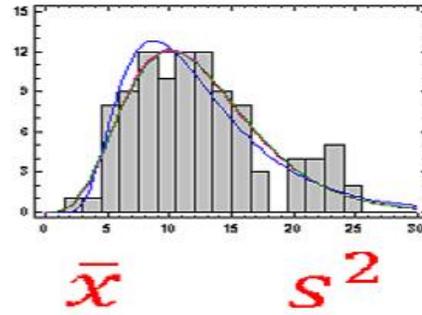




*Seminario Internacional:
Eventos hidrológicos extremos: Análisis y aplicaciones
Santiago – Chile 1/13/2012*

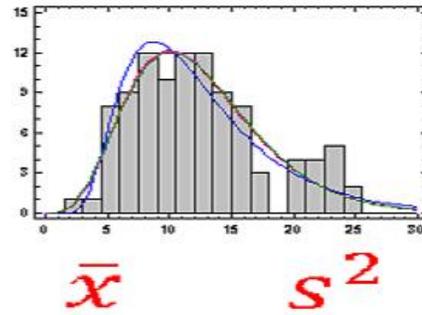
**Métodos estadísticos
tradicionales para el análisis de
eventos extremos**

Dr. Oscar Santiago Vallejos Barra



Función que proporciona la frecuencia relativa

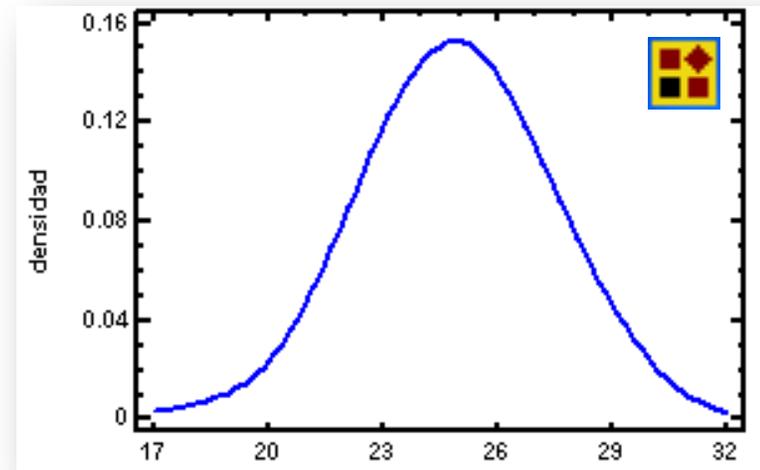
1. FUNCIÓN DE DENSIDAD DE PROBABILIDAD (FDP)

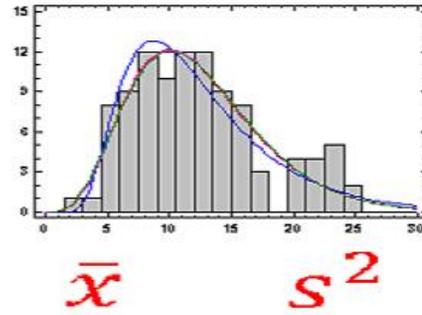


Fdp normal

$$f(x) = \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left(-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2\right)$$

- La media se obtiene del primer momento (\bar{x})
- La varianza se obtiene del segundo momento (s^2)





Primer momento

$$E(x) = \int_{-\infty}^{\infty} x f(x) \partial x$$

Si $z = \frac{x - \mu}{\sqrt{2} \sigma}$

entonces

$$x = z \sqrt{2} \sigma + \mu$$

Por lo tanto

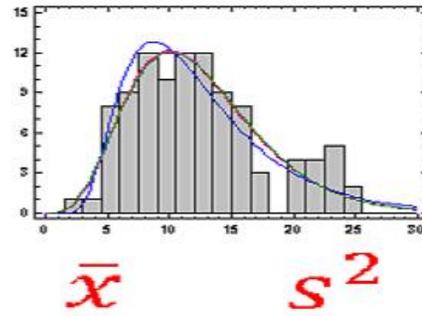
$$\frac{\partial z}{\partial x} = \frac{1}{\sqrt{2} \sigma}$$

y

$$\partial x = \partial z \sqrt{2} \sigma$$

$$E(x) = \frac{\sqrt{2} \sigma}{\sqrt{\pi}} \int_{-\infty}^{\infty} z \exp(-z^2) \partial z + \frac{\mu}{\sqrt{\pi}} \int_{-\infty}^{\infty} \exp(-z^2) \partial z$$

$$E(x) = \frac{\sqrt{2} \sigma}{\sqrt{\pi}} 0 + \frac{\mu}{\sqrt{\pi}} \sqrt{\pi} = \mu$$



