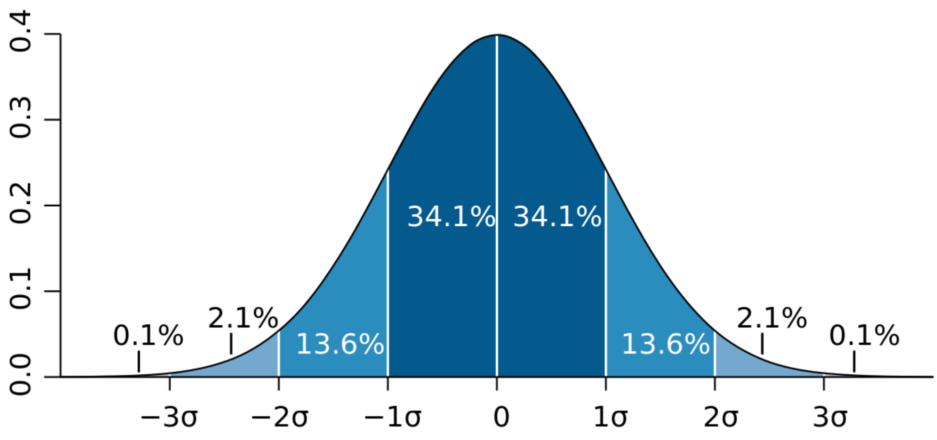
Instituto Tecnológico y de Estudios Superiores de Monterrey

Campus Monterrey

Inteligencia artificial avanzada para la ciencia de datos I
TC3006C.101

Algunas distribuciones de probabilidad



Rodolfo Sandoval Schipper A01720253
Marcelo Márquez A01720588
Arturo Garza Campuzano A00828096

11 ago 2023

1. **Graficar una distribución Normal con media $\mu = 10$, y desviación estándar $\sigma = 2$.**

Sugerencia. Adapte el código siguiente:

```
miu = 0
sigma = 1
x = seq(miu - 4*sigma, miu + 4*sigma, 0.01)
y = dnorm(x,miu, sigma)
plot(x,y, type = "l", col = "red", main = "Normal(0,1)")
```

Solución.-

Primero debemos generar números aleatorios para la distribución normal
set.seed(1)

Utilizando el código del ejemplo podemos declarar que.-
mu = 10
sigma = 2

Definimos la cantidad de muestras de los números aleatorios.-
tam_muestra = 1000

Utilizamos la función norm parecida al ejemplo para generar una muestra de los números aleatorios y así seguir la distribución normal. Utilizamos la media, y la desviación estándar cómo mean y sd.
valor_aleatorio = rnorm(tam_muestra, mean = mu, sd = sigma)

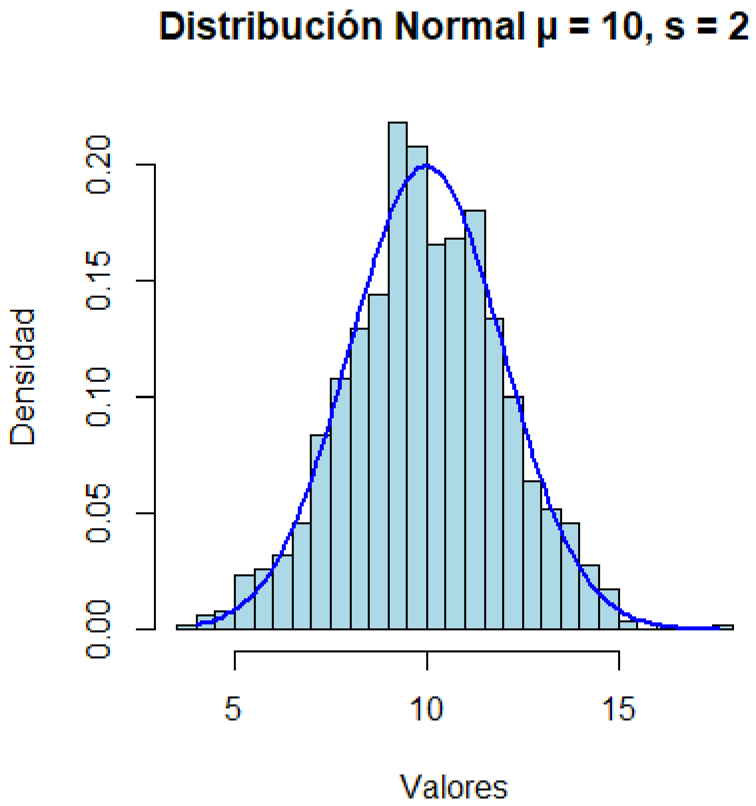
Creamos el histograma con 20 intervalos y sin mostrar las frecuencias

```
hist(valor_aleatorio, breaks = 20, freq = FALSE,
    col = "lightblue", border = "black",
    xlab = "Valores", ylab = "Densidad",
    main = "Solución con μ = 10, σ = 2")
```

