Calculators and one 8.5” by 11” sheet of handwritten notes allowed. Show all work and answers clearly in the space provided. There are 100 points possible.

1. The number of dozing students in a Statistics I lecture in 4 different class meetings is:
   
   8  2  7  1

   Calculate:
   a. (5 pts) Mean \( \bar{x} = \frac{8 + 2 + 7 + 1}{4} = \frac{18}{4} = 4.5 \)

   b. (6 pts) Standard deviation using the formula \( s = \sqrt{\frac{\sum (x_i - \bar{x})^2}{n-1}} \). Calculate by hand and show all work.

   \[
   \begin{array}{c|c|c}
   x_i & x_i - \bar{x} & (x_i - \bar{x})^2 \\
   \hline
   8 & 8 - 4.5 = 3.5 & 3.5^2 = 12.25 \\
   2 & 2 - 4.5 = -2.5 & 6.25 \\
   7 & 7 - 4.5 = 2.5 & 6.25 \\
   1 & 1 - 4.5 = -3.5 & 12.25 \\
   \hline
   & 37 = \sum (x_i - \bar{x})^2 \\
   \end{array}
   \]

   \[
   s^2 = \frac{\sum (x_i - \bar{x})^2}{n-1} = \frac{37}{4-1} = 12.33 \\
   s = \sqrt{12.33} = 3.51
   \]

   c. (3 pts) Two processes, Process A and Process B, are available for manufacturing aspirin tablets labeled 200 mg. However, not all aspirin tablets produced contain 200 mg of aspirin due to variation in the manufacturing process. Here are the summary statistics for aspirin produced by these processes:

<table>
<thead>
<tr>
<th></th>
<th>Process A</th>
<th>Process B</th>
</tr>
</thead>
<tbody>
<tr>
<td>Mean weight</td>
<td>200 mg</td>
<td>200 mg</td>
</tr>
<tr>
<td>Standard deviation</td>
<td>5 mg</td>
<td>20 mg</td>
</tr>
<tr>
<td>Sample Size</td>
<td>1000</td>
<td>1000</td>
</tr>
</tbody>
</table>

   Which process do you choose and why?

   I choose process A. The mean aspirin content is the same for both processes, but Process A has a smaller standard deviation. The aspirin from process A are more likely to be close to the mean.

2. (6 pts) The following data are collected on a group of people. Indicate the type of each variable (categorical, numeric-discrete, numeric-continuous).
   a. Favorite sport   categorical
   b. Number of times person checked e-mail yesterday   numeric-discrete
   c. Hours of sleep last night   numeric-continuous
   d. Amount of time spent watching TV last night   numeric-continuous
   e. Whether or not the person likes bananas   categorical
   f. Favorite radio station   categorical
3. The amount of time it takes a college student to complete a standardized math exam has a bell-shaped distribution with a mean of 45 minutes and standard deviation of 8 minutes.

a. (3 pts) Give an interval where 95% of college students’ times to complete the exam will fall.

\[ \begin{align*}
\text{Empirical Rule} & \quad \bar{x} \pm 2s \\
& \quad 45 \pm 2(8) \\
& \quad 39 \leq x \leq 51
\end{align*} \]

b. (3 pts) Approximately what percent of students will take less than 29 minutes or more than 61 minutes to complete the exam?

\[ (100 - 95)^\% = 5^\% \]

c. (2 pts) What percent of students will take over 61 minutes?

\[ \frac{1}{2}(5^\%) = 2.5^\% \quad \text{due to symmetry} \]

d. (2 pts) Assume the distribution of times to complete the standardized exam is extremely right skewed. At least what percent of the will take between 21 and 69 minutes?

\[ 1 - \frac{1}{2} = 0.98 \]

\[ K = 7.07 \]

\[ 45 \pm 7.07(8) \]

4. The number of students enrolled in 23 mathematics classes at CSUS are:

19, 24, 25, 25, 26, 27, 28, 28, 28, 29, 30, 31, 31, 32, 33, 33, 35, 36, 36, 39, 80

a. (3 pts) Calculate the median. Position

\[ \frac{1}{2}(n+1) = \frac{1}{2}(23+1) = \frac{1}{2}(24) = 12 \text{th} \]

\[ \text{median} = 28 \]

b. (4 pts) Calculate the first quartile and the 65th percentile.

\[ Q_1 = 27 \]

\[ Q_1 = \frac{c_{.25}(n+1)}{4} = \frac{c_{.25}(24)}{4} = 6 \text{th} \]

\[ P_{65} = \frac{31+32}{2} = 31.5 \]

c. (2 pts) Suppose you were to construct a histogram of these data and the first class consists of values 19-<28. What is the relative frequency of this class?

\[ \frac{6}{23} = 0.26 \]

d. (3 pts) Will the standard deviation increase or decrease if the data value 80 is removed from the data. Give a reason for your conclusion. (You should be able to do this without calculating the standard deviation.) The standard deviation will decrease since 80 is far from the bulk of the data. If it is removed, the data become less spread out about the mean. Hence, the standard deviation will decrease.
5. (6 pts) Indicate whether each statement is true (T) or false (F) or can't tell from the given information (C) by circling the appropriate letter. The given information about the data is the boxplot shown below. The dots indicate outliers.

a. The data are symmetric. T F C
b. The mean is greater than the median. F T C
c. The 95th percentile is between 1 and 8. T F C
d. The data between Q1 and the median are more spread out than the data between the median and Q3. F T C
e. The Empirical rule would apply to this data set. T F C
f. The value of Q3 is 3 F T C

---

6. A four-sided die whose faces are numbered 1, 2, 3, 4 and a standard six-sided die are tossed. Assume both dice are fair.

a. (5 pts) Give the sample space, S.

