Problem 1 (30 points). Choose one appropriate answer in each of the following questions.

1. (B) The mean age of five people in a meeting is 30 years. One of the people, whose age is 50 years, leaves the room. The mean age of the remaining four people in the room is
   A) more than 30 years.   B) less than 30 years.   C) still 30 years.   D) can not be determined from the information given.

2. (B) In a class of 100 students, the grades on a statistics test are summarized in the following frequency table.

<table>
<thead>
<tr>
<th>Grade</th>
<th>Frequency</th>
</tr>
</thead>
<tbody>
<tr>
<td>91 - 100</td>
<td>11</td>
</tr>
<tr>
<td>81 - 90</td>
<td>31</td>
</tr>
<tr>
<td>71 - 80</td>
<td>42</td>
</tr>
<tr>
<td>61 - 70</td>
<td>16</td>
</tr>
</tbody>
</table>

   The median grade is in which of the following intervals?

3. (D) A sample of production records for an automobile manufacturer shows the following figures for production per shift: 705, 700, 690, 705. The variance of the above is
   A) 8.66.   B) 7.07.   C) 75.00.   D) 50.00.

4. (D) The correlation coefficient measures
   A) whether there is a relation between two variables.
   B) whether or not a scatterplot shows an interesting pattern.
   C) whether a cause and effect relation exists between two variables.
   D) the strength of a straight line relation between two variables.
5. (B) In a linear regression, a residual plot shows a sine pattern. This says that this linear regression model
   A) fits data well.   B) does not fit data well.   C) both A) and B) are wrong.

   In a statistical course a linear regression equation was computed to predict the final exam score from the score on the first test. The equation is $y = 10 + 0.9x$, where $y$ represents the final exam score and $x$ is the score on the first exam.

6. (C) The first test score is
   A) the intercept.   B) the slope.   C) the explanatory variable.   D) the response variable.

7. (A) Suppose Joe scores a 90 on the first exam. What would be the predicted value of his score on the final exam?
   A) 91,   B) 89,   C) 81.   D) can not be determined.

8. (C) Event $A$ occurs with probability 0.2. Event $B$ occurs with probability 0.8. Suppose they are disjoint (mutually exclusive), then
   A) $P(A \text{ and } B) = 0.16$.   B) $P(A \text{ and } B) = 1$.   C) $P(A \text{ or } B) = 1$.   D) $P(A \text{ and } B) = 0.16$.  

   All human blood can be typed as one of O, A, B, or AB. The distribution of the types varies a bit with race. Choose an African American at random. Here are the approximate probabilities that the person you choose will have blood type O, B, or AB.

   Blood type   O   A   B   AB
   Probability   .50   ?   .20   .05.

9. (B) The probability that the person chosen has blood type A is
   A) 0.4.   B) .25.   C) .27.   D) impossible to determine.

10. (B) The probability that the person chosen has a blood type other than O is
    A) .25   B) .50.   C) .75.   D) 1.00.
**Problem 2 (18 points)** The length of human pregnancies from conception to birth varies according to a distribution that is approximately Normal with mean 266 days and standard deviation 16 days.

(a) What percent of pregnancies last less than 240 days?

(b) What percent of pregnancies last between 240 and 270 days?

(c) How long do the longest 20% of pregnancies last?

**Solution.** (a) $z$-score = $(240 - 266)/16 = -1.625$. So the percentage = $1 - .9484 = 0.0516$;

(b) $z$-score at $270$ = $(270 - 266)/16 = 0.25$; So the percentage = $0.5987 - 0.0516 = 0.5471$;

(c) the 80 percentile = 0.845. So the number is $266 + 0.845 \cdot 16 = 279.52$ days.
**Problem 3 (20 points).** In Professor Bingham’s economics course the correlation between the students’ total scores prior to the final examination and the final examination score is $r = 0.6$. The pre-exam totals for all students in the course have mean 75 and standard deviation 30. The final exam score has mean 75 and standard deviation 8. Professor Bingham has lost John’s final exam but knows that his total before the exam was 300. He decides to predict his final exam score from his pre-exam total.

(a) What is the slope of the least-square regression line of final exam scores on pre-exam total scores in this course? What is the intercept?

(b) Use the regression line to predict John’s final exam score.

(c) Suppose we know (but the professor and John do not know) that John’s true final score is 80, what is the residual in the prediction?

(d) John doesn’t think this method accurately predicts how well he did on the final. Use $r^2$ to argue that his actual score could have been much higher.

**Solution.**

(a) $b = r s_y / s_x = 0.6 \times 8 / 30 = .16$. So $a = \bar{y} - bx = 75 - .16 \times 75 = 63$.

(b) $\hat{y} = 63 + .16 \times 300 = 111$.

(c) $80-111=-31$.

(d) The $r^2 = 0.36$, namely, the proportion of $y$ explained by $x$ is only 36%. The values of $y$ could be very large.
**Problem 4 (16 points).** Choose an American household at random and let the random variable $X$ be the number of cars they own. Here is the probability model if we ignore the few households that own more than five cars.

<table>
<thead>
<tr>
<th>$X$</th>
<th>0</th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
</tr>
</thead>
<tbody>
<tr>
<td>Probability</td>
<td>0.09</td>
<td>0.36</td>
<td>0.35</td>
<td>0.13</td>
<td>0.05</td>
<td>0.02</td>
</tr>
</tbody>
</table>

(a) Verify that this is a legitimate discrete distribution.

(b) Say in words what the event $\{X \geq 1\}$ is. Find $P(X \geq 1)$.

(c) A housing company builds houses with two-car garages. What percent of households have more cars than the garage can hold?

**Solution.** (a) $0.09 + 0.36 + 0.35 + 0.13 + 0.05 + 0.02 = 1$. So it is indeed a probability distribution.

(b) The household has at least one car. $P(X \geq 1) = 0.36 + 0.35 + 0.13 + 0.05 + 0.02 = 0.91$.

(c) $P(X > 2) = P(X \geq 3) = 0.13 + 0.05 + 0.02 = 0.2$. 
Problem 5 (16 points). Peter and Ruth play a die game. Each of Peter and Ruth roll a fair die once.

(a) What is the chance that they both get the Aces (one-dot’s)?

(b) What is the chance that the two numbers tie each other?

(c) What is the chance that Peter’s number is larger than Ruth’s?

(d) Let $A = \{\text{Peter gets either 1, 2, 3 or 4}\}$ and $B = \{\text{Peter gets either 2, 4 or 6}\}$. Calculate $P(A), P(B)$ and $P(A \text{ and } B)$ then conclude that $A$ and $B$ are independent.

Proof. (a) $(1/6) \cdot (1/6) = 1/36$.

(b) $(1/6) \cdot (1/6) + (1/6) \cdot (1/6) + (1/6) \cdot (1/6) + (1/6) \cdot (1/6) + (1/6) \cdot (1/6) + (1/6) \cdot (1/6) + (1/6) \cdot (1/6) = 1/6$.

(c) Except the tie case, half times Peter’s number is larger than Ruth’s. So the chance is $(1/2)\cdot(36-6)/36=15/36$.

(d) $P(A) = 4/6, \ P(B) = 3/6, \ P(A \text{ and } B) = P(\{2, 4\}) = 2/6$. So $P(A \text{ and } B) = P(A) \cdot P(B)$. This shows that $A$ and $B$ are independent.