



PRINCIPIOS DE INFERENCIA ESTADÍSTICA

Klgo. Iván A. Rodríguez Núñez MSc, PhD

Principios de inferencia

- La inferencia pretende generalizar la información contenida en unos pocos datos a un cuerpo de conocimiento mas amplio.
- Si no existiera variabilidad, la inferencia no tendría sentido.
- Evolución epistemológica de la construcción del conocimiento.

• Teoría de muestreo:

• Estudia las relaciones existentes entre la distribución de un carácter en dicha población y las distribuciones de dicho carácter en sus muestras.

BASES DE LA INFERENCIA

Muestra → Población

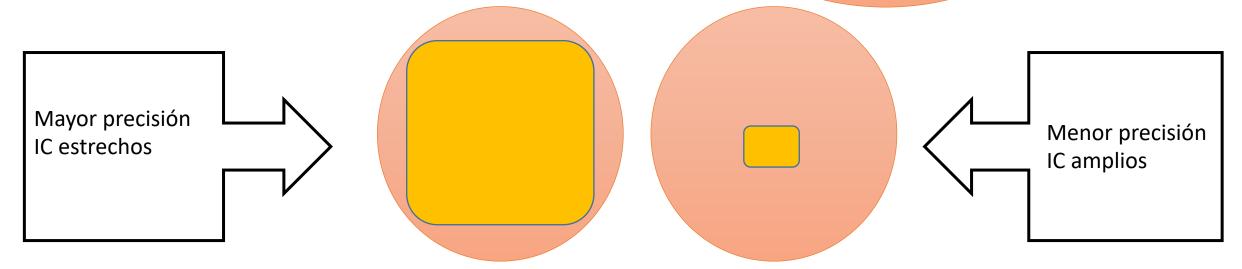
(Relación de probabilidad)

Muestra > Población

- ✓ El conjunto de todos los individuos elegibles se denomina "universo" o "población diana".
- ✓ El marco muestral corresponde a los sujetos accesibles.

Población diana (todos quienes cumplen criterios de elegibilidad)

Población accesible

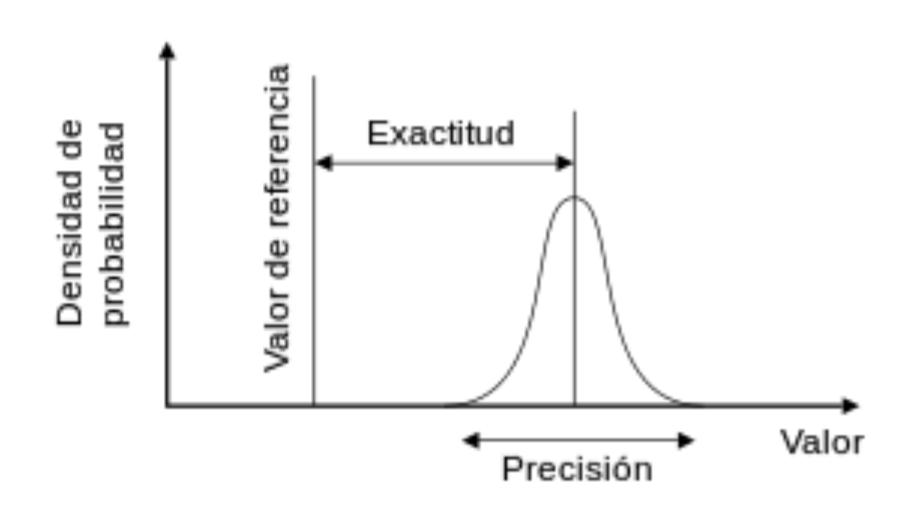


"La frecuencia relativa de un suceso tiende a estabilizarse en torno a un número, a medida que el número de pruebas del experimento crece indefinidamente". (Jakob Bernoulli, 1654 – 1705)

Estimación y muestras representativas

- ✓ La representatividad de la muestra es una exigencia para objetivos descriptivos. Ej: ¿Cuál es el colesterol medio en la población?, ¿Qué porcentaje de personas de la población realiza actividad física 3 veces por semana?.
- ✓ En objetivos analíticos lo importante es que los resultados sean validos. Ej: Estudios comparativos.
- Validez interna: Consiste en que lo hallado coincida con la verdad en la muestra que se estudia (metodología del estudio).
- Validez externa: consiste en que lo hallado coincida con la verdad de la población diana (aplicabilidad del estudio).

Precisión y validez



Axiomas y propiedades de la probabilidad

- Ley de Laplace: "La probabilidad de obtener un resultado será igual al cociente entre casos favorables y casos posibles".
 - Ej: probabilidad de obtener un número par en un dado.

Casos favorables: 3

Casos posibles: 6

Probabilidad 50%

Método de estimación deductiva de probabilidades



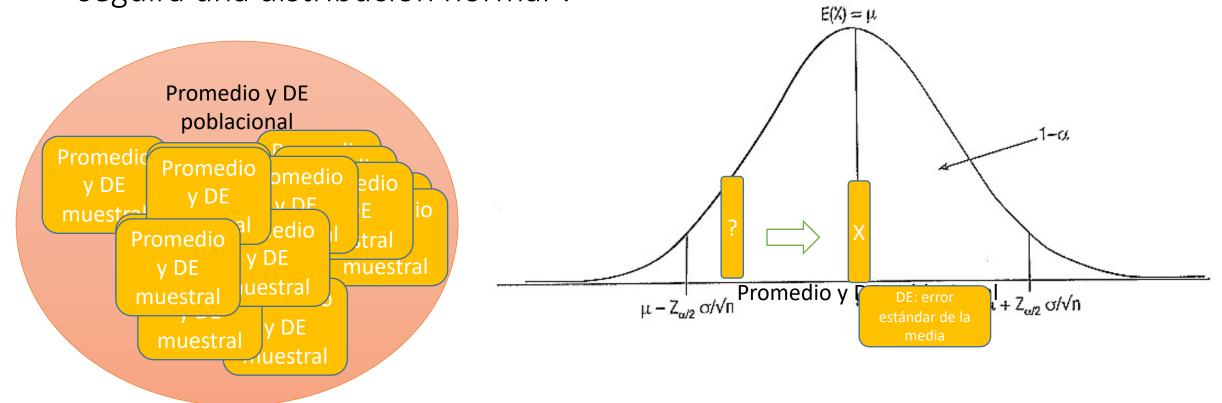
Prevalencia Incidencia



Probabilidad de ocurrencia de un evento de interés (Riesgo absoluto)

Axiomas y propiedades de la probabilidad

• Teorema del límite central: "La distribución de los estimadores que se calculan en sucesivas muestras que se obtengan de la población seguirá una distribución normal".



Errores de estimación

		POBLACIÓN					
		p>0,05 (H ₀)	p<0,05 (H ₁)				
TRA	p>0,05 (H ₀)	Confiabilidad	Error β				
MUESTRA	p<0,05 (H ₁)	Error α	1-β (Potencia)				

Errores de estimación

ERROR ALFA (Tipo 1) Y BETA (Tipo 2)

ERROR ALFA:

- ✓ Equivale al error tipo I.
- ✓ Probabilidad de obtener resultados estadísticamente significativos en la muestra, cuando no lo son en la población.
- ✓ La máxima probabilidad de error tipo 1 aceptado corresponde al 5%.
- ✓ El complemento equivale a la confiabilidad (95%).
- ✓ Si el experimento se repite 100 veces, en 95 los resultados deben encontrarse dentro del intervalo de confianza.

Errores de estimación

ERROR ALFA (Tipo 1) Y BETA (Tipo 2)

• ERROR **BETA**:

- ✓ Equivale al error tipo II.
- ✓ Probabilidad de obtener resultados no estadísticamente significativos en la muestra, cuando en realidad si lo son.
- ✓ Error tipo I usualmente aceptado oscila entre 20 y 30%.
- ✓ El complemento equivale a la potencia estadística (1-β)

Inferencia y estimación del tamaño muestral

- La necesidad de realizar estimación de tamaño muestral se genera desde:
 - ✓ La necesidad de saber la **muestra mínima necesaria** para obtener resultados **estadísticamente significativos** ≠ clínicamente significativos.
 - ✓ Existencia de un riesgo latente de obtener resultados significativos cuando realmente no lo son.
 - ✓ Riesgo latente de trabajar con una muestra mayor a la mínima necesaria lo que incrementa los gastos del estudio.

Principios de inferencia

- Conceptos:
 - Estimador muestral.
 - Parámetro poblacional.
- Promedio muestral (x) \rightarrow Media poblacional (μ).
 - SI el promedio de x equivale a $\mu \rightarrow$ x es un **estimador insesgado de X**.
- Varianza muestral (s²) \rightarrow Varianza poblacional (σ^2)
 - El promedio de s² equivale a $\sigma^2 \rightarrow s^2$ es un **estimador insesgado de \sigma^2**.

Tipos de estimaciones

- Estimación puntual: Corresponde al valor observado de un estimador en una muestra. No provee variabilidad inherente de la estimación.
- Estimación por intervalo: Provee un rango razonable de valores que puede adoptar el parámetro de interés de acuerdo con la información muestral.

Error:
$$e = (1,96)\sqrt{\frac{\sigma^2}{n}}(1-f)$$

Intervalo de confianza (X-e, X+e)

 $1,96= Z_{\alpha}$ (0,05). $\sigma^2 = Varianza poblacional$ n= Tamaño muestral.f = Fracción de muestreo

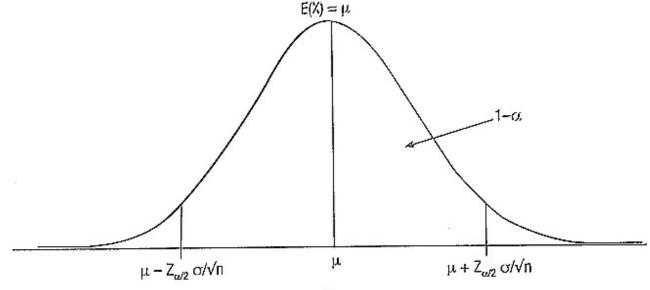


Figura 6-1 Distribución del estimador \overline{X} alrededor del parámetro $E(X) = \mu$. Una proporción $1-\alpha$ de las posibles medias muestrales \overline{X} está incluida entre los límites indicados.

