Dirasat: Human and Social Sciences, Volume **50**, No. 3, 2023

Deanship of Scientific Research

Assessment of Regression Model for Rainfall in Saudi Arabia (1979-2011) Using Dummy Variables

Manahil Eltayeb[1], Sulafa Hag-elsafi*[2]

[1-]Department of Quantitative Analysis, College of Business administration, King Saud University, Riyadh, Saudi Arabia

[2-] Department of Geography, College of Humanities and Social Sciences, King Saud University, Riyadh, Saudi Arabia

Received: 12/8/2021
Revised: 28/9/2021
Accepted: 13/2/2022
Published: 30/5/2023

* *Corresponding author:*
shagelsafi@ksu.edu.sa

Citation: Eltayeb, M., & Hag-elsafi, S. (2023). Assessment of Regression Model for Rainfall in Saudi Arabia (1979-2011) Using Dummy Variables. *Dirasat: Human and Social Sciences*, *50*(3), 167–176.
https://doi.org/10.35516/hum.v50i3.5404

Abstract

Objectives: This study aims to analyze rainfall in Saudi Arabia by designing models based on data from 20 stations across the Kingdom from 1979 to 2011.

Methods: The analysis employed a multiple linear regression model with rainfall as the dependent variable and annual quarters as the independent variables. Dummy variables were utilized in the analysis. The regression model provided valuable insights into the impact of rainfall rates in different quarters across Saudi Arabia. Monthly data was collected from each region of the Kingdom during the study period and categorized into five groups based on average rainfall: Group 1 (5-15 mm), Group 2 (15-25 mm), Group 3 (25-35 mm), Group 4 (35-45 mm), and Group 5 (45-70 mm). Each group was represented by a separate regression model. To reduce the number of dummy variables in the model, the monthly data was converted to quarterly data.

Results: A significant finding of this study is that all models were statistically significant, indicating that rainfall distribution is influenced by the annual quarters. Furthermore, it was observed that the average rainfall in most quarters across different regions was statistically significant, except for the fourth quarter in Group 5 and the third quarter in Groups 1, 2, and 4.

Conclusions: The inclusion of dummy variables as independent variables in the multiple linear regression model proved to be a novel and effective approach for analyzing rainfall time series. The results can serve as a foundation for future studies, enabling prediction and informed decision-making based on the findings.

Keywords: Rainfall, dummy variable, T- test, F- test, Saudi Arabia.

تقييم نموذج الانحدارلهطول الأمطارفي المملكة العربية السعودية (1979-2011) باستخدام المتغيرات الصورية

مناهل الطيب[1]، سلافة حاج الصافي*[2]

[1]قسم التحليل الكمي، كلية إدارة الأعمال، جامعة الملك سعود، الرياض، المملكة العربية السعودية.

[2]قسم الجغرافيا، كلية العلوم الإنسانية والاجتماعية، جامعة الملك سعود، الرياض، المملكة العربية السعودية..

ملخّص

الأهداف: يهدف هذا البحث إلى بناء نماذج تعمل على دراسة هطول الأمطار في أنحاء المملكة العربية السعودية ، وتناولت الدراسة الأمطار في الفترة من 1979 إلى 2011، وشملت 20 محطة.

المنهجية: جرى التحليل على أساس نموذج الانحدار الخطي المتعدد، وعُدّ متغير هطول الأمطار كتابع والأرباع السنوية كمتغيرات مستقلة، واستُخدِمت المتغيرات الصورية للتعبير عنها. يحقق نموذج الانحدار هدف معرفة تأثير الأرباع المختلفة على معدلات الأمطار وبناء نماذج يمكن استخدامها في التنبؤ في دراسات مستقبلية. جرى الحصول على بيانات شهرية لفترة الدراسة وعلى أساسها وُزّعَت مناطق المملكة إلى خمسة مجموعات حسب متوسطات الأمطار(المجموعة 1: من 5-15 ملم ، المجموعة 2: من 15 إلى 25 ملم ، المجموعة 3: من 25 إلى 35 ملم ، المجموعة 4: من 35-45 ملم والمجموعة 5: من 45-70 ملم) كل مجموعة يمثلها نموذج انحدار. عند تقدير النماذج جرى تحويل البيانات الشهرية إلى ربع سنوية لغرض تقليل المتغيرات الصورية في النموذج.

النتائج: من أهم النتائج التي جرى التوصل إليها أن جميع النماذج ذات دلالة إحصائية كلية؛ مما يعني أن توزيع الأمطار يتأثر في الأرباع السنوية، كما جرى التوصل إلى أنّ معظم الأرباع في مختلف المناطق ذات دلالة إحصائية على متوسط الأمطار باستثناء الربع الرابع في المجموعة 5 والربع الثالث في المجموعة 1، 2 و4.

