Tarea 5

EJERCICIO 1 (α=0.05)

1. Primero realizamos los tres modelos y después se analizan conjuntamente. Comenzamos realizando el modelo de regresión simple tomando como variable respuesta la variable sabor y la variable explicativa la concentración de ácido acético.

```
> modeloacet=lm(sabor ~ acetico)
> summary(modeloacet)

Call:
lm(formula = sabor ~ acetico)

Residuals:
     Min       1Q   Median       3Q      Max
-29.4340 -10.6740   0.5875   8.8995  28.7073

Coefficients:
            Estimate Std. Error t value Pr(>|t|)
(Intercept)  9.65407    5.59632   1.725  0.09595 .
acetico      0.05255    0.01708   3.076  0.00477 **
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 14.46 on 27 degrees of freedom
Multiple R-squared:  0.2595,    Adjusted R-squared:  0.2321
F-statistic: 9.461 on 1 and 27 DF,  p-value: 0.004766
```

sabor = 9.65407 + 0.05255 * acético + u(error)

El siguiente modelo corresponde al de variable respuesta sabor y variable explicativa la concentración de ácido sulfhídrico.

```
> modeloh2s=lm(sabor ~ H2S)
> summary(modeloh2s)

Call:
lm(formula = sabor ~ H2S)

Residuals:
    Min      1Q  Median      3Q     Max
-19.759  -8.116  -3.019   4.912  33.172

Coefficients:
             Estimate Std. Error t value Pr(>|t|)
(Intercept) 2.038e+01  2.876e+00   7.087 1.28e-07 ***
H2S         1.575e-03  4.508e-04   3.494  0.00166 **
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 13.94 on 27 degrees of freedom
Multiple R-squared:  0.3114,    Adjusted R-squared:  0.2859
F-statistic: 12.21 on 1 and 27 DF,  p-value: 0.001658
```

sabor = 20.38 + 0.001575 * H2S + u(error)

Y por último el modelo con variable explicativa la concentración de ácido láctico.

```
> modelolac=lm(sabor ~ lactico)
> summary(modelolac)

Call:
lm(formula = sabor ~ lactico)

Residuals:
     Min       1Q   Median       3Q      Max
-20.4685  -9.2754   0.2982   8.6393  26.8697

Coefficients:
            Estimate Std. Error t value Pr(>|t|)
(Intercept)  -30.833     10.533  -2.927  0.00686 **
lactico       38.726      7.181   5.393 1.06e-05 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 11.66 on 27 degrees of freedom
Multiple R-squared:  0.5186,    Adjusted R-squared:  0.5008
F-statistic: 29.08 on 1 and 27 DF,  p-value: 1.061e-05
```

sabor = - 30.833 + 38.726 * láctico + u(error)

A continuación, se analizan conjuntamente las preguntas:

a) Para el modelo del acético:
   - $\boldsymbol{\beta_0}$ **= 9.65407**
   - $\boldsymbol{\beta_1}$ **= 0.05255**

   Para el modelo del H2S:

   - $\boldsymbol{\beta_0}$ **= 20.38**
   - $\boldsymbol{\beta_1}$ **= 0.001575**

   Para el modelo del láctico:

   - $\boldsymbol{\beta_0}$ **= - 30.833**
   - $\boldsymbol{\beta_1}$ **= 38.726**

b) Para el modelo del acético:
   El **p-valor** para el contraste H_0: $\beta_1 = 0$; H_1 : β_1 =! 0 tiene un valor de **0.00477**. Este valor se considera pequeño, por lo que se rechaza la hipótesis inicial, siendo el regresor **significativo.**
   Para el modelo del H2S:
   El **p-valor** para el contraste H_0: $\beta_1 = 0$; H_1 : β_1 =! 0 tiene un valor de **0.00166**. Este valor se considera pequeño, por lo que se rechaza la hipótesis inicial, siendo el regresor **significativo.**
   Para el modelo del acético:
   El **p-valor** para el contraste H_0: $\beta_1 = 0$; H_1 : β_1 =! 0 tiene un valor de **1.06e-05**. Este valor se considera muy pequeño, por lo que se rechaza la hipótesis inicial, siendo el regresor **significativo**.

c) Para el modelo del ácido acético se ha obtenido un coeficiente de $\boldsymbol{\beta_1}$ **= 0.05255**. Esto significa que por una unidad de concentración de ácido acético el sabor aumenta en promedio un 0.05255.

