INTRODUCCION PARA EL TP2 Y TP3

Simulación se encarga de construir **objetos** (**modelos**) que representan a (una parte) la realidad, y que son interpretados por una computadora.

Las técnicas de Simulación ayudan a problemas de Ingeniería, Economía, Medicina, Biología, Ciencias Sociales, Doctorados, Postgrados, Master

(*Todos involucran a técnicas de Simulación*).

1º Se parte de una **Hipótesis**, Aproximación del Problema.

2º Se define el **Modelo** (Deterministas vs. Estocásticos).

3º Ensayar con el **Modelo** (Lenguajes de Simulación vs Lenguajes de Programación).

4º El **resultado es visualmente grafico**, sirven para el análisis y mejora de los sistemas.

(*Analistas importantes eligen un 93% las Estadísticas, un **84% la Simulación** y después (lejos) el resto investigación operativa, programación lineal, PERT/CPM, teoría de inventarios*).

El estudio de las **VARIABLES ALEATORIAS**, *tiene el propósito de replicar el comportamiento*, y para eso está la **Probabilidad y la Estadística**.

(*Es necesario que nos pongamos de acuerdo en algo para comprender esto de Aleatorio o Aleatoriedad*)

- Experimento: es un proceso cuyo resultado no se conoce con certeza. Varios Experimentos forman una Muestra.

- Variable aleatoria*:* es una variable que puede tomar cualquier valor (Finito o Infinito).

- Proceso estocástico: es un proceso que evoluciona en el tiempo, y que involucra una ***variable aleatoria***, de tal modo que el comportamiento del proceso no puede preverse con exactitud.

- Distribución de probabilidad: permite relacionar un conjunto de valores con su frecuencia relativa de aparición.

- Función de densidad de probabilidad f(x)*:* describe la probabilidad que una variable aleatoria X asuma un cierto valor xi: **$f(x_i) = P(X=x_i)$**

- Función de distribución acumulada F(x)*:* describe la probabilidad de que una variable aleatoria X asuma un valor más pequeño o igual que un cierto valor xi: **$F(x_i) = P(X \leq x_i)$**

- Estimación de la media:

$$\overline{X}(n) = \frac{\sum_{i=1}^{n} X_i}{n}$$

- Estimación de la varianza:

$$S^2(n) = \frac{\sum_{i=1}^{n} \left[X_i - \overline{X}(n)\right]^2}{n-1}$$

0.50	3.35	20.85	7.81	0.44	0.03	3.82	7.09	3.02	2.80
2.08	6.53	52.53	10.23	0.76	0.00	28.21	15.51	4.86	10.41
5.25	11.67	46.23	28.06	6.05	4.82	46.36	2.90	5.47	17.42
7.20	41.15	9.54	4.88	19.10	9.17	0.83	7.43	9.98	4.11
10.28	23.44	6.19	2.39	7.57	12.97	12.62	7.65	18.49	6.95
1.08	9.89	5.49	2.16	14.18	11.89	12.73	0.51	14.61	27.01
1.91	18.77	4.98	6.41	1.45	1.71	5.21	2.89	8.38	3.50
2.86	17.60	4.89	11.74	15.31	36.64	3.62	21.78	2.15	6.70
17.13	0.11	17.58	1.30	2.44	9.59	1.74	5.02	6.46	18.76
1.49	7.92	4.03	3.13	1.67	23.31	3.13	9.35	0.10	0.51

Muestra del tiempo entre dos llegadas sucesivas

Todas las variables que no se puedan prever con exactitud se modelan, por eso es **Estocástico** **(<u>Modelo Estadístico</u>)**, sino sería **Determinista**.

⇩

Toma y análisis de los datos

Es la etapa más importante, cara y difícil, permite crear el modelo estadístico (como se comportará)

Se recomienda cuestionar siempre la información obtenida, cual fue el origen?, como la obtuvieron?, tiene sentido?, son suficientes los datos o son excesivos?, es condición indispensable trabajar con buenos datos, y no siempre pasa eso.

↓

Ajuste de una función de distribución

Objetivo es obtener ***f(x) o F(x)*** de los datos que obtuve antes

Representar el histograma de los datos (tabular los datos, establecer el valor de k, construir intervalos, el tamaño de N, y armar la Grafica. *Es una estimación de la Función de densidad **f(x)***

↓

Validación del ajuste

Test estadísticos que validan lo anterior.
Test Chi-Cuadrado y Test Kolmogorov-Smirnov.

