

Assignment 1 R

August 23, 2019

1 Question 1 a)

Estimation for $\hat{\beta}$ using log-likelihood.

$$\begin{aligned}\ell(\beta) &= \sum_{i=1}^N y_i \log\left(\frac{p_i}{1-p_i}\right) + m_i \log(1-p_i) \\ &= \sum_{i=1}^N y_i \log(p) + (m_i - y_i) \log(1-p)\end{aligned}$$

```
In [24]: library(faraway)
         data(orings)
```

```
N = 6
INITIAL_PARAMS = c(10, -0.05)
```

```
In [9]: F = function(eta) pnorm(eta)
        f = function(eta) dnorm(eta)
        iprobit = function(p) qnorm(p)
```

```
In [5]: logL = function(beta, data) {
        n = 6
        y = data$damage
        eta = cbind(1, data$temp) %*% beta
        p = probit(eta)
        return(sum(y * log(p) + (n - y)* log(1 - p)))
      }
```

```
beta.hat = optim(INITIAL_PARAMS, logL, data=orings, control=list(fnscale=-1))$par
beta0 = beta.hat[1]
beta1 = beta.hat[2]
```

Answer to Q1 a)

```
In [6]: cbind(beta0, beta1)
```

| beta0 | beta1 |
|----------|------------|
| 5.597882 | -0.1059069 |

2 Question 1b)

Fisher Information Matrix Let $\eta_i = x_i^T \beta = \beta_0 + \beta_1 x_i$, where β is the parameter vector and x_i is a vector of observed variables.

Probit function is defined as the Standard Normal CDF $p = F(\eta)$, and we take note of the some properties of the Standard Normal CDF:

$$\begin{aligned} F(-x) &= 1 - F(x) \\ F'(x) &= f(x) \\ f'(x) &= -xf(x). \end{aligned}$$

The Score Vector (first derivative) is:

$$\begin{aligned} U(\eta_i) &= \frac{\partial \ell(\beta_0, \beta_1)}{\partial \eta_i} \\ &= \sum_{i=1}^N \left[y_i \frac{f(\eta_i)}{F(\eta_i)} + y_i \frac{f(\eta_i)}{1 - F(\eta_i)} - m_i \frac{f(\eta_i)}{1 - F(\eta_i)} \right] \eta'_i \\ &= \sum_{i=1}^N \left[y_i \frac{f(\eta_i)}{F(\eta_i)} - (m_i - y_i) \frac{f(\eta_i)}{1 - F(\eta_i)} \right] \eta'_i, \end{aligned}$$

The Fisher Information (second derivative using quotient rule) is:

$$\begin{aligned} I(\eta_i) &= -\frac{\partial U(\eta_i)}{\partial \eta_i} \\ &= -\sum_{i=1}^N \left[y_i \frac{f(\eta)f(\eta) + \eta f(\eta)F(\eta)}{F(\eta)^2} - (m_i - y_i) \frac{-f(\eta)f(\eta) - \eta f(\eta)[1 - F(\eta)]}{[1 - F(\eta)]^2} \right] \eta'_i \eta_i'^T \end{aligned}$$

We also have,

$$\begin{aligned} \frac{\partial \eta_i}{\partial \beta_0} &= 1 \\ \frac{\partial \eta_i}{\partial \beta_1} &= x_i, \end{aligned}$$

