

Regresion

2024-08-30

Parte 1

```
M = read.csv("documents/Estatura-peso_HyM.csv")
head(M)
```

```
##   Estatura  Peso Sexo
## 1    1.61 72.21   H
## 2    1.61 65.71   H
## 3    1.70 75.08   H
## 4    1.65 68.55   H
## 5    1.72 70.77   H
## 6    1.63 77.18   H
```

Matriz de Correlacion

```
MM = subset(M,M$Sexo=="M")
MH = subset(M,M$Sexo=="H")
M1=data.frame(MH$Estatura,MH$Peso,MM$Estatura,MM$Peso)

cor(M1)
```

```
##           MH.Estatura    MH.Peso  MM.Estatura    MM.Peso
## MH.Estatura 1.0000000000 0.846834792 0.0005540612 0.04724872
## MH.Peso     0.8468347920 1.0000000000 0.0035132246 0.02154907
## MM.Estatura 0.0005540612 0.003513225 1.0000000000 0.52449621
## MM.Peso     0.0472487231 0.021549075 0.5244962115 1.000000000
```

Interpretar

Obtén medidas

```
n=4 #número de variables
d=matrix(NA,ncol=7,nrow=n)
for(i in 1:n){
  d[i,]<-c(as.numeric(summary(M1[,i])),sd(M1[,i]))
}
m=as.data.frame(d)

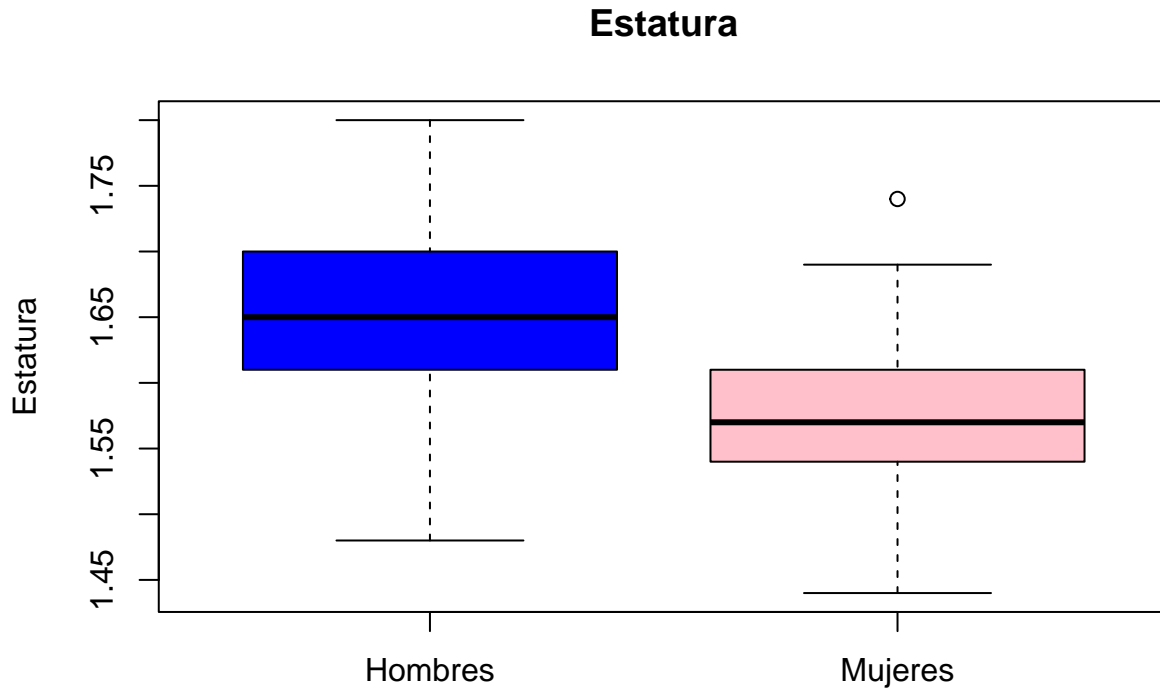
row.names(m)=c("H-Estatura", "H-Peso", "M-Estatura", "M-Peso")
names(m)=c("Minimo", "Q1", "Mediana", "Media", "Q3", "Máximo", "Desv Est")
m
```

```
##           Minimo      Q1 Mediana      Media      Q3 Máximo  Desv Est
## H-Estatura   1.48  1.6100   1.650  1.653727  1.7000   1.80 0.06173088
## H-Peso       56.43 68.2575  72.975 72.857682 77.5225  90.49 6.90035408
```

