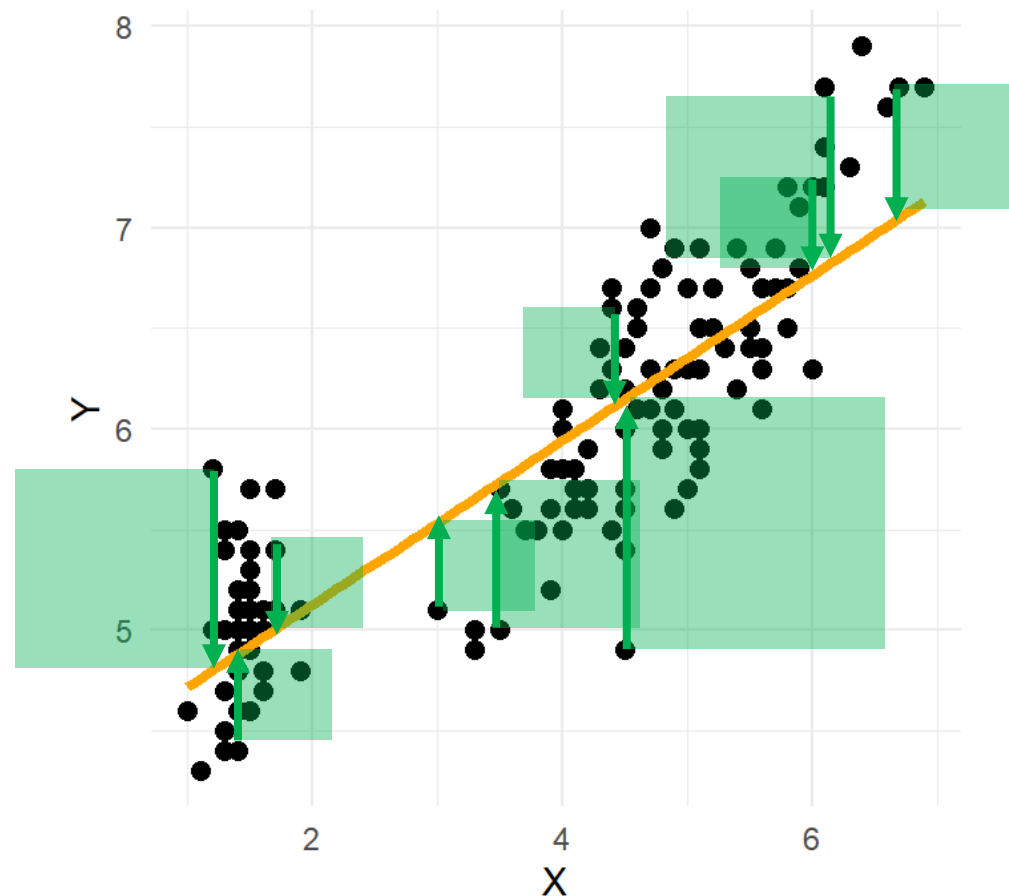
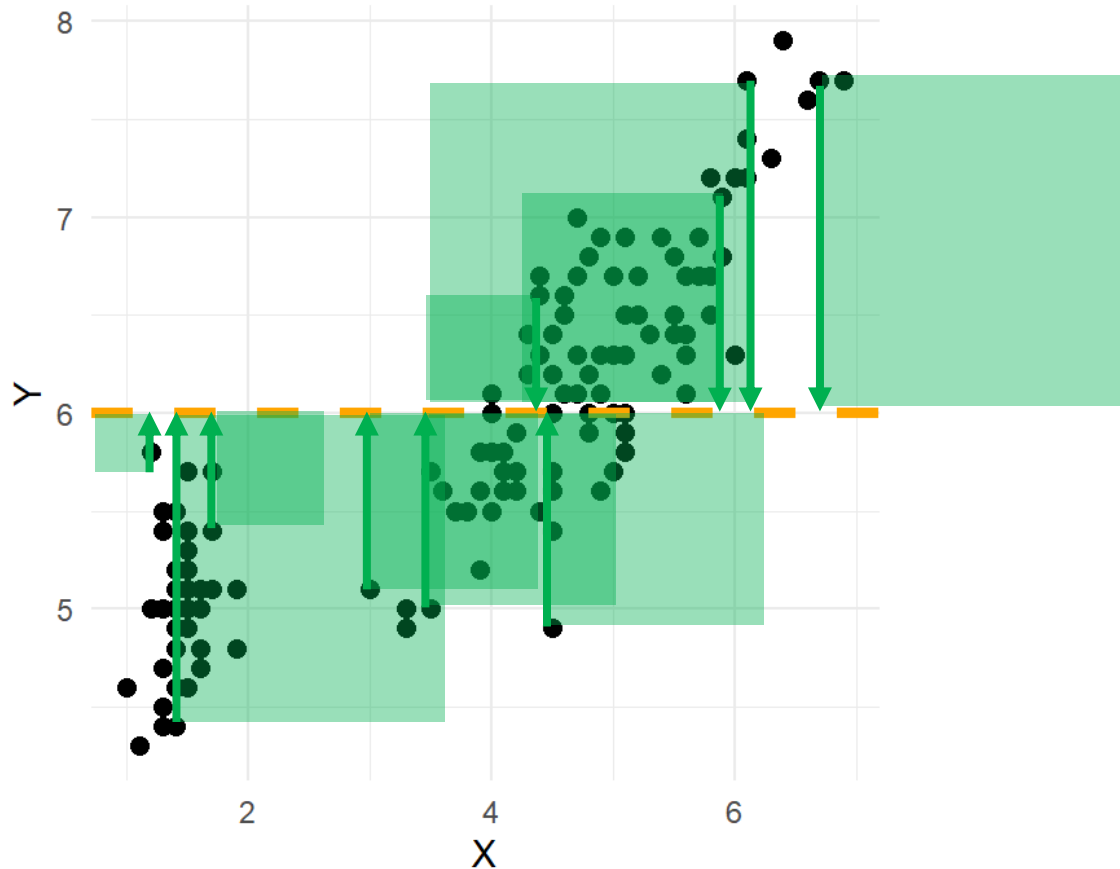