الخلاصة: على الرغم من شيوع استخدام نماذج الانحدار في تحليل السلاسل الزمنية للأمطار، إلا أن استخدام المتغيرات الصورية كمتغيرات مستقلة في نموذج الانحدار الخطي المتعدد يعدُّ إضافة جديدة أدت إلى نتائج إحصائية جيدة. وعليه، إنّ هذه النتائج يمكن أن تشكل أساسا لدراسات أخرى أعمق وأحدث ويمكن استخدامها في التنبؤ واتخاذ قرارات مهمة.

الكلمات الدالة: المتغير الصوري ، اختبار F ، اختبار T، المملكة العربية السعودية ، الهطول المطري.

INTRODUCTION:

The effects of climate change and fluctuations in human existence have long been recognized. As a result, weather and climatic factors are monitored in this study. Rainfall has a significant influence on the natural environment and human existence as one of the most important climatic factors. (Sahriman et al, 2014; Fetene et al, 2018; Hafez, 2019).

Rainfall is one of the most important sources of water for most areas in the world, and Saudi Arabia is no exception. This demand for water resources is due to population growth, and the need to grow more foods to feed the people in the world. Therefore, agriculture and domestics uses for water grew rapidly, according to several studies that were analyzed on rainfall (Modarres & Silva, 2007; Silva, 2007; ELAGIB, 2010). Also different techniques of analysis were used by (Tarawneh, 2016) and Harmonic to analyze the annual, seasonal, and inter-seasonal rainfall in Saudi Arabia. (Tarawneh and Kadioglu, 2002) studied the precipitation and its periodicity in Jordan using Harmonics analysis.

Despite the above cited literature, little is known about the relevance of trends and variability in hydro-climatic data for most Middle Eastern nations, and to measure climate model uncertainty, several studies compared simulated climatic data with observed data. In fact, recent efforts of research coverage in many Middle Eastern countries particularly dried and semiarid regions of the Gulf Peninsula remain relatively poor (Merabtene et al., 2016).

These regions are impacted by a variety of well-known weather variables that affect different locations at different periods of the year, both synoptically and climatologically.

The percentages reflect the impact of these circumstances, which often affect each region independently. (Tarawneh, 2016; Fetene et al, 2018)

The rainfall distribution is not uniform in time or in space, geographic factors, like temperature, pressure, distance from the source moisture and topography are also important. The time factor (i.e., seasonality) is found as one of the important factors (Abdullah and Al-Mazroui, 1998; Hasanean and Almazroui, 2015; Hag-elsafi and El-Tayib, 2016; Hafez, 2019).

Sahriman et al, 2014 showed that the principal component regression (PCR) model was adjusted to account for the inaccuracies created by introducing dummy variables to the equation. For instance, partial least square regression was used to determine the dummy variables. The rainfall prediction was enhanced by using the PCR model together with the dummy variables.

In order to understand the trends and possible changes in the rainfall patterns, it requires the analysis of regional precipitation characteristics over an extended period of time. This might be beneficial to other disciplines as well, such as climate modeling (Luong et al, 2020). Therefore, the aim of this study is to reduce the monthly data of rainfall to four quarter seasons per year, then evaluate the performance of the regression model using dummy variable.

Study area:

Saudi Arabia lies between latitudes 16°21'58"N and 32°9'57"N, and longitudes 34°33'48"E and 55°41'29"E. It has dried desert climate with high temperatures in most parts of the country due to the aridity, and there are wide fluctuations across seasons and places. Annual rainfall in the north fluctuates between 100 and 200 mm, except near the coast, where yearly rainfall drops below 100 mm in the further south and some small areas receives 500 mm/year. (AQUASTAT, 2008).