Segundo momento

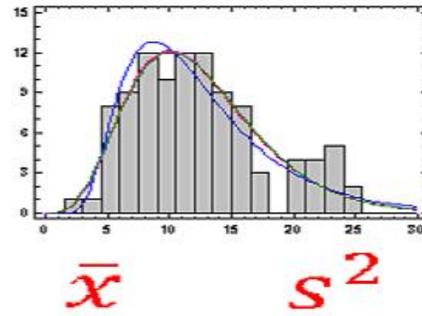
$$E(x^2) = \int_{-\infty}^{\infty} x^2 f(x) dx$$

$$E(x^2) = \frac{2\sigma^2}{\sqrt{\pi}} \int_{-\infty}^{\infty} z^2 \exp(-z^2) dz + \frac{2\sqrt{2}\sigma\mu}{\sqrt{\pi}} \int_{-\infty}^{\infty} z \exp(-z^2) dz + \frac{\mu^2}{\sqrt{\pi}} \int_{-\infty}^{\infty} \exp(-z^2) dz$$

$$E(x^2) = \frac{2\sigma^2\sqrt{\pi}}{\sqrt{\pi} \cdot 2} + \frac{2\sqrt{2}\sigma\mu}{\sqrt{\pi}} 0 + \frac{\mu^2}{\sqrt{\pi}} \sqrt{\pi} = \sigma^2 + \mu^2$$

Por lo tanto

$$\sigma^2 = E(x^2) - (E(x))^2 = \sigma^2 + \mu^2 - \mu^2 = \sigma^2$$

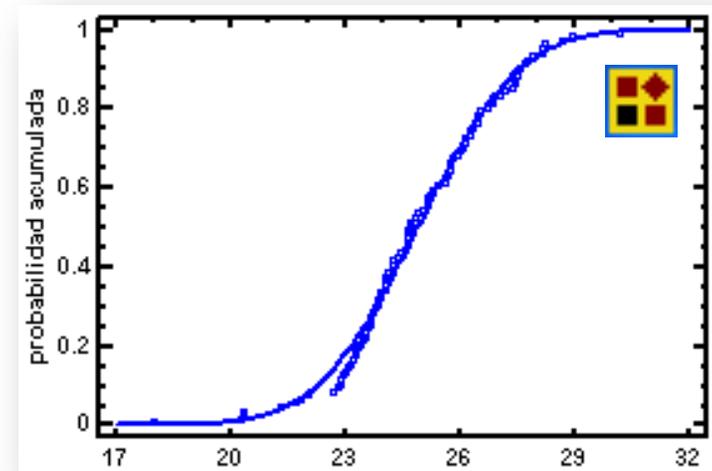


¿Cómo se comprueba normalidad?

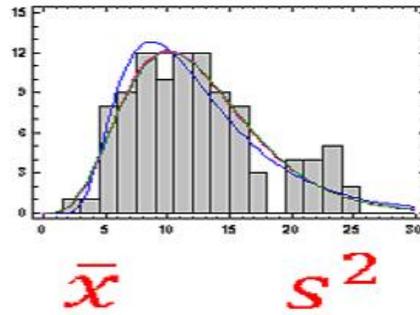
- Prueba Shapiro-Wilk

Pruebas de Normalidad para normal		
Prueba	Estadístico	Valor-P
Estadístico W de Shapiro-Wilk	0.985636	0.813986

0.814



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



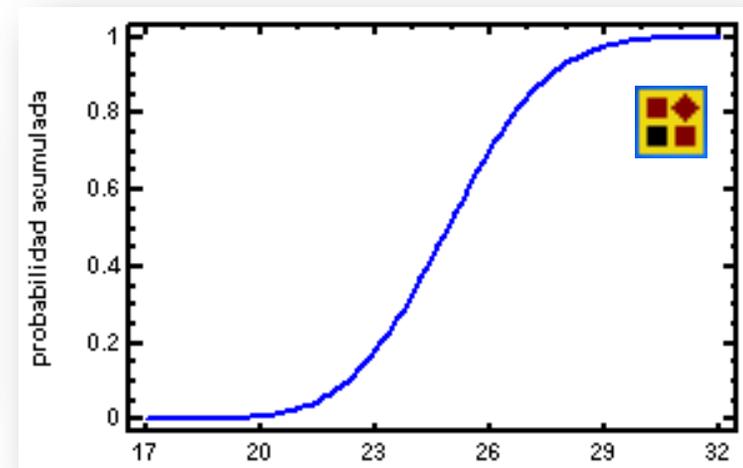
¿Cómo se comprueba normalidad?

- Prueba Kolmogorov-Smirnov

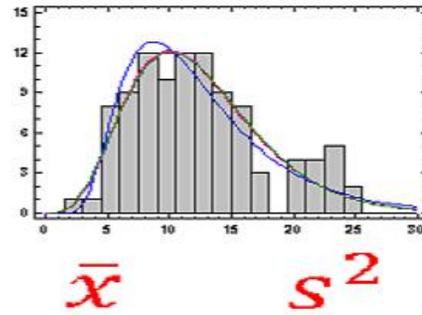
Pruebas de Bondad-de-Ajuste para normal
Prueba de Kolmogorov-Smirnov

	Normal
DMAS	0.0542965
DMENOS	0.0703409
DN	0.0703409
Valor-P	0.705553

0.706



$$D = \max(D^+, D^-) = \max\left(\max_i \left\{ \frac{i}{n} - z_{(i)} \right\}, \max_i \left\{ z_{(i)} - \frac{i-1}{n} \right\} \right)$$



