IRAQI Academic Scientific Journals

University of Kirkuk Journal For Administrative and Economic Science (2025) 15 (4) Part (2): 385-396

UKJAES

University of Kirkuk Journal For Administrative and Economic Science

ISSN:2222-2995 E-ISSN:3079-3521
University of Kirkuk Journal For Administrative and Economic Science

**Abdullah Omiad Saber & Ahmed Sihad Faisal. Effects of CAF vs TAC Chemotherapy and Age at Puberty on IgM Levels in Breast Cancer Patients: A Small-Sample RCBD Study Using ANCOVA and ANOVA Comparison**. *University of Kirkuk Journal For Administrative and Economic Science* **(2025) 15 (4) Part (2):385-396.**

# Effects of CAF vs TAC Chemotherapy and Age at Puberty on IgM Levels in Breast Cancer Patients: A Small-Sample RCBD Study Using ANCOVA and ANOVA Comparison

Omiad Saber Abdullah [1], Sihad Faisal Ahmed [2]

[1,2] *College of Administration & Economic-Salahaddin University-Erbil, Erbil, Iraq*

omiad.abdullah@su.edu.ked [1]
sihadsoran01@gmail.com [2]

**Abstract:** In the Kurdistan Region of Iraq, breast cancer is the most common type of cancer, especially among women in the Erbil province. Unfortunately, society views this type of disease with great concern. To investigate the factors that affect the outcomes of breast cancer, in our study, we had to use a robust statistical method, using multifactor analysis of covariance (ANCOVA). Two independent categorical variables, each of which had two levels, which were chemotherapy and age at puberty, were evaluated for their effects on the measures of continuous outcomes, while controlling for the relevant covariate. Our analysis was integrated with a Randomized Complete Block Design (RCBD) to account for the patient's age as a blocking factor, which allows for precise adjustment of baseline variability. In this study, patient data was collected from Nanakali and Rizgari Teaching Hospitals in Erbil city. The level of Immunoglobulin M (IgM), which was obtained from the blood, serves as the primary outcome measure. Comparing ANCOVA with the traditional ANOVA revealed that ANOVA only identified factor B as significant, whereas ANCOVA, by controlling for the covariate, found the significant effects of factor A and the block that ANOVA had missed. Consequently, the study's findings showed that ANCOVA yields more accurate and reliable results by accounting for covariates, which enhances the precision and interpretation of the analyses of treatment outcomes. The study highlights the importance of targeted and evidence-based strategies in the management of breast cancer, also emphasizing the impact of covariates on clinical outcomes.

**Keywords:** ANCOVA, ANOVA, RCBD, covariate, IgM, breast cancer.

# تأثيرات العلاج الكيميائي (CAF) و (TAC) وعمر البلوغ على مستويات (IgM) في مريضات سرطان الثدي: دراسة عينة صغيرة بتصميم القطع العشوائية الكاملة باستخدام مقارنة (ANCOVA) و (ANOVA)

أ.م.د. أميد صابر عبدالله [١]، الباحثة: سيهاد فيصل أحمد [٢]

[١،٢] *كلية الادارة والاقتصاد/ قسم الإحصاء والمعلوماتية-جامعة صلاح الدين – أربيل، أربيل، العراق*

**المستخلص:** يُعد سرطان الثدي المرض الأكثر شيوعًا في إقليم كردستان العراق، وبالأخص بين النساء في محافظة أربيل، حيث يحظى باهتمام مجتمعي واسع لما يسببه من قلق بالغ. وانطلاقًا من الحاجة إلى فهم أدق للعوامل المؤثرة في مخرجات المرض، اعتمدت هذه الدراسة على منهج إحصائي متقدم هو تحليل التباين المشترك متعدد العوامل (ANCOVA).وقد جرى تقييم عاملين مستقلين فئويين، لكل منهما مستويان، هما نوع العلاج الكيميائي والعمر عند البلوغ، وذلك لقياس تأثيرهما على مؤشرات النتائج المستمرة، مع الأخذ بعين الاعتبار ضبط المتغيرات المصاحبة ذات الصلة.
ولغرض تحقيق ضبط منهجي أكبر للتباين الأولي بين المرضى، استُخدم تصميم الكتل الكاملة العشوائية (RCBD)، حيث اعتُبر العمر كقطاعات. وقد جمعت بيانات الدراسة من مراجعي مستشفيي نانەکەلی و ڕزگاری التعليميين في محافظة أربيل- العراق، واعتمد مستوى الغلوبولين المناعي M (IgM) المستخلص من عينات الدم كمؤشر أساسي لقياس النتائج.
أظهرت المقارنة بين تحليل التباين التقليدي (ANOVA) وتحليل التباين المشترك (ANCOVA) اختلافًا في تفسير النتائج؛ إذ بيّن ANOVA دلالة إحصائية للعامل (B) فقط، بينما كشف ANCOVA بعد ضبط المتغير المشترك عن تأثيرات مهمة لكل من العامل (A) والقطاعات، والتي لم يتمكن ANOVA من رصدها.
تؤكد هذه النتائج أن تحليل التباين المشترك (ANCOVA) يوفر تقديرات أكثر دقة وموثوقية من خلال إدماج المتغيرات المصاحبة في النموذج، مما يثري عملية تفسير نتائج العلاجات ويعزز موثوقيتها. وتبرز الدراسة أهمية اعتماد استراتيجيات علاجية موجهة وقائمة على الأدلة في إدارة سرطان الثدي، مع تسليط الضوء على دور العوامل المتداخلة في التأثير على الاستجابات العلاجية.

