PART - V

Statistical Analysis
CHAPTER VII

ANOVA TECHNIQUE AND REGRESSION ANALYSIS

Introduction

Scientific research is one of the fields where statistics is inevitable. According to A.M. Mood, "Statistics provides tools and techniques for research workers." Experimentation and making inferences are twin essential features of general scientific methodology. Statistics as a scientific discipline is mainly designed to achieve these objectives. Design and analysis of experiments fall in the sphere of data collection and interpretation techniques. Inferences can be made from the experimental data in the best possible manner using suitable statistical techniques. The two powerful tools used in this study are Analysis of Variance (ANOVA) and "Regression Analysis".104-108

7.1 Analysis of Variance (ANOVA)

7.1.1 Methodology

In order to provide a more rigorous basis and a sound statistical treatment, the technique of analysis of variance (ANOVA) is used. The analysis of variance is a powerful statistical tool for tests of significance. The term Analysis of variance was introduced by R.A. Fisher in 1920's to deal with problems in the analysis of agronomical data. According to R.A. Fisher, Analysis of Variance (ANOVA) is the separation of variance ascribable to one group of causes from the variance ascribable to another group. The ANOVA model suitable for two valuable factors say, concentration and temperature on certain physical parameters such as ultrasonic velocity, internal pressure, free volume, solvation number, apparent molal volume, apparent molal compressibility and equivalent conductance is discussed below.

The temperature is kept at different levels 'i' and the molality is also fixed at different levels 'j'.

81
In such two factor experiments, the observations can be arranged in a two way layout or a M x N table, where each row corresponds to a level $A_i$ of $A$ and each column corresponding to a level $B_j$ of $B$. $Y_{ij}$ be the value of the physical parameter on the $(i,j)^{th}$ cell $i = 1, 2, 3 \ldots p$ and $j = 1, 2, 3 \ldots q$.

<table>
<thead>
<tr>
<th>Factor A</th>
<th>$B_1$</th>
<th>$B_2$</th>
<th>$B_j$</th>
<th>$B_N$</th>
<th>Row Total</th>
</tr>
</thead>
<tbody>
<tr>
<td>$A_1$</td>
<td>$Y_{11}$</td>
<td>$Y_{12}$</td>
<td>$Y_{1j}$</td>
<td>$Y_{1N}$</td>
<td>$T_1$</td>
</tr>
<tr>
<td>$A_2$</td>
<td>$Y_{21}$</td>
<td>$Y_{22}$</td>
<td>$Y_{2j}$</td>
<td>$Y_{2N}$</td>
<td>$T_2$</td>
</tr>
<tr>
<td>$A_N$</td>
<td>$Y_{M1}$</td>
<td>$Y_{M2}$</td>
<td>$Y$</td>
<td>$Y_{MN}$</td>
<td>$T_M$</td>
</tr>
<tr>
<td>$A_M$</td>
<td>$Y_{M1}$</td>
<td>$Y_{M2}$</td>
<td>$Y$</td>
<td>$Y_{MN}$</td>
<td>$T_M$</td>
</tr>
<tr>
<td>Column Total</td>
<td>$T_1$</td>
<td>$T_2$</td>
<td>$T_j$</td>
<td>$T_N$</td>
<td>$T$</td>
</tr>
</tbody>
</table>

In the case of two-way classified data with one observation per cell, the interaction effect (i.e. combined effect of temperature and molality) is taken as zero.

The mathematical model suitable here is

$$Y_{ij} = \mu + t_i + c_j + e_{ij}$$

Where $Y_{ij}$ is an observation coming from a unit defined by levels and

$$\Sigma t_i = \Sigma c_j = 0$$

( $i, j , i = 1, 2, \ldots M, j = 1, 2 \ldots N$ ), $\mu$ is the general effect, $t_i$ is the temperature effect and $C_j$ is the concentration effect.

The null hypothesis is
$H_{01}$ : The treatment (concentration / molality) effects on a physical parameter are equal i.e., the physical parameter is unaffected due to change in the levels of concentration potential.