Earlier this week...

Modelos lineales → El **objetivo** es encontrar la línea que mejor defina los datos = *Encontrar los valores de  $b_0$  y  $b_1$  que nos permiten minimizar la suma de los cuadrados de los residuos*

$$y_i = b_0 + b_1 x_i + \varepsilon_i$$

$$SS_{res} = \sum_{i=1}^n (\varepsilon_i)^2$$



Earlier this week...

Modelos lineales → El **objetivo** es encontrar la línea que mejor defina los datos = ***Encontrar los valores de  $b_0$  y  $b_1$  que nos permiten minimizar la suma de los cuadrados de los residuos***

$SS_{res} = \sum_{i=1}^n (\epsilon_i)^2$  → Mide la varianza no explicada por el modelo

$SS_{total} = \sum_{i=1}^n (y_i - \bar{y})^2$  → Mide la varianza del modelo

$R^2 = 1 - \frac{SS_{res}}{SS_{total}}$  → proporción de varianza de la var. respuesta que está explicada por el modelo

- **Asunciones en Modelos Lineales: CON LOS RESIDUOS**

Para obtener residuos necesitamos tener un modelo

```
>Modelito<-lm(y~x)  
>resid(Modelito)  
>hist(resid(Modelito))
```

### T-test

- Comparar **dos** grupos  
H0= las medias de los dos grupos son iguales  
Ha= las medias de los dos grupos son distintas

>t.test( Y ~ X )

### ANOVA (One-way)

- Comparar **más de dos** grupos  
H0= La media de los grupos no difiere  
Ha= La media de los grupos difiere al menos entre dos grupos

>aov( Y ~ X ) %>%summary()

### ANOVA (Two-way)

- Comparar el efecto de la **combinación** de varios factores  
H0= La media de los grupos no difiere  
Ha= La media de los grupos difiere al menos entre dos grupos

>aov( Y ~ X1\* X2 ) %>%summary()

### Modelo Lineal Simple

- Determinar cómo se relacionan dos variables continuas  
H0= No existe relación entre variables →  $b_1 = 0$   
Ha= Existe relación entre variables →  $b_1 \neq 0$

>lm( y ~ x ) %>%summary()



