

Infraestructura computacional

Evaluación de Desempeño

Indicadores de desempeño de máquina

VS.

Indicadores de desempeño de aplicación

Objetivo

- La idea es identificar cómo influyen los diferentes elementos de la infraestructura en el desempeño del sistema
- Esta puede ser una buena herramienta para hacer planeación de capacidades

Análisis de desempeño

vs.

Planeación de capacidad

Análisis de Desempeño

- Una primera idea es identificar los períodos pico (meses, días, horas)

Análisis de Desempeño

- Para entender las relaciones entre diferentes elementos y el desempeño se puede :
 - Medir y registrar información sobre indicadores de la aplicación
 - Medir y registrar información sobre indicadores del sistema

Análisis de Desempeño

- Para entender las relaciones entre diferentes elementos y el desempeño se puede :
 - Determinar las relaciones entre las dos anteriores
 - Encontrar umbrales superiores

Elementos

- Algunos de los elementos importantes de monitorear:
 - Consumo de disco a lo largo del tiempo
 - Qué tanto se están demorando las operaciones
 - Cómo es el patrón de acceso al disco
 - lecturas
 - escrituras

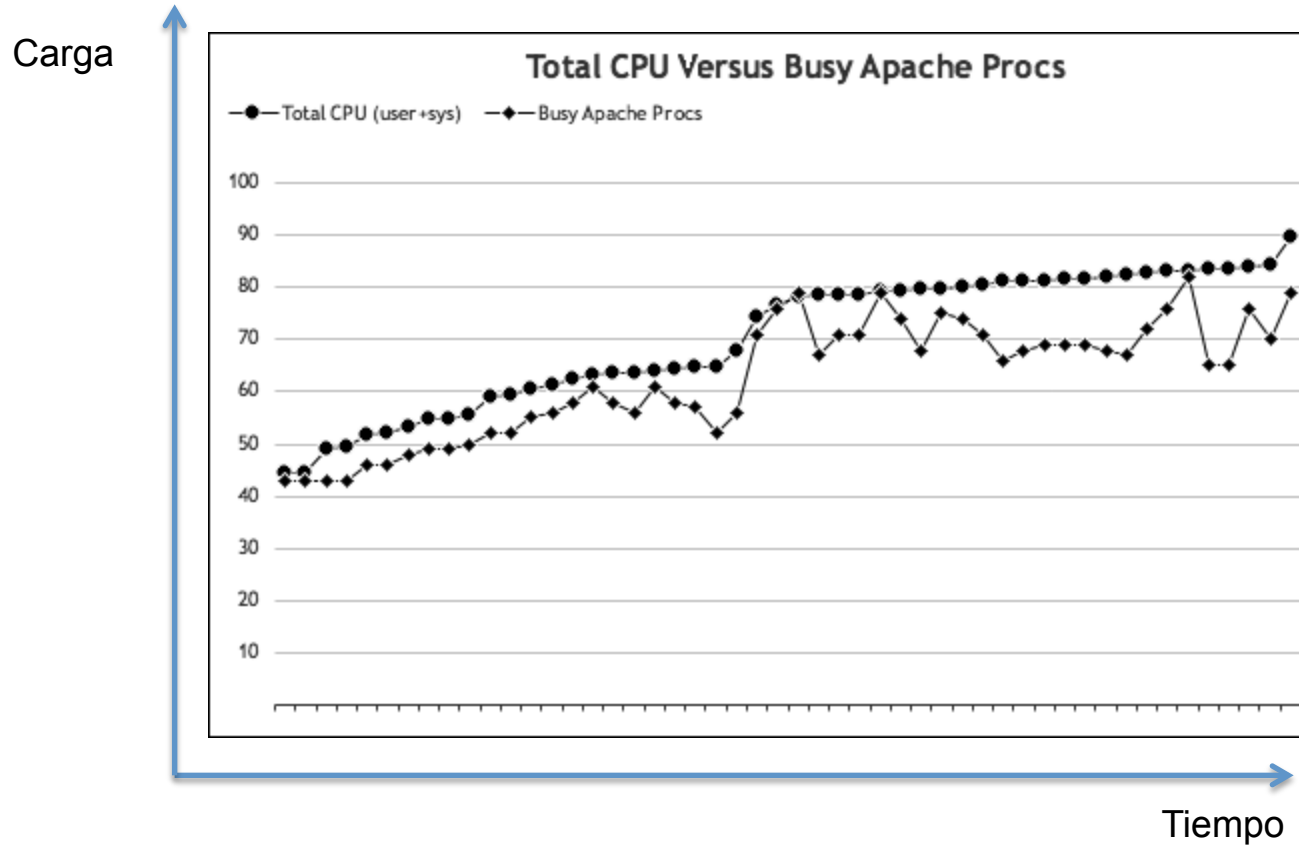
Elementos

- Algunos de los elementos importantes de monitorear:
 - Cuánto espacio ocupan los backups
 - Porcentaje de uso del procesador
 - Uso de la BD

Análisis de Desempeño

- Una idea interesante es mirar la relación entre la carga y el uso de los recursos

Ejemplo



Manejo de Datos

Análisis de Desempeño

- Defina
 - Medidas de desempeño
 - Criterio usado para evaluar el desempeño del sistema
 - Carga
 - Pedidos generados por los usuarios

Análisis de Desempeño

- Carga
 - Real
 - Carga observada en un sistema real bajo operación normal
 - Sintética
 - Carga generada con características similares a las de la carga normal, que puede ser repetida de forma controlada para adelantar estudios

Monitores

- Monitores
 - Un monitor es una herramienta usada para observar la actividad de un sistema
- Términos usados:
 - Evento
 - Rastro
 - Sobrecarga
 - Dominio
 - Resolución

Monitores

- Evento
 - Cambio en el estado del sistema
- Rastro
 - Registro de una secuencia de eventos
- Sobrecarga (overhead)
 - La mayoría de monitores perturban ligeramente el sistema (consumen recursos)

Monitores

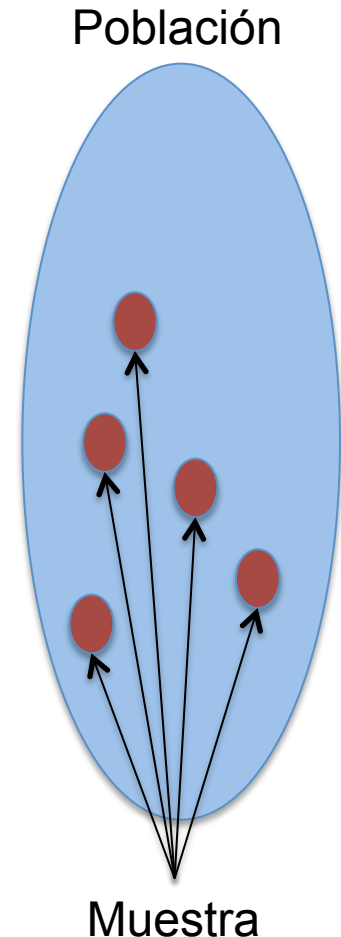
- Dominio
 - Conjunto de actividades observables por el monitor
- Resolución
 - Precisión de la información que se observa