**الكلمات المفتاحية:** تحليل التغاير، تحليل التباين، **RCBD**، المتغير المصاحب، **IgM** ، سرطان الثدي.

**Corresponding Author: E-mail:** omaid.abdullah@su.edu.krd

## Introduction

Two broad approaches to controlling variability resulting from experimental error are direct and statistical. Examples of direct control are blocking or stratifying the experimental units, improving the uniformity of the experimental conditions, and increasing the precision of the measurements. These include replicate experiments and randomized block, repeated measure, split plot, and incomplete block designs. These designs use direct control.

This article will cover approaches that employ indirect or statistical control to enhance experimental precision and eliminate potential sources of bias. The latter purpose is especially significant in scenarios where the investigator is unable to randomly allocate individual units to the experimental conditions, monitoring one or more concurrent variables in addition to the variable of primary interest, statistical control is accomplished. Covariates are the concomitant variables, while the latter will be referred to as the criterion or just the variate. Measurements on the covariates are made to adjust the measurements on the variate in our study on the effects of chemotherapy and pubertal age on (IgM) levels after chemotherapy in breast cancer patients. We aim to evaluate how different treatments and age at puberty affect the immune The dependent variable is IgM levels after chemotherapy. The covariate is baseline IgM levels (before chemotherapy). Why Include a Covariate (e.g., Pre-treatment IgM)in breast cancer, patients naturally differ in their baseline IgM levels due to several extraneous (non-treatment-related)factors, such as Age, Nutritional status, Prior infections or immune disorders, and Genetic predispositions. If we analyze only post-treatment IgM values without accounting for these pre-treatment differences, we risk confounding, where the observed effects may be falsely attributed to the treatments when, in fact, they stem from initial patient variability. ANCOVA solves this problem by including baseline IgM as a covariate, allowing us to statistically adjust post-treatment outcomes. This isolates the true effect of Factor A = chemotherapy type, Factor =B puberty. It shows Thus, the adjusted IgM means reflect differences due to the treatments after controlling for baseline IgM, improving the precision and validity of the comparisons. In this paper, I aim to combine ANCOVA with RCBD. The goal of combining ANCOVA with RCBD.

RCBD (direct control): reduces variability by grouping experimental units into homogeneous blocks, ANCOVA (statistical control): further reduces variability by statistically adjusting for covariates. (Winer et al., 1971), (Rutherford, 2011), (Rutherford, 2000).

**1st: Research questions or Research objective**

1. Does the type of chemotherapy have an influence on the level of IgM?
2. Does the age of puberty have an influence on the level of IgM?
3. Is there any interaction between the type of chemotherapy and the age of puberty on the level of IgM or not?
4. Does the age group or the impact of blocking considerably address the variegations in the IgM responses?
5. Does the initial level of IgM before chemotherapy adjust or clarify the changes?
6. Does the ANCOVA have an influence on changing the factors, and are its results more precise and concise compared to the traditional methods like ANOVA?

**2nd: Literature Review**

In ANCOVA is a form of ANOVA where the effect of a continuous covariate is controlled for on the outcome. It integrates regression and analysis of variance to enhance precision and control for initial differences. ANCOVA is used in many fields such as psychology, agriculture, economics, and medicine. ANCOVA is common in experimental designs as it controls more effectively for treatment effects. This power comes from its capacity to reduce error variance in both randomized and observational studies. The use of ANCOVA was pioneered by Ronald A. Fisher in the 1930s.

(Wang et al., 2019). An experiment was carried out called Analysis of Covariance in Randomized Trials: More Precision and Valid Confidence Intervals, Without Model Assumptions. The main problem the study addressed is the ambiguity and inconsistency surrounding the use of covariate adjustments, particularly Analysis of Covariance (ANCOVA), in randomized clinical trials. To this end, data from three distinct RCTs conducted in other medical specialties were examined: the MCI study (neurology), the METS study (psychiatry), and the TADS study (adolescent psychiatry). They determined that ANCOVA is statistically valid, that is, it produces valid estimates and standard errors even when the linear model is incorrectly specified. It further makes estimates more precise and reduces their variance by 4% to 32% depending on the trial. Researchers can use standard ANCOVA packages, as there is no need for further sophisticated corrections.

