Notas del Curso de Estadística

Carlos E. Martínez Rodríguez

19 de septiembre de 2019

Estimación por intervalos
- Para la media
- Intervalos de confianza sobre la varianza
- Intervalos de confianza para proporciones

Intervalos de confianza para dos muestras
- Varianzas conocidas
- Varianzas desconocidas
  - $\sigma_1^2 = \sigma_2^2 = \sigma$
  - $\sigma_1^2 \neq \sigma_2^2$

Intervalos de confianza para Distribuciones pareadas

Preliminares

Introducción

Gráfico y Lectura de Datos

Estadística Descriptiva

Intervalos de Confianza y Pruebas de Hipótesis

Análisis de Regresión Lineal

Estadística Multivariada

Anexo 1: Simulación de Variables Aleatorias

Distribución Uniforme
Distribución Normal
Distribución Gamma
Distribución Beta
Distribución Exponencial
Distribución t-Student
Distribución χ^2
Distribución Binomial
Distribución Geométrica
Distribución Poisson
Miscelanea de instrucciones

Intervalos de confianza para μ

Recordemos que S^2 es un estimador insesgado de σ^2

Definición

Sean $\hat{\theta}_1$ y $\hat{\theta}_2$ dos estimadores insesgados de θ, parámetro poblacional. Si $\sigma^2_{\hat{\theta}_1} < \sigma^2_{\hat{\theta}_2}$, decimos que $\hat{\theta}_1$ un estimador más eficaz de θ que $\hat{\theta}_2$.

Algunas observaciones que es preciso realizar

a) Para poblaciones normales, $\overline{X}$ y $\tilde{X}$ son estimadores insesgados de μ, pero con $\sigma^2_{\overline{X}} < \sigma^2_{\tilde{X}_2}$.

b) Para las estimaciones por intervalos de θ, un intervalo de la forma $\hat{\theta}_L < \theta < \hat{\theta}_U$, $\hat{\theta}_L$ y $\hat{\theta}_U$ dependen del valor de $\hat{\theta}$.

c) Para $\sigma^2_{\overline{X}} = \frac{\sigma^2}{n}$, si $n \to \infty$, entonces $\hat{\theta} \to \mu$.

Intervalos de confianza para μ

d) Para $\hat{\theta}$ se determinan $\hat{\theta}_L$ y $\hat{\theta}_U$ de modo tal que

$$P\left\{\hat{\theta}_L < \hat{\theta} < \hat{\theta}_U\right\} = 1 - \alpha, \tag{1}$$

con $\alpha \in (0, 1)$. Es decir, $\theta \in \left(\hat{\theta}_L, \hat{\theta}_U\right)$ es un intervalo de confianza del $100\,(1 - \alpha)$ %.

e) De acuerdo con el TLC se espera que la distribución muestral de $\overline{X}$ se distribuye aproximadamente normal con media $\mu_X = \mu$ y desviación estándar $\sigma_{\overline{X}} = \frac{\sigma}{\sqrt{n}}$.

UACM Universidad Autónoma de la Ciudad de México Nada humano me es ajeno.

Intervalos de confianza para μ

Para $Z_{\alpha/2}$ se tiene $P\left\{-Z_{\alpha/2} < Z < Z_{\alpha/2}\right\} = 1-\alpha$, donde $Z = \frac{\overline{X}-\mu}{\sigma/\sqrt{n}}$. Entonces $P\left\{-Z_{\alpha/2} < \frac{\overline{X}-\mu}{\sigma/\sqrt{n}} < Z_{\alpha/2}\right\} = 1-\alpha$ es equivalente a $P\left\{\overline{X} - Z_{\alpha/2}\frac{\sigma}{\sqrt{n}} < \mu < \overline{X} + Z_{\alpha/2}\frac{\sigma}{\sqrt{n}}\right\} = 1-\alpha$

f) Si $\overline{X}$ es la media muestral de una muestra de tamaño n de una población con varianza conocida σ^2, el intervalo de confianza de $100\,(1-\alpha)$ % para μ es $\mu \in \left(\overline{x} - z_{\alpha/2}\frac{\sigma}{\sqrt{n}}, \overline{x} + z_{\alpha/2}\frac{\sigma}{\sqrt{n}}\right)$.

g) Para muestras pequeñas de poblaciones no normales, no se puede esperar que el grado de confianza sea preciso.

h) Para $n \geq 30$, con distribución de forma no muy sesgada, se pueden tener buenos resultados.

UACM Universidad Autónoma de la Ciudad de México Nada humano me es ajeno.

UACM Universidad Autónoma de la Ciudad de México Nada humano me es ajeno.

Intervalos de confianza para μ

Teorema

Si $\overline{X}$ es un estimador de μ, podemos tener $100\,(1-\alpha)$ % de confianza en que el error no excederá a $z_{\alpha/2}\frac{\sigma}{\sqrt{n}}$, error entre $\overline{X}$ y μ.

Teorema

Si $\overline{X}$ es un estimador de μ, podemos tener $100\,(1-\alpha)$ % de confianza en que el error no excederá una cantidad e cuando el tamaño de la muestra es

$$n = \left(\frac{z_{\alpha/2}\sigma}{e}\right)^2.$$

Nota

Para intervalos unilaterales

$$P\left\{\frac{\overline{X}-\mu}{\sigma/\sqrt{n}} < Z_\alpha\right\} = 1-\alpha$$

Intervalos de confianza para μ

equivalentemente

$$P\left\{\mu < \overline{X} + Z_\alpha \frac{\sigma}{\sqrt{n}}\right\} = 1 - \alpha.$$

Si $\overline{X}$ es la media de una muestra aleatoria de tamaño n a partir de una población con varianza σ^2, los límites de confianza unilaterales del $100\,(1-\alpha)$ % de confianza para μ están dados por

- Límite unilateral superior: $\overline{x} + z_\alpha \frac{\sigma}{\sqrt{n}}$
- Límite unilateral inferior: $\overline{x} - z_\alpha \frac{\sigma}{\sqrt{n}}$

UACM Universidad Autónoma de la Ciudad de México Nada humano me es ajeno.