• Ejemplo:

```
1,96= Z_{\alpha} (0,05).

\sigma^2 = Varianza.

n= Tamaño muestral.

f= Fracción de muestreo
```

$$e = (1,96)\sqrt{\frac{\sigma^2}{n}}(1-f)$$

Se desea inferir el promedio de VEF₁ de una población de pacientes con EPOC N=5000.

$$- x = 2 L$$

$$-\sigma VEF_1 = 0.6 L$$

Determine el error de muestreo e intervalo de confianza de la estimación de VEF₁.

$$e = (1,96)\sqrt{\frac{0,36}{100}}(1-0,02)$$
 $e = 0,11$ INTERVALO DE CONFIANZA (2,11 L - 1,88 L)

"Con un IC (1- α =95%) podemos afirmar que":

- El 95% de los casos se encuentra dentro del intervalo.
- Si se repitiera el proceso, el 95% de los casos estarían dentro del intervalo.
- Hay una probabilidad del 5% de que el parámetro no esté en el intervalo.
- Hay una confianza del 95% de que el parámetro esté en el intervalo.

Interdependencia entre el n y el IC.

n	95% Confidence Limits for μ	Length of Interval		
10	$\overline{X} \pm 0.620\sigma$	1.240σ		
100	$\overline{X} \pm 0.196\sigma$	0.392σ		
1000	$\overline{X} \pm 0.062\sigma$	0.124σ		

- Pero.....
- En estadística, el **parámetro** es una **constante**, y no una variable aleatoria.
- En investigación generalmente se miden variables aleatorias.
 - S
 - X

Intervalo 95 % de
$$\overline{X} = \mu \pm \frac{1,96\sigma}{\sqrt{n}}$$

Distribución t de Student

- Sustituir el parámetro σ por el estimador S implica:
 - Sustituir una constante por una variable aleatoria.
 - Las variables aleatorias poseen una distribución de valores denominada t de Student.

$$Z = \frac{\overline{X} - \mu}{\sigma / \sqrt{n}}$$

$$t = \frac{\overline{X} - \mu}{\sqrt{n}}$$

Distribución t de Student

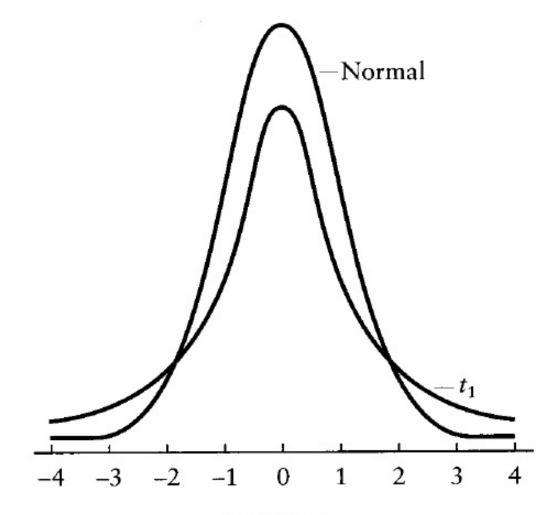


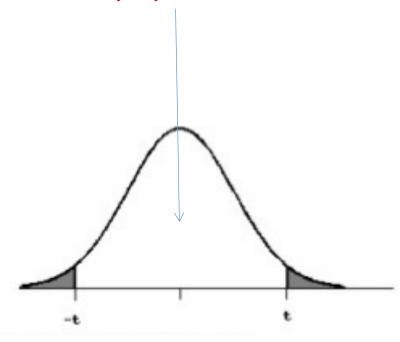
FIGURE 9.2

The standard normal distribution and Student's t distribution with 1 degree of freedom

- (a) El área de las dos colas está sombreada en la figura.
- (b) Si H_A es direccional, las cabeceras de las columnas deben ser divididas por 2 cuando se acota el P-valor.

gl	ÁREA DE DOS COLAS								
	0,20	0,10	0,05	0,02	0,01	0,001	0,0001		
1	3,078	6,314	12,706	31,821	63,657	636,619	6366,198		
2	1,886	2,920	4,303	6,695	9,925	31,598	99,992		
3	1,638	2,353	3,182	4,541	5,841	12,924	28,000		
4	1,533	2,132	2,776	3,747	4,604	8,610	15,544		
5	1,476	2,015	2,571	3,365	4,032	6,869	11,178		
6	1,440	1,943	2,447	3,143	3,707	5,959	9,082		
7	1,415	1,895	2,365	2,998	3,499	5,408	7,885		
8	1,397	1,860	2,306	2,896	3,355	5,041	7,120		
9	1,383	1,833	2,262	2,821	3,250	4,781	6,594		
10	1,372	1,812	2,228	2,764	3,169	4,587	6,211		
11	1,363	1,796	2,201	2,718	3,106	4,437	5,921		
12	1,356	1,782	2,179	2,681	3,055	4,318	5,694		
13	1,350	1,771	2,160	2,650	3,012	4,221	5,513		
14	1,345	1,761	2,145	2,624	2,977	4,140	5,363		
15	1,341	1,753	2,131	2,602	2,947	4,073	5,239		
16	1,337	1,746	2,120	2,583	2,921	4,015	5,134		
17	1,333	1,740	2,110	2,567	2,898	3,965	5,044		
18	1,330	1,734	2,101	2,552	2,878	3,922	4,966		
19	1,328	1,729	2,093	2,539	2,861	3,883	4,897		
20	1,325	1,725	2,086	2,528	2,845	3,850	4,837		
21	1,323	1,721	2,080	2,518	2,831	3,819	4,784		
22	1,321	1,717	2,074	2,508	2,819	3,792	4,736		
23	1,319	1,714	2,069	2,500	2,807	3,767	4,693		
24	1,318	1,711	2,064	2,492	2,797	3,745	4,654		
25	1,316	1,708	2,060	2,485	2,787	3,725	4,619		
26	1,315	1,706	2,056	2,479	2,779	3,707	4,587		
27	1,314	1,703	2,052	2,473	2,771	3,690	4,558		
28	1,313	1,701	2,048	2,467	2,763	3,674	4,530		
29	1,311	1,699	2,045	2,462	2,756	3,659	4,506		
30	1,310	1,697	2,042	2,457	2,750	3,646	4,482		
40	1,303	1,684	2,021	2,423	2,704	3,551	4,321		
60	1,296	1,671	2,000	2,390	2,660	3,460	4,169		
100	1,290	1,660	1,984	2,364	2,626	3,390	4,053		
140	1,288	1,656	1,977	2,353	2,611	3,361	4,006		
00	1,282	1,645	1,960	2,326	2,576	3,291	3,891		

En la distribución t de Student, el 95% de IC depende de los grados de libertad (n-1).



En distribución normal el 95% IC se encuentra a ±1,96 DS de la media (valor Z)

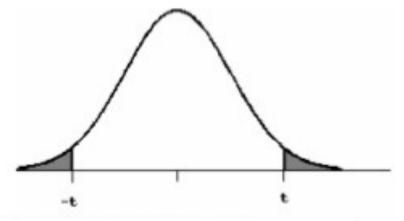
Distribución t de Student

• Ejemplo:

- Considerando 10 niños con BO post infecciosa que poseen una Pimax de 68 cmH₂O, DS: 12 cmH₂O.
- Varianza poblacional es desconocida.
- Calcular el 95% de intervalo de confianza para la variable Pimax.

- (a) El área de las dos colas está sombreada en la figura.
- (b) Si H_A es direccional, las cabeceras de las columnas deben ser divididas por 2 cuando se acota el P-valor.

	ÁREA DE DOS COLAS								
gl	0,20	0,10	0,05	0,02	0,01	0,001	0,0001		
1	3,078	6,314	12,706	31,821	63,657	636,619	6366,198		
2	1,886	2,920	4,303	6,695	9,925	31,598	99,992		
3	1,638	2,353	3,182	4,541	5,841	12,924	28,000		
4	1,533	2,132	2,776	3,747	4,604	8,610	15,544		
5	1,476	2,015	2,571	3,365	4,032	6,869	11,178		
6	1,440	1,943	2,447	3,143	3,707	5,959	9,082		
7	1,415	1,895	2,365	2,998	3,499	5,408	7,885		
8	1,397	1,860	2,306	2,896	3,355	5,041	7,120		
9	1,383	1,833		3,321	3, 250	4,781	6,59		
10	1,372	1,812	2,226	2,764	3,169	4,587	6,21		
11	1,363	1,796	2,201	2,718	3,106	4,437	5,921		
12	1,356	1,782	2,179	2,681	3,055	4,318	5,694		
13	1,350	1,771	2,160	2,650	3,012	4,221	9,51		
14	1,345	1,761	2,145	2,624	2,977	4,140	5,363		
15	1,341	1,753	2,131	2,602	2,947	4,073	5,239		
16	1,337	1,746	2,120	2,583	2,921	4,015	5,134		
17	1,333	1,740	2,110	2,567	2,898	3,965	5,044		
18	1,330	1,734	2,101	2,552	2,878	3,922	4,966		
19	1,328	1,729	2,093	2,539	2,861	3,883	4,897		
20	1,325	1,725	2,086	2,528	2,845	3,850	4,83		
21	1,323	1,721	2,080	2,518	2,831	3,819	4,784		
22	1,321	1,717	2,074	2,508	2,819	3,792	4,736		
23	1,319	1,714	2,069	2,500	2,807	3,767	4,693		
24	1,318	1,711	2,064	2,492	2,797	3,745	4,65		
25	1,316	1,708	2,060	2,485	2,787	3,725	4,619		
26	1,315	1,706	2,056	2,479	2,779	3,707	4,587		
27	1,314	1,703	2,052	2,473	2,771	3,690	4,558		
28	1,313	1,701	2,048	2,467	2,763	3,674	4,530		
29	1,311	1,699	2,045	2,462	2,756	3,659	4,500		
30	1,310	1,697	2,042	2,457	2,750	3,646	4,482		
40	1,303	1,684	2,021	2,423	2,704	3,551	4,321		
60	1,296	1,671	2,000	2,390	2,660	3,460	4,169		
100	1,290	1,660	1,984	2,364	2,626	3,390	4,053		
140	1,288	1,656	1,977	2,353	2,611	3,361	4,000		
00	1,282	1,645	1,960	2,326	2,576	3,291	3,891		



$$\Rightarrow \left(\overline{X} - 2.262 \frac{s}{\sqrt{10}}, \overline{X} + 2.262 \frac{s}{\sqrt{10}}\right).$$



$$(68 - 2.262 \frac{12}{\sqrt{10}}; 68 + 2,262 \frac{12}{\sqrt{10}})$$

(59,42; 76,5)

Con un 95% de confianza, el valor real se encuentra en ese intervalo.

