Nociones de Estadística

Las mediciones tienen siempre asociadas un error experimental (inherente a la resolución del equipamiento empleado, a errores aleatorios y/o a errores sistemáticos). Esto hace que no se puedan sacar conclusiones con total certeza.

Lo que si podemos hacer es utilizar las herramientas que proporciona la estadística para aceptar conclusiones que tienen una alta probabilidad de ser correctas y rechazar aquellas que no. Existen dos tipos de descriptores (estadígrafos) que permiten hacer esta valoración: las medidas de tendencia central y las medidas de dispersión.

Medidas de tendencia central

Moda: Valor (o valores) que aparece con mayor frecuencia en un conjunto de datos.

Mediana: Valor que divide en dos partes iguales al conjunto ordenado de datos. Para un conjunto impar es el valor que ocupa la posición central, para un conjunto par es el promedio de los datos centrales

Media aritmética:

$$\overline{x} = \frac{\sum_{i=1}^{n} x_i}{n}$$

donde x_i representa al componente iésimo del conjunto de mediciones y n es el número total de mediciones

Propiedades de la media:
- Si las mediciones están formadas por un solo valor que se repite, la media es el propio valor.
- Si se suma o se resta una constante a todas las mediciones del conjunto, la media quedará aumentada o disminuida en esa cantidad.
- Si todas las mediciones se multiplican o dividen por una constante, la media quedará multiplicada o dividida por ella.
- La suma de las desviaciones de las mediciones del conjunto con respecto a su media es cero.

Media geométrica

$$G = \left(x_1 x_2 \cdots x_n\right)^{1/n}$$

No se puede usar si uno de los datos es cero o negativo.

Media armónica

$$\frac{1}{H} = \frac{1}{n}\sum_{i=1}^{n}\frac{1}{x_i} \qquad H = \frac{n}{\sum_{i=1}^{n} x_i}$$

Media ponderada: Promedio de las medias teniendo en cuenta el peso relativo (*w*) de cada una de ellas. Con la introducción de estos pesos se da diferente importancia a cada valor.

$$\overline{x} = \frac{\sum_{i=1}^{n} w_i x_i}{\sum_{i=1}^{n} w_i}$$

Medidas de dispersión

Rango o recorrido: Es la diferencia entre el mayor valor y el menor valor del conjunto.

$$R = x_{\max} - x_{\min}$$

Desviación estándar: Mide las desviaciones de los datos con respecto a su media.

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

Para obtener un valor adecuado de S es necesario:
-Excluir heterogeneidades del material.
-Excluir errores subjetivos.
-Establecer condiciones experimentales.
-Realizar cada medición independiente-mente e incluyendo siempre los mismos pasos.
-Emplear los valores experimentales sin aproximarlos.

Dispersión o varianza: Se le llama de este modo al cuadrado de la desviación estándar.

Propiedades de S y S^2:
- Si todas las mediciones son idénticas: $S = S^2 = 0$
- Si las mediciones se multiplican o dividen por una constante, S^2 queda multiplicada o dividida por el cuadrado de esa constante y S queda multiplicada o dividida por esa constante.
- Si a todos los datos se les suma o se les resta una constante, S y S^2 permanecen inalterables

Desviación estándar relativa: También conocida como coeficiente de variación. Es una medida de dispersión de los datos con respecto a la media y suele expresarse en forma porcentual:

$$S_R = \frac{S}{\overline{x}} \times 100$$

El valor correspondiente a la medición de interés suele reportarse como:

$$x = \overline{x} \pm S$$

Otros conceptos estadísticos

Población: Cualquier conjunto (finito o no) de individuos u objetos, que presentan alguna característica en común observable o determinable.

Muestra: Subconjunto de la población al que se le realizan las mediciones.

Muestreo simple o aleatorio: En el que todos los elementos que componen la población tienen igual probabilidad de ser seleccionados para formar parte de la muestra.

Distribución empírica de frecuencias: Es el agrupamiento de los datos en clases, acompañados por sus frecuencias. Se presentan en forma de tablas para facilitar la representación gráfica de los datos, lo que a su vez permite una mejor visualización del comportamiento de dichos datos.