Finalmente agregamos la curva de densidad teórica que también se encuentra en el ejemplo proporcionado.

```
x = seq(min(valor_aleatorio), max(valor_aleatorio), length.out = 100)
y = dnorm(x, mean = mu, sd = sigma)
lines(x, y, col = "blue", lwd = 2)
```

Gráfica →



Hacemos procesos similares para graficar las preguntas 2 - 4.

2. **Graficar una distribución T Student con grados de libertad** $\nu = 12$**.**

set.seed(1)

Grados de libertad
df = 12

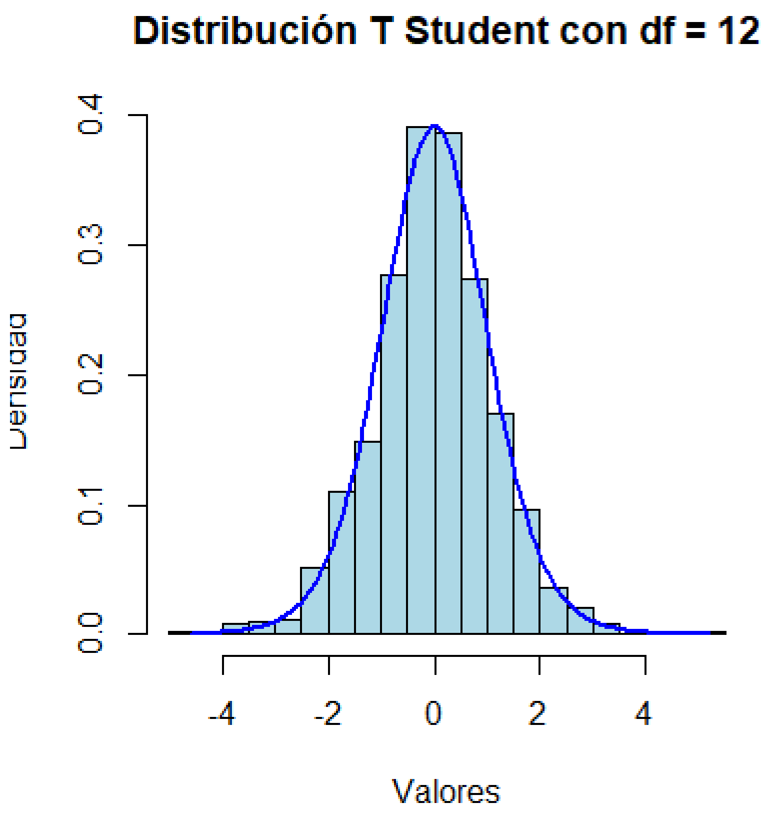
Cantidad de muestras
tam_muestra = 1000

Generar muestra de distribución T Student
valor_aleatorio = rt(tam_muestra, df)