Test de hipótesis

- Hipótesis nula= ausencia de diferencia.
 - $H_0 = \mu = \mu_0$

- Hipótesis alternativa= diferencias significativas.
 - $H_A = \mu \neq \mu_0$

Dado que H₀ es verdadero, el test debe verificar la probabilidad de que la hipótesis nula sea cierta.

Se debe evaluar si existe suficiente evidencia para aceptar o rechazar H₀.

SE ACEPTA LA HIPOTESIS NULA...CASO CERRADO¡¡¡¡



Test de hipótesis

- Hipótesis nula= ausencia de diferencia.
 - $H_0 = \mu = \mu_0$

- Hipótesis alternativa= diferencias significativas.
 - $H_A = \mu \neq \mu_0$

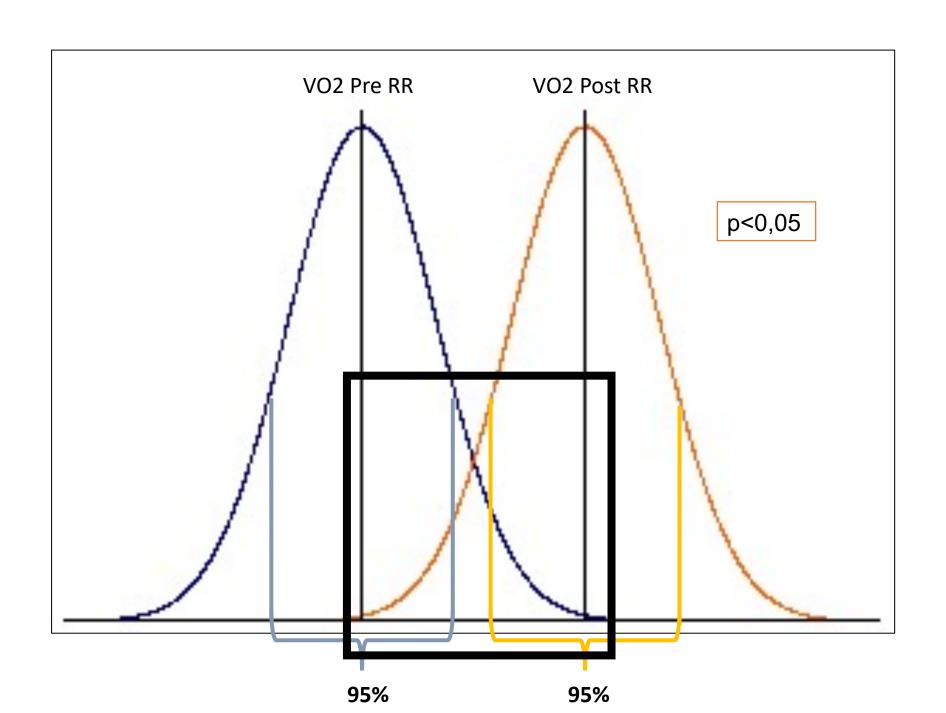
Inicialmente se considera H₀ como verdadero, hasta que el test diga lo contrario.

Test de hipótesis

- Hipótesis de dos caras (two sided test of hypotheses).
 - $H_A = \mu \neq \mu_0$
- Hipótesis de una cara (one sided test of hypotheses).
 - $H_A = \mu < /> \mu_0$

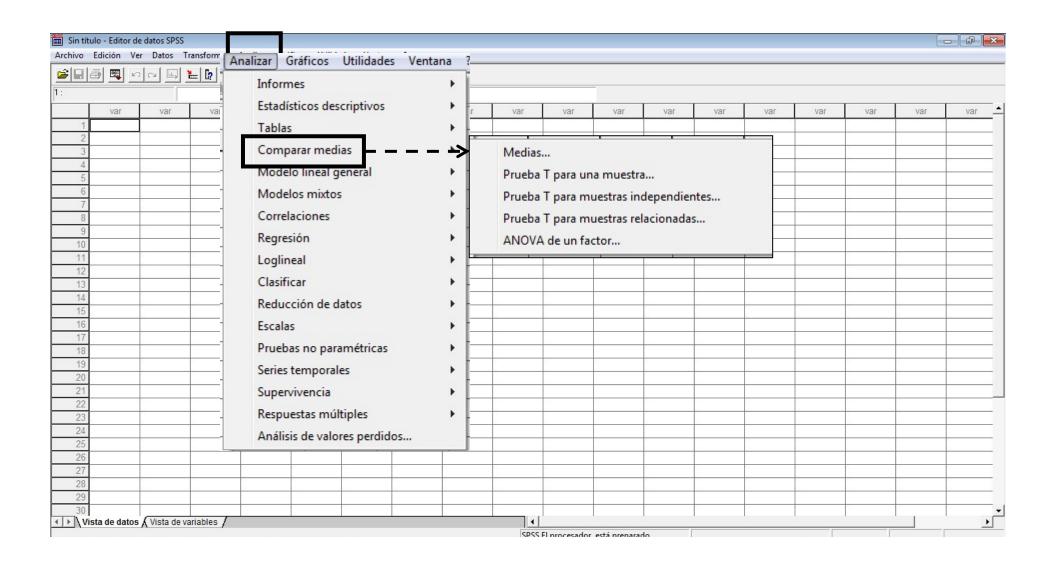
Test de hipótesis para variables cuantitativas

- t test para muestras pareadas.
- t test para muestras independientes.
- NO OLVIDAR....
 - p Value= Probabilidad de que la diferencia obtenida sea producto del azar.
 - Se acepta por consenso un valor de p < 0,05.

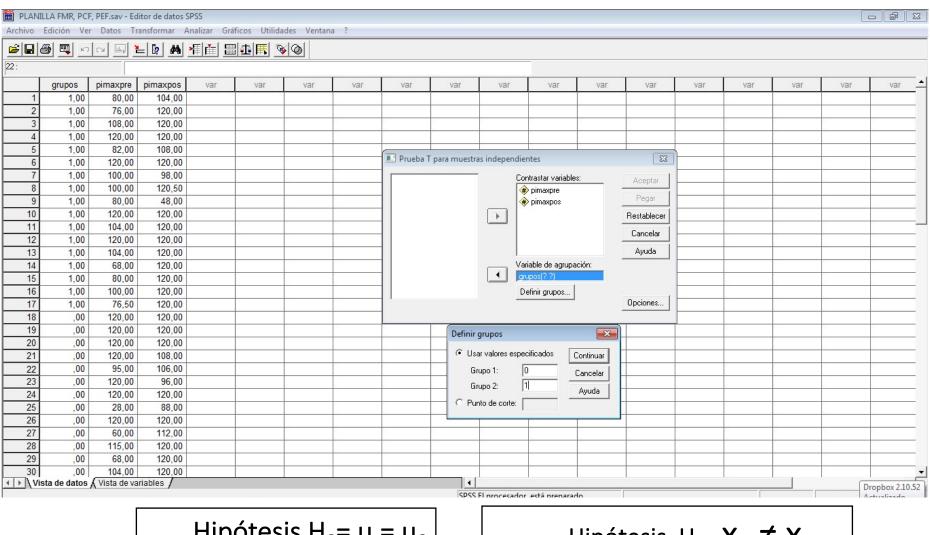


Test de hipótesis para variables cuantitativas

- t test para 2 muestras pareadas.
 - Ej: Diseños de investigación antes y después en el mismo grupo.
- t test para 2 muestras independientes.
 - Presencia de grupo control y experimental
- Análisis de varianza (ANOVA)



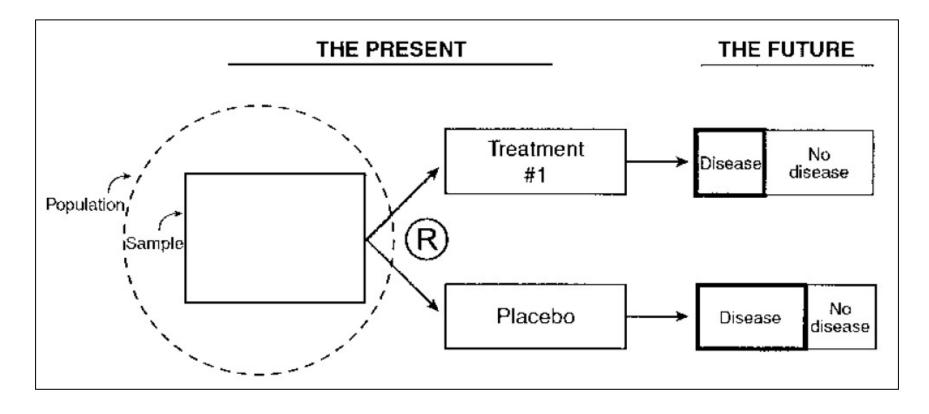
T test para muestras independientes



Hipótesis $H_0 = \mu = \mu_0$

Hipótesis $H_A = X_1 \neq X_0$

T test para muestras independientes



Hipótesis
$$H_0 = \mu = \mu_0$$

Hipótesis $H_A = x_1 \neq x_0$

T test para muestras independientes

Estadísticos de grupo

	Grupo	N	Media	Desviación típ.	Error típ. de la media
PIMAXPRE	control	23	99,6087	28,63999	5,97185
	experimental	17	96,3824	17,94047	4,35120
PIMAXPOS	control	23	108,1304	16,98744	3,54213
	experimental	17	112,8529	18,02755	4,37232