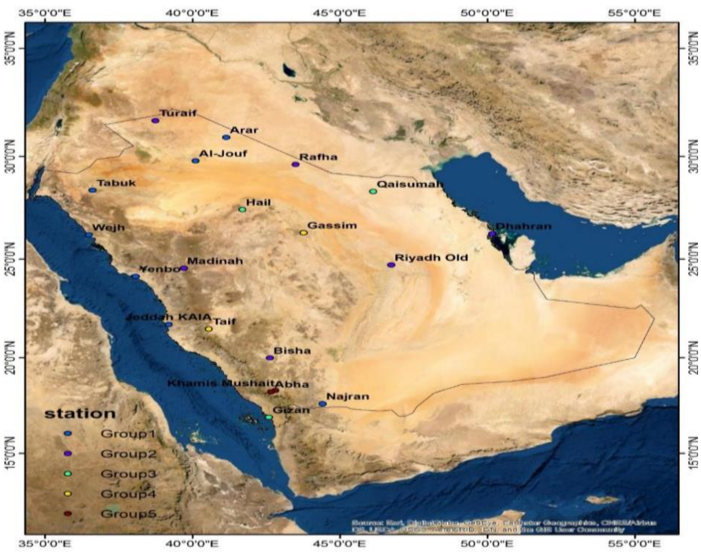

Figure (1): Study area.

Data:

Climatic data were provided by the Presidency of Meteorology and Environment, the Kingdom's official climate agency. The data were collected from 20 meteorological stations across Saudi Arabia as shown in Figure (1), and they were gathered for the duration of (396 months), from 1979 to 2011.

According to a study by Hasanean and Almazroui (2015), there are two seasons in Saudi Arabia, wet and dry. The wet season is from November to April, and the dry season is from June to September respectively. However, due to the geographical and temporal fluctuations in precipitation, the rainfall distribution is not uniform in time or place. Therefore, it should be analyzed and/or evaluated as a spatiotemporal phenomenon, especially in dry and semiarid locations. This is done from the surface observatory stations as well as the related rainfall quantities for the study period:

1. The data were obtained from 396 months and converted into 132 quarterly data, which represented the sample size used to design the models.
2. On the quarterly basis of rainfall then the averages for the study period (1979 to 2011) were calculated as shown in Table (1).
3. The study areas (20 stations) were divided into 5-group according to the annual average as shown in Table (2) and Figure (1).

Table (1): Descriptive Statistics

Descriptive Statistics				
Station name	Longitude	Latitude	N	Mean
Tabuk	36.607	28.376	132	7.38864
Yenbo	38.064	24.14	132	7.84091
Wejh	36.477	26.205	132	9.62576
AlJouf	40.099	29.789	132	13.28333
Jeddah	39.187	21.71	132	13.74015
Arar	41.141	30.902	132	13.94091
Najran	44.414	17.611	132	14.77500
Madinah	39.699	24.548	132	15.44470
Rafha	43.495	29.621	132	21.24015
Bisha	42.619	19.991	132	21.35227
Turaif	38.739	31.688	132	21.89242
Riyadh	46.738	24.711	132	22.17727
Dhahran	50.161	26.259	132	22.73939
Hail	41.691	27.434	132	27.51288
Qaisumah	46.13	28.319	132	30.94394
Gizan	42.585	16.897	132	32.57955
Gassim	43.768	26.308	132	36.05985
Taif	40.549	21.479	132	41.46136
KhamisMushait	42.806	18.299	132	47.40606
Abha	42.661	18.233	132	56.36818
Valid N			132	

Table (2): Groups according to average quarterly rainfall.

The Groups	Average Quarterly Rainfall	Stations
Group 1	5 up to 15	Wejh, Yenbo, Tabuk, Jeddah, AlJouf, Arar and Najran.
Group 2	15 up to 25	Madinah, Bisha, Turaif, Riyadh, Rafha and Dhahran.
Group 3	25 up to 35	Hail, Qaisumah and Gizan.
Group 4	35 up to 45	Taif and Gassim.
Group 5	45 up to 70*	KhamisMushait and Abha.

*The last two categories were merged.

METHODOLOGY:

The annual rainfall records from 1979 to 2011 for the 20 stations (33 years) were used in order to detect rainfall regimes and climatic features of Saudi Arabia, using dummy variable analysis techniques.

4-1 Dummy variables:

Dummy or indicator variables are used to determine the relationships between qualitative independent variables and the dependent variable. There are many ways to identify the classes of a qualitative variable; the values 0 and 1 are used in this research (Hanke and Wichern, 2013).

If the qualitative variable is an ‘m’ category, it introduces only (m – 1) dummy variables. If this rule is not followed, we can fall into what is called the dummy-variable trap, that is, the situation of perfect collinearity or perfect multicollinearity, if there is more than one exact relationship among the variables. This rule is also applied if we have more than one qualitative variable in the model.

With regard to the above situation (the dummy-variable trap), there are two ways to circumvent this trap. First, introduce as many dummy variables as the number of categories of that variable (a dummy for each category and omit the intercept term), and second, include the intercept term and introduce only (m – 1) dummies, where m is the number of categories of the dummy variable with the intercept suppressed. Thus, allowing a dummy variable for each category, and obtaining directly the mean values of the various categories, (Gujarati, 1995).

