20.9. Donner Party.

For females: $\log(\text{odds}) = 3.2 - (0.078 \times age)$

For males: $\log(\text{odds}) = 1.6 - (0.078 \times age)$

(a) The estimated probabilities of survival are:

- For 25 year-old men:

$$\log\left(\frac{\hat{\pi}}{1-\hat{\pi}}\right) = 1.6 - (0.078)(25) = -0.35$$

$$\hat{\pi} = \frac{\exp(-0.35)}{1+\exp(-0.35)} = \mathbf{0.4134}$$

- For 50 year-old men:

$$\log\left(\frac{\hat{\pi}}{1-\hat{\pi}}\right) = 1.6 - (0.078)(50) = -2.3$$

$$\hat{\pi} = \frac{\exp(-2.3)}{1+\exp(-2.3)} = \mathbf{0.0911}$$

- For 25 year-old women:

$$\log\left(\frac{\hat{\pi}}{1-\hat{\pi}}\right) = 3.2 - (0.078)(25) = 1.25$$

$$\hat{\pi} = \frac{\exp(1.25)}{1+\exp(1.25)} = \mathbf{0.7773}$$

- For 50 year-old women:

$$\log\left(\frac{\hat{\pi}}{1-\hat{\pi}}\right) = 3.2 - (0.078)(50) = -0.7$$

$$\hat{\pi} = \frac{\exp(-0.7)}{1+\exp(-0.7)} = \mathbf{0.3318}$$

(b) When the estimated probability of survival is $\hat{\pi} = 0.5$, then

$$\log\left(\frac{\hat{\pi}}{1-\hat{\pi}}\right) = \log\left(\frac{0.5}{0.5}\right) = \log(1) = 0$$

(i) For men, the age corresponding to $\hat{\pi} = 0.5$ is

$$0 = 1.6 - (0.078 \times age) \Rightarrow age = \mathbf{20.51}$$

(ii) For women, the age corresponding to $\hat{\pi} = 0.5$ is

$$0 = 3.2 - (0.078 \times age) \Rightarrow age = \mathbf{41.03}$$

20.10. Odds Ratio.

Let ω_A = the odds at A, and ω_B = the odds at B. Then

$$\log(\omega_A) - \log(\omega_B) = (\beta_0 + \beta_1 A + \beta_2 x_2 + ...) - (\beta_0 + \beta_1 B + \beta_2 x_2 + ...)$$

$$= \beta_1 A - \beta_1 B = \beta_1(A - B)$$

Note, however, that the left-hand side can be written

$$\log(\omega_A) - \log(\omega_B) = \log\left(\frac{\omega_A}{\omega_B}\right)$$

and so taking the anti-logarithm of both sides yields

$$\frac{\omega_A}{\omega_B} = \exp(\beta_1(A - B))$$

HW7 Sample Solution

20.11

```
> shuttle=data.frame(TEMP=ex2011$Temp,FAILURE=ex2011$Failure)
```

(a) R output given below:

```
> mshut <- glm(FAILURE~TEMP,family=binomial,data=shuttle)
> summary(mshut)

Call:
glm(formula = FAILURE ~ TEMP, family = binomial, data = shuttle)

Deviance Residuals:
    Min       1Q   Median       3Q      Max
-1.2125  -0.8253  -0.4705   0.5907   2.0512

Coefficients:
            Estimate Std. Error z value Pr(>|z|)
(Intercept) 10.87535    5.70291   1.907   0.0565 .
TEMP        -0.17132    0.08344  -2.053   0.0400 *
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

(Dispersion parameter for binomial family taken to be 1)

    Null deviance: 28.975  on 23  degrees of freedom
Residual deviance: 23.030  on 22  degrees of freedom
AIC: 27.030

Number of Fisher Scoring iterations: 4
```