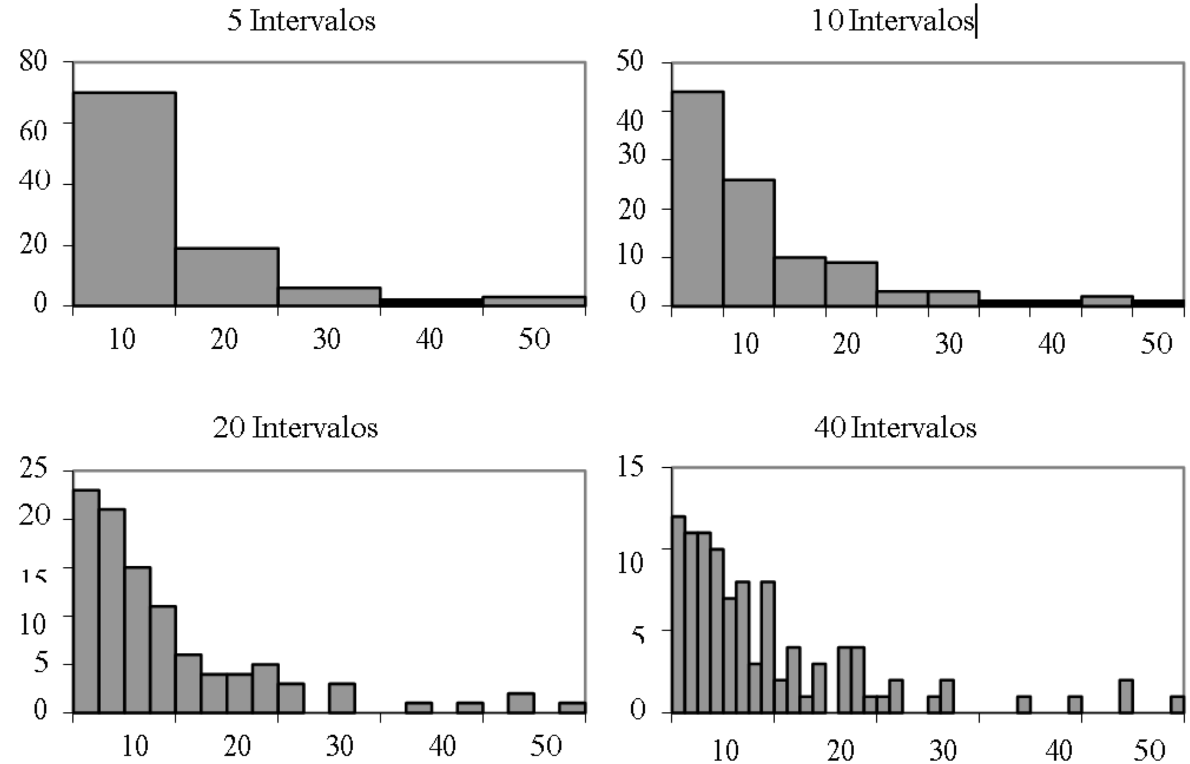
0.50	3.35	20.85	7.81	0.44	0.03	3.82	7.09	3.02	2.80
2.08	6.53	52.53	10.23	0.76	0.00	28.21	15.51	4.86	10.41
5.25	11.67	46.23	28.06	6.05	4.82	46.36	2.90	5.47	17.42
7.20	41.15	9.54	4.88	19.10	9.17	0.83	7.43	9.98	4.11
10.28	23.44	6.19	2.39	7.57	12.97	12.62	7.65	18.49	6.95
1.08	9.89	5.49	2.16	14.18	11.89	12.73	0.51	14.61	27.01
1.91	18.77	4.98	6.41	1.45	1.71	5.21	2.89	8.38	3.50
2.86	17.60	4.89	11.74	15.31	36.64	3.62	21.78	2.15	6.70
17.13	0.11	17.58	1.30	2.44	9.59	1.74	5.02	6.46	18.76
1.49	7.92	4.03	3.13	1.67	23.31	3.13	9.35	0.10	0.51

