Pruebas de hipótesis de una muestra

Sadan De la Cruz Almanza

(Departamento de Economía - Universidad de Pamplona)

Contents

1 ¿Qué es una hipótesis estadística?

La palabra hipótesis puede ser utilizada en diferentes contextos. En general, se puede definir como "un enunciado no verificado, que se intenta confirmar o refutar" (wiki). Desde una perspectiva estadística, la hipotesis tiene una fuerte relación con el concepto de **estimación**. Una hipótesis estadística "es una afirmación con respecto a alguna característica desconocida deuna población de interés" (Canavos, 1988).

En esencia al probar una **hipótesis estadística**, buscamos decidir si la afirmación propuesta se encuentra se encuentra apoyadq por la evidencia experimental que se obtiene a través de una muestra aleatoria. Dicha decisión se fundamenta con base en probabilidades, en otras palabras si contamos con **"suficiente evidencia estadística"** aceptaremos o rechazaremos la hipótesis.

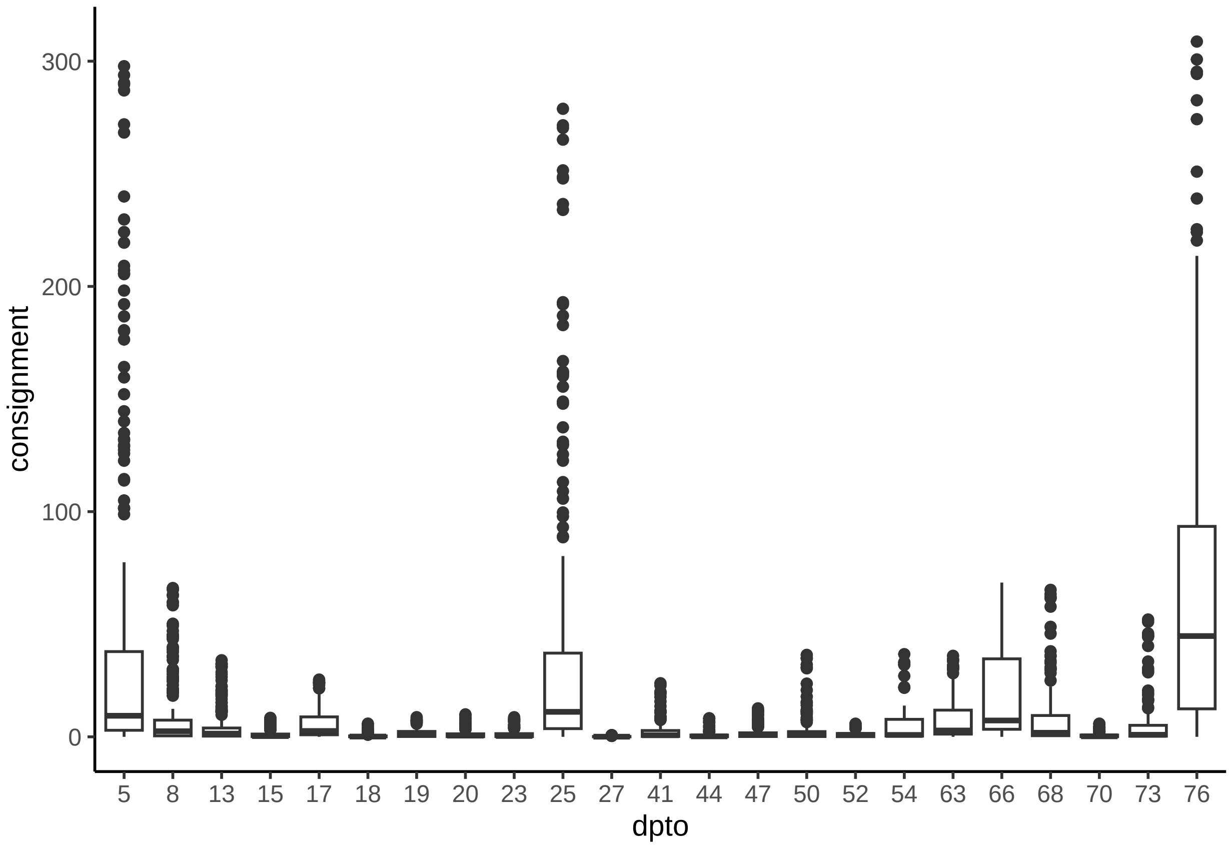
En terminos generales, vamos a identificar la **hipótesis nula (H_0)** y la **hipótesis alternativa (H_1)**. El primer caso se relaciona con la definición inicial, y en el segundo, refleja "el valor posible o intervalo de valores del parámetro de interés si la hipótesis nula es falsa" (Canavos, 1988, p. 308). Por otro lado, **prueba estadística** respecto a "una característica desconocida de la población de interés es cualquier regla para decidir si se rechaza la hipótesis nula con base en una muestra aleatoria de la población" (Canavos, 1988, p. 306).

```r
# paquetes

library(ggplot2)
library(haven)
library(dplyr)
```

