DISEÑO DE EXPERIMENTOS (DOE) PARA LA INVESTIGACIÓN

Profa. Dra. Kelly Johana Dussán Medina
kelly.medina@unesp.br

unesp

Instituto de Química

FACULTAD DE INGENIERÍA

UNIVERSIDAD DE LA REPÚBLICA URUGUAY

1

PRESENTACIÓN

- Ingeniera Química de la Universidad Nacional de Colombia sede Manizales.
- M.Sc. en Ingeniería Química de la Universidad Nacional de Colombia sede Manizales (Inmovilización de enzimas en nanoestructuras magnéticas para aplicaciones industriales).
- Ph.D. en Biotecnología por la Universidade de São Paulo (USP) (Produção de bioetanol a partir de hidrolisado hemicelulósico de bagaço de cana-de-açúcar empregando as leveduras *Scheffersomyces stipitis* NRRL Y-7124 e *Candida shehatae* UFMG HM 52.2 visando à aplicação em bioprocessos com campo eletromagnético).
- Profesora de la Universidade Estadual Paulista (UNESP) del curso de Ingeniería Química (Aprovechamiento Integral de Biomasas Lignocelulósicas).

2

GRUPO DE INVESTIGACIÓN

Profesores Colaboradores


Débora


Maria Angélica

Estudiantes de Doctorado


Sâmilla


Jonas

Estudiantes de Maestría


Henrique


Naira

Estudiantes de Graduación


Bruno


Matheus


Letícia


Mariana


João


Bianca

PROGRAMA

1. INTRODUCCIÓN AL DISEÑO DE EXPERIMENTOS (DOE)

1.1 Conceptos fundamentales

2. DISEÑOS FACTORIALES COMPLETOS

2.1 Conceptos fundamentales

2.2 Diseño con dos factores 2^2

2.3 Diseño con tres factores 2^3

2.4 Diseño con k factores 2^k

2.5 Adición de puntos centrales en el diseño 2^k

PROGRAMA

3. DISEÑOS FACTORIALES FRACCIONADOS
3.1 Introducción
3.2 Diseños factorial fraccionados 2^{k-p}
4. método de la superficie de respuesta
4.1 Conceptos fundamentales
4.2 Diseños de superficie de respuesta
4.3 Técnicas de optimización

5

BIBLIOGRAFIA

- Montgomery, Douglas. (2008). Diseño y Análisis de Experimentos. Limusa-Wiley: México
- H. Gutiérrez-Pulido y R. Vara-Salazar. Análisis y Diseño de Experimentos. McGraw-Hill (2012).
- Castro, Luis. (1980) Diseño experimental sin estadística. Trillas: México
- G.W. Oehlert. A First Course in Design and Analysis of Experiments. University of Minnesota

6

FORMA DE EVALUACIÓN

- Participación y aportación en las clases (50%).
- Resolución y exposición do estudio individual de caso (50%).

CRONOGRAMA

Fecha	Tema
24/10	Introducción al diseño de experimentos
25/10	Diseños factoriales completos
26/10	Diseños factoriales fraccionados
27/10	Método de la superficie de respuesta
30/10	Entrega del estudio individual de caso

¿O QUE ES ESTADÍSTICA?

La estadística asocia datos a problemas, generando información relevante para establecer conclusiones capaces de permitir tomar decisiones entorno de incertidumbre y variación.

ORIGEN DE LA ESTADÍSTICA

ANTIGUIDAD

Registro del número de habitantes, nacimientos y muertes, cuantificación de la riqueza, recaudación de impuestos y realización de otras encuestas cuantitativas.

ORIGEN DE LA ESTADÍSTICA

EDAD MEDIA

Conseguir información con fines fiscales y bélicos.

ORIGEN DE LA ESTADÍSTICA

HOY

La información numérica es necesaria para los ciudadanos y las organizaciones de cualquier tipo y en cualquier parte del mundo globalizado.

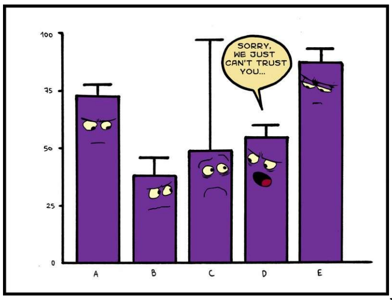
APLICACIONES

- Índices económicos.
- Sondeos de intención de voto.
- Gráficos publicados en los medios de comunicación.
- *Big data.*

IMPORTANCIA

- ¿Qué tipo de información se necesita?
- ¿Cuánta información es suficiente?
- ¿Cómo debe procesarse esta información?

Los métodos científicos son fundamentales para el trabajo de ingenieros, investigadores, médicos, científicos sociales, economistas y otros.



ACTIVIDAD

SCAN ME

RESPUESTAS

- **Media Aritmética**, también llamada promedio o simplemente media, se obtiene con la suma de un conjunto de valores dividida entre el número total de sumandos.

$$M_S = \frac{8{,}2 + 7{,}8 + 10{,}0 + 9{,}5 + 6{,}5}{5}$$

$$M_S = \frac{42}{5} = 8{,}4$$

- La **mediana** representa el valor de la variable de posición central en un conjunto de datos ordenados.