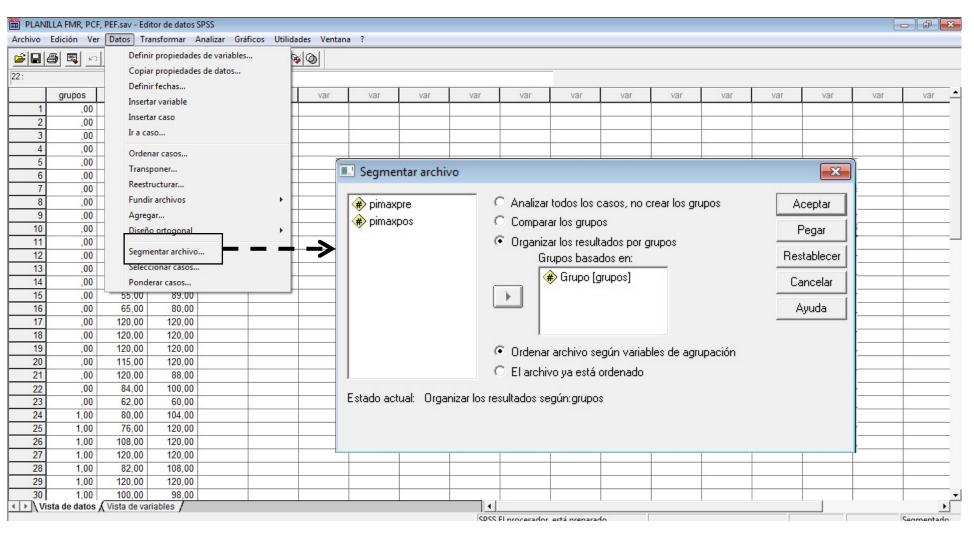
Prueba de muestras independientes

		Prueba de para la igu v aria	ualdad de	Prueba T para la igualdad de medias						
							Dif erencia	Error típ. de	95% Intervalo de confianza para la diferencia	
		F	Sig.	t	gl	Sig. (bilateral)	de medias	la diferencia	Inferior	Superior
PIMAXPRE	Se han asumido v arianzas iguales	5,250	,028	,408	38	,685	3,2263	7,90222	-12,77087	19,22356
	No se han asumido v arianzas iguales			,437	37, 159	,665	3,2263	7,38891	-11,74284	18, 19553
PIMAXPOS	Se han asumido v arianzas iguales	,552	,462	-, 847	38	,402	-4,7225	5,57587	-16,01026	6,56525
	No se han asumido v arianzas iguales			-, 839	33,423	,407	-4,7225	5,62707	-16,16535	6,72034

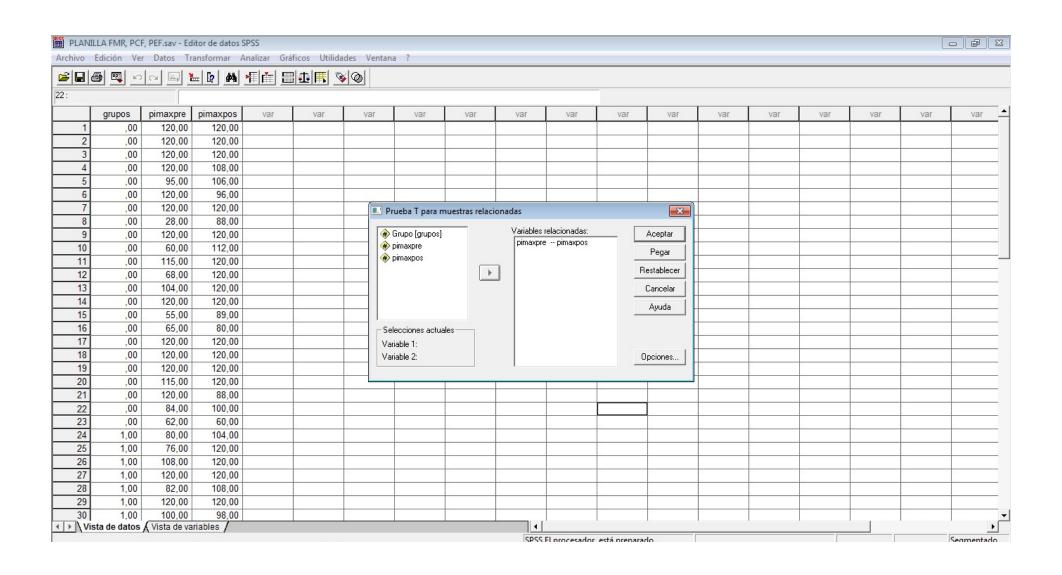
Hipótesis $H_0 = \mu = \mu_0$

T test para muestras pareadas

Primero es necesario segmentar archivo. De lo contrario se analizará ambos grupos como uno solo (pre vs post)



T test para muestras pareadas



T test para muestras pareadas

Estadísticos de muestras relacionadas

		Media	N	Des viación tí p.	Error típ. de la media
Par 1	PIMAXPRE	96,3824	17	17,94047	4,35120
	PIMAXPOS	112,8529	17	18,02755	4,37232

a. Grupo = experimental

Prueba de muestras relacionadas

		Dif ere	ncias relaciona	das				
		Desviación	Error típ. de	confianz	erv alo de a para la encia			
	Media	típ.	la media	Inferior	Superior	t	gl	Sig. (bilateral)
Par 1 PIMAXPRE - PIMAXP	DS -16,4706	21,32595	5,17230	-27,4354	-5,5058	-3,184	16	,006

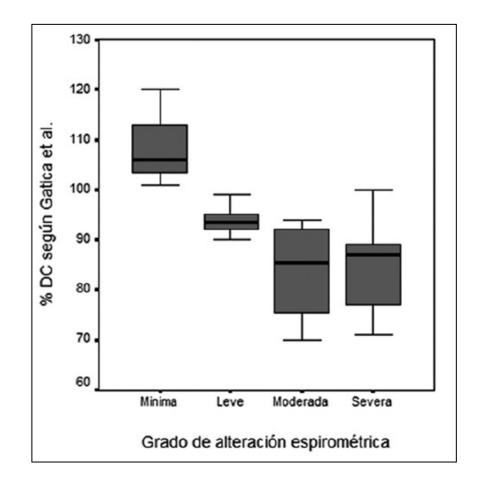
a. Grupo = experimental

Asociación entre una variable categórica y una variable continua (ANOVA)

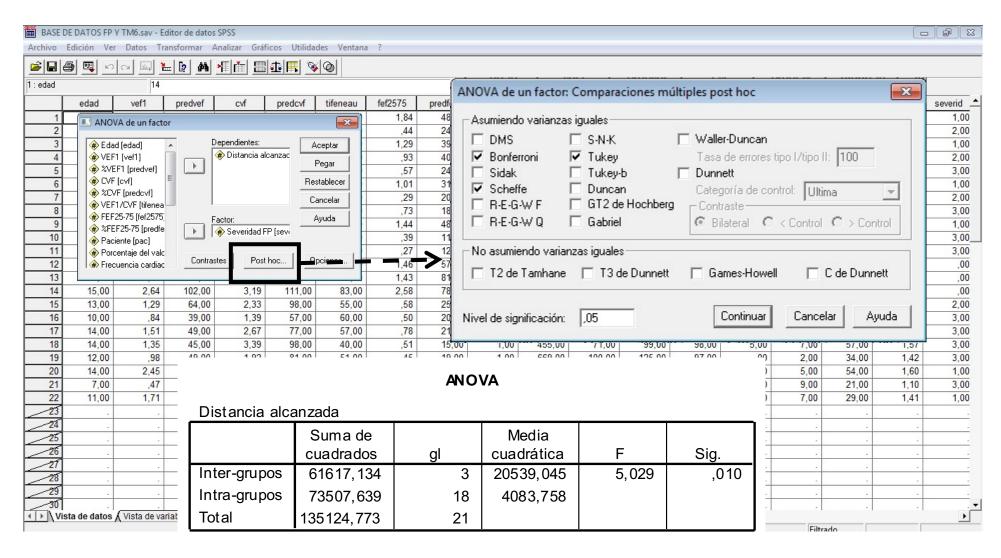
TRABAJO ORIGINAL

Test de caminata de seis minutos y función pulmonar en pacientes con bronquiolitis obliterante post infecciosa

IVÁN RODRÍGUEZ N.*,**,****, SCARLETT HENRÍQUEZ J.**,*****, PAULINA VÁSQUEZ M.**,***** y DANIEL ZENTENO A.*,***,*****.



Asociación entre una variable categórica y una variable continua (ANOVA)



