Family (Last) Name ______________

Given (First) Name Jerry

Student Number ______________

STA 302s13 Quiz 11A

1. (4 points) Consider the following *uncentered* regression model with no intercept.

$$Y_i = \beta_1 x_{i1} + \cdots + \beta_k x_{ik} + \epsilon_i,$$

In the centered version of this model, is the intercept necessarily equal to zero? If the answer is yes, prove it. If the answer is no, give the value of the intercept, which could be called α_0. In either case, show your work. You have more room than you need.

No, the intercept need not be zero.

$$Y_i = \beta_1 (x_{i1} - \bar{x}_1) + \cdots + \beta_k (x_{ik} - \bar{x}_k)$$
$$+ \beta_1 \bar{x}_1 + \cdots + \beta_k \bar{x}_k,$$

So the intercept of the centered model is

$$\alpha_0 = \beta_1 \bar{x}_1 + \cdots + \beta_k \bar{x}_k$$

centered

2. In your analysis of the Census Tract data, the independent variables were `region` and `docs` (with interactions) and the dependent variable was crime rate. For `region`, 1=Northeast, 2=North Central, 3=South and 4=West. The reference category was Northeast.

(a) (2 points) For each region, give the estimated expected crime rate for census tracts with an average (sample mean) number of doctors. Write numbers in the spaces below.

Northeast: 41.26

North Central: 41.26 + 11.61 = 52.87

South: 41.26 + 17.21 = 58.47

West: 41.26 + 28.69 = 69.95

(b) (1 point) You want to know whether, for census tracts with an average (sample mean) number of doctors, there is a diference in expected crime rate between the Northeast and North Central regions. Write the t or F statistic and the p-value (both numbers from your printout) in the table below.

Test Statistic (F or t)	p-value
$t = 4.108$	0.0000693

(c) (1 point) You want to know whether, for census tracts with an average (sample mean) number of doctors, there is a diference in expected crime rate between the North Central and South regions. Write the t or F statistic and the p-value (both numbers from your printout) in the table below.

Test Statistic (F or t)	p-value
$t = -2.27$	0.0247

(d) (1 point) You want to know whether the regression lines for the Northeast and West regions are parallel. Write the t or F statistic and the p-value (both numbers from your printout) in the table below.

Test Statistic (F or t)	p-value
$t = -1.219$	0.22483

(e) (1 point) You want to know whether the regression lines for the four regions are parallel. Write the t or F statistic and the p-value (both numbers from your printout) in the table below.