S = \{(1,1), (1,2), (1,3), (1,4), (1,5), (1,6), (2,1), (2,2), (2,3), (2,4), (2,5), (2,6), (3,1), (3,2), (3,3), (3,4), (3,5), (3,6), (4,1), (4,2), (4,3), (4,4), (4,5), (4,6)\}

What is the probability

b. (3 pts) of any single outcome in S? \[ \frac{1}{24} = 0.042 \]

c. (3 pts) the sum is 7 \[ \frac{4}{24} = \frac{1}{6} \]

d. (3 pts) doubles \[ \frac{4}{24} = \frac{1}{6} \]

e. (3 pts) the value on the six-sided die is greater than the value on the four-sided die \[ \frac{14}{24} = 0.583 \]

f. (3 pts) at least one value face up is odd \[ P(A) = 1 - P(A^c) = 1 - P(\text{both are even}) = 1 - \frac{6}{24} = \frac{18}{24} = 0.75 \]
7. 7 horses are competing in a race. The table below gives the probability of each horse winning. Use the table below to answer the following questions.

<table>
<thead>
<tr>
<th>Horse number</th>
<th>Probability of winning race</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>0.10</td>
</tr>
<tr>
<td>2</td>
<td>0.05</td>
</tr>
<tr>
<td>3</td>
<td>0.40</td>
</tr>
<tr>
<td>4</td>
<td>0.25</td>
</tr>
<tr>
<td>5</td>
<td>0.10</td>
</tr>
<tr>
<td>6</td>
<td>0.06</td>
</tr>
<tr>
<td>7</td>
<td>0.10</td>
</tr>
</tbody>
</table>

Define events as follows:
- E = winning horse wears an even number $\{2, 4, 6, 8\}$
- F = winning horse wears a number less than 4 $\{1, 2, 3\}$
- V = winning horse wears the number 7

Calculate the probability

a. (3 pts) List the outcomes in the event $E^c$ $\{4, 5, 6, 7\}$

b. (3 pts) List the outcomes in the event $E \cap (F^c)$ $\{2, 4, 6\} \cap \{4, 5, 6, 7\} = \{4\}$

c. (2 pts) Calculate $P(V)$

$1 - (0.1 + 0.05 + 0.4 + 0.25 + 0.1 + 0.06) = 1 - 0.76 = 0.24$

d. (2 pts) Calculate $P(E)$

$P(E) = P(2) + P(4) + P(6) = 0.05 + 0.25 + 0.06 = 0.36$

e. (2 pts) Calculate $P(F)$

$P(F) = P(1) + P(2) + P(3) = 0.1 + 0.05 + 0.4 = 0.55$

f. (2 pts) The winning horse wears an even number OR a number less than 4

$\{2, 4, 6\} \cup \{1, 2, 3\} = \{1, 2, 3, 4, 6\}$

$P(E \cup F) = 1 - P(5) - P(7) = 1 - 0.10 - 0.06 = 0.84$

g. (2 pts) $P(E \cap F)$

$P(E \cap F) = P(1) = 0.05$

h. (2 pts) $P(E \cap F)^c$

$1 - P(E \cap F) = 1 - 0.05 = 0.95$

i. (2 pts) Are the events E and F mutually exclusive? Give reasons for your answer.

E and F have outcomes in common since $E \cap F = \{2\}$ so E and F are NOT mutually exclusive.
8. Suppose $k$ is a fixed but unspecified positive number.

a. (5 pts) Show data sets 1 and 2 have the same mean. Show your calculations for credit.

Data set 1: $-k, 0, k$
Data set 2: $-10k, 0, 10k$

\[
\bar{X}_1 = \frac{-k + 0 + k}{3} = \frac{0}{3} = 0
\]

\[
\bar{X}_2 = \frac{-10k + 0 + 10k}{3} = \frac{0}{3} = 0
\]

b. (2 pts) Does data set 1 or 2 have a larger standard deviation? Give a reason for your choice.

Data set 2 has a larger standard deviation since the mean is 0 and the data are further from the mean than in data set 1.

c. (2 pts) Calculate the standard deviation of data set 2. (It will be an expression involving $k$, not a number.)

\[
\begin{array}{c|c|c}
X_i & X_i - \bar{X} & (X_i - \bar{X})^2 \\
-10k & -10k & (10k)^2 = 100k^2 \\
0 & 0 & 0 \\
10k & 10k & (10k)^2 = 100k^2 \\
\end{array}
\]

\[\sum (X_i - \bar{X})^2 = 100k^2 + 0 + 100k^2 \]

\[s = \sqrt{\frac{\sum (X_i - \bar{X})^2}{3 - 1}} = \sqrt{\frac{200k^2}{2}} = \sqrt{100k^2} = 10k
\]

d. (1 pt) What is the standard deviation for Data set 3: $-10k + 5, 5, 10k + 5$?

\[s = 10k
\]

Since Data set 3 is Data set 2 with 5 added to every value, the spread of the data do not change.