

Seminario:

Modelos de regresión lineal

Resumen

Introducir el concepto de comparación de modelos.

- Ajuste de una recta a varios grupos.
- Ajuste de una recta a cada grupo.
- Ajuste de rectas paralelas.
- Comparación de modelos.

Contracepción y disminución del número de hijos: ¿Hasta qué punto existen diferencias entre continentes?

Esquema general

Disponemos de la edad y de la concentración de un biomarcador en hombres y mujeres voluntarias y queremos caracterizar la relación entre la edad y la variación de este metabolito en cuatro tratamientos (A,B,C)

1. Representar la media de biomarcador por sexo y tratamiento.
2. Representar la variación del biomarcador en función de la edad, sexo y tratamiento.
3. Ajustar un modelo de regresión, teniendo en cuenta los distintos factores.
4. Evaluar qué modelo es el más conveniente.

```
dd <- dt %>% group_by(sex,group) %>%
```

```
  summarise(m=mean(biomarker),  
            low=t.test(biomarker)$conf.int[1],  
            upper=t.test(biomarker)$conf.int[2])
```

```
ggplot(dd,aes(x=group,y=m))+  
  geom_errorbar(aes(x=group,ymin=low,ymax=upper,width=.1),color='blue',size=1)+  
  geom_point(size=4,shape=21,aes(fill=sex))+  
  geom_line(aes(x=group,y=m,group=sex,color=sex),size=1)+  
  ylab('95% CI of Mean')
```

```
dt <- readRDS('SeminararioReg (1).R')
```

```
m <- lm(biomarker~group*sex,data=dt)  
summary(m)
```

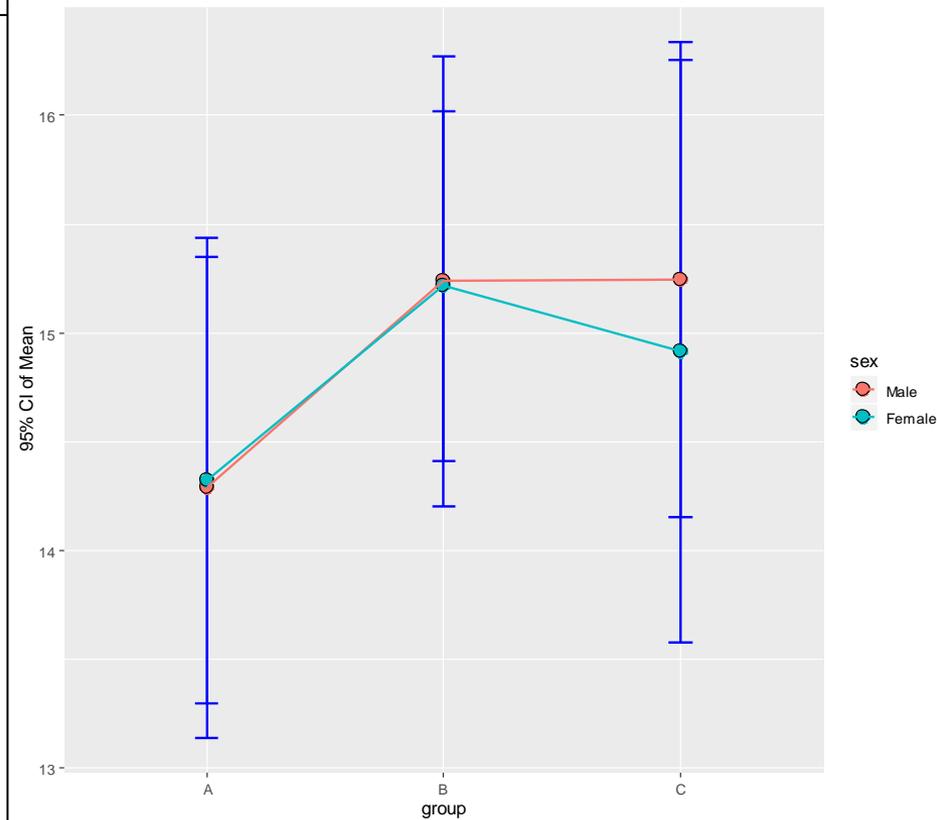
```
Call:  
lm(formula = biomarker ~ group * sex, data = dt)
```

Residuals:

	Min	1Q	Median	3Q	Max
	-4.3235	-1.5145	0.2201	1.2218	5.2824

Coefficients:

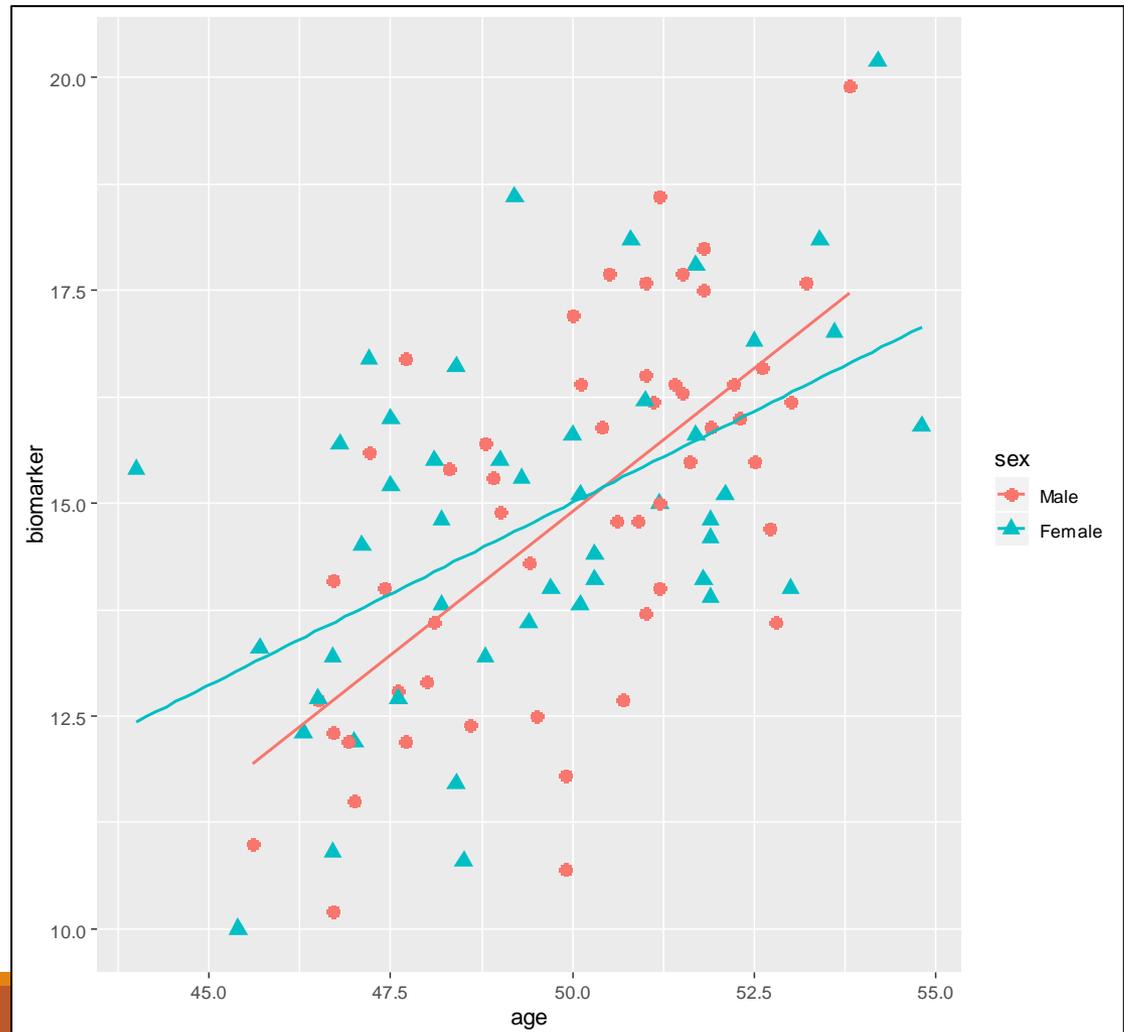
	Estimate	Std. Error	t value	Pr(> t)
(Intercept)	14.28947	0.49311	28.979	<2e-16 ***
groupB	0.94803	0.72931	1.300	0.197
groupC	0.95497	0.70698	1.351	0.180
sexFemale	0.03406	0.71757	0.047	0.962
groupB:sexFemale	-0.05617	1.07659	-0.052	0.958
groupC:sexFemale	-0.36085	1.02144	-0.353	0.725



