0/0

Math 484 Quiz 5 fall 2016 Name____________________

1) Suppose that the regression model is $Y_i = 10 + 2X_{1i} + \beta X_{2i} + e_i$ for $i = 1, ..., n$ where the e_i are iid random variables. The least squares criterion is $Q(\eta) = \sum_{i=1}^{n}(Y_i - 10 - 2X_{1i} - \eta X_{2i})^2$.

a) What is $E(Y_i)$?

$$\boxed{10 + 2X_{1i} + \beta X_{2i}}$$

b) Find the least squares estimator $\hat{\beta}$ of β by setting the first derivative $\frac{d}{d\eta}Q(\eta)$ equal to zero.

$$\frac{dQ}{d\eta} = 2\sum_{i=1}^{n}(Y_i - 10 - 2X_{1i} - \eta X_{2i})(-X_{2i}) \stackrel{set}{=} 0 \quad \text{or}$$

$$2\eta \sum_{i=1}^{n} X_{2i}^2 = 2\sum_{i=1}^{n} X_{2i}(Y_i - 10 - 2X_{1i})$$

$$\text{or} \quad \boxed{\hat{\beta} = \frac{\sum_{i=1}^{n} X_{2i}(Y_i - 10 - 2X_{1i})}{\sum_{i=1}^{n} X_{2i}^2} = \underbrace{\frac{\sum X_{2i} Y_i - 10\sum X_{2i} - 2\sum X_{1i} X_{2i}}{\sum X_{2i}^2}}_{\text{should not simplify}}}$$

better

c) Show that your $\hat{\beta}$ is the global minimizer of the least squares criterion Q by showing that the second derivative $\frac{d^2}{d\eta^2}Q(\eta) > 0$ for all values of η.

$$\frac{d^2 Q}{d\eta^2} = \frac{d}{d\eta} 2\eta \sum X_{2i}^2 = 2\sum_{i=1}^{n} X_{2i}^2 > 0$$

$$\left(= \frac{d}{d\eta}\left[2\eta \sum X_{2i}^2 - 2\sum X_{2i}(Y_i - 10 - 2X_{1i})\right] \right.$$

42

no intercept model

```
Coefficient Estimates
Label       Estimate        Std. Error     t-value    p-value
X           0.585616        0.0185600       31.553     0.0000
Sigma hat: 6.10219 Number of cases: 202 Degrees of freedom: 201
Summary Analysis of Variance Table
Source           df          SS             MS             F      p-value
Regression        1     37071.5        37071.5       995.56     0.0000
Residual        201     7484.59        37.2367
```

2) The above output is for a Cook and Weisberg (1999) *Arc* data set on 102 male and 100 female athletes collected at the Australian Institute of Sport. The response Y is the athlete's *body fat*. The variable X is BMI = body mass index = $weight/(height)^2$. Response and residual plots as well as a scatterplot of the data suggest that regression through the origin is a reasonable model.

a) Predict Y if $X = 17$.

$\hat{Y} = 0.585616(17) = \boxed{9.9555}$

b) Predict Y if $X = 34$.

$\hat{Y} = 0.585616(34) = \boxed{19.9109}$

c) Perform the t test for $\beta_1 = 0$.

$H_0\ \beta_1 = 0 \quad H_A\ \beta_1 \neq 0$

$t_0 = 31.553$

pval $= 0$

reject Ho, X = BMI is needed in the MLR model for Y = bodyfat

d) Perform the no intercept Anova F test.

$H_0\ \beta_1 = 0 \quad H_A\ \beta_1 \neq 0$

$F_0 = 995.56$

pval $= 0$

reject Ho there is an MLR relationship between Y = body fat and X = BMI.

56