STATS 500, HOMEWORK #4, due Wednesday, Feb, 10th

```
set.seed(48109711)
n=200; x=rnorm(2*n); x1=x[1:n]; x2=x[(n+1):(2*n)]; x3=.7*x1+.9*x2+rnorm(n)*.1
beta=c(1, 1, 1, 2)
y1=beta[1]+beta[2]*x1+beta[3]*x2+rnorm(n)*.3
y2=beta[1]+beta[2]*x1+beta[3]*x2+beta[4]*x3+rnorm(n)*.3
dat=data.frame(y1=y1,y2=y2,x1=x1,x2=x2,x3=x3)
outx12 = lm(y1~x1+x2, data=dat)
summary(outx12)
outx123 = lm(y1~x1+x2+x3, data=dat)
library(ellipse)
## Plot the confidence region
par(mfrow=c(1,2))
plot(ellipse(outx12, c("x1", "x2")), type="l")
plot(ellipse(outx123, c("x1", "x2")), type="l")
## Add the estimates to the plot
D1= model.matrix(outx12);D2= model.matrix(outx123);
cov(x1,x2);
V1=solve(t(D1)%*%D1); V2=solve(t(D2)%*%D2);
round(V1[2:3,2:3],4); round(V2[2:3,2:3],4);
```

1. In the codes given above, we generate $n = 200$ $x1$, $x2$ and $x3$, and $x1$, $x2$ are independent of each other. In model A, we generate $y1$ with the predictors $x1$ and $x2$. In model B, we generate $y2$ with all three predictors. In the two plots we generate, we obtain the 95% confidence regions for the coefficients, β_{x1}, β_{x2} from each of the two model. Note that even you generate $x1$ $x2$ independently, the sample correlation still may not be 0, but it tends to be small. Consequently, the major axis and minor axis of the ellipse are roughly parallel with the x-axis and y-axis corresponding to $\hat{\beta}_{x1}$, $\hat{\beta}_{x2}$. In model B, you are using the same $x1$ and $x2$ and you still consider the 95% confidence ellipse for β_{x1}, β_{x2}, do you expect the major axis and minor axis of the ellipse are roughly parallel with the x-axis and y-axis corresponding to $\hat{\beta}_{x1}$, $\hat{\beta}_{x2}$? Please briefly explain your answer using the output you obtain. Basically, in a multiple regression, unless the predictors are completely orthogonal to each other, the estimated coefficients and the

corresponding inference are adjusted to the existence of other predictors.

2. Using the `sat` data as in the previous homework and fit a model with `total` sat score as the response and `takers, ratio` and `salary` as predictors. Let $\alpha = .05$. Re-test H_{A1}: $\beta_{ratio} \neq 0$ and H_{A2}: $\beta_{salary} \neq 0$ by using a permutation test.

3. As above, but now use the permutation test to test H_0: $\beta_{ratio} = \beta_{salary} = 0$ vs the H_A that at least one of the two coefficients are not 0 by a 0.05 permutation test.