Test Statistic (F or t)	p-value
$F = 0.8395$	0.4745

```
> # R work for STA302f13 Assignment 11
> rm(list=ls())
> census =
read.table("http://www.utstat.toronto.edu/~brunner/302f13/code_n_data/hw/CensusTract.data")
> attach(census)
> cdocs = docs-mean(docs)
> crimerate = crimes/pop
> region=factor(region,labels=c("NE","NC","S","W" ))
> fullmod = lm(crimerate ~ cdocs + region + cdocs:region) # I didn't show them this easy way.
> summary(fullmod)

Call:
lm(formula = crimerate ~ cdocs + region + cdocs:region)

Residuals:
     Min       1Q   Median       3Q      Max 
-29.9595  -7.3935  -0.3664   8.7014  24.1772 

Coefficients:
                 Estimate Std. Error t value Pr(>|t|)    
(Intercept)    41.2597926  2.1320421  19.352  < 2e-16 ***
cdocs           0.0013754  0.0004247   3.239  0.00152 ** 
regionNC       11.6136999  2.8270582   4.108 6.93e-05 ***
regionS        17.2139479  2.6790015   6.426 2.16e-09 ***
regionW        28.6894951  2.9834655   9.616  < 2e-16 ***
cdocs:regionNC -0.0004296  0.0008725  -0.492  0.62329    
cdocs:regionS   0.0008174  0.0011545   0.708  0.48015    
cdocs:regionW  -0.0009758  0.0008002  -1.219  0.22483    
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1 

Residual standard error: 10.97 on 133 degrees of freedom
Multiple R-squared: 0.4508, Adjusted R-squared: 0.4219 
F-statistic:  15.6 on 7 and 133 DF,  p-value: 8.1e-15 

> 
> # Remaining pairwise diffs at xbar
> betahat = fullmod$coefficients; betahat
   (Intercept)          cdocs       regionNC        regionS        regionW cdocs:regionNC 
 41.2597926415   0.0013753725  11.6136998625  17.2139478594  28.6894951429  -0.0004295946 
 cdocs:regionS  cdocs:regionW 
  0.0008174386  -0.0009757715 
> dfe = fullmod$df.residual
> V = vcov(fullmod)
> 
> #   NC vs S
> a = rbind(0,0,1,-1,0,0,0,0)
> T = as.numeric( t(a)%*%betahat/sqrt(t(a)%*%V%*%a) ); T
[1] -2.271555
> p = 2*(1-pt(abs(T),dfe)); p
[1] 0.02471922
> 
> #   NC vs W
> a = rbind(0,0,1,0,-1,0,0,0)
> T = as.numeric( t(a)%*%betahat/sqrt(t(a)%*%V%*%a) ); T
[1] -6.113286
> p = 2*(1-pt(abs(T),dfe)); p
[1] 1.01378e-08
> 
> #   S vs W
```

```
> a = rbind(0,0,0,1,-1,0,0,0)
> T = as.numeric( t(a)%*%betahat/sqrt(t(a)%*%V%*%a) ); T
[1] -4.341411
> p = 2*(1-pt(abs(T),dfe)); p
[1] 2.782269e-05
>
> # Test 4 slopes equal
> parallel = lm(crimerate ~ cdocs + region)
> anova(parallel,fullmod)
Analysis of Variance Table

Model 1: crimerate ~ cdocs + region
Model 2: crimerate ~ cdocs + region + cdocs:region
  Res.Df   RSS Df Sum of Sq      F Pr(>F)
1    136 16320
2    133 16017  3     303.3 0.8395 0.4745
>
```

Family (Last) Name ______________________

Given (First) Name Jerry

Student Number ______________________

STA 302s13 Quiz 11B

Consider the following model with random independent variables. Independently for $i = 1, \ldots, n$,

$$\begin{aligned} Y_i &= \alpha + \beta_1 X_{i1} + \cdots + \beta_k X_{ik} + \epsilon_i \\ &= \alpha + \boldsymbol{\beta}' \mathbf{X}_i + \epsilon_i, \end{aligned}$$

where

$$\mathbf{X}_i = \begin{pmatrix} X_{i1} \\ \vdots \\ X_{ik} \end{pmatrix} \text{ with } E(\mathbf{X}_i) = \boldsymbol{\mu}_x,\ cov(\mathbf{X}_i) = \boldsymbol{\Sigma}_x,$$

and $\mathbf{X}_i$ is independent of ϵ_i. Let $\boldsymbol{\Sigma}_{xy}$ denote the $k \times 1$ matrix of covariances between Y_i and X_{ij} for $j = 1, \ldots, k$. Calculate $\boldsymbol{\Sigma}_{xy} = C(\mathbf{X}_i, Y_i)$, obtaining $\boldsymbol{\Sigma}_{xy} = \boldsymbol{\Sigma}_x \boldsymbol{\beta}$.

$$C(X_i, Y_i) = E\{(X_i - \mu_x)(Y_i - E(Y_i))'\}$$

$$= E\{(X_i - \mu_x)(\alpha + \beta' X_i + \varepsilon_i - \alpha - \beta' \mu_x)'\}$$

$$= E\{(X_i - \mu_x)(\beta'(X_i - \mu_x) + \varepsilon_i)'\}$$

$$= E\{(X_i - \mu_x)(\beta'(X_i - \mu_x))'\} + E\{(X_i - \mu_x)\varepsilon_i'\}$$

$$= E\{(X_i - \mu_x)(X_i - \mu_x)'\beta\} + E\{X_i - \mu_x\} E\{\varepsilon_i\}$$ Independence.

$$= E\{(X_i - \mu_x)(X_i - \mu_x)'\}\beta + \underset{\sim}{0} \cdot 0$$

$$= \Sigma_x \beta$$