(Mistry et al., 2022) The study was conducted under the title National Institutes of Health Stroke Scale as an Outcome in Stroke Research: Value of ANCOVA Over Analyzing Change from Baseline. The issue was that the majority of trials examine change from the baseline (ΔNIHSS), an approach that is based on a flawed assumption of a linear, constant relationship between scores at baseline and at follow-up that often introduces bias and inflated variance, as well as being less statistically powerful. The authors examined ΔNIHSS in simulated trials of 2,000 patients with 10,000 simulations and compared it to ANCOVA, which analyzes follow-up NIHSS while controlling for baseline. The outcome indicated that ANCOVA was less variable, more powerful statistically, and more clinically meaningful. The clear conclusion was that ANCOVA is preferable to ΔNIHSS and that future stroke research should use this approach for more precise and reliable results.

(Downs et al., 2023) Study Title: (Differences in total cognition and cerebrovascular function in female breast cancer survivors and cancer-free women). They researched how breast cancer and its treatment impact brain function. Age and BMI-matched controls without a history of cancer were recruited in similar numbers to 15 breast cancer survivors who were 5 years or less post-treatment with stages I–III breast cancer. Participants underwent cognitive, CVR, cardiovascular, and exercise performance, and mood and fatigue testing. Covariates in these analyses included 6-min walk distance, handgrip strength, blood pressures, arterial compliance, heart rate, resting CO2, and ventilation, as well as years of education. They conducted ANCOVA analyses to account for these variables in order to ensure that the observed group differences in cognition and CVR remained significant regardless of fitness, cardiovascular, or educational level and were associated with the experience of surviving breast cancer.

(Jung et al., 2025)The study was conducted under the title, how to construct analysis of covariance in clinical trials: ANCOVA with one covariate in a completely randomized design structure. which tells the researcher when and how to appropriately utilize the Analysis of Covariance (ANCOVA) model in clinical trials. Given that ANCOVA addresses baseline imbalances and attempts to minimize error variance in order to obtain more power and more precise estimates of treatment effects, it is also commonly applied. But it is only valid under certain assumptions: independence, linearity, normality, homogeneity of variance, and homogeneity of regression slopes across groups. They stress that covariates should be identified a priori in the study protocol and should be assessed prior to the treatment e.g., at baseline). We need to fit a model with the interaction, as the adjusted means lose their interpretation.

## 3rd: Materials and Methods

### 1- The Randomized Complete Block Design (RCBD)

Randomized complete block design (RCBD) is a design in which Experimental units are grouped or divided into blocks so that the experimental units within any group or block are relatively homogeneous. The number of experimental units within each block must be equal to the number of treatments required to be studied in the experiment, i.e., each block must contain all treatments, and this is the meaning of complete blocking in the name of the design. The treatments are distributed to the experimental units within each block randomly and independently of the other blocks.
From this, it becomes clear that the ages of the patients are different. There is no doubt that a 26 or 35-year-old patient has a stronger immune response than a 55 or 65-year-old patient. Therefore, we created 3 blocks to control for age so that we could determine the effect of the factors.(Gomez and Gomez, 1984), (Montgomery, 2017).

### 2- Analysis of Covariance (ANCOVA)

Analysis of Covariance (ANCOVA) is a statistical technique used to examine the effect of one or more categorical independent variables (IVs) on a continuous outcome while controlling for covariates that are related to the outcome. The objective of ANCOVA is to identify differences in the mean level of the dependent variable (DV) across the different levels of the IV(s) after controlling for the effect of the covariate(s). The analysis tests whether the mean level of the outcome differs between the levels of the IV after accounting for the effect of the covariate.(Rutherford, 2000),(Huitema, 2011).

### A. Covariate

A covariate is an independent variable that may influence the dependent variable, though it is not the central focus of the study. Covariates can be of continuous type or categorical type, representing levels of the covariate. Including covariates in the study will enhance our ability to isolate the treatment effect and improve the accuracy of our analyses. If covariates are important and we do not include them in the analysis, then it may lead to results in a confounded fashion. Therefore, it is crucial to describe the rationale and methods to select the covariates and control for them in conducting an analysis of covariance. If changes in the covariate are associated with the change in the treatment levels, then omission of covariates will result in biased estimates of the treatment effect. (Winer et al., 1971), (Rutherford, 2000).

### B. Immunoglobulin M (IgM)

Immunoglobulin M (IgM) is an important immune system protein that functions to detect and combat invading pathogens such as bacteria and viruses. IgM is the first antibody produced in a primary immune response. It is a fast-acting protective antibody. It is mainly produced by B cells in the spleen, lymph nodes, and bone marrow. The structure and Function of IgM is the largest antibody and is usually a pentamer, five “Y” shaped antibody units that are connected. This large structure has ten antibody-binding sites, thus providing a high avidity to bind to foreign invaders. The main role of antibodies is binding to these antigens and triggering the complement system, a

cascade of proteins that helps destroy pathogens. The Use of IgM in Medical Diagnosis. The immunoglobulin M, or IgM, test is a blood test that detects levels of IgM in the body. Thus, elevated IgM levels are usually a sign of a recent or active infection. This is because IgM is the first antibody type to be produced in response to a new pathogen. The Role of IgM In Cancer. The role of IgM in cancer is more complicated. Natural IgM antibodies are able to specifically recognize and eliminate some types of malignant cells, indicating that it can be both a biomarker and a target for immunotherapy. Unlike IgG, whose levels are usually downregulated with tumor advancement, IgM levels are not altered, supporting its application as a diagnostic tool in early cancer and as a subject of immunotherapeutic efforts. (Díaz-Zaragoza et al., 2015).

