Tarea 3

Arina Bilan, Víctor Duarte, Benjamín Zúñiga

2024-10-07

Enunciado:

En una planta química hay dos máquinas que envasan detergentes industriales en bidones con un volumen de producto que sigue una distribución normal con desviación estándar de 1 litro. La ingeniera a cargo de la planta debe asegurar que los bidones se están llenando con una media de 10 litros. Pero ella tiene la sospecha de que hay desviaciones en esta media, lo que piensa confirmar usando una muestra aleatoria de 100 envases (50 de cada una de las máquinas). También cree que hay diferencia en el cumplimiento del volumen requerido entre la máquina más antigua y la más moderna, que han de andar por el 90% y 96% de los bidones, respectivamente.

```r
library(ggpubr)
```

```
## Warning: package 'ggpubr' was built under R version 4.4.1
```

```
## Cargando paquete requerido: ggplot2
```

```r
library(pwr)
```

```
## Warning: package 'pwr' was built under R version 4.4.1
```

```r
library(tidyr)
```

Pregunta 1:

Si la ingeniera está segura de que el verdadero volumen medio no puede ser superior a 10 litros y piensa rechazar la hipótesis nula cuando la muestra presente una media menor a 9,8 litros, ¿cuál es la probabilidad de que cometa un error de tipo I? Para responder, generen un gráfico de la distribución muestral de las medias hipotetizada en donde se marque la zona correspondiente a la probabilidad solicitada, para luego, basándose en este gráfico, calcular el área correspondiente.

Respuesta:

Del siguiente enunciado se puede notar varias variables que se conoce: Desviación estandár, Tamaño de muestra, Media.

Se puede notar del enunciado que se trata de una prueba de hipótesis unilateral, por lo que se pueden generar las siguientes hipótesis:

En forma declarativa:

H_0: el volumen medio es igual a 10 litros

H_1: el volumen medio es menor a 10 litros

Matemáticamente:

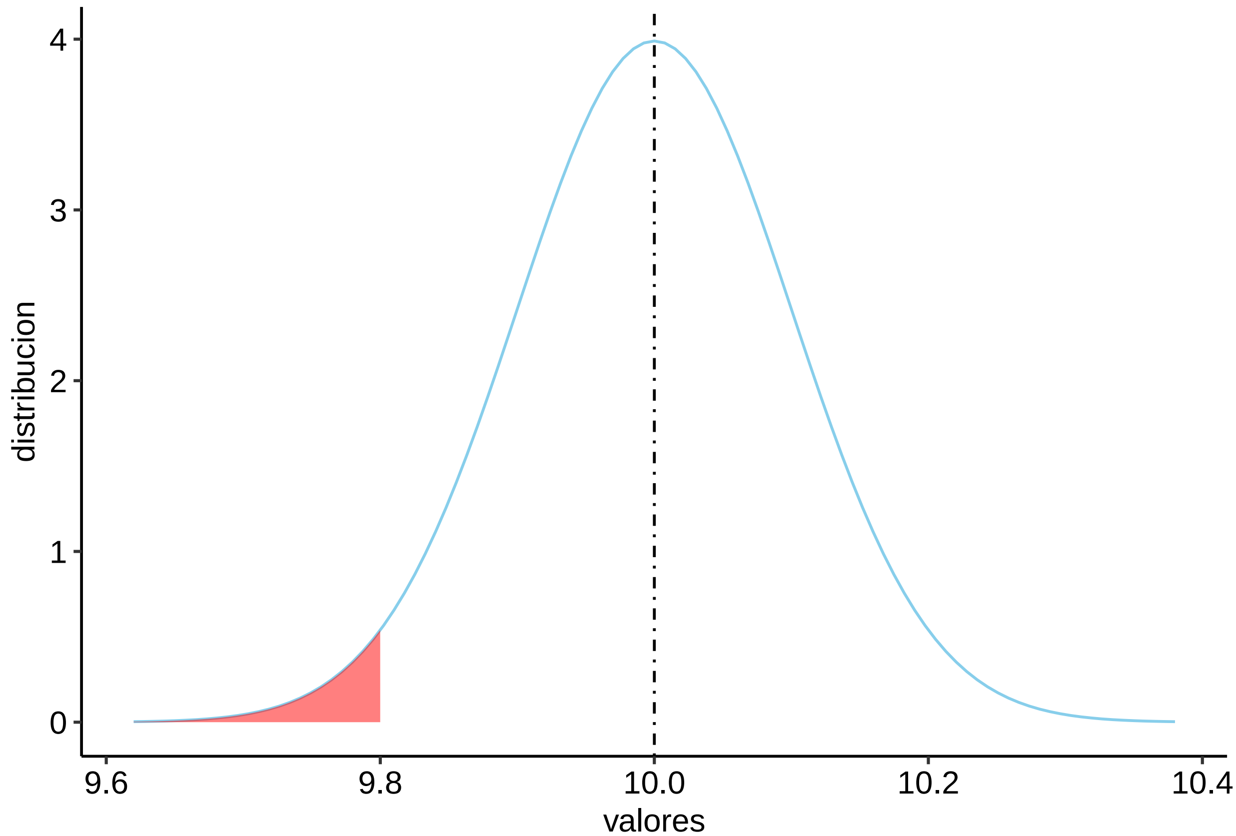
$H_0 : \mu = 10$

$H_1 : \mu < 10$

```
desviacion_estandar = 1
tamano_muestra = 100
valor_nulo = 10
```