# Estadística aplicada en R

## *Modelos Lineales:*

*Regresión simple*  
*Regresión múltiple sin interacción*  
*Regresión múltiple con interacción*

-Febrero 2021-

Carlota Solano  
Álvaro Arredondo

## 4.2. Modelo lineal múltiple con interacción

### 1. ¿Qué es?

Es un modelo que relaciona una variable dependiente con varias variables independientes.

$$y = a + mx$$

→ El objetivo es encontrar la línea que mejor defina los datos = **Encontrar los valores de *intercepto* y *pendiente* que nos permiten minimizar la suma de los cuadrados de los residuos**

### 2. ¿Cuándo se puede utilizar?

Cuando quieres evaluar la influencia que tienen múltiples variables independientes sobre una variable respuesta.

**Correlación no implica causalidad**

*¿Cómo difiere del modelo lineal simple?*

Incluimos más de una variable explicativa para estudiar cómo todas afectan a nuestra variable respuesta

$$y_i = b_0 + b_1x_{i1} + b_2x_{i2} + \varepsilon_i$$

Intercepto



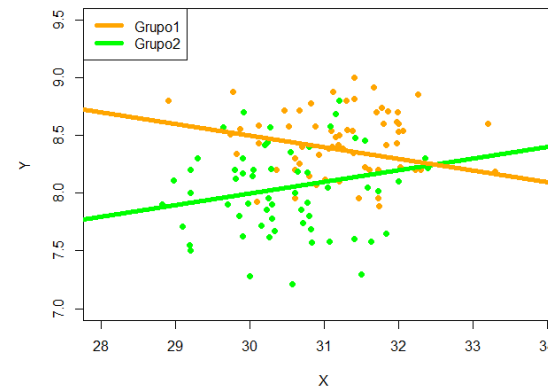
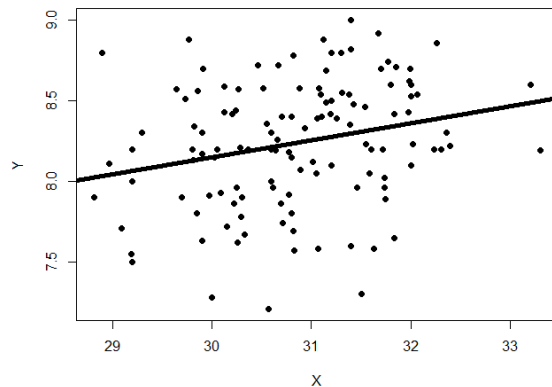
Efecto de x1



Efecto de x2



Residuo



## 4.2. Modelo lineal múltiple con interacción

### 3. ¿Qué tipo de datos se necesitan?

Variable respuesta (dep.; y) → Numérica continua

Variables explicativas (indep.; x) → Continuas y categóricas

← **Variables continuas** → **pendiente**: efecto del incremento de una unidad de x sobre y  
**Variables categóricas** → efecto del cambio de grupo de x sobre y

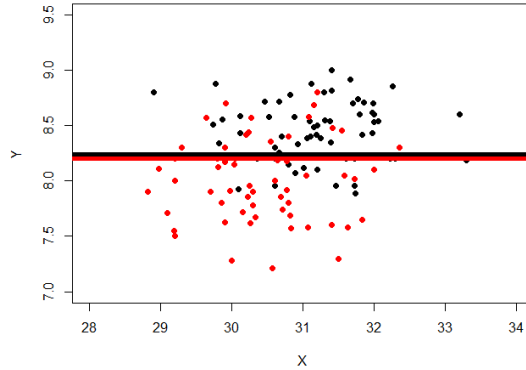
### 4. ¿Qué asunciones tiene?

- Variables explicativas no deben ser colineales → **No correlación** entre variables explicativas
- **Principio de parsimonia** (Navaja de Ockham)
- **Relación lineal** entre variables respuesta y explicativas
- Distribución **normal** de los residuos del modelo.
- Igualdad de **varianza** de los residuos en torno a la línea de la regresión.
- **Independencia** de las observaciones (i.e. de los datos).
- ¡Ojo con los **outliers**!