$H_{02}$ : The block (temperature) effects on a physical parameter are equal i.e., the physical parameter is unaffected due to change in the levels of temperature.

The inferences can be drawn using ANOVA table. The layout of the table is Table 7.1

<table>
<thead>
<tr>
<th>Source</th>
<th>Degree of freedom (DF)</th>
<th>Sum of squares due to (SS)</th>
<th>Mean sum of squares</th>
<th>F-Cal</th>
</tr>
</thead>
<tbody>
<tr>
<td>Due to molality</td>
<td>$M-1$</td>
<td>SSM</td>
<td>MSM=SSM/(M-1)</td>
<td>$F_m = \frac{MSM}{MSE}$</td>
</tr>
<tr>
<td>Due to Temperature</td>
<td>$N-1$</td>
<td>SST</td>
<td>MST=SST/(N-1)</td>
<td>$F_t = \frac{MST}{MSE}$</td>
</tr>
<tr>
<td>Due to error</td>
<td>$(M-1) (N-1)$</td>
<td>SSE</td>
<td>MSE=SSE/(M-1)(N-1)</td>
<td></td>
</tr>
<tr>
<td>Total</td>
<td>$MN-1$</td>
<td>TSS</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
Findings

The findings of the study are presented in Table 7.2.

### 7.2 THE FINDINGS OF ANOVA

<table>
<thead>
<tr>
<th>Physical Parameters</th>
<th>Factors</th>
<th>Group I</th>
<th>Group II</th>
<th>Group III</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>Tar</td>
<td>Pot</td>
<td>Di</td>
</tr>
<tr>
<td>U</td>
<td>Molality</td>
<td>HS</td>
<td>HS</td>
<td>S</td>
</tr>
<tr>
<td></td>
<td>Temperature</td>
<td>HS</td>
<td>HS</td>
<td>HS</td>
</tr>
<tr>
<td>Πi</td>
<td>Molality</td>
<td>HS</td>
<td>HS</td>
<td>HS</td>
</tr>
<tr>
<td></td>
<td>Temperature</td>
<td>HS</td>
<td>HS</td>
<td>HS</td>
</tr>
<tr>
<td>Vf</td>
<td>Molality</td>
<td>HS</td>
<td>HS</td>
<td>HS</td>
</tr>
<tr>
<td></td>
<td>Temperature</td>
<td>HS</td>
<td>HS</td>
<td>HS</td>
</tr>
<tr>
<td>nh</td>
<td>Molality</td>
<td>S</td>
<td>S</td>
<td>S</td>
</tr>
<tr>
<td></td>
<td>Temperature</td>
<td>S</td>
<td>S</td>
<td>S</td>
</tr>
<tr>
<td>φv</td>
<td>Molality</td>
<td>HS</td>
<td>HS</td>
<td>S</td>
</tr>
<tr>
<td></td>
<td>Temperature</td>
<td>S</td>
<td>HS</td>
<td>HS</td>
</tr>
<tr>
<td>φk</td>
<td>Molality</td>
<td>S</td>
<td>S</td>
<td>S</td>
</tr>
<tr>
<td></td>
<td>Temperature</td>
<td>S</td>
<td>HS</td>
<td>HS</td>
</tr>
<tr>
<td>Λc</td>
<td>Molality</td>
<td>HS</td>
<td>HS</td>
<td>HS</td>
</tr>
<tr>
<td></td>
<td>Temperature</td>
<td>S</td>
<td>HS</td>
<td>HS</td>
</tr>
</tbody>
</table>

Expansions for abbreviations:

- **S** = Significant
- **NS** = Not significant
- **HS** = Highly significant

**Note:**

**SSE = TSS - SSM - SST**

Inference: If F-Cal < F-Tab, the null hypothesis is accepted otherwise, it is rejected.
7.2 Findings from ANOVA Tables

Effects of Concentration/Molality and Temperature on Physical Parameters

Chapter 2

For all the systems under study, the effect of temperature on the ultrasonic velocity is highly significant. In the third group consisting of Copper tartrate, calcium tartrate and Barium tartrate solutions the effect of concentration on ultrasonic velocity is insignificant whereas for the other systems, the ultrasonic velocity is highly affected by the changes in concentration.