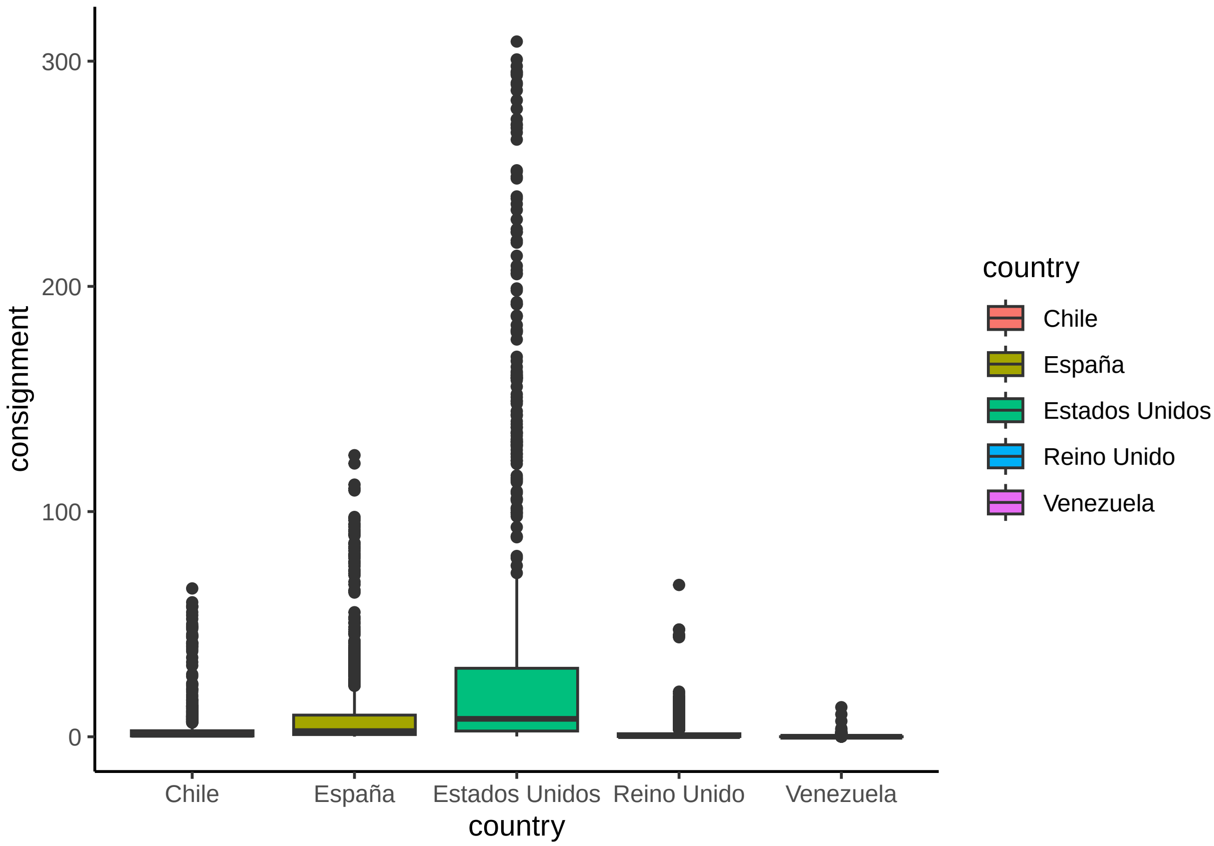
```r
# Remesas, Colombia (2014 - 2023)

# Cargar base de datos en un formato .dta

bd <- read_dta("/Users/macbookait/Documents/Investigación/Investigación/Remesas/datos/bd_remesas.dta")
```

```r
# DIAGRAMA DE CAJA Y BIGOTES

#El diagrama muestra la distribución, la mediana y la variabilidad de los datos.

bd$country <- factor(bd$country)
bd$dpto <- factor(bd$dpto)
```