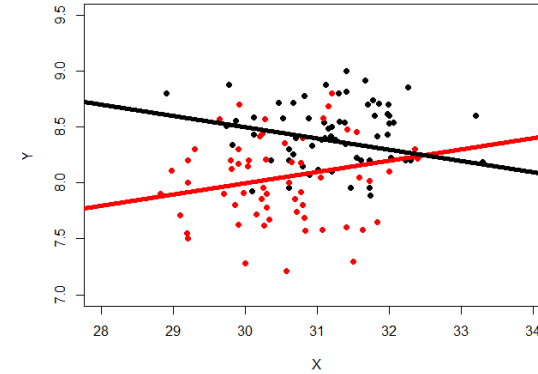
## 4.2. Modelo lineal múltiple con interacción

### 5. ¿Cuál es la hipótesis?

H0: Las vars. explicativas no afectan a la var. respuesta



Ha: Las vars. explicativas afectan a la var. respuesta



### 6. ¿Cómo se corre en R?



```
> aguacate<-lm (y ~ x1 * x2)  
> summary(aguacate)
```



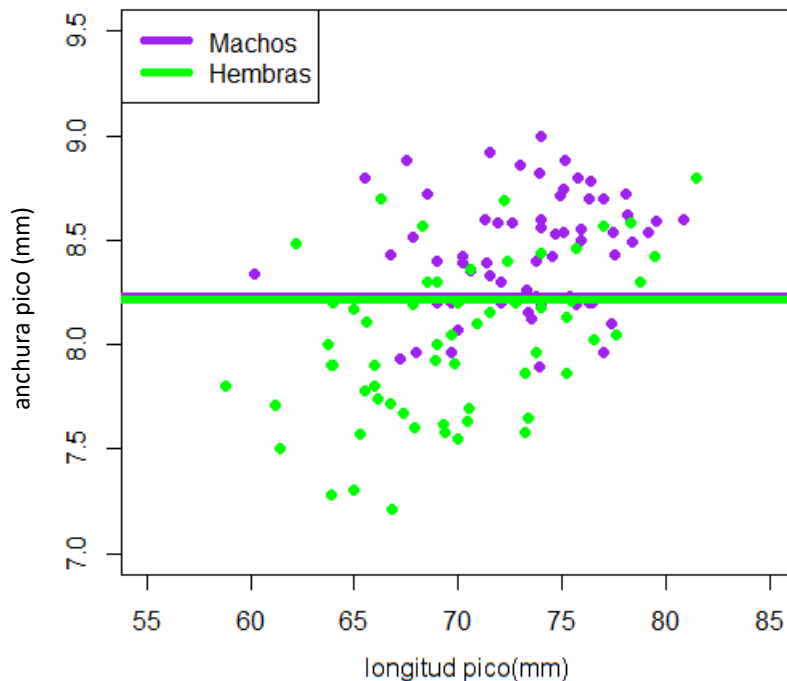
## 4.2. Modelo lineal múltiple con interacción

### 7. ¿Cómo se interpreta el resultado de R?

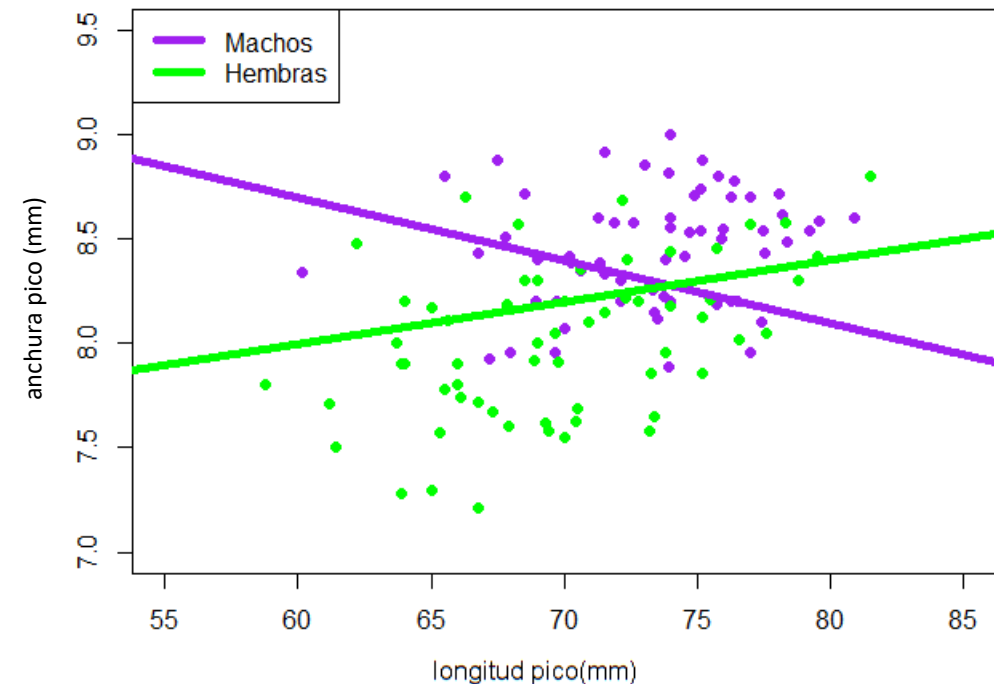
E.g. Queremos estudiar la morfología del pico de los arrendajos azules, concretamente si cuanto más ancho es el pico, su longitud también es mayor. Además, queremos ver si esta relación morfológica difiere entre machos y hembras.