Intervalos o clases: Partes en que se divide el rango total de valores obtenidos. Casi siempre son de igual amplitud para que sean comparables. El número de intervalos (*NI*) suele tomarse como aproximadamente igual a la raíz cuadrada del número total de mediciones (*n*). Por lo tanto el ancho de los intervalos (*ai*) queda como:

$$ai = \frac{R}{NI} \approx \frac{R}{\sqrt{n}}$$

Marca de clase: es un número que pertenece al intervalo y lo caracteriza para el trabajo posterior. Se toma generalmente como el punto medio o alguno de los extremos del intervalo. Las marcas de clase deben ser equidistantes.

Frecuencia absoluta (f): es el número de valores o mediciones que hay en un intervalo. De modo que para *n* mediciones y *j* intervalos se cumple que:

$$n = \sum_{j=1}^{NI} f_j$$

Frecuencia relativa (f_R): Refleja la proporción de observaciones contenidas en un intervalo.

$$f_{Rj} = \frac{f_j}{n}$$

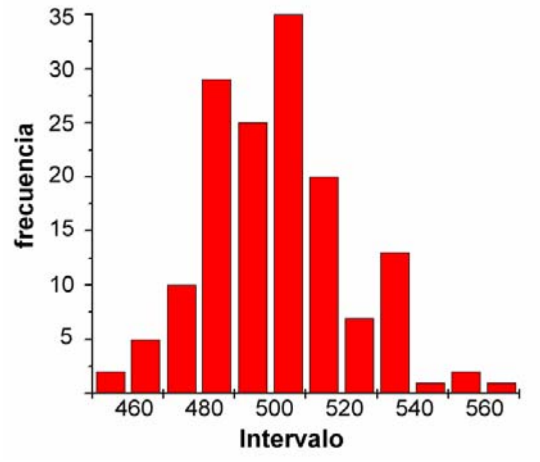
Frecuencia acumulativa (f_A): Es la sumatoria de las frecuencias absolutas de los *k* primeros intervalos.

$$f_{Ak} = \sum_{j=1}^{k} f_j$$

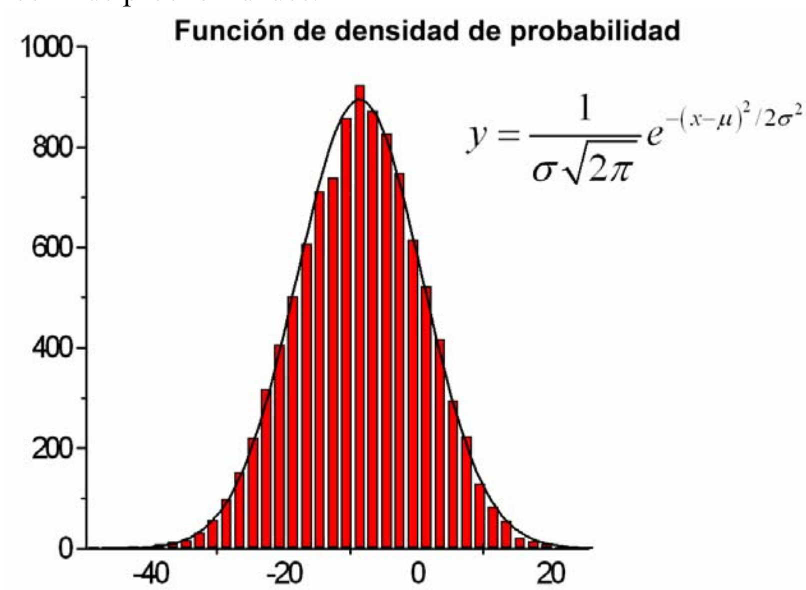
Frecuencia acumulativa relativa (f_{AR}): Proporción de observaciones cuyos valores son menores o iguales al límite superior de la clase *k*.

