1. Answer all questions. Please show the steps in your reasoning and in your calculations.

2. You have two hours to complete this exam.

3. You may use a calculator, the tables of distributions in the back of your text, and 3 formula sheets.

4. Carry out any tests at the 0.05 level of significance.

5. Your test should contain 8 distinct pages (including the one you’re now reading).

6. Point values are given in parentheses.

7. Good luck and have a great summer!

If you would like your grade posted at the course website by a random number that I will provide you with, please sign below. I cannot tell you your grade over the phone or by email. Also, if you lose the random number I give you, I cannot tell you the number by phone or email.

Signature ________________________________
In questions 1-4, please circle the correct answer.

1. (4) A polling agency places registered Texas voters into the distinct categories Republican, Democrat, and Other. To gather information on voting patterns, the agency randomly selects 1000 registered voters from each of these three categories. This an example of
   (a) stratified random sampling.
   (b) single-stage cluster sampling.
   (c) two-stage cluster sampling.
   (d) judgement sampling.
   (e) where a census would have been a much less expensive way of soliciting opinions.

2. (4) A motivation for stratified sampling is that
   (a) it does not involve any random sampling.
   (b) it does not require a frame for the entire population.
   (c) in certain cases it leads to more precise estimation of the population mean.
   (d) all of the above are true.
   (e) it has been shown to be the most effective way to select a good piece of candy from one of those boxes of chocolates that have lots of yucky ones filled with white cream.

3. (4) A large aquarium in a certain city contains 300 fish tanks. To learn about the condition of the fish population at this aquarium, 20 tanks are randomly selected, and then 5 fish are randomly selected from each of these tanks. (Each tank contains more than 5 fish.) This scheme is
   (a) an example of stratified sampling.
   (b) an example of single-stage cluster sampling.
   (c) an example of two-stage cluster sampling.
   (d) an example of judgement sampling.
   (e) how Pixar Studios came up with the storyline for Finding Nemo.

4. (4) The modern statistical method known as the bootstrap is
   (a) not useful unless the data are normally distributed.
   (b) only useful when extremely large sample sizes are available.
   (c) useful in situations where the assumption of normally distributed errors may be violated.
   (d) a means of dealing with extrapolation problems in regression.
   (e) a hip fashion accessory for sartorially challenged statisticians.
5. The skin response $Y$ (in mm) of rats to different concentrations $X$ (in ml/l) of a newly developed vaccine was measured in an experiment. Polynomial models of orders one through three were fit to these data.

(a) A plot of the original data and residual plots for the linear, quadratic, and cubic models are given below. Based on these plots, which model is most appropriate and why?

Use the accompanying SPSS output to answer the remaining parts of this question. The variables $x_2$ and $x_3$ denote $x^2$ and $x^3$, respectively.

(b) Carry out a test to determine whether the cubic model improves significantly on the linear model.
Use the cubic model to answer the remaining parts of this question.

(c) (6) What proportion of the variance in skin response is explained by the cubic relationship with concentration?

(d) (6) Estimate the mean skin response for rats exposed to a concentration of 2 ml/l.

(e) (6) What is the estimated standard deviation of the error terms?

(f) (6) Would one be justified in using this model to estimate the mean skin response for rats exposed to a concentration of 4 ml/l? Explain why or why not.
6. Three different regions on the coast of California are being considered as possible locations for a liquid natural gas (LNG) terminal. To learn about public opinion, three separate random samples of individuals are obtained, one from each region. Each individual is asked whether he/she favors having an LNG terminal built in the region where the individual resides. The observed cell counts are given in the following table:

<table>
<thead>
<tr>
<th>Regions</th>
<th>Favor</th>
<th>Oppose</th>
<th>Sample size</th>
</tr>
</thead>
<tbody>
<tr>
<td>A</td>
<td>198</td>
<td>202</td>
<td>400</td>
</tr>
<tr>
<td>B</td>
<td>140</td>
<td>210</td>
<td>350</td>
</tr>
<tr>
<td>C</td>
<td>133</td>
<td>217</td>
<td>350</td>
</tr>
</tbody>
</table>

(a) (12) Carry out a test of the hypothesis that the proportions of people who favor the LNG project are the same in the three regions.

(b) (10) Construct a 95% confidence interval for the difference between the proportions of individuals in Regions A and C who favor the LNG project.
Short-run supermarket strategies are designed to temporarily increase unit sales of particular products. Two factors in a particular study were Price level (regular, reduced, and at cost to the supermarket) and Display level (normal display space and twice the normal space). A complete factorial experiment with the Price and Display factors was carried out. Each treatment was applied during three randomly selected weeks to a particular product at a particular supermarket, and the unit sales for all three weeks were recorded. To minimize treatment carryover effects, each treatment was preceded and followed by a week in which the product was priced at its regular price and was displayed in its normal manner. The order in which the treatments were applied was random.

(a) (6) How many treatments are there, and what are they?

(b) (15) Use the accompanying SPSS output to perform a complete analysis of the data from this experiment. What conclusions can you make about the effects of different marketing strategies on sales? (Note: Should you need it, Fisher’s 0.05 level LSD for comparing treatment means is 288.3.)