# The Effect of Leadership in Reducing Indiscipline among Employees (Applied Study on the North Refineries Company)

# أثر القيادة في تخفيض حالات عدم الانضباط لدى الموظفين: دراسة تطبيقية على شركة مصافي الشمال

أ.م.د. فارس غانم الطائي أ.م.د. ريم علي الجراح

كلية طب الاسنان / جامعة الموصل

تاريخ استلام البحث 2023/9/19 تاريخ قبول النشر 2023/11/26 تاريخ النشر 2024/ 7 /5

**Abstract:**

The research seeks to demonstrate the impact of leadership in reducing Indiscipline among employees at the Northern Refineries Company. The research was applied to a sample of (239) observations from the aforementioned company.

In our research, applied and multivariate statistical analysis was used to determine the role of leadership in preventing the spread of impolite behavior among workers at the North Refineries Company. The research also addressed the hypotheses that must be available theoretically and practically to use this analysis. A questionnaire was used to obtain the necessary data and ensure that it met the assumptions of the analysis. Finally, significant effects were shown between the variables under discussion, and other results were reached, for instance, the importance continuing to train employees to control their anger and emotions through seminars and workshops.

***Key words:*** *Leadership , Reducing Indiscipline among Employees , North Refineries Company .*

**المستخلص:**

يسعى البحث الى بيان تأثير القيادة في تخفيض حالات عدم الانضباط لدى الموظفين في شركة مصافي الشمال،اذ طبق البحث على عينة مكونة من (239) مشاهدة من الشركة المذكورة. وفي هذا البحث تم استخدام التحليل الإحصائي التطبيقي والمتعدد المتغيرات لتحديد دور القيادة في منع انتشار حالات عدم الانضباط بين العاملين في شركة مصافي الشمال ، كما تناول البحث الفرضيات التي يجب أن تتوافر نظرياً وعملياً لاستخدام هذا التحليل، وتم استخدام استبانة للحصول على البيانات اللازمة والتأكد من أنها تلبي فرضيات التحليل، وأخيراً تم إظهار تأثيرات معنوية بين المتغيرات قيد البحث، وتم التوصل إلى نتائج أخرى، على سبيل المثال، أهمية الاستمرار في تدريب الموظفين على التحكم في غضبهم وانفعالاتهم من خلال الندوات وورش العمل.

***الكلمات المفتاحية:*** *القيادة، تخفيض حالات عدم الانضباط ، شركة مصافي الشمال*

## INTRODUCTION

There is no doubt that impolite behavior, whether explicit or implicit, violates etiquette, and its spread certainly causes great harm among members of society. And according to the hadith of the Holy Prophet, "Kuluukum Rrae Wa Kuluukum Maswowol Ean Raeiatihi", it is the responsibility of the patron or the head of the place to limit this phenomenon.

If you examine two or more random variables that are not equivalent measures, you can determine whether there is a correlation between them. Regression is used to determine relationships, while correlation is used to determine how strong those relationships are.

Regression analysis has three main applications: determining causal relationships, forecasting, and prediction [1].

There are two questions that come to mind while comparing various variables: "Are there relationships between the variables?" and "How strong is that relationship?" Regression and correlation can be used to provide the answers to thèse queries.

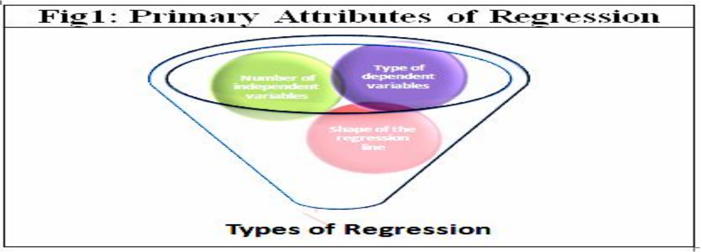
Regression determines the strength of the linear relationship, whether there is a relationship or correlation, and introduces both of thèse concepts.