$$f_{AR_k} = \sum_{j=1}^{k} f_{Rj}$$

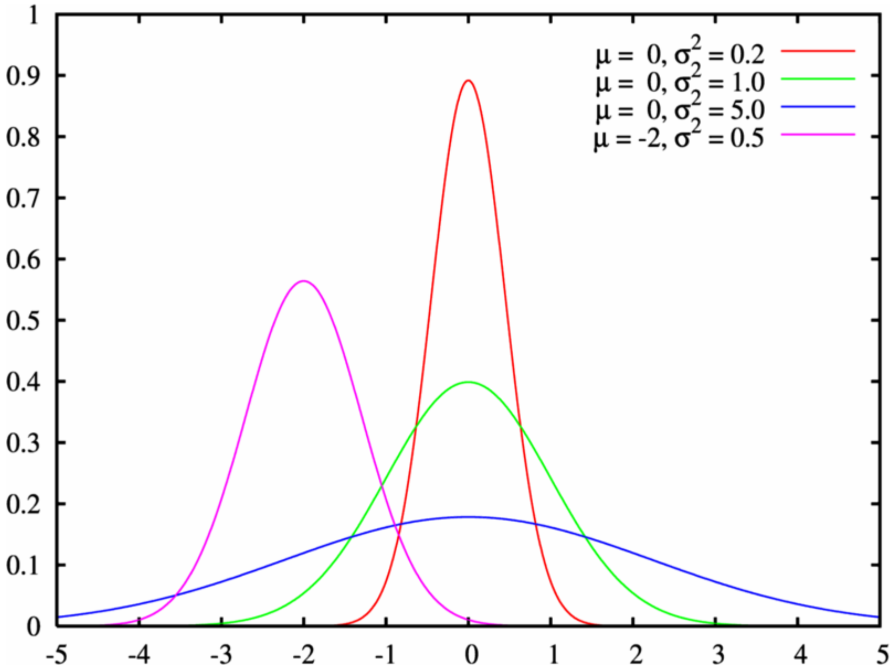
Histograma: Gráfico de barras con las frecuencias (absoluta o relativa) en la ordenada (eje y) y los intervalos o clases en la abscisa (eje x).



Distribución normal o en forma de campana de Gauss: Si una medición se repite un gran número de veces y los errores cometidos son puramente aleatorios los resultados tienden a agruparse simétricamente en torno al valor medio. La función densidad asemeja una campana. Muchas mediciones de fenómenos naturales siguen este tipo de distribución. Es la distribución de probabilidad que con más frecuencia aparece en estadística y teoría de probabilidades.

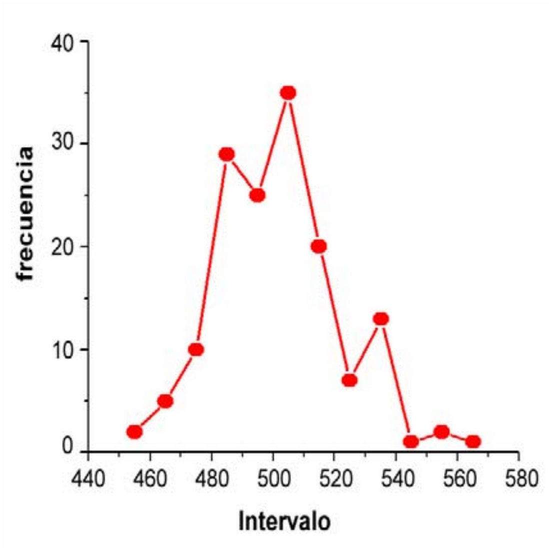


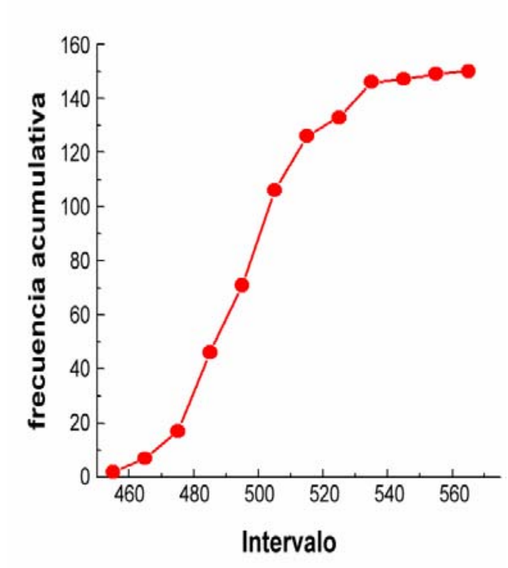
Para una serie finita μ se aproxima a $\overline{x}$ y σ se aproxima a S.



La línea verde corresponde a la distribución normal estándar.

<u>Polígonos de frecuencia</u>: Es la unión sucesiva de las marcas de clases de los diferentes intervalos mediante líneas rectas. Se pueden construir con frecuencias absolutas, relativas, acumulativas o acumulativas relativas.