Intervalos de confianza para μ

- Para σ desconocida recordar que $T = \frac{\overline{x}-\mu}{s/\sqrt{n}} \sim t_{n-1}$, donde s es la desviación estándar de la muestra. Entonces

$$P\left\{-t_{\alpha/2} < T < t_{\alpha/2}\right\} = 1-\alpha, \text{equivalentemente}$$
$$P\left\{\overline{X} - t_{\alpha/2}\frac{s}{\sqrt{n}} < \mu < \overline{X} + t_{\alpha/2}\frac{s}{\sqrt{n}}\right\} = 1-\alpha.$$

- Un intervalo de confianza del $100\,(1-\alpha)$ % de confianza para μ, σ^2 desconocida y población normal es $\mu \in \left(\overline{x} - t_{\alpha/2}\frac{s}{\sqrt{n}}, \overline{x} + t_{\alpha/2}\frac{s}{\sqrt{n}}\right)$, donde $t_{\alpha/2}$ es una t-student con $\nu = n-1$ grados de libertad.
- Los límites unilaterales para μ con σ desconocida son $\overline{X} - t_{\alpha/2}\frac{s}{\sqrt{n}}$ y $\overline{X} + t_{\alpha/2}\frac{s}{\sqrt{n}}$.

Intervalos de confianza para μ

- Cuando la población no es normal, σ desconocida y $n \geq 30$, σ se puede reemplazar por s para obtener el intervalo de confianza para muestras grandes:

$$\overline{X} \pm t_{\alpha/2} \frac{s}{\sqrt{n}}.$$

- El estimador de $\overline{X}$ de μ, σ desconocida, la varianza de $\sigma^2_{\overline{X}} = \frac{\sigma^2}{n}$, el error estándar de $\overline{X}$ es $\sigma/\sqrt{n}$.
- Si σ es desconocida y la población es normal, $s \to \sigma$ y se incluye el error estándar $s/\sqrt{n}$, entonces

$$\overline{x} \pm t_{\alpha/2} \frac{s}{\sqrt{n}}.$$

Intervalos de confianza para σ^2

Supongamos que X se distribuye normal (μ, σ^2), desconocidas. Sea $X_1, X_2, \ldots, X_n$ muestra aleatoria de tamaño n , s^2 la varianza muestral.
Se sabe que $X^2 = \frac{(n-1)s^2}{\sigma^2}$ se distribuye χ^2_{n-1} grados de libertad. Su intervalo de confianza es

$$\begin{aligned} &P\left\{\chi^2_{1-\frac{\alpha}{2},n-1} \leq \chi^2 \leq \chi^2_{\frac{\alpha}{2},n-1}\right\} = 1-\alpha \\ &P\left\{\chi^2_{1-\frac{\alpha}{2},n-1} \leq \frac{(n-1)s^2}{\sigma^2} \leq \chi^2_{\frac{\alpha}{2},n-1}\right\} = 1-\alpha \\ &P\left\{\frac{(n-1)s^2}{\chi^2_{\frac{\alpha}{2},n-1}} \leq \sigma^2 \leq \frac{(n-1)s^2}{\chi^2_{1-\frac{\alpha}{2},n-1}}\right\} = 1-\alpha \end{aligned} \tag{2}$$

es decir

Intervalos de confianza para σ^2

$$\sigma^2 \in \left[\frac{(n-1)\, s^2}{\chi^2_{\frac{\alpha}{2},n-1}}, \frac{(n-1)\, s^2}{\chi^2_{1-\frac{\alpha}{2},n-1}} \right] \tag{3}$$

los intervalos unilaterales son

$$\sigma^2 \in \left[\frac{(n-1)\, s^2}{\chi^2_{\frac{\alpha}{2},n-1}}, \infty \right] - \tag{4}$$

$$\sigma^2 \in \left[-\infty, \frac{(n-1)\, s^2}{\chi^2_{1-\frac{\alpha}{2},n-1}} \right] \tag{5}$$

Intervalos de confianza para proporciones

Supongamos que se tienen una muestra de tamaño n de una población grande pero finita, y supongamos que X, $X \leq n$, pertenecen a la clase de interés, entonces

$$\hat{p} = \frac{\overline{X}}{n}$$

es el estimador puntual de la proporción de la población que pertenece a dicha clase.

n y p son los parámetros de la distribución binomial, entonces $\hat{p} \sim N\left(p, \frac{p(1-p)}{n}\right)$ aproximadamente si p es distinto de 0 y 1; o si n es suficientemente grande. Entonces

$$Z = \frac{\hat{p} - p}{\sqrt{\frac{p(1-p)}{n}}} \sim N\left(0, 1\right), \text{aproximadamente.}$$

UACM Universidad Autónoma de la Ciudad de México Nada humano me es ajeno.

Intervalos de confianza para proporciones

entonces

$$
\begin{aligned}
1-\alpha &= P\left\{-z_{\alpha/2} \leq \frac{\hat{p}-p}{\sqrt{\frac{p(1-p)}{n}}} \leq z_{\alpha/2}\right\} \\
&= P\left\{\hat{p} - z_{\alpha/2}\sqrt{\frac{p\,(1-p)}{n}} \leq p \leq \hat{p} + z_{\alpha/2}\sqrt{\frac{p\,(1-p)}{n}}\right\}
\end{aligned}
$$

con $\sqrt{\frac{p(1-p)}{n}}$ error estándar del estimador puntual p. Una solución para determinar el intervalo de confianza del parámetro p (desconocido) es

Intervalos de confianza para proporciones

$$1-\alpha = P\left\{\hat{p} - z_{\alpha/2}\sqrt{\frac{\hat{p}\,(1-\hat{p})}{n}} \leq p \leq \hat{p} + z_{\alpha/2}\sqrt{\frac{\hat{p}\,(1-\hat{p})}{n}}\right\}$$

entonces los intervalos de confianza, tanto unilaterales como de dos colas son:

- $p \in \left(\hat{p} - z_{\alpha/2}\sqrt{\frac{\hat{p}(1-\hat{p})}{n}}, \hat{p} + z_{\alpha/2}\sqrt{\frac{\hat{p}(1-\hat{p})}{n}}\right)$
- $p \in \left(-\infty, \hat{p} + z_{\alpha/2}\sqrt{\frac{\hat{p}(1-\hat{p})}{n}}\right)$
- $p \in \left(\hat{p} - z_{\alpha/2}\sqrt{\frac{\hat{p}(1-\hat{p})}{n}}, \infty\right)$

para minimizar el error estándar, se propone que el tamaño de la muestra sea $n = \left(\frac{z_{\alpha/2}}{E}\right)^2 p\,(1-p)$, donde $E = \mid p - \hat{p} \mid$.