2, 2, 3, **7**, 8, 9, 9

Mediana = 7

1, 4, 4, **5**, **6**, 7, 7, 7

Mediana = (5+6) ÷ 2

= 5.5

RESPUESTAS

- La **amplitud** es la longitud de un intervalo de clase dado en una distribución agrupada de frecuencias. Por ejemplo, para un intervalo de edad 10-15 años, la amplitud de clase es de 5 años (5=15-10).
- La **media ponderada** es una generalización de la media aritmética. Es una medida de tendencia central, que es apropiada cuando en un conjunto de datos cada uno de ellos tiene una importancia relativa (o peso) respecto de los demás datos.

$$M_p = \frac{3.8{,}2+2.10{,}0+4.9{,}5+2.7{,}8+2.10+3.9{,}5+4.6{,}7}{3+2+4+2+2+3+4}$$

$$M_p = \frac{24{,}6+20+38+15{,}6+20+28{,}5+26{,}8}{20}$$

$$M_p = \frac{173{,}5}{20}$$

$$M_p = 8{,}7$$

RESPUESTAS

- La **media geométrica** de una cantidad arbitraria de números (por decir n números) es la raíz n-ésima del producto de todos los números; es recomendada para datos de progresión geométrica, para promediar razones, interés compuesto y números índice. Indica la tendencia central o valor típico de un conjunto de números mediante el producto de sus valores.

$$M_g = \sqrt[n]{x_1 \cdot x_2 \cdot \ldots \cdot x_n}$$

$$M_g = \sqrt[5]{3 \cdot 9 \cdot 27 \cdot 81 \cdot 243} \quad \textit{fatorando :}$$

$$M_g = \sqrt[5]{3 \cdot 3^2 \cdot 3^3 \cdot 3^4 \cdot 3^5}$$

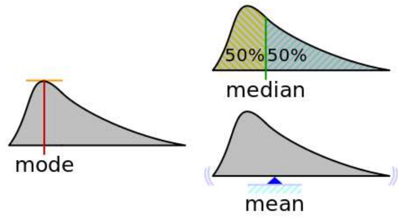
$$M_g = \sqrt[5]{3^{15}}$$

$$M_g = 3^{\frac{15}{5}}$$

$$M_g = 3^3 = 27$$

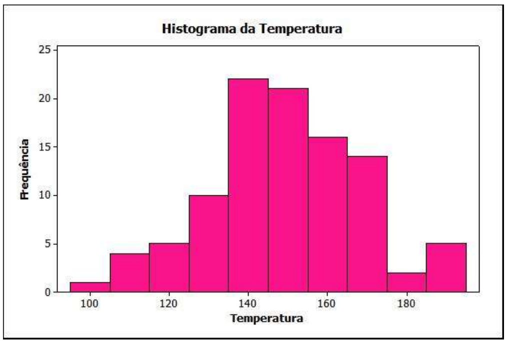
RESPUESTAS

- La **moda** es el valor que aparece con mayor frecuencia en un conjunto de datos. Puede definirse como moda de la muestra y moda de la población. Con respecto a la primera, la moda muestral de un conjunto de datos es el valor que se da con más frecuencia o el valor más común en un conjunto de datos.
- Un **Pictograma** es un tipo de gráfico, que, en lugar de barras, utiliza figuras proporcionales a la frecuencia. Generalmente se emplea para representar variables cualitativas.



RESPUESTAS

- Un **histograma** es una representación gráfica de una variable en forma de barras, donde la superficie de cada barra es proporcional a la frecuencia de los valores representados. Sirven para obtener una "primera vista" general, o panorama, de la distribución de la población, o de la muestra, respecto a una característica, cuantitativa y continua (como la longitud o el peso)



PRINCIPIOS DEL DISEÑO DE EXPERIMENTOS

21

Buen criterio a la hora de realizar experimentos y analizar sus resultados.

Aprenda a realizar sus experimentos y sacar sus conclusiones de forma más económica y eficaz.

La relación costo/beneficio siempre es importante

22

¿CÓMO PUEDE AYUDAR LA ESTADÍSTICA?

Ejemplo

Un químico desea obtener el máximo rendimiento en una reacción, y esta reacción es controlada por dos variables: temperatura y concentración de un determinado reactivo.

En la nomenclatura que usaremos en estas clases:

- La propiedad de interés, que en este caso es el **rendimiento**, la llamaremos de **respuesta**.
- Las variables que en principio influyen en la respuesta (es decir, temperatura y concentración) son los **factores**.
- La función que describe esta influencia se denomina **superficie de respuesta**.

El objetivo del investigador es descubrir qué valores (niveles) de los dos factores producen la mayor respuesta posible.

¿CÓMO RESOLVERÍA ESTE PROBLEMA?

1. Fijamos uno de los factores en un nivel determinado y variamos el otro hasta descubrir qué nivel de ese otro factor produce el mayor rendimiento. Al variar sólo uno de los factores, podemos estar seguros de que cualquier cambio en la respuesta habrá sido causado por el cambio del nivel de ese factor.
2. A continuación, manteniendo este factor en el nivel óptimo encontrado, variamos el nivel del primer factor (el que se había fijado), hasta encontrar el valor de éste que también produce un rendimiento máximo.