Utilidad de los polígonos de frecuencia absoluta:
Al igual que los histogramas, permiten una fácil visualización del comportamiento de los datos obtenidos.

Utilidad de los polígonos de frecuencia acumulativa:
La obtención de cuantiles (valor por debajo del cual se encuentra una determinada proporción de los valoresde la distribución).

Percentil: es el cuantil más común, nos permite saber qué porciento de los datos se encuentran por debajo de un valor dado. Por ejemplo P_{25} indica el valor por debajo del cual se encuentra el 25% de la distribución de valores.

Pruebas de hipótesis

Permiten comparar estadígrafos y decidir si existen diferencias significativas entre ellos. Verifican si la diferencia entre dos valores calculados es real o si se trata de una diferencia casual, aleatoria, debida al azar.

El método clásico de trabajo en Estadística es proponer una hipótesis que contiene una igualdad y es conocida como hipótesis nula (H_0). Por ejemplo:

$H_0: \ \bar{x}_1 = \bar{x}_2$ hipótesis nula

$H_1: \ \bar{x}_1 \neq \bar{x}_2$ hipótesis alternativa

Si se cumple H_0 decimos que no existen diferencias significativas entre los estadígrafos comparados. Si no se cumple H_0 decimos que existen diferencias significativas entre los estadígrafos comparados.

Al tomar la desición de rechazar o no rechazar la hipótesis nula pueden cometerse dos tipos de errores:

tipo 1 (α): rechazar una hipótesis que no debe ser rechazada.

tipo 2 (β): no rechazar una hipótesis que debe ser rechazada.

De modo que:

	si H_0 en realidad es	
Decisión	VERDADERA	FALSA
No rechazar H_0	Decisión Correcta	Error tipo 2 (β)
Rechazar H_0	Error tipo 1 (α)	Decisión Correcta

Procedimiento:

1- Formulación de la hipótesis nula.
2- Reunir muestras observables o calculables.
3- Examinar los resultados para establecer si están en concordancia con lo que plantea nuestra hipótesis. (Si la concordancia es grande no rechazamos la hipótesis nula y si es pequeña la rechazamos)

Para determinar si la concordancia es grande o pequeña se calcula algún estadígrafo y se comparan sus valores. Para realizar esta comparación es importante conocer el nivel de confianza (α). Este no es más que la probabilidad (expresada en tanto por uno) de cometer un error α, o sea de rechazar una hipótesis nula siendo cierta.

Los químicos en general empleamos α=0.05. Esto significa que aproximadamente en 5 ocasiones de cada 100 rechazamos una hipótesis nula que era cierta. La única forma de reducir el riesgo de cometer un error tipo 1 (α) es realizar un contraste más riguroso, por ejemplo tomar α=0.01. Sin embargo debe tenerse en cuenta que a medida que disminuyen las posibilidades de cometer un error α (tipo 1) aumentan las de cometer un error β (tipo 2).

Determinación de errores burdos

Prueba *Q*   (n ≤ 10)

Se utiliza para decidir si una medición que no parece coherente con las restantes se descarta o se retiene.

$$Q_{calc} = \frac{|x_1 - x_2|}{R}$$

donde x_1 es el valor dudoso y x_2 el valor más próximo a x_1. Recordemos que R es el rango o recorrido (R=x_{max}-x_{min})

Luego se compara Q_{calc} con $Q(\alpha,n)$. Si $Q_{calc} > Q(\alpha,n)$ el punto sospechoso se descarta y si $Q_{calc} \leq Q(\alpha,n)$ el punto sospechoso se retiene.

Prueba *NS*   (n > 10)

En este caso es necesario calcular la media aritmética y la desviación estándar a partir de los valores de las mediciones sin incluir el valor del que se duda y se considera que este no es un error burdo si se encuentra en el intervalo:

$$\overline{x} \pm NS$$

donde N se toma frecuentemente como N=4, lo que es equivalente a trabajar con α=99.99%, esto significa que la probabilidad de rechazar la hipótesis nula siendo cierta es de un 0.01%. O sea que al considerar un valor como error burdo es muy poco probable que no lo sea.