Varianzas Conocidas

Sean X_1 y X_2 variables aleatorias independientes. X_1 con media desconocida μ_1 y varianza conocida σ_1^2; y X_2 con media desconocida μ_2 y varianza conocida σ_2^2. Se busca encontrar un intervalo de confianza de $100\,(1-\alpha)$ % de la diferencia entre medias μ_1 y μ_2.

Sean $X_{11}, X_{12}, \ldots, X_{1n_1}$ muestra aleatoria de n_1 observaciones de X_1, y sean $X_{21}, X_{22}, \ldots, X_{2n_2}$ muestra aleatoria de n_2 observaciones de X_2.

Sean $\overline{X}_1$ y $\overline{X}_2$, medias muestrales, entonces el estadśitico

$$Z = \frac{\left(\overline{X}_1 - \overline{X}_2\right) - (\mu_1 - \mu_2)}{\sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}} \sim N\,(0,1)\,, \tag{6}$$

si X_1 y X_2 son normales o aproximadamente normales si se aplican las condiciones del Teorema de Límite Central respectivamente.

Varianzas conocidas

Entonces se tiene

$$
\begin{aligned}
1-\alpha &= P\left\{-Z_{\alpha/2} \leq Z \leq Z_{\alpha/2}\right\} \\
&= P\left\{-Z_{\alpha/2} \leq \frac{(\overline{X}_1 - \overline{X}_2) - (\mu_1 - \mu_2)}{\sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}} \leq Z_{\alpha/2}\right\} \\
&= P\left\{(\overline{X}_1 - \overline{X}_2) - Z_{\alpha/2}\sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}} \leq \mu_1 - \mu_2 \leq \right. \\
&\qquad \left. (\overline{X}_1 - \overline{X}_2) + Z_{\alpha/2}\sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}\right\}
\end{aligned}
$$

UACM Universidad Autónoma de la Ciudad de México Nada humano me es ajeno.

UACM Universidad Autónoma de la Ciudad de México Nada humano me es ajeno.

Varianzas conocidas

Entonces los intervalos de confianza unilaterales y de dos colas al $(1-\alpha)$ % de confianza son

- $\mu_1 - \mu_2 \in$
$$\left[\left(\overline{X}_1 - \overline{X}_2\right) - Z_{\alpha/2}\sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}, \left(\overline{X}_1 - \overline{X}_2\right) + Z_{\alpha/2}\sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}\right]$$
- $\mu_1 - \mu_2 \in \left[-\infty, \left(\overline{X}_1 - \overline{X}_2\right) + Z_{\alpha/2}\sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}\right]$
- $\mu_1 - \mu_2 \in \left[\left(\overline{X}_1 - \overline{X}_2\right) - Z_{\alpha/2}\sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}, \infty\right]$

Varianzas conocidas

Nota

Si σ_1 y σ_2 son conocidas, o por lo menos se conoce una aproximación, y los tamaños de las muestras n_1 y n_2 son iguales, $n_1 = n_2 = n$, se puede determinar el tamaño de la muestra para que el error al estimar $\mu_1 - \mu_2$ usando $\overline{X}_1 - \overline{X}_2$ sea menor que E (valor del error deseado) al $(1 - \alpha)$ % de confianza. El tamaño n de la muestra requerido para cada muestra es

$$n = \left(\frac{Z_{\alpha/2}}{E} \right)^2 \left(\sigma_1^2 + \sigma_2^2 \right).$$

UACM Universidad Autónoma de la Ciudad de México Nada humano me es ajeno.

UACM Universidad Autónoma de la Ciudad de México Nada humano me es ajeno.

Varianzas desconocidas

- Si $n_1, n_2 \geq 30$ se pueden utilizar los intervalos de la distribución normal para varianza conocida
- Si n_1, n_2 son muestras pequeñas, supongase que las poblaciones para X_1 y X_2 son normales con varianzas desconocidas y con base en el intervalo de confianza para distribuciones t-student

$\sigma_1^2 = \sigma_2^2 = \sigma$

Supongamos que X_1 es una variable aleatoria con media μ_1 y varianza σ_1^2, X_2 es una variable aleatoria con media μ_2 y varianza σ_2^2. Todos los parámetros son desconocidos. Sin embargo supóngase que es razonable considerar que $\sigma_1^2 = \sigma_2^2 = \sigma^2$.

Nuevamente sean X_1 y X_2 variables aleatorias independientes. X_1 con media desconocida μ_1 y varianza muestral S_1^2; y X_2 con media desconocida μ_2 y varianza muestral S_2^2. Dado que S_1^2 y S_2^2 son estimadores de σ_1^2, se propone el estimador S de σ^2 como

$$S_p^2 = \frac{(n_1 - 1)\, S_1^2 + (n_2 - 1)\, S_2^2}{n_1 + n_2 - 2},$$

entonces, el estadístico para $\mu_1 - \mu_2$ es

$\sigma_1^2 = \sigma_2^2 = \sigma$

$$t_\nu = \frac{(\overline{X}_1 - \overline{X}_2) - (\mu_1 - \mu_2)}{S_p\sqrt{\frac{1}{n_1} + \frac{1}{n_2}}}$$

donde t_ν es una t de student con $\nu = n_1 + n_2 - 2$ grados de libertad.