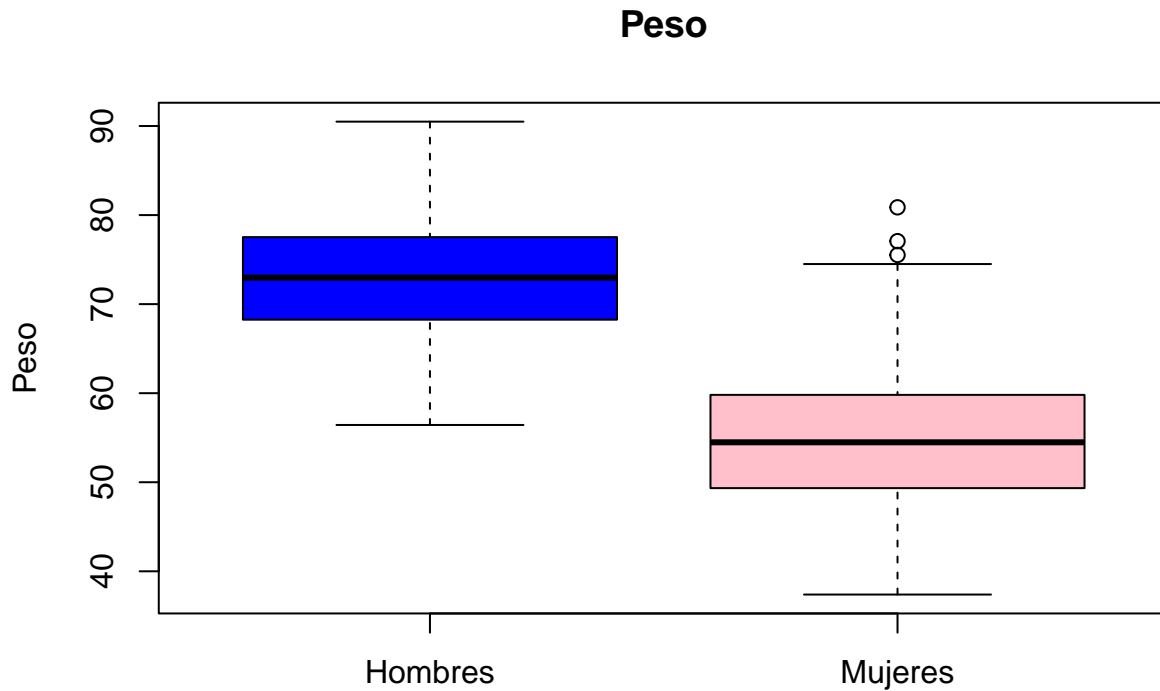
```
## M-Estatura  1.44  1.5400  1.570  1.572955  1.6100  1.74  0.05036758
## M-Peso      37.39 49.3550  54.485 55.083409 59.7950  80.87  7.79278074
```

Boxplot

```
boxplot(M$Estatura~M$Sexo, ylab="Estatura", xlab="", col=c("blue","pink"), names=c("Hombres", "Mujeres"))
```



```
boxplot(M$Peso~M$Sexo, ylab="Peso", xlab="", names=c("Hombres", "Mujeres"), col=c("blue","pink"), main="Peso")
```



Encuentra la ecuación de regresión de mejor ajuste

```
modelo1H = lm(Peso ~ Estatura, data = MH)
modelo1H
```

```
##
## Call:
## lm(formula = Peso ~ Estatura, data = MH)
##
## Coefficients:
## (Intercept)      Estatura
##      -83.68       94.66
```

```
modelo1M = lm(Peso ~ Estatura, data = MM)
modelo1M
```

```
##
## Call:
## lm(formula = Peso ~ Estatura, data = MM)
##
## Coefficients:
## (Intercept)      Estatura
##      -72.56       81.15
```

```
modelo2 = lm(Peso ~ Estatura+Sexo, M)
modelo2
```

```
##
## Call:
## lm(formula = Peso ~ Estatura + Sexo, data = M)
##
## Coefficients:
## (Intercept)      Estatura      SexoM
##      -74.75       89.26      -10.56
```