Crear histograma y curva de densidad
hist(valor_aleatorio, breaks = 20, freq = FALSE,
col = "lightblue", border = "black",
xlab = "Valores", ylab = "Densidad",
main = "Distribución T Student con df = 12")

x = seq(min(valor_aleatorio), max(valor_aleatorio), length.out = 100)
y = dt(x, df)

lines(x, y, col = "blue", lwd = 2)



3. **Grafique la distribución Chi-cuadrada con 8 grados de libertad.**

set.seed(1)

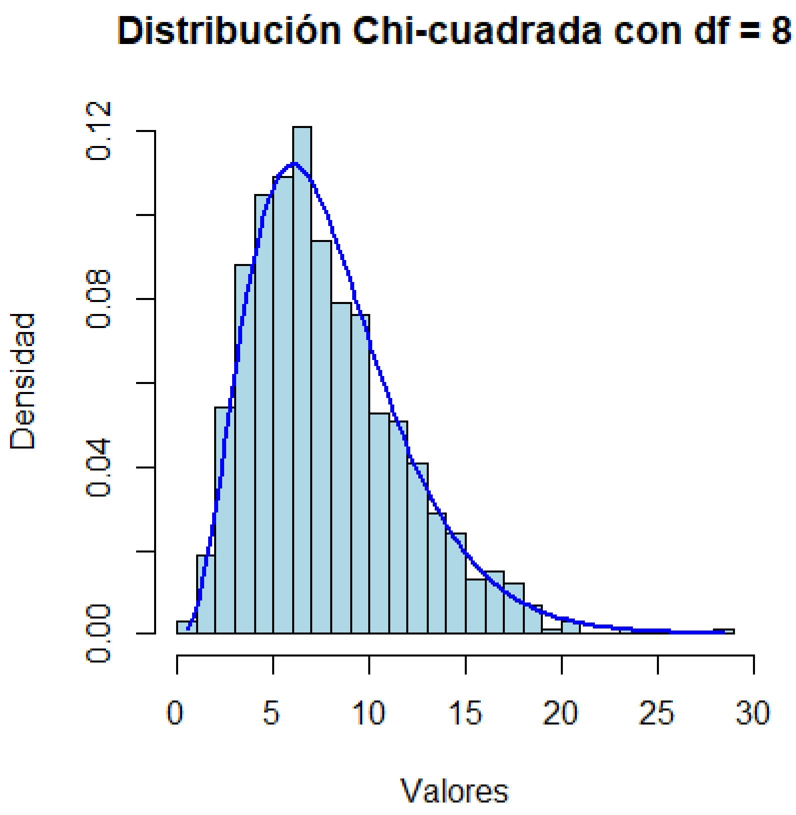
Grados de libertad
df = 8

Cantidad de muestras
tam_muestra = 1000

Generar muestra de distribución Chi-cuadrada
valor_aleatorio = rchisq(tam_muestra, df)

Crear histograma y curva de densidad
hist(valor_aleatorio, breaks = 20, freq = FALSE,
col = "lightblue", border = "black",
xlab = "Valores", ylab = "Densidad",
main = "Distribución Chi-cuadrada con df = 8")

x = seq(min(valor_aleatorio), max(valor_aleatorio), length.out = 100)
y = dchisq(x, df)
lines(x, y, col = "blue", lwd = 2)



4. **Graficar una distribución F con $v1 = 9$, $v2 = 13$.**

    set.seed(1)