Muestra del tiempo entre dos llegadas sucesivas

- Frecuencia observada de los valores

Intervalo	Nº de sucesos	Probabilidad
[0-4)	34	0.34
[4-8)	26	0.26
[8-12)	13	0.13
[12-16)	7	0.07
[16-20)	8	0.08
[20-24)	4	0.04
[24-28)	1	0.01
[28-32)	2	0.02
[32-36)	0	0.00
[36-40)	1	0.01
[>40)	4	0.04

- Histogramas



f(x) - <u>Función de Densidad de Probabilidad</u> - *f(x)*

Cuando el proceso estocástico (*modelo estadístico*) que se quiere describir (lo que quiero modelar) es continuo y no se encuentra limitado a valores discretos, la variable aleatoria utilizada para describir el modelo estadístico puede tomar un número infinito de valores posibles.
Teniendo en cuenta que la probabilidad de obtener un valor cualquiera dentro del rango de valores para el cual ha sido definida la variable aleatoria es 1, la probabilidad de que la variable aleatoria asuma un cierto valor concreto P(X= xi*)* será 0. Este es uno de los motivos por los cuales las propiedades estadísticas de una variable aleatoria X se suelen describir mediante la función de densidad de probabilidad ***f(x)***.
En lugar de describir la probabilidad de que la variable aleatoria asuma un valor determinado, ***se describe la probabilidad de que la variable aleatoria asuma un valor dentro de un rango de valores (entre a y b)***:

Esta información (***f(x)***) puede obtenerse directamente de la función de distribución de probabilidad observada (***¡¿¿de dónde??!***) de los datos de la Tabla de Frecuencias Observadas.
Y también se pude calcular la Distribución de Probabilidad Acumulada

$$P(a \leq X \leq b) = \int_a^b f(x)dx$$

Intervalo	Nº de sucesos	Probabilidad	Acumulativa
[0-4)	34	0.34	0.34
[4-8)	26	0.26	0.60
[8-12)	13	0.13	0.73
[12-16)	7	0.07	0.80
[16-20)	8	0.08	0.88
[20-24)	4	0.04	0.92
[24-28)	1	0.01	0.93
[28-32)	2	0.02	0.95
[32-36)	0	0.00	0.95
[36-40)	1	0.01	0.96
[>40)	4	0.04	1

Función de Distribución Acumulada - *F(x)*

$$F(x) = \int_a^b f(x)dx$$

Hasta ahora hemos visto como trabajar con valores observados, *sucede que en general las muestras con las que contamos suelen ser chicas como para poder representar el comportamiento de lo que queremos estudiar, no alcanza, los datos que dispongo no muestran todos los aspectos*, entonces ¿Qué hago?, **Ahora es cuando hago uso las Funciones Teóricas.**

Funciones Discretas: *Poisson* vs **Funciones Continuas**: *Exponencial*, *Normal*, *Uniforme*.

Para cada una de las funciones se incluye una tabla resumen (la función de densidad de probabilidad ***f(x)***, la función de distribución acumulada ***F(x)***, la media y la varianza).

Función con Distribución Exponencial (λ):

Se usa para modelar ***el tiempo entre llegadas*** o ***el tiempo que tarda en producirse un evento*** y también se emplea para modelar tiempos de servicio que son muy variables, por ejemplo, la duración de una llamada telefónica.