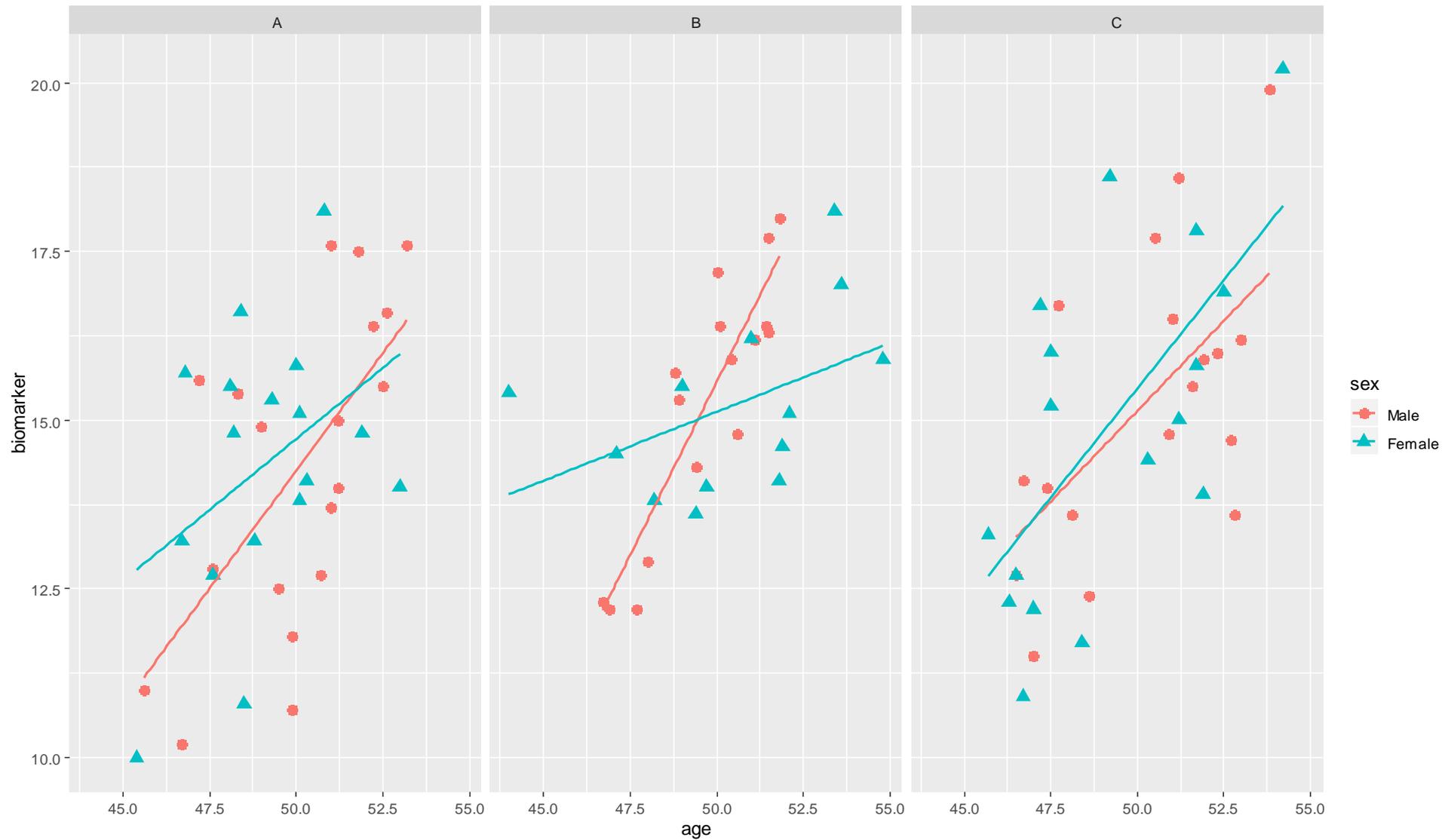
De acuerdo con este análisis, la media del biomarcador no varía con los tratamientos y es la misma en hombres y mujeres.

```
dt <- readRDS('SeminararioReg (1).R')
```

```
ggplot(dt, aes(x=age, y=biomarker, group=sex))+  
  geom_point(aes(color=sex, shape=sex), size=3)+  
  geom_smooth(aes(color=sex), method=lm, se=F)
```



```
ggplot(dt, aes(x=age, y=biomarker, group=sex))+  
  geom_point(aes(color=sex, shape=sex), size=3)+  
  geom_smooth(aes(color=sex), method=lm, se=F)+  
  facet_wrap(.~group)
```



```
m.1 <- lm(biomarker~sex*group*age,data=dt)
summary(m.1)
```

Call:

```
lm(formula = biomarker ~ sex * group * age, data = dt)
```

Residuals:

```
    Min       1Q   Median       3Q      Max
-3.4793 -1.0110 -0.1914  1.0901  3.6482
```

Coefficients:

	Estimate	Std. Error	t value	Pr(> t)	
(Intercept)	-20.6264	9.3195	-2.213	0.029467	*
sexFemale	14.3115	14.3045	1.000	0.319816	
groupB	-15.2413	16.0274	-0.951	0.344236	
groupC	8.9901	12.6101	0.713	0.477775	
age	0.6975	0.1860	3.750	0.000316	***
sexFemale:groupB	26.4807	21.0980	1.255	0.212754	
sexFemale:groupC	-19.4667	18.4685	-1.054	0.294745	
sexFemale:age	-0.2768	0.2889	-0.958	0.340577	
groupB:age	0.3313	0.3216	1.030	0.305790	
groupC:age	-0.1621	0.2513	-0.645	0.520623	
sexFemale:groupB:age	-0.5480	0.4241	-1.292	0.199682	
sexFemale:groupC:age	0.3866	0.3723	1.039	0.301869	

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

```
Residual standard error: 1.711 on 88 degrees of freedom
```

```
Multiple R-squared:  0.4298,    Adjusted R-squared:  0.3585
```

```
F-statistic:  6.03 on 11 and 88 DF,  p-value: 2.949e-07
```

El sexo y el grupo no son factores significativos. En este caso, solo la edad tiene una relación significativa con la concentración.

Ajustaremos un modelo solo con la edad.

```
m.0 <- lm(biomarker~age,data=dt)
summary(m.0)

anova(m.0,m.1)
```

```
Call:
lm(formula = biomarker ~ age, data = dt)

Residuals:
    Min       1Q   Median       3Q      Max
-4.226 -1.149 -0.149  1.128  4.048

Coefficients:
            Estimate Std. Error t value Pr(>|t|)
(Intercept) -11.69124    3.77843  -3.094  0.00257 **
age           0.53340    0.07587   7.031 2.77e-10 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 1.751 on 98 degrees of freedom
Multiple R-squared:  0.3353,    Adjusted R-squared:  0.3285
F-statistic: 49.43 on 1 and 98 DF,  p-value: 2.774e-10
```

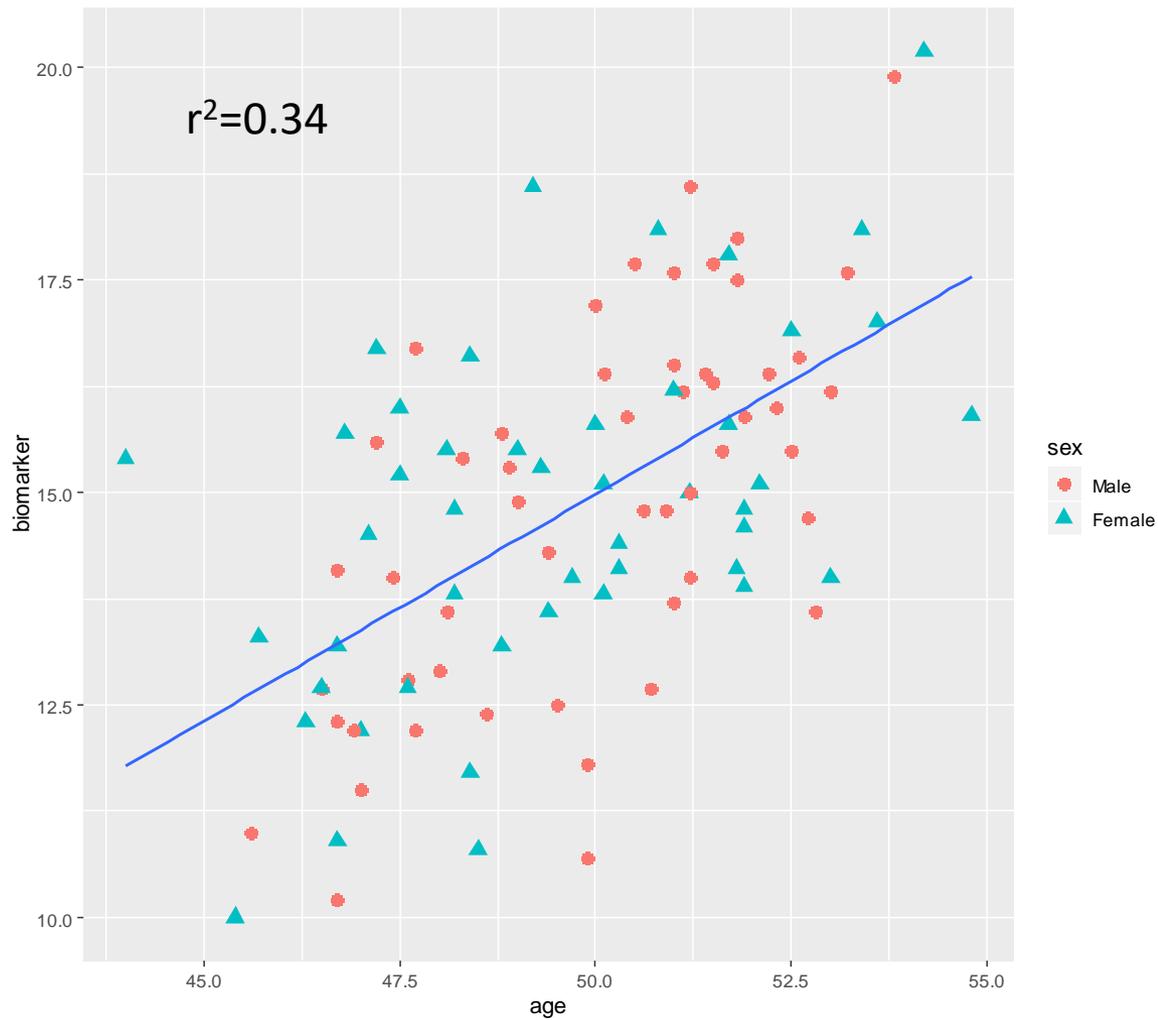
Analysis of Variance Table

```
Model 1: biomarker ~ age
Model 2: biomarker ~ sex * group * age
  Res.Df  RSS Df Sum of Sq    F Pr(>F)
1     98 300.43
2     88 257.73 10    42.702 1.458 0.1689
```

El modelo que considera solo la edad es suficiente.

