Tarea 1 Análisis y Diseño de Experimentos

Alexandro Mayoral, Daniela Chávez y Gerardo Vazquez

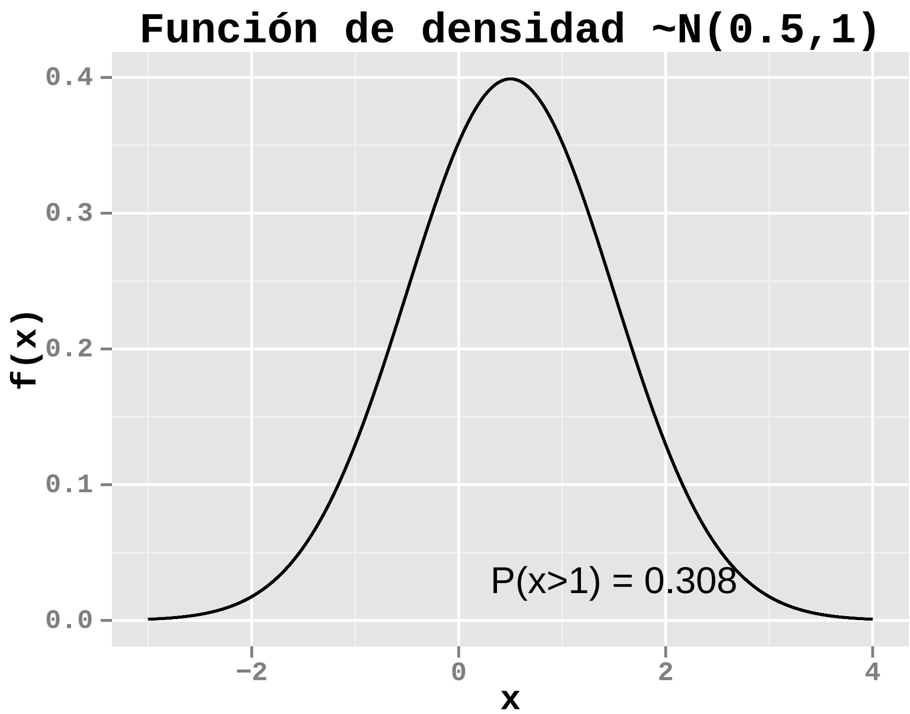
15/04/2015

Problema 1 Considere una variable aleatoria $X \sim N(0.5, 1)$ determine la probabilidad de que $X > 1$. Elabore la gráfica correspondiente.

```
# Obtenemos la probailidad de que X > 1
pnorm(1, mean = 0.5, sd = sqrt(1), lower.tail = F)

## [1] 0.3085
```

A continuación observamos la gráfica:



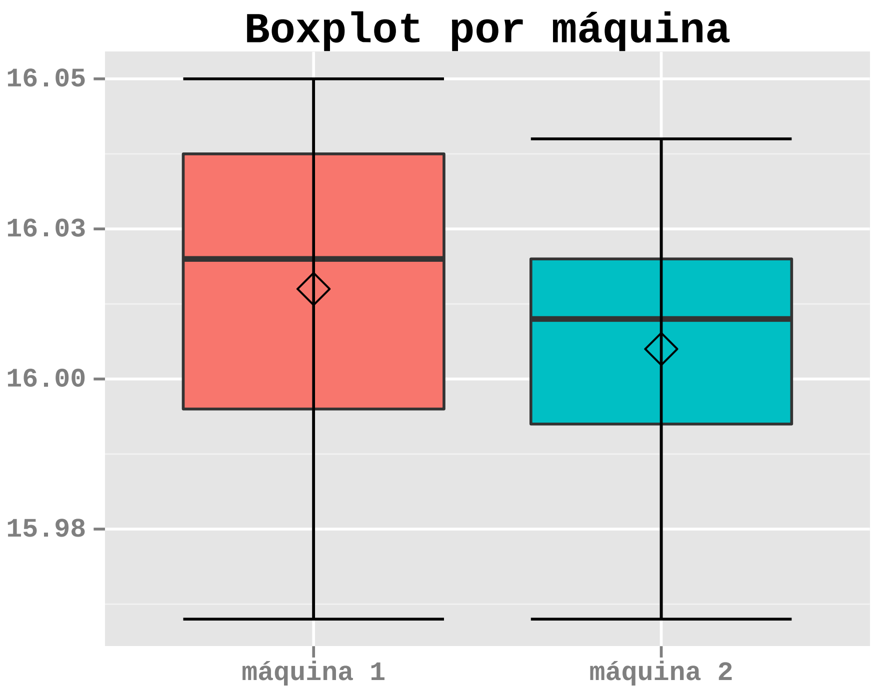
Problema 2 Se utilizan dos máquinas para llenar botellas de plástico con un volumen neto de 16 onzas. El proceso de llenado se puede suponer normal. Los ingenieros del departamento de calidad sospechan que las dos máquinas llenan a diferente volumen, por lo que se realizó un experimento tomando una muestra al azar de 10 botellas llenadas por cada una de las dos máquinas.

