STATS 500, HOMEWORK #3, due Wednesday, Oct. 6, 3 PM

1. Using the dataset `gala` we discussed in class. Consider a regression model with "Endemics" as the response and "Area", "Elevation", "Nearest", "Scruz", "Adjacent" as predictors.

   (a) What would be the H_0 and H_A if you wish to claim that an island with a large highest elevation level tends to have more endemic species. What would be the test statistic, p-value and conclusion for your test if α-level is 0.01?

   (b) For the regression model above, find 99% confidence intervals for $\beta_{Elevation}$ and $\beta_{Nearest}$, respectively.

   (c) For $\alpha = 0.05$, conduct a test for H_0 $\beta_{Nearest} = \beta_{Scruz} = 0$. What would be the p-value for this test? Based on your analysis, do you feel any of these predictors have an effect on the response? Without drawing the 95% simultaneous confidence region for $(\beta_{Nearest}, \beta_{Scruz})$, please make a guess whether $(0, 0)$ would be inside this confidence region or not. Briefly explain your answer.

2. Use the `sat` data (see `help(sat)` for the description of variables). Fit a model with `total` sat score as the response and `takers, ratio` and `salary` as predictors. Answer the following question using the output provided here:

```
> var(sat$total)
[1] 5598.116
> tmp=lm(total~takers+ratio+salary, sat)
> summary(tmp)
Call:
lm(formula = total ~ takers + ratio + salary, data = sat)
Residuals:
    Min      1Q  Median      3Q     Max
-89.244 -21.485  -0.798  17.685  68.262
Coefficients:
             Estimate Std. Error t value Pr(>|t|)
(Intercept) 1057.8982    44.3287  23.865   <2e-16 ***
takers        -2.9134     0.2282 -12.764   <2e-16 ***
```

```
ratio         -4.6394     2.1215  -2.187   0.0339 *
salary         2.5525     1.0045   2.541   0.0145 *
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
Residual standard error: 32.41 on 46 degrees of freedom
Multiple R-squared:  XX
```

(a) What would be the H_0 and H_A if you wish to claim that a higher value of average pupil/teacher ratio (ratio) tends to lead to a lower sat score. What would be the numerical value of the test statistic, p-value and conclusion for your test if α-level is 0.01?

(b) Let σ^2 denote the variance of random errors in the regression model (model tmp in R) , based on the output, what should be the estimates of σ^2 and R^2 (XX value in Multiple R-squared) – you do not need to carry out the calculation but make sure that I can get the correct numbers using your answers and a plain calculator.

3. Use the `sat` data and fit a model with `total` sat score as the response and `takers, ratio` and `salary` as predictors. Let $\alpha = .05$. Conduct a test with H_A: $\beta_{ratio} \neq 0$ by using a permutation test and report the testing result. Using the same permutation outcomes, what would be the p-value for the test you consider in Problem 2(a) above?