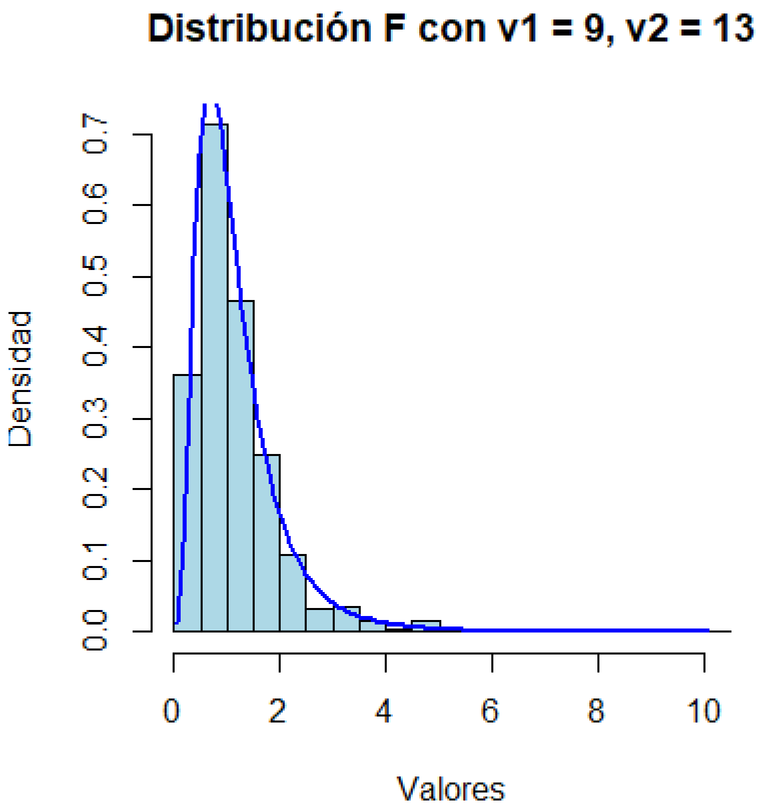
    Grados de libertad
    v1 = 9
    v2 = 13

    Cantidad de muestras
    tam_muestra = 1000

    Generar muestra de distribución F
    valor_aleatorio = rf(tam_muestra, v1, v2)

    Crear histograma y curva de densidad
    hist(valor_aleatorio, breaks = 20, freq = FALSE,
    col = "lightblue", border = "black",
    xlab = "Valores", ylab = "Densidad",
    main = "Distribución F con v1 = 9, v2 = 13")

x = seq(min(valor_aleatorio), max(valor_aleatorio), length.out = 100)
y = df(x, v1, v2)
lines(x, y, col = "blue", lwd = 2)



5. **Si Z es una variable aleatoria que se distribuye normalmente con media 0 y desviación estándar 1, hallar los procedimientos de:**
   a) $P(Z > 0.7) = 0.2419637$

   Para calcular $P(Z > 0.7)$ se necesita encontrar el área bajo la curva de la distribución normal estándar a la derecha de $Z = 0.7$**.** Para realizar el cálculo se puede ejecutar el siguiente comando en un script R: $p\_a <- 1 - pnorm(0.7)$. El resultado de este cálculo es 0.2419637**.**

   b) $P(Z < 0.7) = 0.7580363$

   Para calcular $P(Z < 0.7)$ se necesita encontrar el área bajo la curva de la distribución normal estándar a la izquierda de $Z = 0.7$**.** Para realizar el cálculo se puede ejecutar el siguiente comando en un script R: $p\_b <- pnorm(0.7)$. El resultados de este cálculo es 0.7580363.

   c) $P(Z = 0.7) = 0$

La probabilidad de que una variable aleatoria continua como Z tome un valor específico es en general cero. En este caso, $P(Z = 0.7) = 0$**.**

6. **Cuando lo que se quiere es hallar el valor de Z dada el área a la izquierda bajo la curva se usa qnorm(área izq.). Hallar el valor de Z que tiene al 45% de los demás valores inferiores a ese valor.**

   Área a la izquierda bajo la curva (probabilidad)
   area_izquierda = 0.45

   Encontrar el valor de Z correspondiente al área
   valor_z = qnorm(area_izquierda)

   El valor correspondiente al 45% de los valores inferiores es: $-0.125661346855074.$

7. **Hallar el procedimiento para verificar los siguientes resultados si se sabe que X se distribuye normalmente con una media de 100 y desviación estándar de 7.**
   Sugerencia. Utilice la función *pnorm(x, miu, sigma)* de R
   a) $P(X < 87) = 0.031645$
   Para calcular *P(X < 87),* utilizaremos la función \`pnorm\` para calcular probabilidades acumulativas con los valores dados de *media = 100* y *desv = 7.* Poniéndolo en R nos da: *prob1 = pnorm(x1, mean = media, sd = desv)* que iguala a 0.03164542.

