tercer parcial

Andres Felipe Beltran Rodriguez

2/3/2022

# primer punto

## a

El gobierno otorga fondos para los departamentos de agricultura de nueve universidades para probar las capacidades de rendimiento de dos nuevas variedades de trigo. Cada variedad se planta en parcelas de área igual en cada universidad y el rendimiento, en kilogramos por parcela para una muestra tomada aleatoriamente, se registra como sigue:

sureno <- c(39, 23, 35, 41, 44, 29, 37, 31, 38)  
escudo <- c(44, 25, 31, 38, 50, 33, 36, 40, 43)

Pruebe si la variedad escudo es mejor en cuanto a rendimiento en Kg. Suponga que las poblaciones de las cuales fueron tomadas las muestras se distribuyen normalmente y son independientes. Calcular el p valor y concluir. Usar . (10 PUNTOS)

Primero revisamos si las varianzas son diferentes:

var.test(sureno,escudo)

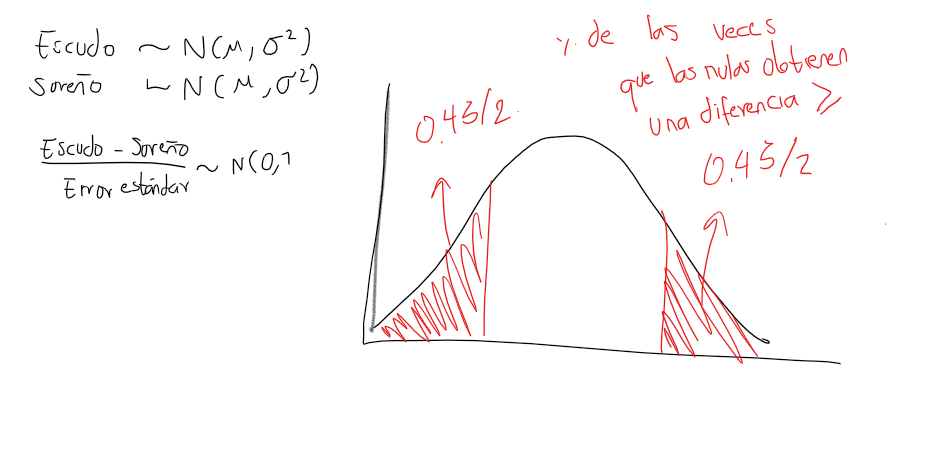
##   
## F test to compare two variances  
##   
## data: sureno and escudo  
## F = 0.74976, num df = 8, denom df = 8, p-value = 0.6935  
## alternative hypothesis: true ratio of variances is not equal to 1  
## 95 percent confidence interval:  
## 0.1691207 3.3238636  
## sample estimates:  
## ratio of variances   
## 0.7497561

Dado que la relación entre las varianzas es diferente a 1, usamos el argumento var.equal=F en la función t.test

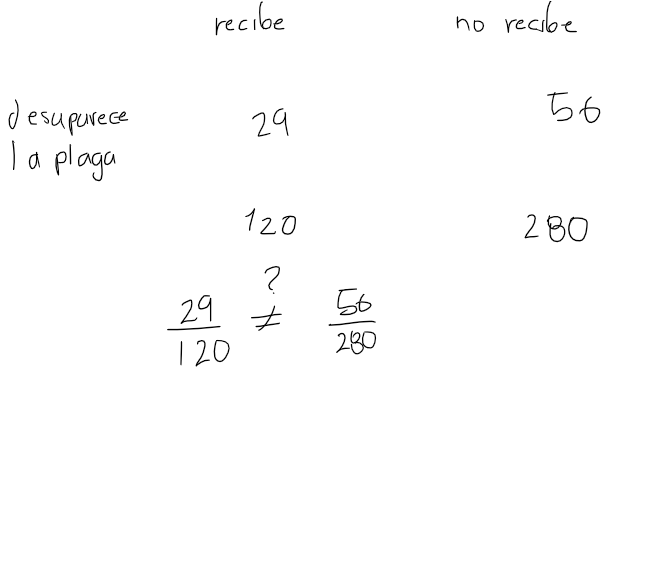
t.test(sureno,escudo,var.equal = F,conf.level = 0.95,alternative = c("two.sided"))

##   
## Welch Two Sample t-test  
##   
## data: sureno and escudo  
## t = -0.76805, df = 15.679, p-value = 0.4539  
## alternative hypothesis: true difference in means is not equal to 0  
## 95 percent confidence interval:  
## -9.620880 4.509769  
## sample estimates:  
## mean of x mean of y   
## 35.22222 37.77778

