Taller 1:Diseño de experimentos

Sergio Andres Diaz Vera

2022-08-21

Análisis y Comparación

En un estudio realizado en India en mayo de 2021 se quiería saber si hay diferencia significativa en la carga viral entre personas infectadas con Sars-Cov-2 con la variante beta y la variante delta.

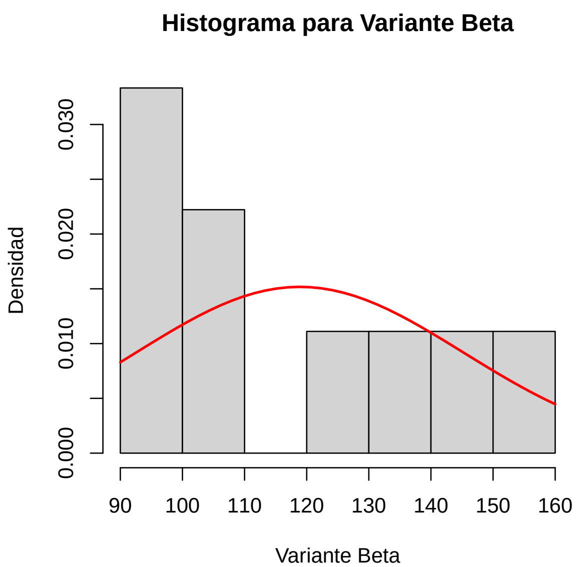
1. Se procede a observar la carga viral entre personas infectadas con la variante beta y delta

	1	2	3	4	5	6	7	8	9
Variante Beta	100	102	130	140	150	160	90	103	95

	1	2	3	4	5	6	7	8	9	10	11
Variante Delta	128	156	100	98	120	160	150	120	129	137	140

Análisis descriptivo Variante Beta

Es posible observar , aunque con pocos datos , la variante beta posee un comportamiento normal además de no presentar datos atípicos (observado en el boxplot) presentando un sesgo a derecha y colas pesadas.



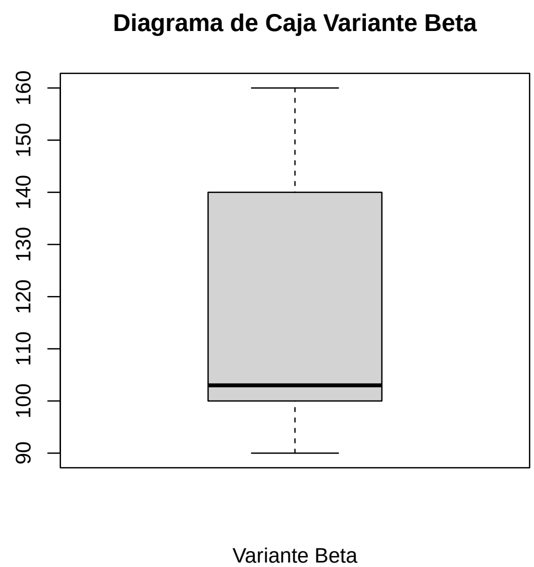
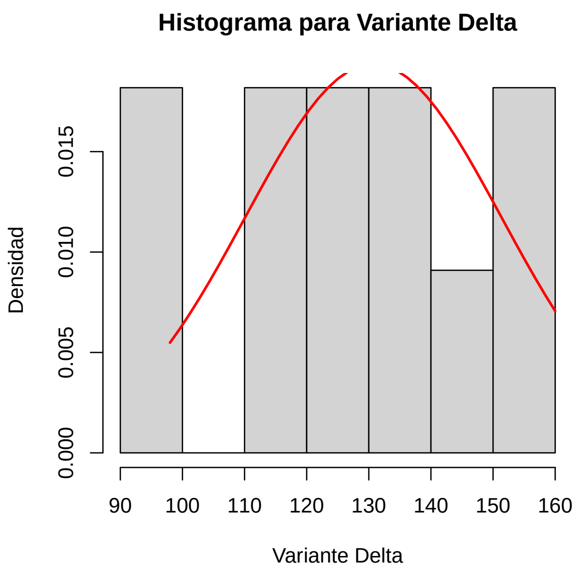


Tabla 1 : Descripción numérica para la variante beta

	Estimaciones
Promedio	118.89
Mediana	103.00
Varianza	690.86
Desvi.Estandar	26.28
Coef.Variación	22.11
Q1	100.00
Q2	103.00
Q3	140.00
Mínimo	90.00
Máximo	160.00

Análisis descriptivo Variante Delta

Es posible observar , aunque con pocos datos , la variante beta posee un comportamiento normal además de no presentar datos atípicos (observado en el boxplot) presentando simetria.



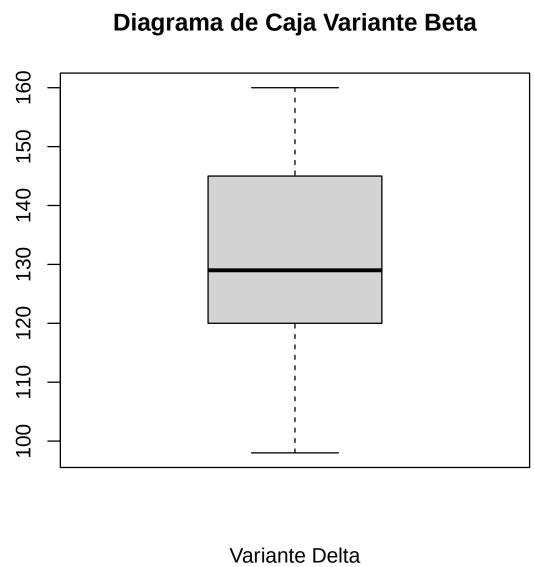


Tabla 2 : Descripción numérica para la Variante Delta

	Estimaciones
Promedio	130.73
Mediana	129.00
Varianza	424.82
Desvi.Estandar	20.61
Coef.Variación	15.77
Q1	120.00
Q2	129.00
Q3	145.00
Mínimo	98.00
Máximo	160.00

Del análsis descriptivo se observa una diferencia sustancial entre las medias y las varianzas de las personas infectadas con la variante beta y delta .