Por lo tanto

$$1 - \alpha = P\left\{-t_{\alpha/2,\nu} \leq t \leq t_{\alpha/2,\nu}\right\}$$

$$= P\left\{(\overline{X}_1 - \overline{X}_2) - t_{\alpha/2,\nu}S_p\sqrt{\frac{1}{n_1} + \frac{1}{n_2}} \leq \right.$$

$$\left. t \leq (\overline{X}_1 - \overline{X}_2) + t_{\alpha/2,\nu}S_p\sqrt{\frac{1}{n_1} + \frac{1}{n_2}}\right\}$$

$\sigma_1^2 = \sigma_2^2 = \sigma$

luego, los intervalos de confianza del $(1-\alpha)$ % para $\mu_1 - |mu_2$ son

- $\mu_1 - \mu_2 \in \left[\left(\overline{X}_1 - \overline{X}_2\right) - t_{\alpha/2,\nu} S_p \sqrt{\frac{1}{n_1} + \frac{1}{n_2}}, \left(\overline{X}_1 - \overline{X}_2\right) + t_{\alpha/2,\nu} S_p \sqrt{\frac{1}{n_1} + \frac{1}{n}}\right.$
- $\mu_1 - \mu_2 \in \left[-\infty, \left(\overline{X}_1 - \overline{X}_2\right) + t_{\alpha/2,\nu} S_p \sqrt{\frac{1}{n_1} + \frac{1}{n_2}}\right]$
- $\mu_1 - \mu_2 \in \left[\left(\overline{X}_1 - \overline{X}_2\right) - t_{\alpha/2,\nu} S_p \sqrt{\frac{1}{n_1} + \frac{1}{n_2}}, \infty\right]$

UACM Universidad Autónoma de la Ciudad de México Nada humano me es ajeno.

Varianzas desconocidas

Si no se tiene certeza de que $\sigma_1^2 = \sigma_2^2$, se propone el estadístico

$$t^* = \frac{(\overline{X}_1 - \overline{X}_2) - (\mu_1 - \mu_2)}{\sqrt{\frac{S_1^2}{n_1} + \frac{S_2^2}{n_2}}} \tag{7}$$

que se distribuye t-student con ν grados de libertad, donde

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

Varianzas desconocidas

Entonces el intervalo de confianza de aproximadamente el $100\,(1-\alpha)\ \%$ para $\mu_1-\mu_2$ con $\sigma_1^2 \neq \sigma_2^2$ es

$$\mu_1-\mu_2 \in \left[\left(\overline{X}_1-\overline{X}_2\right)-t_{\alpha/2,\nu}\sqrt{\frac{S_1^2}{n_1}+\frac{S_2^2}{n_2}},\right.$$

$$\left.\left(\overline{X}_1-\overline{X}_2\right)+t_{\alpha/2,\nu}\sqrt{\frac{S_1^2}{n_1}+\frac{S_2^2}{n_2}}\right]$$

Distribuciones Pareadas

Supongamos que se toman dos muestras aleatorias independientes de las dos poblaciones de interés

```
> a <- -2;
> b <- 2;
> N <- 10000;
> t <- runif(N, min =a, max=b)
> hist(t,freq = FALSE,
+      col = "blue",
+      xlab = 'Graficando uniformes [-1,1]',
+      density = 30,
+      main = "DISTRIBUCION UNIFORME EN EL INTERVALO
> curve(dunif(x, min = a, max = b),
+       from = -5, to = 5,
+       n = N,
+       col = "darkblue",
+       lwd = 2,
+       add = TRUE,
+       yaxt = "n",
+       ylab = 'probability')
> #------------------------------------------------
```

```
> # utilizando la libreria ggplot2
> #-------------------------------------------------------
> library(ggplot2)
> #-------------------------------------------------------
> xvals <- data.frame(x = c(a, b)) #Range for x-values
> ggplot(data.frame(x = xvals), aes(x = x)) +
+   xlim(c(a, b)) +
+   ylim(c(0, 1/4)) +
+   stat_function(fun = dunif,
+                 args = list(min = a, max = b),
+                 geom = "area",
+                 fill = "green",
+                 alpha = 0.35) +
+   labs(x = "\n u",
+        y = "f(u) \n",
+        title = "Uniform Distribution With a = -2 and
+   theme(plot.title = element_text(hjust = 0.5),
+         axis.title.x = element_text(face="bold",
```

```
+                                        colour="blue",
+                                        size = 12),
+          axis.title.y = element_text(face="bold",
+                                        colour="blue",
+                                        size = 12)) +
+    geom_vline(xintercept = a,
+               linetype = "dashed",
+               colour = "red") +
+    geom_vline(xintercept = b,
+               linetype = "dashed",
+               colour = "red")
> #-------------------------------------------------------
> # ggplot Histogram
> # Uniform Distribution With a = -2 and b = 2
> #-------------------------------------------------------
> unifs <- runif(n = 10000, min = a, max = b)
> ggplot(data = NULL, aes(x = unifs)) +
+    geom_histogram(binwidth = 0.25, boundary = 2) +
```

