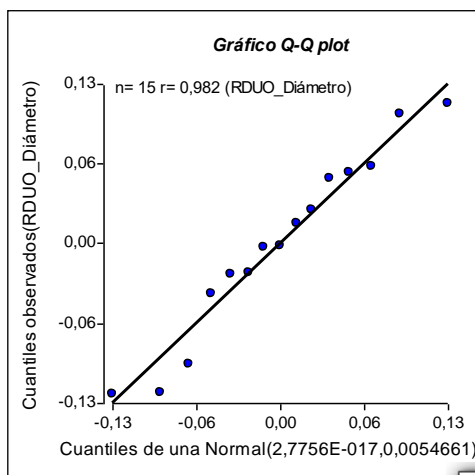
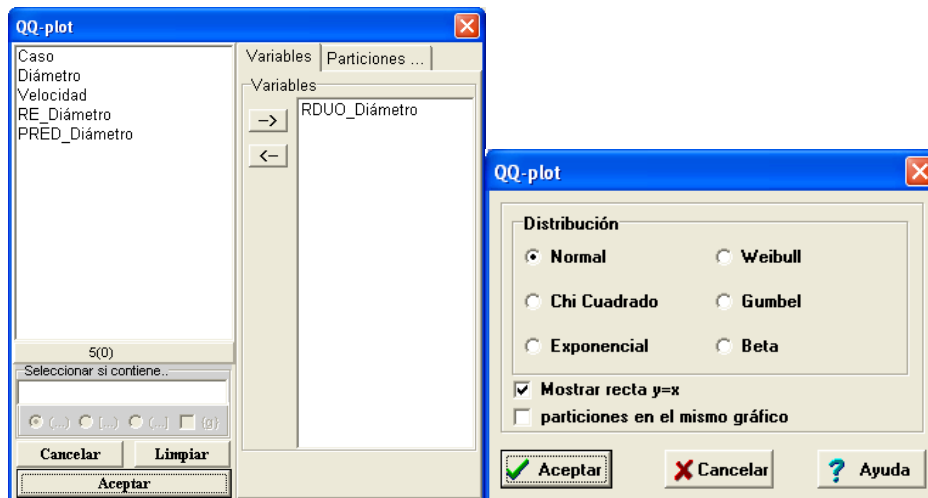
- F.V.: Fuente de Variación
- SC: Suma de Cuadrados

- CM: Cuadrados Medios
- gl: grados de libertad
- el estadístico F para la prueba de hipótesis sobre el modelo global. En cada caso se hipotetiza H_0 : El modelo ajustado no es satisfactorio vs. H_1 : El modelo ajustado es satisfactorio
- el p-valor asociado dichas prueba

Las dos primeras filas tienen los mismos valores, ya que como el modelo tiene una sola variable explicativa, lo que explica el modelo coincide con lo que explica esa variable. Los valores de la tabla permiten calcular con facilidad la varianza residual y el coeficiente de determinación. El p-valor en este caso es inferior a cualquier nivel de significación convencional, por lo que se rechaza la hipótesis nula, aceptando que el modelo globalmente es satisfactorio para explicar la relación entre ambas variables.

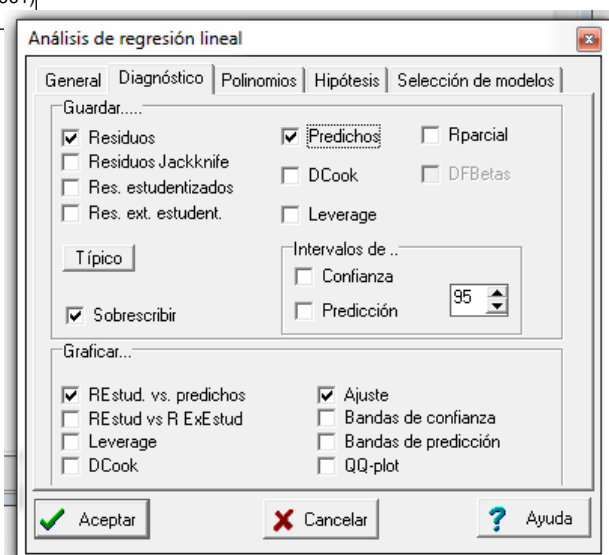
Validación de los supuestos

Normalidad



El Q-Q plot se realiza con los residuos del modelo de regresión y usando como distribución teórica la Normal. Los puntos se disponen en una recta a 45° indicando que el supuesto distribucional para los residuos se cumple.

y varianza constante de los

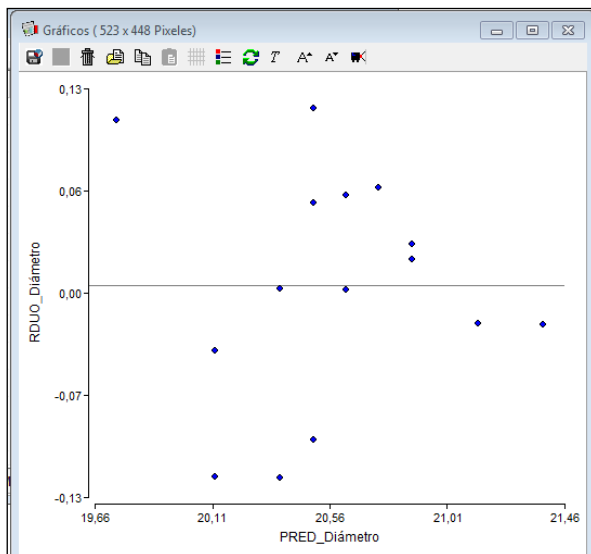


Independencia, media cero residuos

Al ejecutar el menú “Regresión lineal”, en la solapa “Diagnóstico” hay que marcar los casilleros correspondientes a los Residuos y los valores Predichos. Al ejecutar el menú se generan las siguientes columnas:

Caso	Diámetro	Velocidad	RDUO_Diámetro	PRED_Diámetro
1	19,85	43	0,10	19,75
2	20,00	46	-0,12	20,12
3	20,08	46	-0,04	20,12
4	20,25	48	-0,12	20,37
5	20,37	48	0,00	20,37
6	20,40	49	-0,10	20,50
7	20,55	49	0,05	20,50
8	20,61	49	0,11	20,50
9	20,62	50	0,00	20,62
10	20,68	50	0,06	20,62
11	20,81	51	0,06	20,75
12	20,89	52	0,02	20,87
13	20,90	52	0,03	20,87
14	21,10	54	-0,02	21,12
15	21,35	56	-0,02	21,37

Luego se realiza un diagrama de los residuos vs. los valores predichos, mediante el cual se puede evaluar visualmente los supuestos de independencia, media cero y varianza constante de los residuos.



La inspección del diagrama muestra que los residuos se distribuyen aleatoriamente alrededor del cero, sin presentar ningún patrón de comportamiento, lo que permite suponer que son independientes. Además se ve a simple vista que la media es cero y como la dispersión alrededor de este valor se mantiene aproximadamente constante (no aumenta no disminuye sistemáticamente), es razonable el supuesto de varianza constante.