Monitores

Criterio	Monitor de Hardware	Monitor de Software
Dominio	Eventos de hardware	Eventos de sistema operacional o programas
Tasa de entrada	10^5 por segundo	La tasa está limitada por el manejo que el sistema operacional haga
Resolución	Nanosegundos o más fino	Milisegundos
Sobrecarga	No	Si (varia)
Costo	Alto	Medio

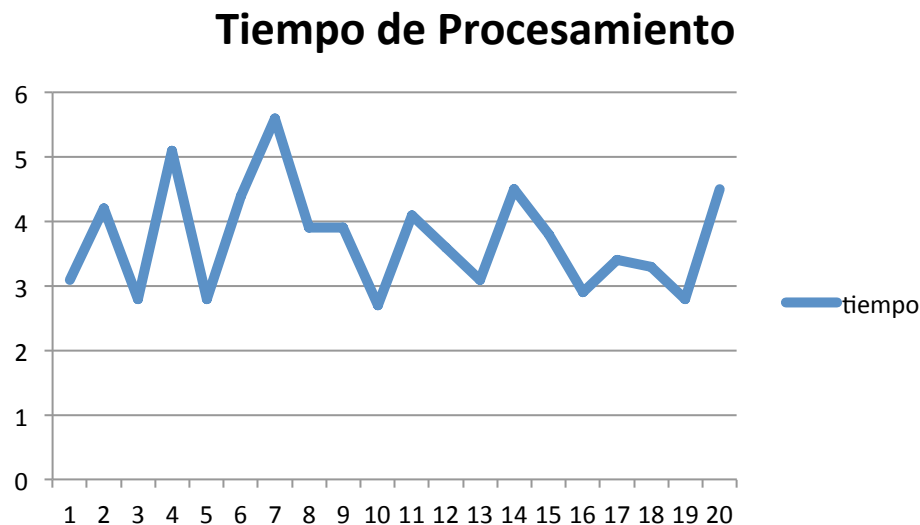
Medidas Registradas

- Muestra
 - Conjunto de observaciones sobre una población
- Población
 - Todos los elementos en el universo que se está estudiando
- Muestra vs. Población
 - Las características de la población son desconocidas
 - El objetivo es estimar esas características



Medidas Registradas

- Variable aleatoria
 - Una variable que toma un valor de un conjunto de valores (con cierta probabilidad)



Medidas Registradas

- Varianza
 - El cuadrado de la distancia entre los valores de x y su promedio
 - una medida de dispersión

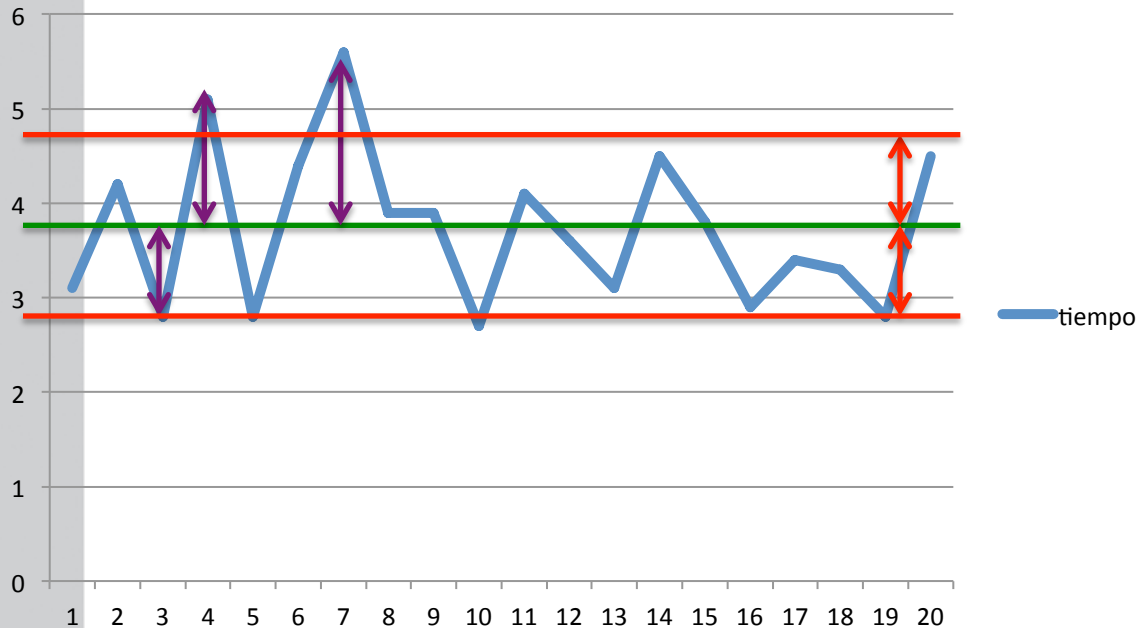
$$Var(X) = \frac{1}{n} \sum_{i=1}^n (x_i - \bar{x})^2$$

Medidas Registradas

- Varianza
 - El cuadrado de la distancia entre los valores de x y su promedio
 - una medida de dispersión
 - La raíz cuadrada de la varianza es la desviación estándar
 - una medida de distancia a la media

Medidas Registradas

Tiempo de Procesamiento



Medida	Valor
promedio	3,725
varianza	0,67
desviación estándar	0,82

Medidas Registradas

- Media Aritmética
 - Promedio de un conjunto de valores
- Mediana
 - El valor central de una variable (50% de los valores son menores y 50% son mayores)
- Moda
 - El valor más frecuente

valor
3,1
4,2
2,8
5,1
2,8
4,4

Medida	Valor
media/promedio	3,73
mediana	3,65
moda	2,8

Resumen de los Datos

- Resumen de los datos observados:
 - El valor más usado es el promedio

valor
2,2
5,2
2,2
5,2
2,2
5,2
2,2

¿Qué nos dice este promedio acerca de los datos?