Chapter 3

The effect of temperature on internal pressure of the solutions studied here is highly significant. In the case of all solutions, the internal pressure is highly affected by the various levels of concentrations and this effect is insignificant in Sodium antimony tartrate and Barium tartrate solutions.

The effect of temperature on free volume of all the systems studied here is highly significant. The effect of changing the levels of molality is highly significant in all the systems studied here with the exception of Sodium antimony tartrate where the effect is only significant.
(a) Different levels of concentrations significantly affect the solvation number of almost all the solutions studied. In the case of sodium antimony tartrate and barium tartrate solution the effect is highly significant. It may be noted that the solvation number of all solutions studied here is significantly affected by different levels of temperatures.

(b) The effect of concentration on apparent molal volume of Tartaric acid, Potassium bitartrate, Sodium antimony tartrate and Barium tartrate is highly significant and the effect is significant in the case of Disodium tartrate, Copper tartrate and the effect is insignificant in the case Sodium potassium tartrate solutions. The changes in temperature have significant effect on apparent molal volume, whereas in the case of Tartaric acid, Copper tartrate solutions and for the remaining system effect is highly significant.

(c) Apparent molal compressibility of almost all the solutions are significantly affected by the changes in the concentration with the exception Sodium potassium tartrate for which the effect is insignificant similar to the variation of apparent molal volume. The effect of temperature on apparent molal compressibility is almost similar to the variation of apparent molal volume with respect to temperature.

Chapter 6

Various levels of concentration significantly affect equivalent conductance in the case of Copper tartrate, Calcium tartrate and Barium tartrate. The effect is highly significant for the remaining systems studied here. The effect of temperature on the equivalent conductance is highly significant in Disodium tartrate, Copper tartrate and Barium tartrate solutions and it is significant for Tartaric acid, Potassium Bitartrate, Sodium potassium tartrate and Sodium antimony tartrate.
In general the effect of temperature on all the physical parameters taken for statistical analysis in the present work, is highly significant.

7.3 Regression Analysis

7.3.1 Introduction

Regression analysis is a branch of statistical theory that is widely used in almost all the scientific disciplines. Sir Francis Galton first used the term regression in 1877 while studying the relationship between the height of fathers and sons. The definition of the term regression by Ya-Lum Chou is as follows.

"Regression Analysis attempts to establish the 'nature of the relationship' between variables—that is to study the functional relationship between the variable and thereby provide a mechanism for prediction, or forecasting."

This section deals with the joint variation of two variables namely concentration and physical parameters (Internal pressure, Apparent molal compressibility, etc.,) under different controlled temperatures. Concentration is considered as the independent variable, which is denoted by $X$ and the physical parameter is considered as dependent variable, denoted by $Y$. The situation now poses the question of how to setup the required equation.

If there is a fairly uniform increase in physical parameter with uniform increases in levels of concentration, the response of physical parameter to different levels of concentration appears to be linear, i.e., can be represented by a straight line, of the form

$$Y = A + BX.$$

On the other hand, if there is rise and fall in the response then it is non-linear.
In general the functions that can be used for fitting are divided into two groups: (1) polynomials and (2) exponentials. The exponentials are often referred to as logarithmic curves because they are transformed to logarithms before fitting. Typical examples are

**Polynomials**

**Linearizable function**

Y = A + BX (linear)  
Y = A\text{e}^{BX}

Y = A + BX + CX^2 (quadratic)  
Y = AX^B

Y = A + BX + CX^2 + DX^3 (cubic)  
Y = A + B \log X

Y = X / (AX - B)  
Y = e^A + BX / (1 + e^A) + BX

with a number of variations etc.

Selecting a suitable regression equation for a non-linear trend is not a simple problem because there are large numbers of equations. One requires the selection of the particular equation that is reasonable and gives a good fit.