α=99%   (N=2.58)
α=95%   (N=1.96)

Intervalo de confianza de la media

Es importante notar que esto no constituye una prueba de hipótesis. Si solo disponemos de un número limitado de mediciones no podemos hallar la verdadera media (μ) ni la verdadera desviación estándar (σ) de la población. Lo que podemos hacer es calcular la media ($\overline{x}$) y la desviación estándar (S) de la muestra. El intervalo de confianza nos brinda los límites, para un α dado, dentro de los cuales debe encontrarse el valor medio real (poblacional) compatible con la media muestral ($\overline{x} \pm \Delta\overline{x}$).

$$\Delta\overline{x} = \frac{t(\alpha, f)S}{\sqrt{n}}$$

donde f representa los grados de libertad ($f = n$-1), y $t(\alpha,f)$ es conocida como t de Sudent. Sus valores se encuentran tabulados para diferentes niveles de confianza (α) y grados de libertad.

El valor del intervalo de confianza depende del número de determinaciones y disminuye considerablemente al aumentar el número de determinaciones de 2 a 3 ó de 3 a 4. Al aumentar el número de determinaciones a más de 4 esta disminución no es considerable, por lo que en general un gran incremento del número de determinaciones no es justificable (desde el punto de vista del intervalo de confianza) en comparación con el aumento de gastos y tiempo necesario para la realización de los experimentos.

Comparación de dispersiones muestrales (Prueba F)

En química es frecuente que sea necesario comparar dos magnitudes proporcionales a los errores aleatorios presentes en dos condiciones de trabajo dadas, esto es: comparar dos dispersiones. O sea se hace necesario decidir si la diferencia entre dos dispersiones S_1^2 y S_2^2 se encuentra dentro de los límites de los errores casuales, en cuyo caso pueden considerarse como estimaciones de la misma dispersión poblacional (σ^2) o si no es así.

La hipótesis nula en este caso sería: $H_0: \; S_1^2 = S_2^2$

Y la hipótesis alternativa sería: $H_1: \; S_1^2 \neq S_2^2$

Para verificar la hipótesis nula se emplea la distribución *F* de Fisher. Primeramente se calcula F_{calc} según:

$$F_{calc} = \frac{S_1^2}{S_2^2}$$

donde $S_1^2 > S_2^2$. Además f_1 y f_2 son los grados de libertad de las muestras 1 y 2 para los cuales aparece tabulada. $F(\alpha, f_1, f_2)$.
Luego se compara F_{calc} con $F(\alpha, f_1, f_2)$. Si F_{calc} es mayor que $F(\alpha, f_1, f_2)$ rechazamos la hipótesis nula y si es menor o igual la aceptamos.

Comparación de medias muestrales

Se usa para comparar dos conjuntos de mediciones y decidir si son o no diferentes. En este caso:

La hipótesis nula sería: $H_0: \; \overline{x}_1 = \overline{x}_2$

Y la hipótesis alternativa sería: $H_1: \; \overline{x}_1 \neq \overline{x}_2$

Existen tres casos diferentes donde es importante esta comparación ya que nos permite decidir si dos resultados son o no coincidentes dentro del error eperimental:

Caso 1: Se mide una cantidad varias veces y se obtiene el valor medio y la desviación estándar correspondiente y se necesita comparar el resultado obtenido con un valor conocido y aceptado.

Primeramente se obtiene la *t* de Student calculada según:

$$t_{calc} = \frac{\left|\overline{x} - x_0\right|}{S}\sqrt{n}$$

donde x_0 representa al valor conocido.

Luego se compara t_{calc} con $t(\alpha, f)$. Si t_{calc} es mayor rechazamos la hipótesis nula, o sea las mediciones realizadas y el valor aceptado no coinciden. Si es Si $t_{calc} \leq t(\alpha, f)$ menor que aceptamos la hipótesis nula, o sea consideramos que mediciones realizadas y el valor aceptado coinciden para un nivel de confianza α.

Caso 2: Se mide una cantidad varias veces con dos métodos diferentes y se calculan los valores medios y las desviaciones estándar correspondientes.

En este caso la hipótesis nula sería considerar que ambos conjuntos de mediciones producen resultados equivalentes y la hipótesis alternativa sería que ambos conjuntos de mediciones producen resultados diferentes.

Los valores n_1, $\bar{x}_1$ y S_1 corresponden al primer conjunto de mediciones (método 1) y los valores n_2, $\bar{x}_2$ y S_2 corresponden al segundo conjunto de mediciones (método 2).