Medida	Valor
media/promedio	3,48

Resumen de los Datos

- Resumen de los datos observados:
 - El valor más usado es el promedio
 - Para que sea significativo este valor debería representar la mayor parte del conjunto de datos

valor
2,2
5,2
2,2
5,2
2,2
5,2
2,2

Medida	Valor
media/promedio	3,48



Resumen de los Datos

- Resumen de los datos observados:
 - El valor más usado es el promedio
 - Para que sea significativo este valor debería representar la mayor parte del conjunto de datos
 - Otra opción es presentar la media, la mediana y la moda

valor
2,2
5,2
2,2
5,2
2,2
5,2
2,2

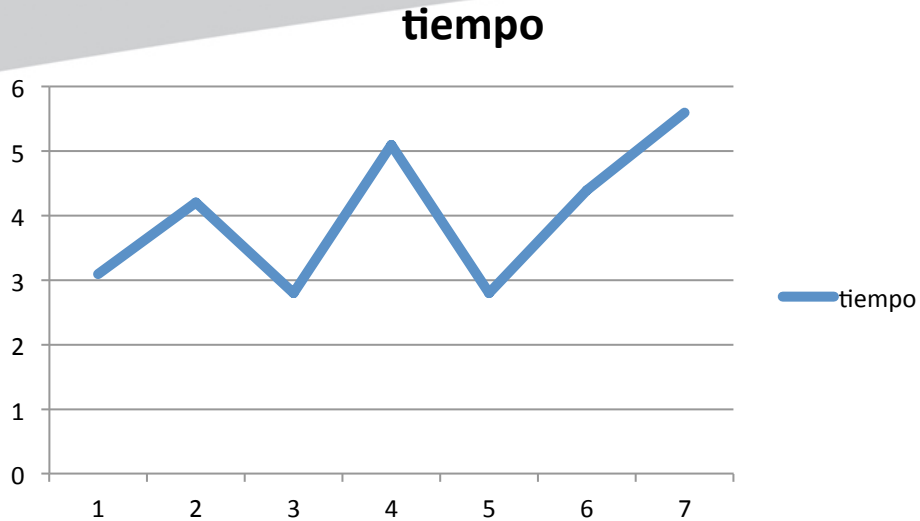
Medida	Valor
media/promedio	3,48
mediana	2,2
moda	2,2

Resumen de los Datos

- Resumen de los datos observados:
 - La desviación estándar nos indica la agrupación de los datos alrededor de un valor central
 - Si los valores están muy alejados del valor no podemos tener certeza sobre una predicción

Resumen de los Datos

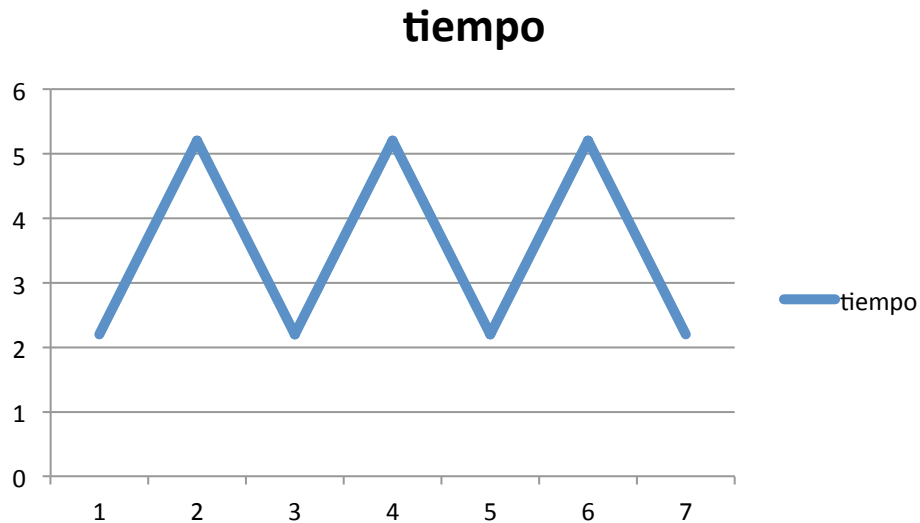
tiempo
3,1
4,2
2,8
5,1
2,8
4,4
5,6



Medida	Valor
media/promedio	4
varianza	1,27
desviación estándar	1,12

Resumen de los Datos

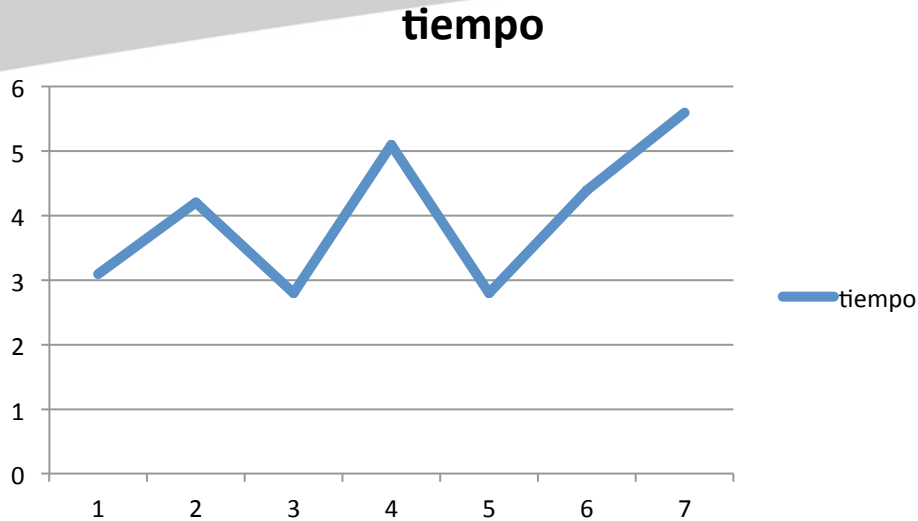
tiempo
2,2
5,2
2,2
5,2
2,2
5,2
2,2



Medida	Valor
media/promedio	3,48
varianza	2,57
desviación estándar	1,60

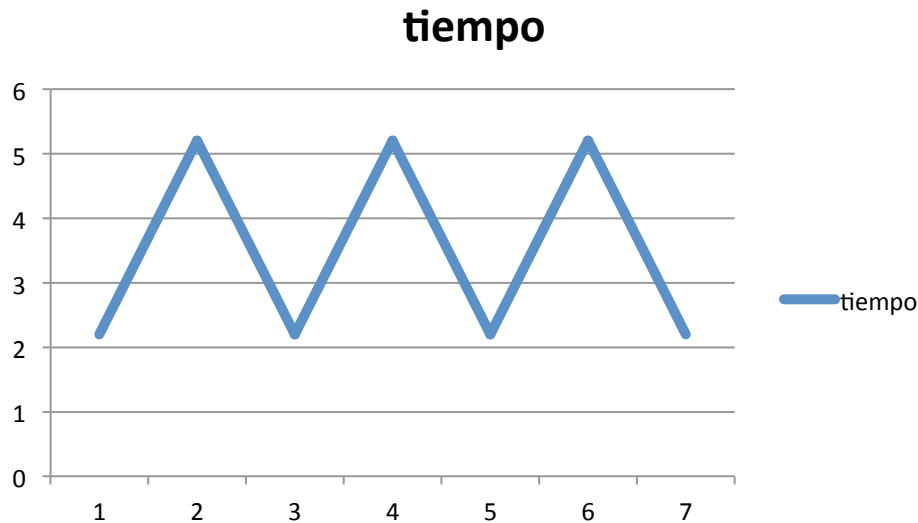
Resumen de los Datos

tiempo
3,1
4,2
2,8
5,1
2,8
4,4
5,6



Medida	Valor
media/promedio	4
varianza	1,27
desviación estándar	1,12

tiempo
2,2
5,2
2,2
5,2
2,2
5,2
2,2



Medida	Valor
media/promedio	3,48
varianza	2,57
desviación estándar	1,60

Promedio

- Errores comunes:
 - ¿Cuál es el tiempo esperado de ejecución de un programa a con una configuración C?
 - Medida:
 - El tiempo promedio de ejecución es 20 seg.
 - Conclusión:
 - Está dentro de los límites aceptables para el servicio. Compre una segunda plataforma para correr el mismo servicio.