## 3- ANCOVA with RCBD

Analysis of Covariance (ANCOVA) is a method that merges characteristics from ANOVA and regression, in which means of groups are compared and the effects of one or more continuous covariates are adjusted for. This change increases the accuracy of treatment effect estimates by controlling for variation due to pre-treatment characteristics of experimental units. Although ANCOVA also assumes independence of observations, it makes additional assumptions of linearity of the covariate with the dependent variable and normality of residuals, and homogeneity of regression slopes. ANCOVA is even further enhanced by its application within a Randomized Complete Block Design (RCBD). RCBD is an experimental design that accounts for heterogeneity by grouping similar experimental units in blocks that receive all treatment combinations. This is done because the design reduces error variance between the groups and therefore increases the statistical power of the test. The use of RCBD in conjunction with ANCOVA combines the advantage of adjustment for covariates with the economy of blocking. (Montgomery, 2017), (Huitema, 2011).

## 4- Statistical Analysis

**Table (1):** Multi-factor ANCOVA in RCBD

| S.O.V. | d.f. | SS and SCP | | | d.f.´ | SS´ | MS´ | $F_{cal.}$ |
|---|---|---|---|---|---|---|---|---|
| | | XX | XY | YY | | | | |
| Blocks | r-1 | $r_{xx}$ | $r_{xy}$ | $r_{yy}$ | r-1 | $r'_{yy}$ | $MS'_{Block}$ | $F_{Block}$ |
| Treat. Com. | ab-1 | $t_{xx}$ | $t_{xy}$ | $t_{yy}$ | ab-1 | $t'_{yy}$ | $MS'_{t}$ | |
| $F_{actor}$ A | a-1 | $A_{xx}$ | $A_{xy}$ | $A_{yy}$ | a-1 | $A'_{yy}$ | $MS'_{A}$ | $F_{A}$ |
| $F_{actor}$ B | b-1 | $B_{xx}$ | $B_{xy}$ | $B_{yy}$ | b-1 | $B'_{yy}$ | $MS'_{B}$ | $F_{B}$ |
| $I_{ntraction}$ AB | (a-1) (b-1) | $AB_{xx}$ | $AB_{xy}$ | $AB_{yy}$ | (a-1) (b-1) | $AB'_{yy}$ | $MS'_{AB}$ | $F_{AB}$ |
| $E_{rror}$ | (r-1) (ab-1) | $e_{xx}$ | $e_{xy}$ | $e_{yy}$ | $df_E$-1 | $e'_{yy}$ | $MS'_{e}$ | |
| $T_{otal}$ | abr-1 | $T_{xx}$ | $T_{xy}$ | $T_{yy}$ | $df_T$-1 | $T'_{yy}$ | | |
| $A+E_{rror}$ | $df_A+df_E$ | $A_{xx}+e_{xx}$ | $A_{xy}+e_{xy}$ | $A_{yy}+e_{yy}$ | $df_{A+E}$ | $(A_{yy}+e_{yy})'$ | | |
| $B+E_{rror}$ | $df_B+df_E$ | $B_{xx}+e_{xx}$ | $B_{xy}+e_{xy}$ | $B_{yy}+e_{yy}$ | $Df_{B+E}$ | $(B_{yy}+e_{yy})'$ | | |
| $AB+E_{rror}$ | $df_{AB}+df_E$ | $AB_{xx}+e_{xx}$ | $AB_{xy}+e_{xy}$ | $AB_{yy}+e_{yy}$ | $df_{AB+E}$ | $(AB_{yy}+e_{yy})'$ | | |

**Explanation of all the rules and the way they are calculated, Sum of squares for (yy)**

$$T_{yy} = \sum y_{ijk}^2 - \frac{(Y...)^2}{abr} \quad ...(1)$$

$$A_{yy} = \frac{\sum y_{i..}^2}{br} - \frac{(Y...)^2}{abr} \quad ...(2)$$

$$B_{yy} = \frac{\sum y_{.j.}^2}{ar} - \frac{(Y...)^2}{abr} \quad ...(3)$$

$$AB_{yy} = \frac{\sum y_{ij.}^2}{r} - \frac{\sum y_{i..}^2}{br} - \frac{\sum y_{.j.}^2}{ar} + \frac{(Y...)^2}{abr} \quad ...(4)$$

$$r_{yy} = \frac{\sum y_{..k}^2}{ab} - \frac{(Y...)^2}{abr} \quad ...(5)$$

$$e_{yy} = T_{yy} - A_{yy} - B_{yy} - AB_{yy} - r_{yy} \quad ...(6)$$

Sum of squares for (XX), $T_{xx} = \sum x_{ijk}^2 - \frac{(X...)^2}{abr}$ ...(7)