Statgraphics Centurion

STATGRAPHICS Centurion - expo 2011.sgp

Archivo Editar Graficar Describir Comparar Relacionar Pronósticos CEP DDE SnapStats!! Herramientas Ver Ventana Ayuda

Libro de Datos StatAdvisor StatGallery StatReporter Comentarios del StatF Ajuste de Datos No C

Ajuste de Distributions (Datos No Censurados)

normal Datos: normal (Selección:)

Ordenar nombres de columna

Aceptar Cancelar Borrar Transformar... Ayuda

Este análisis muestra los resultados de ajustar una distribución normal a los datos de 1

Pruebas de Normalidad para normal

Prueba	Estadístico	Valor-P
Estadístico W de Shapiro-Wilk	0.985636	0.813986

El StatAdvisor
Esta ventana muestra los resultados de diversas pruebas realizadas para determinar si no normal. La prueba de Shapiro-Wilk está basada en la comparación de los cuartiles de la c
Debido a que el valor-P más pequeño de las pruebas realizadas es mayor ó igual a 0.05, n distribución normal con 95% de confianza.

Pruebas de Bondad-de-Ajuste para normal
Prueba de Kolmogorov-Smirnov

	Normal
DMAS	0.0542965
DMENOS	0.0703409
DN	0.0703409
Valor-P	0.705553

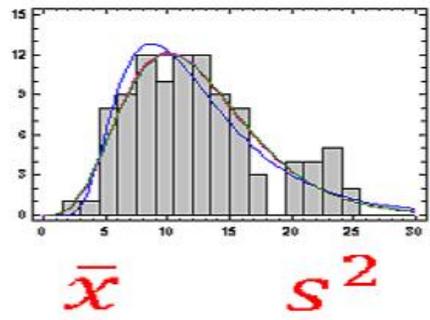
El StatAdvisor
Esta ventana muestra los resultados de diversas pruebas realizadas para determinar si normal.

Gráfico de Densidad

Gráfico de Cuantiles

Distribución Acumulada

Use el botón derecho del mouse para seleccionar opciones



¿Otras fdp?

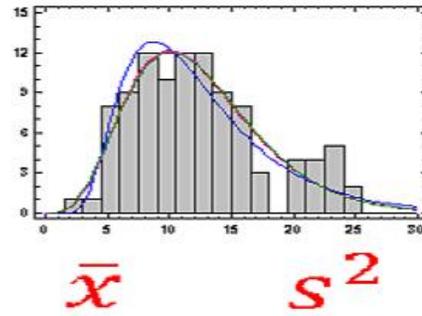
Comparación de Distribuciones Alternas					
Distribución	Parámetros Est.	KS D	Distribución	Parámetros Est.	KS D
Loglogística	2	0.0012511	Exponencial	2	0.104922
Logística Generalizada	3		Valor Extremo Más Grande	2	
Loglogística (3-Parámetros)	3		Triangular	3	
Logística	2		Gamma Generalizada	3	0.186599
Potenciación Exponencial	3	0.0010278	Maxwell	2	0.204184
Lognormal (3-Parámetros)	3	0.0683286	Rayleigh	2	0.277018
Beta (4-Parámetros)	4	0.068679	Chi-Cuadrada	1	0.306682
Normal Plegada	2	0.0691249	Uniforme	2	0.306704
Normal	2	0.0800400	Mitad Normal (2-Parámetros)	2	0.408262
Gamma (3-Parámetros)	3		Exponencial (2 Parámetros)	2	0.414011
Gamma	2		Pareto (2-Parámetros)	2	0.434235
t No Centrada	2	0.0733547	Chi-Cuadrada no Centrada	2	0.536767
Gaussiana Inversa	2	0.0745294	Exponencial	1	0.547223
Birnbaum-Saunders	2	0.0810000	Pareto	1	0.598305
Lognormal	2		F	2	0.705393
Weibull (3-Parámetros)	3		F No Centrada	3	0.772953
Weibull	2	0.0914972	Erlang	2	1.0
Cauchy	2	0.0949638	Beta	<sin ajuste>	
Valor Extremo Más Chico	2	0.0994014	t de Student	<sin ajuste>	