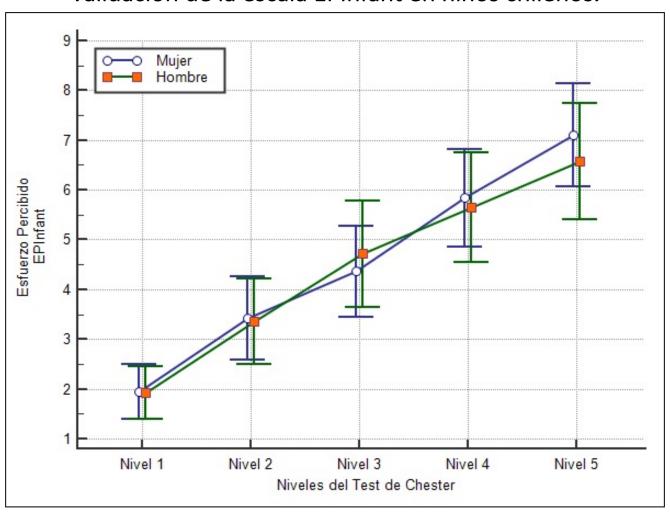
Comparaciones múltiples

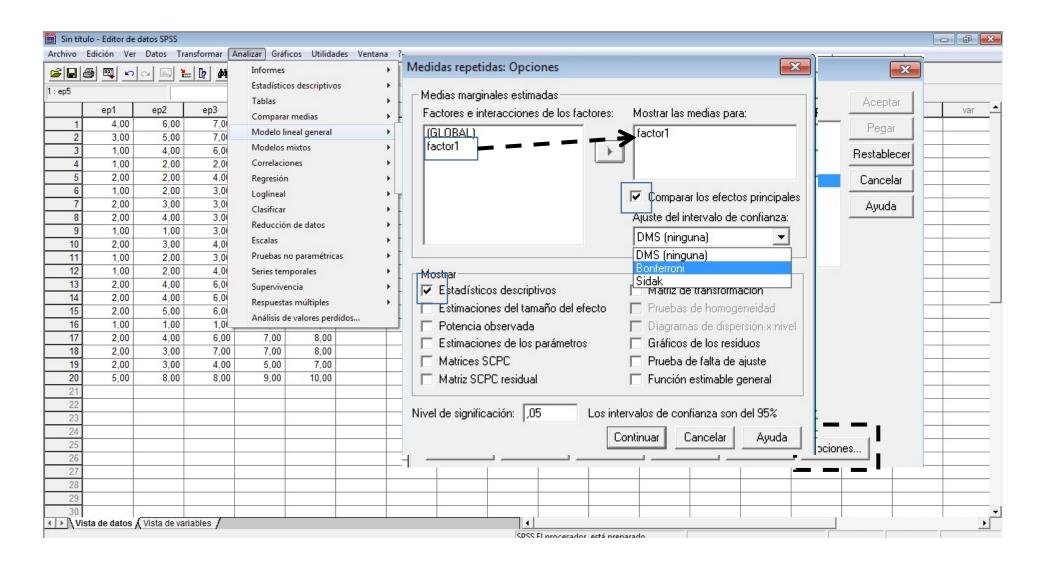
Variable dependiente: Distancia alcanzada

Valiable depond	diente: Distancia ai	ourizada		i		Intervalo de co 95%	
			Diferencia de				Límite
	(I) Sev eridad FP	(J) Sev eridad FP	medias (I-J)	Error típico	Sig.	Lí mite inferior	superior
HSD de Tukey	Minima	Lev e	64,0000	45, 18715	,506	-63,7120	191,7120
		Moderada	172,9167*	48,80770	,011	34,9719	310,8614
		Sev era	117,5556	42,60286	,057	-2,8525	237,9636
	Lev e	Minima	-64,0000	45,18715	,506	-191,7120	63,7120
		Moderada	108,9167	41,25004	,072	-7,6679	225,5012
		Sev era	53, 5556	33,68052	,409	-41,6354	148,7465
	Moderada	Minima	-172,9167*	48,80770	,011	-310,8614	-34,9719
		Lev e	-108,9167	41,25004	,072	-225,5012	7,6679
		Sev era	-55,3611	38,40170	,491	-163,8954	53, 1732
	Sev era	Minima	-117,5556	42,60286	,057	-237,9636	2,8525
		Lev e	-53,5556	33,68052	,409	-148,7465	41,6354
		Moderada	55, 3611	38,40170	,491	-53,1732	163,8954
Scheffé	Minima	Lev e	64,0000	45, 18715	,582	-75,1274	203,1274
		Moderada	172,9167*	48,80770	,021	22,6419	323,1915
		Sev era	117,5556	42,60286	,089	-13,6151	248,7262
	Lev e	Minima	-64,0000	45,18715	,582	-203,1274	75, 1274
		Moderada	108,9167	41,25004	,109	-18,0887	235,9221
		Sev era	53, 5556	33,68052	,488	-50,1439	157,2550
	Moderada	Minima	-172,9167*	48,80770	,021	-323,1915	-22,6419
		Lev e	-108,9167	41,25004	,109	-235,9221	18,0887
		Sev era	-55,3611	38,40170	,568	-173,5967	62,8745
	Sev era	Minima	-117,5556	42,60286	,089	-248,7262	13,6151
		Lev e	-53,5556	33,68052	,488	-157,2550	50, 1439
		Moderada	55, 3611	38,40170	,568	-62,8745	173,5967
Bonf erroni	Minima	Lev e	64,0000	45, 18715	1,000	-69,8775	197,8775
		Moderada	172,9167*	48,80770	,014	28,3124	317,5209
		Sev era	117,5556	42,60286	,077	-8,6654	243,7765
	Lev e	Minima	-64,0000	45,18715	1,000	-197,8775	69,8775
		Moderada	108,9167	41,25004	,100	-13,2962	231,1296
		Sev era	53,5556	33,68052	,775	-46,2309	153,3420
	Moderada	Minima	-172,9167*	48,80770	,014	-317,5209	-28,3124
		Lev e	-108,9167	41,25004	,100	-231,1296	13,2962
		Sev era	-55,3611	38,40170	,999	-169,1351	58,4129
	Sev era	Minima	-117,5556	42,60286	,077	-243,7765	8,6654
		Lev e	-53,5556	33,68052	,775	-153,3420	46,2309
		Moderada	55,3611	38,40170	,999	-58,4129	169,1351

^{*.} La diferencia entre las medias es significativa al nivel .05.

Validación de la escala EPInfant en niños chilenos.





Comparaciones por pares

Medida:	MEASURE	1

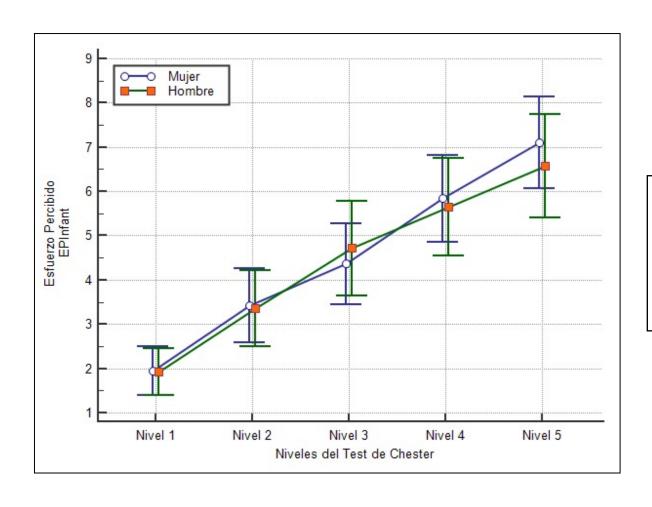
IVIEUTUA. IVIEAS						
		Dif erencia			Intervalo de cor % para dife	
		entre				Lí mite
(I) FACTOR1	(J) FACTOR1	medias (I-J)	Error típ.	Significación ^a	Límite inferior	superior
1	2	-1,450*	,211	,000	-1,892	-1,008
	3	-2,700*	,317	,000	-3,364	-2,036
	4	-3,700*	,333	,000	-4,397	-3,003
	5	-4,950*	,432	,000	-5,854	-4,046
2	1	1,450*	,211	,000	1,008	1,892
	3	-1,250*	,250	,000	-1,773	-,727
	4	-2,250*	,239	,000	-2,751	-1,749
	5	-3,500*	,344	,000	-4,220	-2,780
3	1	2,700*	,317	,000	2,036	3,364
	2	1,250*	,250	,000	,727	1,773
	4	-1,000*	,145	,000	-1,304	-, 696
	5	-2,250*	,331	,000	-2,944	-1,556
4	1	3,700*	,333	,000	3,003	4,397
	2	2,250*	,239	,000	1,749	2,751
	3	1,000*	,145	,000	,696	1,304
	5	-1,250*	,260	,000	-1,795	-, 705
5	1	4,950*	,432	,000	4,046	5,854
	2	3,500*	,344	,000	2,780	4,220
	3	2,250*	,331	,000	1,556	2,944
	4	1,250*	,260	,000	,705	1,795

Basadas en las medias marginales estimadas.

^{*-} La diferencia de las medias es significativa al nivel ,05.

a. Ajuste para comparaciones múltiples: Diferencia menos significativa (equivalente a la ausencia de ajuste).

Validación de la escala EPInfant en niños chilenos.

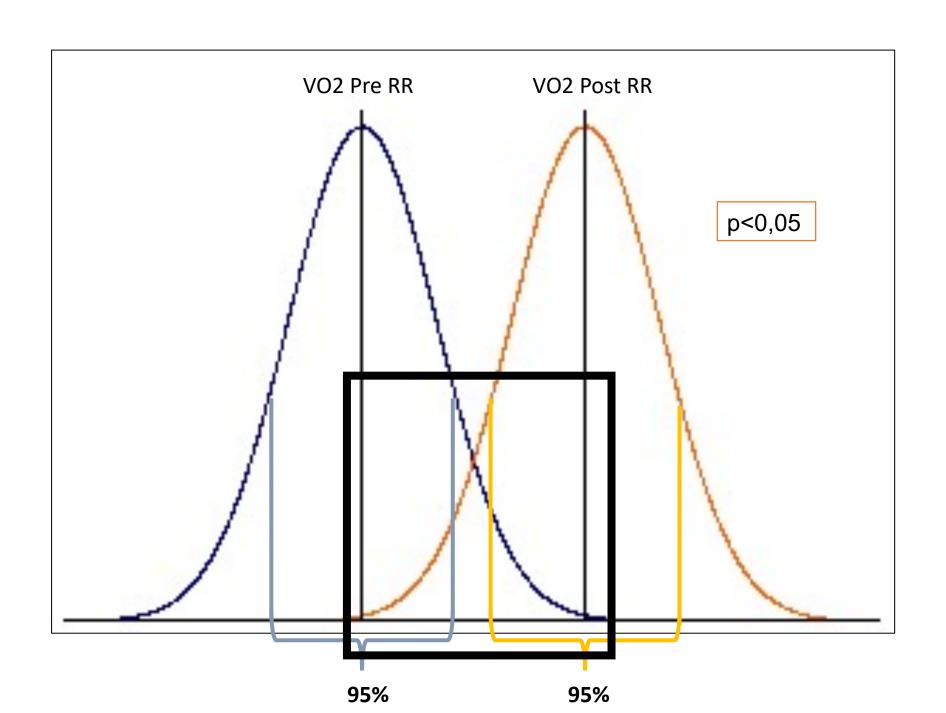


Existe suficiente evidencia para confirmar la existencia de diferencia estadísticamente significativa entre el EP medido en cada nivel del test de Chester.