Test de Normalidad

El test de Shapiro-Wilks plantea la hipótesis nula que una muestra proviene de una distribución normal. Se elige un nivel de significanza, por ejemplo 0,05, y tenemos una hipótesis alternativa que sostiene que la distribución no es normal.

Así:

$$H_0 : X \sim N(\mu, \sigma^2)$$
$$H_1 : X \nsim N(\mu, \sigma^2)$$

Ahora el test Shapiro-Wilks intenta rechazar la hipotesis nula al nivel de significancia y puesto que el tamaño de la muestra para ambos grupos es de menos de 50 individuos se usará el test de normalidad de Shapiro-Wilks al ser el mas potente.

```
##
##  Shapiro-Wilk normality test
##
## data:  vBeta$VBeta
## W = 0.87727, p-value = 0.1469
```

```
##
##  Shapiro-Wilk normality test
##
## data:  vDelta$VDelta
## W = 0.95095, p-value = 0.6561
```

dado los p-value de las pruebas siendo 0.1469 para el grupo con la variante Beta y 0.6561 el grupo con la variante Delta , con una significancia de 0.05 es posible decir que no existe evidencia para rechazar la hipotesis nula de normalidad para cada grupo . Es decir con una confianza del 95% los grupos se distribuyen aproxiamadamente normal.

Test de varianzas

Al optener la normalidad de los grupos es posible realizar la prueba F de comparación de varianzas para grupo normales

```
##
##  F test to compare two variances
##
## data:  vBeta$VBeta and vDelta$VDelta
## F = 1.6263, num df = 8, denom df = 10, p-value = 0.4635
## alternative hypothesis: true ratio of variances is not equal to 1
## 95 percent confidence interval:
##  0.421867 6.984956
## sample estimates:
## ratio of variances
##           1.626251
```

El valor p de la prueba es de 0.4653, con una significancia de 0.05 no hay evidencia para rechazar la hipotesis nula de la igualdad de varianzas , es decir con una confianza del 95% los grupos provienen de poblaciones con varianzas iguales.

Test de Comparación de Medias con Varianzas iguales pero desconocidas

Definamos las variables X_1 =“cantidad de carga viral en personas infectadas con la variante Beta” y X_1 =“cantidad de carga viral en personas infectadas con la variante Beta” entonces sabemos por los puntos anteriores que $X_1 \sim N(\mu_1, \sigma_1^2)$ y $X_2 \sim N(\mu_2, \sigma_2^2)$ y además que $\sigma_1^2 = \sigma_2^2$ pero desconocidas , así se plantea el contraste de hipótesis:

$$H_0 : \mu_1 - \mu_2 = 0 \quad VS \quad H_1 : \mu_1 - \mu_2 \neq 0$$

para hallar una diferencia significativa entre los grupos de personas infectadas , veamos el estadístico de prueba

$$t_c = \frac{\bar{x}_1 - \bar{x}_2 - d_0}{S_p\sqrt{1/n_1 + 1/n_2}}$$

donde $S_p = \frac{(n_1-1)s_1^2+(n_1-1)s_2^2}{n_1+n_2-2}$ al tener esto en cuenta se usará la prueba para comparación en R

```
##
##  Two Sample t-test
##
## data:  vBeta$VBeta and vDelta$VDelta
## t = -1.1302, df = 18, p-value = 0.2732
## alternative hypothesis: true difference in means is not equal to 0
## 95 percent confidence interval:
##  -33.84387  10.16711
## sample estimates:
## mean of x mean of y
##  118.8889  130.7273
```

puesto que el valor P es de 0.2732 con una significancia del 0.05 no hay evidencai para recahzar la hipótesisi nula de la igualdad de medias ,por lo tanto con una confianza del 95% no existe evidencia significativa para decir que hay una cantidad media de particulas del virus en la variantes de cada grupo.

Poder estadśitico y tamaño del efecto

```
##
##      t test power calculation
##
##              n1 = 9
##              n2 = 11
##               d = 1.424502
##       sig.level = 0.05
##           power = 0.85
##     alternative = two.sided
```

El poder de la prueba realizada es de aproximadamente el 85% es decir la probabilidad de rechazar la hipotesis nula cuando esta es falasa.

El tamaño del efecto representa el grado en que la hipótesis nula es falsa.El efecto de la prueba de medias para grupos independientes es la de d de cohen calculada de la siguiente manera

$$d = \frac{\bar{X}_1 - \bar{X}_2}{S_{comun}}$$

donde $S_{comun} = \sqrt{\frac{(n_1-1)s_1^2+(n_2-1)s_2^2}{n_1+n_2-2}} = \sqrt{\frac{(9-1)26.28^2+(11-1)20.61^2}{9+11-2}} = 23.3010$