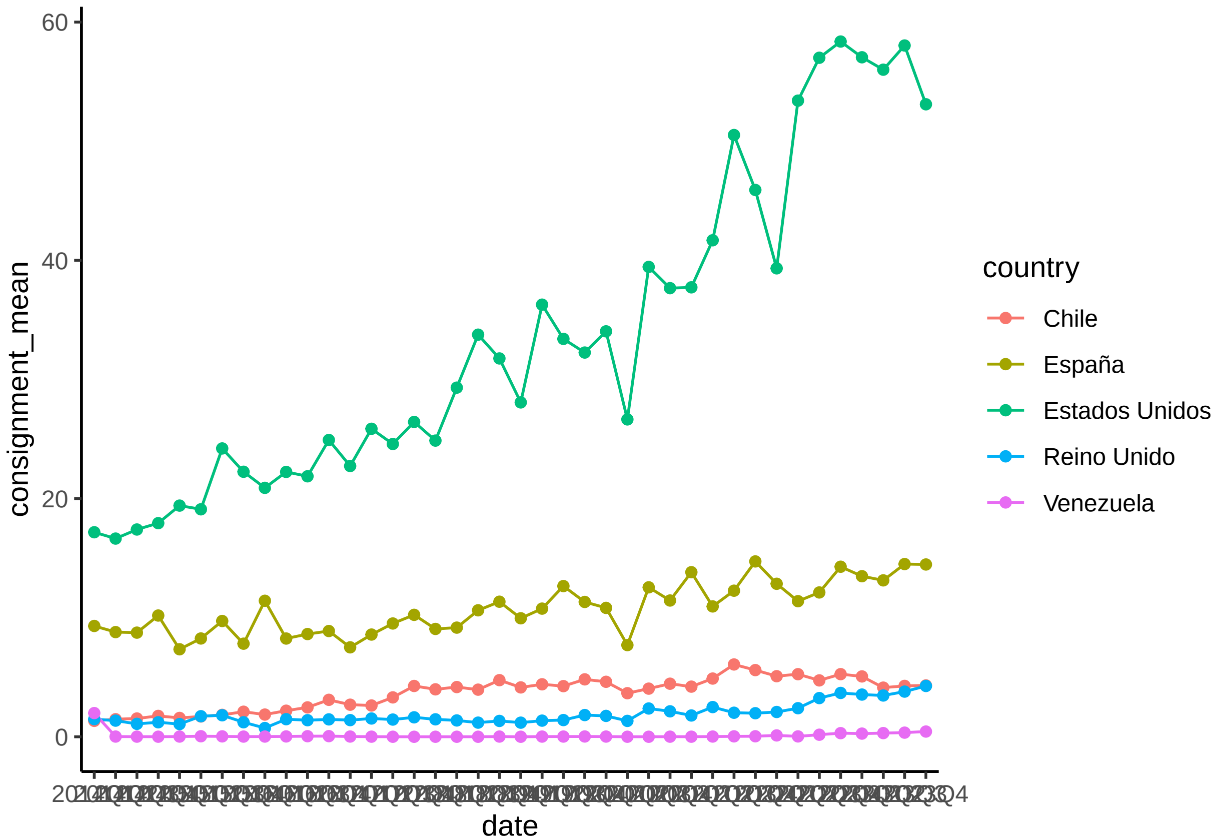
```r
# Departamentos
ggplot(bd, aes(x = dpto, y = consignment)) + geom_boxplot() + theme_classic()
```



