1. There are 8 problems and a total of 9 pages in this booklet.

2. Show all of your work and label your answers clearly.

3. If you use a table, state the table used: for example, 1.887 (from Table A).

4. If you use a function on the TI 83 (or TI 89), write out the command you entered as well as the result: for example, 0.0668 (normalcdf (-10, -1.5, 0.1)).

<p>| | |</p>
<table>
<thead>
<tr>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td></td>
</tr>
<tr>
<td>2</td>
<td></td>
</tr>
<tr>
<td>3</td>
<td></td>
</tr>
<tr>
<td>4</td>
<td></td>
</tr>
<tr>
<td>5</td>
<td></td>
</tr>
<tr>
<td>6</td>
<td></td>
</tr>
<tr>
<td>7</td>
<td></td>
</tr>
<tr>
<td>8</td>
<td></td>
</tr>
<tr>
<td>Total</td>
<td></td>
</tr>
</tbody>
</table>
1. (12 pts) Suppose that a student measuring the boiling temperature of a certain liquid observes the six readings and obtained the sample mean of 101.8 (in degrees Celsius). Also assume that the standard deviation for this procedure (σ) is 1.2 degrees.

(a) Construct a 95% confidence interval for the mean boiling temperature of this liquid.

(b) Suppose that you were dissatisfied with the width of the confidence interval, and wanted to cut the margin of error to ±0.5, keeping the same confidence level of 95%. How many readings (total) would have to be included in the new sample?
2. (13 pts) The level of calcium in the blood of healthy young adults follows a normal distribution with mean $\mu = 10$ milligrams per deciliter and standard deviation $\sigma = 0.5$. A clinic measures the blood calcium of 27 healthy pregnant young women at their first visit for prenatal care. The mean of these 27 measurements is $\bar{x} = 9.8$. Is this evidence that the mean calcium level in the population of healthy pregnant young women is less than 10?

(a) Set up the null and alternative hypotheses that you will be testing regarding the above question.

(b) Assuming that the distribution of calcium level measurements for pregnant young women is a normal distribution with $\sigma = 0.5$, find the value of the test statistic and compute the P-value.

(c) Is this statistically significant (circle your answers)

at the 5% level? (Yes, No); at the 1% level? (Yes, No)

(d) At the 5% level of significance, state your conclusion in words.
3. (12 pts) Consider three events, $E$, $F$ and $G$ such that $P(E) = P(F) = P(G) = 0.6$.

(a) Are the two events $E$ and $F$ disjoint? Explain why or why not.

(b) If $P(F|E) = 0.9$, are the two events $E$ and $F$ independent? Why or why not?

(c) If $E$ and $G$ were known to be independent, what is the probability that they occur simultaneously? That is, what is $P(E$ and $G)$?

(d) If $E$ and $G$ were known to be independent, what is the probability that at least one of them occurs? That is, what is $P(E$ or $G)$?
4. (13pts) A rare genetic disease is discovered and it is known that only one in 100 people has it. A test for the disease is developed. The test is positive for 99% of people who have the disease, while the test is also positive for 10% of people who do not have the disease.

(a) Construct a 2 step tree diagram for a randomly selected person (whether he/she has the disease and then whether the test is positive or not) and label the branches of the tree with appropriate probabilities.

(b) What is the probability that a randomly selected person shows a positive test result? Give your answer to four decimal places.

(c) Suppose that a person has a positive test result. What is the probability that the person actually has the disease? Give your answer to two decimal places.
5. (13 pts) The proportion of students who own a cell phone on college campuses across the country has increased tremendously over the past few years. It is estimated that approximately 90% of students now own a cell phone.

(a) Seven students are to be selected at random from a large university. Assume that the proportion of students who own a cell phone at this university is the same as nationwide. What is the probability that all students in a simple random sample of 7 students own a cell phone? Give your answer to four decimal places.

(b) Consider a simple random sample of size 100 taken from this university. Compute the mean and standard deviation of the number of students among 100, who own a cell phone.

(c) Consider a simple random sample of 100 students taken from this university. Use the normal approximation to find the probability that out of 100 students, at least 84 students own a cell phone. Give your answer to four decimal places and indicate whether the computation was done with or without the continuity correction.