máquina 1		**máquina 2**	
16.03	16.01	16.02	16.03
16.04	15.96	15.97	16.04
16.05	15.98	15.96	16.02
16.05	16.02	16.01	16.01
16.02	15.99	15.99	16.00

a) Pruebe la hipótesis de que el promedio de llenado de las dos máquinas es el mismo vs. es diferente. Use $\alpha = 0.05$. Primero suponiendo que las varianzas son homogéneas y después suponiendo que no son homogéneas.

b) Calcule un intervalo del 95% de confianza para la diferencia de las medias de volumen de llenado de las máquinas.

Veamos un boxplot de los datos por máquina:




```r
# Pruebe la hipótesis de que el promedio de llenado de las dos máquinas
# es el mismo vs. es diferente
# Bajo el supuesto de varianzas iguales:
# Método largo
# Manipulamos los datos para los cálculos
datos <- data.frame(maquina1=subset(datos, trat == "máquina 1")[, "vol"],
                    maquina2=subset(datos, trat == "máquina 2")[, "vol"])
```

```r
# Cálculo de las medias
colMeans(datos)

## maquina1 maquina2
##    16.02    16.00

# Cálculo de la diferencia de medias
dif <- as.numeric(colMeans(datos)[1] - colMeans(datos)[2])
S2_1 <- var(datos[, 1])
S2_2 <- var(datos[, 2])
S2_p <- ((n1 - 1) * S2_1 + (n2 - 1) * S2_2) / (n1 + n2 - 2)

t_0 <- dif/(sqrt(S2_p*(1/n1 + 1/n2))); t_0

## [1] 0.7989

# Región de rechazo
alpha <- 0.05
gl <- n1 + n2 - 2
t <- qt(1 - (alpha/2), gl)
# P-value de 2 colas
pt(t_0, gl, lower.tail = FALSE) * 2

## [1] 0.4347

# Método corto
t.test(datos[, 1], datos[, 2])

##
##  Welch Two Sample t-test
##
## data:  datos[, 1] and datos[, 2]
## t = 0.7989, df = 17.49, p-value = 0.435
## alternative hypothesis: true difference in means is not equal to 0
## 95 percent confidence interval:
##  -0.01635  0.03635
## sample estimates:
## mean of x mean of y
##     16.02     16.00
```

Por lo que bajo la hipótesis $H_0 : \mu_1 = \mu_2 (t = 0.7989, p > 0.05)$ y asumiendo varianzas iguales, no existe evidencia significativa para rechazar que el promedio del llenado de botellas de la máquina 1 y 2 son iguales. Ahora veamos el caso cuando no consideramos que las varianzas son homogéneas.

```r
# Bajo el supuesto de varianzas diferentes:
# Método largo
# t welch
tw <- dif / (sqrt(S2_1/n1 + S2_2/n2)); tw
```

```
## [1] 0.7989
```

```r
#región de rechazo
gl_tw <- (S2_1/n1 + S2_2/n2)**2 / (((S2_1/n1)**2/(n1 - 1)) + ((S2_2/n2)**2/(n2-1)))
t2 <- qt(1-(alpha/2), gl_tw); t2
```

```
## [1] 2.105
```

```r
# P-value
pt(tw, gl_tw, lower.tail = FALSE) * 2
```

```
## [1] 0.435
```

```r
# Método Corto
t.test(datos[, 1], datos[, 2], var.equal=F)
```

```
##
##  Welch Two Sample t-test
##
## data:  datos[, 1] and datos[, 2]
## t = 0.7989, df = 17.49, p-value = 0.435
## alternative hypothesis: true difference in means is not equal to 0
## 95 percent confidence interval:
##  -0.01635  0.03635
## sample estimates:
## mean of x mean of y
##     16.02     16.00
```