Hipotesis:

$$H_0 : \beta_1 = 0 \quad H_1 : \beta_1 \neq 0$$

```
summary(modelo1H)
```

```
##
## Call:
## lm(formula = Peso ~ Estatura, data = MH)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -8.3881 -2.6073 -0.0665  2.4421 11.1883
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)  -83.685      6.663  -12.56  <2e-16 ***
## Estatura      94.660      4.027   23.51  <2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 3.678 on 218 degrees of freedom
## Multiple R-squared:  0.7171, Adjusted R-squared:  0.7158
```

```
## F-statistic: 552.7 on 1 and 218 DF, p-value: < 2.2e-16
```

```
summary(modelo1M)
```

```
##
## Call:
## lm(formula = Peso ~ Estatura, data = MM)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -21.3256  -4.1942   0.4004   4.2724  17.9114
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)  -72.560     14.041  -5.168 5.34e-07 ***
## Estatura      81.149       8.922   9.096 < 2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 6.65 on 218 degrees of freedom
## Multiple R-squared:  0.2751, Adjusted R-squared:  0.2718
## F-statistic: 82.73 on 1 and 218 DF, p-value: < 2.2e-16
```

```
summary(modelo2)
```

```
##
## Call:
## lm(formula = Peso ~ Estatura + Sexo, data = M)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -21.9505  -3.2491   0.0489   3.2880  17.1243
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept) -74.7546     7.5555  -9.894 <2e-16 ***
## Estatura     89.2604     4.5635  19.560 <2e-16 ***
## SexoM       -10.5645     0.6317 -16.724 <2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 5.381 on 437 degrees of freedom
## Multiple R-squared:  0.7837, Adjusted R-squared:  0.7827
## F-statistic: 791.5 on 2 and 437 DF, p-value: < 2.2e-16
```

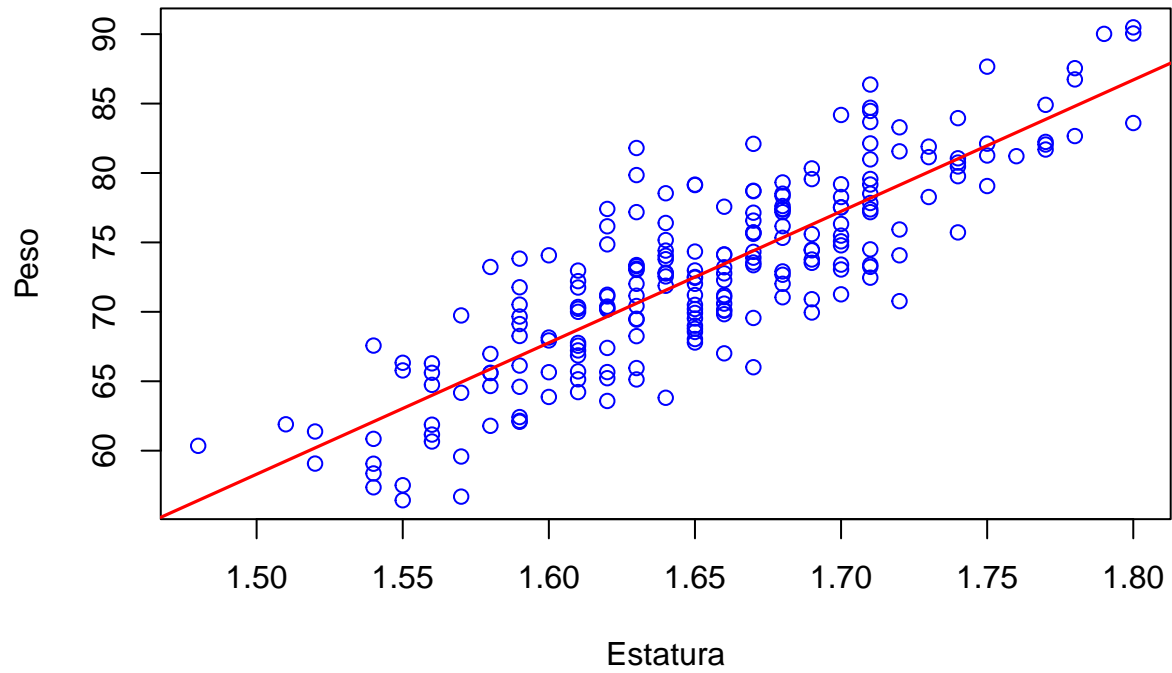
A 0.05 si es significativo si los modelos:

Hombres:

Dibuja el diagrama de dispersión de los datos y la recta de mejor ajuste.