Researchers can analyze relationships between one dependent variable (DV) and one or more independent variables (IDVs) using regression analysis. The outcome we care about is often the (DV), and the (IDVs) are the tools we have to use to get there. Regression analysis's main advantages are that it can:

- State whether there is a meaningful relationship between an IDVs and DV
- illustrate the relative strength of various IDVs' impacts on a DV.
- Make predictions [2, pp.193-233].

There are various different types of regression models to create predictions. These techniques are mostly based on three key factors, as shown in Fig. 1, which are: first, the number of IDV; second, the nature of DV; and lastly, the shape of the regression line.

Fig1: Primary Attributes of Regression




So, the types of regression analysis are:

- Linear Regression.
- Polynomial Regression.
- Logistic Regression.
- Ridge Regression.
- Lasso Regression.
- Bayesian Linear Regression.

Anyway, Linear Regression is the most basic and commonly used regression technique and is of two types: Simple and Multiple Regression.

You can use Simple linear regression when there is a single dependent and a single independent variable.

Multiple linear regression on the other hand can be used when we have one continuous dependent variable and two or more independent variables. Importantly, the independent variables could be quantitative or qualitative [1].

## MATRIAL & METHODS :

### Assumption about the data in Regression Analysis:

Before starting a regression analysis, a number of data needs and assumptions must be taken into account.

The first is that we require a sufficiently enough sample size as our initial source of data. Acceptable sample sizes are determined by the smallest sample size at which you have a reasonable chance of finding significant results, if any are already available. The second is that a regression model cannot be estimated if the variables have no variance. Regression is not necessary, in particular, because we already know the value of the dependent variable if it is constant (i.e., there is no variation in it). Similar to how if an independent variable does not change, it cannot account for any variation in the dependent variable. The dependent variable must be scaled either on an interval or on a ratio basis, which is the third data requirement. Alternative methods of regression must be employed if the data are not interval or ratio scaled, If the dependant variable is binary and can only take on two values (for instance, zero and one), you should use binary logistic regression, If multinomial logistic regression is appropriate. The last data criterion is the absence of or a minimal amount of collinearity - two independent variables are strongly connected-. [3].

### The linear Regression Model:

A statistical model is a mathematically represented abstraction of reality. All statistical models include assumptions in order to achieve this simplification, no exemption applies to linear regression (LR).

When applied properly, (LR) is a potent statistical instrument that may explain and predict real-world phenomena; yet, a misunderstanding of its underlying assumptions can result in false and deceptive inferences.

(LR) models serve a variety of functions and are helpful during both the planning and analysis phases of research. Forecasting, describing or explaining data, estimating parameters, choosing or screening variables, and controlling output [4, pp. 590-596].

**Simple Linear Regression (SLR):**

When modeling a relationship between two variables using simple linear regression (SLR), we can use a model of the following form::

$$Y = \beta_0 + \beta_1 X + \varepsilon \qquad (3.1)$$

Where: Y is the "dependent or response" variable, and X is the "independent or predictor" variable. The model's error term is a random variable, in this sense, the term "error" refers to random fluctuations, measurement errors, or the impact of outside forces rather than a mistake. The model in (3.1)'s linearity is a presumption. Other presumptions, such as the independence of the observed values of Y and the distribution of the error terms, are frequently added. We estimate 0 and 1 using observed values of X and Y, and we draw conclusions for 0 and 1 by testing hypotheses and calculating confidence ranges. In the event that we decide to forecast or anticipate the value of Y for a specific value of X, a measure of forecasting accuracy may also be of relevance.

**Multiple Linear Regression (MLR):**

Multiple predictor variables frequently affect the answer Y. The response Y is modeled linearly in relation to various predictors as follows

$$Y = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \cdots + \beta_n X_n + \varepsilon \qquad (3.2)$$

Provides for random variation in Y not explained by the X variables. This random variation may be due partly to other variables that affect Y but are not known or not observed.