$$y = -11.69 + 0.53 \times age$$

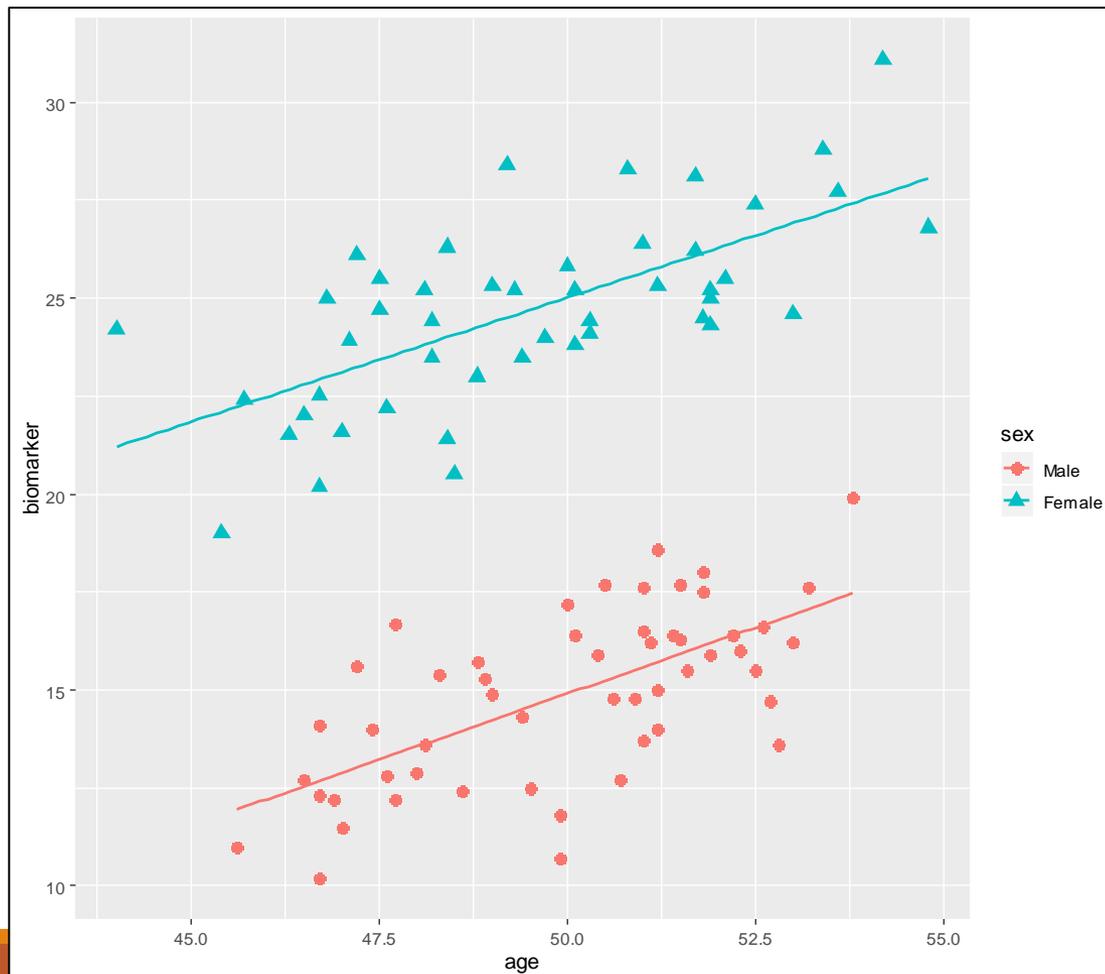
La comparación de modelos muestra que la introducción de más factores y sus interacciones no reduce significativamente la RSS.



El ajuste no es muy bueno. Pueden existir otros factores que mejoren la predicción.

Caso 2:

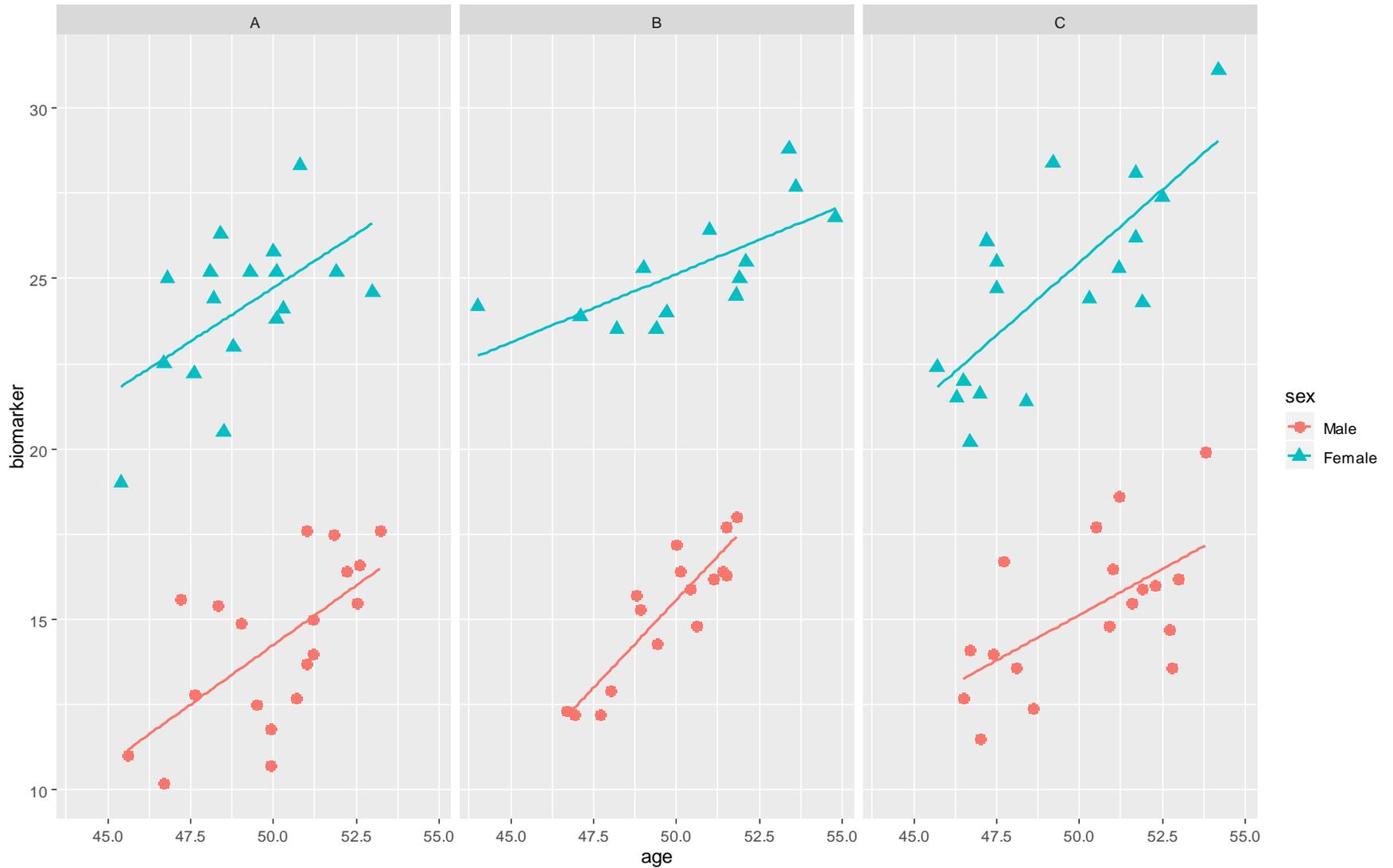
SeminarioReg (2).R



En este caso, parece claro que la concentración del biomarcador aumenta con la edad y que los hombres tienen valores inferiores a las mujeres.

La pendiente de ambas rectas parece la misma (no interacción).

Parece que hay cierta variación según el tratamiento. Deberíamos verificar si existe interacción.



```
m.1 <- lm(biomarker~sex*group*age,data=dt)
summary(m.1)
```

```
Call:
lm(formula = biomarker ~ sex * group * age, data = dt)
```

```
Residuals:
    Min       1Q   Median       3Q      Max
-3.4793 -1.0068 -0.2275  1.0829  3.6136
```

```
Coefficients:
                Estimate Std. Error t value Pr(>|t|)
(Intercept)    -20.62637     9.32189  -2.213  0.029508 *
sexFemale       13.81223    14.30819   0.965  0.337021
groupB        -15.24126    16.03162  -0.951  0.344360
groupC         8.99012    12.61337   0.713  0.477889
age             0.69751     0.18606   3.749  0.000317 ***
sexFemale:groupB  27.15413    21.10345   1.287  0.201568
sexFemale:groupC -19.38807    18.47330  -1.050  0.296813
sexFemale:age    -0.06665     0.28897  -0.231  0.818135
groupB:age       0.33128     0.32169   1.030  0.305915
groupC:age      -0.16210     0.25139  -0.645  0.520732
sexFemale:groupB:age -0.56151    0.42423  -1.324  0.189066
sexFemale:groupC:age  0.38486     0.37235   1.034  0.304160
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```
Residual standard error: 1.712 on 88 degrees of freedom
Multiple R-squared:  0.9111,    Adjusted R-squared:  0.8999
F-statistic: 81.95 on 11 and 88 DF,  p-value: < 2.2e-16
```

En el modelo de interacciones completas, solo la edad aparece como significativa.

```
m.2 <- lm(biomarker~sex*group+age,data=dt)
summary(m.2)
```

Call:

```
lm(formula = biomarker ~ sex * group + age, data = dt)
```

Residuals:

Min	1Q	Median	3Q	Max
-3.4858	-1.1031	-0.1238	1.0743	3.6241

Coefficients:

	Estimate	Std. Error	t value	Pr(> t)	
(Intercept)	-18.56797	3.85822	-4.813	5.74e-06	***
sexFemale	10.50160	0.58090	18.078	< 2e-16	***
groupB	1.19935	0.58599	2.047	0.0435	*
groupC	0.85805	0.56744	1.512	0.1339	
age	0.65639	0.07667	8.561	2.25e-13	***
sexFemale:groupB	-0.93999	0.87471	-1.075	0.2853	
sexFemale:groupC	-0.31008	0.81968	-0.378	0.7061	

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

Residual standard error: 1.725 on 93 degrees of freedom

Multiple R-squared: 0.9046, Adjusted R-squared: 0.8984

F-statistic: 146.9 on 6 and 93 DF, p-value: < 2.2e-16

En este modelo, consideramos la posible interacción entre sexo y grupo, más el efecto de la edad.