4-2 The use of dummy variables in seasonal analysis:

It is often desirable to remove the seasonal factor, or component, from a time series so that one can concentrate on the other components. The process of removing the seasonal component from the time series is known as deseasonalization or seasonal adjustment, and the time series thus obtained is called the deseasonalized, or seasonally adjusted time series.

There are several methods of deseasonalizing the time series, but only one of these methods was used, namely, the method of dummy variables.

The basic model is (Gujarati, 1995):

$$Y_t = \beta_1 D_{1t} + \beta_2 D_{2t} + \beta_3 D_{3t} + \beta_4 D_{4t} + u_t \text{ ....} (1)$$

Where

Y_t: The quarterly rain.

D’s are the dummies, taking a value of 1 in the relevant quarter and 0 otherwise.

D_1 : The first quarter.

D_2 : The second quarter.

D_3 : The third quarter.

D_4 : The fourth quarter.

$\hat{\beta}'s$: The average rainfall in the each quarter (1, 2, 3, 4)

4-3 Multiple linear regression assumptions:

The study employed the regression model to analyze each group. It shows the relationship between average rainfall (dependent variable) and quarters (independent variables), and it determines the quarter, which has the greatest average rainfall.

The regression analysis is also used to predict the value of the dependent and the independent variables, based on the value of at least one independent variable, which explains the impact of changes the independent variable has on the dependent variable (s).

When running a multiple regression, there are several assumptions that must be checked to make sure they meet the data requirements. The most important assumptions are:

1: The relationship between the dependent variable and the independent variables must be linear.

2: There should be no multicollinearity in the data. This assumption is essential to ensure that the independent variables are not too highly correlated with one another.

3: The values of the residuals must be independent. This assumption can be tested using the Ljung–Box test (Hanke and Wichern, 2013).

The Ljung–Box test is a type of statistical test which determines whether any of a group of autocorrelations of a time series is different from zero. Instead of testing randomness at each distinct lag, it tests the "overall" randomness based on a number of lags, and it is therefore a portmanteau test. The hypotheses of this test are:

H_0: The data are independently (random)

H_1: The data are not independently at any time lag.

We accept H_0 if $P-value$ "Sig" $> \alpha$ "0.05"

4: The variance of the residuals has to be constant, and this is called homoscedasticity.

If all of the independent variables in the model are dummy variables, the standard errors of the estimated coefficients are constant (take the same value for all coefficients) because all independent variables only take a value of 1 or zero.

5: The values of the residuals are normally distributed. As the sample size gets large enough (n ≥ 30) the sampling distribution of the sample statistic becomes almost normal regardless of the shape of the population (Levine et al, 2016).

In this research the sample size =132 quarters.

The coefficient of determination (R^2) is the portion of the total variation in the dependent variable that is explained by the variation in the independent variable.

F Test: The F Test is for the overall significance of the model, and it is to show if there is a significant linear relationship between all of the independent variables considered together and the dependent variable. If the p-value is less than (α) we reject the null hypotheses which mean the model is significance or the independent variables (D's) have significant effect on the dependent variable

t test: The t-test is used for individual variable slopes. It shows if there is a linear relationship between the one independent variable and the dependent variable, as well as having constant effects of other independent variables. If the p-value is less than (α) we reject the null hypotheses which mean the specified independent variables (D_i) have significant effect on the dependent variable .

RESULTS AND DISCUSSION:

According to the Regression Model Analysis of each group of the data, the following results were reported:

Group 1: The cities (Wejh, Yenbo, Tabuk, Jeddah, AlJouf, Arar and Najran), average quarterly rainfall was 5 up to 15