Función de Densidad $f(x) = \lambda e^{-\lambda x} \longrightarrow x \geq 0$

Media $\mu = \dfrac{1}{\lambda}$

Varianza $\sigma^2 = \dfrac{1}{\lambda^2}$

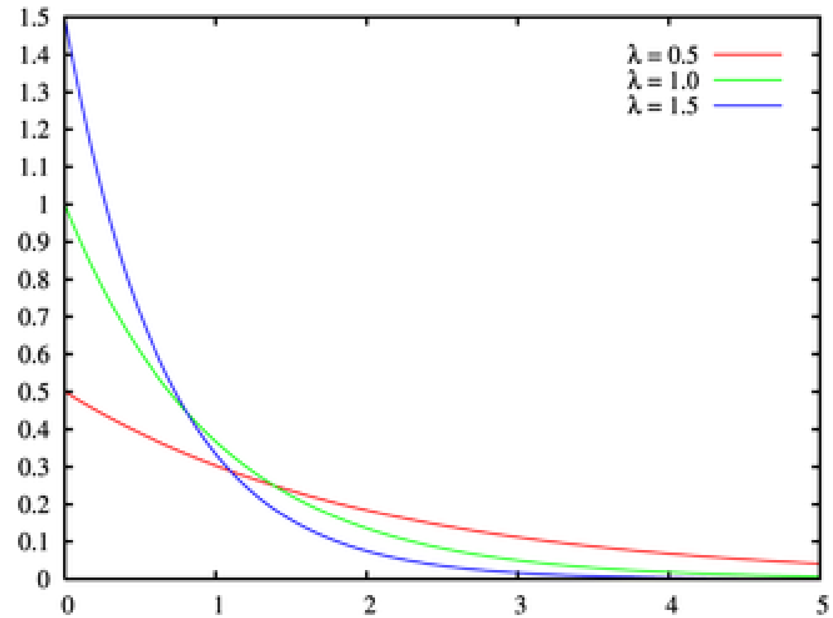
Función Acumulada $F(x) = 1 - e^{-\lambda x}$ para $x > 0$.

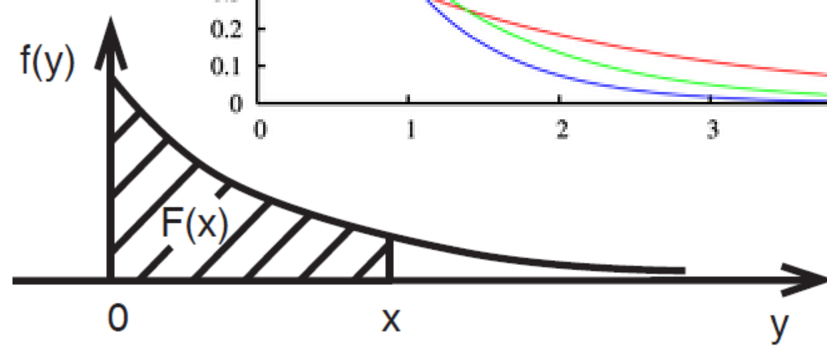
Función Generadora de Rnd con Distribución Exponencial (λ)

$$\text{Rnd} = F(x) = 1 - e^{-\lambda x}$$

$$x = -\frac{1}{\lambda}\ln(1 - \text{Rnd})$$

$$\text{Rnd} \in [0, 1)$$





Función con Distribución Normal (μ ; σ):

Se utiliza para modelar sistemas tales que el 70% de los datos muestreados se encuentran a una distancia inferior a **σ** (desviación estándar) del valor promedio **μ**, y la frecuencia de aparición de los datos se encuentra distribuida simétricamente respecto al valor promedio. Un ejemplo de utilización de una función de distribución normal es el modelado del tiempo de producción de las máquinas cuando no se considera la posibilidad de fallos o errores de diversos tipos.

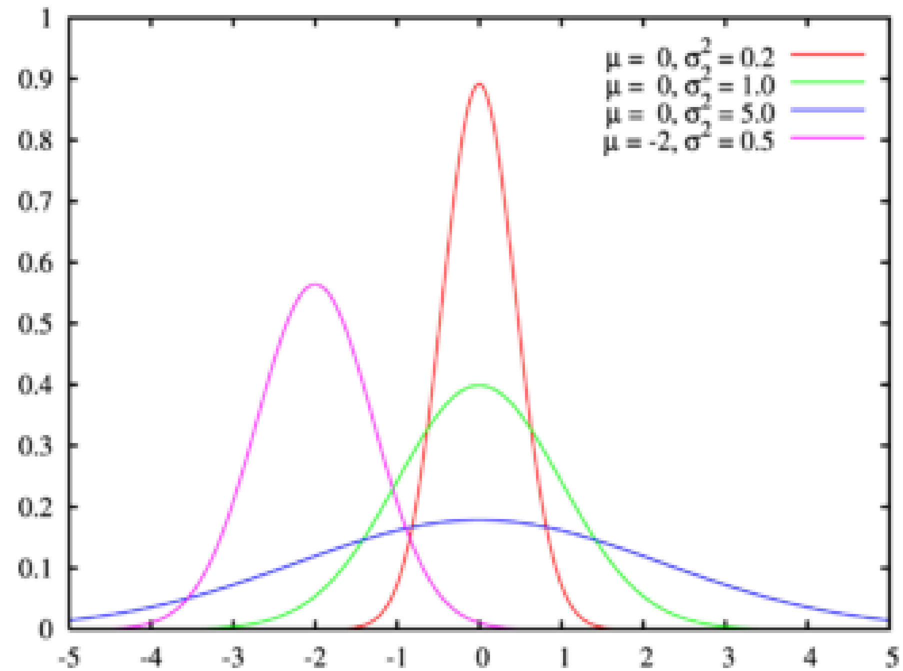
Función de Densidad $f(x) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2}$