Los efectos principales son significativos, pero no las interacciones.

```
m.3 <- lm(biomarker~sex+group+age,data=dt)
summary(m.3)
```

```
Call:
lm(formula = biomarker ~ sex + group + age, data = dt)
```

```
Residuals:
    Min       1Q   Median       3Q      Max
-3.6642 -1.1358 -0.1716  1.1196  3.6568
```

```
Coefficients:
              Estimate Std. Error t value Pr(>|t|)
(Intercept) -17.7016     3.7561  -4.713 8.35e-06 ***
sexFemale    10.1148     0.3463  29.207 < 2e-16 ***
groupB        0.7748     0.4298   1.803  0.0746 .
groupC        0.7141     0.4078   1.751  0.0831 .
age           0.6426     0.0751   8.557 1.99e-13 ***
```

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

```
Residual standard error: 1.717 on 95 degrees of freedom
Multiple R-squared:  0.9034,    Adjusted R-squared:  0.8993
F-statistic: 222 on 4 and 95 DF, p-value: < 2.2e-16
```

El grupo de tratamiento no es muy significativo. Podemos ajustar un modelo solo con sexo y edad.

```
ggplot(dt, aes(x=age, y=biomarker))+  
  geom_point(aes(color=group, shape=sex), size=3)+  
  geom_abline(slope=0.65, intercept=-17.67, color='red')+  
  geom_abline(slope=0.65, intercept=-17.67+10.12, color='turquoise')
```



```
m.4 <- lm(biomarker~sex+age,data=dt)
summary(m.4)
```

```
Call:
lm(formula = biomarker ~ sex + age, data = dt)
```

```
Residuals:
    Min       1Q   Median       3Q      Max
-4.1398 -1.1488 -0.1406  1.3175  3.9007
```

```
Coefficients:
            Estimate Std. Error t value Pr(>|t|)
(Intercept) -17.66966   3.79479  -4.656 1.02e-05 ***
sexFemale    10.11554   0.35031  28.876 < 2e-16 ***
age           0.65149   0.07576   8.600 1.40e-13 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```
Residual standard error: 1.738 on 97 degrees of freedom
Multiple R-squared:  0.899,    Adjusted R-squared:  0.8969
F-statistic: 431.6 on 2 and 97 DF,  p-value: < 2.2e-16
```

```
> anova(m.4,m.3,m.2,m.1)
Analysis of Variance Table
```

```
Model 1: biomarker ~ sex + age
Model 2: biomarker ~ sex + group + age
Model 3: biomarker ~ sex * group + age
Model 4: biomarker ~ sex * group * age
```

	Res.Df	RSS	Df	Sum of Sq	F	Pr(>F)
1	97	292.85				
2	95	280.17	2	12.6882	2.1651	0.1208
3	93	276.68	2	3.4881	0.5952	0.5537
4	88	257.86	5	18.8181	1.2844	0.2779

La comparación de los distintos modelos muestra que el modelo con sexo y edad es suficiente.

Veremos si deberíamos considerar la interacción entre estos dos factores.

```
m.4.1 <- lm(biomarker~sex*age,data=dt)
summary(m.4.1)
```

```
Call:
lm(formula = biomarker ~ sex * age, data = dt)

Residuals:
    Min       1Q   Median       3Q      Max
-4.1377 -1.1576 -0.1292  1.3000  3.8957

Coefficients:
            Estimate Std. Error t value Pr(>|t|)
(Intercept) -18.80186    5.73258   -3.280  0.00145 **
sexFemale    12.13345    7.63718    1.589  0.11541
age          0.67414    0.11457    5.884 5.84e-08 ***
sexFemale:age -0.04055    0.15330   -0.265  0.79196
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 1.746 on 96 degrees of freedom
Multiple R-squared:  0.8991,    Adjusted R-squared:  0.8959
F-statistic: 285 on 3 and 96 DF,  p-value: < 2.2e-16
```

```
> anova(m.4,m.4.1,m.3,m.2,m.1)
Analysis of Variance Table

Model 1: biomarker ~ sex + age
Model 2: biomarker ~ sex * age
Model 3: biomarker ~ sex + group + age
Model 4: biomarker ~ sex * group + age
Model 5: biomarker ~ sex * group * age
  Res.Df  RSS Df Sum of Sq    F Pr(>F)
1      97 292.85
2      96 292.64  1    0.2133 0.0728 0.78796
3      95 280.17  1   12.4749 4.2573 0.04203 *
4      93 276.68  2    3.4881 0.5952 0.55366
5      88 257.86  5   18.8181 1.2844 0.27785
```

No es necesario introducir la interacción entre sexo y edad.

Por lo tanto, nos quedamos con el modelo con sexo y edad, sin interacción.

```

Call:
lm(formula = biomarker ~ sex + age, data = dt)

Residuals:
    Min       1Q   Median       3Q      Max
-4.1398 -1.1488 -0.1406  1.3175  3.9007

Coefficients:
            Estimate Std. Error t value Pr(>|t|)
(Intercept) -17.66966    3.79479  -4.656 1.02e-05 ***
sexFemale    10.11554    0.35031  28.876 < 2e-16 ***
age          0.65149    0.07576   8.600 1.40e-13 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 1.738 on 97 degrees of freedom
Multiple R-squared:  0.899,    Adjusted R-squared:  0.8969
F-statistic: 431.6 on 2 and 97 DF,  p-value: < 2.2e-16

```

Hombres

$$y = -17.67 + 0.65 \times edad$$

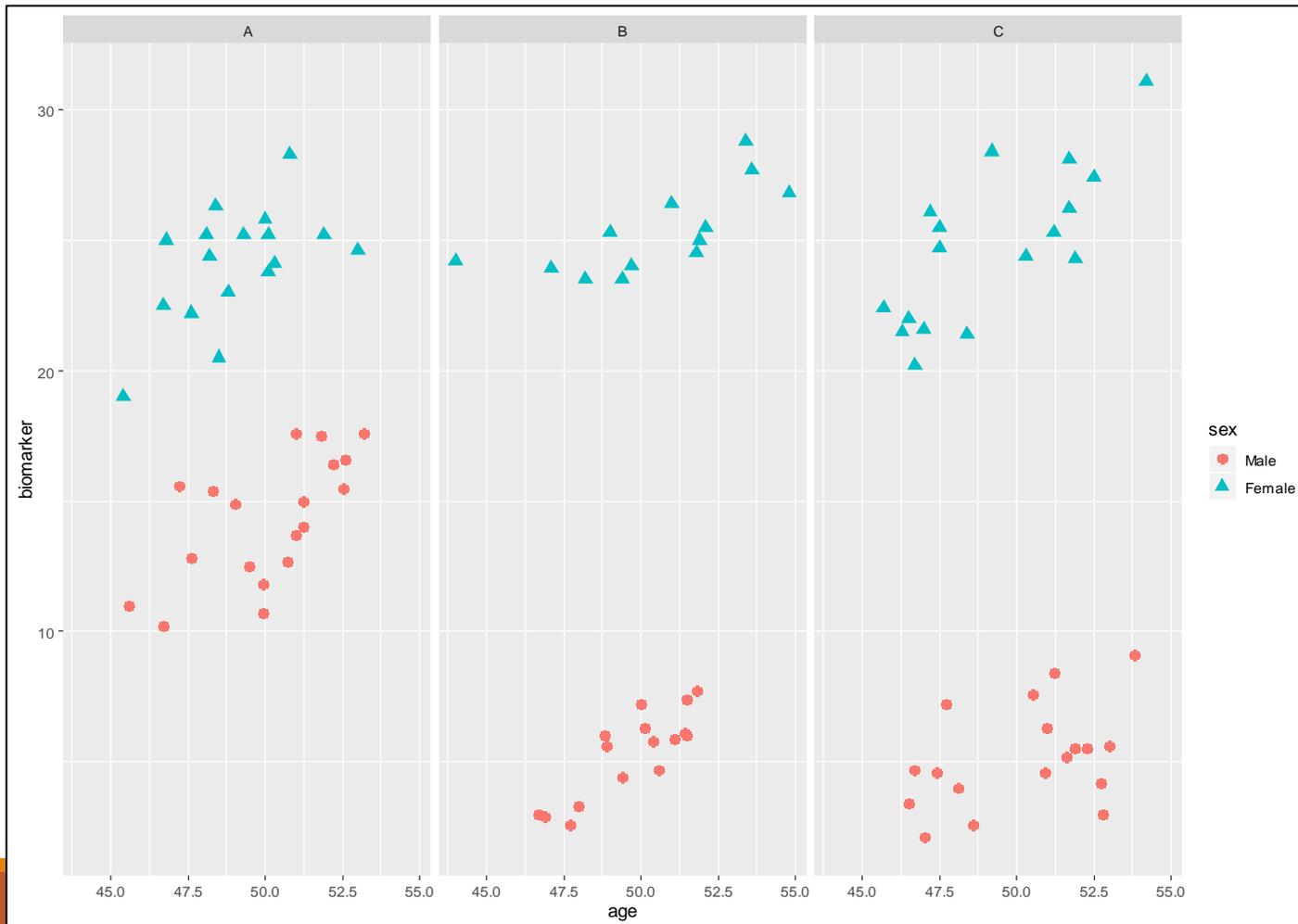
Mujeres

$$y = (-17.67 + 10.12) + 0.65 \times edad$$

El modelo final corresponde a rectas paralelas de pendiente 0.65 para hombre y mujeres, con un efecto constante de un aumento e 10.12 unidades en la concentración esperada del biomarcador en las mujeres respecto a los hombres a cualquier edad.

Caso 3:

SeminarioReg (3)



Analysis of Variance Table

Model 1: biomarker ~ sex + age
 Model 2: biomarker ~ sex * age
 Model 3: biomarker ~ sex + group + age
 Model 4: biomarker ~ sex * group + age
 Model 5: biomarker ~ sex * group * age

	Res. Df	RSS	Df	Sum of Sq	F	Pr(>F)	
1	97	1267.03					
2	96	1266.87	1	0.17	0.0564	0.8128	
3	95	791.27	1	475.60	161.6063	<2e-16	***
4	93	280.52	2	510.75	86.7742	<2e-16	***
5	88	258.98	5	21.53	1.4635	0.2099	

El mejor modelo incluye la interacción de grupo de tratamiento con el sexo y el efecto de la edad.

Call:
 lm(formula = biomarker ~ sex * group + age, data = dt)

Residuals:

Min	1Q	Median	3Q	Max
-3.7510	-1.1200	-0.2435	1.1074	3.6272

Coefficients:

	Estimate	Std. Error	t value	Pr(> t)	
(Intercept)	-15.6366	3.8849	-4.025	0.000116	***
sexFemale	10.4431	0.5849	17.854	< 2e-16	***
groupB	-8.7543	0.5900	-14.837	< 2e-16	***
groupC	-9.1777	0.5714	-16.063	< 2e-16	***
age	0.5978	0.0772	7.744	1.16e-11	***
sexFemale:groupB	9.0958	0.8808	10.327	< 2e-16	***
sexFemale:groupC	9.7309	0.8254	11.790	< 2e-16	***

```
Call:
lm(formula = biomarker ~ sex * group + age, data = dt)
```

```
Residuals:
```

```
      Min       1Q   Median       3Q      Max
-3.7510 -1.1200 -0.2435  1.1074  3.6272
```

```
Coefficients:
```

	Estimate	Std. Error	t value	Pr(> t)	
(Intercept)	-15.6366	3.8849	-4.025	0.000116	***
sexFemale	10.4431	0.5849	17.854	< 2e-16	***
groupB	-8.7543	0.5900	-14.837	< 2e-16	***
groupC	-9.1777	0.5714	-16.063	< 2e-16	***
age	0.5978	0.0772	7.744	1.16e-11	***
sexFemale:groupB	9.0958	0.8808	10.327	< 2e-16	***
sexFemale:groupC	9.7309	0.8254	11.790	< 2e-16	***

Mujeres con tratamiento B

$$y = (-15.64 + 10.44 - 8.75 + 9.09) + (0.60) \times age$$

Hombre con tratamiento B

$$y = (-15.64 - 8.75) + (0.60) \times age$$

Puntos importantes

El modelo lineal (lm) permite incluir tanto variables cuantitativas como cualitativas.

