HW5_solution

Problem 1 (a)

```
shuttle = read.table("Shuttle.csv", header = TRUE, fill = TRUE, sep=",")
shuttle$ThermalDistress = as.factor(shuttle$ThermalDistress)
summary(shuttle)
```

```
##   Temperature    ThermalDistress
##  Min.   :53.00   0:16
##  1st Qu.:67.00   1: 7
##  Median :70.00
##  Mean   :69.57
##  3rd Qu.:75.00
##  Max.   :81.00
```

```
fit_prob1 = glm(formula = ThermalDistress ~ ., family = binomial(link = "logit"), data = shuttle)
summary(fit_prob1)
```

```
##
## Call:
## glm(formula = ThermalDistress ~ ., family = binomial(link = "logit"),
##     data = shuttle)
##
## Deviance Residuals:
##     Min       1Q   Median       3Q      Max
## -1.0611  -0.7613  -0.3783   0.4524   2.2175
##
## Coefficients:
##             Estimate Std. Error z value Pr(>|z|)
## (Intercept)  15.0429     7.3786   2.039   0.0415 *
## Temperature  -0.2322     0.1082  -2.145   0.0320 *
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## (Dispersion parameter for binomial family taken to be 1)
##
##     Null deviance: 28.267  on 22  degrees of freedom
## Residual deviance: 20.315  on 21  degrees of freedom
## AIC: 24.315
##
## Number of Fisher Scoring iterations: 5
```

(b) -0.2322; one unit increment in the temperature would result in an increment of $exp(-0.2322) = 0.7927875$ to the odds ratio.

(c)

```
confint(fit_prob1)
```

```
## Waiting for profiling to be done...
```

```
##                   2.5 %      97.5 %
## (Intercept)  3.3305848 34.34215133
## Temperature -0.5154718 -0.06082076
```

There is 95% chance that the true β_1 would be in between [-1.135886, 4.1795680].

(d)

```
newdata = data.frame(Temperature = 31)
predict(fit_prob1, newdata, type="response")
```

```
##         1
## 0.9996088
```

(e)

```
newdata2 = data.frame(Temperature = seq(60,80,by=0.1))
y = predict(fit_prob1, newdata2, type="response")
```

When the temperature is about 64.8 F, the probability is about .5.

Problem 2 (a)

```
adolescent= read.table("adolescent.csv", header = TRUE, fill = TRUE, sep=",")
```

```
## Warning in read.table("adolescent.csv", header = TRUE, fill = TRUE, sep =
## ","): incomplete final line found by readTableHeader on 'adolescent.csv'
```

```
summary(adolescent)
```

```
##     Race      Gender       Yes              No
##  Black:2   Female:2   Min.   :22.0   Min.   : 23.00
##  White:2   Male  :2   1st Qu.:25.0   1st Qu.: 32.75
##                       Median :27.5   Median : 85.00
##                       Mean   :30.0   Mean   : 85.50
##                       3rd Qu.:32.5   3rd Qu.:137.75
##                       Max.   :43.0   Max.   :149.00
```

```
fit_prob2 = glm(cbind(Yes,No)~factor(Race)+factor(Gender), family = binomial, data = adolescent)
summary(fit_prob2)
```

```
##
## Call:
## glm(formula = cbind(Yes, No) ~ factor(Race) + factor(Gender),
##     family = binomial, data = adolescent)
##
## Deviance Residuals:
##        1         2         3         4
## -0.08867   0.10840   0.14143  -0.13687
##
```

```
## Coefficients:
##                     Estimate Std. Error z value Pr(>|z|)
## (Intercept)          -0.4555     0.2221  -2.050  0.04032 *
## factor(Race)White    -1.3135     0.2378  -5.524 3.32e-08 ***
## factor(Gender)Male    0.6478     0.2250   2.879  0.00399 **
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## (Dispersion parameter for binomial family taken to be 1)
##
##     Null deviance: 37.516984  on 3  degrees of freedom
## Residual deviance:  0.058349  on 1  degrees of freedom
## AIC: 25.186
##
## Number of Fisher Scoring iterations: 3
```

(b)

```
confint(fit_prob2)
```

```
## Waiting for profiling to be done...
```

```
##                         2.5 %      97.5 %
## (Intercept)        -0.8971266 -0.02385449
## factor(Race)White  -1.7824267 -0.84865350
## factor(Gender)Male  0.2105773  1.09436472
```

The confidence interval is $[exp(0.2105773), exp(1.09436472)] = [1.23439, 2.987284]$.

(c) The confidence interval is $[exp(-1.7824267), exp(-0.84865350)] = [0.1682294, 0.4279908]$.

(d) Since the test statistic is 37.516984-0.058349 = 37.45864, which is greater than the critical value under the chi-squared distribution, we reject the null hypotheis.

(e) Since the p-value is smaller than α, we reject the null hypothesis.