Media $\mu = 0$

Varianza $\sigma^2 = 1$

Función Acumulada $\nexists$

Función Generadora de Rnd con Distribución Normal (μ ; σ)



Box-Muller

$$N_1 = [\sqrt{-2.\ln(RND_1)}.\cos(2.\pi.RND_2)] * \sigma + \mu$$
$$N_2 = [\sqrt{-2.\ln(RND_1)}.sen(2.\pi.RND_2)] * \sigma + \mu$$

Convolución (*N = 12*)

$$Z_i = \left(\sum_{i=1}^{N} RND_i - 6\right).\sigma + \mu$$

<u>Función con Distribución Uniforme (a ; b):</u>

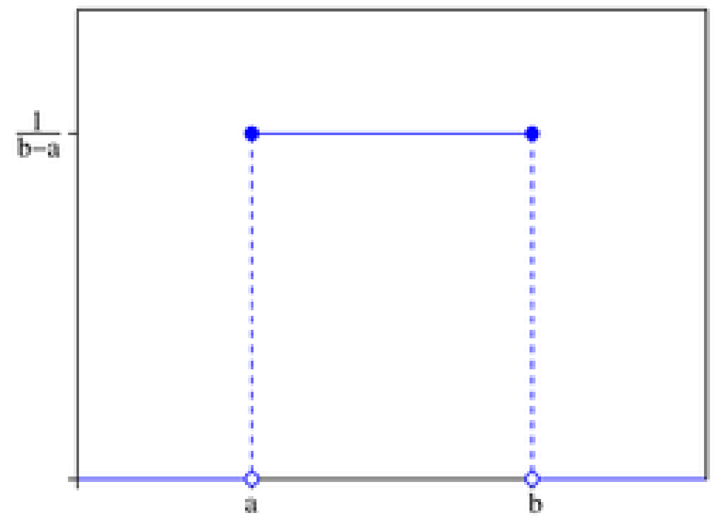
Se usa para establecer valores aleatorios que tiene la misma probabilidad de ocurrencia en cualquier punto de un rango de valores. Se define especificando un límite inferior **a** y un límite superior **b** del rango.

Función de Densidad
$$f(x) = \begin{cases} \dfrac{1}{b-a} \longrightarrow a \leq x \leq b \\ 0 \longrightarrow resto \end{cases}$$

Media
$$\mu = \frac{b+a}{2}$$

Varianza
$$\sigma^2 = \frac{(b-a)^2}{12}$$

Función Acumulada
$$\mathrm{F(x)} = \begin{cases} 0 & \mathrm{x < a} \\ \dfrac{\mathrm{x-a}}{\mathrm{b-a}} & \mathrm{a \leq x \leq b} \\ 1 & \mathrm{b < x} \end{cases}$$



Función Generadora de Rnd con Distribución Uniforme (a,b)
$$X = A + RND(B - A)$$

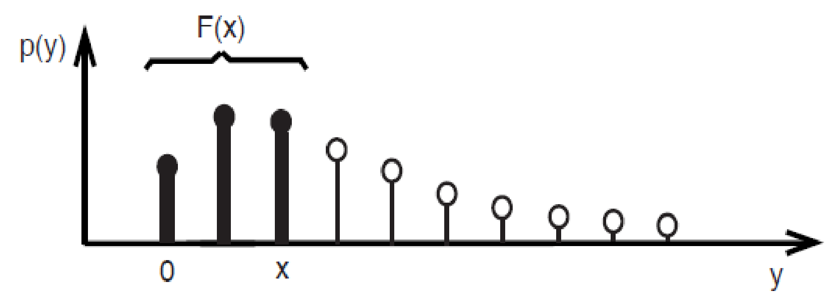
Función con Distribución Poisson (λ):

Poisson solo observa las frecuencias de un evento por intervalo de tiempo. Los valores aleatorios generado por Poisson solo puede ser valores enteros, incluido el cero. **λ** debe ser un valor positivo y no necesariamente entero. **λ** es el numero esperado de ocurrencias en un intervalo de tiempo.

