

Para estimar con R la regresión a través del origen utilizaremos los datos `cars`, ya cargados en R, y estudiaremos la distancia de frenado (`dist`) en función de la velocidad (`speed`).

```
data("cars")
cars.lm <- lm(dist ~ speed, data = cars)
```

Podemos imponer que la ordenada  $\beta_0$  sea igual a 0 de dos maneras: añadiendo un 0 como si fuera un regresor o poniendo -1 después del regresor:

```
cars.lm2 <- lm(dist ~ 0 + speed, data = cars)
summary(cars.lm2)
cars.lm3 <- lm(dist ~ speed -1, data = cars)
summary(cars.lm3)
```

Compara el modelo de regresión estimado con y sin  $\beta_0$ .

### Solución:

El modelo con ordenada en el origen es preferible a la regresión a través del origen, primero porque la desviación típica residual es menor en el primero ( $s_R = 15.38$ ) que en el segundo ( $s_R = 16.26$ ). El  $R^2$  no es comparable en ambos casos porque en la regresión a través del origen se calcula con una expresión diferente (por ejemplo, se toma  $TSS = \sum_{i=1}^n y_i^2$ ). Además, en el modelo completo, el término independiente  $\beta_0$  es significativo individualmente a nivel 0.05 (p-valor=0.0123). Por otro lado, la distribución normal de media 0 y varianza igual a la residual  $s_R^2$  parece que ajusta mejor a los residuos en el modelo completo que en el modelo a través del origen. Y además la recta del modelo completo se ajusta mejor a los datos que la del modelo reducido (a pesar de que una recta tampoco parece el mejor modelo de todos, dada la curvatura de los datos).

```
summary(cars.lm)
```

Call:

```
lm(formula = dist ~ speed, data = cars)
```

Residuals:

Min	1Q	Median	3Q	Max
-29.069	-9.525	-2.272	9.215	43.201

Coefficients:

	Estimate	Std. Error	t value	Pr(> t )
(Intercept)	-17.5791	6.7584	-2.601	0.0123 *
speed	3.9324	0.4155	9.464	1.49e-12 ***

---

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

Residual standard error: 15.38 on 48 degrees of freedom

Multiple R-squared: 0.6511, Adjusted R-squared: 0.6438

F-statistic: 89.57 on 1 and 48 DF, p-value: 1.49e-12

```
summary(cars.lm3)
```

```
Call:
```

```
lm(formula = dist ~ speed - 1, data = cars)
```

```
Residuals:
```

```
    Min      1Q  Median      3Q     Max
-26.183 -12.637 -5.455  4.590  50.181
```

```
Coefficients:
```

```
      Estimate Std. Error t value Pr(>|t|)
speed  2.9091    0.1414   20.58  <2e-16 ***
```

```
---
```

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

```
Residual standard error: 16.26 on 49 degrees of freedom
```

```
Multiple R-squared:  0.8963, Adjusted R-squared:  0.8942
```

```
F-statistic: 423.5 on 1 and 49 DF, p-value: < 2.2e-16
```

```
anova(cars.lm, cars.lm3)
```

```
Analysis of Variance Table
```

```
Model 1: dist ~ speed
```

```
Model 2: dist ~ speed - 1
```

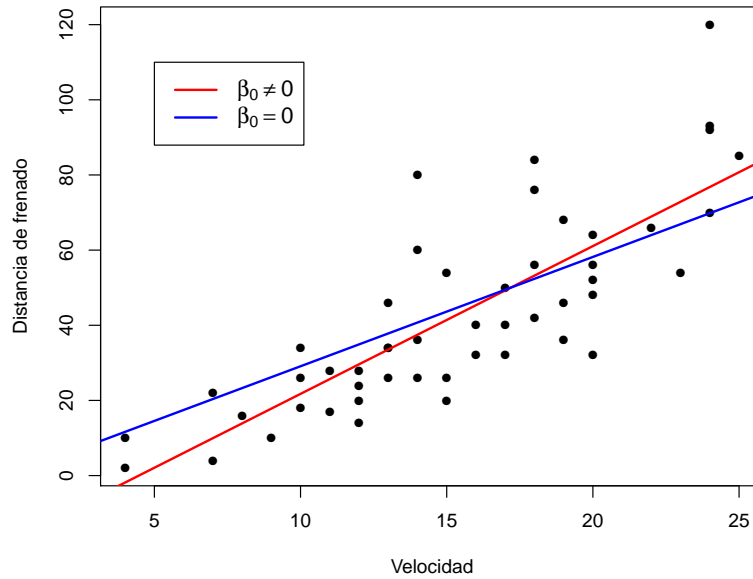
	Res.Df	RSS	Df	Sum of Sq	F	Pr(>F)
1	48	11354				
2	49	12954	-1	-1600.3	6.7655	0.01232 *

```
---
```