The parameters $(\beta_i, i = 0,1, \dots, n)$ -the beta coefficients- that we are regarded in estimating. These estimates are usually given by a method known as ordinary least squares (OLS).

LR models are used for various purposes, Prediction, Data Description or explanation, Parameter Estimation, Variable Selection or Screening, & Control of Output [5, p.479].

**The Efficiency Model:**

Regression models are implemented in various ways, and the data structure has an impact on the effectiveness of the model. Different algorithmic patterns aid in identifying the parameters required for making predictions.

There are various approaches to selecting a model. More effective models have higher adjusted and predicted R-squared values. A statistically significant result for the hypothesis is indicated by independent variable p-values that are lower than the significance level.

In other hand, Mean Square root of the variance of the residuals (MSE) is a useful indicator of how precisely the model predicts the behavior. Better fit is indicated by lower MSE values. [2 pp. 499-510, [6],[7].

**Important Statistics:**

**Collinearity Statistics tests:**

Collinearity and Multi-Collinearity negatively affect multiple regression in a number of ways, including how the findings are acquired and how they are interpreted.

**Tolerance test (TOL)**: as a measurement of collinearity and multi-collinearity

$$TOL_i = 1 - R_i^2 \qquad \ldots \quad (4-1)$$

Where: $R_i^2$ : is the coefficient that determines how well the other predictor variables predict the variable i-th.

As the tolerance value approaches zero, it indicates that the variable is highly predicted (collinear) with the other predictor variables, depending on how close to 1 or 0 the coefficient of determination between the variables is. [8].

**Variance Inflation Factor (VIF)**: used as a measurement of how other predictor variables affect the regression coefficient. [9, p.130].

$$VIF_i = (TOL_i)^{-1} \qquad \ldots . \quad (4-2)$$

The $(VIF_i)^{1/2}$ measures how much multicollinearity has increased the standard error of the regression coefficient. A high level of collinearity or multi-collinearity among the independent variables is indicated by large VIF values, with a typical threshold of 10.0.
.

**Autocorrelation Statistics tests:**

The idea of testing for auto-correlation is to determine whether the real disturbances are connected in any way, which can be done by looking at the auto-correlations of the Least Squares Residuals (LSR).

Durbin-Watson statistic test (DW), used to determine whether or not the mistakes have a first-order auto-correlation. It is the first formally described method for determining autocorrelation using residuals from the least squares method. For a data collection of size n, the test statistic for the Durbin-Watson test is given by: [9, p.962].

$$DW = \frac{\sum_{t=2}^{n}(e_t - e_{t-1})^2}{\sum_{t=1}^{n} e_t^2} \; \ldots\ldots (4-3)$$

The residuals from the Ordinary least Squares (OLS) fit are: $e_t = y_t - \hat{y}_t$.

$DW \approx 2(1-r)$, where (r) is the sample auto-correlation coefficient from residuals based on OLS, is the formula for big n.

The DW test has a range of 0 to 4, with values between 0 and 2 indicate positive auto-correlation, 2 indicates zero auto-correlation, and values between 2 and 4 indicate negative auto-correlation [9].

**Hetroscedactisty (homoscedasticity) Statistics Tests:**

It depends on homoscedasticity and heteroscedasticity whether the variances of the predictions produced by regression change or stay unchanged. However, that does not mean that examples of linear regression through the origin are the only situations in which heteroscedasticity is taken into account. [10, pp. 101-102].

**Breusch Pagan LM test:**

Used to detect whether Homoscedasticity is existing in a regression model or not. Here, the null and alternative hypotheses are:

($H_0$): Homoscedasticity is present. *v.s.* ($H_1$): Heteroscedasticity is present

$$T_{LM} = NR^2 \; \ldots\ldots (4-4)$$

Where: N: Sample Size, R2: coefficient of determination of assistance regression (squared residuals as the response values) [11, pp.277-280].