La media y la varianza de una variable aleatoria discreta X que sigue una función de densidad ***f(x)*** se calcula mediante

$$\mu = E(X) = \sum_{x} x p(x) \qquad \sigma^2 = V(X) = E(X-\mu)^2$$

Función de Densidad $p(x) = \begin{cases} \dfrac{e^{-\lambda}\lambda^{x}}{x!} & \text{si } x \in \{0,1,\ldots,\} \\ 0 & \text{resto} \end{cases}$

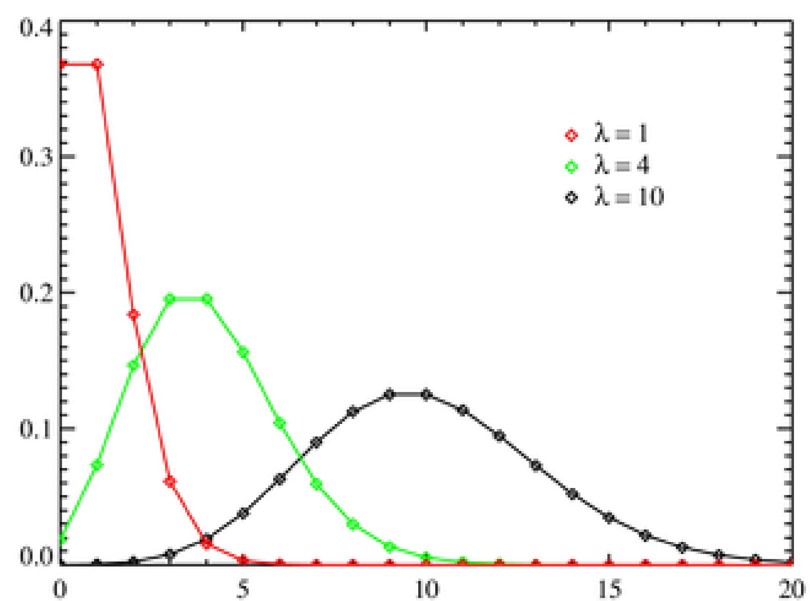


Media = Varianza λ

Función Acumulada $F(x) = \begin{cases} 0 & \text{si } x < 0 \\ e^{-\lambda} \displaystyle\sum_{i=0}^{\lfloor x \rfloor} \frac{\lambda^{i}}{i!} & \text{si } 0 \le x \end{cases}$

Función Generadora de Rnd con Distribución Poisson (λ)

```
P = 1;
X = -1;
A = e^-λ;
Hacer
{
    Generar U = RND(0,1);
    P = P * U;
    X = X + 1;
} mientras (P >= A);
Devolver X;
```



Chi-cuadrado:

- Frecuencia esperada debe ser 5 o más, de no alcanzar ésa cifra se agrupa en intervalos adyacentes.
- El método más adecuado para muestras mayores o iguales a 30 - 50 elementos.
- Si λ^2 calculado es menor o igual al λ^2 tabulado, entonces no se rechaza la hipótesis.

$$\lambda^2 = \sum_{1}^{k} \frac{(f_0 - f_e)^2}{f_e}$$

$$v = k - 1 - m$$

f

- $\mathbf{f_o}$: frecuencia observada
- $\mathbf{f_e}$: frecuencia esperada
- **k**: cantidad de clases o intervalos
- **m**: cantidad de datos empíricos
- **v**: grados de libertad

Kolmogorov-Smirnov

- Solo se aplica a variables continuas.
- En esta prueba ***los grados de libertad están dados por el tamaño de muestra***.
- La prueba es más adecuada para muestras pequeñas, entre 10 y 30 elementos.
- Si *KS* calculado es menor o igual al *KS* tabulado, entonces no se rechaza la hipótesis.

$$KS = \max_{1-K}(|P(f_o)_{AC} - P(f_e)_{AC}|)$$

- $\mathbf{f_o}$: frecuencia observada
- $\mathbf{f_e}$: frecuencia esperada
- **k**: cantidad de clases o intervalos
- $\mathbf{P()_{AC}}$: probabilidad acumulada

Ejemplo: usé el generador de Rnd de una calculadora y obtuve 30 valores posibles. ¿son buenos?