2. Primero realizamos los tres modelos y después se analizan conjuntamente.
   El primer modelo de regresión lineal múltiple corresponde con el que tiene por variable respuesta el sabor y regresores el ácido acético y el H2S.

```
modelomult1=lm( sabor ~ acetico + H2S)
summary(modelomult1)
```

```
Call:
lm(formula = sabor ~ acetico + H2S)

Residuals:
     Min       1Q   Median       3Q      Max
-23.5465  -9.9548  -0.2527   5.2678  25.3864

Coefficients:
             Estimate Std. Error t value Pr(>|t|)
(Intercept) 9.8157510  4.9247137   1.993   0.0568 .
acetico     0.0396575  0.0156438   2.535   0.0176 *
H2S         0.0012750  0.0004281   2.978   0.0062 **
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 12.72 on 26 degrees of freedom
Multiple R-squared:  0.4479,    Adjusted R-squared:  0.4054
F-statistic: 10.55 on 2 and 26 DF,  p-value: 0.0004432
```

sabor = 9.8157510 + 0.0396575 * acético + 0.0012750 * H2S + u(error)

El segundo modelo tiene por regresores el ácido acético y el láctico.

```
modelomult2=lm( sabor ~ acetico + lactico)
summary(modelomult2)
```

```
Call:
lm(formula = sabor ~ acetico + lactico)

Residuals:
     Min       1Q   Median       3Q      Max
-19.2145  -7.1292   0.7287   9.9050  24.3869

Coefficients:
             Estimate Std. Error t value Pr(>|t|)
(Intercept) -28.73165   10.78299  -2.665 0.013067 *
acetico       0.01581    0.01668   0.948 0.351895
lactico      34.09596    8.69552   3.921 0.000574 ***
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 11.68 on 26 degrees of freedom
Multiple R-squared:  0.5347,    Adjusted R-squared:  0.4989
F-statistic: 14.94 on 2 and 26 DF,  p-value: 4.797e-05
```

sabor = - 28.73165 + 0.01581 * acético + 34.09596 * lactico + u(error)

El tercer modelo tiene por regresores el ácido H2S y el láctico.

```
modelomult3=lm( sabor ~ H2S + lactico)
summary(modelomult3)
```

```
Call:
lm(formula = sabor ~ H2S + lactico)

Residuals:
    Min      1Q  Median      3Q     Max
-18.194  -6.415  -1.875   7.307  28.827

Coefficients:
             Estimate Std. Error t value Pr(>|t|)
(Intercept) -2.271e+01  1.133e+01  -2.005 0.055519 .
H2S          7.085e-04  4.280e-04   1.655 0.109917
lactico      3.169e+01  8.154e+00   3.887 0.000627 ***
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 11.3 on 26 degrees of freedom
Multiple R-squared:  0.5645,    Adjusted R-squared:  0.531
F-statistic: 16.85 on 2 and 26 DF,  p-value: 2.029e-05
```

sabor = - 22.71 + 0.0007085 * H2S + 31.69 * lactico + u(error)

A continuación, se analizan conjuntamente las preguntas:

a) Para el modelo de regresores acético y H2S:
   - β_0 **= 9.8157510**
   - β_1 **= 0.0396575 (acético)**
   - β_2 **= 0.0012750 (H2S)**

   Para el modelo de regresores acético y láctico:

   - β_0 **= - 28.73165**
   - β_1 **= 0.01581 (acético)**
   - β_2 **= 34.09596 (láctico)**

   Para el modelo de regresor H2S y láctico:

   - β_0 **= - 22.71**
   - β_1 **= 0.0007085 (H2S)**
   - β_2 **= 31.69 (láctico)**

b) Para el modelo de acético y H2S:

   El **p-valor** para el contraste H_0: $\beta_1 = 0$; H_1 : $\beta_1 \neq 0$ tiene un valor de **0.0176**. Este valor se considera pequeño, por lo que se rechaza la hipótesis inicial, siendo el regresor **significativo.**
   El **p-valor** para el contraste H_0: $\beta_2 = 0$; H_1 : $\beta_2 \neq 0$ tiene un valor de **0.0062**. Este valor se considera pequeño, por lo que se rechaza la hipótesis inicial, siendo el regresor **significativo.**

   Para el modelo del acético y láctico:

   El **p-valor** para el contraste H_0: $\beta_1 = 0$; H_1 : $\beta_1 \neq 0$ tiene un valor de **0.351895**. Este valor se considera grande, por lo que se acepta la hipótesis inicial, siendo el regresor **no significativo.**

El **p-valor** para el contraste H_0: $\beta_2 = 0$; H_1 : $\beta_2 \neq 0$ tiene un valor de **0.000574**. Este valor se considera pequeño, por lo que se rechaza la hipótesis inicial, siendo el regresor **significativo.**

Para el modelo del H2S y láctico:

El **p-valor** para el contraste H_0: $\beta_1 = 0$; H_1 : $\beta_1 \neq 0$ tiene un valor de **0.109917**. Este valor se considera grande, por lo que se acepta la hipótesis inicial, siendo el regresor **no significativo**.
El **p-valor** para el contraste H_0: $\beta_2 = 0$; H_1 : $\beta_2 \neq 0$ tiene un valor de **0.000627**. Este valor se considera pequeño, por lo que se rechaza la hipótesis inicial, siendo el regresor **significativo.**

c) En el modelo simple del acético el regresor tiene un valor de 0.05255, en cambio en los dos modelos múltiples en los cuales aparece de nuevo toma un valor de 0.0396575 (modelo con H2S) y de 0.01581 (modelo con láctico).
Se puede observar que el coeficiente disminuye respecto del modelo lineal.
El otro cambio notable es el de los p-valores. En el modelo simple tiene un p-valor de 0.00477. En los modelos múltiples tiene unos p-valores de 0.0176 (modelo con H2S) y de 0.351895 (modelo con láctico).
Podemos observar que el p-valor respecto del modelo simple aumenta.
Otra diferencia relacionada con el p-valor es que para el modelo simple el coeficiente es significativo, cosa que en el modelo múltiple con H2S también sucede, pero en el modelo múltiple con el láctico no es significativo.

3. Realizamos el modelo de regresión lineal múltiple con variable respuesta la variable sabor y como regresores los tres ácidos.

```
modelomult=lm(sabor ~ acetico + H2S + lactico)
summary(modelomult)
```

```
Call:
lm(formula = sabor ~ acetico + H2S + lactico)

Residuals:
    Min      1Q  Median      3Q     Max
-16.860  -5.321  -1.422   8.116  26.275

Coefficients:
             Estimate Std. Error t value Pr(>|t|)
(Intercept) -2.041e+01  1.154e+01  -1.769  0.08917 .
acetico      1.642e-02  1.613e-02   1.018  0.31836
H2S          7.183e-04  4.278e-04   1.679  0.10564
lactico      2.679e+01  9.467e+00   2.830  0.00905 **
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 11.29 on 25 degrees of freedom
Multiple R-squared:  0.5818,    Adjusted R-squared:  0.5316
F-statistic: 11.59 on 3 and 25 DF,  p-value: 5.947e-05
```

sabor = - 20.41 + 0.01642 * acético + 0.0007183 * H2S + 26.79 * lactico + u(error)

Para el modelo obtenido tenemos los siguientes coeficientes:

- $\boldsymbol{\beta_1}$ **= 0.01642.** Este primer coeficiente indica que, por cada aumento de una unidad de concentración de ácido acético, manteniéndose constantes el resto de los ácidos, el sabor aumenta en promedio 0.01642 unidades.
  El **p-valor** para el contraste H_0: $\beta_1 = 0$; H_1 : $\beta_1 \neq 0$ tiene un valor de **0.31836**. Este valor se considera grande, por lo que se acepta la hipótesis inicial, siendo el regresor **no significativo**.
- $\boldsymbol{\beta_2}$ **= 0.0007183.** Este coeficiente indica que, por cada aumento de una unidad de concentración de H2S, manteniéndose el resto de los ácidos constantes, el sabor aumenta en promedio 0.0007183 unidades.
  El **p-valor** para el contraste H_0: $\beta_2 = 0$; H_1 : $\beta_2 \neq 0$ tiene un valor de **0.10564**. Este valor se considera grande, por lo que se acepta la hipótesis inicial, siendo el regresor **no significativo**.
- $\boldsymbol{\beta_3}$ **= 26.79.** Este coeficiente indica que, por cada aumento de una unidad de concentración de ácido láctico, siendo el resto de los regresores constantes, el sabor aumenta en promedio 26.79 unidades.
  El **p-valor** para el contraste H_0: $\beta_3 = 0$; H_1 : $\beta_3 \neq 0$ tiene un valor de **0.00905**. Este valor se considera pequeño, por lo que se rechaza la hipótesis inicial, siendo el regresor **significativo**.

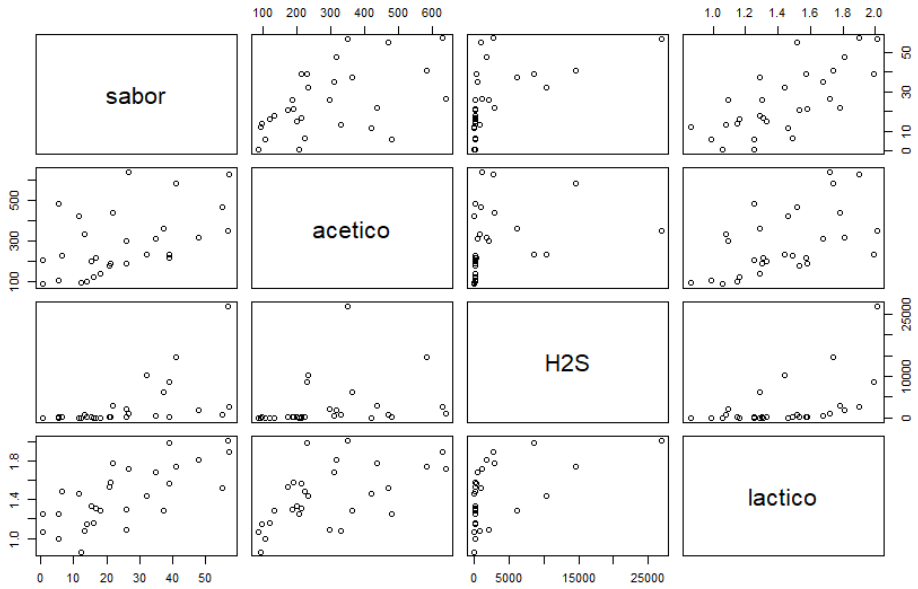
4. Ente los 7 modelos estimados:

   - El más adecuado para explicar el efecto del regresor acético es el modelo de regresión simple de este. En este modelo el coeficiente estimado para el regresor tiene un error menor, un estadístico t mayor y un p-valor más pequeño. Con ese **p-valor (0.00477)** sabemos que el regresor acético tiene un valor significativo sobre el sabor.

   - Para elegir el mejor modelo para predecir el sabor nos tenemos que fijar en el R^2 ajustado. Comparando los 7 modelos mencionados anteriormente podemos observar que el modelo cuyo $\boldsymbol{R^2}$ **ajustado** es mayor es el modelo de regresión múltiple que incluye los tres ácidos, con un valor de **0.5316**. Además, el modelo múltiple con regresores H2S y láctico se queda muy próximo ya que tiene un R^2 ajustado de 0.531.

5. Realizando en RStudio los cálculos pedidos se obtiene:

```
pairs(datos[,])
```




```
> cor(datos[,])
              sabor   acetico       H2S   lactico
sabor    1.0000000 0.5094020 0.5580302 0.7201267
acetico  0.5094020 1.0000000 0.2767023 0.5616802
H2S      0.5580302 0.2767023 1.0000000 0.5209933
lactico  0.7201267 0.5616802 0.5209933 1.0000000
```

Como podemos observar en la matriz de correlación (simétrica) los regresores tienen alta correlación entre ellos, se da multicolinealidad. Esto hace que los errores de los regresores y sus p-valores sean mayores. Algunos regresores que en sus modelos simples son significativos en este no lo son debido al efecto de la multicolinealidad. También se ve este efecto en los estadísticos t, que se ven reducidos.