$$A_{xx} = \frac{\sum x_{i..}^2}{br} - \frac{(X...)^2}{abr} \quad ...(8)$$

$$B_{xx} = \frac{\sum x_{.j.}^2}{ar} - \frac{(X...)^2}{abr} \quad ...(9)$$

$$AB_{xx} = \frac{\sum x_{ij.}^2}{r} - \frac{\sum x_{i..}^2}{br} - \frac{\sum x_{.j.}^2}{ar} + \frac{(X...)^2}{abr} \quad ...(10)$$

$$r_{xx} = \frac{\sum x_{..k}^2}{ab} - \frac{(X...)^2}{abr} \quad ...(11)$$

$$e_{xx} = T_{xx} - A_{xx} - B_{xx} - AB_{xx} - r_{xx} \quad ...(12)$$

Sum of squares for (XY), $T_{xy} = \sum x_{ijk} y_{ijk} - \frac{(X...)(Y...)}{abr}$ ...(13)

$$A_{xy} = \frac{\sum x_{i..} y_{i..}}{br} - \frac{(X...)(Y...)}{abr} \quad ...(14)$$

$$B_{xy} = \frac{\sum x_{.j.} y_{.j.}}{ar} - \frac{(X...)(Y...)}{abr} \quad ...(15)$$

$$AB_{xy} = \frac{\sum x_{ij.}}{r} - \frac{\sum x_{i..} y_{i..}}{br} - \frac{\sum x_{.j.} y_{.j.}}{ar} + \frac{(X...)(Y...)}{abr} \quad ...(16)$$

$$r_{xy} = \frac{\sum x_{..k} y_{..k}}{ab} - \frac{(X...)(Y...)}{abr} \quad ...(17)$$

$$e_{xy} = T_{xy} - A_{xy} - B_{xy} - AB_{xy} - r_{xy} \quad ...(18)$$

Sum of squares for adjusted, $A'_{yy} = (A_{yy} + e_{yy})' - e'_{yy}$ ...(19)

$$B'_{yy} = (A_{yy} + e_{yy})' - e'_{yy} \quad ...(20)$$

$$AB'_{yy} = (AB_{yy} + e_{yy})' - e'_{yy} \quad ...(21)$$

$$e'_{yy} = e_{yy} - \frac{(e_{xy})^2}{e_{xx}} \quad ...(22)$$

$$(A_{yy} + e_{yy})' = (A_{yy} + e_{yy}) - \frac{(A_{xy} + e_{xy})^2}{(A_{xx} + e_{xx})} \quad ...(23)$$

$$(B_{yy} + e_{yy})' = (B_{yy} + e_{yy}) - \frac{(B_{xy} + e_{xy})^2}{(B_{xx} + e_{xx})} \quad ...(24)$$

$$(AB_{yy} + e_{yy})' = (AB_{yy} + e_{yy}) - \frac{(AB_{xy} + e_{xy})^2}{(AB_{xx} + e_{xx})} \quad ...(25)$$

Mean square

$$MSA' = \frac{A'_{yy}}{a-1}, \quad MSB' = \frac{B'_{yy}}{b-1}, \quad MS(AB)' = \frac{AB'_{yy}}{(a-1)(b-1)}, \quad MSe' = \frac{e'_{yy}}{(r-1)(ab-1)-1}$$

Calculating F

$$F_A = \frac{MSA'}{MSe'}, \quad F_B = \frac{MSB'}{MSe'}, \quad F_{AB} = \frac{MS(AB)'}{MSe'}.$$ (Winer et al., 1971).

**5- The general linear model**

$$Y_{ijk} = \mu + \alpha_i + \beta_j + \rho_k + (\alpha\beta)_{ij} + \gamma(x_{ijk} - \bar{x}...) + \varepsilon_{ijk} \begin{cases} i = 1,2,...a \\ j = 1,2,..b \\ k = 1,2,...r \end{cases} ...(26)$$

μ : $\overline{Y}_{...}$

β : common regression for all treatments. (Winer et al., 1971)

**4th: Results**

From 2020 to 2024, recent epidemiological reports show a rising trend of cancer in the Kurdistan Region. This is more noticeable in Erbil, where the highest number of new cases are reported each year among the three governorates.

Disaggregated by gender, the impact of this perpetual increase on both men and women can be further appreciated. Regardless, the prevalence of some cancers is still higher among female patients at a disproportionately high rate (Karwan et al., 2022). Most notably, breast cancer is the most common cancer among women, ranking first in incidence, and remains a public health problem. This highlights the importance of the region having effective screening programs and early detection initiatives, as well as cancer control policies in place.