### 7.3.2 Identifying Suitable Regression Equation

#### 1. Graphical Method or Scatter Diagram

This method is an approximate procedure. According to this method, the molality values are taken along the X-axis and the physical parameter values are taken along the Y-axis. All pairs of values can be plotted as points in a X-Y plane and a curve can be drawn connecting these points. Then comparing this curve with the curves available in mathematical literature, the appropriate curve may be identified.

#### 2. Principle of Least Square

The terms “best fit” is interpreted in accordance with LEGENDER’s principle of least squares which consists in minimizing the sum of squares of the deviations of the actual values of Y from their estimated values as given by the line of best fit. This method is a
powerful tool in the hands of a scientific researcher but it cannot be applied when the line of best fit may be of a growth curve.

3. Iterative Method

The growth curves are curves belong to exponential family with a constant C. A few of them are given below.

\[ Y = C + AB^X \] Modified exponential curve.

\[ Y = CA^{(B^X)} \] Gompertz curve.

\[ Y = \frac{C}{1 + e^{(B-X)}} \] Logistic curve.

\[ Y = \frac{C}{1 + e^{(a+bx)}} \] Logistic curve

\[ Y = \frac{C}{1 + BA^X} \] Sigmoid.

In all the above-mentioned curves one can’t apply a transformation to linearize it, because, three constants are to be found out with the help of two variables. To solve such kind of problem an iterative method can be followed. One must specify initial values for all parameters. Good initial values are important to provide a better solution in little iteration.

A Measure for Identifying the Best Regression

To find the best regression, the study considers a very useful measure known as percentage of fit, \( R^2 \). The ability of the fitted regression equation to explain, or account for, the variation in the dependent variable is quantified by \( R^2 \).

\[ \text{Percentage of fit} = R^2 = 100 \, r^2 \]

Where \( r \) is the correlation coefficient between the dependent variable and the independent variable. A percentage of fit of 98% tells us that the variations due to concentration (X) of the solutions
studied here, has accounted for 98% of the variation in the physical parameter(Y).

**Fitting a suitable regression equation**

In the present investigation \( \log (1/ V_f) \) is taken along X axis and the \( \log \pi_i \) values are taken along Y axis all pairs of values are plotted for a given concentration at all temperatures in an X Y plane and a straight line is drawn connecting these points. It is identified that the straight line is more suitable for the data.

![Sample graph - Tartaric acid](image)

Therefore the suggested Linear equation is

\[
\log \pi_i = \alpha \log (1/ V_f) + \log k
\]

which leads to \( \pi_i V_f^\alpha = k \).

The author has made use of the statistical software packages available for this study. The regression function which gives \( R^2 \) values more or less 99%, for different levels of molality are selected as the best regression function for that solution.
Table 7.3

<table>
<thead>
<tr>
<th>SAMPLE</th>
<th>$R^2$</th>
<th>$\pi V_t^\alpha = k$</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>$\alpha$</td>
<td>$K$</td>
</tr>
<tr>
<td>Tartaric acid</td>
<td>0.9998</td>
<td>0.2726</td>
</tr>
<tr>
<td>Potassium bitartrate</td>
<td>0.9998</td>
<td>0.2688</td>
</tr>
<tr>
<td>Disodium tartrate</td>
<td>0.9998</td>
<td>0.2709</td>
</tr>
<tr>
<td>Sodium potassium tartrate</td>
<td>0.9995</td>
<td>0.2661</td>
</tr>
<tr>
<td>Sodium antimony tartrate</td>
<td>0.9991</td>
<td>0.2731</td>
</tr>
<tr>
<td>Copper tartrate</td>
<td>0.9995</td>
<td>0.2796</td>
</tr>
<tr>
<td>Calcium tartrate</td>
<td>0.9998</td>
<td>0.2712</td>
</tr>
<tr>
<td>Barium tartrate</td>
<td>0.9997</td>
<td>0.2682</td>
</tr>
</tbody>
</table>

7.4 Findings of the study

The values of $\alpha$, $k$ and $R^2$ are presented in the table 7.3. Thus on the basis of statistical analysis $\pi V_t^\alpha = k$ is the appropriate mathematical model to the salts under study.
SUMMARY