Primeramente se analizan las dispersiones (varianzas) de ambos conjuntos (prueba F). Si no existen diferencias significativas entre las dispersiones S_1^2 y S_2^2, para el nivel de confianza empleado, se calcula la t de Student calculada según:

$$t_{calc} = \frac{\left|\bar{x}_1 - \bar{x}_2\right|}{\sqrt{S_C}}\sqrt{\frac{n_1 n_2}{n_1 + n_2}}$$

donde S_C representa la desviación estándar combinada que se obtiene a partir de:

$$S_C = \sqrt{\frac{S_1^2 f_1 + S_2^2 f_2}{f_C}}$$

y a su vez f_C son los grados de libertad combinados, o sea los grados de libertad de los dos conjuntos tratados como un todo:

$$f_C = f_1 + f_2 = n_1 + n_2 - 2$$

Luego se comparan t_{calc} y $t(\alpha,f)$. Si t_{calc} es mayor rechazamos la hipótesis nula, o sea los dos conjuntos de mediciones no producen resultados coincidentes. Si es Si $t_{calc} \leq t(\alpha,f)$ menor aceptamos la hipótesis nula, o sea consideramos que ambos conjuntos de mediciones coinciden para un nivel de confianza α.

Si existen diferencias significativas entre las dispersiones S_1^2 y S_2^2, para el nivel de confianza empleado, entonces se calcula la t de Student calculada según:

$$t_{calc} = \frac{\left|\bar{x}_1 - \bar{x}_2\right|}{\sqrt{n_2 S_1^2 + n_1 S_2^2}}\sqrt{n_1 n_2}$$

Y se compara con $t(\alpha)$ calculada como:

$$t(\alpha) = \frac{n_2 S_1^2 t(\alpha, f_1) + n_1 S_2^2 t(\alpha, f_2)}{n_2 S_1^2 + n_1 S_2^2}$$

Otra alternativa es comparar t_{calc} con la t de Student tabulada para f_C igual a:

$$f_C = \left\{ \frac{\left(S_1^2 / n_1 + S_2^2 / n_2\right)^2}{\frac{\left(S_1^2 / n_1\right)^2}{n_1 + 1} + \frac{\left(S_2^2 / n_2\right)^2}{n_2 + 1}} \right\} - 2$$

Caso 3: Se mide la muestra 1 una sola vez la con el método A y una sola vez con el método B y no se obtiene el mismo resultado. Se repite el proceso para la muestra 2 y nuevamente los resultados obtenidos por el método A y el método B no coinciden. Se repite el proceso para n muestras.

En este caso no se replica ninguna medición, por lo que se aplica la prueba *t* a las diferencias individuales (d_i) entre los resultados de cada muestra:

$$t_{calc} = \frac{\bar{d}}{S_d}\sqrt{n}$$

donde

$$S_d = \sqrt{\frac{\sum_{i=1}^{n}\left(d_i - \bar{d}\right)^2}{n-1}}$$

con $\bar{d} = \frac{\sum_{i=1}^{n} d_i}{n}$ y para cada muestra $d_i = \left|x_{iA} - x_{iB}\right|$

Valores críticos de **Q** para el test de Dixon, en función del número de determinaciones y del nivel de confianza

n	Nivel de confianza 90%	95%	99%
3	0.941	0.970	0.994
4	0.765	0.829	0.926
5	0.642	0.710	0.821
6	0.560	0.625	0.740
7	0.507	0.568	0.680
8	0.468	0.526	0.634
9	0.437	0.493	0.598
10	0.412	0.466	0.568
12	0.376	0.426	0.522
14	0.349	0.396	0.508
16	0.329	0.374	0.463
18	0.263	0.356	0.442
20	0.300	0.342	0.425
25	0.277	0.317	0.393
30	0.260	0.298	0.372