Promedio

- Errores comunes:
 - La utilidad del promedio depende del número de valores y de la varianza
 - No es útil decir que el tiempo promedio de ejecución de un programa es 20 s. como resultado de dos medidas (10,30)

Promedio

- Errores comunes:**

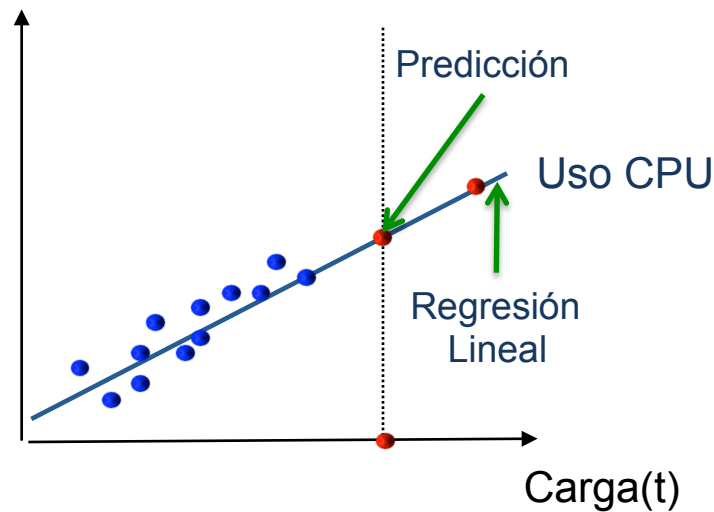
A	B
10	5
9	5
11	5
10	4
10	31

Suma	50	50
Promedio	10	10
Moda	10	5

Carga Sintética

Generación

- Tome medidas del sistema
- Cree un modelo
 - Permite la predicción

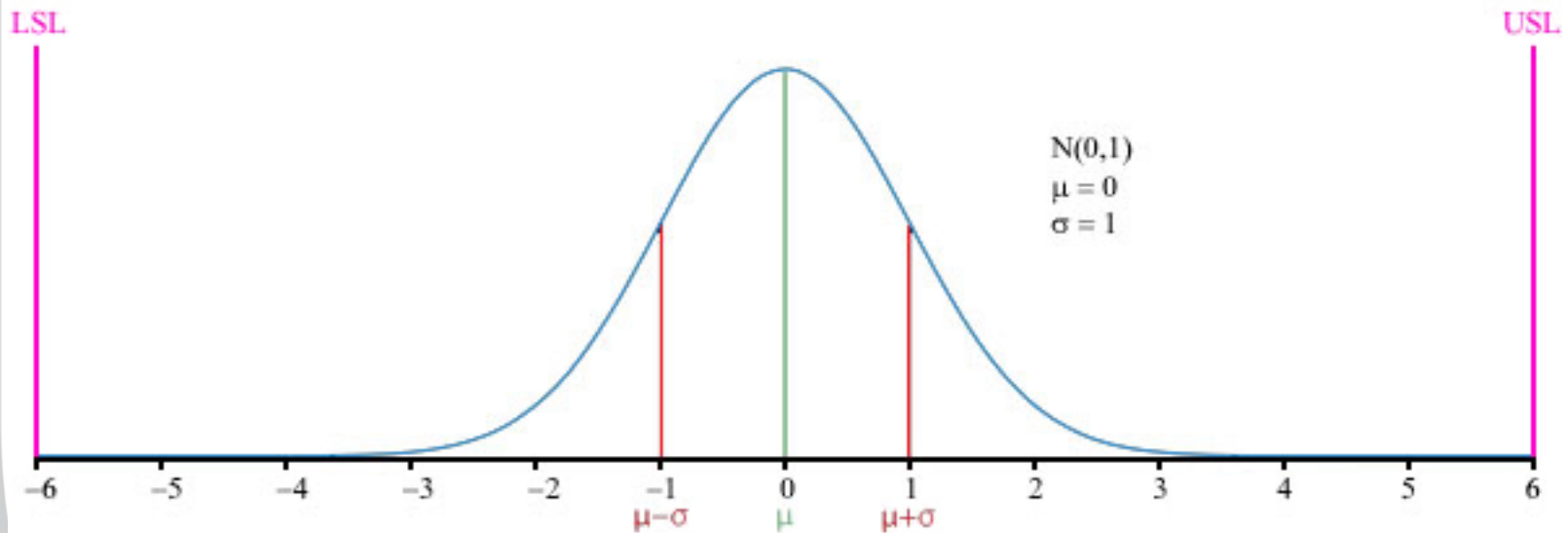


Modelos

- Regresión lineal
- Función de distribución de probabilidad
 - Bernoulli
 - Exponencial
 - Normal