LISTO!!!

El experimento ha terminado y hemos encontrado los valores óptimos que buscábamos, ¿verdad?

ERRADO!!

Esto puede ser de sentido común, pero desde luego no lo es.

Casi todo el mundo está de acuerdo en que el procedimiento que acabamos de describir era "**el más lógico**", y sin embargo hay una forma más eficaz de realizar experimentos.

Es mejor variar todos los factores al mismo tiempo.

- Las variables pueden influir mutuamente, y el valor ideal de una de ellas puede depender del valor de la otra.
- Este comportamiento, que denominamos interacción entre factores, es un fenómeno que se produce con mucha frecuencia.
- Raras son las situaciones en las que dos factores actúan de forma realmente independiente.

¡El sentido común puede inducir a error!

25

¿ CÓMO PUEDE AYUDAR LA ESTADÍSTICA ?

¿Cómo podemos investigar los efectos de todos estos factores, minimizando el trabajo necesario y el costo de los experimentos?

¿Cómo podemos mejorar la calidad del producto resultante?

¿Qué factores experimentales debemos controlar para garantizar la calidad del producto?

Las investigaciones realizadas con el objetivo de dar respuesta a estas preguntas suelen llevar varios meses de trabajo a investigadores y técnicos, con un costo bastante elevado en salarios, reactivos, análisis químicos, etc.

26

- El uso de **conocimientos estadísticos** puede ayudar a responder a estas preguntas de forma racional y económica.
- Utilizando **diseño de experimentos** basados en principios estadísticos, los investigadores pueden extraer la máxima cantidad de información útil del sistema, realizando un **número mínimo** de experimentos.
- Estos métodos son poderosas herramientas con las que se pueden alcanzar diversos objetivos específicos. Podemos fabricar productos con mejores características, **reducir** el tiempo de desarrollo, aumentar la productividad de los procesos, **minimizar** la sensibilidad de los productos a las variaciones de las condiciones ambientales, etc.
- Las herramientas estadísticas, aunque valiosas, son sólo un **complemento de los conocimientos técnicos** sobre el sistema estudiado. Lo ideal es que ambos -**conocimiento básico del problema y estadística - vayan de la mano.**

Intentemos ayudar al investigador a alcanzar sus objetivos más rápidamente y a menor costo. Ya sabe que **temperatura** y **concentración**, así como tipo de **catalizador**, afectan al rendimiento.

¿Cómo se podrían ajustar los valores de temperatura y concentración para obtener una **mayor cantidad de producto**?

Variando los factores, ¿sería posible **maximizar** el rendimiento?

¿Los cambios en estos factores, provocarían cambios similares en los rendimientos si el catalizador es diferente?

¿Qué experimentos deberíamos realizar para conocer mejor el sistema?

¿Cómo podemos **cuantificar** la eficacia de los catalizadores en diferentes condiciones de temperatura y concentración?

¿Cómo pueden modificarse los valores de los factores para obtener el mayor rendimiento posible sin que las propiedades mecánicas del producto final incumplan sus especificaciones?

MODELOS EMPÍRICOS

- Un químico tiene que diseñar una planta piloto basada en una reacción concreta desarrollada recientemente en el laboratorio.
- Sabe que el comportamiento de esta reacción puede ser influenciada por muchos factores: cantidades iniciales de reactivos, pH, tiempo de reacción, carga del catalizador, velocidad con la que se introducen los reactivos en el reactor, presencia o ausencia de luz, etc.
- Aunque exista un modelo cinético para la reacción, es poco probable que pueda tener en cuenta la influencia de todos estos factores, así como de otros que suelen aparecer.

- Se trata de una situación muy complicada, para la que es difícil ser optimista sobre la posibilidad de descubrir un modelo mecanicista completo y eficaz. El investigador tiene que recurrir **a modelos empíricos**, modelos que sólo intentan describir, a partir de pruebas experimentales, el comportamiento del proceso estudiado.
- Los **modelos empíricos** son modelos locales. Utilizarlos para hacer predicciones sobre situaciones desconocidas es bajo tu propia responsabilidad.

PLANIFICACIÓN Y OPTIMIZACIÓN DE EXPERIMENTOS

- La esencia de una buena planificación es **diseñar** un experimento de forma que proporcione exactamente el tipo de información que **buscamos.**
- Para ello, primero tenemos que **saber qué es lo que buscamos.** De nuevo, parece obvio, pero en realidad no lo es.
- Incluso podríamos decir que un buen experimentador es, ante todo, una persona que **sabe lo que quiere.**
- Dependiendo de lo que quiera, algunas técnicas serán más **ventajosas,** mientras que otras serán simplemente **inocuas.**

Por tanto, si quieres convertirte en un buen diseñador de experimentos, empieza por preguntarte:

¿Qué me gustaría saber cuándo termine el experimento?