H0: La anchura y la longitud del pico no están relacionado ni en hembras ni en machos



Ha: La anchura y la longitud del pico están relacionadas, y esta relación difiere entre machos y hembras



## 4.2. Modelo lineal múltiple con interacción

### 7. ¿Cómo se interpreta el resultado de R?

E.g. Queremos estudiar la morfología del pico de los arrendajos azules, concretamente si cuanto más ancho es el pico, su longitud también es mayor. Además, queremos ver si esta relación morfológica difiere entre machos y hembras.



**Grupos: Female & Male (Hembras y machos)**

```
> lm(BJ$BillDepth~BJ$BillLength*BJ$KnownSex)%>%summary()
```

Call:

```
lm(formula = BJ$BillDepth ~ BJ$BillLength * BJ$KnownSex)
```

Residuals:

Min	1Q	Median	3Q	Max
-0.67412	-0.19634	0.00585	0.23708	0.65073

Coefficients:

	Estimate	Std. Error	t value	Pr(> t )	
(Intercept)	3.54664	0.98028	3.618	0.000437	***
BJ\$BillLength	0.18465	0.04050	4.559	1.25e-05	***
BJ\$KnownSexM	3.38540	1.37280	2.466	0.015086	*
BJ\$BillLength:BJ\$KnownSexM	-0.12525	0.05534	-2.263	0.025438	*

---

signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 0.3003 on 119 degrees of freedom

Multiple R-squared: 0.4221, Adjusted R-squared: 0.4076

F-statistic: 28.98 on 3 and 119 DF, p-value: 3.876e-14

## 4.2. Modelo lineal múltiple con interacción

### 7. ¿Cómo se interpreta el resultado de R?

E.g. Queremos estudiar la morfología del pico de los arrendajos azules, concretamente si cuanto más ancho es el pico, su longitud también es mayor. Además, queremos ver si esta relación morfológica difiere entre machos y hembras.



**Grupos: Female & Male (Hembras y machos)**

```
> lm(BJ$BillDepth~BJ$BillLength*BJ$KnownSex)%>%summary()
```

Call:

```
lm(formula = BJ$BillDepth ~ BJ$BillLength * BJ$KnownSex)
```

Residuals:

Min	1Q	Median	3Q	Max
-0.67412	-0.19634	0.00585	0.23708	0.65073

Coefficients:

	Estimate	Std. Error	t value	Pr(> t )	
(Intercept)	3.54664	0.98028	3.618	0.000437	***
BJ\$BillLength	0.18465	0.04050	4.559	1.25e-05	***
BJ\$KnownSexM	3.38540	1.37280	2.466	0.015086	*
BJ\$BillLength:BJ\$KnownSexM	-0.12525	0.05534	-2.263	0.025438	*

---

signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 0.3003 on 119 degrees of freedom

Multiple R-squared: 0.4221, Adjusted R-squared: 0.4076

F-statistic: 28.98 on 3 and 119 DF, p-value: 3.876e-14

**Estimación de coeficientes que definen las líneas de regresión**

- **Intercepto de grupo de referencia**: Valor de "y" cuando x=0

## 4.2. Modelo lineal múltiple con interacción

### 7. ¿Cómo se interpreta el resultado de R?

E.g. Queremos estudiar la morfología del pico de los arrendajos azules, concretamente si cuanto más ancho es el pico, su longitud también es mayor. Además, queremos ver si esta relación morfológica difiere entre machos y hembras.