En resumen

- t test para 2 muestras pareadas.
 - Ej: Diseños de investigación antes y después en el mismo grupo.
- t test para 2 muestras independientes.
 - Presencia de grupo control y experimental.
- Análisis de varianza (ANOVA)
 - Asociación entre una variable categórica y una cuantitativa.
 - Análisis de medidas repetidas.

IMPORTANTE RECORDAR¡¡¡¡¡Para el empleo de t test y ANOVA debe cumplirse el paradigma de normalidad en la distribución de los datos, por lo tanto en la métodología se debe declarar la prueba de normalidad empleada.



¿Y si no existe distribución normal?... Pruebas no paramétricas.

- Test de los signos de Wilcoxon.
 - Muestras pareadas. (Equivale a t de Student para muestras pareadas).
- Test U de Mann Whitney.
 - Muestras independientes. (Equivale a t de Student para muestras independientes).
- Kruskal Wallis.
 - Asociación entre una variable categórica y una cuantitativa. (Equivale a ANOVA).

Principios.....

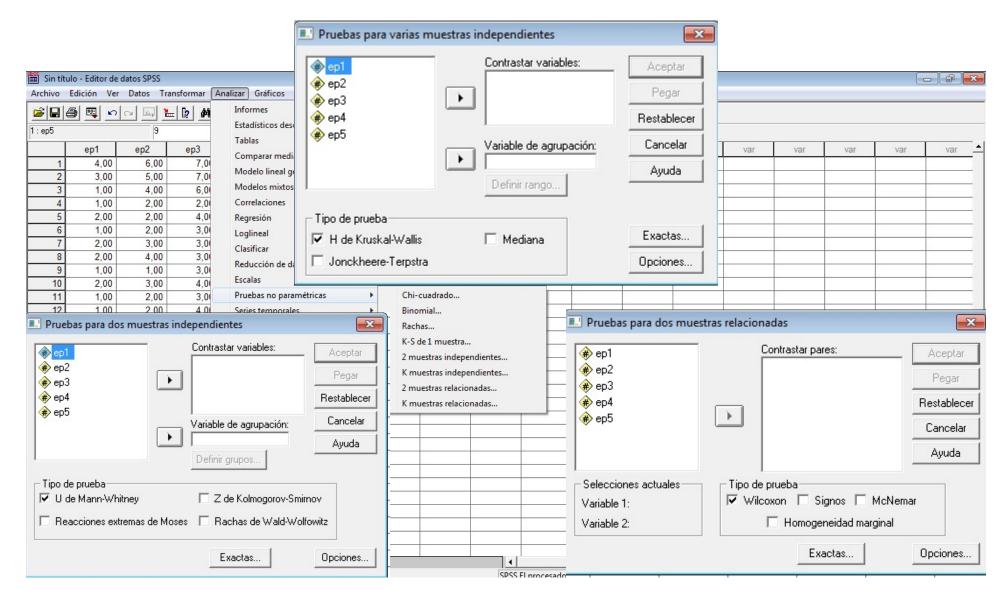
• Se dispone de observaciones de una misma variable X en dos poblaciones.

Población 1:
$$x_1^1, x_2^1, ..., x_{n_1}^1$$

Población 2: $x_1^2, x_2^2, ..., x_{n_2}^2$

- Se ordenan las observaciones de menor a mayor asignándole un rango.
- Si hay empate, se asigna el promedio de los rangos asignados como si no lo hubiera.
- Se evalúa la diferencia de la suma de rangos entre los grupos.

¿Y si no existe distribución normal?... Pruebas no paramétricas.



Test de hipótesis para variables categóricas

- Chi cuadrado de Pearson (X²).
- Test exacto de Fisher.

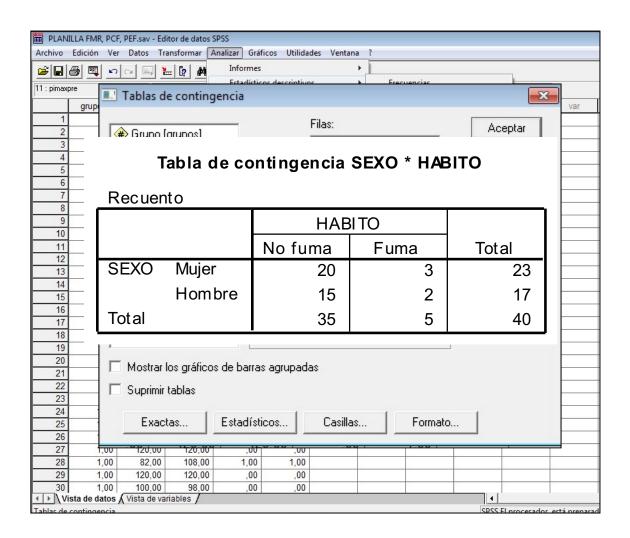
Chi cuadrado de Pearson (X²)

- Evalúa la asociación o independencia entre variables categóricas.
- Es necesaria la construcción de tablas de contingencia.

Tipo de consulta	Invierno	Verano	Total
Accidentes	1 418 (a)	2 221 (b)	3 639
No accidentes	20 133 (c)	14 269 (d)	34 402
Total	21 551	16 490	38 041
Tasa de incidencia §	6,6%	13,5%	9,6%

Chi cuadrado de Pearson (X²)

TABLA DE CONTINGENCIA



Chi Cuadrado de Pearson (X²)

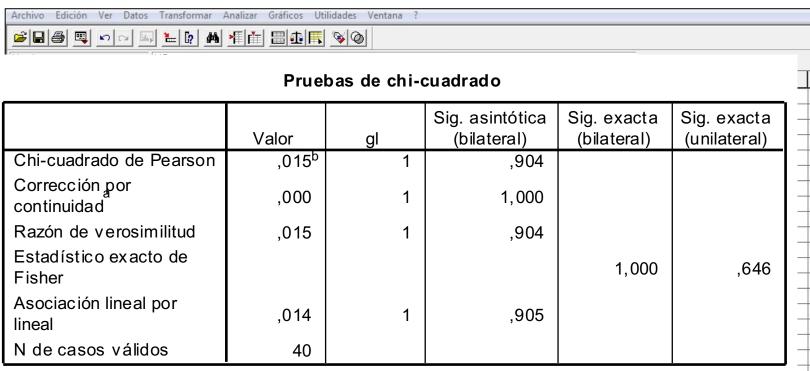
HIPOTESIS NULA (H₀)

- Hipótesis de independencia entre variables.
- La proporción de sujetos que presentan y no presentan el evento de interés es igual a la proporción de sujetos que ha sido expuesta y no ha sido expuesta al factor de exposición.
- p>0,05

• HIPOTESIS ALTERNATIVA (H₁)

- Hipótesis de asociación entre variables.
- Las proporciones de personas que han sido expuestas al evento de interés y factor de exposición son las mismas.
- P<0,05

Chi cuadrado de Pearson (X²)



- a. Calculado sólo para una tabla de 2x2.
- b. 2 casillas (50,0%) tienen una frecuencia esperada inferior a 5. La frecuencia mínima esperada es 2,13.

25	1,00	76,00	120,00	,00	,00		0	(a)	2	() ()	
26	1,00	108,00	120,00	,00	,00						
27	1,00	120,00	120,00	,00	,00	10					
28	1,00	82,00	108,00	1,00	1,00						
29	1,00	120,00	120,00	,00	,00						
30	1,00	100,00	98,00	,00	,00		0				

Test de hipótesis para variables categóricas

CONSIDERACIONES:

- El calculo de Ji-cuadrado de Pearson se basa en calcular las diferencias entre observaciones "observadas" y esperadas cuando existe condición de independencia (50/50).
- Para considerar correcta la significancia calculada por Ji-cuadrado en cada casilla deben existir mas de 5 valores.
- Si en alguna casilla existen menos de 5 valores se sugiere utilizar test exacto de Fisher.

Test de hipótesis para 2 variables cuantitativas.

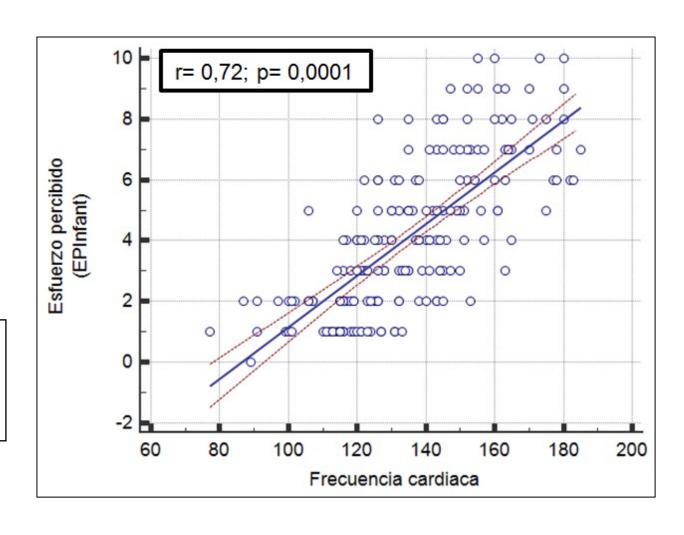
- Estudia la asociación entre dos variables cuantitativas y el sentido de la correlación.
 - Coeficiente de correlación r de Pearson \rightarrow muestras con distribución normal.
 - Coeficiente de correlación rho de Spearman

 muestras sin distribución normal.
- Predicción de una variable en función de la otra/s
 - Regresión lineal simple.
 - Regresión lineal múltiple.

La forma mas sencilla de evaluar la correlación de dos variables cuantitativas es mediante el grafico de correlación o dispersión.