**Table (2):** Yearly Distribution of Cancer Cases by Gender in Erbil, Sulaimani, and Duhok, Kurdistan Region of Iraq (2020–2024)

| Gov | Year | Female | | Male | | Total | |
|---|---|---|---|---|---|---|---|
| | | Frequency | Row N % | Frequency | Row N % | Frequency | |
| Erbil | 2020 | 1593 | 52.2% | 1460 | 47.8% | 3053 | 22593 |
| | 2021 | 3605 | 54.4% | 3019 | 45.6% | 6624 | |
| | 2022 | 897 | 54.6% | 747 | 45.4% | 1644 | |
| | 2023 | 3481 | 55.7% | 2763 | 44.3% | 6244 | |
| | 2024 | 2877 | 57.2% | 2151 | 42.8% | 5028 | |
| Sulaimani | 2020 | 1293 | 50.6% | 1261 | 49.4% | 2554 | 16430 |
| | 2021 | 1689 | 52.6% | 1522 | 47.4% | 3211 | |
| | 2022 | 1976 | 52.5% | 1787 | 47.5% | 3763 | |
| | 2023 | 1876 | 52.2% | 1716 | 47.8% | 3592 | |
| | 2024 | 1807 | 54.6% | 1503 | 45.4% | 3310 | |
| Duhok | 2020 | 379 | 49.4% | 388 | 50.6% | 767 | 5700 |
| | 2021 | 557 | 54.8% | 459 | 45.2% | 1016 | |
| | 2022 | 794 | 55.4% | 640 | 44.6% | 1434 | |
| | 2023 | 633 | 53.6% | 547 | 46.4% | 1180 | |
| | 2024 | 727 | 55.8% | 576 | 44.2% | 1303 | |
| Total | | 24184 | 54.1% | 20539 | 45.9% | 44723 | |

From 2020-2024, the available cancer registry data from the Kurdistan Region-Iraq (KRG) indicate a rising trend of cancer cases in the three main governorates (Erbil, Sulaimani, Duhok) with a consistently higher burden among females versus males. Erbil had the highest overall cases, also reported in 2021 (6624 ) cases, with a consistent female preponderance of 52-57%. Sulaimani had a slightly upward trend as well, with a few years with over 3,000 cases/year, and female patients comprised approximately 52-55% of cases. While the overall total in Duhok was again less than in the other two governorates, there is a gradual increase, and females also represented slightly more in Duhok (54-56%). In summary, a total of 44,723 cancer cases were diagnosed in the KRG over these five years, with 54.1% being female and 45.9% male, which reflects both the increasing burden of cancer as well as the stability of the gender gap in incidence.

Given that breast cancer represents the most prevalent malignancy among women in Erbil according to the previously mentioned source, our analyses were primarily directed toward this domain. Accordingly, we employed a factorial ANCOVA model to ensure a more rigorous and comprehensive assessment.

This study included 12 female breast cancer patients at different stages after surgery. The data of this thesis can be clarified as follows: it was collected on female breast cancer patients from both Nanakaly Hospital and Rizgary Teaching Hospital in Erbil. The level of Immunoglobulin M (IgM) was measured from blood samples. Factor (A), which represents the type of chemotherapy treatment, was divided into two groups. The first group included 6 patients who received the (CAF regimen) of chemotherapy, which consists of (Cyclophosphamide (C) 500 mg/m²),(Doxorubicin (A) 50 mg/m²), and (5-Fluorouracil (F) 500 mg/m²) ($a_1$). This regimen was administered in two sessions, each including 3 cycles, for a total of 6 cycles, and each cycle lasting all of 21 days. The second group included 6 patients who received the (TAC regimen) chemotherapy, which consists of (Docetaxel (T) 75 mg/m²), (Doxorubicin (A) 50 mg/m²), and (Cyclophosphamide (C) 500 mg/m²) ($a_2$). Factor (B) was assigned based on puberty status. Six patients were included who reached puberty at the age of twelve or younger ($b_1$), and six patients who reached puberty at the age of thirteen or older ($b_2$). The ages of the patients were considered as blocks and divided into three groups, with each block containing four patients. The first IgM test without treatment was considered as the covariate (X), and the response variable (Y) was the IgM test after chemotherapy.

The change in IgM levels and the significance and result of comparing the factors were clarified in the practical section. The analyses were performed using the software program: SAS-9.4.

**Table (3):** Represents real data of breast cancer

| **Factor B** | | | **b1** | | | **b2** | | | |
|---|---|---|---|---|---|---|---|---|---|
| Block (Age Groups) | | | block1 (26-37) | block2 (38-47) | block3 (48-69) | block1 (26-37) | block2 (38-47) | block3 (48-69) | $\bar{a}$ |
| Factor A | a1 | x | 118 | 180 | 80 | 255 | 118 | 255 | 167.7 |
| | | y | 125 | 125 | 80 | 245 | 125 | 180 | 146.7 |
| | a2 | x | 150 | 178 | 100 | 120 | 300 | 250 | 183 |
| | | y | 255 | 180 | 80 | 250 | 245 | 250 | 210 |
| $\overline{block}$ | | x | 134 | 179 | 90 | 187.5 | 209 | 252.5 | |
| | | y | 190 | 152.5 | 80 | 247.5 | 185 | 215 | |
| $\bar{b}$ | | x | | 134.3 | | | 216.3 | | |
| | | y | | 140.8 | | | 215.8 | | |

## 1- Testing Model Assumptions

Five main assumptions must be met. If one of these five assumptions is not met, the results may be inaccurate or biased