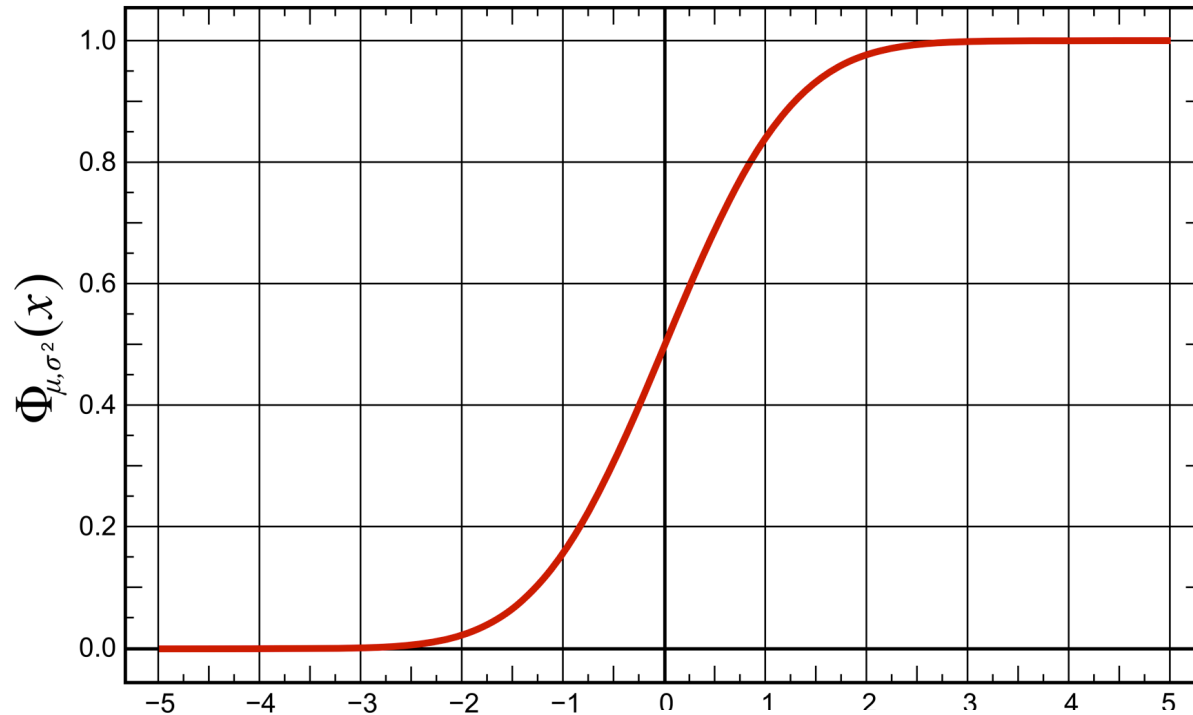
Función de Distribución de Probabilidad

- Describe la probabilidad que una variable aleatoria X tome un valor x

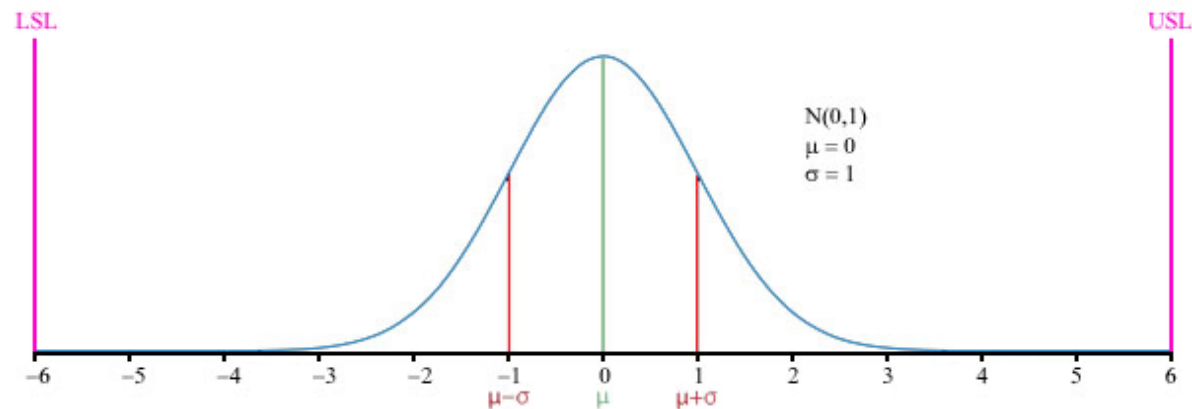


Cumulative distribution function

for normal distribution, mean (μ) = 0 and standard deviation (σ) = 1



Describe la probabilidad que una variable X tome un valor $\leq x$

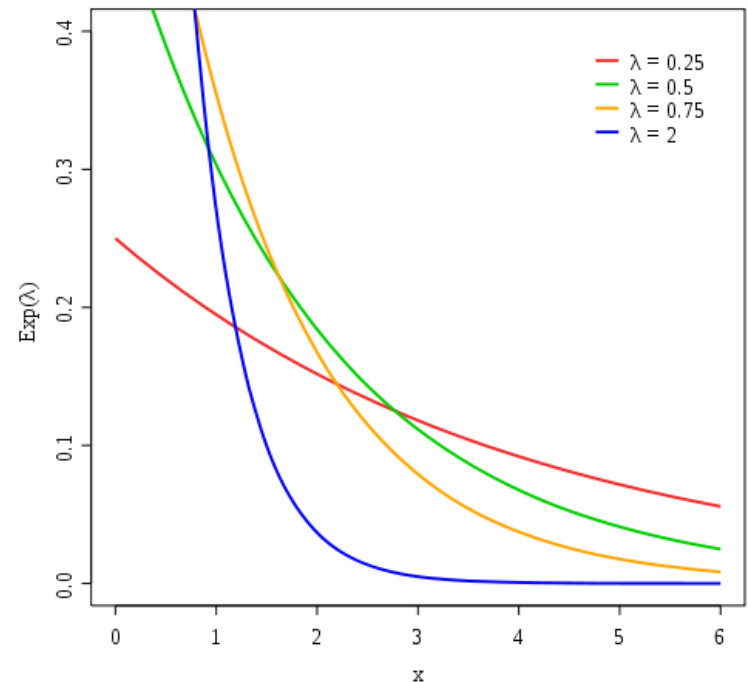


Distribución Bernoulli

- Distribución discreta más simple
 - $x = 0$
 - $x = 1$
 - p : probabilidad de éxito ($x=1$)
 - $1-p$: probabilidad de falla
- Ejemplos:
 - Un sistema esté up|down
 - Un paquete llega a sus destino si|no
 - El paquete tiene errores si|no