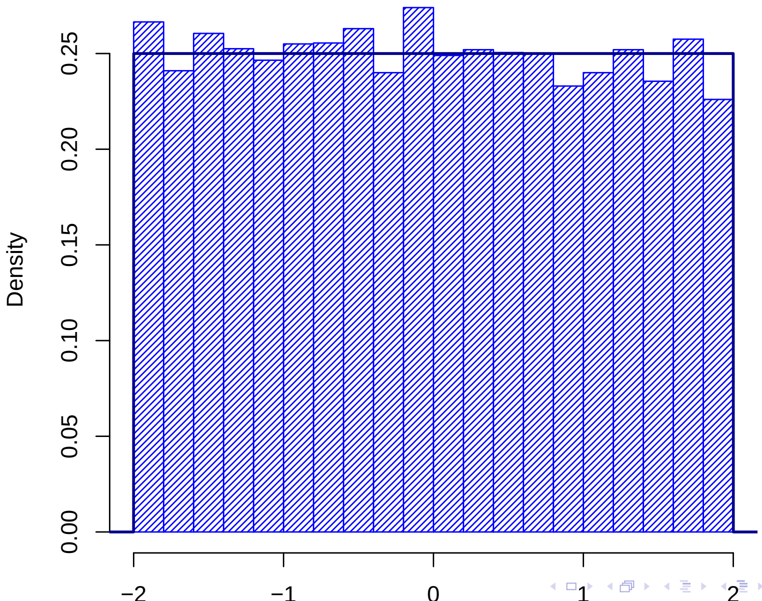
```
+   xlim(c(-3, 3)) +
+   ylim(c(-10, 800)) +
+   labs(x = "\n u",
+        y = "f(u) \n",
+        title = "Uniform Distribution With a = -2 and
+   theme(plot.title = element_text(hjust = 0.5),
+         axis.title.x = element_text(face="bold",
+                                     colour="brown",
+                                     size = 12),
+         axis.title.y = element_text(face="bold",
+                                     colour="brown",
+                                     size = 12))
> #--------------------------------------------------
> # Standard Uniform Distribution With a = 0 and b = 1
> #--------------------------------------------------
> a=0;
> b=1;
> std_unifs <- runif(n = N, min = a, max = b)
```

```
> #------------------------------------------------------------
> # ggplot Histogram
> #------------------------------------------------------------
> ggplot(data = NULL, aes(x = std_unifs)) +
+   geom_histogram(binwidth = 0.05, boundary = 1) +
+   xlim(c(-0.05, 1.05)) +
+   labs(x = "\n u", y = "f(u) \n",
+        title = "Uniform Distribution With a = 0 and
+   theme(plot.title = element_text(hjust = 0.5),
+         axis.title.x = element_text(face="bold",
+                                     colour="brown",
+                                     size = 12),
+         axis.title.y = element_text(face="bold",
+                                     colour="brown",
+                                     size = 12))
> #------------------------------------------------------------
> # otra alternativa
> #------------------------------------------------------------
```

```
> t <- runif(N)*10; head(t)
[1] 0.3427805 0.6261653 5.7132067 0.9506013 1.7658213 7
> tf<- floor(t);head(t)
[1] 0.3427805 0.6261653 5.7132067 0.9506013 1.7658213 7
> table(tf)
tf
   0    1    2    3    4    5    6    7    8    9
 980 1014 1018  964 1003 1020  937 1031 1029 1004
> barplot(table(tf),
+         xlab="NUMEROS GENERADOS",
+         ylab="NUMERO DE OCURRENCIAS",
+         border="red",
+         main = "Ejemplo de generacion de Variables A
> #-----------------------------------------------------
> # EJERCICIO 1) GENERAR UNA FUNCION QUE GRAFIQUE VARI
> # VARIANDO EL NUMERO DE V.A. GENERADAS
```

```
> # EJERCICIO 2) MEJORAR LOS GRAFICOS
> #-------------------------------------------------
```

DISTRIBUCION UNIFORME EN EL INTERVALO [−1,1]



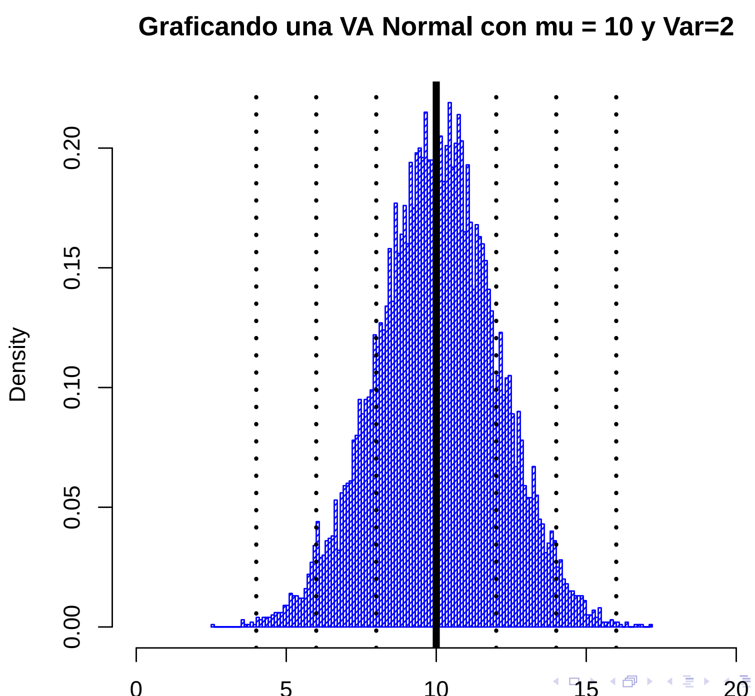
```
> #setwd("~/Nextcloud/TodoMundoFiles/Curso Estadistica
> # ------------------------------------------------
> # PARA UNA DISTRIBUCION NORMAL ESTANDAR
> # ------------------------------------------------
> # Para determinar los valores de Z_{alpha/2}
> # Ejecutar los siguientes comandos
> # ------------------------------------------------
> z <- qnorm(0.05);  print(z)
[1] -1.644854
> z <- qnorm(0.01);  print(z)
[1] -2.326348
> z <- qnorm(0.025); print(z)
[1] -1.959964
> z <- qnorm(0.1);   print(z)
[1] -1.281552
```

```
> z <- qnorm(0.001); print(z)

[1] -3.090232

> # ----------------------------------------------------
> # Para una distribucion Normal(mu,sigma2), los cuart
> # Q1 <- 0.25
> # Q2 <- 0.5
> # Q3 <- 0.75
> # la funcion a utilizar es: qnorm(p, mean= mu, sd =
> # ----------------------------------------------------
> mu     <- 2;
> sigma2 <- 4;
> q1     <- 0.25;Q1 <- qnorm(q1,mean=mu,sd = sigma2);p

[1] -0.697959

> q2     <- 0.5; Q2 <- qnorm(q2,mean=mu,sd = sigma2);p

[1] 2

> q3     <- 0.75;Q3 <- qnorm(q3,mean=mu,sd = sigma2);p
```

```
[1] 4.697959
> # ------------------------------------------------
> # Comparemos los resultados anteriores con las salic
> # comandos p(norm)
> z <- pnorm(1.64);  print(z)
[1] 0.9494974
> z <- pnorm(2.32);  print(z)
[1] 0.9898296
> z <- pnorm(1.96);  print(z)
[1] 0.9750021
> z <- pnorm(1.28);  print(z)
[1] 0.8997274
> # ------------------------------------------------
> # ------------------------------------------------
> # Ahora proporcionemos una bonita grafica de una dis
```

```
> # ----------------------------------------------------
> N      <- 10000
> mu     <- 10
> sigma2 <- 2
> x      <- rnorm(N,mean=mu, sd=sigma2)
> hist(x,
+      breaks=150,
+      xlim=c(0,20),
+      freq=FALSE,
+      col = "blue",
+      xlab = 'N(mu,sigma2)',
+      density = 30,
+      main = "Graficando una VA Normal con mu = 10 y
> abline(v=mu, lwd=5)
> abline(v=c(4,6,8,12,14,16), lwd=3,lty=3)
> #
> #
> #
```