### A. Linearity of covariance(X)-Response(Y) Relationship

The first assumption of ANCOVA states that there should be a linear relationship between the covariate and the response variable. In SAS, we can check this assumption by using Scatter plots with regression lines, Correlation analysis

$H_0: \rho = 0$ $H_1: \rho \neq 0$

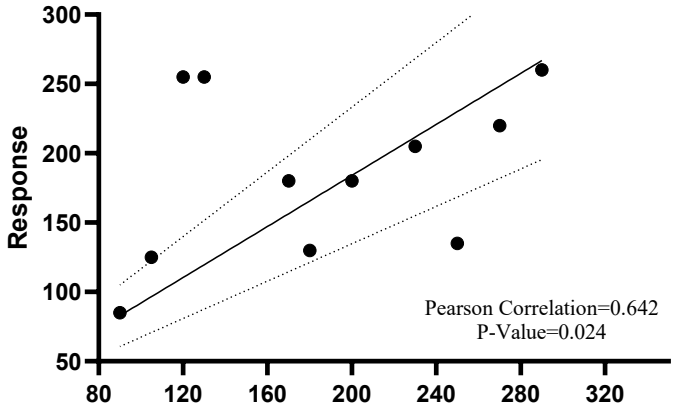


**Figure (1):** Scatter plots with regression line

*From the scatter plot above, we see that there is a linear relationship between X and Y, and the Pearson correlation value of r=0.642. This suggests a strong relationship exists between covariate X and response variable Y. So, we reject $H_0$ and accept $H_1$, so there is a linear relationship between X and Y . (McLouth, 2018), (Shieh, 2017), (Rutherford, 2000).*

### B. Normality of Residuals

The second assumption of ANCOVA is that the residuals (errors) follow a normal distribution. This is important because ANCOVA relies on normally distributed errors for valid statistical inference. To check this assumption, we can: Perform a normality test (Shapiro-Wilk test). $H_0$: Residuals are normally distributed, $H_1$: Residuals are not normally distributed

**Table (4):**| Tests for Normality of Residuals

| Test | Statistic | P-Value | Sig. |
|---|---|---|---|
| Shapiro-Wilk | 0.957 | 0.745 | Normal |
| Kolmogorov-Smirnov | 0.210 | 0.146 | Normal |
| Cramer-von Mises | 0.065 | >0.250 | Normal |
| Anderson-Darling | 0.353 | >0.250 | Normal |

In the table above (tests for Normality), it's visible that all tests have a P-Value greater than $\alpha$=0.05, but since we have $n \leq 50$ in this test, we use the Shapiro-Wilk, $p > 0.05$ then we fail to reject $H_0$ So the residues are normally distributed. (McLouth, 2018), (Huitema, 2011).

### C. Homogeneity of Regression Slopes

The homogeneity of regression slopes assumption states that the relationship between the covariate and the response should be the same across all levels of the categorical independent variable (e.g., Factor_A or Factor_B). To test this, we check if there is a significant interaction between the covariate and the factor(s). If the interaction is not significant, the assumption holds, meaning the effect of the covariate is consistent across groups. Factor A has two levels =$a_1$, $a_2$. Factor B has two levels =$b_1$, $b_2$. This gives us 2 × 2 = 4 combinations (cells), each with its slope $\beta_i$, e.g., $A_1B_1$, $A_1B_2$, $A_2B_1$, $A_2B_2$

$$H_0: \beta_1 = \beta_2 = \beta_3 = \beta_4 \quad H_1: at\ least\ one\ \beta_i \neq \beta_j\ for\ i \neq j$$

**Table (5):** Testing Homogeneity of Regression Slopes (interaction between covariate and Factors)

| Source | df | SS | Mean Square | F Value | P-Value |
|---|---|---|---|---|---|
| Factor A | 1 | 2323.26 | 2323.26 | 0.86 | 0.39 |
| Factor B | 1 | 3480.28 | 3480.28 | 1.29 | 0.30 |
| Covariate | 1 | 7273.18 | 7273.18 | 2.70 | 0.15 |
| Covariate*Factor A | 1 | 237.91 | 237.91 | 0.09 | 0.78 |
| Covariate*Factor B | 1 | 1647.37 | 1647.37 | 0.61 | 0.46 |

In the above table, to verify the assumption of equality of regression slopes in the ANCOVA, interaction terms between the covariate and each factor were examined. Results from Type III sum of squares indicated that neither covariate × factor A (p = 0.78) nor covariate × factor B (p = 0.46) interactions were statistically significant. So we fail to reject $H_0$, thus, this assumption is also met. (McLouth, 2018), (Shieh, 2017), (Leon et al., 1998).