$$Y_t = 18.315D_{1t} + 7.47D_{2t} + 1.682_3D_{3t} + 18.588D_{4t}$$

$$Se = 1.813 \quad 1.813 \quad 1.813 \quad 1.813$$

$$t = 10.100 \quad 4.119 \quad 0.927 \quad 10.251$$

$$P-value\text{ “}Sig\text{”} = 0.000 \quad 0.000 \quad 0.355 \quad 0.000$$

$$F = 56.229 \qquad P-value\text{”}Sig\text{”} = 0.000$$

$$R^2 = 0.64$$

$$Q-test = 490.652 \qquad P-value\text{”}Sig\text{”} = 0.000$$

	The first quarter D_1	The second quarter D_2	The third quarter D_3	The fourth quarter D_4
Coefficients	The average rainfall in the first quarter = $\hat{\beta}_1 = 18.315$	The average rainfall in the second quarter = $\hat{\beta}_2 = 7.47$	The average rainfall in the third quarter = $\hat{\beta}_3 = 1.682$	The average rainfall in the fourth quarter = $\hat{\beta}_4 = 18.588$
$Var(\hat{\beta}_i) = Se^2_{\hat{\beta}_i}$	For all regression coefficients, standard error = 1.813 it follows that the variance is constant			
t- test	the first quarter is statistically significant because the $p-value = 0.000 < \alpha$ "0.05"	the second quarter is statistically significant because the $p-value = 0.000 < \alpha$ "0.05"	the third quarter is non - statistically significant because the $p-value = 0.355 > \alpha$ "0.05"	the fourth quarter is statistically significant because the $p-value = 0.000 < \alpha$ "0.05"
F-test	According to the F test $F = 56.229$, $p-value = 0.000 < \alpha$ "0.05" The overall model is significant.			
R^2	The independent variable (D's) explained 64% from the variation in the rainfall "Y" other variables explained about 36%.			
Q-test	According to the Q-test we accept the null hypothesis that there is no auto-correlation ($Q-test = 490.652$ and $P-value$ "Sig" = 0.000 > α "0.05")			

Group 2: The cities (Madinah, Bisha, Turaif, Riyadh, Rafha and Dhahran), average quarterly rainfall was 15 up to 25

$$Y_t = 37.268D_{1t} + 21.667D_{2t} + 1.369D_{3t} + 22.925D_{4t}$$

$$Se = 2.785 \quad 2.785 \quad 2.785 \quad 2.785$$

$$t = 13.383 \quad 7.781 \quad 0.492 \quad 8.232$$

$$P-value\text{ "}Sig\text{" } = 0.000 \quad 0.000 \quad 0.624 \quad 0.000$$

$$F = 76.917 \qquad P-value\text{"}Sig\text{" } = 0.000$$

$$R^2 = 0.71$$

$$Q-test = 711.016 \qquad P-value\text{"}Sig\text{" } = 0.000$$

	The first quarter D_1	The second quarter D_2	The third quarter D_3	The fourth quarter D_4
Coefficients	The average rainfall in the first quarter = $\hat{\beta}_1 = 37.268$	The average or rainfall in the second quarter = $\hat{\beta}_2 = 21.667$	The average rainfall in the third quarter = $\hat{\beta}_3 = 1.369$	The average rainfall in the fourth quarter = $\hat{\beta}_4 = 22.925$
$Var(\hat{\beta}_i) = Se^2_{\hat{\beta}_i}$	For all regression coefficients, standard error = 2.785 it follows that the variance is constant			
t- test	the first quarter is statistically significant because the $p-value = 0.000 < \alpha$ "0.05"	the second quarter is statistically significant because the $p-value = 0.000 < \alpha$ "0.05"	the third quarter is non - statistically significant because the $p-value = 0.624 > \alpha$ "0.05"	the fourth quarter is statistically significant because the $p-value = 0.000 < \alpha$ "0.05"
F-test	According to the F test $F = 76.917$, $p-value = 0.000 < \alpha$ "0.05" The overall model is significant.			
R^2	The independent variable (D's) explained 71% from the variation in the rainfall "Y" other variables explained about 29%.			
Q-test	According to the Q-test we accept the null hypothesis that there is no auto-correlation ($Q-test = 711.016$ and $P-value$ "Sig" = 0.000 > α "0.05")			

Group 3: The cities (Hail, Qaisumah and Gizan.).