"Scatter Plot"

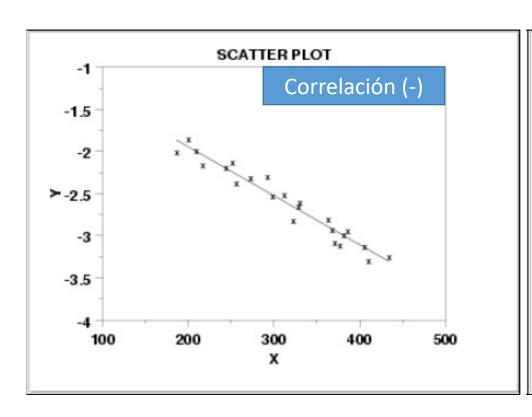
No permite cuantificar la fuerza de asociación

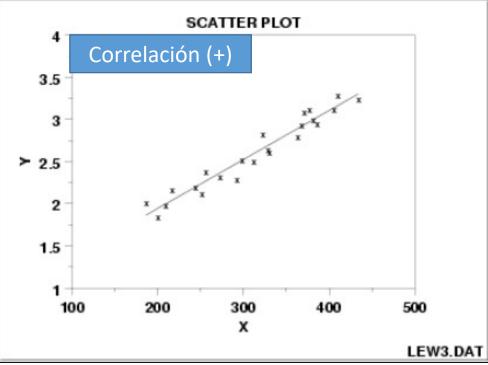


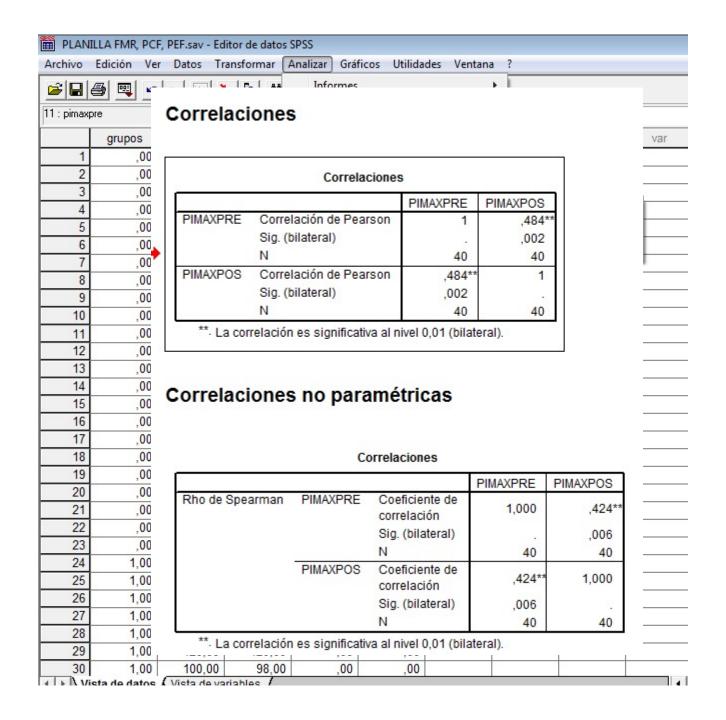
El calculo del coeficiente de correlación r de Pearson permite:

- comprobar que existe una relación lineal entre dos variables aleatorias, antes de proceder al análisis de regresión.
- Resumir en un solo número (r) la intensidad de la relación lineal entre estas dos variables.
- El coeficiente de correlación r no debe utilizarse para.
- Establecer relaciones causales entre dos variables.
- suplantar el análisis de regresión.
- analizar la coherencia entre mediciones.

- El coeficiente de correlación r no tiene unidades.
- El intervalo de posibles valores de r es: −1≤ r ≤1.
- En el análisis de la correlación no es aplicable la distinción entre variable "dependiente" "independiente" como sucede en el modelo de regresión: aquí ambas variables son "dependientes".







Regresión lineal simple

• ¿Es posible predecir el Pef tos a partir de la Pemax en adolescentes sanos?

pimax	pemax	fem	peftos	abd
80,00	48,00	450,00	250,00	23,00
76,00	72,00	400,00	330,00	20,00
108,00	48,00	480,00	400,00	18,00
120,00	84,00	390,00	350,00	15,00
82,00	84,00	520,00	550,00	26,00
120,00	60,00	390,00	390,00	26,00
100,00	80,00	600,00	470,00	23,00
100,00	60,00	200,50	330,00	17,00
80,00	40,00	200	300,00	22,00
120,00	80,00	460,00	490,00	26,00
104,00	84,00	540,00	480,00	29,00
120,00	80,00	470,00	490,00	23,00
104,00	68,00	410,00	350,00	32,00
68,00	60,00	370,00	400,00	22,00
80,00	100,00	370,00	400,00	23,00
100,00	72,00	390,00	410,00	28,00
76,50	60,00	440,00	430,00	27,00
120,00	120,00	560,00	400,00	34,00
120,00	104,00	600,00	500,00	23,00
120.00	116.00	500.00	330.00	26.00

Regresión lineal simple

Primero ¿Existe asociación entre la Pemax y PEF tos?

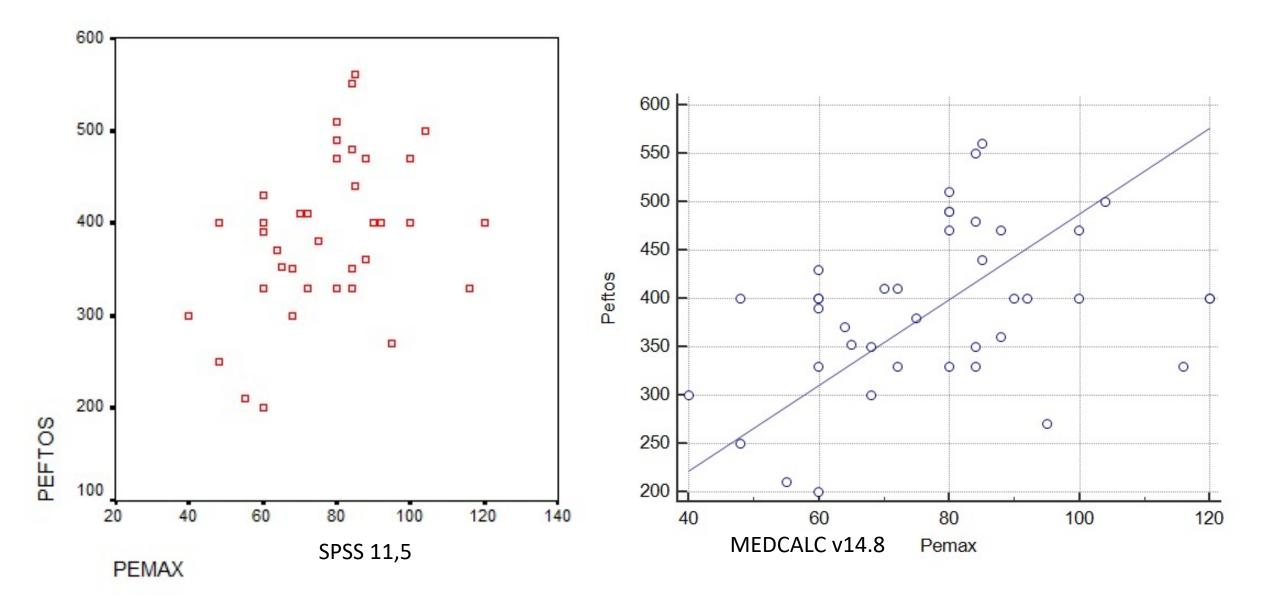
Correlaciones

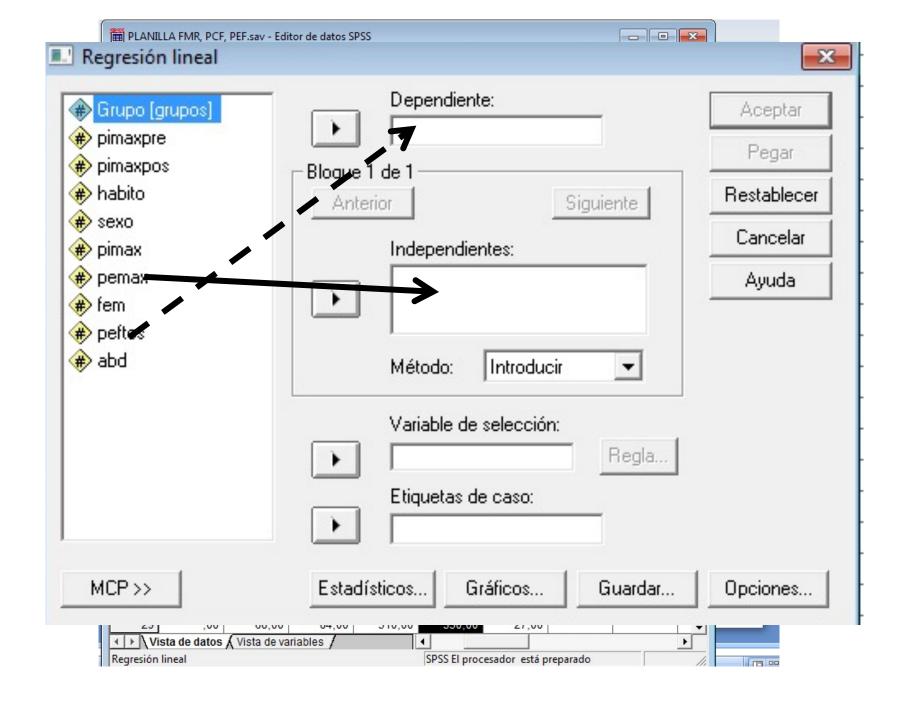
		PEMAX	PEFTOS
PEMAX	Correlación de Pearson	1	,357*
	Sig. (bilateral)		,024
	N	40	40
PEFTOS	Correlación de Pearson	,357*	1
	Sig. (bilateral)	,024	
	N	40	40

^{*} La correlación es significante al nivel 0,05 (bilateral).

Regresión lineal simple

- En el análisis de regresión existe:
 - Variable respuesta (dependiente)→ Y
 - Variable explicativa, predictora (independiente) \rightarrow X
- Pemax: variable predictora (independiente).
- PEFtos: variable respuesta (dependiente).
- $Y = a + bX \rightarrow PEF tos = a + bPemax$.







→ Regresión

Variables introducidas/eliminadasb

Modelo	Variables introducidas	Variables eliminadas	Método
1	PEMAXa		Introducir

- a. Todas las variables solicitadas introducidas
- b. Variable dependiente: PEFTOS

Resumen del modelo

Modelo	R	R cuadrado	R cuadrado corregida	Error típ. de la estimación
1	,357ª	,127	,104	79,72760

a. Variables predictoras: (Constante), PEMAX

ANOVAb

Modelo	**************************************	Suma de cuadrados	gl	Media cuadrática	F	Sig.
1	Regresión	35193,786	1	35193,786	5,537	,024ª
1979	Residual	241546,614	38	6356,490	-11,10000	100000
	Total	276740,400	39		22	S.