El coeficiente de correlación del acético y el láctico en el modelo múltiple de los tres ácidos es de 0.5616802. Valor alto, por lo que hay bastante correlación entre ambos regresores por esos los p-valores aumentan y el acético en este modelo resulta ser no significativo.

EJERCICIO 2 (α=0.05)

1. El modelo de regresión simple resultante es:

anchurapetalo = 3.1569 – 0.6403 * anchurasepalo + u(error)

```
> modelo=lm(Petal.Width ~ Sepal.Width)
> summary(modelo)

Call:
lm(formula = Petal.Width ~ Sepal.Width)

Residuals:
     Min       1Q   Median       3Q      Max
-1.38424 -0.60889 -0.03208  0.52691  1.64812

Coefficients:
            Estimate Std. Error t value Pr(>|t|)
(Intercept)   3.1569     0.4131   7.642 2.47e-12 ***
Sepal.Width  -0.6403     0.1338  -4.786 4.07e-06 ***
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 0.7117 on 148 degrees of freedom
Multiple R-squared:  0.134,     Adjusted R-squared:  0.1282
F-statistic: 22.91 on 1 and 148 DF,  p-value: 4.073e-06
```

Para este modelo resultan las siguientes estimaciones:

- $\boldsymbol{\beta_0}$ **= 3.1569**
- $\boldsymbol{\beta_1}$ **= - 0.6403**

2. Según el modelo anterior.

   Para $\boldsymbol{\beta_0}$ **:** se tiene un **estadístico t** de **7.642**. El **p-valor** para el contraste H_0: β_0 = 0 ; H_1 : β_0 =! 0 tiene un valor de **2.47e-12**. Este valor se considera muy pequeño, por lo que se rechaza la hipótesis inicial, siendo el regresor **significativo**.

   Para $\boldsymbol{\beta_1}$ **:** se tiene un **estadístico t** de **- 4.786**. El **p-valor** para el contraste H_0: β_1 = 0 ; H_1 : β_1 =! 0 tiene un valor de **4.07e-06**. Este valor se considera muy pequeño, por lo que se rechaza la hipótesis inicial, siendo el regresor **significativo**.

   El **contraste** general tiene por hipótesis inicial que los valores estimados no sean significativos y como contraste que al menos uno de ellos lo sea, cosa que sucede en este modelo. Se rige por una distribución t-student de 148 grados de libertad. Ésta tiene un **p-valor** general de **4.073e-06**, confirmando lo anterior, rechazando la hipótesis inicial y confirmando que al menos un estimador es significativo. Además, su **estadístico** es **22.91**.

   La **bondad de ajuste** del modelo es R^2 = 13.4 % que no es un valor muy alto, por lo que sabemos que la anchura del sépalo explica un 13.4% de la anchura del pétalo.

3. El modelo de regresión múltiple resultante (con la variable cualitativa especie setosa como variable de referencia) es:

```
modelomult=lm( Petal.Width ~ Sepal.Width + Species)
summary(modelomult)
```

```
Call:
lm(formula = Petal.Width ~ Sepal.Width + Species)

Residuals:
     Min       1Q   Median       3Q      Max
-0.51998 -0.09476 -0.01406  0.10161  0.42333

Coefficients:
                  Estimate Std. Error t value Pr(>|t|)
(Intercept)       -0.72577    0.15298  -4.744 4.94e-06 ***
Sepal.Width        0.28348    0.04400   6.443 1.59e-09 ***
Speciesversicolor  1.26653    0.04638  27.306  < 2e-16 ***
Speciesvirginica   1.90870    0.04138  46.127  < 2e-16 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 0.1812 on 146 degrees of freedom
Multiple R-squared:  0.9446,    Adjusted R-squared:  0.9435
F-statistic: 830.2 on 3 and 146 DF,  p-value: < 2.2e-16
```

anchurapetalo = - 0.72577 + 0.28348 * anchurasepalo + 1.26653 * especieversicolor + 1.90870 * especievirginica + u(error)