**Grupos:** Female & Male (Hembras y machos)

```
> lm(BJ$BillDepth~BJ$BillLength*BJ$KnownSex)%>%summary()
```

Call:

```
lm(formula = BJ$BillDepth ~ BJ$BillLength * BJ$KnownSex)
```

Residuals:

Min	1Q	Median	3Q	Max
-0.67412	-0.19634	0.00585	0.23708	0.65073

Coefficients:

	Estimate	Std. Error	t value	Pr(> t )	
(Intercept)	3.54664	0.98028	3.618	0.000437	***
BJ\$BillLength	0.18465	0.04050	4.559	1.25e-05	***
BJ\$KnownSexM	3.38540	1.37280	2.466	0.015086	*
BJ\$BillLength:BJ\$KnownSexM	-0.12525	0.05534	-2.263	0.025438	*

---

signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 0.3003 on 119 degrees of freedom

Multiple R-squared: 0.4221, Adjusted R-squared: 0.4076

F-statistic: 28.98 on 3 and 119 DF, p-value: 3.876e-14

**Estimación de coeficientes que definen las líneas de regresión**

- **Intercepto de grupo de referencia:** Valor de “y” cuando x=0

- **Var. Explicativa continua = pendiente de grupo de referencia:** efecto del incremento de una unidad de “x” sobre “y” para el grupo de referencia

## 4.2. Modelo lineal múltiple con interacción

### 7. ¿Cómo se interpreta el resultado de R?

E.g. Queremos estudiar la morfología del pico de los arrendajos azules, concretamente si cuanto más ancho es el pico, su longitud también es mayor. Además, queremos ver si esta relación morfológica difiere entre machos y hembras.



**Grupos:** Female & Male (Hembras y machos)

```
> lm(BJ$BillDepth~BJ$BillLength*BJ$KnownSex)%>%summary()
```

Call:

```
lm(formula = BJ$BillDepth ~ BJ$BillLength * BJ$KnownSex)
```

Residuals:

Min	1Q	Median	3Q	Max
-0.67412	-0.19634	0.00585	0.23708	0.65073

Coefficients:

	Estimate	Std. Error	t value	Pr(> t )
(Intercept)	3.54664	0.98028	3.618	0.000437 ***
BJ\$BillLength	0.18465	0.04050	4.559	1.25e-05 ***
BJ\$KnownSexM	3.38540	1.37280	2.466	0.015086 *
BJ\$BillLength:BJ\$KnownSexM	-0.12525	0.05534	-2.263	0.025438 *

---  
signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 0.3003 on 119 degrees of freedom  
Multiple R-squared: 0.4221, Adjusted R-squared: 0.4076  
F-statistic: 28.98 on 3 and 119 DF, p-value: 3.876e-14

**Estimación de coeficientes que definen las líneas de regresión**

- **Intercepto de grupo de referencia:** Valor de “y” cuando x=0
- **Var. Explicativa continua** = pendiente de grupo de referencia: efecto del incremento de una unidad de “x” sobre “y” para el grupo de referencia
- **Var. Explicativa categórica** = intercepto del segundo grupo: efecto del cambio de grupo respecto al grupo de referencia



## 4.2. Modelo lineal múltiple con interacción

### 7. ¿Cómo se interpreta el resultado de R?

E.g. Queremos estudiar la morfología del pico de los arrendajos azules, concretamente si cuanto más ancho es el pico, su longitud también es mayor. Además, queremos ver si esta relación morfológica difiere entre machos y hembras.