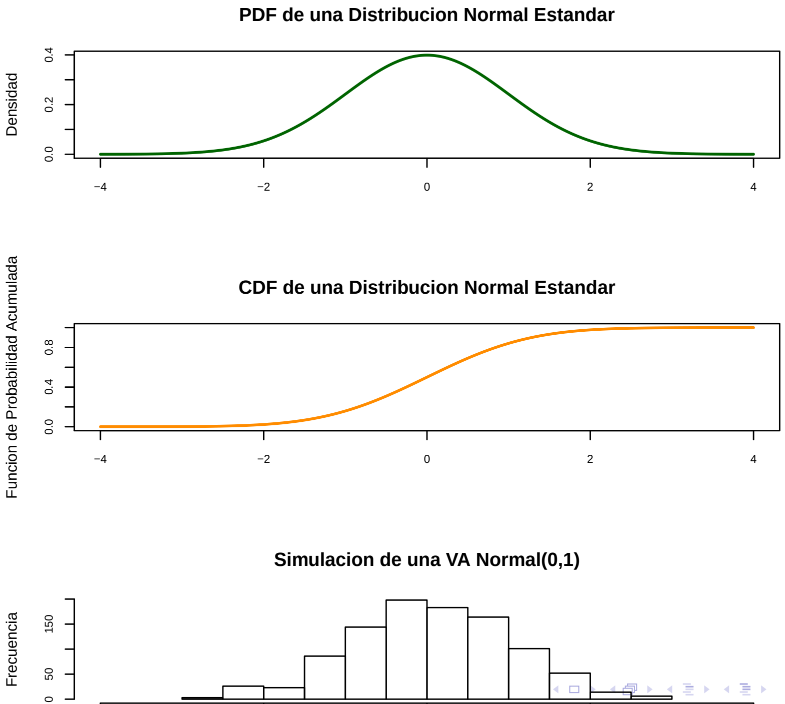
```
> # --------------------------------------------------
```


Graficando una VA Normal con mu = 10 y Var=2
Density
0.00
0.05
0.10
0.15
0.20
0
5
10
15
20

```
> set.seed(3000)
> xseq            <- seq(-4,4,.01)
> densities       <- dnorm(xseq, 0,1)
> cumulative      <- pnorm(xseq, 0, 1)
> randomdeviates <- rnorm(1000,0,1)
> par(mfrow=c(3,1))
> #par(mfrow=c(1,3), mar=c(3,4,4,2))
> plot(xseq,
+      densities,
+      col="darkgreen",
+      xlab="",
+      ylab="Densidad",
+      type="l",
+      lwd=2,
+      cex=2,
+      main="PDF de una Distribucion Normal Estandar",
+      cex.axis=.8)
> plot(xseq,
```

```
+       cumulative,
+       col="darkorange",
+       xlab="",
+       ylab="Funcion de Probabilidad Acumulada",
+       type="l",
+       lwd=2,
+       cex=2,
+       main="CDF de una Distribucion Normal Estandar",
+       cex.axis=.8)
> x<- rnorm(1000,0,1)
> hist(x,
+       main="Simulacion de una VA Normal(0,1)",
+       ylab= 'Frecuencia',
+       xlab='Histograma de una VA Normal(0,1)',
+       cex.axis=.8,
+       xlim=c(-4,4))
```


UACM Universidad Autónoma de la Ciudad de México Nada humano me es ajeno.


PDF de una Distribucion Normal Estandar
Densidad
0.0
0.2
0.4
−4
−2
0
2
4
CDF de una Distribucion Normal Estandar
Funcion de Probabilidad Acumulada
0.0
0.4
0.8
−4
−2
0
2
4
Simulacion de una VA Normal(0,1)
Frecuencia
0
50
150

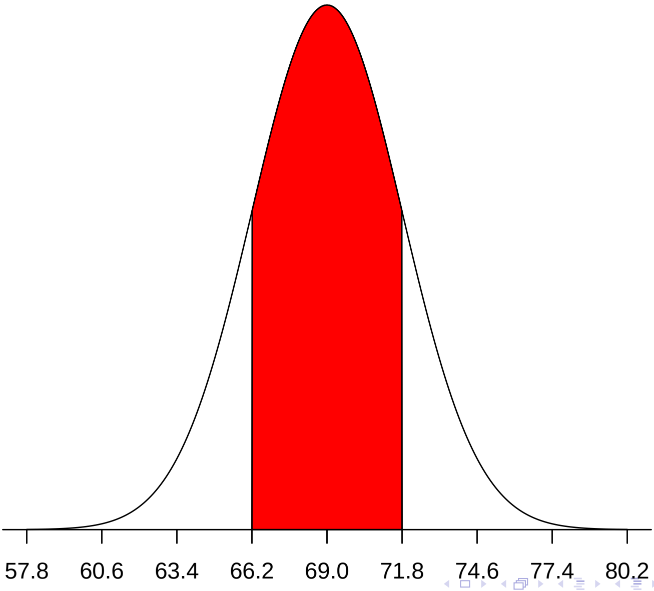
Ahora veamos un ejemplo más práctico

```
> mu           <- 69;
> sigma2       <- 2.8;
> DE_rellenar <- 1;
> LI <- mu-sigma2*DE_rellenar
> LS <- mu+sigma2*DE_rellenar
> #Se generan valores equiespaciados entre -4 y 4 desv
> #alrededor de la media
> x <- seq(-4, 4, length = 1000) * sigma2 + mu
> y <- dnorm(x, mu, sigma2)
> plot(x, y,
+      type="n",
+      xlab = "Longitud (inches)",
+      ylab = "",
+      main = "Distribucion de las alturas de hombres
+      axes = FALSE)
> lines(x, y)
> bounds_filter <- x >= LI & x <= LS
```

```
> x_within_bounds <- x[bounds_filter]
> y_within_bounds <- y[bounds_filter]
> x_polygon <- c(LI, x_within_bounds, LS)
> y_polygon <- c(0, y_within_bounds, 0)
> polygon(x_polygon, y_polygon, col = "red")
> probability_within_bounds <- pnorm(LS, mu, sigma2) -
> text <- paste("p(", LI, "< height <", LS, ") =", sig
> mtext(text)
> sd_axis_bounds = 5
> axis_bounds <- seq(-sd_axis_bounds * sigma2 + mu, sd
> axis(side = 1, at = axis_bounds, pos = 0)
```

