FACULTAD DE MATEMÁTICAS
PONTIFICIA UNIVERSIDAD
CATÓLICA DE CHILE

# EYP1113 - Probabilidad y Estadística

## Laboratorio 07


Pilar Tello Hernández
pitello@uc.cl

Facultad de Matemáticas
Departamento de Estadística
Pontificia Universidad Católica de Chile


Segundo Semestre 2021

# Función `lm()`

La función `lm()` sirve para crear modelos de regresión lineal. Para efectos de este laboratorio se utilizará para estimar el intercepto y la pendiente de una recta que prediga `y` respecto a `x`, es decir: $y = \beta_0 + \beta_1 x$. La función se utiliza de la siguiente manera: `lm(y ~ x)`.

# Introducción

El modelo de distribución de probablidad apropiado para describir un fenómeno es generalmente, desconocido.

Bajo ciertas circunstancias, las propiedades básicas del proceso físico subyacente del fenómeno aleatorio sugiere la forma de la distribución de probabilidades.

**Ejemplos**

- Cumple vs. No cumple $\rightarrow$ Bernoulli.
- Número de "eventos" en periodos $\rightarrow$ Poisson.
- Tiempos de duración o espera $\rightarrow$ Exponencial.
- Suma de eventos individuales $\rightarrow$ Normal.
- Condiciones extremas de un proceso $\rightarrow$ Valor extremo.

# Introducción

En muchas situaciones, la distribución de probabilidad debe ser determinada empíricamente a partir de los datos.

Inicialmente, aproximaciones gráficas (Histograma v/s Densidad) nos pueden ayudar a inferir "visualmente" sobre la distribución.

También, con datos disponibles, pueden obtenerse los gráficos de probabilidad (Probability Papers) para distribuciones dadas (si los puntos están en línea recta, la distribución es apropiada).

Por último, dada una distribución a priori puede evaluarse la "bondad de ajuste" (Test $\chi^2$, Test de Kolmogorov-Smirnov o el Test de Anderson-Darling, entre otros).

En este laboratorio nos enfocaremos en la construcción de un gráfico de probabilidad.

# Gráficos de probabilidad

Es la representación gráfica de los datos observados y sus correspondientes frecuencias acumuladas.

Para un conjunto de $N$ observaciones, $x_1, \ldots, x_N$ ordenados de menor a mayor, el $m$-ésimo valor es graficado contra la probabilidad acumulada de $\frac{m}{N+1}$.

La utilidad del "papel" de probabilidad es reflejar "el ajuste" que presentan los datos con respecto a la distribución subyacente.

La linealidad o falta de esta nos indica lo adecuado o inadecuado de la distribución.

# Gráficos de probabilidad: Distribución Normal

Sean $x_{(1)}, \ldots, x_{(N)}$ observaciones ordenadas de menor a mayor y $p_1 = \frac{1}{N+1}, \ldots, p_N = \frac{N}{N+1}$ sus respectivas probabilidades empíricas.

Calculemos los percentiles teóricos $\Phi^{-1}(p_i)$ de una distribución Normal Estándar para cada $p_i$, con $i = 1, \ldots, N$.

Si los $x$'s distribuyen $Normal(\mu, \sigma)$ entonces la siguiente relación lineal se debe cumplir:

$$x_{(q)} = \mu + \sigma \cdot \Phi^{-1}(p_q)$$

# Gráficos de probabilidad: Distribución Log-Normal

Sean $x_{(1)}, \ldots, x_{(N)}$ observaciones ordenadas de menor a mayor y $p_1 = \frac{1}{N+1}, \ldots, p_N = \frac{N}{N+1}$ sus respectivas probabilidades empíricas.

Calculemos los percentiles teóricos $\Phi^{-1}(p_i)$ de una distribución Normal Estándar para cada $p_i$, con $i = 1, \ldots, N$.

Si los $x$'s distribuyen $log - Normal(\lambda, \zeta)$ entonces la siguiente relación lineal se debe cumplir:

$$\ln x_{(q)} = \lambda + \zeta \cdot \Phi^{-1}(p_q)$$

# Aplicación

1. Construya un gráfico de probabilidad para una distribución Normal. Simule datos provenientes de una distribución Normal para probar el gráfico construido.
2. Construya un gráfico de probabilidad para una distribución Log-Normal. Simule datos provenientes de una distribución Log-Normal para probar el gráfico construido.
3. Simule datos provenientes de una distribución $Gamma(k = 10, \nu = 0{,}1)$ y ajuste los modelos anteriores en un histograma según la estimación obtenida del gráfico de probabilidad.

Usando el siguiente código, se puede obtener las estimaciones para $\beta$ (shape) y $\eta$ (scale) de la distribución Weibull.

```
QQ.Weibull = function(y){
x=sort(y)
N=length(y)
p=(1:N)/(N+1)
plot(log(x)~log(-log(1-p)),
pch = 20, col = "darkblue", bty = "n", las = 1,
main = expression("QQ-Weibull"),
ylab = expression(log(x[p])),
xlab = expression(log(-log(1-p))))
abline(lm(log(x) ~ log(-log(1-p))), lwd = 3, col = "darkorange")
aux = lm(log(x) ~ log(-log(1-p)))
aux
}

eta = as.numeric(exp(QQ.Weibull(X)$coef[1]))
beta = as.numeric(1/QQ.Weibull(X)$coef[2])
```

# ¿Por qué?

La función de probabilidad acumulada de la distribución Weibull es $1 - e^{-(x/\eta)^\beta}$ por lo que al igualar ésta expresión a $p$ e intentar obtener una función lineal, se tiene

$$\begin{aligned} p &= 1 - e^{-\left(\frac{x}{\eta}\right)^\beta} \\ -log(1-p) &= \left(\frac{x}{\eta}\right)^\beta \\ log(\eta) + \frac{1}{\beta} log(-log(1-p)) &= log(x) \end{aligned}$$

Luego, la regresión lineal $log(x) \sim log(-log(1-p))$ entrega pendiente $= 1/\beta$ y la intercepto $= log(\eta)$.
Por lo tanto, $\hat{\beta} = 1/\text{pendiente}$ y $\hat{\eta} = exp(\text{intercepto})$

# Actividad

Simule una muestra aleatoria Weibull y estime sus parámetros. Grafique su histograma y agregue línea de densidad Weibull con los parámetros obtenidos.