- a. Variables predictoras: (Constante), PEMAX
- b. Variable dependiente: PEFTOS

Coeficientes

		Coeficientes no estandarizados		Coeficientes estandarizad os			Intervalo de cor B al 95	•
								Límite
Modelo		В	Error típ.	Beta	t	Sig.	Límite inferior	superior
1	(Constante)	267,058	53,872		4,957	,000	158,001	376,116
	PEMAX	1,578	,671	,357	2,353	,024	,220	2,936

a. Variable dependiente: PEFTOS

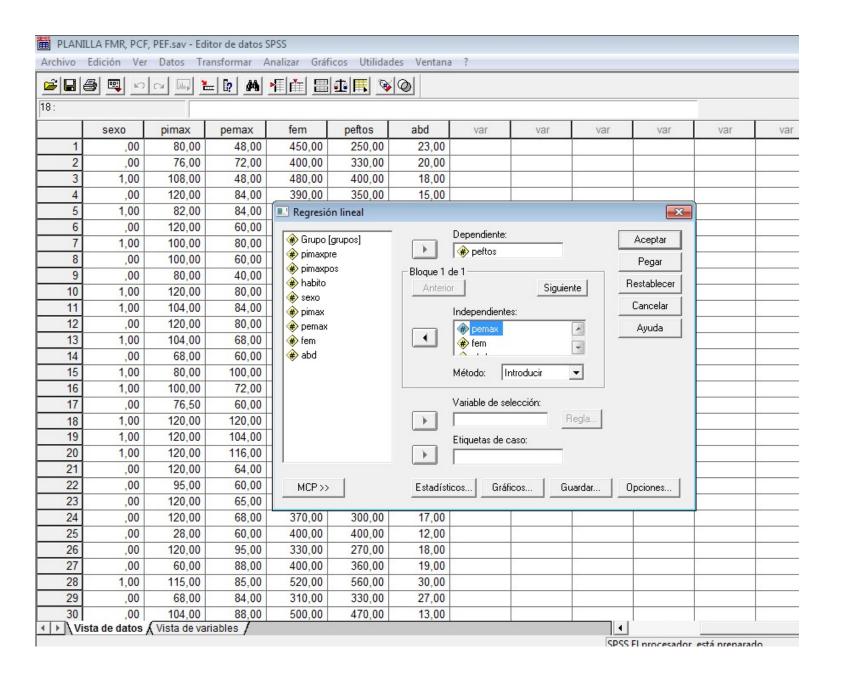
Pef tos= 267,058+1,578Pemax.

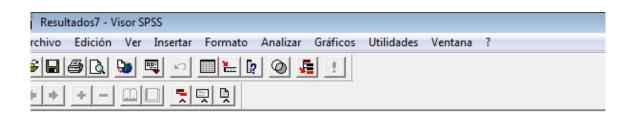
Si un adolescente tiene una Pemax de 90 cmH2O, ¿Cuál será el PEF tos estimado?.

PEF tos= 267,058+1,578×90 El PEF tos sería 409,078 L/Min

Regresión lineal múltiple

- En el análisis de regresión existe:
 - 1 Variable respuesta (dependiente) → Y
 - 2 o mas variables explicativas, predictoras (independientes) > X1, X2, X3, etc.
- Pemax: variable predictora 1 (indepediente)
- FEM: variable predictora 2 (indepediente)
- Abd: variable predictora 3 (independiente)
- Peftos: variable respuesta (dependiente).
- $Y = a + bX1 + cX2 + dX3 \rightarrow$
 - PEF tos= a+bPemax+cFEM+dAbd





→ Regresión

Variables introducidas/eliminadasb

Modelo	Variables introducidas	Variables eliminadas	Método
1	ABD, FEM, PEMAX	2/	Introducir

a. Todas las variables solicitadas introducidas

b. Variable dependiente: PEFTOS

Resumen del modelo

400000000	9151		R cuadrado	Error típ. de la
Modelo	R	R cuadrado	corregida	estimación
1	,750ª	,563	,527	57,96158

a. Variables predictoras: (Constante), ABD, FEM, PEMAX

ANOVA^b

Modelo		Suma de cuadrados	gl	Media cuadrática	F	Sig.
1	Regresión	155796,797	3	51932,266	15,458	,000a
	Residual	120943,603	36	3359,545	1,000	1.02
	Total	276740,400	39			

a. Variables predictoras: (Constante), ABD, FEM, PEMAX

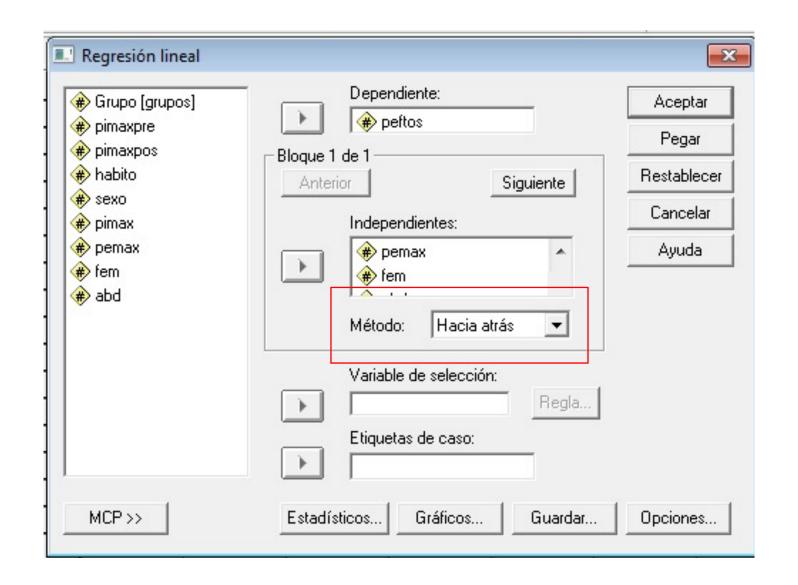
b. Variable dependiente: PEFTOS

Coeficientes

		Coeficientes no estandarizados		Coeficientes estandarizad os				tervalo de confianza para B al 95%	
Modelo		В	Error típ.	Beta	t	Sig.	Lí mite inferior	Lí mite superior	
1	(Constante)	90,266	51,621	Dota	1,749	,089	-14,425	194,958	
	PEMAX	-, 095	,571	-, 022	-, 167	,868	-1,253	1,062	
	FEM	,647	,115	,718	5,632	,000	,414	,879	
	ABD	1,564	1,742	,108	,898	,375	-1,970	5,097	

a. Variable dependiente: PEFTOS

• Para que el modelo sea significativo, todas las variables predictoras deben ser significativas.



Variables introducidas/eliminadas

Modelo	Variables introducidas	Variables eliminadas	Método
1	ABD, F _a EM, PEMAX		Introducir
2		PEMAX	Hacia atrás (criterio: Prob. de F para eliminar >= ,100).
3	·	ABD	Hacia atrás (criterio: Prob. de F para eliminar >= ,100).

a. Todas las variables solicitadas introducidas

b. Variable dependiente: PEFTOS

Resumen del modelo

			R cuadrado	Error típ. de la
Modelo	R	R cuadrado	corregida	estimación
1	,750 ^a	,563	,527	57,96158
2	,750 ^b	,563	,539	57, 19505
3	,744 ^c	,553	,541	57,04595

a. Variables predictoras: (Constante), ABD, FEM, PEMAX

b. Variables predictoras: (Constante), ABD, FEM

c. Variables predictoras: (Constante), FEM

ANOVA^d

Modelo		Suma de cuadrados	gl	Media cuadrática	F	Sig.
1	Regresión	155796,797	3	51932,266	15,458	,000 ^a
	Residual	120943,603	36	3359,545		
	Total	276740,400	39			
2	Regresión	155703,255	2	77851,627	23,799	,000 ^b
	Residual	121037, 145	37	3271,274		
	Total	276740,400	39			
3	Regresión	153079,250	1	153079,250	47,040	,000 ^c
	Residual	123661,150	38	3254,241		
	Total	276740,400	39			

a. Variables predictoras: (Constante), ABD, FEM, PEMAX

b. Variables predictoras: (Constante), ABD, FEM

c. Variables predictoras: (Constante), FEM

d. Variable dependiente: PEFTOS

Coeficientes

		Coeficientes no estandarizados		Coeficientes estandarizad os			Intervalo de confianza para B al 95%	
Modelo		В	Error típ.	Beta	t	Sig.	Lí mite inferior	Lí mite superior
1	(Constante)	90, 266	51,621		1,749	,089	-14,425	194,958
	PEMAX	-, 095	,571	-, 022	-, 167	,868	-1,253	1,062
	FEM	,647	,115	,718	5,632	,000	,414	,879
	ABD	1,564	1,742	,108	,898	,375	-1,970	5,097
2	(Constante)	87,723	48,667		1,803	,080	-10,886	186,332
	FEM	,639	,104	,709	6, 153	,000	,429	,849
	ABD	1,492	1,666	,103	,896	,376	-1,883	4,867
3	(Constante)	109,747	41,888		2,620	,013	24,949	194,545
	FEM	,670	,098	,744	6,859	,000	,472	,868

a. Variable dependiente: PEFTOS

Variables excluida§

						Estadísticos de
					Correlación	colinealidad
Modelo		Beta dentro	t	Sig.	parcial	Tolerancia
2	PEMAX	-, 022 ^a	-, 167	,868	-, 028	,730
3	PEMAX	,007 ^b	,057	,955	,009	,777
	ABD	,103 ^b	,896	,376	,146	,889

a. Variables predictoras en el modelo: (Constante), ABD, FEM

b. Variables predictoras en el modelo: (Constante), FEM

c. Variable dependiente: PEFTOS





PRINCIPIOS DE INFERENCIA ESTADÍSTICA

Klgo. Iván A. Rodríguez Núñez. MSc, PhD