Graficando:

```
#como se conoce desviacion poblacional entonces se puede calcular error estandar
error_estandar = desviacion_estandar/sqrt(tamano_muestra)

#se ocupa distribucion normal porque se conoce desviacion poblacional

valores = seq(valor_nulo - 3.8 * error_estandar, valor_nulo + 3.8 * error_estandar, length.out = 7000)

distribucion = dnorm(valores, mean = valor_nulo, sd = error_estandar)
distribucion_normal = data.frame(valores, distribucion)

grafico = ggplot(data = distribucion_normal,
                 aes(valores)) + theme_pubr() + stat_function(fun = dnorm, args = list(mean = valor_nul
  geom_area(data = subset(distribucion_normal, valores < 9.8),aes(y = distribucion),fill = "red", alpha

grafico
```




```r
probabilidad = pnorm(9.8, mean = valor_nulo, sd = error_estandar, lower.tail = TRUE)

probabilidad
```

```
## [1] 0.02275013
```

Por lo tanto la probabilidad de que cometa un error de tipo I, es decir, rechazar la hipótesis nula cuando esta es verdadera, es de $\alpha = 0,02275013$, la cual está representada en rojo en el gráfico y fue calculada a través de la función pnorm.

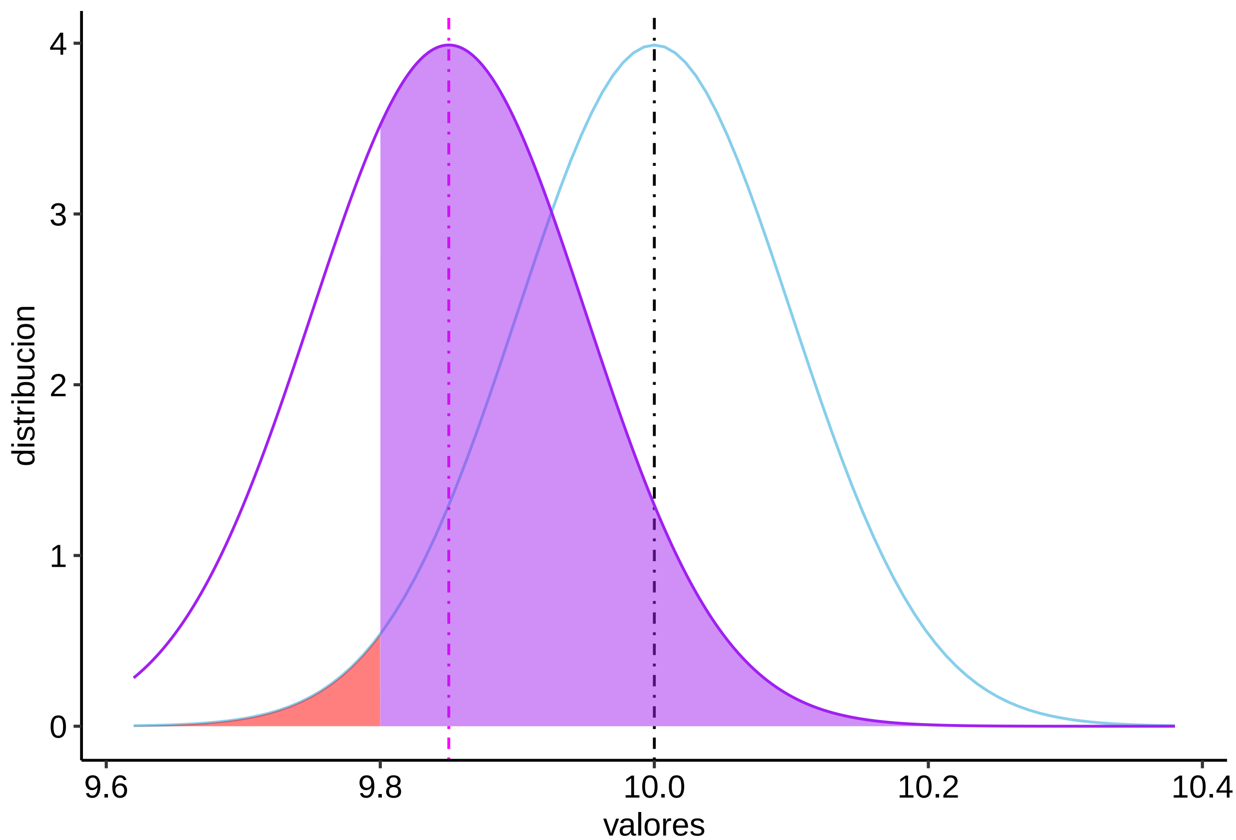
Pregunta 2

Si el verdadero volumen medio de los bidones fuera de 9,85 litros, ¿cuál sería la probabilidad de que la ingeniera, que obviamente no conoce este dato, cometa un error de tipo II? Para responder, agregue al gráfico anterior la verdadera distribución muestral de las medias y marquen (con otro color) la zona correspondiente a la probabilidad solicitada, para luego, basándose en este gráfico, calcular el área correspondiente.

Respuesta:

El error de tipo II sería la probabilidad de no rechazar H_0 que dice que $\mu = 10$, cuando en realidad es 9.85, para esto simulamos distribución normal con la media de 9.85 y añadimos en el gráfico anterior.

```
media = 9.85
valores_x <- seq(media - 3.8 * error_estandar, media + 3.8 * error_estandar, length.out = 7000)

valores_y = dnorm(valores_x, mean = media, sd = error_estandar)

distribucion_n = data.frame(x = valores_x, y = valores_y)

area = subset(distribucion_n, x > 9.8)

grafico_2 = grafico +
  stat_function(fun = dnorm, n = 7000, args = list(mean = media, sd = error_estandar),colour ="purple")
  geom_vline(xintercept = media, colour = "magenta", linetype = "dotdash") +
  geom_area(fill = "purple", alpha = 0.5,
            data = area, aes(x = x, y = y))




grafico_2
```



Para calcular la probabilidad de error de tipo II, se debe considerar que el poder estadístico está representado por el área bajo la curva de la distribución real (en color púrpura) desde el límite de la cola inferior hasta la región de rechazo, que está determinada por el nivel de significancia, α, asociado al error de tipo I. El error de tipo II corresponde al área en el extremo opuesto de la distribución (la cola superior, en este caso) del mismo gráfico púrpura. Para calcular esta probabilidad, se utiliza la función pnorm de la siguiente manera:

```r
probabilidad = pnorm(9.8, mean = media, sd = error_estandar, lower.tail = FALSE)
probabilidad
```

```
## [1] 0.6914625
```

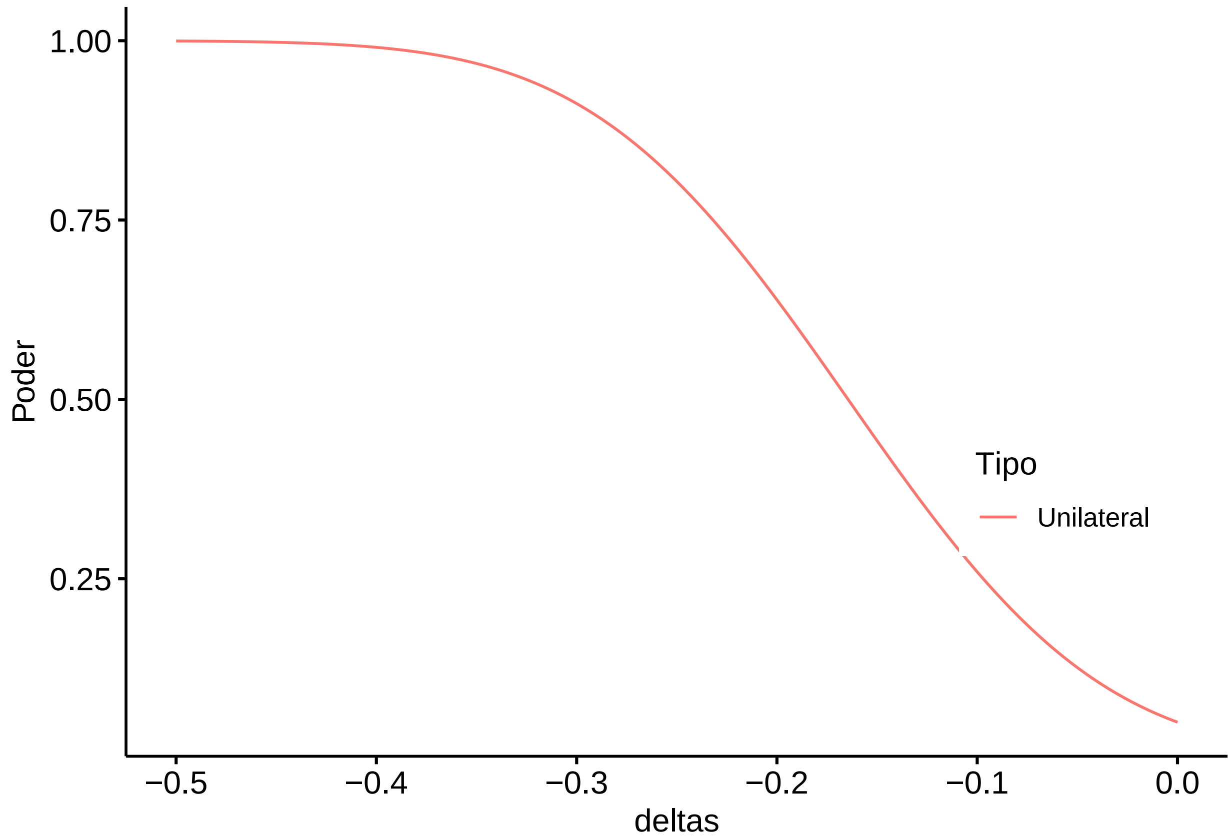
La probabilidad de que la ingeniera cometa un error de tipo II, es decir, no rechazar la hipótesis nula cuando esta es falsa, es de $\beta = 0,6914625$, la cual está representada en el gráfico con el área púrpura.