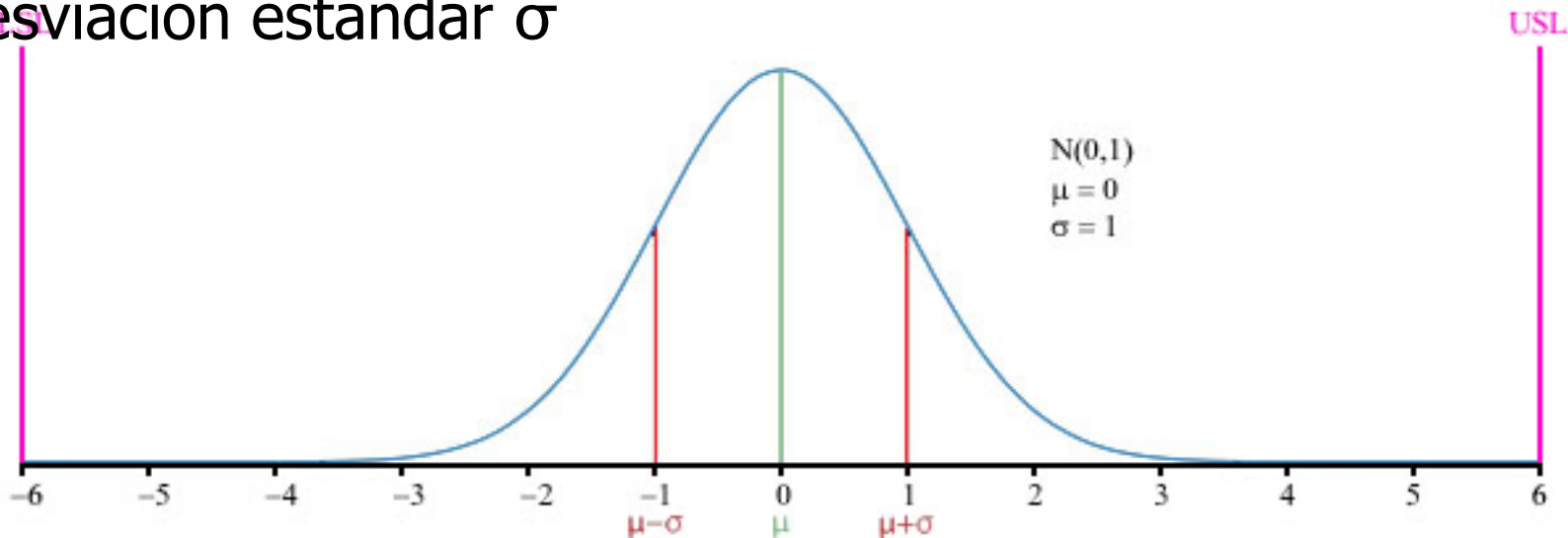
Distribución Exponencial

- Distribución más usada en teoría de colas
 - El tiempo entre llegadas de trabajos a un sistema es exponencial con valor esperado $1/\lambda$
- Ejemplos:
 - Tiempo entre llegadas de pedidos a un dispositivo
 - Tiempo entre fallas de un dispositivo



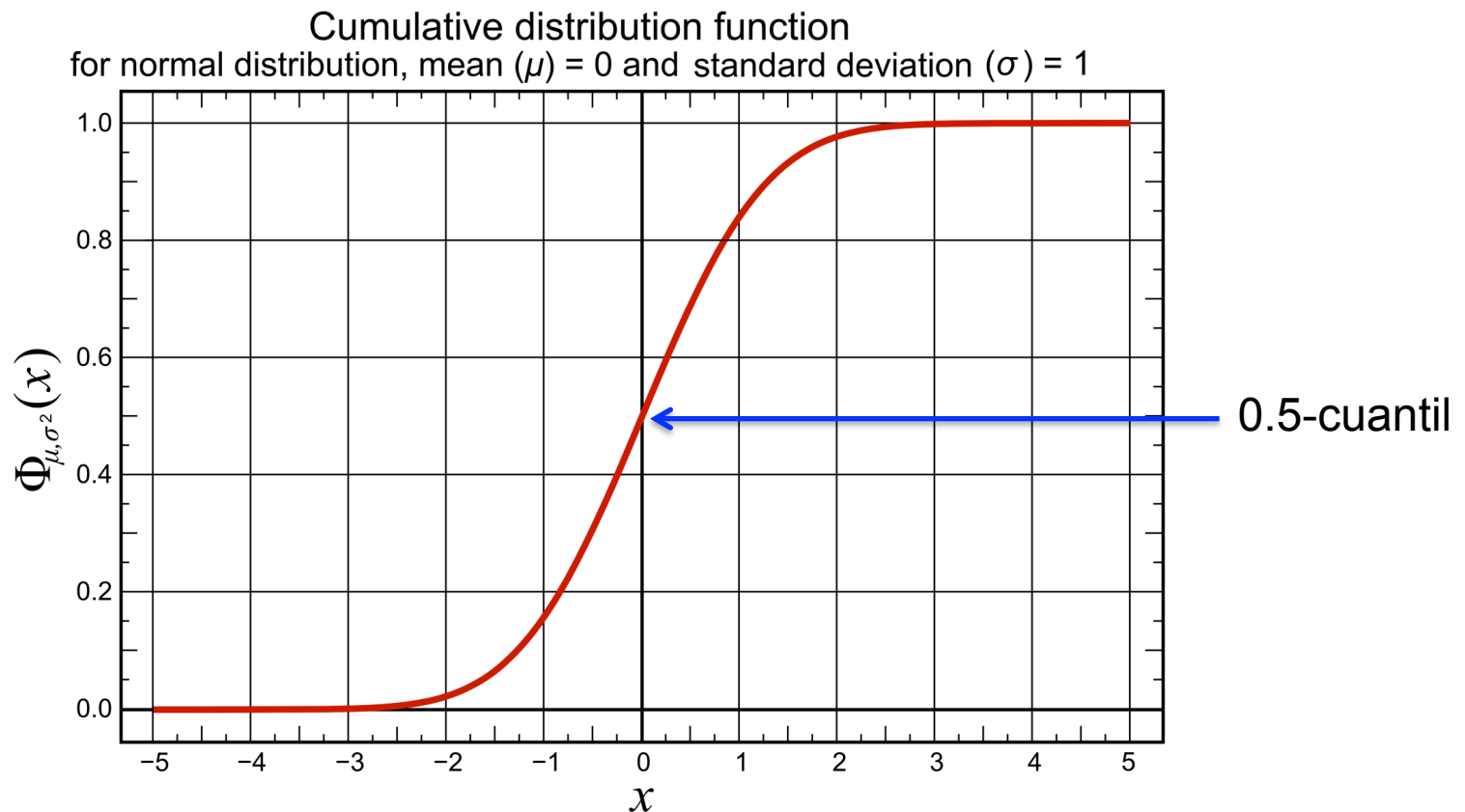
Distribución Normal

- Distribución usada cuando el comportamiento aleatorio es causado por varias fuentes independientes, como errores en la medida, errores por factores que no están incluidos en el modelo, muestreo sobre un gran número de valores.
 - Valor esperado μ
 - Desviación estándar σ



α -cuantil

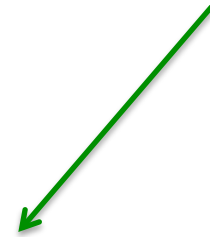
- El valor x al cual la función de distribución acumulada CDF toma el valor α .



Intervalo de Confianza

- *No es posible calcular el valor exacto del valor promedio de una población con base en una muestra.*
- Pero podemos ubicar límites de probabilidad.
 - Podemos determinar límites $c1$ y $c2$ tal que existe una probabilidad $(1-\beta)$ de que el promedio de la población se encuentra en el intervalo $(c1, c2)$.
 - El intervalo $(c1, c2)$ es el intervalo de confianza.