Si no sabes adónde vas, acabarás en otro sitio. Yogi Berra

Objetivo	Técnica
Muchas variables	Diseños fraccionarios
Evaluación de la influencia de variables	Diseños factoriales completos
Construcción de modelos empíricos	Modelaje por mínimos cuadrados
Optimización	Metodología de superficie de respuesta (RSM) Simplex

No basta construir modelos empíricos. También tenemos que evaluar si son realmente adecuados para el sistema que estamos describiendo. Sólo entonces podremos intentar sacar conclusiones de estos modelos. Un modelo mal ajustado forma parte de la ciencia ficción, no de la ciencia.

¡ALGUNOS CONCEPTOS!

ERRORES

- Para obtener datos experimentales confiables, necesitamos llevar a cabo un procedimiento bien definido, con detalles operativos que dependen de la finalidad del experimento.
- Por ejemplo, determinar la concentración de ácido acético en una muestra de vinagre.
  - Preparar la solución estándar;
  - Usarla para estandarizar la solución de hidróxido de sodio;
  - Realizar la titulación.

Operaciones básicas: pesar, diluciones y lecturas de volumen.

TIPOS DE ERRO

- **Errores gruesos:** Olvidarse de una variable, por ejemplo, en el caso de la titulación del vinagre olvidarse de añadir el indicador.

- **Errores sistemáticos:** Errores que afectan siempre al resultado en la misma dirección, ya sea hacia arriba o hacia abajo. Por ejemplo, cambiar el tipo de indicador en la titulación, que el estándar esté adulterado, que la balanza esté descalibrada, que la pipeta no fue usada correctamente, etc.

Aunque se eliminen los errores sistemáticos, no todo está realmente bajo control. Alguna fuente de error, por pequeña que parezca, afecta a los resultados.

DISTRIBUCIÓN NORMAL

DISTRIBUCIÓN NORMAL ESTÁNDAR

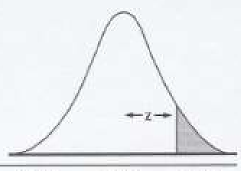
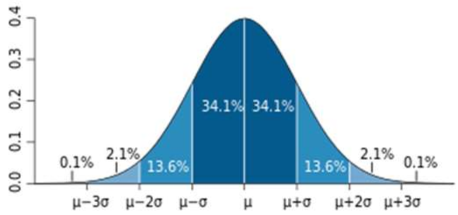


Tabela A.1 Área da cauda da distribuição normal padronizada

z	0,00	0,01	0,02	0,03	0,04	0,05	0,06	0,07	0,08	0,09
0,0	0,5000	0,4960	0,4920	0,4880	0,4840	0,4801	0,4761	0,4721	0,4681	0,4641
0,1	0,4602	0,4562	0,4522	0,4483	0,4443	0,4404	0,4364	0,4325	0,4286	0,4247
0,2	0,4207	0,4168	0,4129	0,4090	0,4052	0,4013	0,3974	0,3936	0,3897	0,3859
0,3	0,3821	0,3783	0,3745	0,3707	0,3669	0,3632	0,3594	0,3557	0,3520	0,3483
0,4	0,3446	0,3409	0,3372	0,3336	0,3300	0,3264	0,3228	0,3192	0,3156	0,3121
0,5	0,3085	0,3050	0,3015	0,2981	0,2946	0,2912	0,2877	0,2843	0,2810	0,2776
0,6	0,2743	0,2709	0,2676	0,2643	0,2611	0,2578	0,2546	0,2514	0,2483	0,2451
0,7	0,2420	0,2389	0,2358	0,2327	0,2296	0,2266	0,2236	0,2206	0,2177	0,2148
0,8	0,2119	0,2090	0,2061	0,2033	0,2005	0,1977	0,1949	0,1922	0,1894	0,1867

- Normalmente los valores de “z” na tabla van de 0,00 a 3,99, que se denomina de **área de la curva** (a la derecha) de la distribución normal estándar.
- La primera columna da el valor de “z” hasta el primero decimal y la fila superior de la Tabla da el segundo decimal.
- Para saber el área de la curva que corresponde a un cierto “z”, tenemos que buscar en la Tabla el valor localizado en la intersección de la línea con la columna.
- El valor que corresponde a z = 1,96, por ejemplo, está en la intersección de la línea z = 1,9 con la columna 0,06. Este valor, 0,025 es la fracción del área total bajo la curva que se encuentra a la derecha de z = 1,96.



Como la curva es simétrica alrededor de la media, un área idéntica se encuentra a la izquierda de z = -1,96 en la otra mitad de la curva Gaussiana. La suma de las dos áreas da 5% del área total.