**Outlier Statistic Tests:**

An item of data that differs abnormally from other values is known as an outlier. It is information that is different from the other values, in other words. Measures are used to identify situations where the independent variables have an unusually large number of values and situations where the regression model may be significantly affected.

**Mahalanobis Distance ($D^2_{(i)M}$):**

based on a chi-square distribution, evaluated using using p < 0.001, with 'k' degrees of freedom, this metric expresses how much a case's values for the independent variables deviate from the average of all cases.

A case is known to have extreme values for one or more of the independent variables when the Mahalanobis distance is substantial.

$$D^2_{(i)M} = (X - m)^T . C^{-1} . (X - m) \dots \dots (4 - 5)$$

Where X is the observation's vector, m is the vector of independent variable mean values, and C is the inverse covariance matrix of the same independent variables.

**Cook's Distance, $D^2_{(i)C}$:**

a measurement of how much the residuals of all cases would alter if a particular instance were eliminated from the calculation of the regression coefficient. A high number indicates that the regression statistic is essentially changed when a case is not included in the computation. [12, p.400].

$$D^2_{(i)C} = \frac{\sum_{j=1}^{n} (\hat{Y}_j - \hat{Y}_{j(i)})^2}{k * MSE} \dots \dots (4 - 6)$$

Where: The jth fitted response value is given by $\hat{Y}_j$

The jth fitted response value, where the fit excludes observation i, is denoted by $\hat{Y}_{j(i)}$

$MSE$: Mean Square Error.

In each observation, compare the Cooks value to (4/n), where n is the total number of observations. Values higher than this reflect observations that may pose a concern. [13].

**Practical Part:**

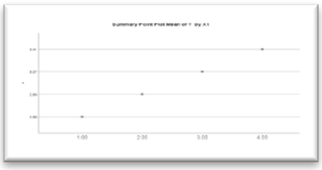
**Sample Research:**

A questionnaire was distributed to the employees of the North Refineries Company / Beeji, which included several axes with their variables shown in Appendix (A) below.

**The Assumptions:**

**Assumption 1:** The relationship between the IVs and the DV is linear. We can see that a straight line can be used to model the relationship between the (IV) and the (DV), indicating that the relationship between these variables is linear..

Relationship between the $x_1$ and the Y

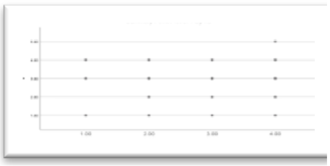


Relationship between the $x_2$ and the Y

Relationship between the $x_3$ and the Y

Relationship between the $x_4$ and the Y

Relationship between the $x_5$ and the Y



**Assumption 2:** There is no multi-collinearity in your data.

Table (1): Spearman Correlations

| Var. | Y | $X_1$ | $X_2$ | $X_3$ | $X_4$ | $X_5$ |
|---|---|---|---|---|---|---|
| Y | 1 | 0.235** | 0.072 | 0.136* | 0.219** | 0.149* |
| $X_1$ | 0.235** | 1 | 0.293** | 0.438** | 0.491** | 0.256** |
| $X_2$ | 0.072 | 0.293** | 1 | 0.268** | 0.339** | 0.325** |
| $X_3$ | 0.136* | 0.438** | 0.268** | 1 | 0.488** | 0.335** |
| $X_4$ | 0.219** | 0.491** | 0.339** | 0.488** | 1 | 0.474** |
| $X_5$ | 0.149* | 0.256** | 0.325** | 0.335** | 0.474** | 1 |

(*) : Significant at $P \leq 0.05$ (**): Highly Significant at $P \leq 0.01$

As the highest correlation is (r = 0.491), there is a weak correlation between the variables under study, in other hand. To evaluate this premise, we can perform the (VIF) and (TOL) statistical test for collinearity. We need (VIF) scores to be under 10 and (TOL) scores to be above (0.2) for the assumption to hold, which they do in our data (see table (2).

