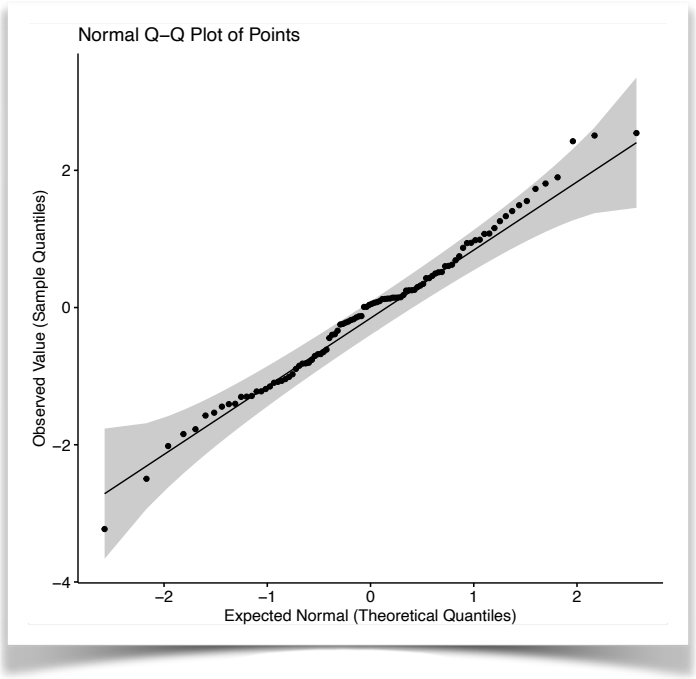
Week 7. Analysis III

SOLUTIONS TO PREPARATORY QUESTIONS.

Q1. Examine the following Quantile-Quantile (Q-Q) plot, and indicate which answer correctly describes the distribution of data.



A. Q-Q plots are used to visually assess homogeneity of variance (homoscedasticity), and the data here is homoscedastic.
B. Q-Q plots are used to visually assess homogeneity of variance (homoscedasticity), and the data here is heteroscedastic.
C. Q-Q plots are used to visually assess normality, and the data is almost normally distributed. **[Correct]**
D. Q-Q plots are used to visually assess normality, and the data is *not at all* normally distributed.

Q2. Closely examine the following table obtained in SPSS after conducting Levene's test, and choose the correct answer below.

	Levene Statistic	df1	df2	Sig.
Variable	8.1106	1	30	0.007871

A. Levene's test is used to assess the homogeneity of variance, and the results show that 'Variable' is homoscedastic.
B. Levene's test is used to assess the homogeneity of variance, and the results show that 'Variable' is heteroscedastic. **[Correct]**
C. Levene's test is used to assess the normality of distribution, and the results show that 'Variable' is normally distributed.
D. Levene's test is used to assess the normality of distribution, and the results show that 'Variable' is not normally distributed.

Q3. You have collected data on vitamin D levels for groups of Finish, Dutch and Italians. Your data is non-normally distributed. Which non-parametric test are you going to use to find if there are any significant differences among these groups?

A. Independent Samples *t*-test
B. Kruskal-Wallis test **[Correct]**
C. Friedman's test
D. Chi-square test

Q4. Regression analysis was used to examine the relationship between the amount of time students spent (x; in hours) studying and their exam scores (y) for high school students. The following regression equation was obtained:
$y = 31.9 + 0.84\ x$

Based on the estimated regression equation above, if the amount of time were increased by 1 hour, which of the following choices would apply?

A. Exam Score will increase by 0.84 **[Correct]**
B. Exam Score will increase by 31.9
C. Exam Score will increase by 32.74
D. Exam Score will decrease by 0.84

Q5. Which of the following is not a measure of the goodness of fit of a linear regression?

A. Coefficient of Determination
B. Intercept **[Correct]**
C. Residuals
D. Standard Error

Q6. A Pearson's correlation of $r = 0.7$ was found between time spent studying and exam score. What is the proportion of variance in exam scores that can be explained by time spent studying?

A. 0.70
B. 0.49 **[Correct]**; Proportion of variance is measured by R^2 (the coefficient of determination)

C. 0.07
D. 0.14

Q7. Consider the following multiple regression:
$z = 21.3 - 3.4x + 0.6y$
If the coefficient of determination (R^2) for this regression model is 0.81, what does the value indicate?

A. The standard deviation of the residuals is equal to 0.81.
B. The proportion of variance explained by this model is 81%. **[Correct]**
C. The model is not statistically significant.
D. The goodness of fit is 0.81.