   b) $P(X > 87) = 0.968354$
   Para calcular *P(X > 87),* utilizaremos la función \`pnorm\` para calcular probabilidades acumulativas con los valores dados de *media = 100* y *desv = 7.* Poniéndolo en R nos da: *prob2 = pnorm(x1, mean = media, sd = desv, lower.tail = FALSE)* que iguala a 0.9683546.

   c) $P(87 < X < 110) = 0.89179$
   Para calcular *P(87 < X < 110)* tendremos que utilizar la función \`pnorm\` dos veces con fin de calcular las probabilidades acumulativas y luego restamos para conseguir la probabilidad en el rango dado utilizando los valores de *media = 100* y *desv = 7* al igual que usando *x3_lower = 87* y *x3_upper = 110.*
   Poniéndolo en R nos da:
   prob3_lower = pnorm(x3_lower, mean = media, sd = desv)
   prob3_upper = pnorm(x3_upper, mean = media, sd = desv)
   prob3 = prob3_upper - prob3_lower
   Lo cual se iguala a 0.8917909.

```{r}
media = 100
desv = 7

# Calcular P(X < 87)
x1 = 87
prob1 = pnorm(x1, mean = media, sd = desv)
prob1

# Calcular P(X > 87)
x2 = 87
prob2 = pnorm(x2, mean = media, sd = desv, lower.tail = FALSE)
prob2

# Calcular P(87 < X < 110)
x3_lower = 87
x3_upper = 110
prob3_lower = pnorm(x3_lower, mean = media, sd = desv)
prob3_upper = pnorm(x3_upper, mean = media, sd = desv)
prob3 = prob3_upper - prob3_lower
prob3
```

```
[1] 0.03164542
[1] 0.9683546
[1] 0.8917909
```

8. **Hallar el procedimiento para verificar los siguientes resultados si se sabe que X se distribuye T Student con $gl = 10$, hallar:**
   Sugerencia. Utilice pt(x, gl) y qt(área izq, gl)
   a) $P(X < 0.5) = 0.6860532$
      Para calcular *P(X < 0.5)* utilizamos la función 'pt' con el valor dado de *x1 = 0.5* y el valor *gl = 10* (grados de libertad). Juntándose en R nos da: *prob1 = pt(x1, df = gl)* que iguala a 0.6860532.

   b) $P(X > 1.5) = 0.082253$
      Para calcular *P(X > 1.5)* utilizamos la función 'pt' con el valor dado de *x2 = 1.5* y el valor *gl = 10* (grados de libertad). Juntándose en R nos da: *prob2 = pt(x2, df = gl, lower.tail = FALSE)* que iguala a 0.08225366.

   La t que sólo el 5% son inferiores a ella. (t = -1.812461)

```{r}
gl = 10
# Calcular P(X < 0.5)
x1 = 0.5
prob1 = pt(x1, df = gl)
prob1

# Calcular P(X > 1.5)
x2 = 1.5
prob2 = pt(x2, df = gl, lower.tail = FALSE)
prob2

# La t que sólo el 5% son inferiores a ella.  (t = -1.812461)
p = 0.05
t = qt(p, df = gl)
t
```

```
[1] 0.6860532
[1] 0.08225366
[1] -1.812461
```

9. **Hallar el procedimiento para verificar los siguientes resultados si se sabe que X se distribuye Chi-cuadrada con $gl = 6$, hallar:**

   Sugerencia. Utilice pchisq(x, gl) y qchisq(área izq., gl)

   a) $P(X2 < 3) = 0.1911532$

      Para poder calcular *P(X2 < 3)* se utilizará la función 'pchisq' con los valores *gl = 6* (grados de libertad) y *x1 = 3* (que es el valor que se nos dio). Todo junto ponemos en R *prob1 = pchisq(x1, df = gl)* que nos regresa *0.1911532*.

   b) $P(X2 > 2) = 0.9196986$

      Para poder calcular *P(X2 > 2)* se utilizará la función 'pchisq' con los valores *gl = 6* (grados de libertad) y *x2 = 2* (que es el valor que se nos dio). Todo junto ponemos en R *prob2 = pchisq(x2, df = gl, lower.tail = FALSE)* que nos regresa *0.9196986*.