Circle one: (with, without)
6. (15pts) The following stemplot displays the proportion (in the form of a percentage) of the population that is non-Hispanic white for the 50 states of the US and the District of Columbia. Note that there are 51 data values. (USA Today, August 15, 2006). In the stemplot, the stem represents the tens digit.

```
2 3 9
3
4 2 3 8
5 9 9 9
6 0 0 0 1 2 3 5 5 6 8 8 9 9
7 1 2 5 6 7 8 8 9 9
8 1 1 2 3 3 3 4 5 6 6 6 7 8 8 9
9 0 1 3 4 6 6
```

(a) Give the five number summary of these data.

(b) Are there any suspected outliers by the 1.5 x IQR criterion? Show your work.

(c) With no further computations you should be able to decide which of the following statements are valid about this data set. Circle all valid statements.

(i) The distribution is skewed to the left.
(ii) mean > median
(iii) The data would approximately follow the 68-95-99.7 rule.
7. (7pts) Suppose that you like to select five samples from the 50 states and the District of Columbia in the previous problem for further study.

<table>
<thead>
<tr>
<th></th>
<th>Alabama</th>
<th>Alaska</th>
<th>Arizona</th>
<th>Arkansas</th>
<th>California</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>2</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>3</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>4</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>5</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>6</td>
<td>Colorado</td>
<td>Connecticut</td>
<td>Delaware</td>
<td>District of Columbia</td>
<td>Florida</td>
</tr>
<tr>
<td>7</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>8</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>9</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>10</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>11</td>
<td>Georgia</td>
<td></td>
<td>Idaho</td>
<td>Illinois</td>
<td>Iowa</td>
</tr>
<tr>
<td>12</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>Kansas</td>
</tr>
<tr>
<td>13</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>Kentucky</td>
</tr>
<tr>
<td>14</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>Louisiana</td>
</tr>
<tr>
<td>15</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>Maine</td>
</tr>
<tr>
<td>41</td>
<td>South Carolina</td>
<td>42.South Dakota</td>
<td>43.Tennessee</td>
<td>44.Texas</td>
<td>45.Utah</td>
</tr>
<tr>
<td>46</td>
<td>Vermont</td>
<td>47.Virginia</td>
<td>48.Washington</td>
<td>49.West Virginia</td>
<td>50.Wisconsin</td>
</tr>
<tr>
<td>51</td>
<td>Wyoming</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

(a) Use the above alphabetical ordering to select a simple random sample of size 5. If you use Table B, start at line 140. If you use TI 83, write out the command you entered and the five numbers generated by your calculator.

(b) Suppose that you decided to divide the United States into five regions, East, South East, Mid West, West, and South West, and then to randomly select one state from each of the five regions. Which one of the following best describes this sampling?

- Simple random sample
- Stratified sample
- Systematic sample

8. (15pts) A simple random sample of eight drivers was selected. All eight drivers are insured with the same insurance company, and all have similar auto insurance policies. The following table lists their driving experiences (in years) and monthly auto insurance premiums.

<table>
<thead>
<tr>
<th>$x$: Driving experience (years)</th>
<th>$y$: Monthly auto insurance premium ($)</th>
</tr>
</thead>
<tbody>
<tr>
<td>5</td>
<td>64</td>
</tr>
<tr>
<td>2</td>
<td>77</td>
</tr>
<tr>
<td>12</td>
<td>50</td>
</tr>
<tr>
<td>9</td>
<td>71</td>
</tr>
<tr>
<td>15</td>
<td>44</td>
</tr>
<tr>
<td>6</td>
<td>56</td>
</tr>
<tr>
<td>25</td>
<td>42</td>
</tr>
<tr>
<td>16</td>
<td>60</td>
</tr>
</tbody>
</table>

(a) Make a scatter plot of these data. Note that the scales in the $x$ and $y$ axis are different.
Consider the following summary statistics:

\[ \bar{x} = 11.25, \ s_x = 7.40, \ \bar{y} = 58.00, \ s_y = 12.48, \ r = -0.775. \]

(b) Find the equation of the regression line of \( y \) on \( x \).

c) If driving experience increases by one year, how much of a change in the average monthly auto insurance premium do you expect? Specify if the change in the premium is an increase or a decrease.

d) Predict the monthly auto insurance premium for a driver with 20 years of driving experience.

e) Based on the residual plot given below, does the linear regression seem appropriate? Explain briefly.