where,

$$\eta'_i \eta_i'^T = X_i X_i^T.$$

Substitute $\hat{\eta}$ and $\hat{p} = E[y_i]$,

$$I(\hat{\eta}) = -\sum_{i=1}^N \left[6\hat{p} \frac{f(\hat{\eta})f(\hat{\eta}) + \hat{\eta}f(\hat{\eta})F(\hat{\eta})}{F(\hat{\eta})^2} - (m_i - 6\hat{p}) \frac{-f(\hat{\eta})f(\hat{\eta}) - \hat{\eta}f(\hat{\eta})[1 - F(\hat{\eta})]}{[1 - F(\hat{\eta})]^2} \right] X_i X_i^T \quad (1)$$

```
In [7]: X = cbind(1, orings$temp)
        x = orings$temp
        y = orings$damage
        eta = X %*% beta.hat
```

```
f.dash = function(eta) {
  constants = -eta / sqrt(2 * pi)
  exponential = (-eta^2)/2
  return(constants * exp(exponential))
}
```

```
In [8]: LEFT = (f.dash(eta) * F(eta) - f(eta)^2) / (F(eta)^2)
        RIGHT = (f.dash(eta) * (1 - F(eta)) + f(eta)^2) / (1 - F(eta))^2

        I11 = (N * F(eta) * LEFT) + (N * F(eta) - N) * RIGHT
        I12 = I21 = x %*% I11
        I22 = x^2 %*% I11

        I = -matrix(c(sum(I11), sum(I21), sum(I12), sum(I22)), 2, 2)
```

```
In [9]: I
```

```
30.16329    1931.384
1931.38421  125084.505
```

```
In [10]: Iinv = solve(I)
```

```
beta0.error = sqrt(Iinv[1,1])
beta1.error = sqrt(Iinv[2,2])
```

```
In [11]: Iinv
```

```
2.92831155  -0.0452149904
-0.04521499  0.0007061428
```

```
In [12]: z.alpha = qnorm(0.975)
```

```
CI.beta0 = beta0 + c(-1, 1) * z.alpha * beta0.error
CI.beta1 = beta1 + c(-1, 1) * z.alpha * beta1.error
```

Answer to Q1 b)

```
In [13]: rbind(CI.beta0)
        rbind(CI.beta1)
```

```
CI.beta0 | 2.243931  8.951834
CI.beta1 | -0.1579897 -0.05382405
```

3 Question 1c)

LRT for the temperature parameter β_1

```
In [14]: p.hat = sum(y) / sum(rep(n, length(y)))

F.model = logL(beta.hat, orings)
R.model = sum(y)*log(p.hat) + sum(N - y)*log(1 - p.hat)

LR = -2 * (R.model - F.model)
```

Answer to Q1 c)

```
In [15]: pchisq(LR, df=1, lower=FALSE)
```

```
5.18671169167522e-06
```

At the $\alpha = 0.05$ significance level, the LRT concludes that the β_1 parameter is significantly relevant.

4 Question 1d)

Estimate when temperature = $31^\circ F$. Let $t^T = [1 \ 31]$.

```
In [16]: TEMP = 31
eta = beta0 + beta1 * TEMP

t = c(1, 31)
damage.estimate = F(eta)

CI.ttbeta = F(eta + c(-1, 1) * c(z.alpha * sqrt(t(t) %*% Iinv %*% t)))
```

Answer to Q1 d)

```
In [17]: rbind(damage.estimate)
rbind(CI.ttbeta)

damage.estimate | 0.9896872
CI.ttbeta | 0.7115098 0.9999767
```

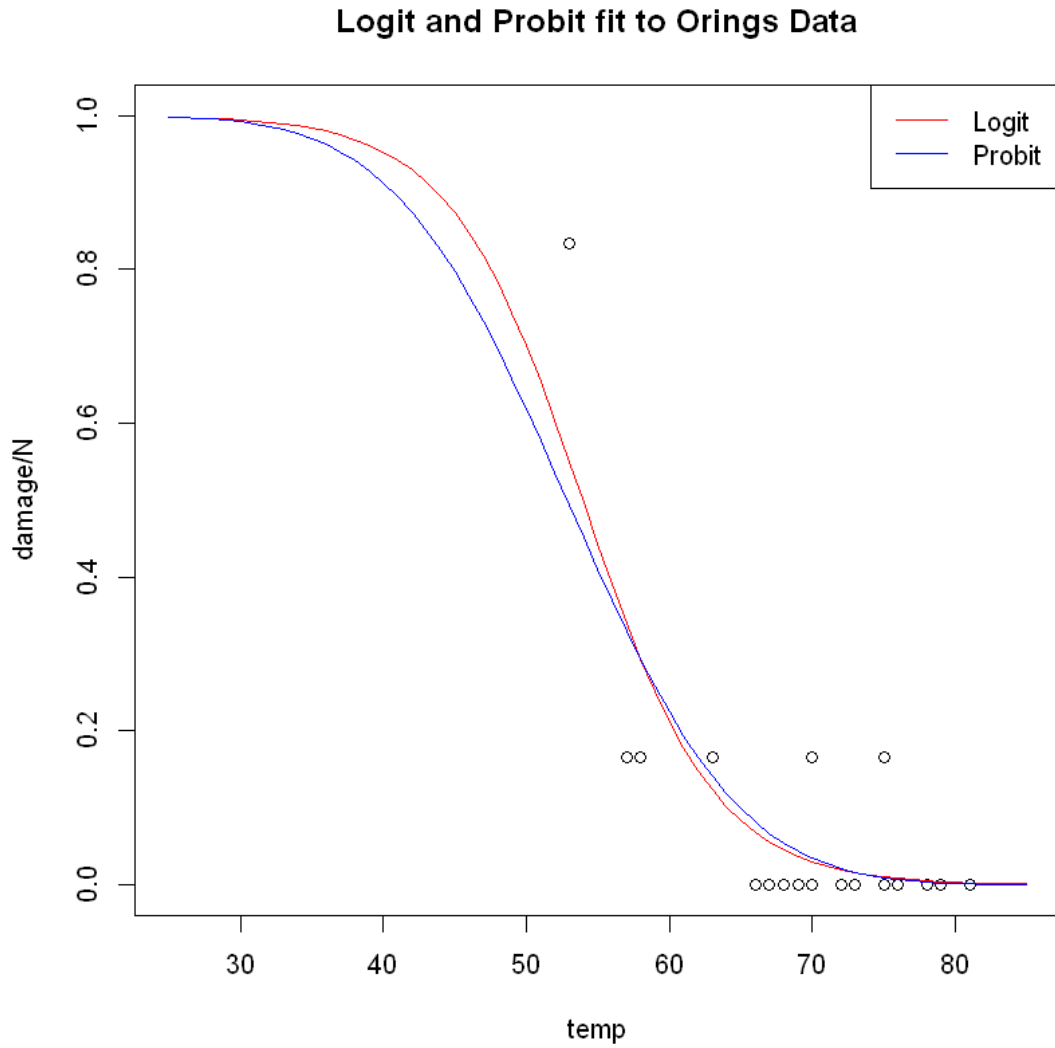
5 Question 1e)