   El valor x de chi que sólo el 5% de los demás valores de x es mayor a ese valor.

```{r}
gl = 6
# Calcular P(X2 < 3)
x1 = 3
prob1 = pchisq(x1, df = gl)
prob1

# Calcular P(X2 > 2)
x2 = 2
prob2 = pchisq(x2, df = gl, lower.tail = FALSE)
prob2

# El valor x de chi que sólo el 5% de los demás valores de x es mayor a ese valor ( Resp. 12.59159)
p = 0.05
x = qchisq(1 - p, df = gl)
x
```

```
[1] 0.1911532
[1] 0.9196986
[1] 12.59159
```

10. Hallar el procedimiento para verificar los siguientes resultados si se sabe que X se distribuye F con $v1 = 8$, $v2 = 10$, **hallar:**

a) $P(X < 2) = 0.8492264$

Para resolver *P(X < 2)* utilizaremos la función 'pf' la cual devuelve la probabilidad acumulativa a la izquierda de un valor en una distribución F. Utilizamos los valores *v1 = 8* (grados de libertad del numerador), *v2 = 10* (grados de libertad del denominador) y el valor dado *x1 = 2* juntandolo en R nos da: *prob1 = pf(x1, df1 = v1, df2 = v2)* y nos regresa: 0.8492264.

b) $P(X > 3) = 0.05351256$

Para resolver *P(X < 3)* utilizaremos la función 'pf' la cual devuelve la probabilidad acumulativa a la izquierda de un valor en una distribución F. Utilizamos los valores *v1 = 8* (grados de libertad del numerador), *v2 = 10* (grados de libertad del denominador) y el valor dado *x1 = 3* juntandolo en R nos da: *prob2 = pf(x2, df1 = v1, df2 = v2, lower.tail = FALSE)* y nos regresa: 0.05351256.

El valor x que sólo el 25% de los demás valores es inferior a él.

```{r}
v1 = 8
v2 = 10

# Calcular P(X < 2)
x1 = 2
prob1 = pf(x1, df1 = v1, df2 = v2)
prob1

# Calcular P(X < 3)
x2 = 3
prob2 = pf(x2, df1 = v1, df2 = v2, lower.tail = FALSE)
prob2

# El valor x que sólo el 25% de los demás valores es inferior a él.
# 'qf' devuelve el valor crítico en la distribución F para un nivel de probabilidad dado
p = 0.25
x = qf(p, df1 = v1, df2 = v2)
x
```

```
[1] 0.8492264
[1] 0.05351256
[1] 0.6131229
```

11. **Una compañía de reparación de fotocopiadoras encuentra, revisando sus expedientes, que el tiempo invertido en realizar un servicio, se comporta como una variable normal con media de 65 minutos y desviación estándar de 20 minutos. Calcula la proporción de servicios que se hacen en menos de 60 minutos. Resultado en porcentaje con dos decimales, ejemplo 91.32%.**

    Considerando el tiempo invertido en realizar un servicio como una variable aleatoria normal X, se tienen los siguientes datos: $\mu = 65$ y $\sigma = 20$.

    Y, como se quiere calcular la proporción de servicios que se hacen en menos de 60 minutos esto se puede representar de la siguiente manera: $P(X < 60)$. Se ejecuta la siguiente línea de código de R para obtener el cálculo:

    $p <- pnorm(x, miu, sigma)$

    El resultado es 0.4013, lo cual se puede representar por medio del siguiente porcentaje: 40.13%.