- Si los predictores son cualitativos (en general grupos de intervención, niveles de gravedad, sexo (Hombre/Mujer), etc.) estamos en un caso de ANOVA.
- Si los predictores son cuantitativos, estamos en un caso de regresión lineal.
- Si analizamos un diseño experimental con factores fijos y consideramos la posible influencia de la edad (los ejemplos que hemos discutido en esta presentación), nos encontramos en un modelo especial de ANOVA denominado análisis de la covarianza (ANCOVA). En este caso, la selección de modelos muchas veces nos lleva a comparar distintas rectas de regresión.

El procedimiento anova() nos permite comparar modelos y sirve de ayuda para seleccionar el modelo más adecuado.

Contracepción y disminución del número de hijos

Tasa reproductiva y contracepción

Se dispone de datos de la tasa de reproducción (número medio de hijos por mujer) y del porcentaje de mujeres que usan métodos contraceptivos en diversos países agrupados por grandes zonas geográficas (Asia, Latinoamérica, Próximo Oriente y Africa).

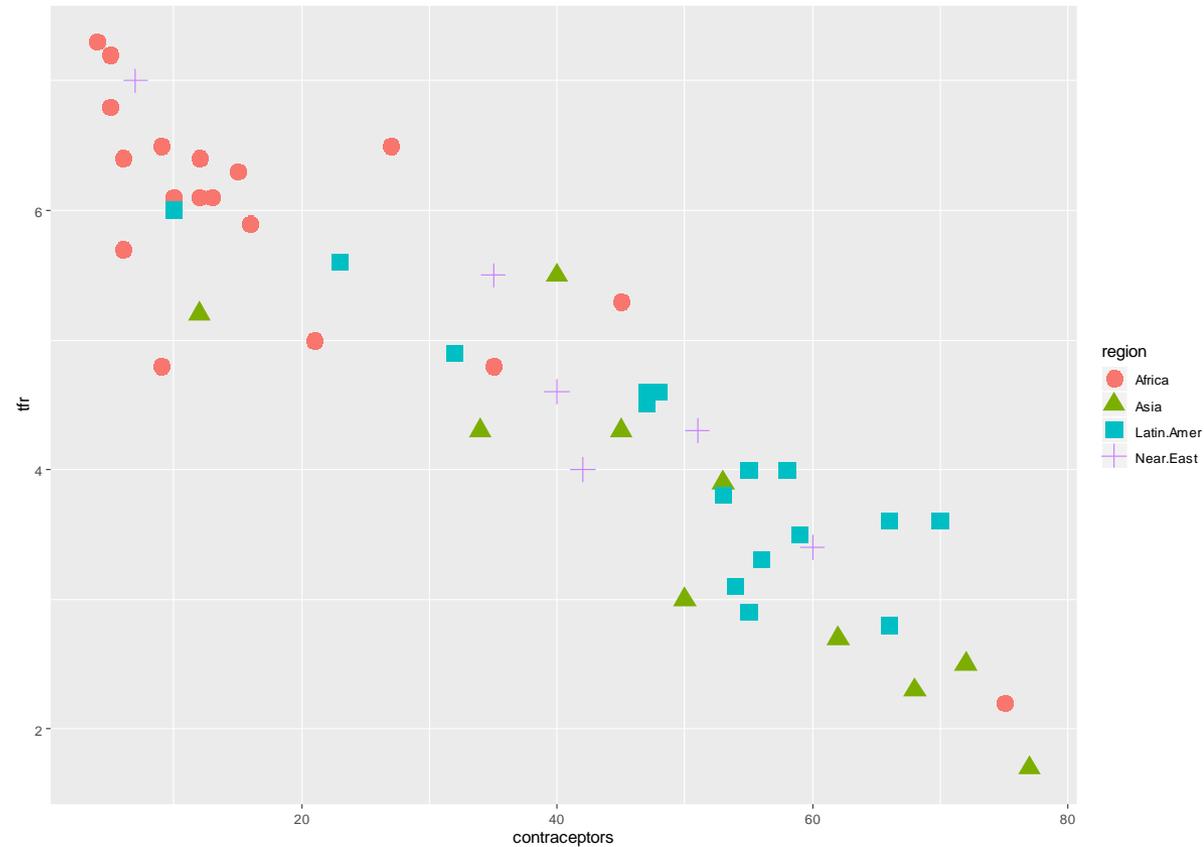
Estamos interesados en caracterizar la tendencia entre una aumento de la contracepción y la disminución de la tasa reproductiva y verificar si las tendencias son similares en las distintas áreas geográficas.

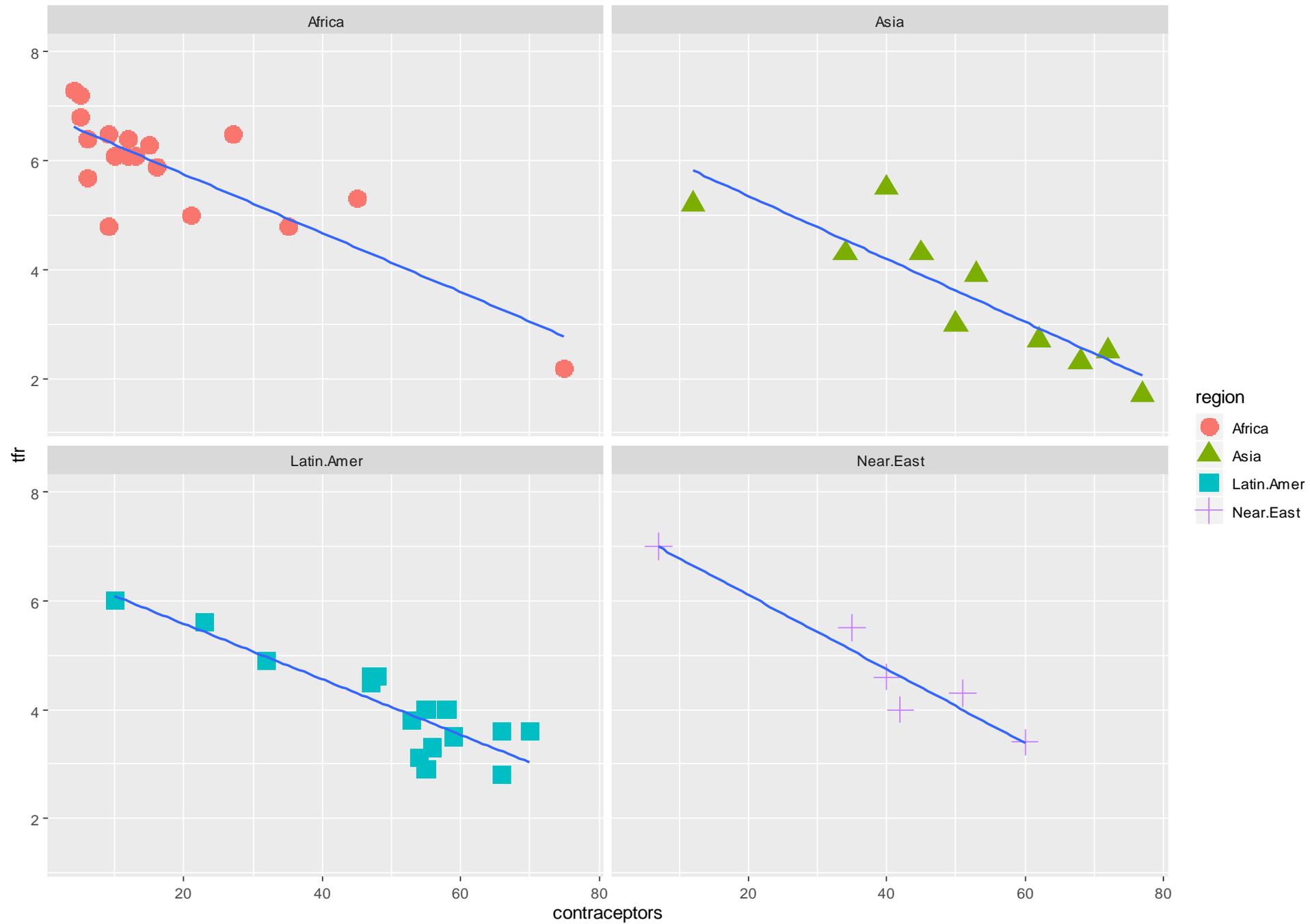
Los datos están en un data.frame denominado ***Robey*** de la librería ***car***.

```
library(car)
data(Robey)
Robey[1:20,]
```

	region	tfr	contraceptors
Botswana	Africa	4.8	35
Burundi	Africa	6.5	9
Cameroon	Africa	5.9	16
Ghana	Africa	6.1	13
Kenya	Africa	6.5	27
Liberia	Africa	6.4	6
Mali	Africa	6.8	5
Mauritius	Africa	2.2	75
Niger	Africa	7.3	4
Nigeria	Africa	5.7	6
Senegal	Africa	6.4	12
Sudan	Africa	4.8	9
Swaziland	Africa	5.0	21
Tanzania	Africa	6.1	10
Togo	Africa	6.1	12
Uganda	Africa	7.2	5
Zambia	Africa	6.3	15
Zimbabwe	Africa	5.3	45
Bangladesh	Asia	5.5	40
China	Asia	2.5	72

```
ggplot(Robey, aes(x=contraceptors,
                  y=tfr,
                  shape=region))+
  geom_point(aes(color=region), size=5)
```





```
> res.1 <- lm(tfr~contraceptors,Robey)
> summary(res.1)
```

```
Call:
lm(formula = tfr ~ contraceptors, data = Robey)
```

```
Residuals:
```

```
      Min       1Q   Median       3Q      Max
-1.5493 -0.3013  0.0254  0.3957  1.2021
```

```
Coefficients:
```

```
              Estimate Std. Error t value Pr(>|t|)
(Intercept)   6.875085   0.156860   43.83  <2e-16 ***
contraceptors -0.058416   0.003584  -16.30  <2e-16 ***
```

```
---
```