90% de las muestras
tienen un valor entre
 $c1$ y $c2$



Intervalo de Confianza

- Intervalo de confianza del 90%
 $100(1-\beta)\% = 100(1-0.1)\%$

$$[\bar{x} - (z_{1-\alpha/2})(s / \sqrt{n}), \bar{x} + (z_{1-\alpha/2})(s / \sqrt{n})]$$

s es la desviación estándar

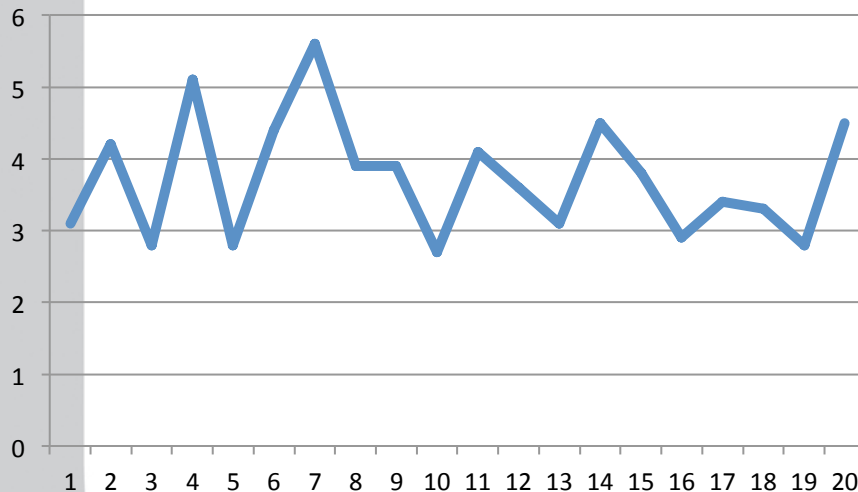
n es el tamaño de la muestra

$Z_{1-\beta/2}$ es el $(1-\beta/2 = 0,95)$ cuantil de la distribución normal unitaria

Intervalo de Confianza

$$[\bar{x} - (z_{1-\alpha/2})(s / \sqrt{n}), \bar{x} + (z_{1-\alpha/2})(s / \sqrt{n})]$$

Tiempo de Procesamiento



$$\bar{x} = 3.90$$

$$s = 0.95$$

$$n = 32$$

$$z_{1-\alpha/2} = 1.645$$

$$ic90\% = 3.90 \pm (1.645)(0.95) / \sqrt{32}$$

$$ic90\% = (3.62, 4.17)$$

Tabla de cuantiles
para la distribución
normal unitaria

Decimos con el 90% de confianza que el valor de la media de la población está entre 3.62 y 4.17. Si se toman 100 muestras, el 90% de las veces la media estará en el intervalo (3.62, 4.17)

Presentación de Datos

Presentación de Datos

- Se muestran los ejes y tienen etiquetas
- Las etiquetas son claras y concisas
- Se muestran las escalas y las divisiones para todos los ejes
- Se presentan el mínimo y el máximo
- El número de curvas debe ser razonable
- Todos los gráficos deberían usar la misma escala (o se maneja con cuidado)

Presentación de Datos

- Asignar etiqueta a cada curva
- Todos los símbolos en el gráfico se explican
- Indicar las unidades de medida
- Incluir un título para el gráfico
- El título debe ser claro y conciso
- El gráfico debe adicionar información
- El reporte que incluye la figura debe hacer referencia a la misma

Evaluación Comparativa (Benchmarking)

Evaluación Comparativa

- Las evaluaciones comparativas de desempeño se corren para comparar el comportamiento de dos sistemas
 - Las cargas se generan de forma automática

Evaluación Comparativa

- Errores comunes
 - La carga de prueba solo representa el comportamiento típico
 - Solo representa el comportamiento promedio sin la varianza. Lo cual puede conducir a conclusiones equivocadas
 - El sesgo de demandas sobre algunos dispositivos es ignorado
 - Se supone que los pedidos sobre dispositivos de entrada/salida están distribuidos uniformemente

Evaluación Comparativa

- Errores comunes
 - Ajuste equivocado de los parámetros
 - El proceso de generación de carga puede ser ajustado por medio de parámetros para aumentar o disminuir la carga. Cambiar el valor de dos parámetros diferentes puede conducir APARENTEMENTE al mismo escenario.

Ejemplo: un sistema que maneja número de usuarios, el tiempo de respuesta de los usuarios, la demanda de recursos por usuario.

Uno puede verse tentado a aumentar el tiempo de respuesta de los usuarios en vez del número de usuarios, pero estos dos parámetros no conducen a los mismos resultados.

- Las fallas de caché son (en principio) menores en el sistema con menos usuarios.

- Variar la demanda de recursos por usuario puede alejar al sistema del comportamiento típico del sistema real.

Evaluación Comparativa

- Errores comunes
 - Ignorar los efectos del cache
 - El comportamiento de los cachés varía debido al orden de llegada de los pedidos. Como consecuencia es importante modelar el orden de llegada de los pedidos.
 - Ignorar los efectos del buffering
 - El tamaño de las áreas temporales de almacenamiento tiene una fuerte incidencia en los tiempos de respuesta de los dispositivos de entrada/salida.

Evaluación Comparativa

- Errores comunes
 - Imprecisiones por el muestreo
 - Los datos recolectados para el modelaje de cargas no son, por definición, completos. Es necesario evaluar su validez, con métodos estadísticos por ejemplo.

Ejemplo: la probabilidad de tener un dato que está más allá de la media más o menos una desviación estándar es mayor al 40%.

Evaluación Comparativa

- Errores comunes
 - Ignorar la sobrecarga del monitor
 - El efecto de un monitor en el sistema no debería pasar cierto umbral. Al afectar el comportamiento del sistema, un monitor también afecta las variaciones en los resultados.

Evaluación Comparativa

- Errores comunes
 - Validación de los datos
 - Revisar los resultados y evaluar si aparecen comportamientos extraños (¿Cuál sería la justificación de tales comportamientos?)
 - Condiciones iniciales
 - Si los resultados son sensibles a las condiciones iniciales (por ejemplo, si dependen del espacio en disco), crear un mecanismo para arrancar los experimentos en estados iniciales equivalentes

Evaluación Comparativa

- Errores comunes
 - Recoger muchos datos y adelantar poco análisis
 - La recolección de datos es solo el primer paso, su verdadera utilidad corresponde al análisis de los mismos

Introducción a Diseño Experimental

Diseño Experimental

- Variable de respuesta
 - Es el resultado de un experimento.
Generalmente es una medida del desempeño de un sistema
- Factores
 - Variables que pueden tomar diferentes valores y cuya variación afecta las variables de respuesta
- Niveles
 - Los valores que un factor puede asumir

Diseño Experimental

- Replicación
 - Repeticiones de los experimentos
- Diseño
 - Especificar uno o varios experimentos, las combinaciones de factores en cada caso y las replicaciones

Diseño Experimental

- Diseño simple
 - Se empieza con una configuración típica y solo se varía un factor al tiempo para evaluar cómo afecta el desempeño.
 - Dados k factores, con n_i niveles cada factor i , un diseño simple requiere n experimentos

$$n = 1 + \sum_{i=1}^k (n_i - 1)$$

No es adecuado si los factores interactúan.