The fitted logistic regression model is

$$\log\left(\frac{\hat{\pi}}{1-\hat{\pi}}\right) = 10.875 - 0.171(temp)$$

HW7 Sample Solution

(b) Wish to test H_0: $\beta_1 = 0$ vs. H_a: $\beta_1 < 0$. The test statistic is

$$z = \frac{\hat{\beta}_1}{\mathrm{SE}(\hat{\beta}_1)} = \frac{-0.171}{0.083} = \mathbf{-2.053}$$

The one-sided p-value is

$$P(Z < -2.053) = \mathbf{0.020}$$

We reject H_0 and conclude that the odds of O-ring failure *decrease* with temperature.

```
> pnorm(-2.053)
[1] 0.02003629
```

(c) Wish to test H_0: $\beta_1 = 0$ vs. H_a: $\beta_1 \neq 0$, using the drop-in-deviance test. The test statistic is the drop in deviance:

$$\chi^2 = D_{reduced} - D_{full} = 28.975 - 23.030 = \mathbf{5.945}$$

Under H_0, χ^2 follows a chi-squared distribution with $23 - 22 = 1$ degree of freedom. The p-value for this test is

$$P(\chi_1^2 > 5.945) = \mathbf{0.015}$$

Hence we reject H_0 and conclude the full model (where $\beta_1 \neq 0$) is a better fit.

```
> 1-pchisq(5.945,df=1)
[1] 0.01475909
```

(d) A 95% confidence interval for β_1 is given by

$$\hat{\beta}_1 \pm z(.975)\mathrm{SE}(\hat{\beta}_1) = -0.1713 \pm (1.96)(0.0834)$$

$= \mathbf{-0.335 \text{ to } -0.0078}$.

(e) The estimated logit of failure probability at 31° F is

$$\log\left(\frac{\hat{\pi}}{1-\hat{\pi}}\right) = 10.875 - 0.1713(31) = \mathbf{5.56}$$

(This corresponds to an estimated probability of O-ring failure of 0.9962).

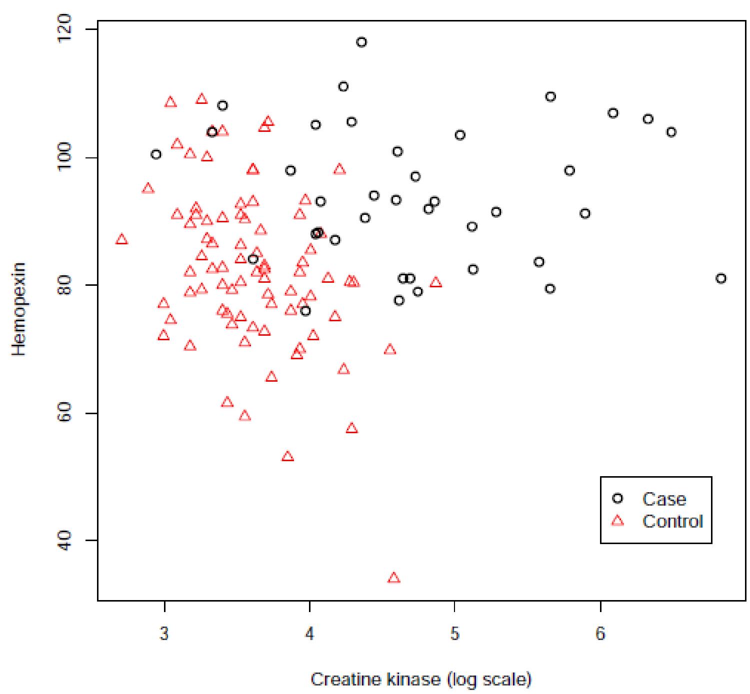
(f) The data used to fit the logistic regression model contained temperatures ranging from 51° F to 81° F. To make a prediction based on an observation falling so far from these data is an **extrapolation** and can yield very inaccurate results.