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

```
Residual standard error: 0.5745 on 48 degrees of freedom
Multiple R-squared:  0.847,    Adjusted R-squared:  0.8438
F-statistic: 265.7 on 1 and 48 DF,  p-value: < 2.2e-16
```

Un modelo para
todas las regiones

```
> anova(res.1)
```

```
Analysis of Variance Table
```

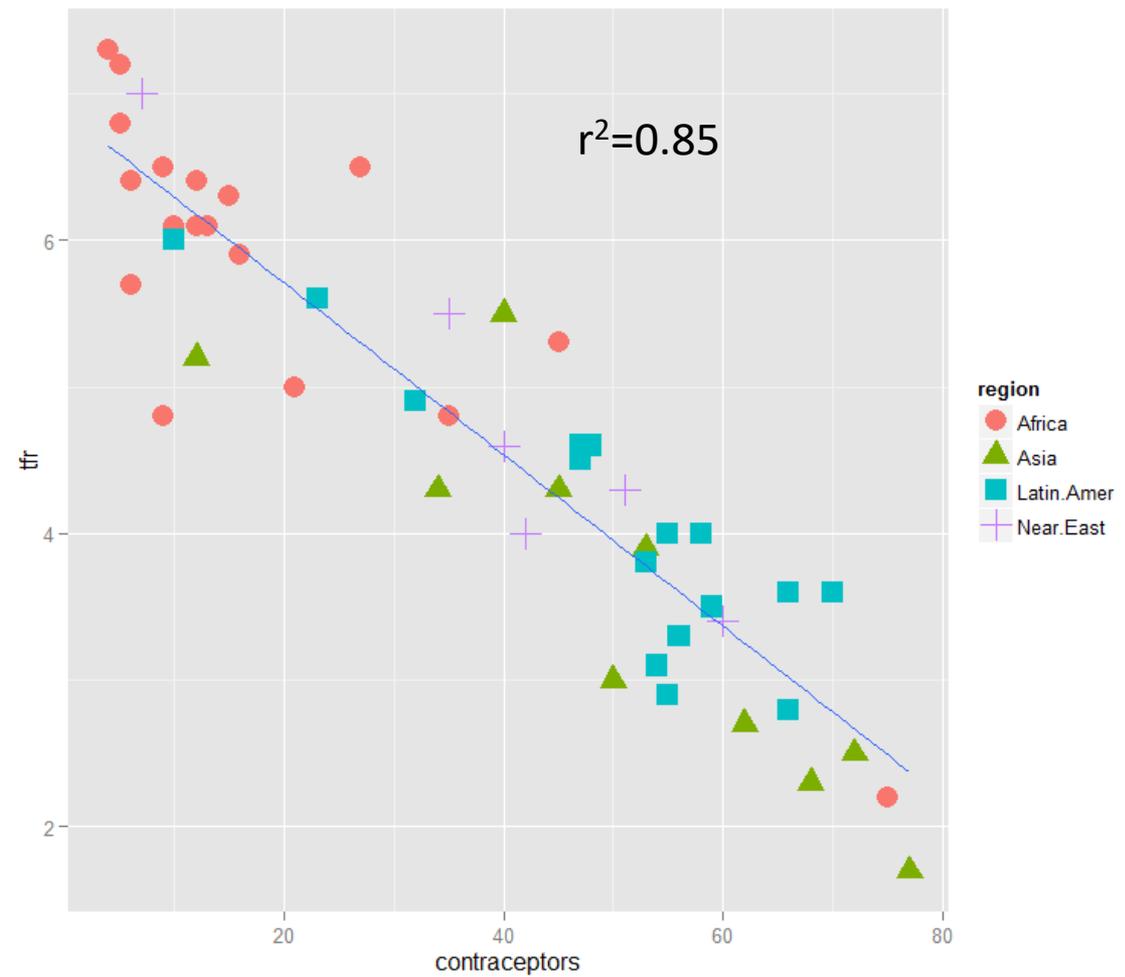
```
Response: tfr
```

```
              Df Sum Sq Mean Sq F value    Pr(>F)
contraceptors  1  87.672   87.672  265.67 < 2.2e-16 ***
Residuals     48  15.840    0.330
```

```
---
```

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

```
> ggplot(Robey, aes(x=contraceptors, y=tfr)) +  
+ geom_point(aes(colour=region, shape=region), size=5) +  
+ geom_smooth(method=lm, se=F)
```



```

> res.par <- lm(tfr~contraceptors+region,Robey)
> summary(res.par)

Call:
lm(formula = tfr ~ contraceptors + region, data = Robey)

Residuals:
    Min       1Q   Median       3Q      Max
-1.56044 -0.30085 -0.05744  0.39619  1.32998

Coefficients:
            Estimate Std. Error t value Pr(>|t|)
(Intercept)   6.86223    0.15674  43.782 < 2e-16 ***
contraceptors -0.05575    0.00466 -11.963 1.42e-15 ***
regionAsia    -0.46203    0.27012  -1.710  0.0941 .
regionLatin.Amer -0.02800    0.24338  -0.115  0.9089
regionNear.East  0.12148    0.28217   0.431  0.6689
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 0.561 on 45 degrees of freedom
Multiple R-squared:  0.8632,    Adjusted R-squared:  0.851
F-statistic: 70.97 on 4 and 45 DF,  p-value: < 2.2e-16

```

La misma
pendiente para
cada región

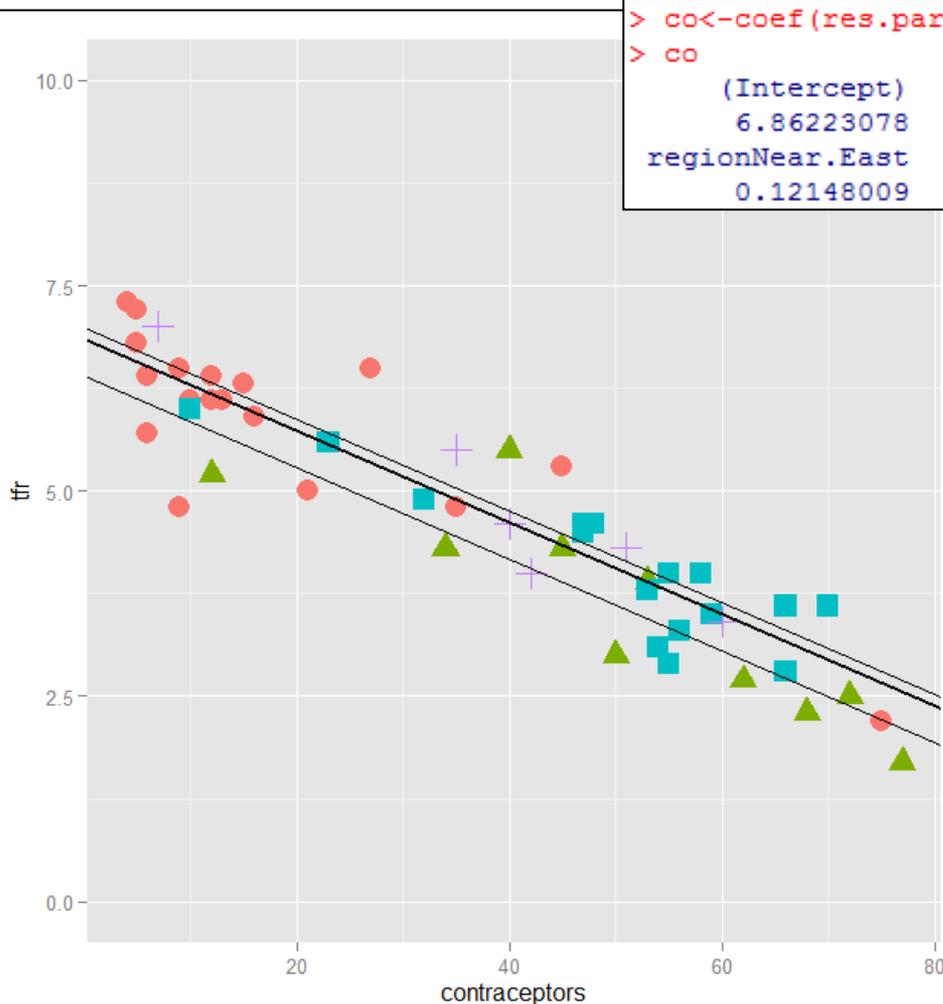
```

> anova(res.par)
Analysis of Variance Table

Response: tfr
          Df Sum Sq Mean Sq  F value Pr(>F)
contraceptors  1  87.672   87.672  278.5590 <2e-16 ***
region         3   1.677    0.559   1.7764 0.1652
Residuals     45  14.163    0.315
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

```

```
> ggplot(Robey, aes(x=contraceptors, y=tfr, shape=region)) +
+ geom_point(aes(colour=region), size=5) +
+ geom_abline(intercept=co[1], slope=co[2]) +
+ geom_abline(intercept=co[1]+co[3], slope=co[2]) +
+ geom_abline(intercept=co[1]+co[4], slope=co[2]) +
+ geom_abline(intercept=co[1]+co[5], slope=co[2]) +
+ expand_limits(y=c(0,10))
```



```
> co<-coef(res.par)
> co
```

	(Intercept)	contraceptors	regionAsia	regionLatin.Amer
	6.86223078	-0.05575432	-0.46203415	-0.02799942
regionNear.East	0.12148009			

El modelo de rectas paralelas no proporciona un ajuste mejor que el de la recta única.

```
> anova(res.1, res.par)
Analysis of Variance Table
```

	Model	Res.Df	RSS	Df	Sum of Sq	F	Pr(>F)
1	tfr ~ contraceptors	48	15.840				
2	tfr ~ contraceptors + region	45	14.163	3	1.6772	1.7764	0.1652

```
> res.grup <- lm(tfr~contraceptors*region,Robey)
> summary(res.grup)
```

Call:

```
lm(formula = tfr ~ contraceptors * region, data = Robey)
```

Residuals:

Min	1Q	Median	3Q	Max
-1.54546	-0.26527	-0.04661	0.34689	1.30579

Coefficients:

	Estimate	Std. Error	t value	Pr(> t)	
(Intercept)	6.832351	0.194090	35.202	< 2e-16	***
contraceptors	-0.054099	0.007718	-7.009	1.41e-08	***
regionAsia	-0.322375	0.563627	-0.572	0.570	
regionLatin.Amer	-0.237356	0.520948	-0.456	0.651	
regionNear.East	0.631733	0.632999	0.998	0.324	
contraceptors:regionAsia	-0.003795	0.012389	-0.306	0.761	
contraceptors:regionLatin.Amer	0.003136	0.012044	0.260	0.796	
contraceptors:regionNear.East	-0.013920	0.016141	-0.862	0.393	

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

Residual standard error: 0.5732 on 42 degrees of freedom

Multiple R-squared: 0.8667, Adjusted R-squared: 0.8445

Una recta para
cada region

```
> anova(res.grup)
```

Analysis of Variance Table

Response: tfr

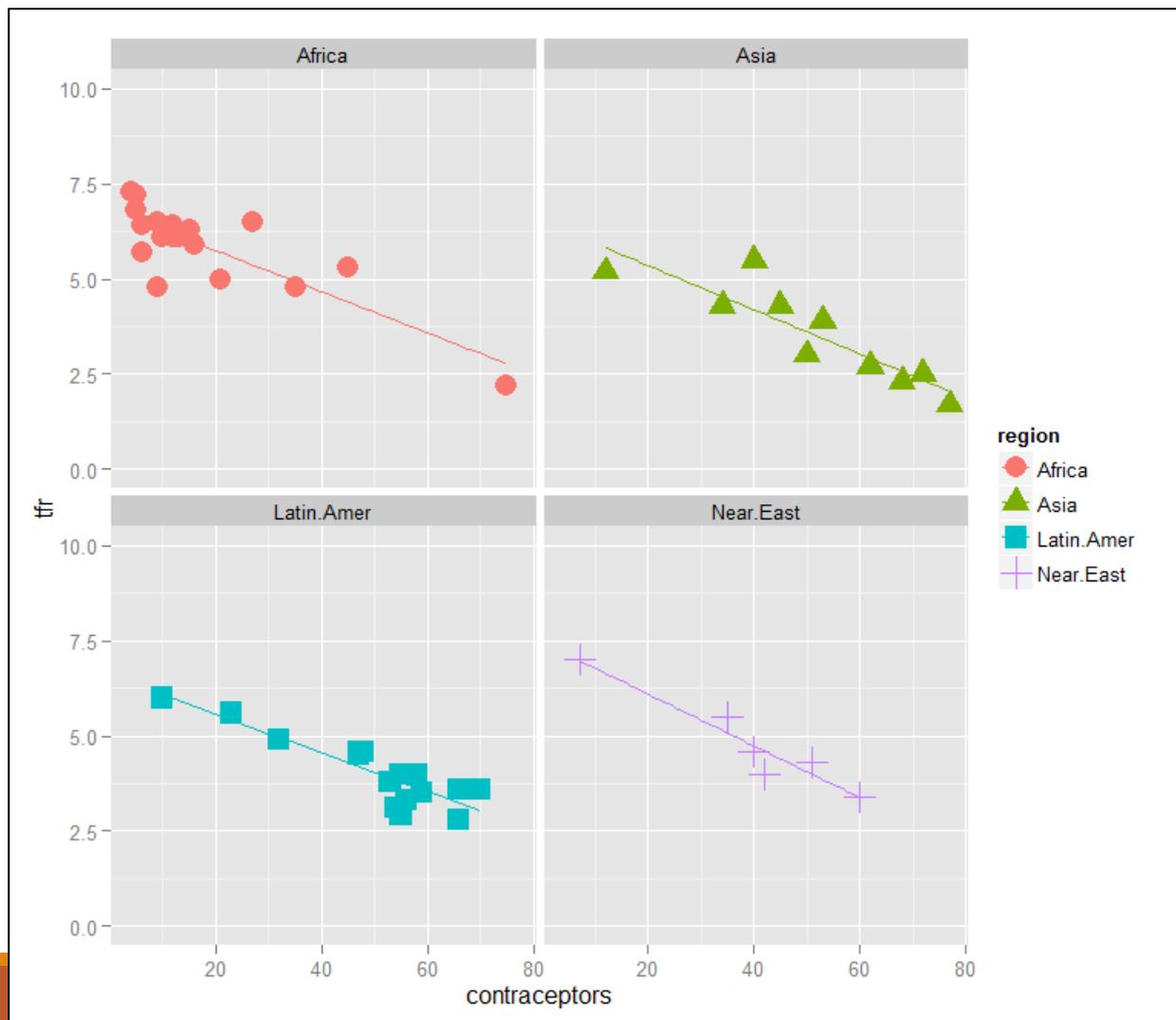
	Df	Sum Sq	Mean Sq	F value	Pr(>F)	
contraceptors	1	87.672	87.672	266.8706	<2e-16	***
region	3	1.677	0.559	1.7018	0.1812	
contraceptors:region	3	0.365	0.122	0.3706	0.7746	
Residuals	42	13.798	0.329			

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

```

> ggplot(Robey, aes(x=contraceptors, y=tfr, shape=region)) +
+ geom_point(aes(colour=region), size=5) +
+ geom_smooth(method="lm", se=F, aes(colour=region)) +
+ expand_limits(y=c(0,10)) +
+ facet_wrap(~region, ncol=2)

```



Comparación de los modelos

```
> anova(res.l, res.par, res.grup)
Analysis of Variance Table

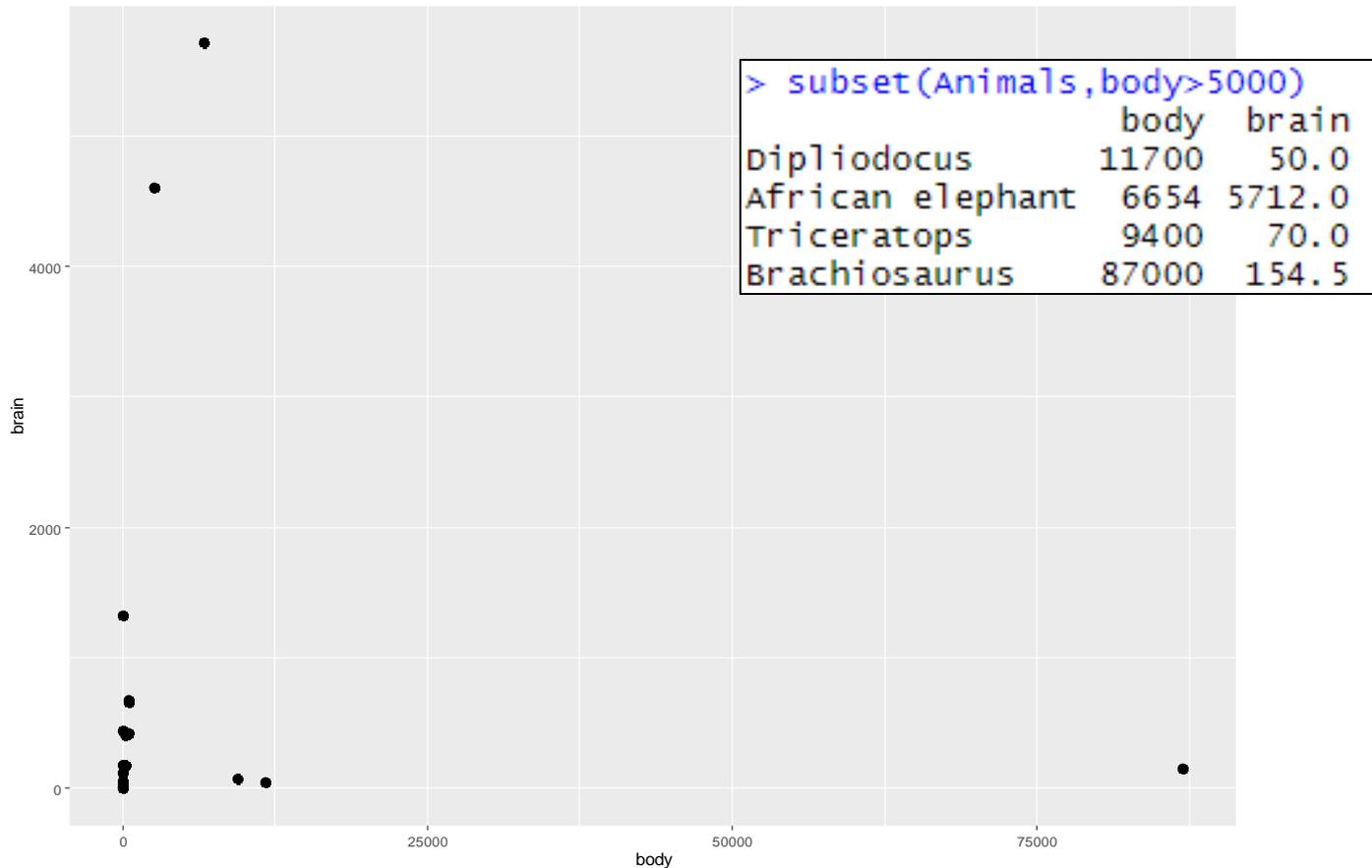
Model 1: tfr ~ contraceptors
Model 2: tfr ~ contraceptors + region
Model 3: tfr ~ contraceptors * region
  Res.Df    RSS Df Sum of Sq      F Pr(>F)
1      48 15.840
2      45 14.163  3   1.67724 1.7018 0.1812
3      42 13.798  3   0.36524 0.3706 0.7746
```

Los resultados pueden explicarse con el modelo de recta única. Por lo tanto, concluimos que la disminución de la tasa de fecundidad es similar en las distintas regiones al aumentar el uso de contraceptivos.

Con el modelo de recta única, $r^2=0.84$. Por lo tanto, la disminución de la tasa de fecundidad queda muy bien explicada por la relación lineal que se observa respecto al aumento de contracepción.