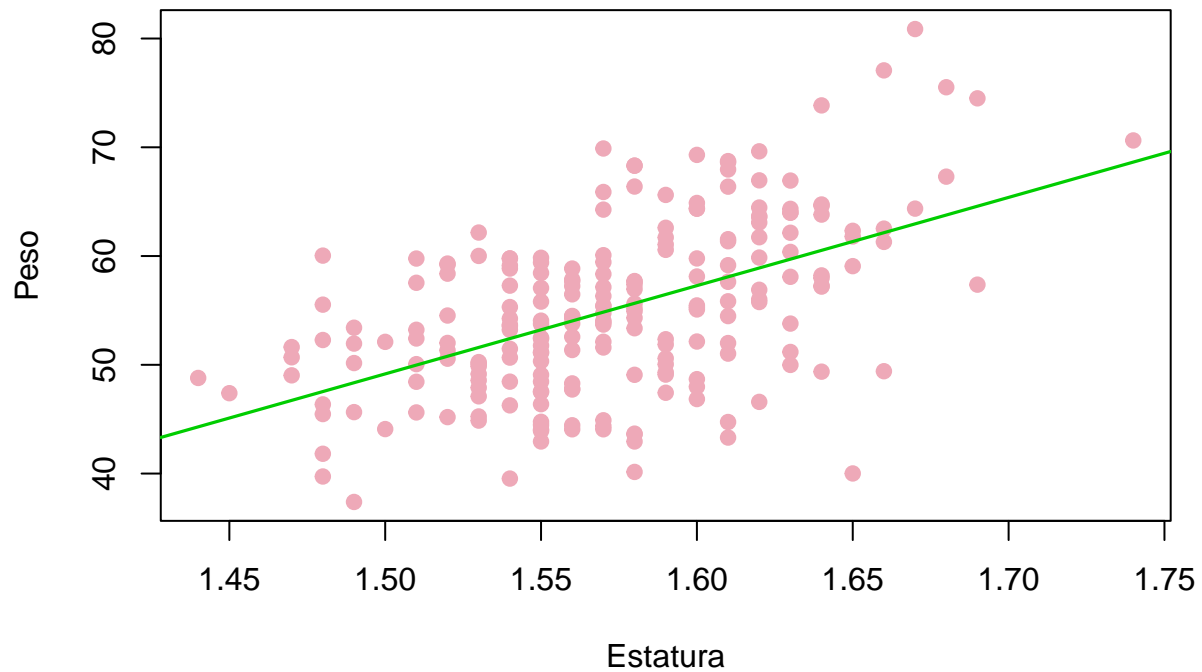
```
plot(MH$Estatura, MH$Peso, col="blue", main = "Estatura vs Peso \n Hombres", ylab="Peso", xlab="Estatura")
abline(modelo1H, col="red", lwd=1.6)
```

Estatura vs Peso Hombres



```
plot(MM$Estatura, MM$Peso, col="pink2", pch=19, main = "Estatura vs Peso \n Mujeres", ylab="Peso", xlab=
abline(modelo1M, col="green3", lwd=1.6)
```

