**REPETITION – Statistical Data Processing**

1. **What does the median represent in a data set?**
a) The most frequent value in the data set
b) The average value of the data set
c) The middle value when the data is sorted in ascending order
d) The difference between the largest and smallest value
2. **Explain what variance represents in a data set and how it differs from standard deviation.**
3. **Given the data set: 5, 8, 12, 15, 20, calculate the mean, median, and mode.**
4. **Which of the following is NOT a measure of variability?**
a) Standard deviation
b) Variance
c) Median
d) Range
5. **What is the purpose of the test of independence (chi-square test)?**
a) To determine if two variables are independent
b) To compare the means of two data sets
c) To assess the correlation between two variables
d) To estimate variance in a data set
6. **What is the main assumption for using the test of independence?**
a) Data must be nominal or ordinal
b) Data must be interval-scale
c) Data must follow a normal distribution
7. **Example**:
The following table shows preferences for two chocolate brands by gender:

|  | **Brand A** | **Brand B** |
| --- | --- | --- |
| Men | 30 | 20 |
| Women | 40 | 10 |

Test the hypothesis that preferences for the brands are independent of gender at a 5% significance level.

1. **What is the purpose of the goodness-of-fit test (chi-square test)?**
a) To test a hypothesis about a mean
b) To test if observed data matches an expected distribution
c) To test for normality of a distribution
2. **Example with multiple-choice answers**:
Data on car colors passing an intersection is collected. The theoretical distribution is 40% white, 30% red, and 30% other. The observed data is as follows:
White: 50 cars
Red: 25 cars
Other: 25 cars

Test if the observed data matches the theoretical distribution at a 5% significance level.

1. **What is the purpose of analysis of variance (ANOVA)?**
a) To compare differences between two means
b) To compare differences between more than two means
c) To determine the relationship between two variables
2. **When can one-way ANOVA be used?**

**Answer**: When comparing the means of three or more independent groups to test if at least one mean is significantly different.

1. **Example**:
Three groups of students were tested using three different teaching methods. The test scores are as follows:
	* Group 1: 85, 89, 90, 100
	* Group 2: 78, 82, 84, 85
	* Group 3: 92, 94, 88, 90

Test if there is a statistically significant difference in the test scores among the groups at a 5% significance level.

1. **What is the purpose of regression analysis?**
a) To determine the relationship between a dependent and an independent variable
b) To test the match between theoretical and observed distributions
c) To compare multiple means in the data
2. **Example with multiple-choice answers**:
The following data shows the relationship between study hours and scores achieved:
	* Study hours: 2, 4, 6, 8, 9
	* Scores achieved: 20, 35, 50, 65, 66

What is the slope of the regression line?

1. **What is the difference between simple linear regression and multiple regression?**