Acoustics and spectroscopy are found to be the two important fields to study the molecular interactions in the liquid system. In the first part of the thesis, the fundamental nature of the thermodynamic parameters namely, the internal pressure and free volume in determining the interactions in the systems and hence the behaviour of liquids has been well established i.e., temperature, internal pressure and free volume are the three basic thermodynamic properties of liquid systems similar to temperature, pressure and volume for gaseous systems. A detailed analysis of the variation of internal pressure and free volume with temperature and concentration has been carried out. The results obtained in the present investigation support the idea that the systems namely, Tartaric acid, Potassium bitartrate, Disodium tartrate, Sodium potassium tartrate, Copper tartrate and Calcium tartrate salts in Formamide behave as structure makers and Sodium antimony tartrate and Barium tartrate salts in Formamide behave as structure breakers.

Acoustical parameters such as ultrasonic velocity, adiabatic compressibility, intermolecular free length, specific acoustic impedance, Rao's constant, Wada's constant and van der Waal's constant have been presented and their variations with respect to concentration and temperature are interpreted in terms of various molecular interactions.

The second part of the thesis highlights the work on solvation number, using compressibility method. Intermolecular interactions i.e., solute-solute, solute-solvent and solvent-solvent interactions are discussed in detail based on the negative, positive and zero values of solvation number. This part of the thesis also deals with the study of apparent molal volume and apparent molal compressibility. They have been proven to be very useful in explaining the structural interaction
occurring in solutions. The variation of apparent molal volume with concentration and temperature explains the solute-solvent interaction in the solutions of all the systems in Formamide. The higher apparent molal compressibility of Tartaric acid, Potassium bitartrate, Disodium tartrate, Sodium potassium tartrate, and Copper tartrate and Calcium tartrate solutions indicates that cation must cause this effect. The Structure promotion is evidently increased in such systems. In Sodium antimony tartrate and Barium tartrate solutions; there is a decrease in apparent molal compressibility despite the lower electrostatic field due to larger ions.

The third part of the thesis elaborates the study of spectroscopic behaviour of solution from the measurement of chemical shift using proton magnetic resonance. The various interactions in solutions have been confirmed by $^1$HNMR and IR spectral analysis. Spectroscopic study reveals that Tartaric acid, Potassium bitartrate, Disodium tartrate, Sodium potassium tartrate, Copper tartrate and Calcium tartrate salts have a strong solute-solvent interactions and the Sodium antimony tartrate and Barium tartrate salts have a weak solute-solvent interactions. The acoustical analysis also reveals the same fact.

The fourth part of the thesis is concerned with the electrochemical nature of the tartrate salts in Formamide. The relationship between the equivalent conductance and internal pressure is analysed. The results obtained again reveal the significance of internal pressure in the study of liquid systems.

The statistical analysis in part five epitomizes the following results. ANOVA technique is applied to all systems under study. The acoustic parameter ultrasonic velocity, the thermodynamic parameters viz. the internal pressure and free volume, thermo chemical parameters such as apparent molal volume and apparent molal compressibility, solvation number and the conductivity parameter are extremely
affected by variations with respect to temperature. The variation of all the above parameters with respect to concentrations is also analysed. The salient features of the present work may be summarized as follows:

i. Use of microprocessor based ultrasonic interferometric system.

ii. Application of ANOVA technique i.e., the analysis of variance to study the effects of different levels of molality and temperature on a physical parameter.

iii. Focus on the study of the systems from low temperature to high temperature.

iv. Use of computer software for data compilation, computation and graphical representation of the results.

v. Spectroscopic study of the samples using 'HNMR and IR spectra.

vi. Comparison of acoustic and thermo dynamical results with spectroscopic analysis of the samples.
REFERENCES


70. Eller J.T., Hist. Acad. Roy. Berlin, 6, (1750) 82
72. Holkerd S., Phil Magazine, 27 (1844) 207
74. Tammann G., Z. Phys. Chem., 11, (1893) 676