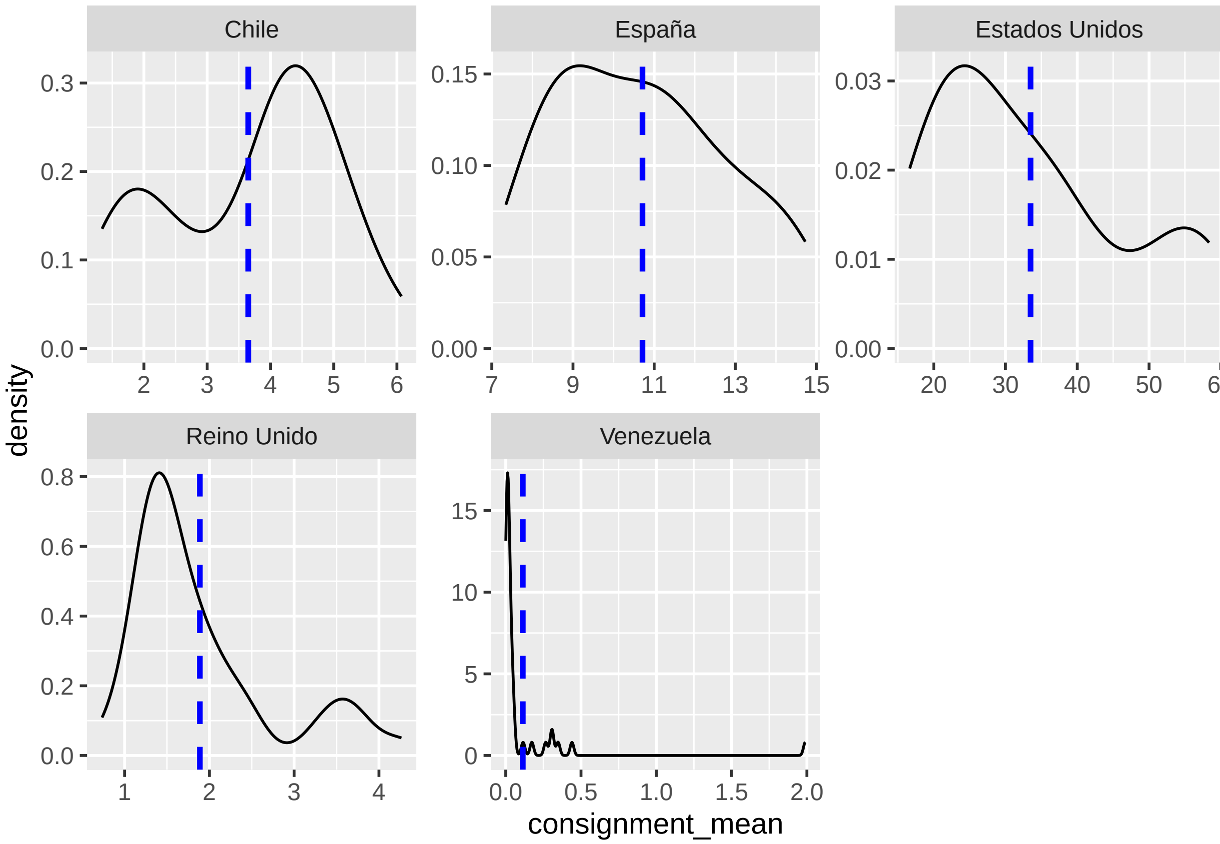

```r
# Países

ggplot(bd, aes(x = country, y = consignment, fill=country)) +
  geom_boxplot() + theme_classic()
```




```r
# SERIE DE TIEMPO

bd <- bd %>%
  mutate(date = paste(year, "Q", quarter, sep = ""))

bd1 <- bd %>%
  group_by(country, date) %>%
  summarize(consignment_mean = mean(consignment), .groups = 'drop')

bd1 <- bd1 %>%
  mutate(date = factor(date, levels = unique(date)))

ggplot(bd1, aes(x = date, y = consignment_mean, color = country, group = country)) +
  geom_line() + geom_point() + theme_classic()
```




```r
# GRÁFICA DE DENSIDAD POR GRUPOS

means_by_country <- bd1 %>%
  group_by(country) %>%
  summarize(mean_consignment = mean(consignment_mean))

ggplot(bd1, aes(x = consignment_mean)) + geom_density() +
  geom_vline(data = means_by_country, aes(xintercept = mean_consignment),
             color="blue", linetype="dashed", size = 1) +
  facet_wrap(~ country, scales = "free")
```



1.1 Errores de inferencia

En estadística inferencial saber "aceptar" o "rechazar" la hipótesis propuesta es importante respecto al "que decir" de los datos.

1. **Error tipo I:** El rechazo de H_0 cuando en realidad H_0 es cierto. Esto es posible cuando la decisión es rechazar la hipótesis nula.
2. **Error tipo II:** Equivocarse al rechazar H_0 cuando en realidad H_0 es falsa. Esto es posible cuando la decisión es no rechazar la hipótesis nula.

Como se menciono en clases anteriores, la idea de la estadística es encontrar el estado "natural" de nuestros fenomenos económicos (población) a partir de una muestra. En este caso, la siguiente tabla proporciona un resumen respecto a la decisión tomada:

	Población ("Verdad")	
Muestra	H_0 es verdadero	H_0 es falsa
No rechazarH_0	**Decisión correcta** (Pr = 1 - α)	Error tipo II (Pr = α)
Rechazar H_0	Error tipo I (Pr = β)	**Decisión correcta** (Pr = 1 - β)

En todo caso, $0 \leq \alpha \leq 1$ y $0 \leq \beta \leq 1$. α, β permiten medir la posibilidad de cometer alguno de estos errores. ()

Nota: De acuerdo con Canavos, "la decisión de rechazar H_0 no necesariamente significa que H_0 sea falsa", no obstante, la muestra utilizada para tomar la decisión proporciona un grado de confiabilidad con el que puede procederse como si H_0 fuese falsa.