Average Quarterly Rainfall was 25 up to 35

$$Y_t = 42.166D_{1t} + 22.669D_{2t} + 13.757D_{3t} + 42.791D_{4t}$$

$$Se \;=\; 2.244 \quad 2.244 \quad 2.244 \quad 2.244$$

$$t = 9.935 \quad 5.341 \quad 3.241 \quad 10.082$$

$$P - value \text{ “}Sig\text{”} = 0.000 \quad 0.000 \quad 0.002 \quad 0.000$$

$$F = 59.847 \qquad P - value\text{”}Sig\text{”} = 0.000$$

$$R^2 = 0.65$$

$$Q - test = 245.892 \qquad P - value\text{”}Sig\text{”} = 0.000$$

	The first quarter $\boldsymbol{D_1}$	**The second quarter** $\boldsymbol{D_2}$	**The third quarter** $\boldsymbol{D_3}$	**The fourth quarter** $\boldsymbol{D_4}$
Coefficients	The average rainfall in the first quarter = $\hat{\beta}_1 = 42.166$	The average rainfall in the second quarter = $\hat{\beta}_2 = 22.669$	The average rainfall in the third quarter = $\hat{\beta}_3 = 13.757$	The average rainfall in the fourth quarter = $\hat{\beta}_4 = 42.791$
$Var(\hat{\beta}_i) = Se^2_{\hat{\beta}_i}$	For all regression coefficients, standard error = 2.244 it follows that the variance is constant			
t- test	the first quarter is statistically significant because the $p - value = 0.000 < \alpha$ “0.05”	the second quarter is statistically significant because the $p - value = 0.000 < \alpha$ “0.05”	the third quarter is statistically significant because the $p - value = 0.002 < \alpha$ “0.05”	the fourth quarter is statistically significant because the $p - value = 0.000 < \alpha$ “0.05”
F-test	According to the F test $F = 59.847$, $p - value = 0.000 < \alpha$ “0.05” The overall model is significant.			
R^2	The independent variable (D’s) explained 65% from the variation in the rainfall “Y” other variables explained about 35%.			
Q-test	According to the Q-test we accept the null hypothesis that there is no auto-correlation ($Q - test = 245.892$ *and* $P - value$ “Sig” = 0.000 $> \alpha$ “0.05”)			

Group 4: The cities (Taif and Gassim.).

Average Quarterly Rainfall was 35 up to 45

$$Y_t = 51.865D_{1t} + 28.102D_{2t} + 0.292D_{3t} + 35.645D_{4t}$$

$$Se = 4.5 \quad 4.5 \quad 4.5 \quad 4.5$$

$$t = 11.526 \quad 6.245 \quad 0.065 \quad 7.921$$

$$P - value \text{ “}Sig\text{”} = 0.000 \quad 0.000 \quad 0.948 \quad 0.000$$

$$F = 58.647 \qquad P - value\text{”}Sig\text{”} = 0.000$$

$$R^2 = 0.65$$

$$Q - test = 555.181 \qquad P - value\text{”}Sig\text{”} = 0.000$$

	The first quarter $\boldsymbol{D_1}$	**The second quarter** $\boldsymbol{D_2}$	**The third quarter** $\boldsymbol{D_3}$	**The fourth quarter** $\boldsymbol{D_4}$
Coefficients	The average rainfall in the first quarter = $\hat{\beta}_1 = 51.865$	The average rainfall in the second quarter = $\hat{\beta}_2 = 28.102$	The average rainfall in the third quarter = $\hat{\beta}_3 = 0.292$	The average rainfall in the fourth quarter = $\hat{\beta}_4 = 35.645$

$Var(\hat{\beta}_i) = Se^2_{\hat{\beta}_i}$	For all regression coefficients, standard error = 4.5 it follows that the variance is constant			
t- test	the first quarter is statistically significant because the $p-value = 0.000 < \alpha$ "0.05"	the second quarter is statistically significant because the $p-value = 0.000 < \alpha$ "0.05"	the third quarter is non - statistically significant because the $p-value = 0.948 > \alpha$ "0.05"	the fourth quarter is statistically significant because the $p-value = 0.000 < \alpha$ "0.05"
F-test	According to the F test $F = 58.647$, $p-value = 0.000 < \alpha$ "0.05" The overall model is significant.			
R^2	The independent variable (D's) explained 65% from the variation in the rainfall "Y" other variables explained about 35%.			
Q-test	According to the Q-test we accept the null hypothesis that there is no auto-correlation ($Q-test = 555.181$ and $P-value$ "Sig" $= 0.000 > \alpha$ "0.05")			

Group 5: The cities (KhamisMushait and Abha).