Para utilizar la tabla, es necesario saber el concepto de normalización. Por definición, normalizar una variable aleatoria “x” de media “μ” y varianza “σ^2” significa construir a partir de ella una nueva variable aleatoria "z", cuyos valores se obtienen restando la media poblacional de cada valor de "x" y dividiendo el resultado por la desviación estándar.

$$z = \frac{x-\mu}{\sigma}$$

x = variable aleatoria con distribución N (μ, σ^2)

z = variable aleatoria con distribución N(0,1)

$$s^2 = \frac{1}{N-1}\sum_{i=1}^{N}(x_i-\bar{x})^2 \quad y \quad \sigma = \sqrt{s^2}$$

Intervalo de confianza para la media de la población a partir de una observación

$$x_i - z\,\sigma < \mu < x_i + z\,\sigma$$

Intervalo de confianza para la media de la población a partir de una distribución normal

$$\bar{x} - z\frac{\sigma}{\sqrt{N}} < \mu < \bar{x} + z\frac{\sigma}{\sqrt{N}}$$

μ = Media poblacional
x_i = Una observación
σ = desviación estándar de la población
z = punto de distribución N(0,1) correspondiente al nivel de confianza deseado.

DISTRIBUCIÓN T (DE STUDENT)

La distribución t (de student) es una distribución de probabilidad que surge del problema de estimar la media de una población normalmente distribuida cuando el tamaño de la muestra es pequeño y la desviación estándar poblacional es desconocida.

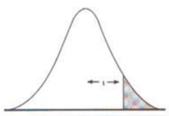


Tabela A.2 Pontos de probabilidade da distribuição *t* com ν graus de liberdade

	Área de probabilidade									
ν	0,4	0,25	0,1	0,05	0,025	0,01	0,005	0,0025	0,001	0,0005
1	0,325	1,000	3,078	6,314	12,706	31,821	63,657	127,32	318,31	636,62
2	0,289	0,816	1,886	2,920	4,303	6,965	9,925	14,089	22,326	31,598
3	0,277	0,765	1,638	2,353	3,182	4,541	5,841	7,453	10,213	12,924
4	0,271	0,741	1,533	2,132	2,776	3,747	4,604	5,598	7,173	8,610
5	0,267	0,727	1,476	2,015	2,571	3,365	4,032	4,773	5,893	6,869
6	0,265	0,718	1,440	1,943	2,447	3,143	3,707	4,317	5,208	5,959
7	0,263	0,711	1,415	1,895	2,365	2,998	3,499	4,029	4,785	5,408
8	0,262	0,706	1,397	1,860	2,306	2,896	3,355	3,833	4,501	5,041
9	0,261	0,703	1,383	1,833	2,262	2,821	3,250	3,690	4,297	4,781
10	0,260	0,700	1,372	1,812	2,228	2,764	3,169	3,581	4,144	4,587

MUESTREO ALEATORIO EN POBLACIONES NORMALES

Considerando muestras de N elementos, extraídas al azar de una población normal con media μ y varianza σ^2. Puede demostrarse que los valores muestrales $\bar{x}$ y s^2 obedecen a lo siguiente:

- Las medias muestrales $\bar{x}$ también se distribuyen normalmente, con la misma media μ, pero con una varianza igual a σ^2/N.
- La variable aleatoria , t, es definida por:

$$t = \frac{\bar{x} - \mu}{s/\sqrt{N}}$$

Siguen la **distribuicion t** con N-1 grados de libertad.

La siguiente tabla contiene los valores "t" de algunas zonas de la curva a la derecha de la distribución de Student.

La distribución t también es simétrica alrededor de la media, como la distribución normal estándar, por lo que sólo necesitamos un lado de la curva.

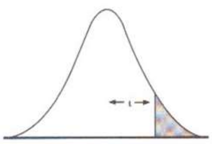


Tabela A.2 Pontos de probabilidade da distribuição *t* com ν graus de liberdade

Grados de Libertad →

ν	Área de probabilidade 0,4	0,25	0,1	0,05	0,025	0,01	0,005	0,0025	0,001	0,0005
1	0,325	1,000	3,078	6,314	12,706	31,821	63,657	127,32	318,31	636,62
2	0,289	0,816	1,886	2,920	4,303	6,965	9,925	14,089	22,326	31,598
3	0,277	0,765	1,638	2,353	3,182	4,541	5,841	7,453	10,213	12,924
4	0,271	0,741	1,533	2,132	2,776	3,747	4,604	5,598	7,173	8,610
5	0,267	0,727	1,476	2,015	2,571	3,365	4,032	4,773	5,893	6,869
6	0,265	0,718	1,440	1,943	2,447	3,143	3,707	4,317	5,208	5,959
7	0,263	0,711	1,415	1,895	2,365	2,998	3,499	4,029	4,785	5,408
8	0,262	0,706	1,397	1,860	2,306	2,896	3,355	3,833	4,501	5,041
$\boxed{9}$	0,261	0,703	1,383	1,833	$\boxed{2,262}$	2,821	3,250	3,690	4,297	4,781
10	0,260	0,700	1,372	1,812	2,228	2,764	3,169	3,581	4,144	4,587

Por lo tanto, con la distribución de Student, podemos calcular un nuevo intervalo de confianza utilizando sólo los valores de la muestra.