Diseño Experimental

- Diseño factorial completo
 - Evalúa todas las combinaciones posibles de todos los factores.
 - Dados k factores, con n_i niveles cada factor i , un diseño factorial completo requiere n experimentos

$$n = \prod_{i=1}^k n_i$$

Es adecuado si los factores interactúan.

Diseño Experimental

- Diseño factorial completo
 - Considera todas las combinaciones posibles
 - El principal problema es el costo de realizar todos estos experimentos
 - Buscar una forma para reducir el número de experimentos de una forma inteligente

$$n = \prod_{i=1}^k n_i$$

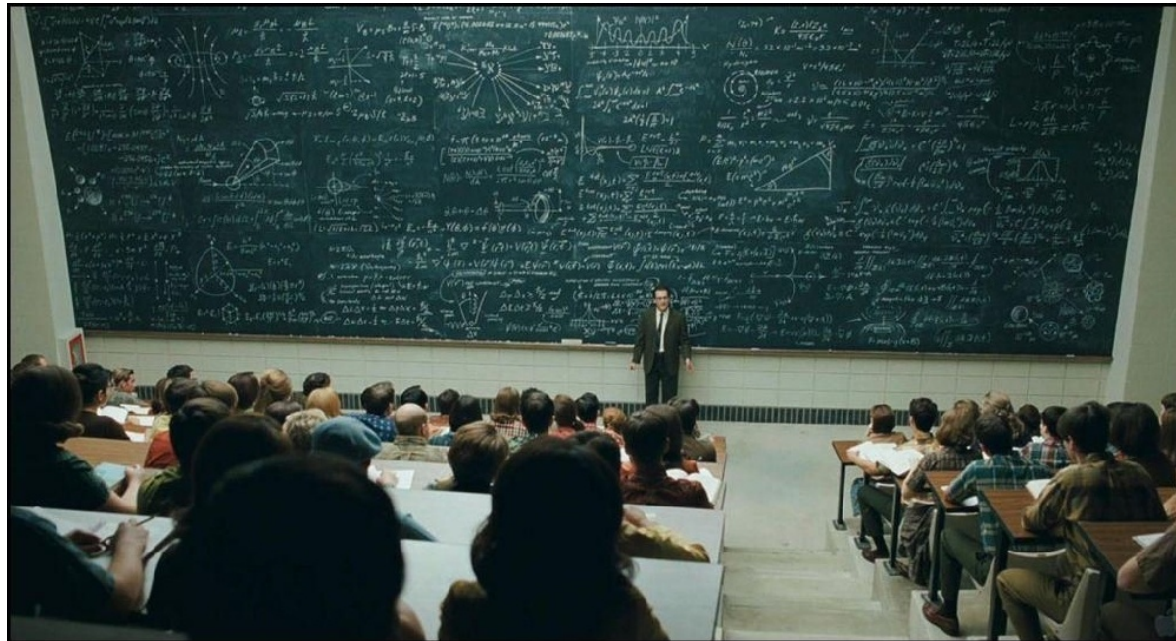
Diseño Experimental

- Diseño factorial parcial
 - El costo de un diseño factorial completo es muy costoso
 - Reduzca los factores o los niveles por factor

$$n = \prod_{i=1}^{k'} n_i'$$

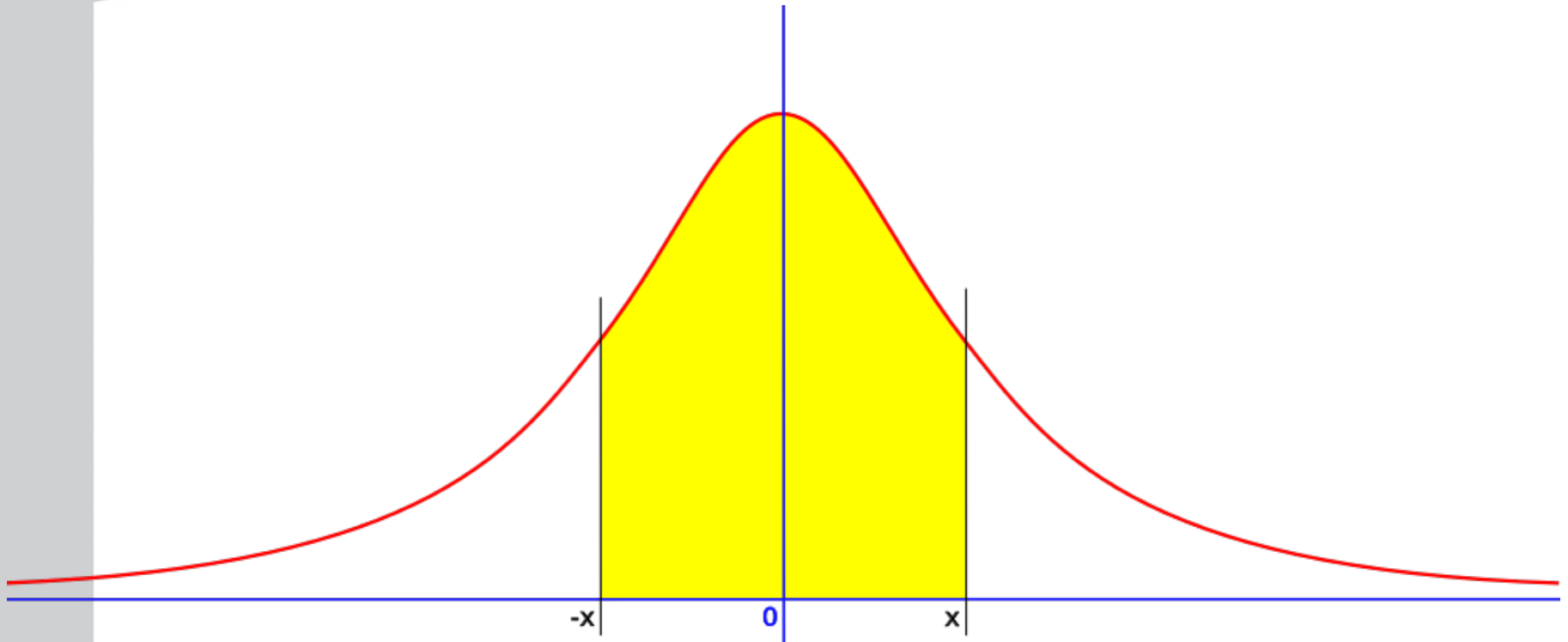
Referencias

- ***Planeación de Capacidades***: The Art of Capacity Planning, Scaling Web Resources, John Allspaw. 2008, capítulo 3
- ***Análisis de Desempeño***: The Art of Computer Systems Performance Analysis. Raj Jain. 1991.



Preguntas

Distribución t



- Distribución de probabilidad que se usa para estimar el promedio de una población con una distribución normal cuando el tamaño de la muestra es pequeño.