Average Quarterly Rainfall was 45 up to 70

$$Y_t = 63.048D_{1t} + 79.312D_{2t} + 52.589D_{3t} + 12.598D_{4t}$$

$$Se = 7.517 \quad 7.517 \quad 7.517 \quad 7.517$$

$$t = 8.386 \quad 10.549 \quad 6.995 \quad 1.676$$

$$P-value \text{ "Sig"} = 0.000 \quad 0.000 \quad 0.000 \quad 0.096$$

$$F = 58.333 \qquad P-value \text{"Sig"} = 0.000$$

$$R^2 = 0.65$$

$$Q-test = 341.38 \quad P-value \text{"Sig"} = 0.000$$

	The first quarter D_1	The second quarter D_2	The third quarter D_3	The fourth quarter D_4
Coefficients	The average rainfall in the first quarter = $\hat{\beta}_1 = 63.048$	The average rainfall in the second quarter = $\hat{\beta}_2 = 79.312$	The average rainfall in the third quarter = $\hat{\beta}_3 = 52.589$	The average rainfall in the fourth quarter = $\hat{\beta}_4 = 12.598$
$Var(\hat{\beta}_i) = Se^2_{\hat{\beta}_i}$	For all regression coefficients, standard error = 7.517 it follows that the variance is constant			
t- test	the first quarter is statistically significant because the $p-value = 0.000 < \alpha$ "0.05"	the second quarter is statistically significant because the $p-value = 0.000 < \alpha$ "0.05"	the third quarter is statistically significant because the $p-value = 0.355 > \alpha$ "0.05"	the fourth quarter is non –statistically significant because the $p-value = 0.096 < \alpha$ "0.05"
F-test	According to the F test $F = 58.333$, $p-value = 0.000 < \alpha$ "0.05" The overall model is significant.			
R^2	The independent variable (D's) explained 65% from the variation in the rainfall "Y" other variables explained about 35%.			
Q-test	According to the Q-test we accept the null hypothesis that there is no auto-correlation ($Q-test = 341.38$ and $P-value$ "Sig" $= 0.000 > \alpha$ "0.05")			

Based on the analysis, we accept the null hypothesis that there is no auto-correlation according to the findings of the Q-test for all models. In group 1: The average rainfall was the highest in the first and fourth quarters, the Coefficients were (18.3 and 18.6) while in the second and third quarters the Coefficients were (7.5 and 1.7) respectively. In group 2: The highest average rainfall was in the first and fourth quarters, as shown in the Coefficients (37.3 and 22.9) while in the second and third quarters the Coefficients were (21.7 and 1.4). In group 3: The highest average rainfall was in the first and fourth quarters the Coefficients were (42.2 and 42.8) while in the second and third quarters the Coefficients were (22.7 and 13.8). In group 4: The highest average rainfall was in the first and fourth quarters the Coefficients were (42.2 and 42.8) while in the second and third quarters the Coefficients were (28.1 and 0.3). In group 5: The highest average rainfall was in the first, second and third quarters the Coefficients were (63.0, 79.3 and 52.6) while in the fourth quarter the Coefficient was (12.6). According to all the models, the highest average rainfall was in the first and fourth quarters except for the fifth model, the highest average rainfall was in the first, second and third quarters. The Kingdom of Saudi Arabia (KSA) has a very arid environment and a significant water shortage. In the dry and semiarid zones of the nation, rainfall distribution, quantity, and duration extremely varied on a regional and temporal scale. Rainfall is not uniform in time or place, and all geographic elements such as temperature, pressure, distance from the source of moisture, and terrain have a role to play (Noy-Meir, 1973; Subyani, 2004; Kwarteng et al., 2009; El-Nesr et al., 2010). (Abdullah and Al-Mazroui, 1998; Hasanean and Almazroui, 2015).

The T-test on Group1 showed that all the quarters have statistically significant effect on the average rainfall except for the third quarter. Group2 showed that all the quarters have statistically significant effect on the average rainfall except for the third quarter. Group3 showed that all the quarters have statistically significant effect on the average rainfall. Group4 showed that all the quarters have statistically significant effect on the average rainfall except for the third quarter. Group5 showed that all the quarters have statistically significant effect on the average rainfall except for the fourth quarter. That is the mean in the models (first, second, and fourth) all the quarters have statistically significant effect on the average rainfall except for the third quarter. In the third model, all the quarters have a statistically significant effect on average rainfall. In the fifth model, all the quarters have statistically significant effect on the average rainfall except for the fourth quarter.

According to the F- test, all the models are statistically significant.

According to R^2 for all the models, the independent variable (D's) explained 65% to 71% of the variation in rainfall "Y" whereas other factors explained around 35%.

CONCLUSION:

The goal of the study is to evaluate rainfall in Saudi Arabia, and there were 5-group categories. Each group was tested to determine the quarter average rainfall, using Regression Model and a dummy variable technique. The process described the link between average rainfall (dependent variable) and quarters (independent variables), with the goal of knowing which quarter has the highest average rainfall in terms of statistical significance. The results show that the highest rainfall for all the groups was in the first and fourth quarters of the years that the study analyzed, and the mean in the months (October till March) is nearly similar to the climate seasons in Saudi Arabia, except group 5.Meanwhile, Khamis Mushait and Abha stations which have rainfall in nearly all months, the results show that the first, second and third quarters had the highest rainfall. The F- test results show that all the models are statistically significant. The T- test indicates that all the quarters have statistically significant effect on the average rainfall, except for the fourth quarter in group5 and the third quarter in groups 1, 2 and 4. For all the models R^2 shows that the independent variable (D’s) explained 65% -71% of the variation in the rainfall “Y”, other variables explained about 35%.