Estatura vs Peso Mujeres



```

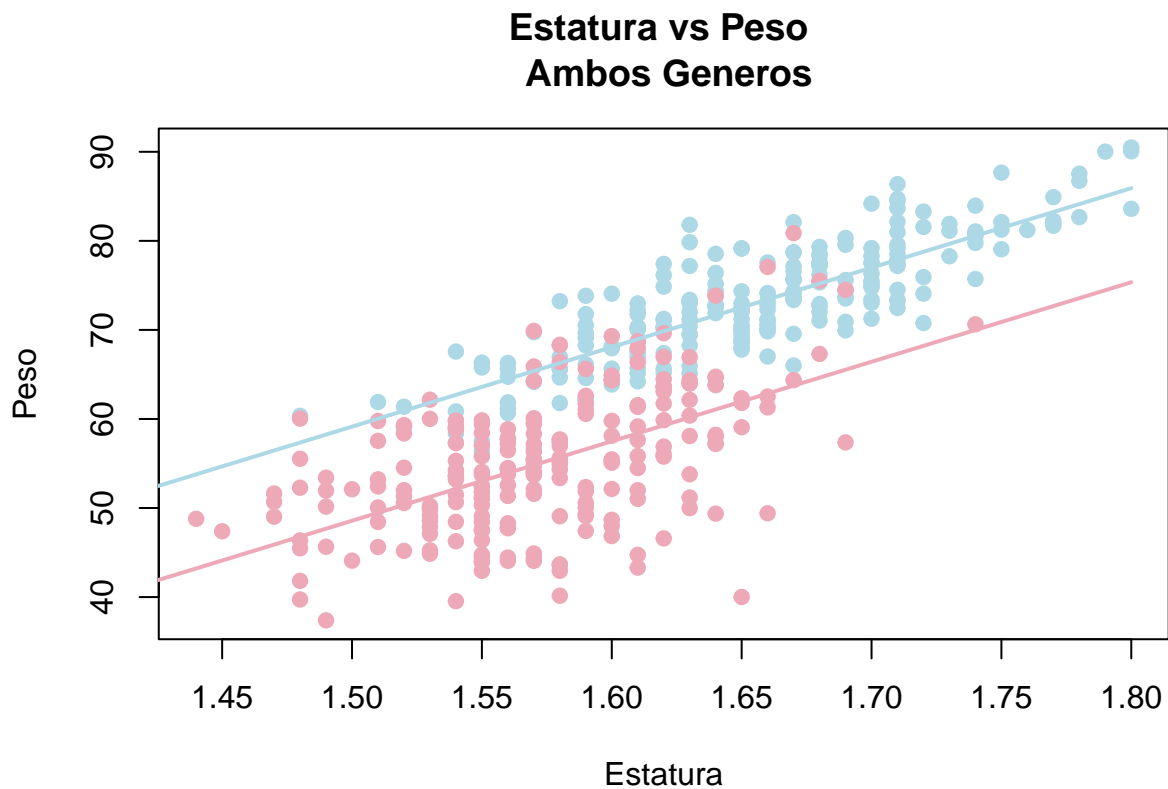
b0 = modelo2$coefficients[1]
b1 = modelo2$coefficients[2]
b2 = modelo2$coefficients[3]

ym = function(x){b0+b2+b1*x}
yh = function(x){b0+b1*x}

colores = c("lightblue", "pink2")

plot(M$Estatura, M$Peso, col=colores[factor(M$Sexo)], pch=19, main = "Estatura vs Peso \n Ambos Generos")
x = seq(1.40, 1.80, 0.01)
lines(x, ym(x), col="pink2", lwd=2)
lines(x, yh(x), col="lightblue", lwd=2)

```



Conclusion

En ambos generos se ve una fuerte relacion entre peso y estatura, para los hombres la estatura ayuda a predecir el peso, viendo un 71.7% de su variabilidad, en mujeres la estatura tambien es un factor significativo, pero solo vemos un 27.5% de la variabilidad en el peso, lo que significa que otros factores influyen.

Con esto podemos ver que en ambos la estatura y el sexo es un facto predictor para el peso, con la estatura siendo una mayor influencia.

Parte 2

```

modelo3 = lm(Peso ~ Estatura*Sexo, M)
modelo3

```

```
##
## Call:
## lm(formula = Peso ~ Estatura * Sexo, data = M)
##
## Coefficients:
##      (Intercept)      Estatura      SexoM  Estatura:SexoM
##          -83.68         94.66         11.12         -13.51
```

```
A = summary(modelo3)
```

```
A
```

```
##
## Call:
## lm(formula = Peso ~ Estatura * Sexo, data = M)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -21.3256  -3.1107   0.0204   3.2691  17.9114
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)   -83.685     9.735  -8.597  <2e-16 ***
## Estatura       94.660     5.882  16.092  <2e-16 ***
## SexoM          11.124    14.950   0.744   0.457
## Estatura:SexoM -13.511     9.305  -1.452   0.147
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 5.374 on 436 degrees of freedom
## Multiple R-squared:  0.7847, Adjusted R-squared:  0.7832
## F-statistic: 529.7 on 3 and 436 DF,  p-value: < 2.2e-16
```

```
b0_A=A$coefficients[1]
```

```
b1_A=A$coefficients[2]
```

```
b2_A=A$coefficients[3]
```

```
b3_A=A$coefficients[4]
```

```
Ym=function(x){b0_A+b2_A+(b1_A+b3_A)*x}
```

```
Yh=function(x){b0_A+b1_A*x}
```

```
colores=c("#66BD63", "#FDAE61")
```

```
plot(M$Estatura, M$Peso, col=colores[factor(M$Sexo)], pch=19, ylab="Peso", xlab="Estatura", main="Relación en")
```