Logística Generalizada

Person III

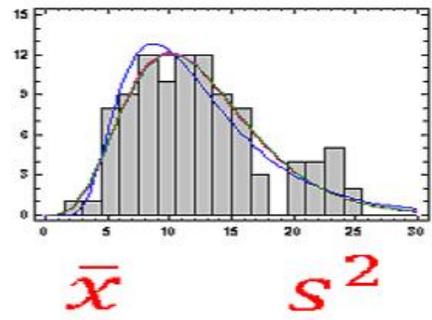
Lognormal

Gumbel

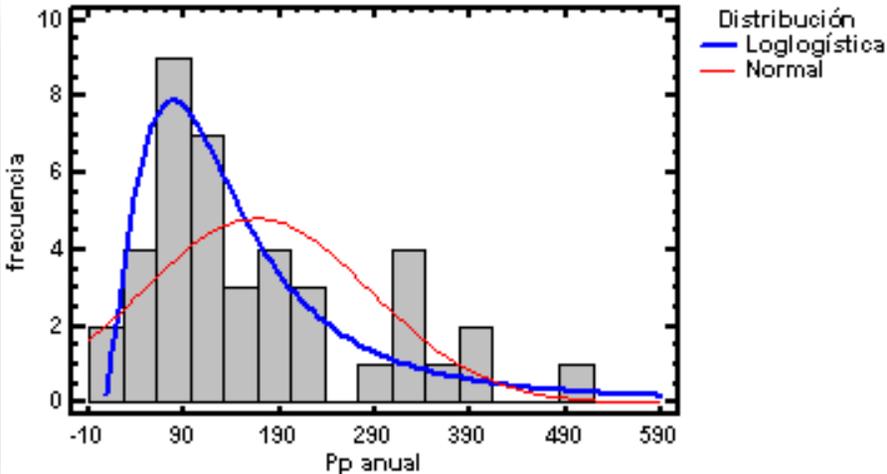


Datos reales... Región de Coquimbo

	normal	Año	Pp anual	Col_4	Col_5	Col_6	Co
1	27.7842	1964	139				
2	27.0703	1965	331.9				
3	25.6281	1966	94				
4	28.674	1967	61.5				
5	24.7079	1968	36.5				



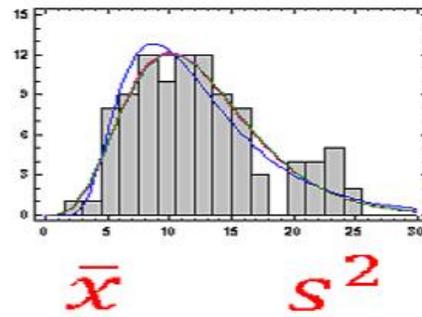
Datos reales... Región de Coquimbo



Comparación de Distribuciones Alternas

Distribución	Parámetros Est.	KS D
Loglogística (3-Parámetros)	3	0.0825376
Gamma (3-Parámetros)	3	0.0860286
Lognormal	2	0.0865011
Loglogística	2	0.0874081
Weibull (3-Parámetros)	3	0.0884479
Lognormal (3-Parámetros)	3	0.0921898

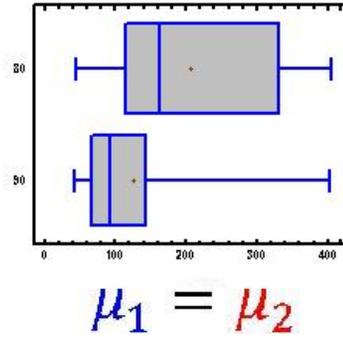
Distribución	Parámetros Est.	KS D
Weibull	2	0.0928508
Valor Extremo Más Grande	2	0.11928
Logística Generalizada	3	0.120014
Logística	2	0.127187
Normal	2	0.157679



Datos reales... Región de Coquimbo

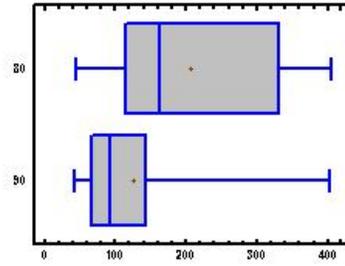
Áreas de Cola para Pp anual		
Área Cola Inferior (<=)		
X	Loglogística (3-Parámetros)	Normal
47.6898	0.1	0.16326
80.0	0.267312	0.238313
95.0	0.348001	0.278714
100.0	0.373951	0.292885
200.0	0.725548	0.613653

Valores Críticos para Pp anual		
Área Cola Inferior (<=)	Loglogística (3-Parámetros)	Normal
0.01	19.4578	-113.641
0.1	47.6898	11.6616
0.5	126.773	165.359
0.9	359.016	319.056



¿Hay diferencias estadísticamente significativas?

2. COMPARACIÓN DE DOS MUESTRAS



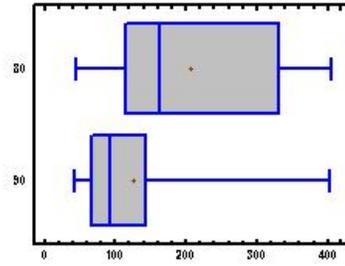
$$\mu_1 = \mu_2$$



Comparación por décadas

	80	90
1	301.5	67.5
2	68	169.5
3	330.5	143.5
4	167	74
5	352.5	114.5
6	114.5	65.5
7	160.2	43
8	406	404
9	45.5	51
10	121	113.5