Intervalo de confianza para la media de la población, basado en la distribución de Student

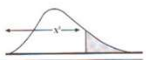
$$\bar{x} - t_{N-1}\frac{s}{\sqrt{N}} < \mu < \bar{x} + t_{N-1}\frac{s}{\sqrt{N}}$$

45

DISTRIBUCIÓN CHI-CUADRADA (χ^2)

- La distribución chi cuadrada es un caso especial de la distribución gamma y es una de las distribuciones de probabilidad más usadas en Inferencia Estadística, principalmente en pruebas de hipótesis y en la construcción de intervalos de confianza.
- La distribución chi-cuadrado se utiliza en la estimación del intervalo de confianza para una desviación estándar poblacional de una distribución normal a partir de una desviación estándar muestral.

Tabela A.3 Pontos de probabilidade da distribuição χ^2 com ν graus de liberdade



	Área de probabilidade													
ν	0,995	0,99	0,975	0,95	0,9	0,75	0,5	0,25	0,1	0,05	0,025	0,01	0,005	0,001
1	---	---	---	---	0,016	0,102	0,455	1,32	2,71	3,84	5,02	6,63	7,88	10,8
2	0,010	0,020	0,051	0,103	0,211	0,575	1,39	2,77	4,61	5,99	7,38	9,21	10,6	13,8
3	0,072	0,115	0,216	0,352	0,584	1,21	2,37	4,11	6,25	7,81	9,35	11,3	12,8	16,3
4	0,207	0,297	0,484	0,711	1,06	1,92	3,36	5,39	7,78	9,49	11,1	13,3	14,9	18,5
5	0,412	0,554	0,831	1,15	1,61	2,67	4,35	6,63	9,24	11,1	12,8	15,1	16,7	20,5
6	0,676	0,872	1,24	1,64	2,20	3,45	5,35	7,84	10,6	12,6	14,4	16,8	18,5	22,5
7	0,989	1,24	1,69	2,17	2,83	4,25	6,35	9,04	12,0	14,1	16,0	18,5	20,3	24,3
8	1,34	1,65	2,18	2,73	3,49	5,07	7,34	10,2	13,4	15,5	17,5	20,1	22,0	26,1
9	1,73	2,09	2,70	3,33	4,17	5,90	8,34	11,4	14,7	16,9	19,0	21,7	23,6	27,9
10	2,16	2,56	3,25	3,94	4,87	6,74	9,34	12,5	16,0	18,3	20,5	23,2	25,2	29,6

46

La variable aleatoria χ^2, es definida por:

$$\chi^2 = (N-1)\frac{s^2}{\sigma^2}$$

Sigue la **distribución chi-cuadrada,** También com N-1 grados de libertad.

Distribución de las estimaciones muestrales en poblaciones normales

$$\bar{x} \approx N\left(\mu, \frac{\sigma^2}{N}\right)$$

$$\frac{\bar{x}-\mu}{s/\sqrt{N}} \approx t_{N-1}$$

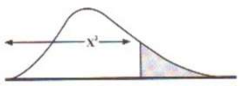
$$(N-1)\frac{s^2}{\sigma^2} \approx \chi^2_{N-1}$$

Partiendo de la siguiente ecuación:

$$(N-1)\frac{s^2}{\sigma^2} \approx \chi^2_{N-1}$$

Tenemos una distribución chi-cuadrada que también tiene una forma asimétrica, más alargada hacia la derecha.

Tabela A.3 Pontos de probabilidade da distribuição χ^2 com ν graus de liberdade



Grados de Libertad

	Área de probabilidade													
ν	0,995	0,99	0,975	0,95	0,9	0,75	0,5	0,25	0,1	0,05	0,025	0,01	0,005	0,001
1	---	---	---	---	0,016	0,102	0,455	1,32	2,71	3,84	5,02	6,63	7,88	10,8
2	0.010	0,020	0,051	0,103	0,211	0,575	1,39	2,77	4,61	5,99	7,38	9,21	10,6	13,8
3	0,072	0,115	0,216	0,352	0,584	1,21	2,37	4,11	6,25	7,81	9,35	11,3	12,8	16,3
4	0,207	0,297	0,484	0,711	1,06	1,92	3,36	5,39	7,78	9,49	11,1	13,3	14,9	18,5
5	0,412	0,554	0,831	1,15	1,61	2,67	4,35	6,63	9,24	11,1	12,8	15,1	16,7	20,5
6	0,676	0,872	1,24	1,64	2,20	3,45	5,35	7,84	10,6	12,6	14,4	16,8	18,5	22,5
7	0,989	1,24	1,69	2,17	2,83	4,25	6,35	9,04	12,0	14,1	16,0	18,5	20,3	24,3
8	1,34	1,65	2,18	2,73	3,49	5,07	7,34	10,2	13,4	15,5	17,5	20,1	22,0	26,1
$\boxed{9}$	1,73	2,09	$\boxed{2{,}70}$	3,33	4,17	5,90	8,34	11,4	14,7	16,9	$\boxed{19{,}0}$	21,7	23,6	27,9
10	2,16	2,56	3,25	3,94	4,87	6,74	9,34	12,5	16,0	18,3	20,5	23,2	25,2	29,6

- La comparación de las varianzas de dos poblaciones es muy importante para evaluar la calidad del ajuste de diversos modelos estadísticos.
- Consideremos dos muestras aleatorias extraídas de dos distribuciones normales posiblemente diferentes. La varianza de cada una de ellas sigue su propia distribución chi-cuadrado, por lo que basándonos en la siguiente ecuación:

$$(N-1)\frac{s^2}{\sigma^2} \approx \chi^2_{N-1}$$

Podemos escribir:

$$s_1^2/\sigma_1^2 \approx \chi_1^2/\nu_1 \quad \text{e} \quad s_2^2/\sigma_2^2 \approx \chi_2^2/\nu_2$$

Donde ν son los grados de libertad.