Answer to Q1 e)

```
In [18]: logit = function(eta) exp(eta) / (1 + exp(eta))

logit.beta = glm(cbind(damage, N - damage) ~ temp, family=binomial(link='logit'), orings)
plot(damage/N ~ temp, orings, xlim=c(25,85), ylim=c(0,1), main="Logit and Probit fit t")

xrange = 25:85
lines(xrange, logit(logit.beta[1] + logit.beta[2] * xrange), col='red')
lines(xrange, F(beta0 + beta1 * xrange), col='blue')
legend("topright", c("Logit", "Probit"), col=c("red", "blue"), lty=c(1,1))
```



6 Question 2a)

We have,

$$\begin{aligned}\log(o_i) &= \log\left(\frac{p_i}{1 - p_i}\right) \\ &= \beta_0 + \beta_1 x_i.\end{aligned}$$

Hence an increase in 10 for bmi is:

$$\begin{aligned}\log(o)_2 - \log(o)_1 &= [\beta_0 + \beta_1(x_i + 10)] - [\beta_0 + \beta_1 x_i] \\ &= 10\beta_1\end{aligned}$$

```
In [24]: load("assign1.Robj", verbose=TRUE)
```

```
INCREASE = 10  
odds = function(p) p / (1 - p)
```

Loading objects:
pima_subset

```
In [25]: br = glm(cbind(test, 1 - test) ~ ., data=pima_subset, family=binomial)  
beta.hat = br$coeff  
  
beta0 = beta.hat[1]  
beta1 = beta.hat[2]
```

Answer to Question 2 a)

```
In [26]: INCREASE * beta1
```

bmi: 0.997168404895279

Estimated increase in log(odds) is 0.9972 when the bmi increase by 10.

7 Question 2b)

```
In [34]: # Standard Error is sqrt(variance)  
ilogit = function(eta) exp(eta) / (1 + exp(eta))  
x = pima_subset$bmi  
eta = beta0 + beta1 * x  
  
p.hat = ilogit(eta)  
  
I11 = sum(p.hat * (1 - p.hat))  
I12 = I21 = sum(x * p.hat * (1 - p.hat))  
I22 = sum(x^2 * p.hat * (1 - p.hat))  
  
Iinv = solve(matrix(c(I11, I12, I21, I22), 2, 2))  
  
beta1.error = sqrt(Iinv[2,2])
```

```
In [36]: CI.10beta1 = 10 * beta1 + c(-1, 1) * z.alpha * 10 * beta1.error
```

Answer to Question 2 b)

```
In [37]: rbind(CI.10beta1)
```

CI.10beta1 | 0.6976052 1.296732