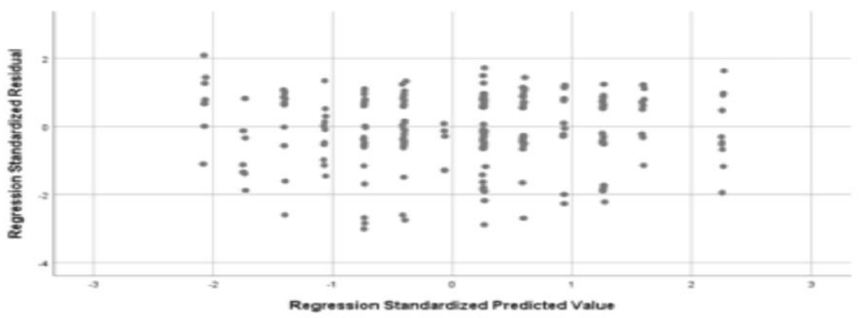
Table (2): Collinearity Statistics

| Variables | Tolerance | VIF |
|---|---|---|
| $X_1$ | 0.695 | 1.438 |
| $X_2$ | 0.827 | 1.209 |
| $X_3$ | 0.695 | 1.440 |
| $X_4$ | 0.573 | 1.744 |
| $X_5$ | 0.735 | 1.361 |

**Assumption 3:** The values of the residuals are independent. By use the Durbin-Watson statistic test (DW. = 1.743), our residuals are independent (or uncorrelated).

**Assumption 4:** The variance of the residuals is constant (homoscedasticity). By using visual tools, as shown in Fig. (2), It seems that the shape is quite far from the funneled form, i.e.: the assumption is probably met.

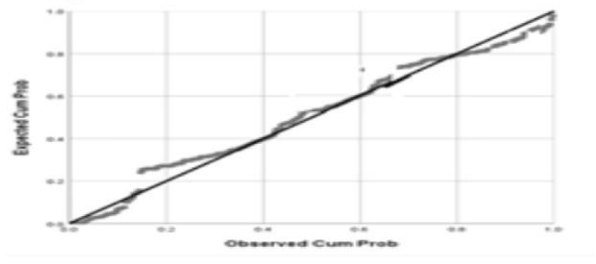
Fig(2): Scatter Plot




On the other hand. By use the Breusch Pagan LM test, since (TLM = 8.381) < ($\chi 2$ (4, 0.05) = 9.49), this means that there is no significant evidence for the difference in variance at 5%.

**Assumption 5:** The values of the residuals are normally distributed. Also, by using visual tools, from Fig. (3) Below, the scatter plot show:

Fig (3) : Normal P-P Plot of Regression Standardized Residual




**Assumption 6:** There are no influential cases biasing your model. From table (3) below: **First:** Mahal. Distance, the max value of $D^2_{(i)M}$ is (18.633) less than $\chi^2_{(5,0.01)} = 20.52$, which indicate no outlier point. **Second:** Cook's Distance, the max value of $D^2_{(i)C}$ is (0.091) less than (1), which indicate no undue influence on the model.

**Table (3): Residuals Statistics**

| | Min. | Max. | Mean | Std. Dev. | N |
|---|---|---|---|---|---|
| Mahal. Dis. | 0.427 | 18.633 | 4.983 | 3.773 | 289 |
| Cook's Dis. | 0.000 | 0.091 | 0.004 | 0.009 | 289 |

**Regression Model:**

By using a Standard Multiple Regression, in which the set of IVs that make up the model were determined. We obtain the result of the ANOVA of the regression. The regression model is significant according to ($F_{(4,283)}$: 4.454) and (sig. value: $0.001 < 0.05$).

It is time to examine the individual coefficients after determining how well the model fits overall.

These can be found in Table (4). First, while examining the various parameters, t-values check to see if the regression coefficients are all zero. The parameter is important if this is the case.