Distribucion de las alturas de hombres

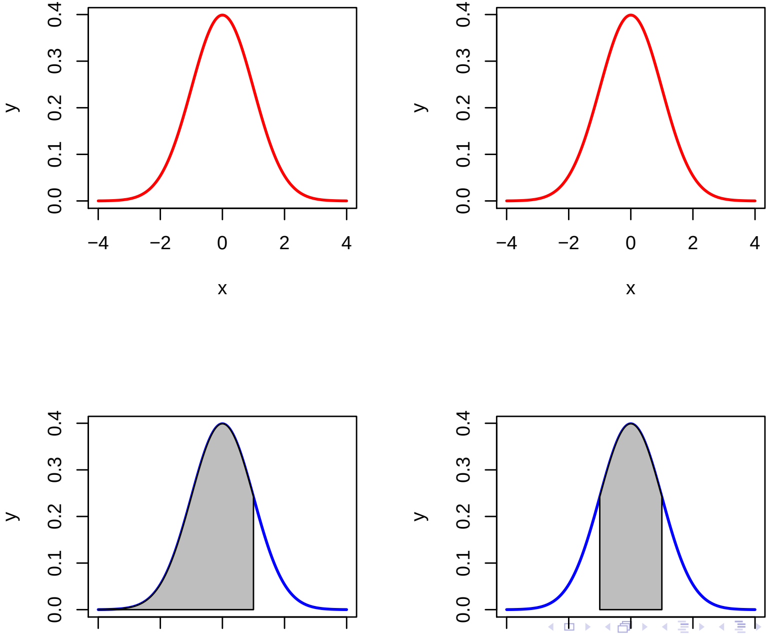
p(66.2 < height < 71.8) = 0.683



Otras instrucciones que realizan algo semejante a lo hasta ahora mostrado son las siguientes:

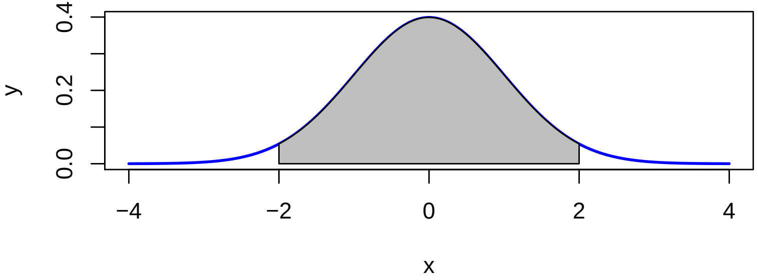
```
> # Parametros iniciales
> par(mfrow=c(2,2))
> a       <- -4;
> b       <- 4;
> N       <- 200;
> DE      <- 1;
> mu      <- 0;
> sigma2 <- 1;
> # Comencemos a graficar
> x=seq(a,b,length=N)
> y=1/sqrt(2*pi)*exp(-x^2/2)
> # Primer Grafica
> plot(x,y,type="l",lwd=2,col="red")
> # Segunda Grafica
> x=seq(a,b,length=N)
> y=dnorm(x,mean=mu,sd=sigma2)
```

```
> plot(x,y,type="l",lwd=2,col="red")
> # Tercer Grafica
> x=seq(a,b,length=N)
> y=dnorm(x)
> plot(x,y,type="l", lwd=2, col="blue")
> x=seq(a,DE,length=N)
> y=dnorm(x)
> polygon(c(a,x,DE),c(0,y,0),col="gray")
> #Cuarta grafica
> x=seq(a,b,length=N)
> y=dnorm(x)
> plot(x,y,type="l", lwd=2, col="blue")
> x=seq(-DE,DE,length=100)
> y=dnorm(x)
> polygon(c(-DE,x,DE),c(0,y,0),col="gray")
```

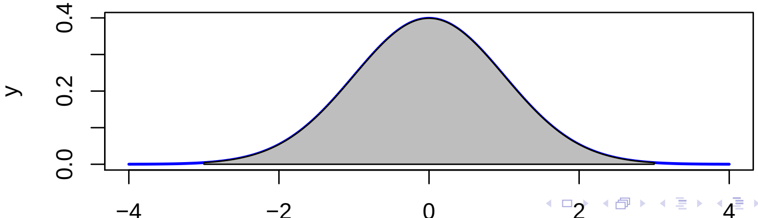

y
0.0
0.1
0.2
0.3
0.4
x
−4
−2
0
2
4

```
> # Primera grafica
> par(mfrow=c(2,1))
> x=seq(a,b,length=N)
> y=dnorm(x)
> plot(x,y,type="l", lwd=2, col="blue")
> x=seq(-2*DE,2*DE,length=N)
> y=dnorm(x)
> polygon(c(-2*DE,x,2*DE),c(0,y,0),col="gray")
> # Segunda grafica
> x=seq(a,b,length=N)
> y=dnorm(x)
> plot(x,y,type="l",lwd=2,col="blue")
> x=seq(-3*DE,3*DE,length=N)
> y=dnorm(x)
> polygon(c(-3*DE,x,3*DE),c(0,y,0),col="gray")
```


UACM Universidad Autónoma de la Ciudad de México Nada humano me es ajeno.