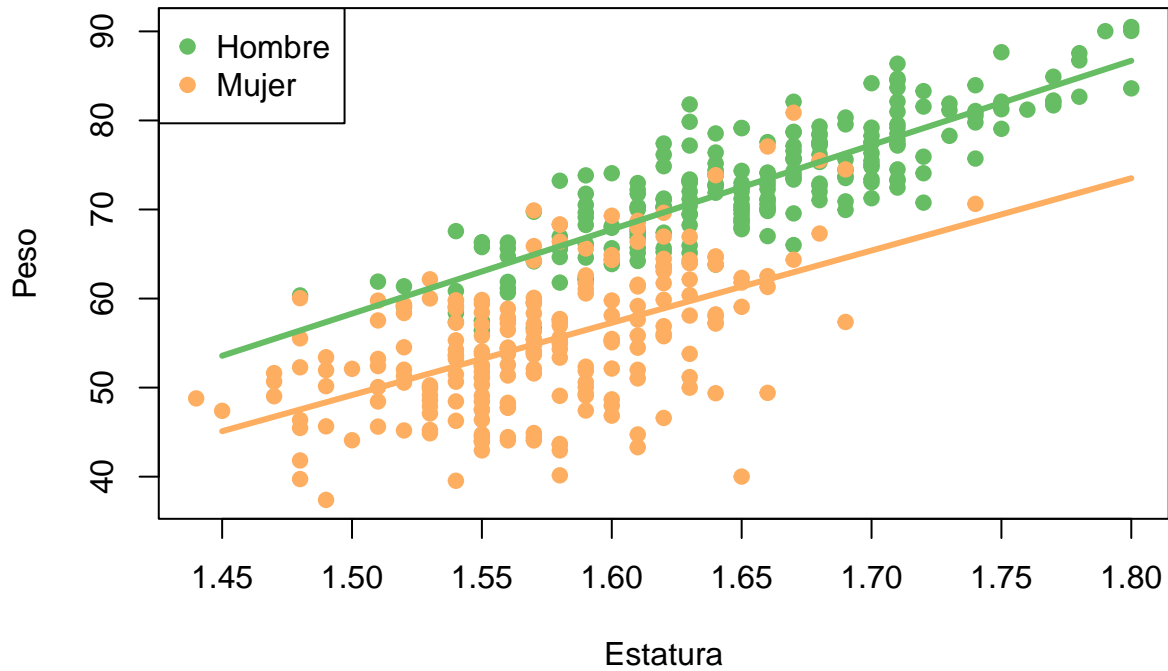
```
x=seq(1.45, 1.80, 0.01)
```

```
lines(x, Ym(x), col="#FDAE61", lwd=3)
```

```
lines(x, Yh(x), col="#66BD63", lwd=3)
```

```
legend("topleft", legend=c("Hombre", "Mujer"), pch=19, col=c("#66BD63", "#FDAE61"))
```

Relación entre estatura y peso



Conclusion

En el modelo 4 vemos la interacción entre estatura con sexo, esto nos proporciona una mayor flexibilidad, también vemos como el intercepto β_0 sigue representando el peso promedio de una mujer con una estatura de 0 y también vemos como β_1 es el cambio en peso por altura para las mujeres pero vemos como en los hombres β_3 ajusta la pendiente y esto permite que la relación entre estatura y peso sea diferente para cada sexo.

Viendo estos resultados el modelo 4 es el mejor porque el modelo es más flexible, esto permite que la pendiente y el intercepto varíen según el sexo, esto nos puede ayudar a tener una mejor representación de los datos. También el modelo 3 es uno de los mejores pero esto depende de cómo nuestras variables afectan, el modelo 3 es más simple y más fácil de interpretar, la diferencia que tiene el modelo 3 del modelo 4 es que este modelo reconoce la diferencia en peso entre sexos, pero asume que la relación entre estatura y peso es la misma para ambos.