**Table (4): Coefficients of Regression**

| | Un-Standardized Coefficients | | Standardized Coefficients | T | Sig. |
|---|---|---|---|---|---|
| | B | Std. Error | Beta | | |
| Constant | 2.232 | 0.260 | | 8.587 | 0.000 |
| $X_1$ | 0.180 | 0.071 | 0.175 | 2.547 | 0.011 |
| $X_2$ | -0.033 | 0.055 | -0.038 | -0.600 | 0.549 |
| $X_3$ | -0.010 | 0.065 | -0.011 | -0.157 | 0.875 |
| $X_4$ | 0.119 | 0.073 | 0.122 | 1.617 | 0.107 |
| $X_5$ | 0.062 | 0.066 | 0.062 | 0.930 | 0.353 |

One significant coefficient with a p-value below the threshold of 0.05 is found in the aforementioned model.

By examining the standardized coefficient values, it is evident that the association between ($X_1$: *Adopting a method of direct and indirect communication to address undisciplined behaviors among employees*) and (Y: *Tension and anger over inappropriate behavior and negative signals among employees*) is the strongest..

The collection of IVs is included to or excluded from the model using a Progressive Multiple regression (Stepwise approach) based on the strength of the correlation between them and the DV. We get the ANOVA and regression results. The regression model is highly significant according to ($F_{(2,286)}$: 10.677) and (sig. value: $0.000 < 0.001$.

From table (5) bellow, it is concluded two significant coefficients, with p-values below the used level of 0.05.

By looking at the standardized coefficients values, both of IVs ($X_1$: *Adopting a method of direct and indirect communication to address undisciplined behaviors among employees*) and ($X_4$: *Training employees on anger management and addressing negative behaviors through seminars and workshops*), respectively, are have clearly the strongest relationship with DV (Y: *Tension and anger over inappropriate behavior and negative signals among employees*).

**Table (5): Significant Coefficients of Regression**

| | Un-standardized Coefficients | | Standardized Coefficients | T | Sig. |
|---|---|---|---|---|---|
| | B | Std. Error | Beta | | |
| Constant | 2.286 | 0.224 | | 10.196 | 0.000 |
| $X_1$ | 0.1725 | 0.067 | 0.167 | 2.560 | 0.011 |
| $X_4$ | 0.1333 | 0.064 | 0.137 | 2.095 | 0.037 |

Finally, the equation of regression can be obtained from the un-standardized coefficients values, so the resulting equations from Enter and Stepwise methods, respectively are:

Enter method:

$$y = 2.232 + 0.180X_1 - 0.033X_2 - 0.010X_3 + 0.119X_4 + 0.06X_5$$

Stepwise method:

$$y = 2.286 + 0.1725X_1 + 0.133X_4$$

Evaluation of Multiple Regression Equations**:**

Table (6): Goodness of fit

| Methods | Coefficient | Adj. $R^2$ | MSE |
|---|---|---|---|
| Enter method | $X_1$, $X_2$, $X_3$, $X_4$ | 0.057 | 0.676 |
| Stepwise method | $X_1$, $X_4$ | 0.063 | 0.672 |

**Conclusions:**

1. The result of the statistical analysis showed a significant effect for both of (Adopting a method of direct and indirect communication to address undisciplined behaviors among employees), (Training employees on anger management and addressing negative behaviors through seminars and workshops) over (Tension and anger over inappropriate behavior and negative signals among employees).

2. There are highly significant relationships for both (Adopting a method of direct and indirect communication to address undisciplined behaviors among employees), (Training employees on anger management and addressing negative behaviors through seminars and workshops) with (Tension and anger over inappropriate behavior and negative signals among employees).
3. Although there is a statistically significant relationship for both (Reducing psychological stress during work through dialogue with workers, despite their different cultural and intellectual levels) and (The use of hard and soft style in dealing with inappropriate behavior among employees) with (Tension and anger over inappropriate behavior and negative signals among employees), they are not included in the regression model.
4. Interest in organizing workshops and seminars to train employees on anger management and positive behavior.