Tamaño cerebral y peso corporal en animales

Library MASS: Animals

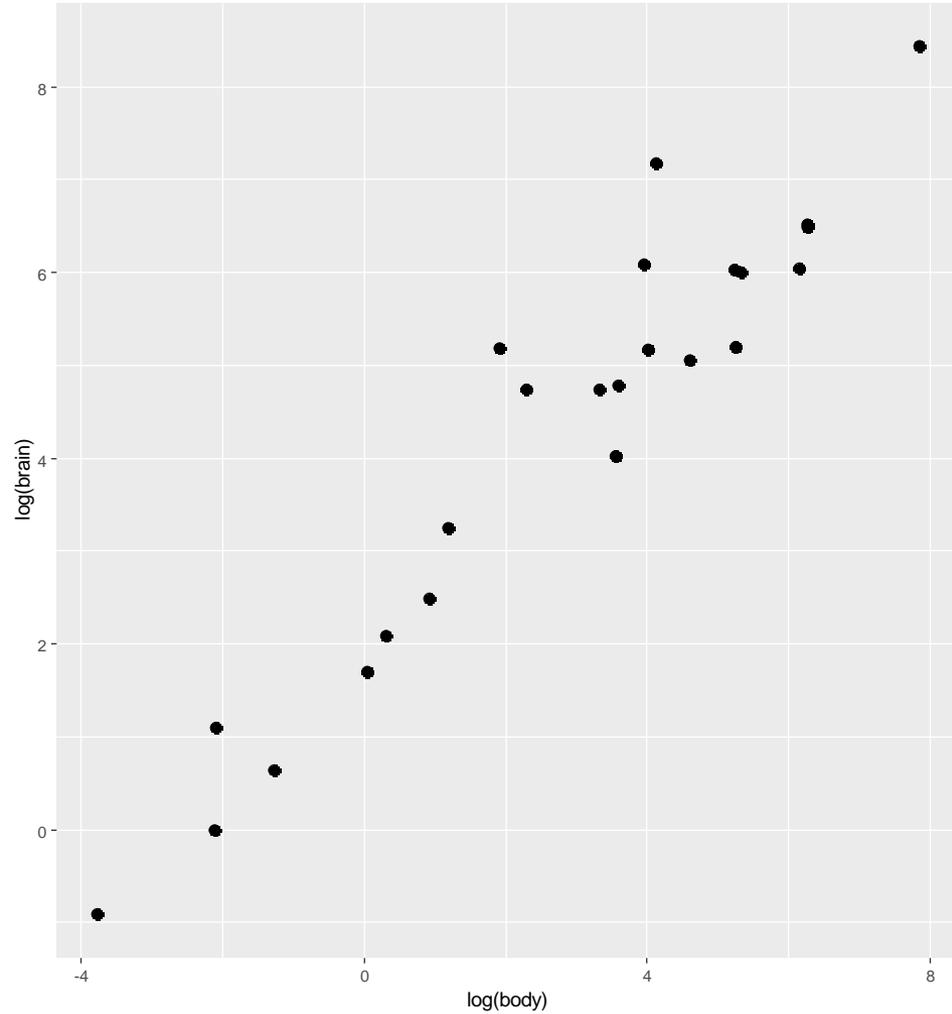
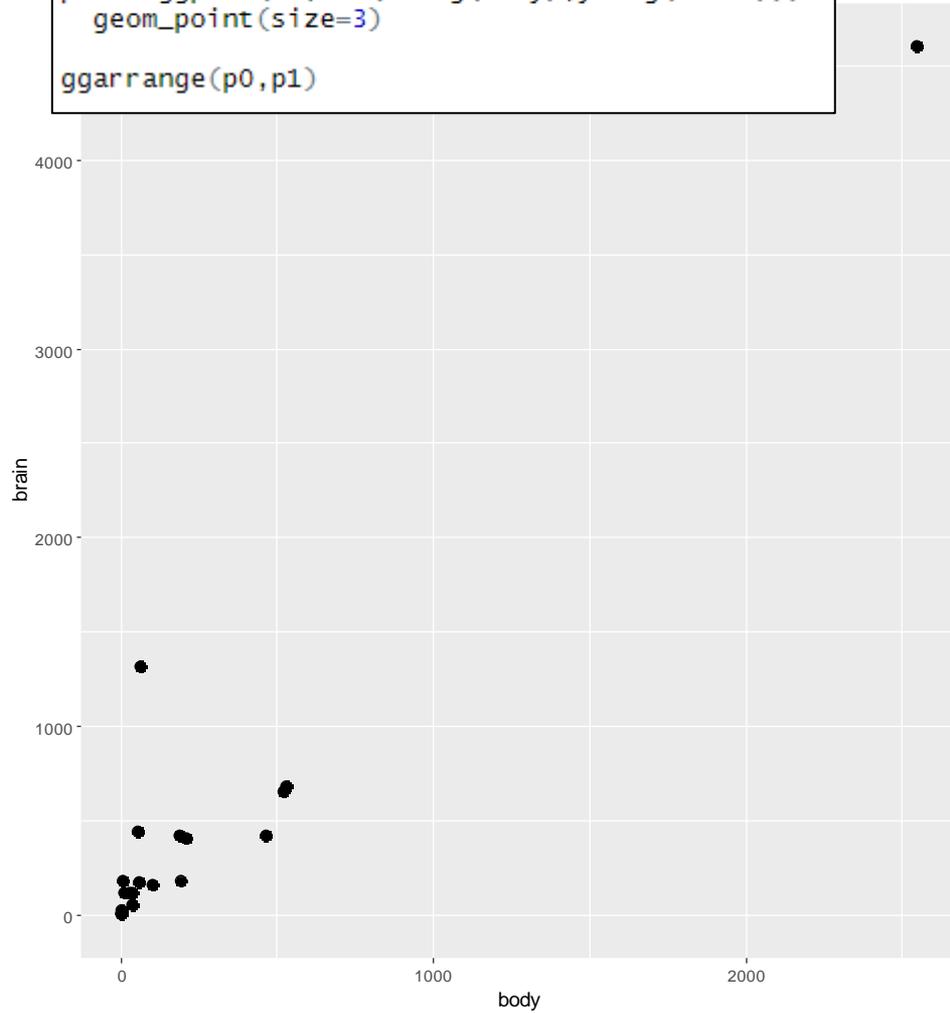


```
df <- Animals %>% filter(body<5000)
```

```
p0 <- ggplot(df, aes(x=body, y=brain))+  
  geom_point(size=3)
```

```
p1 <- ggplot(df, aes(x=log(body), y=log(brain)))+  
  geom_point(size=3)
```

```
ggarrange(p0, p1)
```



```
model <- lm(log(brain)~log(body),data=df)
summary(model)
```

```
ggplot(df,aes(x=log(body),y=log(brain)))+
  geom_point(size=3)+
  geom_smooth(method=lm)
```

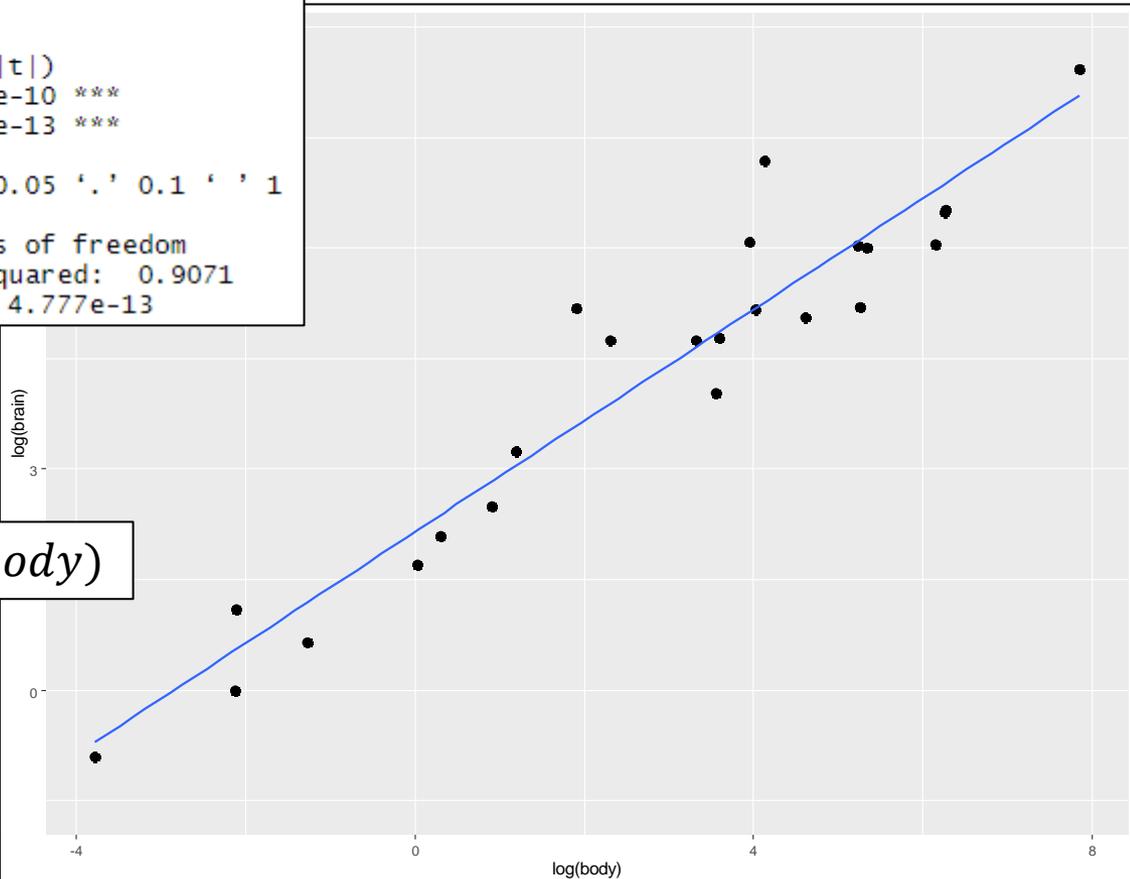
```
Call:
lm(formula = log(brain) ~ log(body), data = df)
```

```
Residuals:
    Min       1Q   Median       3Q      Max
-0.9259 -0.4888 -0.1906  0.2363  1.9207
```

```
Coefficients:
            Estimate Std. Error t value Pr(>|t|)
(Intercept)  2.14607    0.20635   10.40 5.89e-10 ***
log(body)    0.75564    0.05031   15.02 4.78e-13 ***
```

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

```
Residual standard error: 0.7416 on 22 degrees of freedom
Multiple R-squared:  0.9112,    Adjusted R-squared:  0.9071
F-statistic: 225.6 on 1 and 22 DF,  p-value: 4.777e-13
```



$$\log(\text{brain}) = 2.15 + 0.76 \times \log(\text{body})$$

$$\text{brain} = 8.58 \times \text{body}^{0.76}$$