Math 583 old Quiz 6 Name___________

Table 1: Number of Times All Outlier Distances > Clean Distances, otype=3

n	p	γ	steps	pm	covmb2	diag
100	1000	0.25	0	8	5	86
100	500	0.25	9	10	1	92
100	100	0.4	1	11.9	71	85

1) The above table gives counts for the number of times all outlier distances were greater than all clean data distances for two estimators `covmb2` and `diag`. The `covmb2` estimator uses $D(T_{covmb2}, \boldsymbol{I}_p)$ while the `diag` estimator uses $D(T_{covmb2}, diag(\boldsymbol{C}_{covmb2}))$. Higher counts are better than lower counts. The outlier proportion is γ, and the larger pm is, the farther the outliers are from the clean data. For outliers of the type used in the simulation that produced the table, which estimator is better?

Soln: diag

2) Assuming the assumptions of the least squares central limit theorem hold, what is the limiting distribution of $\sqrt{n}\ \ (\hat{\boldsymbol{\beta}} - \boldsymbol{\beta})$ if $(\boldsymbol{X}'\boldsymbol{X})/n \to \boldsymbol{W}^{-1}$ as $n \to \infty$?

soln:

$$\sqrt{n}\ \ (\hat{\boldsymbol{\beta}} - \boldsymbol{\beta}) \xrightarrow{D} N_p(\boldsymbol{0}, \sigma^2 \boldsymbol{W})$$

```
Label      Estimate        Std. Error    t-value    p-value
Constant  -2.93739         1.42523        -2.061     0.0422
mlife      1.12359         0.0229362      48.988     0.0000

R Squared:  0.96424 Sigma hat: 2.11667
Number of cases: 91 Degrees of freedom: 89

Summary Analysis of Variance Table
Source         df       SS            MS           F    p-value
Regression      1   10751.8       10751.8     2399.80    0.0000
Residual       89   398.746       4.48029
```

3) The output above is for predicting $Y =$ *female life expectancy* from $X_2 =$ *male life expectancy*.

a) Predict Y if $X_2 = 67$.

soln: $\hat{Y} = -2.93739 + 1.12359(67) = 72.343$

b) Find the residual if $Y = 75.000$

soln: $r = Y - \hat{Y} = 75 - 72.343 = 2.657$