Meanwhile, the Dummy Variable Method allows the monthly data of 33 years (396 month) to be analyzed in 4-quarter of a year. Thus, reducing to 132 quarters which also reduces the time of the data analysis and errors. Another important fact is most rainfall records have missing data or inaccurate reading, so it useful to minimize the errors. In general, all water harvesting projects depend on the seasonal rainfall. Therefore, it is essential to enhance this technique in future for forecasting purposes.

REFERENCES

Abdullah, M. A., & Al-Mazroui, M. A. (1998). Climatological study of the southwestern region of Saudi Arabia. I. Rainfall analysis. *Climate Research*, *9*(3), 213-223.

Elagib, N. A. (2011). Changing rainfall, seasonality and erosivity in the hyper-arid zone of Sudan. *Land degradation & development*, *22*(6), 505-512.

El-Nesr, M., Abdulrahman, A., & Abu-Zreig, M. (2010). Analysis of evapotranspiration variability and trends in the Arabian Peninsula. *American journal of environmental sciences*, *6*(6), 535-547.

Fetene, Z. A., Weldegerima, T. M., Zeleke, T. T., & Nigussie, M. (2018). Harmonic analysis of precipitation time series in Lake Tana Basin, Ethiopia. *Advances in Meteorology*, *2018*.

Damodar, G. N. (2003). *Basic Econometric*. (4th ed.). New York: MeGraw-Hill.

Hafez, Y. (2019). A recent study concerning the climatic variability over the kingdom Saudi Arabia for the period 1948-2018. *Journal of Geoscience and Environment Protection*, *7*(8), 268-289.

Hag-elsafi, S., & El-Tayib, M. (2016). Spatial and statistical analysis of rainfall in the Kingdom of Saudi Arabia from 1979 to 2008. *Weather*, *71*(10), 262-266.

Hanke, J. E., & Wichern, D. W. (2013). *Business Forecasting: Pearson New International Edition*. (9th ed.). Pearson Higher Ed.

Hasanean, H., & Almazroui, M. (2015). Rainfall: features and variations over Saudi Arabia: A review. *Clim*, 3, 578–626.

Kwarteng, A. Y., Dorvlo, A. S., & Vijaya Kumar, G. T. (2009). Analysis of a 27-year rainfall data (1977–2003) in the Sultanate of Oman. *International Journal of Climatology: A Journal of the Royal Meteorological Society*, *29*(4), 605-617.

Levine, D. M., Berenson, M. L., & Stephan, D. (2000). Estatística: teoria e aplicações. *Rio de janeiro: LTC*, *811*.

Luong, T. M., Dasari, H. P., & Hoteit, I. (2020). Extreme precipitation events are becoming less frequent but more intense over Jeddah, Saudi Arabia. Are shifting weather regimes the cause?. *Atmospheric Science Letters*, *21*(8), e981.

Merabtene, T., Siddique, M., & Shanableh, A. (2016). Assessment of seasonal and annual rainfall trends and variability in Sharjah City, UAE. *Advances in Meteorology*, *2016*, 1-13.

Modarres, R., & da Silva, V. D. P. R. (2007). Rainfall trends in arid and semi-arid regions of Iran. *Journal of arid environments*, *70*(2), 344-355.

Noy-Meir, I. (1973). Desert ecosystems: environment and producers. *Annual review of ecology and systematics*, *4*(1), 25-51.

Sahriman, S., Djuraidah, A., & Wigena, A. H. (2014). Application of principal component regression with dummy variable in statistical downscaling to forecast rainfall. *Open Journal of Statistics*, *4*(09), 678.

da Silva, V. D. P. R. (2004). On climate variability in Northeast of Brazil. *Journal of Arid Environments*, *58*(4), 575-596.

Subyani, A. M. (2004). Geostatistical study of annual and seasonal mean rainfall patterns in southwest Saudi Arabia/Distribution géostatistique de la pluie moyenne annuelle et saisonnière dans le Sud-Ouest de l'Arabie Saoudite. *Hydrological Sciences Journal*, *49*(5).

Tarawneh, Q., & Kadıoğlu, M. (2003). An analysis of precipitation climatology in Jordan. *Theoretical and Applied Climatology*, *74*, 123-136.

Tarawneh, Q. (2016). Harmonic analysis of precipitation climatology in Saudi Arabia. *Theoretical and applied climatology*, *124*, 205-217.