**Recommendations:**
1. Conducting the research on a more comprehensive set of variables using multilevel regression analysis.
2. Interest in organizing workshops and seminars to train employees on anger management and positive behavior.

**References:**

[1] NJ Gogtay, SP Deshpande, UM Thatte: “Principles of Regression Analysis”, Journal of the Association of Physicians, India, 2017, vol. 65.

[2] Marko Sarstedt, Erik Mooi,: “A Concise Guide to Market Research”, Springer-Verlag Berlin Heidelberg, 2014,ch.7, pp. 193-233.

[3] Jason W. Osborne and Elaine Waters,: “Four Assumptions of Multiple Regression That Researchers Should Always Test”, Practical Assessment, Research, and Evaluation, 2017 , 8(2).

[4] Robert J. Casson D. Phil Franzco, Lachlan DM Farmer: “Understanding and checking the assumptions of linear regression: a primer for medical researchers”, Clinical and Experimental Ophthalmology, MBBS 2014, 42: 590–596.

[5] Alvin C. Rencher and G. Bruce,: “Linear Models in Statistics”, 2ed edition, John Wiley & Sons Inc. ,2008 P. 479.

[6] Field, A.: “Discovering statistics using SPSS” ,4th ed., London, Sage, 2013,

[7] Aiken, L. S., & West, S. G. :“ Multiple regression: Testing and interpreting interactions”, Thousand Oaks, CA: Sage. 1991,

[8] Oke, J. A., Akinkunmi, W. B. Etebefia, S. O,: “ Use of Correlation, Tolerance and Variance Inflation Factor for Multi-collinearity Test”, GSJ: Volume 7, Issue 5, May 2019, ISSN 2320-918, www.Global Scientific Journal.com.

[9] William H. Greene: “Econometric Analysis ” , Prentice Hall, 7th ed., 2012, Pp.130-132.

[10] Kleiber, Christian; Zeileis, Achim :“Applied Econometrics with R.” New York: Springer. ISBN 978-0-387-77316-2 2008, Pp. 101–102.

[11] Bade H, Batagi QIL, “ A joint test for serial correlation and random individual effects”, Statistics & Probability Letters Journal, Vol. 11, Iss. 3 1991, pp. 277-280.

[12] Neter, J., M. H. Kutner, C. J. Nachtsheim, and W. Wasserman. “Applied Linear Statistical Models”. 4th ed. Chicago: Irwin, 1996, P.400.

## Appendix (A): The Questionnaire of Sample

| Variables | The Description | Saturation | Cod |
|---|---|---|---|
| Y | Tension and anger over inappropriate behavior and negative signals among employees | Never | 1 |
| | | Rarely | 2 |
| | | Occasionally | 3 |
| | | Always | 4 |
| $X_1$ | Adopting a method of direct and indirect communication to address undisciplined behaviors among employees | Never | 1 |
| | | Rarely | 2 |
| | | Occasionally | 3 |
| | | Always | 4 |
| $X_2$ | Encouraging social communication between workers in various events and forums | Never | 1 |
| | | Rarely | 2 |
| | | Occasionally | 3 |
| | | Always | 4 |
| $X_3$ | Reducing psychological stress during work through dialogue with workers, despite their different cultural and intellectual levels | Never | 1 |
| | | Rarely | 2 |
| | | Occasionally | 3 |
| | | Always | 4 |
| $X_4$ | Training employees on anger management and addressing negative behaviors through seminars and workshops | Never | 1 |
| | | Rarely | 2 |
| | | Occasionally | 3 |
| | | Always | 4 |
| $X_5$ | The use of hard and soft style in dealing with inappropriate behavior among employees | Never | 1 |
| | | Rarely | 2 |
| | | Occasionally | 3 |
| | | Always | 4 |