

Data should be either in comma delimited (`sep=","`) or tab delimited (`sep="\t"`) format. Data should live in the **same** directory as the R program (`.RData`).

```
> births <- read.table("NCBIRTH800.csv", header=T, sep=",")
> dim(births)
> names(births)
```

Notice that the variable names we'll use are `mage`, `tounces`, `gained`, and `smoke`.

```
> attach(births)
```

When we `attach` we create new variables from the data set. That is, we can now use each of the columns as their own variables / vectors.

3. We're interested in predicting a baby's size from the mother's weight gain.

```
> oz.g.lm <- lm(tounces ~ gained)
> summary(oz.g.lm)
> anova(oz.g.lm)
```

Analysis of Variance Table

Response: tounces

	Df	Sum Sq	Mean Sq	F value	Pr(>F)
gained	1	18856	18856	40.293	3.718e-10 ***
Residuals	775	362671	468		

---

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

4. What if we include smoking status as a variable?

```
> oz.gs.lm <- lm(tounces ~ gained + smoke)
> summary(oz.gs.lm)
> anova(oz.gs.lm)
```

Analysis of Variance Table

Response: tounces

	Df	Sum Sq	Mean Sq	F value	Pr(>F)
gained	1	18856	18856	40.946	2.709e-10 ***
smoke	1	6241	6241	13.553	0.0002480 ***
Residuals	774	356429	461		

5. What if we let smoking and weight gain *interact*?

```
> oz.gis.lm <- lm(tounces ~ gained * smoke)
> summary(oz.gis.lm)
> anova(oz.gis.lm)
```

Analysis of Variance Table

Response: tounces

	Df	Sum Sq	Mean Sq	F value	Pr(>F)	
gained	1	18856	18856	40.964	2.688e-10	***
smoke	1	6241	6241	13.559	0.0002472	***
gained:smoke	1	613	613	1.332	0.2487989	
Residuals	773	355816	460			

6. What happens to the model if we add in another quantitative variable?

```
> oz.gsm.lm <- lm(tounces ~ gained + smoke + mage)
> summary(oz.gsm.lm)
> anova(oz.gsm.lm)
```

Analysis of Variance Table

Response: tounces

	Df	Sum Sq	Mean Sq	F value	Pr(>F)	
gained	1	18856	18856	42.029	1.603e-10	***
smoke	1	6241	6241	13.912	0.0002056	***
mage	1	9629	9629	21.462	4.234e-06	***
Residuals	773	346801	449			

7. What happens to the model if we add in another quantitative variable with the interaction?

```
> oz.gism.lm <- lm(tounces ~ gained * smoke + mage)
> summary(oz.gism.lm)
> anova(oz.gism.lm)
```

Analysis of Variance Table

Response: tounces

	Df	Sum Sq	Mean Sq	F value	Pr(>F)	
gained	1	18856	18856	42.0240	1.607e-10	***
smoke	1	6241	6241	13.9102	0.0002058	***
mage	1	9629	9629	21.4596	4.240e-06	***
gained:smoke	1	412	412	0.9177	0.3383793	
Residuals	772	346389	449			

8. On a different note... consider the two different multiple regression model with race as a quantitative variable (a) and race as a categorical variable (b):

```
(a) > oz.gr.lm <- lm( tounces ~ gained + racemom)
> summary(oz.gr.lm)
```

Coefficients:

	Estimate	Std. Error	t value	Pr(> t )
(Intercept)	106.47125	2.18549	48.717	< 2e-16 ***
gained	0.36200	0.05693	6.359	3.46e-10 ***
racemom	-0.58526	0.81427	-0.719	0.473

Residual standard error: 21.64 on 774 degrees of freedom

(23 observations deleted due to missingness)

Multiple R-squared: 0.05006, Adjusted R-squared: 0.0476

F-statistic: 20.39 on 2 and 774 DF, p-value: 2.340e-09

```
(b) > oz.grc.lm <- lm( tounces ~ gained + factor(racemom) )
> summary(oz.grc.lm)
```

Coefficients:

	Estimate	Std. Error	t value	Pr(> t )
(Intercept)	108.05629	1.92957	56.000	< 2e-16 ***
gained	0.35050	0.05602	6.257	6.50e-10 ***
factor(racemom)2	-9.75036	1.87327	-5.205	2.49e-07 ***
factor(racemom)3	-6.62293	6.19723	-1.069	0.2855
factor(racemom)4	2.17102	15.07064	0.144	0.8855
factor(racemom)7	-22.62293	21.26517	-1.064	0.2877
factor(racemom)8	11.89066	6.46737	1.839	0.0664 .

Residual standard error: 21.25 on 770 degrees of freedom

(23 observations deleted due to missingness)

Multiple R-squared: 0.08894, Adjusted R-squared: 0.08184

F-statistic: 12.53 on 6 and 770 DF, p-value: 1.66e-13

Write out the regression model (comparing baby's weight and how much the mother gained) for a mother who is Chinese (category 4) for each of the two different models above.

```
> table(racemom)
```

```
racemom
 1   2   3   4   7   8
604 169  12   2   1  12
```