Cuando se cumple la hipótesis nula, es decir que el peso medio de la variante sureño y el peso medio de la variante escudo son iguales, se obtiene una diferencia igual o mayor a la obtenida al experimento un 45.39 % de las veces. Es decir, una diferencia tal como la observada en el experimento, se podría haber observado aún así los pesos medios de las dos variantes fueran iguales. En este caso, el resultado de esta proporción es lejano al esperado para afirmar que la variedad escudo tiene un mayor rendimiento en kg, uno menor del 5 %.

 ## b

En un cultivo atacado por una plaga, una compañía farmacéutica bien conocida estudió 2000 plantas para determinar si el nuevo insecticida de la compañía era efectivo después de dos días. A 120 plantas que tenían la plaga se les administro el nuevo insecticida, de estas en 29 desapareció la plaga dentro de dos días. Entre 280 plantas que tenían el insecto pero que no recibieron el nuevo insecticida, en 56 desapareció la plaga dentro de dos días. ¿Hay alguna indicación significativa que apoye la afirmación de la compañía de la efectividad del nuevo insecticida?, utilice una confianza del 95% e interprete el intervalo. (10 PUNTOS)



Cargamos los datos:

x <-c(29,56)  
n <- c(120,280)  
prop.test(x,n,conf.level = 0.95)

##   
## 2-sample test for equality of proportions with continuity correction  
##   
## data: x out of n  
## X-squared = 0.64026, df = 1, p-value = 0.4236  
## alternative hypothesis: two.sided  
## 95 percent confidence interval:  
## -0.05407319 0.13740653  
## sample estimates:  
## prop 1 prop 2   
## 0.2416667 0.2000000

Dado que el intervalo de cofianza para la diferencia de proporciones contiene el cero, se puede afirmar con un 95% de confianza que el uso del insecticida no presenta resultados significativamente diferentes al caso en el que no se usa, ya que la proporción de plantas que presentan mejoría respecto al total, no es diferente a cuando no se usa el insecticida.

2. Se midió los porcentajes de nitrógeno (), cloro(), potasio () y log de combustión foliar en segundos() para 30 muestras de tabaco tomadas en campos de agricultores. **(20 PUNTOS)**

1. Determine la ecuación de regresión ajustada e interprete el coeficiente correspondiente a cloro obtenido del modelo.



El cloro tiene un efecto inversamente proporcional al logaritmo de la combustión foliar. A su vez, al mantener las demás variables constantes, al aumentar una unidad de porcentaje de cloro, el logaritmo de combustión foliar disminuye en 0.44 unidades. Además, presenta un p’valor de 2.34e-06, lo cual indica que es significativo e importante para el modelo de regresión

1. Plantear claramente las hipótesis de interés del ANOVA y concluya a través de la salida mostrada. Utilice un nivel de significancia de 0.05.

Para el análisis de varianzas se proponen las hipótesis

El resultado de la prueba F en el análisis de varianza presenta un p-valor de 6.454e-10, mucho menor que el nivel de significancia de 0.05. Lo cual nos permite rechazar la hipótesis nula, indicando que por lo menos uno de los coeficientes, el intercepto o una de las pendientes es importante y contribuye al modelo de regresión

1. Interprete el coeficiente de determinación (el que sea apropiado). Diga si el modelo es bueno.

El valor de R ajustado, el cual tiene en cuenta los grados de libertad y permite comparar con otros modelos es de 0.8025, es decir que 80.25% de la variación es explicada por el modelo.

1. Verifique si se cumple el supuesto de normalidad en el modelo con un nivel de significancia de 0.05. Concluya en términos del experimento. **(Utilice la prueba más potente y especifique la usada en su conclusión)**.

El valor de p para la prueba de shapiro wilk de los residuales es de 0.1135, un valor mayor al nivel de significancia de 0.05. Esto no permite rechazar la hipótesis nula es decir, los residuales presentan distribución normal. El error del modelo al predecir el logaritmo del tiempo de combustión no es dependiente del porcentaje de cloro, nitrógeno o potasio.