**Grupos:** Female & Male (Hembras y machos)

```
> lm(BJ$BillDepth~BJ$BillLength*BJ$KnownSex)%>%summary()
```

Call:

```
lm(formula = BJ$BillDepth ~ BJ$BillLength * BJ$KnownSex)
```

Residuals:

Min	1Q	Median	3Q	Max
-0.67412	-0.19634	0.00585	0.23708	0.65073

Coefficients:

	Estimate	Std. Error	t value	Pr(> t )	
(Intercept)	3.54664	0.98028	3.618	0.000437	***
BJ\$BillLength	0.18465	0.04050	4.559	1.25e-05	***
BJ\$KnownSexM	3.38540	1.37280	2.466	0.015086	*
BJ\$BillLength:BJ\$KnownSexM	-0.12525	0.05534	-2.263	0.025438	*

---

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

Residual standard error: 0.3003 on 119 degrees of freedom

Multiple R-squared: 0.4221, Adjusted R-squared: 0.4076

F-statistic: 28.98 on 3 and 119 DF, p-value: 3.876e-14

**Estimación de coeficientes que definen las líneas de regresión**

- **Intercepto de grupo de referencia:** Valor de “y” cuando x=0

- **Var. Explicativa continua** = pendiente de grupo de referencia: efecto del incremento de una unidad de “x” sobre “y” para el grupo de referencia

- **Var. Explicativa categórica** = intercepto del segundo grupo: efecto del cambio de grupo respecto al grupo de referencia

- **Interacción entre vars. explicativas**= pendiente del segundo grupo: efecto del incremento de una unidad de “x” sobre “y” para el segundo grupo

## 4.2. Modelo lineal múltiple con interacción

### 7. ¿Cómo se interpreta el resultado de R?

E.g. Queremos estudiar la morfología del pico de los arrendajos azules, concretamente si cuanto más ancho es el pico, su longitud también es mayor. Además, queremos ver si esta relación morfológica difiere entre machos y hembras.



Grupos: Female & Male (Hembras y machos)

```
> lm(BJ$BillDepth~BJ$BillLength*BJ$KnownSex)%>%summary()
```

Call:

```
lm(formula = BJ$BillDepth ~ BJ$BillLength * BJ$KnownSex)
```

Residuals:

Min	1Q	Median	3Q	Max
-0.67412	-0.19634	0.00585	0.23708	0.65073

Coefficients:

	Estimate	Std. Error	t value	Pr(> t )	
(Intercept)	3.54664	0.98028	3.618	0.000437	***
BJ\$BillLength	0.18465	0.04050	4.559	1.25e-05	***
BJ\$KnownSexM	3.38540	1.37280	2.466	0.015086	*
BJ\$BillLength:BJ\$KnownSexM	-0.12525	0.05534	-2.263	0.025438	*

---

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

Residual standard error: 0.3003 on 119 degrees of freedom  
Multiple R-squared: 0.4221, Adjusted R-squared: 0.4076  
F-statistic: 28.98 on 3 and 119 DF, p-value: 3.876e-14

- Las **hembras** tienen una anchura de pico de 3.54 mm cuando tienen un pico de longitud de 0 mm
- El pico de las **hembras** incrementa 0.18 mm anchura al incrementar 1 mm de longitud
- Los **machos** tienen una anchura de pico de 3.54+3.38 mm cuando tienen un pico de longitud de 0 mm
- El pico de los **machos** incrementa en 0.18-0.12 mm de anchura al incrementar 1 mm de longitud

### Ecuaciones

Anchura pico hembras=  $3.54 + 0.18 \cdot \text{Long. pico}$   
Anchura pico machos=  $(3.54+3.38) + (0.18-0.12) \cdot \text{Long. pico}$

## 4.2. Modelo lineal múltiple con interacción

### 7. ¿Cómo se interpreta el resultado de R?

E.g. Queremos estudiar la morfología del pico de los arrendajos azules, concretamente si cuanto más ancho es el pico, su longitud también es mayor. Además, queremos ver si esta relación morfológica difiere entre machos y hembras.