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

Análisis gráfico:

```
plot(cars$speed,cars$dist,pch=16,col="black",xlab="Velocidad",ylab="Distancia de
frenado")
abline(cars.lm,col="red",lwd=2)
abline(cars.lm3,col="blue",lwd=2)
legend(5, 110, legend=c(expression(beta[0] != 0), expression(beta[0] = 0)),
      col=c("red", "blue"), lty=c(1,1), cex=1.2,lwd=c(2,2))
```



Histogramas de residuos y comparación con la normal:

```
par( mfrow = c( 1, 2 ) )
# Residuos de reg.lm
evec = seq(-30,45,0.01)
sR = summary(cars.lm)$sigma # Desviacion tipica residual
hist(cars.lm$residuals,freq=F,main="")
lines(evec,dnorm(evec,mean=0,sd=sR))
# Residuos de reg.lm3
evec3 = seq(-30,60,0.01)
sR3 = summary(cars.lm3)$sigma # Desviacion tipica residual
hist(cars.lm3$residuals,freq=F,main="")
lines(evec3,dnorm(evec3,mean=0,sd=sR3))
par( mfrow = c( 1, 1 ) )
```

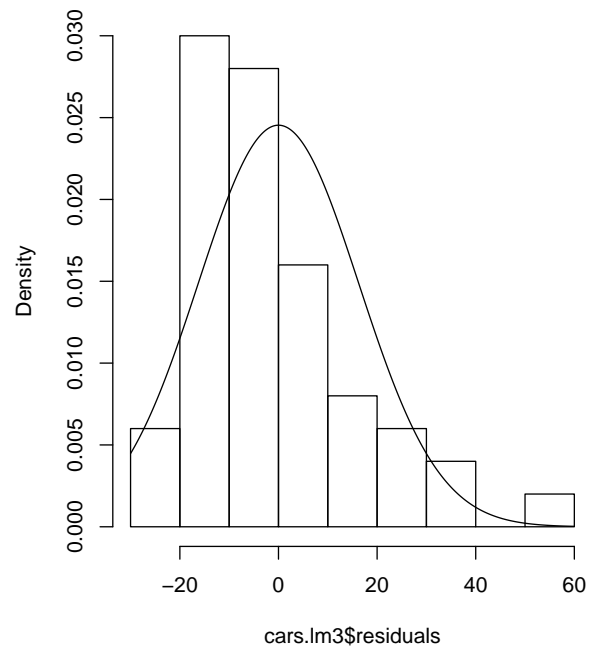
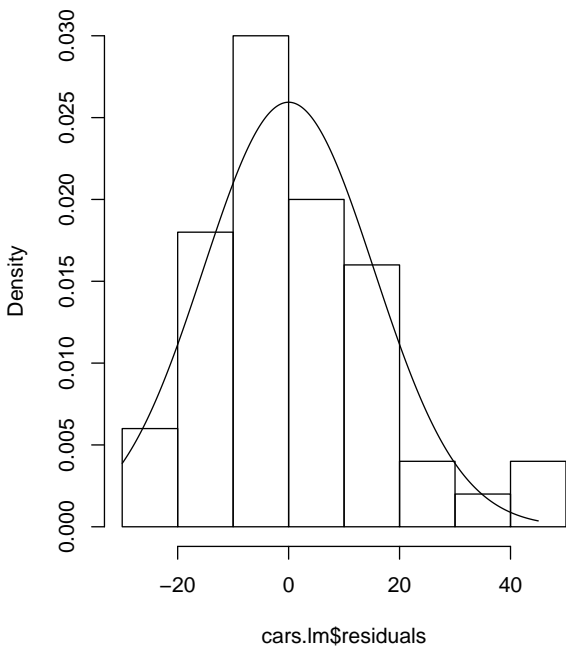


Diagrama de dispersión de los residuos frente a los regresores:

```
par( mfrow = c( 1, 2 ) )  
plot(cars$speed,cars.lm$residuals,pch=16,col="black", xlab="Velocidad",  
      ylab="Residuos",main=expression(paste("Modelo con ",beta[0] != 0)))  
abline(h=0)
```

```
plot(cars$speed,cars.lm3$residuals,pch=16,col="black", xlab="Velocidad",  
      ylab="Residuos",main=expression(paste("Modelo con ",beta[0] == 0)))  
abline(h=0)  
par( mfrow = c( 1, 1 ) )
```

