Name Jerry

Student Number _____

STA 312f2012 Quiz 7

1. (5 Points) For a multiple logistic regression model, let $P(Y_i = 1|x_{i,1}, \ldots, x_{i,p-1}) = \pi(\mathbf{x}_i)$. Find an expression for $\pi(\mathbf{x}_i)$ in terms of the vector of regression parameters $\boldsymbol{\beta}$. Show your work.

$$\log \frac{\pi(x_i)}{1-\pi(x_i)} = x_i'\beta$$

$$\iff \frac{\pi(x_i)}{1-\pi(x_i)} = e^{x_i'\beta}$$

$$\iff \pi(x_i) = (1-\pi(x_i))e^{x_i'\beta}$$

$$\iff \pi(x_i) = e^{x_i'\beta} - \pi(x_i)e^{x_i'\beta}$$

$$\iff \pi(x_i)(1+e^{x_i'\beta}) = e^{x_i'\beta}$$

$$\iff \pi(x_i) = \frac{e^{x_i'\beta}}{1+e^{x_i'\beta}}$$

2. (5 points) This question is based on your printout for the bird lung data ***Write your answers in the spaces below.***

Controlling for all the other variables in the full model, is how many years the person has been smoking related to the chances of getting lung cancer?

(a) Give the value of the test statistic. Your answer is a single number from the printout.

$Z = 2.751$ 1 pt

(b) What is the p-value? The answer is a single number from your printout.

$P = 0.00594$ 1 pt

(c) Do you reject H_0 at $\alpha = 0.05$? Answer Yes or No.

Yes 1 pt

(d) In plain, non-statistical language, what do you conclude from this hypothesis test?

All other things being equal, the longer the person has been smoking, the greater the chances of cancer.

2 points - saying there's a relationship without saying what it is gets zero. Any mention of H_0 at least -1 pt

Attach your printout for Question 2 (Homework Question 7). Make sure your name is written on the printout.

312f12 Quiz 7 R Output

```
> attach(birthwt)
> race <- factor(race, labels = c("white", "black", "other"))
> fullmod = lm(bwt ~ age + lwt + race); summary(fullmod)

Call:
lm(formula = bwt ~ age + lwt + race)

Residuals:
     Min       1Q   Median       3Q      Max
-2103.50  -429.68    41.74   486.10  1902.20

Coefficients:
            Estimate Std. Error t value Pr(>|t|)
(Intercept) 2461.147    314.722   7.820 3.97e-13 ***
age            1.299     10.108   0.128  0.89789
lwt            4.620      1.788   2.584  0.01054 *
raceblack   -447.615    161.369  -2.774  0.00611 **
raceother   -239.357    115.189  -2.078  0.03910 *
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 704.9 on 184 degrees of freedom
Multiple R-squared: 0.08536,  Adjusted R-squared: 0.06548
F-statistic: 4.293 on 4 and 184 DF,  p-value: 0.00241

> # Switch contrasts to test Black vs other
> race2 = race
> contrasts(race2) = contr.treatment(3,base=3) # Other will be ref
> summary(lm(bwt ~ age + lwt + race2))

Call:
lm(formula = bwt ~ age + lwt + race2)

Residuals:
     Min       1Q   Median       3Q      Max
-2103.50  -429.68    41.74   486.10  1902.20
```

2a

2b

```
Coefficients:
             Estimate Std. Error t value Pr(>|t|)
(Intercept) 2221.791    293.446   7.571 1.73e-12 ***
age           1.299     10.108   0.128   0.8979
lwt           4.620      1.788   2.584   0.0105 *
race21      239.357    115.189   2.078   0.0391 *
race22     -208.258    170.455  -1.222   0.2234
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 704.9 on 184 degrees of freedom
Multiple R-squared: 0.08536,  Adjusted R-squared: 0.06548
F-statistic: 4.293 on 4 and 184 DF,  p-value: 0.00241

> # Race controlling for age and weight
> red1 = lm(bwt ~ age + lwt)
> anova(red1,fullmod)
Analysis of Variance Table

Model 1: bwt ~ age + lwt
Model 2: bwt ~ age + lwt + race
  Res.Df      RSS Df Sum of Sq      F   Pr(>F)
1    186 96186834
2    184 91436202  2   4750632 4.7799 0.009467 **
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
> # Age and weight controlling for race
> red2 = lm(bwt ~ race)
> anova(red2,fullmod)
Analysis of Variance Table

Model 1: bwt ~ race
Model 2: bwt ~ age + lwt + race
  Res.Df      RSS Df Sum of Sq      F  Pr(>F)
1    186 94953931
2    184 91436202  2   3517728 3.5394 0.03102 *
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
> # Race controlling for age and weight with a general linear test.
> # Need car. Compare F = 4.7799
> L = rbind(c(0,0,0,1,0),
+           c(0,0,0,0,1) )
```

```
> linearHypothesis(fullmod,L)
Linear hypothesis test

Hypothesis:
raceblack = 0
raceother = 0

Model 1: restricted model
Model 2: bwt ~ age + lwt + race

  Res.Df      RSS Df Sum of Sq      F   Pr(>F)
1    186 96186834
2    184 91436202  2   4750632 4.7799 0.009467 **
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
>
> # With White as the reference category, repeat the test comparing
> # Black to Other controlling for age and weight.
> # Compare  F = (-1.222)^2 = 1.493284
> L = rbind(c(0,0,0,1,-1))
> linearHypothesis(fullmod,L)
Linear hypothesis test

Hypothesis:
raceblack - raceother = 0

Model 1: restricted model
Model 2: bwt ~ age + lwt + race

  Res.Df      RSS Df Sum of Sq      F Pr(>F)
1    185 92177997
2    184 91436202  1    741795 1.4927 0.2234
>
```

Also okay for 2c