Con esto, puede demostrarse que la relación $(\chi_1^2/\nu_1)/(\chi_2^2/\nu_2)$ sigue una **distribución F.**

- Esto nos permite escribir la siguiente ecuación:

$$\frac{s_1^2/\sigma_1^2}{s_2^2/\sigma_2^2} \approx F_{\nu_1,\nu_2}$$

$$\frac{s_1^2}{s_2^2} \approx \frac{\sigma_1^2}{\sigma_2^2} F_{\nu_1,\nu_2}$$

- Podemos utilizar esta última expresión para probar hipótesis sobre la relación entre las varianzas poblacionales. En particular, para probar la posibilidad de que sean idénticas, es decir, que $\sigma_1^2/\sigma_2^2 = 1$.
- Para ello, necesitaremos la tabla de distribución F, que muestra los puntos correspondientes a algunas zonas de la curva.

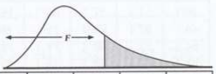


Tabela A.4 Pontos de percentagem da distribuição F, 25%

v_2 \ v_1	1	2	3	4	5	6	7	8	9	10	12	15	20	24	30	40	60	120	∞
1	5,83	7,50	8,20	8,58	8,82	8,98	9,10	9,19	9,26	9,32	9,41	9,49	9,58	9,63	9,67	9,71	9,76	9,80	9,85
2	2,57	3,00	3,15	3,23	3,28	3,31	3,34	3,35	3,37	3,38	3,39	3,41	3,43	3,43	3,44	3,45	3,46	3,47	3,48
3	2,02	2,28	2,36	2,39	2,41	2,42	2,43	2,44	2,44	2,44	2,45	2,46	2,46	2,46	2,47	2,47	2,47	2,47	2,47
4	1,81	2,00	2,05	2,06	2,07	2,08	2,08	2,08	2,08	2,08	2,08	2,08	2,08	2,08	2,08	2,08	2,08	2,08	2,08
5	1,69	1,85	1,88	1,89	1,89	1,89	1,89	1,89	1,89	1,89	1,89	1,89	1,88	1,88	1,88	1,88	1,88	1,88	1,88
6	1,62	1,76	1,78	1,79	1,79	1,78	1,78	1,78	1,77	1,77	1,77	1,76	1,76	1,75	1,75	1,75	1,74	1,74	1,74
7	1,57	1,70	1,72	1,72	1,71	1,71	1,70	1,70	1,69	1,69	1,68	1,68	1,67	1,67	1,66	1,66	1,65	1,65	1,65
8	1,54	1,66	1,67	1,66	1,66	1,65	1,64	1,64	1,63	1,63	1,62	1,62	1,61	1,60	1,60	1,59	1,59	1,58	1,58
9	1,51	1,62	1,63	1,63	1,62	1,61	1,60	1,60	1,59	1,59	1,58	1,57	1,56	1,56	1,55	1,54	1,54	1,53	1,53
10	1,49	1,60	1,60	1,59	1,59	1,58	1,57	1,56	1,56	1,55	1,54	1,53	1,52	1,52	1,51	1,51	1,50	1,49	1,48
11	1,47	1,58	1,58	1,57	1,56	1,55	1,54	1,53	1,53	1,52	1,51	1,50	1,49	1,49	1,48	1,47	1,47	1,46	1,45
12	1,46	1,56	1,56	1,55	1,54	1,53	1,52	1,51	1,51	1,50	1,49	1,48	1,47	1,46	1,45	1,45	1,44	1,43	1,42
13	1,45	1,55	1,55	1,53	1,52	1,51	1,50	1,49	1,49	1,48	1,47	1,46	1,45	1,44	1,43	1,42	1,42	1,41	1,40
14	1,44	1,53	1,53	1,52	1,51	1,50	1,49	1,48	1,47	1,46	1,45	1,44	1,43	1,42	1,41	1,41	1,40	1,39	1,38
15	1,43	1,52	1,52	1,51	1,49	1,48	1,47	1,46	1,46	1,45	1,44	1,43	1,41	1,41	1,40	1,39	1,38	1,37	1,36
16	1,42	1,51	1,51	1,50	1,48	1,47	1,46	1,45	1,44	1,44	1,43	1,41	1,40	1,39	1,38	1,37	1,36	1,35	1,34
17	1,42	1,51	1,50	1,49	1,47	1,46	1,45	1,44	1,43	1,43	1,41	1,40	1,39	1,38	1,37	1,36	1,35	1,34	1,33
18	1,41	1,50	1,49	1,48	1,46	1,45	1,44	1,43	1,42	1,42	1,40	1,39	1,38	1,37	1,36	1,35	1,34	1,33	1,32
19	1,41	1,49	1,49	1,47	1,46	1,44	1,43	1,42	1,41	1,41	1,40	1,38	1,37	1,36	1,35	1,34	1,33	1,32	1,30
20	1,40	1,49	1,48	1,47	1,45	1,44	1,43	1,42	1,41	1,40	1,39	1,37	1,36	1,35	1,34	1,33	1,32	1,31	1,29
21	1,40	1,48	1,48	1,46	1,44	1,43	1,42	1,41	1,40	1,39	1,38	1,37	1,35	1,34	1,33	1,32	1,31	1,30	1,28
22	1,40	1,48	1,47	1,45	1,44	1,42	1,41	1,40	1,39	1,39	1,37	1,36	1,34	1,33	1,32	1,31	1,30	1,29	1,28
23	1,39	1,47	1,47	1,45	1,43	1,42	1,41	1,40	1,39	1,38	1,37	1,35	1,34	1,33	1,32	1,31	1,30	1,28	1,27
24	1,39	1,47	1,46	1,44	1,43	1,41	1,40	1,39	1,38	1,38	1,36	1,35	1,33	1,32	1,31	1,30	1,29	1,28	1,26
25	1,39	1,47	1,46	1,44	1,42	1,41	1,40	1,39	1,38	1,37	1,36	1,34	1,33	1,32	1,31	1,29	1,28	1,27	1,25
26	1,38	1,46	1,45	1,44	1,42	1,41	1,39	1,38	1,37	1,37	1,35	1,34	1,32	1,31	1,30	1,29	1,28	1,26	1,25
27	1,38	1,46	1,45	1,43	1,42	1,40	1,39	1,38	1,37	1,36	1,35	1,33	1,32	1,31	1,30	1,28	1,27	1,26	1,24
28	1,38	1,46	1,45	1,43	1,41	1,40	1,39	1,38	1,37	1,36	1,34	1,33	1,31	1,30	1,29	1,28	1,27	1,25	1,24
29	1,38	1,45	1,45	1,43	1,41	1,40	1,38	1,37	1,36	1,35	1,34	1,32	1,31	1,30	1,29	1,27	1,26	1,25	1,23
30	1,38	1,45	1,44	1,42	1,41	1,39	1,38	1,37	1,36	1,35	1,34	1,32	1,30	1,29	1,28	1,27	1,26	1,24	1,23
40	1,36	1,44	1,42	1,40	1,39	1,37	1,36	1,35	1,34	1,33	1,31	1,30	1,28	1,26	1,25	1,24	1,22	1,21	1,19
60	1,35	1,42	1,41	1,38	1,37	1,35	1,33	1,32	1,31	1,30	1,29	1,27	1,25	1,24	1,22	1,21	1,19	1,17	1,15
120	1,34	1,40	1,39	1,37	1,35	1,33	1,31	1,30	1,29	1,28	1,26	1,24	1,22	1,21	1,19	1,18	1,16	1,13	1,10
∞	1,32	1,39	1,37	1,35	1,33	1,31	1,29	1,28	1,27	1,25	1,24	1,22	1,19	1,18	1,16	1,14	1,12	1,08	1,00