1. Verifique si se cumple el supuesto de la homogeneidad de varianzas, con un nivel de significancia de 0.05. **(Utilice la prueba más potente y especifique la usada en su conclusión).**

Dado que el resultado en la prueba de Bartlett presenta un estadístico de 0.7913, no es posible rechazar la hipótesis nula, en la que no hay diferencia entre las varianzas, es decir no hay diferencia en el logaritmo del tiempo de combustión relativo al contenido de cloro, nitrógeno o potasio

1. Plantee la hipótesis de interés para ver si hay autocorrelación o no y pruébela bajo . Concluya.

Hipótesis nula = no correlación entre los resuduales

Hipótesis alternativa = los residuales están auto correlacionados

Dado que el p valor para la prueba de durbin y Watson es mayor a 0.05 no es posible rechazar la hipótesis nula, por ende los residuales no están correlacionados.

1. Cuál es el número de log de combustión foliar que se esperaría encontrar si se tiene 1.10 de % de cloro, 3.52 % de nitrógeno, y 3.17% de potasio. Concluya.



1. Construya un intervalo de confianza del 95% para el valor promedio de *Y,* con los valores del ítem anterior, e intérprete el resultado obtenido.
2. Construya un intervalo de confianza del 95% para el parámetro asociado a cloro e interprete.



Dado que el intervalo de confianza contiene el cero, el parámetro asociado a cloro no es diferente de cero.

j. Interpretar los gráficos adjuntos.

En los graficos

=MATRIZ INVERSA

|  |  |  |  |
| --- | --- | --- | --- |
| 1.71457789 | -0.32895337 | 0.09529059 | -0.14605239 |
| -0.32895337 | 0.10623365 | -0.02105557 | -0.00050401 |
| 0.09529059 | -0.02105557 | 0.11706378 | -0.02595908 |
| -0.14605239 | -0.00050401 | -0.02595908 | 0.03624479 |

**SALIDA DE R PARA EL EJERCICIO PLANTEADO JUNTO CON LA VALIDACIÓN DE SUPUESTOS**

Call:

lm(formula = logcombfoliar ~ nitrogeno + cloro + potasio, data = datos)

Residuals:

Min 1Q Median 3Q Max

-0.41728 -0.10707 0.04296 0.11911 0.30544

Coefficients:

Estimate Std. Error t value Pr(>|t|)

(Intercept) 1.81104 0.27952 6.479 7.24e-07 \*\*\*

nitrogeno -0.53146 0.06958 -7.638 4.16e-08 \*\*\*

cloro -0.43964 0.07304 -6.019 2.34e-06 \*\*\*

potasio 0.20898 0.04064 5.142 2.31e-05 \*\*\*

---

Signif. codes: 0 ‘\*\*\*’ 0.001 ‘\*\*’ 0.01 ‘\*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 0.2135 on 26 degrees of freedom

Multiple R-squared: 0.8229, Adjusted R-squared: 0.8025

F-statistic: 40.27 on 3 and 26 DF, p-value: 6.454e-10

Start: AIC=-88.95

logcombfoliar ~ nitrogeno + cloro + potasio

Df Sum of Sq RSS AIC

<none> 1.1848 -88.949

- potasio 1 1.2049 2.3897 -69.901

- cloro 1 1.6511 2.8358 -64.766

- nitrogeno 1 2.6587 3.8435 -55.644

Call:

lm(formula = logcombfoliar ~ nitrogeno + cloro + potasio, data = datos)

Coefficients:

(Intercept) nitrogeno cloro potasio

1.8110 -0.5315 -0.4396 0.2090

**studentized Breusch-Pagan test**

data: modelo

BP = 2.9649, df = 3, p-value = 0.3971

**Bartlett test of homogeneity of variances**

data: modelo$residuals and grupos

Bartlett's K-squared = 1.6969, df = 4, p-value = 0.7913

**inflación de la varianza**

nitrogeno cloro potasio

1.045928 1.243327 1.199083

**Durbin y Watson**

lag Autocorrelation D-W Statistic p-value

1 -0.2797321 2.426087 0.304

**Shapiro-Wilk normality test**

data: modelo$residuals

W = 0.94357, p-value = 0.1135

**One-sample Kolmogorov-Smirnov test**

data: n.qo

D = 0.11815, p-value = 0.7529

alternative hypothesis: two-sided

Imagen que contiene Calendario

Descripción generada automáticamente



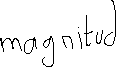
Gráfico

Descripción generada automáticamente con confianza media



Gráfico, Gráfico de líneas

Descripción generada automáticamente



Gráfico, Gráfico de líneas

Descripción generada automáticamente