Pregunta 3:

Como no se conoce el verdadero volumen medio, genere un gráfico del poder estadístico con las condiciones anteriores, pero suponiendo que el verdadero volumen medio podría variar de 9,5 a 10 litros. Hay un ejemplo de este tipo de gráfico en la lectura sobre poder estadístico.

Respuesta:

```r
alfa<-0.05
n<-100
media_nula<-10
sigma<-1

medias_verdaderas<-seq(9.5, 10, 0.001)
deltas<- medias_verdaderas-media_nula
deltas_norm<- deltas / sigma

f_u<- function(x) pwr.norm.test(x,n=n, sig.level = alfa, alternative = "less")[["power"]]

poder_u <- sapply(deltas_norm, f_u)

datos_a<- data.frame(deltas,poder_u)
datos_l<- datos_a %>% pivot_longer(-deltas, names_to = "Tipo", values_to = "Poder")
datos_l[["Tipo"]]<- factor(datos_l[["Tipo"]],labels = c("Unilateral"))

g <- ggline(datos_l, x = "deltas", y = "Poder",
            color = "Tipo", numeric.x.axis = TRUE, plot_type = "l")

g <- ggpar(g,legend = c(.85, .35))
print(g)
```



Pregunta 4

Considerando un volumen medio de 10 litros, ¿Cuántos bidones deberían revisarse para conseguir un poder estadístico de 0,75 y un nivel de significación de 0,05?

Respuesta:

Las hipótesis son las siguientes

H_0 : la media poblacional es de 10 litros.

H_1 : la media poblacional es menor a 10 litros.

Planteadas de manera matemática:

$H_0 : \mu = 10$

$H_1 : \mu < 10$

```
#d = (mediaReal - valorNulo) / desviación estándar
d =(9.85-10)
n = pwr.norm.test(d = d,n = NULL, sig.level = 0.05, power = 0.75,alternative = "less")$n
n = ceiling(n)
n
```

```
## [1] 240
```

Deberían revisarse al menos 240 bidones para conseguir un poder estadístico de 0,75 y una significancia de 0,05.

Pregunta 5

¿Alcanzaría esta muestra para detectar la diferencia que la ingeniera sospecha que existe entre las dos máquinas de la planta con al menos las mismas probabilidades de cometer errores?

Respuesta:

Las hipótesis planteadas son las siguientes

H_0 : no existe diferencia entre el cumplimiento del volumen requerido entre la máquina más antigua y la más moderna.

H_1 : existe diferencia entre el cumplimiento del volumen requerido entre la máquina más antigua y la más moderna.

Matemáticamente:

$H_0 : p_1 = p_2$

$H_1 : p_1 \neq p_2$

en donde p_1 representa la proporción de error de la máquina moderna y p_2 la proporción de error de la máquina más antigua.

Si tenemos las mismas probabilidades para cometer errores, entonces α y β deben ser los mismos que la pregunta anterior (como $poder = 1 - \beta$ el poder también debe ser el mismo).

```
h = ES.h(0.04, 0.1)
n = pwr.2p.test(h=h,n = NULL, sig.level = 0.05, power = 0.75,alternative = "two.sided")$n
ceiling(n)
```

```
## [1] 240
```

Por lo tanto se necesita una muestra de 240 envases para cada máquna para tener un $\alpha = 0,05$ y un $\beta = 0,25$ o que es equivalente $poder = 0,75$ entonces no alcanza con una muestra de 50 envases para cada máquina.