

## Example CH7: Multiple regression (Estimation)

R Package: faraway

Dataset: gala

Dependent variable: Species (number of species of tortoise found on Galápagos Islands)

Independent variables:

- Endemics: the number of endemic species  $x_1$
- Area: the area of the island  $x_2$
- Elevation: the highest elevation of the island  $x_3$
- Nearest: the distance from the nearest island  $x_4$
- Scrutz: the distance from Santa Cruz Island  $x_5$
- Adjacent: the area of the adjacent island  $x_6$

```
> library(faraway)
```

```
> data(gala)
```

```
> gala
```

	Species	Endemics	Area	Elevation	Nearest	Scrutz	Adjacent
Baltra	58	23	25.09	346	0.6	0.6	1.84
Bartolome	31	21	1.24	109	0.6	26.3	572.33
Caldwell	3	3	0.21	114	2.8	58.7	0.78
Champion	25	9	0.10	46	1.9	47.4	0.18
Coamano	2	1	0.05	77	1.9	1.9	903.82
Daphne.Major	18	11	0.34	119	8.0	8.0	1.84
Daphne.Minor	24	0	0.08	93	6.0	12.0	0.34
Darwin	10	7	2.33	168	34.1	290.2	2.85
Eden	8	4	0.03	71	0.4	0.4	17.95
Enderby	2	2	0.18	112	2.6	50.2	0.10
Espanola	97	26	58.27	198	1.1	88.3	0.57
Fernandina	93	35	634.49	1494	4.3	95.3	4669.32
Gardner1	58	17	0.57	49	1.1	93.1	58.27
Gardner2	5	4	0.78	227	4.6	62.2	0.21
Genovesa	40	19	17.35	76	47.4	92.2	129.49
Isabela	347	89	4669.32	1707	0.7	28.1	634.49
Marchena	51	23	129.49	343	29.1	85.9	59.56
Onslow	2	2	0.01	25	3.3	45.9	0.10
Pinta	104	37	59.56	777	29.1	119.6	129.49
Pinzon	108	33	17.95	458	10.7	10.7	0.03
Las.Plazas	12	9	0.23	94	0.5	0.6	25.09
Rabida	70	30	4.89	367	4.4	24.4	572.33
SanCristobal	280	65	551.62	716	45.2	66.6	0.57
SanSalvador	237	81	572.33	906	0.2	19.8	4.89
SantaCruz	444	95	903.82	864	0.6	0.0	0.52
SantaFe	62	28	24.08	259	16.5	16.5	0.52
SantaMaria	285	73	170.92	640	2.6	49.2	0.10
Seymour	44	16	1.84	147	0.6	9.6	25.09
Tortuga	16	8	1.24	186	6.8	50.9	17.95
Wolf	21	12	2.85	253	34.1	254.7	2.33

```
> dim(gala) # number of rows by number of columns
```

```
[1] 30 7
```

$n=30$   $k=5$

lm

$$y_i = \beta_0 + \beta_1 x_{i1} + \beta_2 x_{i2} + \dots + \beta_5 x_{i5} + \epsilon_i$$

$i = 1, \dots, n, n = 30$

$$\underline{\epsilon} = (\epsilon_1, \dots, \epsilon_n)' \sim N(0, \sigma^2 I)$$

Perform multiple regression on the gala dataset

```
> mdl <- lm(Species~Area+Elevation + Nearest + Scrub + Adjacent, data=gala)
> summary(mdl)
```

Call:

```
lm(formula = Species ~ Area + Elevation + Nearest + Scrub + Adjacent,
    data = gala)
```

Residuals:

Min	1Q	Median	3Q	Max
-111.679	-34.898	-7.862	33.460	182.584

Coefficients:

	Estimate	Std. Error	t value	Pr(> t )
(Intercept)	7.068221	19.154198	0.369	0.715351
Area	-0.023938	0.022422	-1.068	0.296318
Elevation	0.319465	0.053663	5.953	3.82e-06 ***
Nearest	0.009144	1.054136	0.009	0.993151
Scrub	-0.240524	0.215402	-1.117	0.275208
Adjacent	-0.074805	0.017700	-4.226	0.000297 ***

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

Residual standard error: 60.98 on 24 degrees of freedom

Multiple R-squared: 0.7658, Adjusted R-squared: 0.7171

F-statistic: 15.7 on 5 and 24 DF, p-value: 6.838e-07

#Extract the residual standard error, the estimate of sigma

```
> names(summary(mdl))
```

[1] "call"	"terms"	"residuals"	"coefficients"	"aliases"
[6] "sigma"	"df"	"r.squared"	"adj.r.squared"	
"fstatistic"				

```
[11] "cov.unscaled"
```

```
> mdls <- summary(mdl)
```

```
> mdls$sigma
```

```
[1] 60.97519
```

```
# (x'x)-1
```

```
> mdls$cov.unscaled
```

	(Intercept)	Area	Elevation	Nearest	Scrub	Adjacent
(Intercept)	9.867829e-02	3.778242e-05	-1.561976e-04	-2.339027e-04	-3.760293e-04	2.309832e-05
Area	3.778242e-05	1.352247e-07	-2.593617e-07	1.294003e-06	-4.913149e-08	4.620303e-08
Elevation	-1.561976e-04	-2.593617e-07	7.745339e-07	-3.549366e-06	3.080831e-07	-1.640241e-07
Nearest	-2.339027e-04	1.294003e-06	-3.549366e-06	2.988732e-04	-3.821077e-05	1.424729e-06
Scrub	-3.760293e-04	-4.913149e-08	3.080831e-07	-3.821077e-05	1.247941e-05	-1.958356e-07
Adjacent	2.309832e-05	4.620303e-08	-1.640241e-07	1.424729e-06	-1.958356e-07	8.426543e-08

#Alternate approaches to obtain standard errors of coefficients

```
> sqrt(diag(mdls$cov.unscaled))*mdls$sigma
```

(Intercept)	Area	Elevation	Nearest	Scrub	Adjacent
19.15419782	0.02242235	0.05366280	1.05413595	0.21540225	0.01770019

```
> mdls$coef[,2]
```

(Intercept)	Area	Elevation	Nearest	Scrub	Adjacent
19.15419782	0.02242235	0.05366280	1.05413595	0.21540225	0.01770019