expo 2011 B C D E F G



$$\mu_1 = \mu_2$$



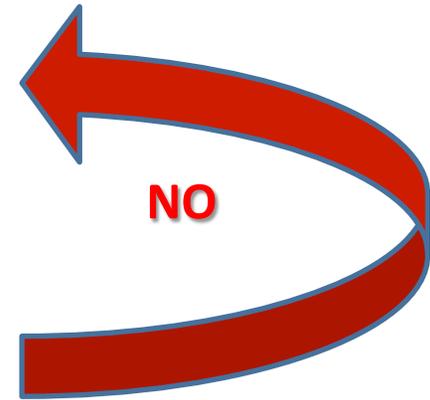
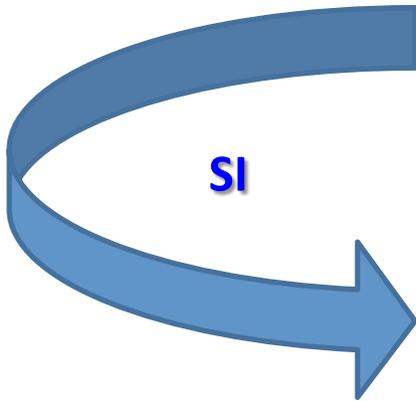
Prueba de medianas
(W de Mann-Whitney)

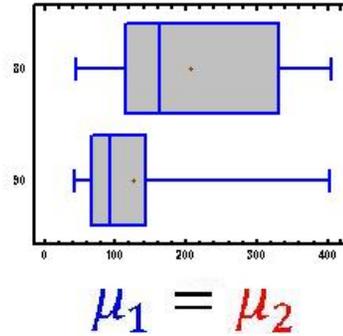
¿NORMALIDAD?

Prueba de medias
(t)

NO

SI





Prueba t

$$t = \frac{(\bar{x}_1 - \bar{x}_2)}{s_p \sqrt{\frac{1}{n_1} + \frac{1}{n_2}}}$$

$$t = \frac{(\bar{x}_1 - \bar{x}_2)}{\sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}}}$$

Comparación de Dos Muestras - 80 & 90

Comparación de Desviaciones Estándar

	80	90
Desviación Estándar	129.142	106.485
Varianza	16677.6	11339.1
Gl	9	9

Razón de Varianzas= 1.4708

Intervalos de confianza del 95.0%

Desviación Estándar de 80: [88.8283; 235.763]

Desviación Estándar de 90: [73.2443; 194.4]

Razones de Varianzas: [0.365327; 5.92145]

Prueba-F para comparar Desviaciones Estándar

Hipótesis Nula: $\sigma_1 = \sigma_2$

Hipótesis Alt.: $\sigma_1 \neq \sigma_2$

F = 1.4708 valor-P = 0.574654

No se rechaza la hipótesis nula para alfa = 0.05.

0.575

Comparación de Dos Muestras - 80 & 90

Comparación de Medias

Intervalos de confianza del 95.0% para la media de 80: 206.67 +/- 92.3827 [114.237; 299.053]

Intervalos de confianza del 95.0% para la media de 90: 124.6 +/- 76.1751 [48.4249; 200.775]

Intervalos de confianza del 95.0% intervalo de confianza para la diferencia de medias sin suponer varianzas iguales: 82.07 +/- 111.494 [-29.4239; 193.564]

Prueba t para comparar medias

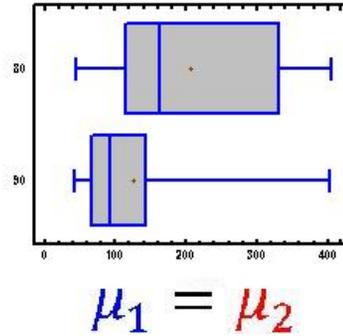
Hipótesis nula: $\mu_1 = \mu_2$

Hipótesis Alt.: $\mu_1 \neq \mu_2$

sin suponer varianzas iguales: t = 1.55051 valor-P = 0.139046

No se rechaza la hipótesis nula para alfa = 0.05.

0.139



Prueba W de Mann-Whitney

$$W = n_1 n_2 + \frac{n_1(n_1 + 1)}{2} - T_x$$

Comparación de Dos Muestras - 80 & 90

Comparación de Medianas
Mediana de muestra 1: 163.6
Mediana de muestra 2: 93.75

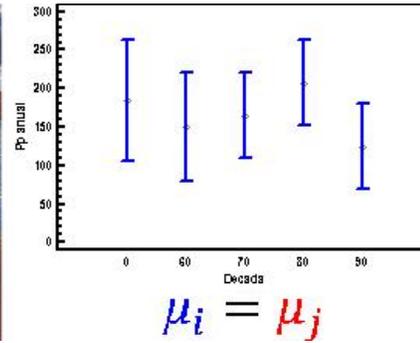
Prueba W de Mann-Whitney (Wilcoxon) para comparar medianas
Hipótesis Nula: mediana1 = mediana2
Hipótesis Alt.: mediana1 <> mediana2

Rango Promedio de muestra 1: 12.65
Rango Promedio de muestra 2: 8.35

W = 28.5 valor-P = 0.112275

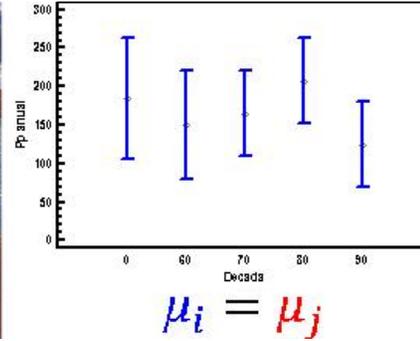
No se rechaza la hipótesis nula para alfa = 0.05.