Para el modelo estimado:

-β_0 **= - 0.72577.** Carece de sentido físico puesto que la anchura de un pétalo no puede ser negativa. Dependiendo de los valores que tomen las variables cualitativas y sus correspondientes α este valor se verá afectado. Es significativo puesto que su p-valor es muy pequeño.
-β_1 **= 0.28348.** Explica, a igualdad del resto de variables, la relación entre la anchura del pétalo y el regresor anchura del sépalo. Por cada aumento de una unidad de anchura del sépalo, la anchura del pétalo aumenta en promedio 0.028348. Su valor es significativo debido a su p-valor.
-$\alpha_{\text{versicolor}}$ **= 1.26653.** Este valor explica cómo las flores versicolor tienen una anchura de pétalo 1.26653 mayor, a igualdad de variables, y respecto de una flor de especie setosa. Debido a su p-valor, es significativo su valor.
-$\alpha_{\text{virgínica}}$ **= 1.90870.** Este valor explica cómo las flores virgínicas tienen una anchura de pétalo 1.90870 mayor, a igualdad de variables, y respecto de una flor de especie setosa. Debido a su p-valor, es significativo su valor.

4. Según el modelo anterior.

Para $\boldsymbol{\beta_0}$ **:** se tiene un **estadístico t** de **-4.744**. El **p-valor** para el contraste H_0: $\beta_0 = 0$; H_1 : β_0 =! 0 tiene un valor de **4.94e-06**. Este valor se considera muy pequeño, por lo que se rechaza la hipótesis inicial, siendo el regresor **significativo**.

Para $\boldsymbol{\beta_1}$ **:** se tiene un **estadístico t** de **6.443**. El **p-valor** para el contraste H_0: $\beta_1 = 0$; H_1 : β_1 =! 0 tiene un valor de **1.59e-09**. Este valor se considera muy pequeño, por lo que se rechaza la hipótesis inicial, siendo el regresor **significativo**.

Para $\alpha_{versicolor}$**:** se tiene un **estadístico t** de **27.306**. El **p-valor** para el contraste H_0: $\alpha_{versicolor}$ = 0 ; H_1 : $\alpha_{versicolor}$ =! 0 tiene un valor menor de **2e-16**. Este valor se considera muy pequeño, por lo que se rechaza la hipótesis inicial, siendo el regresor **significativo**.

Para $\alpha_{virgínica}$**:** se tiene un **estadístico t** de **46.127**. El **p-valor** para el contraste H_0: $\alpha_{virgínica}$ = 0 ; H_1 : $\alpha_{virgínica}$ =! 0 tiene un valor menor de **2e-16**. Este valor se considera muy pequeño, por lo que se rechaza la hipótesis inicial, siendo el regresor **significativo**.

El **contraste general** tiene por hipótesis inicial que los valores estimados no sean significativos y como contraste que al menos uno de ellos lo sea, cosa que sucede en este modelo. Se rige por una distribución t-student de 146 grados de libertad. Ésta tiene un **p-valor** general **menor** de **2.2e-16**, confirmando lo anterior, rechazando la hipótesis inicial y confirmando que al menos un estimador es significativo. Además, su **estadístico** es **830.2**.