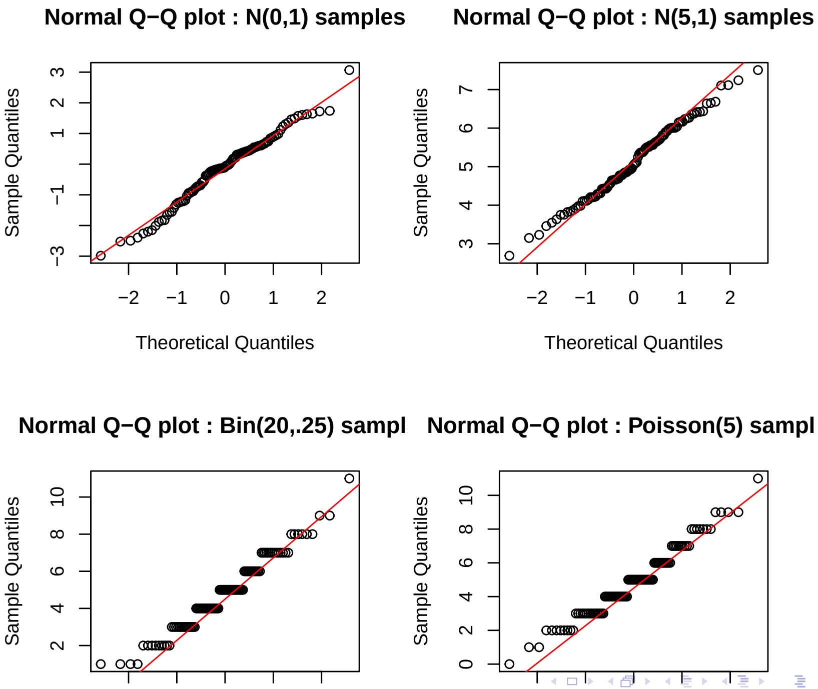


0.4
0.2
0.0
y
−4
−2
0
2
4
x

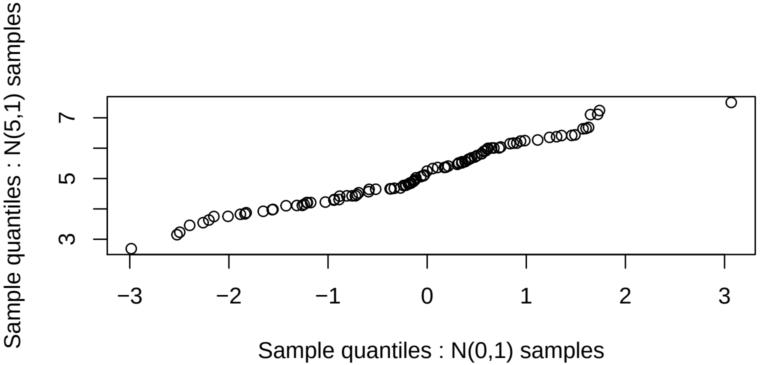


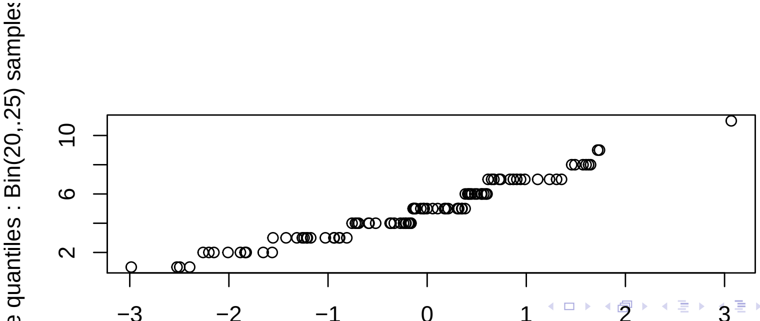
0.4
0.2
0.0
y
−4
−2
0
2
4

```
> stdnormsamp <-rnorm(100,mean=0,sd=1)
> normsamp <- rnorm(100,mean=5,sd=1)
> binomsamp <-rbinom(100,size=20,prob=.25)
> poissamp <- rpois(100,5)
> par(mfrow=c(2,2))
> qqnorm(stdnormsamp,main="Normal Q-Q plot : N(0,1) sa
> qqline(stdnormsamp,col=2)
> qqnorm(normsamp,main="Normal Q-Q plot : N(5,1) sampl
> qqline(normsamp,col=2)
> qqnorm(binomsamp,main="Normal Q-Q plot : Bin(20,.25)
> qqline(binomsamp,col=2)
> qqnorm(poissamp,main="Normal Q-Q plot : Poisson(5) s
> qqline(poissamp,col=2)
```


Normal Q−Q plot : N(0,1) samples
Normal Q−Q plot : N(5,1) samples
Normal Q−Q plot : Bin(20,.25) sampl
Normal Q−Q plot : Poisson(5) sampl
Sample Quantiles
Theoretical Quantiles
−3
−1
1
2
3
−2
−1
0
1
2
3
4
5
6
7
2
4
6
8
10
0

```
> par(mfrow=c(2,1))
> qqplot(stdnormsamp,
+        normsamp,
+        xlab = "Sample quantiles : N(0,1) samples",
+        ylab = "Sample quantiles : N(5,1) samples")
> qqplot(stdnormsamp,
+        binomsamp,
+        xlab = "Sample quantiles : N(0,1) samples",
+        ylab = "Sample quantiles : Bin(20,.25) sample
```


Sample quantiles : N(5,1) samples
3
5
7
−3
−2
−1
0
1
2
3
Sample quantiles : N(0,1) samples


e quantiles : Bin(20,.25) samples
2
6
10
−3
−2
−1
0
1
2
3

- Normal Distribution Functions, *R-bloggers*, https://www.r-bloggers.com/normal-distribution-functions/
- Plotting a Normal Distribution with R, *Matt Mazur Home Page*, https://mattmazur.com/2014/10/25/plotting-a-normal-distribution-with-r/
- Normal Distribution, *R Tutorial: An Introduction to Statistics*, http://www.r-tutor.com/elementary-statistics/probability-distributions/normal-distribution
- The Standard Normal Distribution in R, *Department of Mathematics, College of the Redwoods*,http://msenux2.redwoods.edu/MathDept/R/StandardNor
- Mihael Minn Home Page, *The Normal Distribution in R*, http://michaelminn.net/tutorials/r-normal-rank-order/

UACM Universidad Autónoma de la Ciudad de México Nada humano me es ajeno.