HW7 Sample Solution

20.12

(a) Scatterplot provided on the following page.

```
> ex2012$lck <- with(ex2012,log(CK))
> attach(ex2012)
> plot(H~lck,xlab="Creatine kinase (log scale)",ylab="Hemopexin",
  col=as.numeric(GROUP),pch=as.numeric(GROUP))
> legend(6,50,col=c(1,2),pch=c(1,2),legend=c("Case","Control"))
```



There is a clear separation between cases and controls, in terms of these two enzymes. So yes, these enzymes will be useful predictors of whether a woman is a carrier.

(b) R output given below:

```
> m1 <- glm(GROUP~CK+I(CK^2),family=binomial)
> summary(m1)

Call:
glm(formula = GROUP ~ CK + I(CK^2), family = binomial)

Deviance Residuals:
     Min        1Q    Median        3Q       Max
-2.50614  -0.03892   0.37943   0.51824   2.27518

Coefficients:
              Estimate Std. Error z value Pr(>|z|)
(Intercept)  4.181e+00  7.272e-01   5.749 8.96e-09 ***
CK          -5.805e-02  1.301e-02  -4.460 8.18e-06 ***
I(CK^2)      5.060e-05  3.286e-05   1.540    0.124
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

    Null deviance: 149.840  on 119  degrees of freedom
Residual deviance:  85.435  on 117  degrees of freedom
AIC: 91.435
```

The CK-squared term does **not** differ significantly from 0.

R output given below:

```
> m2 <- glm(GROUP~lck+I(lck^2),family=binomial)
> summary(m2)

Call:
glm(formula = GROUP ~ lck + I(lck^2), family = binomial)

Deviance Residuals:
     Min        1Q    Median        3Q       Max
-2.39251  -0.03075   0.38037   0.50190   2.28852

Coefficients:
            Estimate Std. Error z value Pr(>|z|)
(Intercept)   -9.830     16.309  -0.603    0.547
lck            8.568      8.366   1.024    0.306
I(lck^2)      -1.453      1.064  -1.365    0.172

    Null deviance: 149.84  on 119  degrees of freedom
Residual deviance:  84.98  on 117  degrees of freedom
AIC: 90.98
```

The squared term does **not** significantly differ from 0.

The model with the untransformed CK has more significant p-values for the terms. However, the AIC is smaller for the model with the log-transformed CK. Also, the distribution of log(CK) is more symmetric than the distribution of CK, which is highly skewed to the right.

(c) R output given below:

```
> m3 <- glm(GROUP~lck+H,family=binomial)
> summary(m3)

Call:
glm(formula = GROUP ~ lck + H, family = binomial)

Deviance Residuals:
     Min        1Q    Median        3Q       Max
-2.60372  -0.09903   0.16697   0.38782   1.89707

Coefficients:
            Estimate Std. Error z value Pr(>|z|)
(Intercept) 28.91300    5.80030   4.985 6.20e-07 ***
lck         -4.02041    0.82909  -4.849 1.24e-06 ***
H           -0.13652    0.03654  -3.736 0.000187 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

    Null deviance: 149.840  on 119  degrees of freedom
Residual deviance:  61.992  on 117  degrees of freedom
AIC: 67.992
```

(d) Wish to test H_0: $\beta_1 = \beta_2 = 0$ vs. H_a: some $\beta_j \neq 0$, $j = 1, 2$. The test statistic is

$$\chi^2 = 149.840 - 61.992 = \mathbf{87.848}$$

Under H_0, χ^2 follows a chi-squared distribution on $119 - 117 = 2$ degrees of freedom. The p-value for this test is **approximately zero**, so we reject H_0 and conclude that log(CK) and H are useful predictors.

```
> 1-pchisq(87.848,df=2)
[1] 0
```

(e) The difference in log odds is

$$(28.913 - 4.0204(\log(80)) - 0.1365(85)) - (28.913 - 4.0204(\log(300)) - 0.1365(100))$$

$$= 4.0204(\log(300) - \log(80)) + 0.1365(100 - 85) = 7.36$$

so the odds ratio is

$$e^{7.36} = \mathbf{1574.67}$$

The odds she is a carrier are **over 1,500 times** as great.