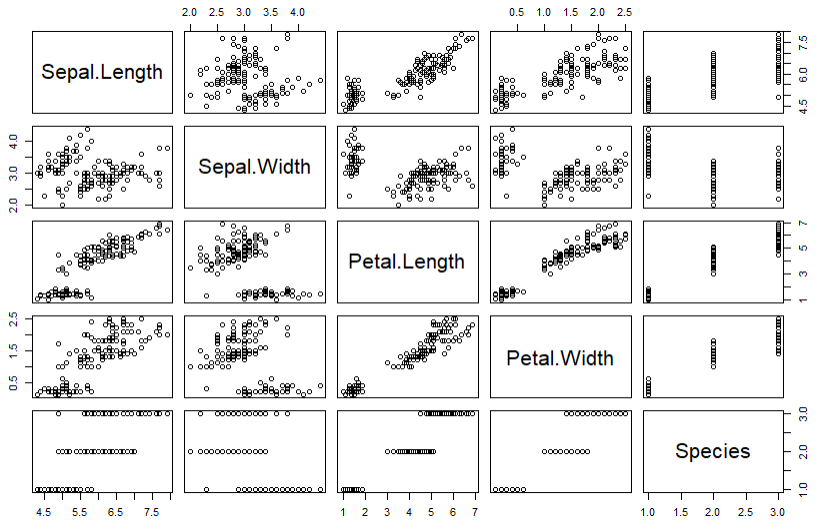
El **intervalo de confianza** para la **varianza** es: **[0.02643417 , 0.04188792]**

```
> chia=qchisq(0.025, 146, lower.tail=T)
> chib=qchisq(0.975, 146, lower.tail=T)
> LInferior=146*0.1812^2/chib
> LSuperior=146*0.1812^2/chia
> LInferior
[1] 0.02643417
> LSuperior
[1] 0.04188792
```

5. Para el modelo de regresión simple tenemos una bondad del ajuste ajustada de 0.1282, para el modelo de regresión múltiple es 0.9435. Lo que significa que el modelo múltiple explica un mayor porcentaje de la variación de la anchura del pétalo, concretamente un 94,46%.

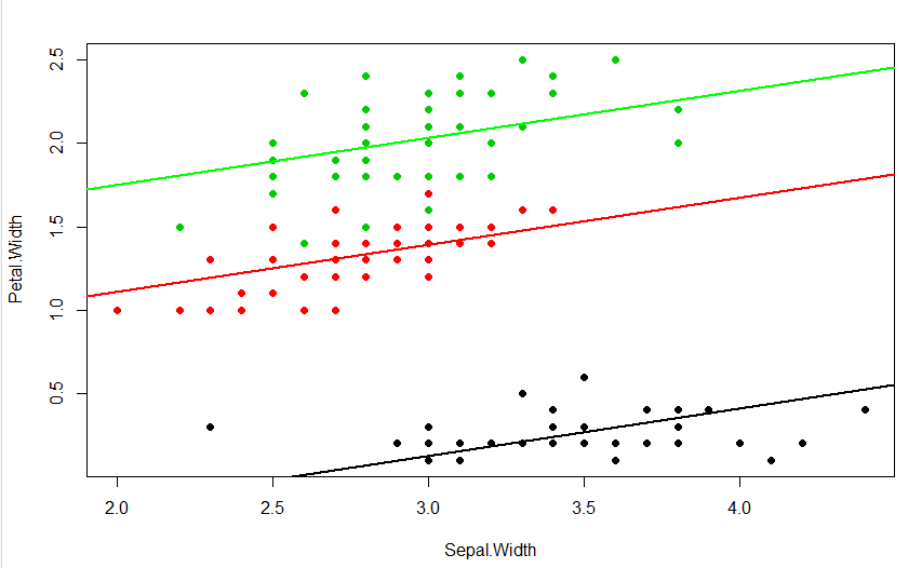
   En el modelo de regresión simple el coeficiente de regresión de la anchura del sépalo es de – 0.6403 y en el múltiple es 0.28348. Estas diferencias pueden deberse a la multicolinealidad y la alta correlación entre los regresores.

```
> cor(datos[,1:4])
             Sepal.Length Sepal.Width Petal.Length Petal.Width
Sepal.Length    1.0000000  -0.1175698    0.8717538   0.8179411
Sepal.Width    -0.1175698   1.0000000   -0.4284401  -0.3661259
Petal.Length    0.8717538  -0.4284401    1.0000000   0.9628654
Petal.Width     0.8179411  -0.3661259    0.9628654   1.0000000
```




```
plot(Petal.Width ~ Sepal.Width, pch=19, data=datos, col=Species)

abline(c(-0.72577, 0.28348), col='black', lwd=2)
abline(c( -0.72577 + 1.26653 , 0.28348), col='red', lwd=2)
abline(c(-0.72577 + 1.90870, 0.28348), col='green', lwd=2)
```



Como se observa en este gráfico la especie influye en el valor por eso el coeficiente también varia del simple, por separado cada especie, al múltiple, todas juntas.