Biostat II: Lab 15, Analyzing the Challenger data by logistic regression `R`
Date: 16 May 2008

In today's lab, we look at data related to failure of the space flight *Challenger*, which had a catastrophic failure due to burning of an O-ring on January 28, 1986. Of the 24 previous shuttle flights, 7 had incidents of damage to joints, 16 had no incidents of damage, and 1 was unknown. After observing these flights, scientists began to suspect that O-ring failure could be related to temperature at the time of the launch. We will use logistic regression to make our own judgment.

Below, find data for the 23 previous shuttle flights with known outcomes:

Temperature (degrees F)	66	70	69	68	67	72	73	70	57	63	70
O-ring Failure	0	1	0	0	0	0	0	0	1	1	1

Temperature (degrees F)	78	67	53	67	75	70	81	76	79	75	76	58
O-ring Failure	0	0	1	0	0	0	0	0	0	1	0	1

1. Read the data into `R` using the following `c()` commands.

```
Temp <- c(66, 70, 69, 68, 67, 72, 73, 70, 57, 63, 70, 78, 67, 53, 67, 75, 70, 81,
 76, 79, 75, 76, 58)

Failure <- c(0, 1, 0, 0, 0, 0, 0, 0, 1, 1, 1, 0, 0, 1, 0, 0, 0, 0, 0, 0, 1, 0, 1)
```

2. Calculate the overall failure rate based on these data. Hint: look at `table(Failure)`

3. Calculate the observed proportion of shuttle flights with O-ring failure among those with temperature below 65 degrees Fahrenheit.
   Hint: it may help to look at a table such as `table(Failure[Temp<65])` although you could also extract the information manually from the table above.

4. Calculate the observed proportion of shuttle flights with O-ring failure among those with temperature above 65 degrees Fahrenheit. Do the failure rates vary much when we stratify by temperature?

5. Now, create a plot of jittered O-ring outcomes versus temperature. You can use code such as:

```
plot(Temp, jitter(Failure))
```

6. Add a lowess line to the plot you've just created:

```
lines(lowess(Failure~Temp))
```

   According to the plot you've created, how does probability of failure vary by temperature?

7. Now, perform logistic regression of Failure using an intercept only model:

```
summary(out.int <- glm(Failure ~ 1, family=binomial))
```

Here, we use the `glm` command to perform any generalized linear model, and we specify that we are doing logistic regression using the option `family=binomial`.

8. Write down the model you've just fit, specifying both systematic and random components.

9. Interpret the fitted coefficient β_0 for the intercept only model.

10. Transform the fitted log odds to probability, and compare to the overall probability of failure you calculated in exercise 2.

11. Now, fit a logistic regression model of Failure on Temperature centered at 50 degrees Fahrenheit.

```
summary(out.temp <- glm(Failure~I(Temp-50), family=binomial))
```

12. Calculate the likelihood of the intercept only model and the simple logistic regression on temperature:

```
logLik(out.int)
logLik(out.temp)
```

13. Perform a likelihood ratio test to determine if the Temperature term was a statistically significant addition to the model. Write down you null and alternative hypotheses, and choose your own level of alpha. Look up the p-value in `R` by typing:

```
pchisq(-2*(logLik(out.int) - logLik(out.temp)), df=1, lower.tail=F)
```

Note: You can perform the same test using the shortcut method given in lecture today, with which you don't have to really 'think' about what you're doing so much:

```
anova(out.int, out.temp, test="Chi")
```

14. Report the level of significance according to the Wald test, as reported in the table. Compare to your likelihood ratio test.

Generally, the likelihood ratio test should agree with the Wald test, as they are asymptotically equivalent. However, in small sample sizes we may see that they disagree notably. In these cases, the likelihood ratio test results are probably more trustworthy than the Wald test.

15. The catastrophic failure of the *Challenger* mission happened at a temperature of 31 degrees Fahrenheit. Use the logistic regression of Failure on Temperature that you just fit to predict the log odds of failure at a temperature of 31 degrees.

16. Now, transform your predicted log odds at 31 degrees to obtain the predicted probability of failure at 31 degrees. Interpret.