¿COMO DETERMINAR EL TAMAÑO DE LA MUESTRA?

- Cuando disponemos de una estimación de la desviación estándar obtenida a partir de una serie histórica, la diferencia entre la distribución t y la distribución normal deja de ser importante.
- Esta es la situación más habitual en los laboratorios de análisis, donde cada día se repiten los mismos procedimientos.
- Para estimar el tamaño de la muestra en estos casos, podemos utilizar la expresión

$$N \geq \left(\frac{z \times \sigma}{L}\right)^2$$

Donde L es la precisión deseada, σ es la desviación estándar y z es el punto de la distribución normal estándar para el nivel de confianza elegido.

Supongamos que queremos determinar un intervalo de confianza del 95% para el peso de un grano de frijoles, de forma que la diferencia entre los valores extremos del intervalo sea la desviación estándar muestral. ¿Cuántos granos debemos pesar?

$$\bar{x} - t_{N-1}\frac{s}{\sqrt{N}} < \mu < \bar{x} + t_{N-1}\frac{s}{\sqrt{N}}$$

$$s = \left(\bar{x} + t_{N-1}\frac{s}{\sqrt{N}}\right) - \left(\bar{x} - t_{N-1}\frac{s}{\sqrt{N}}\right)$$

$$s = 2 * t_{N-1}\frac{s}{\sqrt{N}}$$

$$0{,}5 = \frac{t_{N-1}}{\sqrt{N}}$$

Encontrar valores, en la Tabla de la distribución t, de N e t_{N-1} que la división llegue al valor de 0,5

EJERCICIO

¿Cuántos granos de frijol hay en un 1kg?

- Una posible solución es contar todos los granos (**Descartado**).
- Utilizar herramientas estadísticas para encontrar un intervalo de confianza.
- Fueron realizadas 140 pesajes de granos de frijoles.

CALCULAR

- Histograma
- Media
- Desviación estándar
- Varianza
- Intervalos de confianza para z y t (95%)