0.112



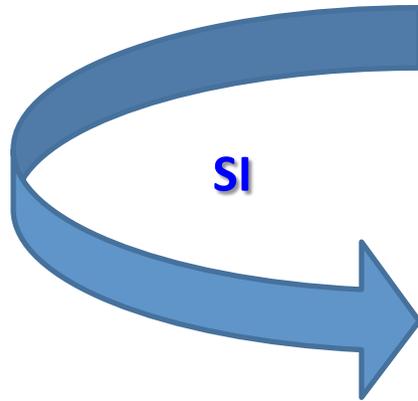
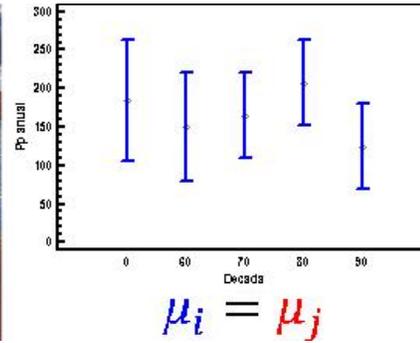
¿Hay diferencias estadísticamente significativas?

3. COMPARACIÓN DE MÁS MUESTRAS



Comparación por décadas

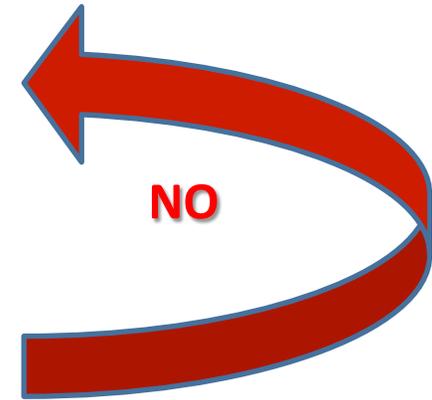
	Decada	Pp anual
1	60.0	139
2	60.0	331.9
3	60.0	94
4	60.0	61.5
5	60.0	36.5
6	60.0	237
7	70.0	73.5
8	70.0	124
9	70.0	24.4
10	70.0	21

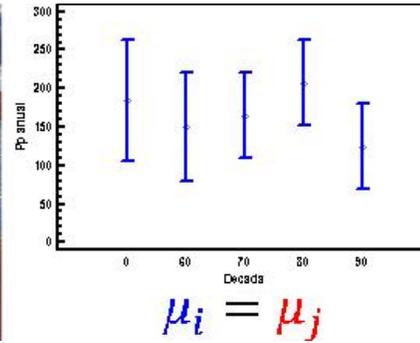


Prueba de medianas
(Kruskal-Wallis)

¿NORMALIDAD
Y
HOMOCEASTICIDAD?

Prueba de medias
(Andeva)





Prueba de Homocedasticidad

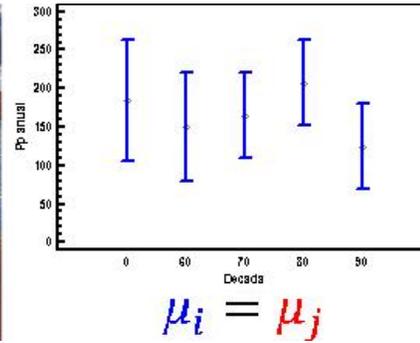
$$Z_{\bar{y}} = |y_{\bar{y}} - \bar{y}_j|$$

ANOVA Simple - Pp anual por Decada

Verificación de Varianza

	Prueba	Valor-P
Levene's	0.586834	0.67423

0.674



Andeva

ANOVA Simple - Pp anual por Decada

Tabla ANOVA para Pp anual por Decada

Fuente	Suma de Cuadrados	Gl	Cuadrado Medio	Razón-F	Valor-P
Entre grupos	36957.5	4	9239.38	0.62	0.6527
Intra grupos	538372.	36	14954.8		
Total (Corr.)	575329.	40			