Por lo qué observamos que tampoco existe evidencia para rechazar la hipótesis nula ($H_0 : \mu_1 = \mu_2 (t = 0.7989, p > 0.05)$), asumiendo que se tienen varianzas distintas. A continuación mostramos la prueba F para evaluar la exitencia de homocedasticidad

```r
var.test(datos[, 1], datos[, 2])
```

```
##
##  F test to compare two variances
##
## data:  datos[, 1] and datos[, 2]
## F = 1.41, num df = 9, denom df = 9, p-value = 0.6168
## alternative hypothesis: true ratio of variances is not equal to 1
## 95 percent confidence interval:
##  0.3503 5.6777
```

```
## sample estimates:
## ratio of variances
##               1.41
```

Por lo que no hay evidencia para rechazar la hipótesis de homocedasticidad a un nivel de significancia del 0.05.

Ahora calculamos los intervalos del 95% de confianza para la diferencia de medias:

```
# Bajo el supuesto de homocedasticidad
ic_vi <- cbind(dif - t * sqrt(S2_p*(1/n1 + 1/n2)),
               dif + t * sqrt(S2_p*(1/n1 + 1/n2)))
colnames(ic_vi) <- cbind("límite inferior","límite superior");ic_vi

##      límite inferior límite superior
## [1,]         -0.0163          0.0363

# Amplitud del intervalo
diff(as.vector(ic_vi))

## [1] 0.05259

# Bajo el supuesto de no homocedasticidad
ic_vd <- cbind(dif - t2 * sqrt(S2_1/n1 + S2_2/n2),
               dif + t2 * sqrt(S2_1/n1 + S2_2/n2))
colnames(ic_vd) <- cbind("límite inferior","límite superior");ic_vd

##      límite inferior límite superior
## [1,]        -0.01635         0.03635

# Amplitud del intervalo
diff(as.vector(ic_vd))

## [1] 0.0527
```

Por lo anterior vemos que el intervalo bajo el supuesto de homocedasticidad es más pequeño, además ambos intervalos contienen al cero, por lo que tenemos una confianza del 95% de que no existe diferencia en el promedio de llenado de botellas en ambas máquinas. Es importante también mencionar que uno de los puestos que se usó al realizar los contrastes de hipótesis fue el de normalidad de los datos, por lo qué se debería aplicar una prueba para evaluar este supuesto.

Problema 3 Considere un diseño de experimento unifactorial (efectos fijos) de 4 tratamientos donde se tienen un total de 24 unidades experimentales y conteste lo siguiente:

- Complete la siguiente tabla ANOVA:

FV	g.l.	S.C.	C.M.	F
Modelo				0.5655
Error			16.02	
Total				

La tabla la podemos completar de la siguiente forma:

FV	g.l.	S.C.	C.M.	F
Modelo	t - 1 = 4 - 1 = 3	SSt = CMt * (t - 1) = 27.178	CMt = F * CME = 9.059	F = CMt / CME = 0.5655
Error	n - t = 96 - 4 = 92	SSE = CME * (n - t) = 16.02 * 92 = 1473.84	CME = 16.02	
Total	n - 1 = 96 -1 = 95	SStotal = SSt + SSE		

- Determine el valor $\hat{\sigma^2}$

De la tabla sabemos que $CME = \hat{\sigma^2} = 16.02$

- Con $\alpha = 0.05$ se rechaza o no la hipótesis de igualdad de medias?

```
alpha <- 0.05
qf(1 - alpha, df1= 3, df2 = 92)

## [1] 2.704
```

Entonces sabemo que valores grandes de F llevan a rechazar la hipótesis de nula de igualdad de medias, pero como $F = 0.5655 < F^{0.05}_{t-1,n-t} = 2.704$) no tenemos evidencia para rechzar la hipótesis nula a una significancia del 5%.

- Con $\alpha = 0.01$ se rechaza o no la hipótesis de igualdad de medias?

```
alpha <- 0.01
qf(1 - alpha, df1= 3, df2 = 92)

## [1] 4.002
```

En este caso $F = 0.5655 < F^{0.05}_{t-1,n-t} = 4.002$) por lo que tampoco tenemos evidencia para rechazar la hipótesis nula a una significancia del 5%.