This worksheet deals with scores on exams in Calculus III: `http://web02.gonzaga.edu/faculty/axon/scores.txt`.

1. I'd like to know if final exams really matter in the calculation of students' grades. Enter the exam scores into R and fit a multiple linear regression line to the data:

```
e1 = c(0.79,0.62,...,0.73)
e2 = c(0.96,0.68,...,0.83)
⋮
final = c(0.8935,0.6389,...,0.7315)
model = lm(final ~ e1 + e2 + e3 + e4)
```

Use `qqnorm()` to check that it's reasonable to assume that all the exam scores are normally distributed. Finally, explore `model$residuals` and decide if final exams are important.

2. The command `summary(model)` should give you a list of P-values. If the independent variables in the model (the scores on exams 1-4) were statistically independent, then these would tell you which exams (if any) were correlated with final exam score (the null hypothesis is that they aren't). Compare these P-values with the P-values coming from simple linear regression models (e.g. `lm(final~e1)`). Are any of the independent variables actually not correlated with the final exam score?

3. It seems like we could also include homework scores in our regression model. The catch is that most of our regression statistics assume that populations are normally distributed. Would it be reasonable to assume that homework scores are normally distributed?

4. Now we're going to test the null hypothesis that scores on exams don't depend on the student or on the exam. This can be tested using Analysis of Variance (ANOVA). To implement this in R, we must put all of our data in a frame:

```
student=c("sa","la",...,"vv")
scores=c(e1, e2, e3, e4, final)
scores.df=data.frame(Student=rep(student,5), Exam=c(rep("E1",23), rep("E2",23), rep("E3",23),
rep("E4",23), rep("Final",23)), Score=scores)
```

Now we can get R to do the necessary calculations and show us the results:

```
scores.aov=aov(Score ~ Student + Exam, data=scores.df)
summary(scores.aov)
```

What does it all mean?