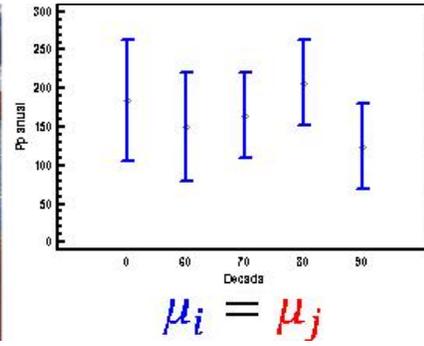
ANOVA Simple - Pp anual por Decada

Pruebas de Múltiple Rangos para Pp anual por Decada

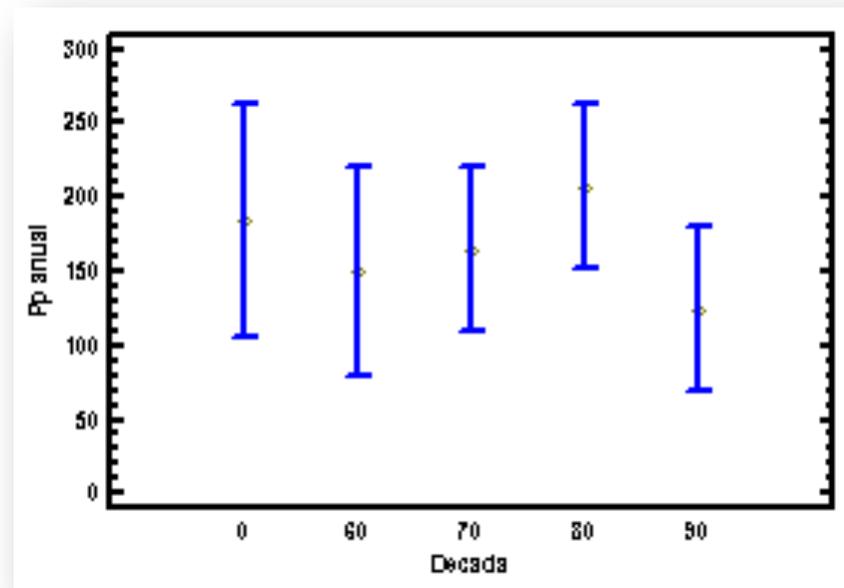
Método: 95.0 porcentaje LSD

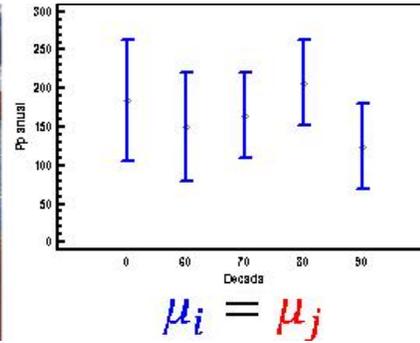
Decada	Casos	Media	Grupos Homogéneos
90	10	124.6	X
60	6	149.983	X
70	10	164.41	X
0	5	184.6	X
80	10	206.67	X

0.653

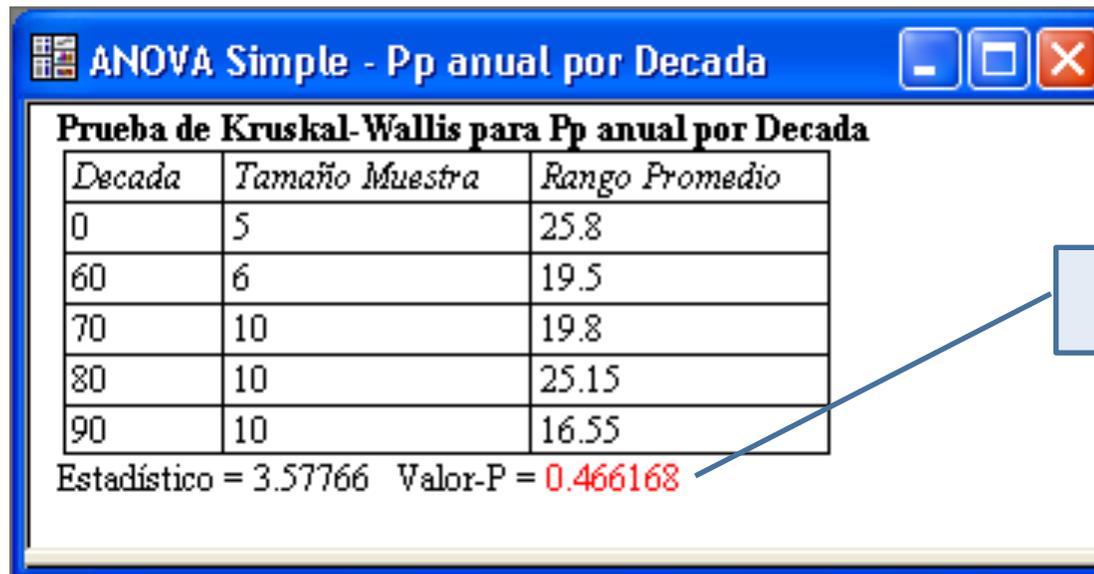


Andeva

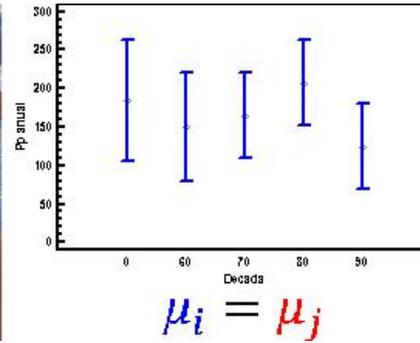




Kruskal-Wallis

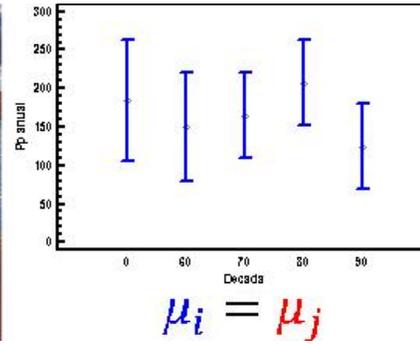


0.466



¿Hay diferencias estadísticamente significativas?