Valores de la *t* de Student

Grados de libertad	Nivel de confianza (%)						
	50	90	95	98	99	99,5	99,9
1	1,000	6,314	12,706	31,821	63,657	127,32	636,619
2	0,816	2,920	4,303	6,965	9,925	14,089	31,598
3	0,765	2,353	3,182	4,541	5,841	7,453	12,924
4	0,741	2,132	2,776	3,747	4,604	5,598	8,610
5	0,727	2,015	2,571	3,365	4,032	4,773	6,869
6	0,718	1,943	2,447	3,143	3,707	4,317	5,959
7	0,711	1,895	2,365	2,998	3,500	4,029	5,408
8	0,706	1,860	2,306	2,896	3,355	3,832	5,041
9	0,703	1,833	2,262	2,821	3,250	3,690	4,781
10	0,700	1,812	2,228	2,764	3,169	3,581	4,587
15	0,691	1,753	2,131	2,602	2,947	3,252	4,073
20	0,687	1,725	2,086	2,528	2,845	3,153	3,850
25	0,684	1,708	2,060	2,485	2,787	3,078	3,725
30	0,683	1,697	2,042	2,457	2,750	3,030	3,646
40	0,681	1,684	2,021	2,423	2,704	2,971	3,551
60	0,679	1,671	2,000	2,390	2,660	2,915	3,460
120	0,677	1,658	1,980	2,358	2,617	2,860	3,373
∞	0,674	1,645	1,960	2,326	2,576	2,807	3,291

Valores críticos de $F = s_1^2/s_2^2$ para un nivel de confianza del 95%

Grados de libertad de s_2	Grados de libertad de s_1: 2	3	4	5	6	7	8	9	10	12	15	20	30	∞
2	19,0	19,2	19,2	19,3	19,3	19,4	19,4	19,4	19,4	19,4	19,4	19,4	19,5	19,5
3	9,55	9,28	9,12	9,01	8,94	8,89	8,84	8,81	8,79	8,74	8,70	8,66	8,62	8,53
4	6,94	6,59	6,39	6,26	6,16	6,09	6,04	6,00	5,96	5,91	5,86	5,80	5,75	5,63
5	5,79	5,41	5,19	5,05	4,95	4,88	4,82	4,77	4,74	4,68	4,62	4,56	4,50	4,36
6	5,14	4,76	4,53	4,39	4,28	4,21	4,15	4,10	4,06	4,00	3,94	3,87	3,81	3,67
7	4,74	4,35	4,12	3,97	3,87	3,79	3,73	3,68	3,64	3,58	3,51	3,44	3,38	3,23
8	4,46	4,07	3,84	3,69	3,58	3,50	3,44	3,39	3,35	3,28	3,22	3,15	3,08	2,93
9	4,26	3,86	3,63	3,48	3,37	3,29	3,23	3,18	3,14	3,07	3,01	2,94	2,86	2,71
10	4,10	3,71	3,48	3,33	3,22	3,14	3,07	3,02	2,98	2,91	2,84	2,77	2,70	2,54
11	3,98	3,59	3,36	3,20	3,10	3,01	2,95	2,90	2,85	2,79	2,72	2,65	2,57	2,40
12	3,88	3,49	3,26	3,11	3,00	2,91	2,85	2,80	2,75	2,69	2,62	2,54	2,47	2,30
13	3,81	3,41	3,18	3,02	2,92	2,83	2,77	2,71	2,67	2,60	2,53	2,46	2,38	2,21
14	3,74	3,34	3,11	2,96	2,85	2,76	2,70	2,65	2,60	2,53	2,46	2,39	2,31	2,13
15	3,68	3,29	3,06	2,90	2,79	2,71	2,64	2,59	2,54	2,48	2,40	2,33	2,25	2,07
16	3,63	3,24	3,01	2,85	2,74	2,66	2,59	2,54	2,49	2,42	2,35	2,28	2,19	2,01
17	3,59	3,20	2,96	2,81	2,70	2,61	2,55	2,49	2,45	2,38	2,31	2,23	2,15	1,96
18	3,56	3,16	2,93	2,77	2,66	2,58	2,51	2,46	2,41	2,34	2,27	2,19	2,11	1,92
19	3,52	3,13	2,90	2,74	2,63	2,54	2,48	2,42	2,38	2,31	2,23	2,16	2,07	1,88
20	3,49	3,10	2,87	2,71	2,60	2,51	2,45	2,39	2,35	2,28	2,20	2,12	2,04	1,84
30	3,32	2,92	2,69	2,53	2,42	2,33	2,27	2,21	2,16	2,09	2,01	1,93	1,84	1,62
∞	3,00	2,60	2,37	2,21	2,10	2,01	1,94	1,88	1,83	1,75	1,67	1,57	1,46	1,00