Q1 (10 points). Choose the most suitable option.

1. A residual is defined as
   A. Error sum of square
   B. Regression sum of squares
   C. Type I Error
   D. $Y - \hat{Y}$

2. The relationship between correlation coefficient and coefficient of determination is that
   A. both are unrelated
   B. both are equal
   C. The coefficient of determination is the square root of the coefficient of correlation
   D. The coefficient of determination is the coefficient of correlation squared

3. The percent of total variation of the dependent variable $Y$ explained by the set of independent variables $X$ is measured by
   A. Coefficient of Correlation
   B. Coefficient of Skewness
   C. Standard Error of Estimate
   D. Coefficient of Determination

4. A company receives 60% of its orders over the Internet. Within a collection of 18 independently placed orders, what is the probability that between eight and ten (both inclusive) of the orders are received over the Internet?
   A. 0.60
   B. 0.6211
   C. 0.3622
   D. 0.3789

5. The prediction interval is used to estimate the value of the response variable for a fixed value of the predictor in a regression model.
   A. median
   B. mean
   C. IQR
   D. individual

6. The time in days between breakdowns of a machine is exponentially distributed with $\lambda = 0.2$. What is the expected time between machine breakdowns?
   A. 5
   B. 0.8
   C. 1.25
   D. 10

7. A word of five letters is transmitted by code to a receiver. The transmission signal is weak, and there is a 5% probability that any letter is in error independently of the others. What is the probability that the word is received correctly?
   A. 0.774
   B. 0.05
   C. 0.50
   D. 0.95

8. On average there are about 25 imperfections in 100 meters of optical cable. What is the probability that there is no more than one imperfection in 1 meter of cable?
   A. 0.9735
   B. 0.0265
   C. 0.025
   D. 0.750

9. There are 11 items of a product on a shelf in a retail outlet, and unknown to the customers, 4 of the items are outdated. Suppose that a customer takes 3 items at random. What is the probability that none of the outdated products are selected by the customer?
   A. 7/33
   B. 26/33
   C. 15/33
   D. 17/33

10. With reference to regression model, the prediction interval is always the confidence interval for a fixed value of the predictor.
    A. wider than
    B. narrower than
    C. same as
    D. None of the above
Q2 (10 points). The tensile strength of a paper product is related to the amount of hardwood in the pulp. Ten samples are produced in the pilot plant, and the data obtained are shown in the following table.

<table>
<thead>
<tr>
<th>Strength</th>
<th>Percent Hardwood</th>
</tr>
</thead>
<tbody>
<tr>
<td>160</td>
<td>10</td>
</tr>
<tr>
<td>171</td>
<td>15</td>
</tr>
<tr>
<td>175</td>
<td>15</td>
</tr>
<tr>
<td>182</td>
<td>20</td>
</tr>
<tr>
<td>184</td>
<td>20</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Strength</th>
<th>Percent Hardwood</th>
</tr>
</thead>
<tbody>
<tr>
<td>181</td>
<td>20</td>
</tr>
<tr>
<td>188</td>
<td>25</td>
</tr>
<tr>
<td>193</td>
<td>25</td>
</tr>
<tr>
<td>195</td>
<td>28</td>
</tr>
<tr>
<td>200</td>
<td>30</td>
</tr>
</tbody>
</table>

By fitting a linear regression model relating strength to percent hardwood, we obtained the following analysis (MINITAB output is given). Fill in all the missing numbers in this output.

---

**Regression Analysis: Strength versus Hardwood**

The regression equation is

\[
\text{Strength} = 144 + 1.88 \text{ Hardwood}
\]

<table>
<thead>
<tr>
<th>Predictor</th>
<th>Coef</th>
<th>SE Coef</th>
<th>T</th>
<th>P</th>
</tr>
</thead>
<tbody>
<tr>
<td>Constant</td>
<td>--------</td>
<td>2.522</td>
<td>--------</td>
<td>------</td>
</tr>
<tr>
<td>Hardwood</td>
<td>--------</td>
<td>0.1165</td>
<td>--------</td>
<td>0.000</td>
</tr>
</tbody>
</table>

**Analysis of Variance**

<table>
<thead>
<tr>
<th>Source</th>
<th>DF</th>
<th>SS</th>
<th>MS</th>
<th>F</th>
<th>P</th>
</tr>
</thead>
<tbody>
<tr>
<td>Regression</td>
<td>--</td>
<td>1262.1</td>
<td>1262.1</td>
<td>260.00</td>
<td>0.000</td>
</tr>
<tr>
<td>Residual Error</td>
<td>--</td>
<td>38.8</td>
<td>-------</td>
<td>-------</td>
<td>------</td>
</tr>
<tr>
<td>Total</td>
<td>--</td>
<td>1300.9</td>
<td>-------</td>
<td>---------</td>
<td>------</td>
</tr>
</tbody>
</table>

\[S = -------\]

R-Sq = \[-------\]
Q3 (5 points). A shipment of 20 calculators contains 4 defective sets. A sample of three calculators is selected in succession without replacement from this shipment.

(a) What is the probability that sample will contain at least one defective calculator given the number of defective calculators is not three?

(b) What is the probability that the calculator selected at the second draw is not defective but the third draw is defective?

Q4 (5 points). The probabilities that Ahmad and Mohammad will succeed in assembling the computer are 56% and 71%, respectively, and that the probability that both will succeed is 39%. Compute the probability that

(a) Exactly one of them will succeed in assembling the computer.

(b) Mohammad will not succeed in assembling the computer given that Ahmad has not succeeded in assembling the computer.
Q5 (7 points).

The manager of a gasoline station wants to study gasoline purchasing habits of motorists at his station. In particular, he decides to focus on two variables:

- The amount purchased by the motorists
- Whether the motorists would consider purchase premium – grade gasoline.

a). He selects a random sample of 60 motorists during a certain week, he found that the mean amount purchased was 11.3 gallons, with standard deviation 3.1 gallons. Set up a 99% confidence interval estimate of the population mean purchased. Also interpret it.

b). How many customers need to be selected to have 90% confidence of estimating the population proportion of motorists who purchase premium – grade gasoline to be within ±0.04?

Q6 (3 points). A random sample of \( n = 36 \) is selected from a population with mean 2 and standard deviation \( \sqrt{\frac{2}{3}} \). Find the probability that the sample mean is greater than 2.1 but less than 2.5.
Q7 (5 points).
Two suppliers manufacture a plastic gear used in a laser printer. The impact strength of these gears measured in foot-pounds is an important characteristic. A random sample of 10 gears from supplier 1 results in $\bar{x}_1 = 321$ and $s_1 = 22$, while another random sample of 16 gears from the second supplier results in $\bar{x}_2 = 290$ and $s_2 = 20$. Using this sample information, test that there is no difference between mean impact strength of gears by the two suppliers, using Critical value approach (assuming equal variances).
As machines are used over long periods of time, the output product can get off target. Data is collected on the average value of how much off target a product is getting manufactured as a function of machine use. The intensity of being off target (Y) is measured in Millimeters and the time of machine use (X) is measured in hours. Based on seven randomly collected observations we have the following summary quantities.

\[
\begin{align*}
    n &= 7, \\
    \sum y_i &= 8.91, \\
    \sum x_i &= 260, \\
    \sum y_i^2 &= 11.4159, \\
    \sum x_i^2 &= 9852, \\
    \sum x_i y_i &= 334.68
\end{align*}
\]

a) Find the estimated regression line.

b) What change in mean intensity of being off target would be expected for a 1 hour change in the time of machine use?

c) What are the assumptions required for the different components of regression analysis?