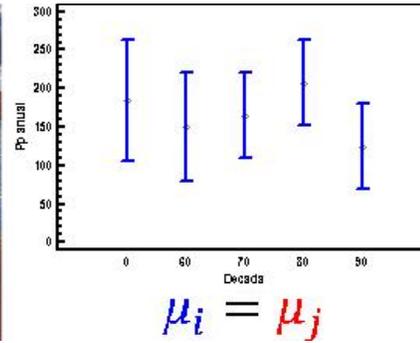
4 COMPARACIÓN DE MÁS MUESTRAS REGISTRO DE DIVERSAS ESTACIONES



Comparación por décadas

	estación	decada_1	Pp anual_1	Co: ▲
1	EL TRAPICHE	70	3	
2	EL TRAPICHE	80	64.5	
3	EL TRAPICHE	80	24	
4	EL TRAPICHE	80	46.1	
5	EL TRAPICHE	80	67.3	
6	EL TRAPICHE	80	93.1	
7	EL TRAPICHE	80	27	

expo 2011 | B | C | D | E | F | G | H | I | J | K



Andeva

0.297

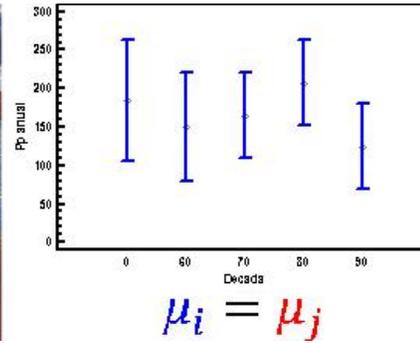
ANOVA Multifactorial - Pp anual_1

Análisis de Varianza para Pp anual_1 - Suma de Cuadrados Tipo III

Fuente	Suma de Cuadrados	Gl	Cuadrado Medio	Razón-F	Valor-P
EFECTOS PRINCIPALES					
A:decada_1	48026.8	4	12006.7	1.24	0.2974
B:estación	335444.	4	83860.9	8.64	0.0000
RESIDUOS	1.54348E6	159	9707.4		
TOTAL (CORREGIDO)	1.91751E6	167			

Todas las razones-F se basan en el cuadrado medio del error residual

0.000



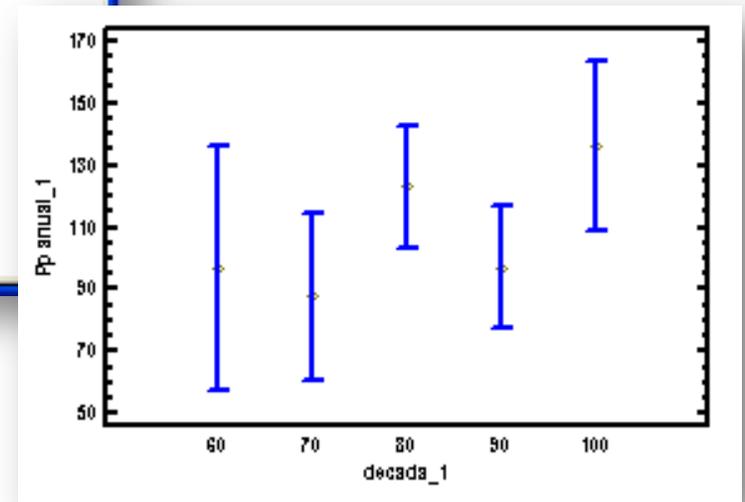
Andeva

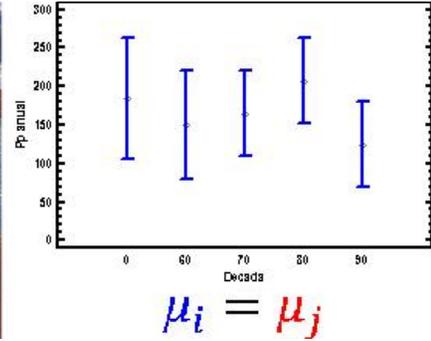
ANOVA Multifactorial - Pp anual_1

Pruebas de Múltiple Rangos para Pp anual_1 por decada_1

Método: 95.0 porcentaje LSD

decada_1	Casos	Media LS	Sigma LS	Grupos Homogéneos
70	29	87.675	19.2007	X
60	14	97.0187	28.1524	X
90	50	97.046	13.9337	X
80	50	122.868	13.9337	X
100	25	135.976	19.7052	X





Andeva

ANOVA Multifactorial - Pp anual_1

Pruebas de Múltiple Rangos para Pp anual_1 por estación

Método: 95.0 porcentaje LSD

estación	Casos	Media LS	Sigma LS	Grupos Homogéneos
EL TRAPICHE	26	41.7093	20.7477	X
ALMENDRAL	41	88.4541	15.489	X
LA SERENA (EA)	34	89.561	17.952	X
LA ORTIGA	26	153.186	20.7477	X
LA LAGUNA EMB.	41	167.673	15.5784	X