**Grupos: Female & Male (Hembras y machos)**

```
> lm(BJ$BillDepth~BJ$BillLength*BJ$KnownSex)%>%summary()
```

Call:

```
lm(formula = BJ$BillDepth ~ BJ$BillLength * BJ$KnownSex)
```

Residuals:

Min	1Q	Median	3Q	Max
-0.67412	-0.19634	0.00585	0.23708	0.65073

Coefficients:

	Estimate	Std. Error	t value	Pr(> t )	
(Intercept)	3.54664	0.98028	3.618	0.000437	***
BJ\$BillLength	0.18465	0.04050	4.559	1.25e-05	***
BJ\$KnownSexM	3.38540	1.37280	2.466	0.015086	*
BJ\$BillLength:BJ\$KnownSexM	-0.12525	0.05534	-2.263	0.025438	*

---  
Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 0.3003 on 119 degrees of freedom  
Multiple R-squared: 0.4221, Adjusted R-squared: 0.4076  
F-statistic: 28.98 on 3 and 119 DF, p-value: 3.876e-14

P-valor:

Significancia de los valores en el grupo de ref: Valor distinto de cero

Significancia de los valores en otros grupos: Valor distinto del de grupo de referencia

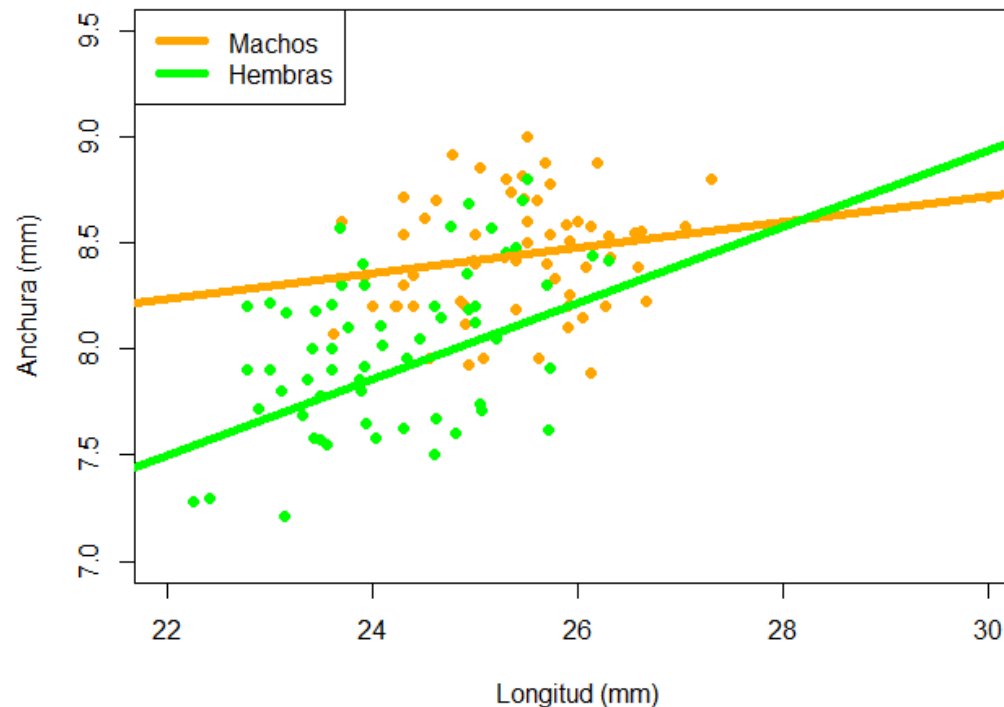
El intercepto de las líneas de regresión es distinto

La pendiente de las líneas de regresión es distinta == La relación entre las variables estudiadas difiere entre grupos



## 4.2. Modelo lineal múltiple con interacción

### 8. ¿Cómo se puede representar?



```
>plot(BJ$BillDepth[BJ$KnownSex=="M"]~BJ$BillLength[BJ$KnownSex=="M"],  
      xlim=c(22,30),ylim=c(7,9.5),col="orange",pch=19,  
      xlab="Longitud(mm)",ylab="Anchura (mm)")  
>points(BJ$BillDepth[BJ$KnownSex=="F"]~BJ$BillLength[BJ$KnownSex=="F"],col="green",pch=19)  
  
>abline(a=3.54,b=0.18,col="green",lwd=5)  
>abline(a=3.54+3.38,b=0.18-0.12,col="orange",lwd=5)  
  
>legend("topleft",legend = c("Machos","Hembras"), col=c("orange","green"),lwd=5)
```

## 4.1. Ejercicios de Modelos Lineales

Ejercicio: 4. Ejer\_LMs (segunda parte)

Ejercicio: 5. Ejer\_LMConInteracción