Homework 6

Due Friday Oct 27 by 3:00

1. Load the ChickWeight data:

```
data(ChickWeight)
names(ChickWeight)
attach(ChickWeight)
```

(a) Plot the data. You might find this command useful:

```
coplot(weight ~ Time | Chick, data = ChickWeight,type = "b", show.given = FALSE)
```

(b) Extract the data corresponding to the sixth chick. (In other words, where ChickWeight$Weight==6). Fit a linear model using time to predict weight. Check the residuals. Does the model fit well? Now try fitting a polynomial model. Show, using a residual plot, that you have improved the fit. Summarize the resulting fitted model.

(c) Now use all the data. Ignore the variables `Chick` and `Diet`. Fit a linear model to predict weight from time. Summarize the fitted model. Based on a residual plot, do you think the model fits well? See if you can get a better fit using polynomials. Explain why this does or does not help.

(d) Repeat the analysis in part (c) but include the variable `Diet`. Make sure you treat `Diet` as a categorical variable. Comment on the model and the residuals. You should plot the residuals versus Time and versus Diet.

2. Consider the following data on income Y (in thousands of dollars) and country of birth (France, Italy or USA):

Y	33	36	35	35	31	29	31	29	37	39	36	36
X_1	1	1	1	1	0	0	0	0	0	0	0	0
X_2	0	0	0	0	1	1	1	1	0	0	0	0
X_3	0	0	0	0	0	0	0	0	1	1	1	1

Note that X_1, X_2 and X_3 are dummy variables for France, Italy and USA, respectively.

(a) Show that, if we include all three of these covariates (and an intercept) in the model, then then $\mathbf{X}^T\mathbf{X}$ is not invertible. Explain why it is not invertible.

(b) Fit the model:

$$Y = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \epsilon.$$

Summarize the fitted model.

(c) Estimate the mean income of each of the three countries in terms of $\widehat{\beta}_0$, $\widehat{\beta}_1$ and $\widehat{\beta}_2$.

(d) Using the regression output, construct a 95 percent confidence interval for the mean income of France.
Hint 1: in general, if $W = (W_1, W_2) \sim N(\mu, \Sigma)$, then $a^T W \sim N(a^T\mu, a^T\Sigma a)$.
Hint 2: check out the `vcov` command in R.

3. Load the birthweight data:

```
library(MASS)
help(birthwt)
```

The goal is to predict birthweight from the other variables. Ignore the variable called `low`. Some of the covariates are discrete factors. You should use the `factor` command to turn these into factors (discrete variables). In particular, the variables `race`, `ptl`, `ftv`. So, for example, the command `factor(race)` tells R to treat this variable as a factor.

(a) Plot the data using the pairs command.

(b) Fit a multiple regression model. Summarize the fitted model. (Remember to ignore the variable called `low`.)

(c) Check the assumptions by plotting residuals. Note that the residual plots for the discrete variables will be quite different than the residual plots you are used to. But they are still useful for checking for constant variance and for looking for outliers.

(d) Briefly but clearly summarize your conclusions. In particular, summarize your fitted model and interpret the coefficients.