1.2 Región crítica y reglas de decisión

Para ciertos valores de la estadística de prueba, la decisión será el rechazar la hipótesis nula, dichos valores se conocen como región crítica. Las reglas de decisión se relacionan con los postulados de H_0 y H_1. En este caso es importante mencionar que en muchas situaciones el error tipo I se considera como un error mucho más grave que el error tipo II, aceptar la premisa del error tipo I es mucho más serio que el error tipo II.

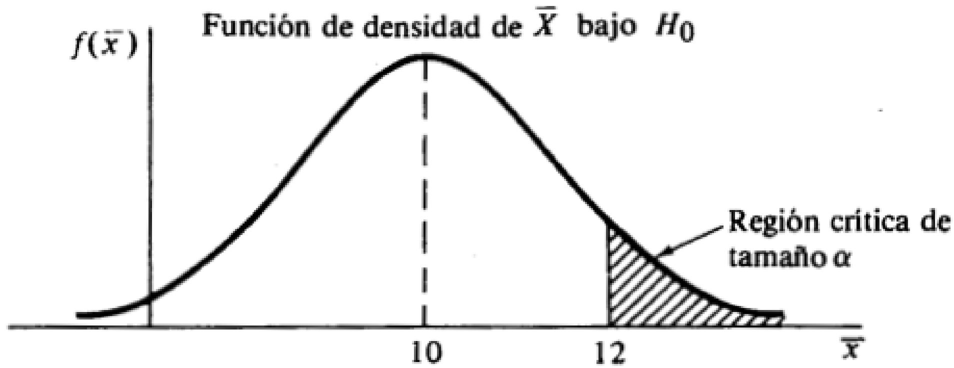


FIGURA 9.1 La región crítica como un área

Un principio sencillo y razonable al obtener reglas de decisión para la prueba de hipótesis estadística es seleccionar aquel procedimiento de prueba que tenga el tamaño más pequeño para el error tipo II entre todos los procedimientos que tengan el mismo tamaño para el error tipo I. El valor de α no puede hacerse muy pequeño sin que se incremente el valor de β. En todo caso, el valor de β es igual a la probabilidad de equivocarse al rechazar H_0 cuando H_1 es cierta.

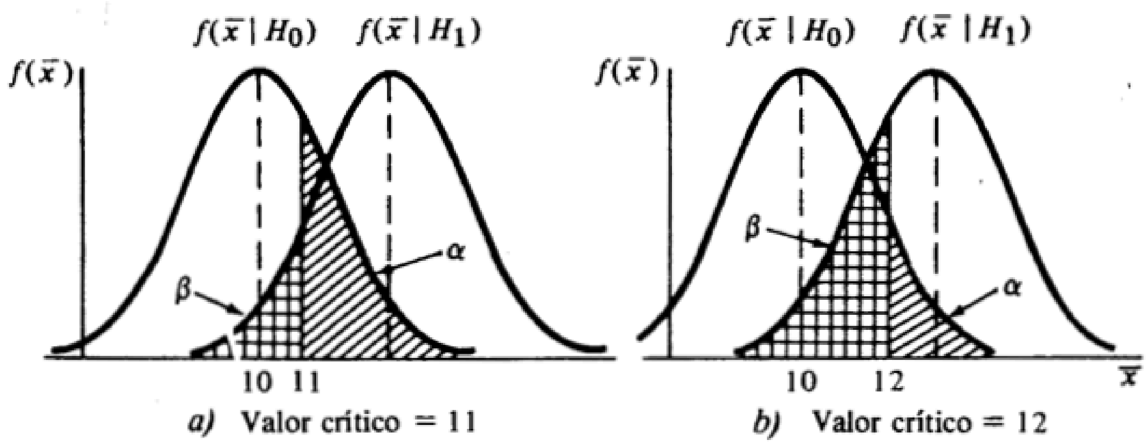


FIGURA 9.2 El efecto sobre α y β al cambiar el valor crítico

La gráfica ilustra como el tamaño de error tipo I disminuye, pero crece el tamaño del error tipo II. La probabilidad de α del error tipo I también se conoce como el **nivel de significancia estadística.**