0,15 – 0,22 – 0,41 – 0,65 – 0,84 – 0,81 – 0,62 – 0,45 – 0,32 – 0,07 – 0,11 – 0,29 – 0,58 – 0,73 – 0,93 – 0,97 – 0,79 – 0,55 – 0,35 – 0,09 – 0,99 – 0,51 – 0,35 – 0,02 – 0,19 – 0,24 – 0,98 – 0,10 – 0,31 – 0,17

Intervalo	fo	fe	Po	Pe	Po_{AC}	Pe_{AC}	$\mid Po_{AC} - Pe_{AC} \mid$	$\max(\mid Po_{AC} - Pe_{AC} \mid)$
0,0 - 0,2	8	6	0,27	0,20	0,27	0,20	0,0667	0,0067
0,2 - 0,4	7	6	0,23	0,20	0,50	0,40	0,1000	0,1000
0,4 - 0,6	5	6	0,17	0,20	0,67	0,60	0,0667	0,1000
0,6 - 0,8	4	6	0,13	0,20	0,80	0,80	0,0000	0,1000
0,8 - 1,0	6	6	0,20	0,20	1,00	1,00	0,0000	0,1000

El valor obtenido de ***KS*** de la tabla es de **0,1** y decimos el nivel de significancia es del 10% o hay un riesgo de equivocarnos o existe una aceptación de un 90% que la muestra es uniforme.
En la Tabla de ***KS*** las fila son los grados de libertad (= tamaño de la muestra) y las columnas son los niveles de riesgo.

KS Calculado es **0,1** y es menor al ***KS Tabulado*** es **0,218**.
Entonces decimos que no se rechaza la hipótesis de que eso valores son uniformes

	α						α				
n	0′2	0′1	0′05	0′02	0′01	n	0′2	0′1	0′05	0′02	0′01
1	0′900	0′950	0′975	0′990	0′995	21	0′226	0′259	0′287	0′321	0′344
2	0′684	0′776	0′842	0′900	0′929	22	0′221	0′253	0′281	0′314	0′337
3	0′565	0′636	0′780	0′785	0′829	23	0′216	0′247	0′275	0′307	0′330
4	0′493	0′565	0′624	0′689	0′734	24	0′212	0′242	0′269	0′301	0′323
5	0′447	0′509	0′563	0′627	0′669	25	0′208	0′238	0′264	0′295	0′317
6	0′410	0′468	0′519	0′577	0′617	26	0′204	0′233	0′259	0′290	0′311
7	0′381	0′436	0′483	0′538	0′576	27	0′200	0′229	0′254	0′284	0′305
8	0′358	0′410	0′454	0′507	0′542	28	0′197	0′225	0′250	0′279	0′300
9	0′339	0′387	0′430	0′480	0′513	29	0′193	0′221	0′246	0′275	0′295
10	0′323	0′369	0′409	0′457	0′489	30	0′190	0′218	0′242	0′270	0′290
11	0′308	0′352	0′391	0′437	0′468	31	0′187	0′214	0′238	0′266	0′285
12	0′296	0′338	0′375	0′419	0′449	32	0′184	0′211	0′234	0′262	0′281
13	0′285	0′325	0′361	0′404	0′432	33	0′182	0′208	0′231	0′258	0′277
14	0′275	0′314	0′349	0′390	0′418	34	0′179	0′205	0′227	0′254	0′273
15	0′266	0′304	0′338	0′377	0′404	35	0′177	0′202	0′224	0′251	0′269
16	0′258	0′295	0′327	0′366	0′392	36	0′174	0′199	0′221	0′247	0′265
17	0′250	0′286	0′318	0′355	0′381	37	0′172	0′196	0′218	0′244	0′262
18	0′244	0′279	0′309	0′346	0′371	38	0′170	0′194	0′215	0′241	0′258
19	0′237	0′271	0′301	0′337	0′361	39	0′168	0′191	0′213	0′238	0′255
20	0′232	0′265	0′294	0′329	0′352	40	0′165	0′189	0′21	0′235	0′252
						> 40	$\frac{1'07}{\sqrt{n}}$	$\frac{1'22}{\sqrt{n}}$	$\frac{1'36}{\sqrt{n}}$	$\frac{1'52}{\sqrt{n}}$	$\frac{1'63}{\sqrt{n}}$