### D. Homoscedasticity (Equal variance of Residuals)

The fourth assumption of ANCOVA is homoscedasticity, meaning that the residuals should have constant variance across levels of the independent variable(s).To check this, we can perform Levene's test (a statistical test for homogeneity of variances)

$$H_0: \sigma_1^2 = \sigma_2^2 = \sigma_3^2 = \sigma_4^2 \quad H_1: at\ least\ two\ variances\ not\ equal$$

**Table (6):** Levene's Test for Equality of Variance

| | | F | P-Value | Sig. |
|---|---|---|---|---|
| Factor A | X Equal variances assumed | 0.002 | 0.962 | Homogeneity |
| | Y Equal variances assumed | 0.218 | 0.650 | Homogeneity |
| Factor B | X Equal variances assumed | 4.397 | 0.062 | Homogeneity |
| | Y Equal variances assumed | 0.245 | 0.631 | Homogeneity |

According to Levene's Test for Equality of Variances, the P-Value for factor A is (0.962, 0.650) and factor B is (0.062, 0.631) greater than $\alpha$=0.05, then we accept $H_0$ all variances are equal. (McLouth, 2018), (Shieh, 2017), (Gomez and Gomez, 1984).

### E. Independence of Residuals

There should be no autocorrelation of the residuals. You can test this assumption by analyzing a scatter plot of the residuals over time or other variables (if you have time series or ordered data) or by looking at the Durbin-Watson statistic. But given that our data do not appear to be related in time or order, we will concern ourselves with confirming the independence of the residuals only by observing scatter plots for any obvious pattern, such as autocorrelation in time series data or clustering of residuals. One way to visualize this is to see whether residuals appear to be randomly distributed. $H_0$: Residuals are independent $H_1$: Residuals are autocorrelated.

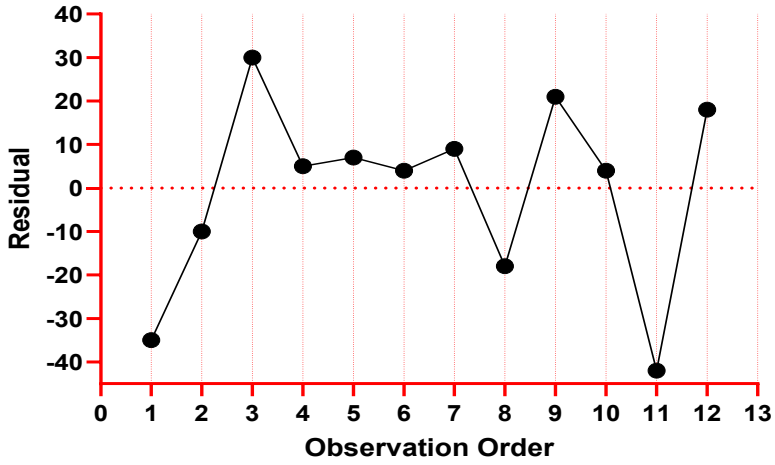


**Figure (2):** Scatter Plots for Independence of Residuals

In this graph, the residual versus observation order has been created. Clearly, there is no shape. The residues have gone up and down randomly around zero. This means that the residues are independent and have nothing to do with each other. Therefore, we reject $H_1$ and accept $H_0$. The residuals are independent, so our condition has been fulfilled .So we met all five assumptions, we can implement ANCOVA. (Leon et al., 1998), (Gomez and Gomez, 1984).

### 2- ANCOVA Fixed Effect Model in Real Data

**Table (7):** ANCOVA Summary Table Assessing the Effects of Factors A and B and Their Interaction on the Dependent Variable, Controlling for Covariate (X)

| Source | Sum of Squares | df | Mean Square | F-Ratio | P-Value |
|---|---|---|---|---|---|
| Corrected Model | 47780.706 | 6 | 7963.451 | 7.753 | $0.020^{(S)}$ |
| Block | 13334.700 | 2 | 6667.330 | 6.490 | $0.041^{(S)}$ |
| Factor A | 9226.090 | 1 | 9226.090 | 8.980 | $0.030^{(S)}$ |
| Factor B | 2454.990 | 1 | 2454.990 | 2.390 | $0.183^{(NS)}$ |
| Interaction AB | 15.926 | 1 | 15.930 | 0.016 | $0.906^{(NS)}$ |
| Covariate (X) | 8159.870 | 1 | 8159.870 | 7.940 | $0.037^{(S)}$ |
| Error | 5135.960 | 5 | 1027.190 | | |
| TOTAL | 52916.700 | 11 | | | |

In the ANCOVA summary table evaluating the effects of factors A and B and their interactions on the dependent variable, controlling for the covariate (X), each of the factors A, Block, and covariate with a P-Value smaller than $\alpha$=0.05 So we reject $H_0$ and accept $H_1$, which says that each of these affects the response variable means significantly, but factor B and interaction AB have no significant effect on the response variable (Y).

A two-factor ANOVA in an RCBD may produce unreliable results, resulting in erroneous conclusions.

**Table (8):** ANOVA Summary Table Assessing the Effects of Factors A and B and Their Interaction on the Dependent Variable

| Source | Sum of Squares | df | Mean Square | F-Ratio | P-Value |
|---|---|---|---|---|---|
| Block | 10704.20 | 2 | 5352.08 | 2.420 | $0.170^{(NS)}$ |
| Factor A | 12033.30 | 1 | 12033.30 | 5.430 | $0.059^{(NS)}$ |
| Factor B | 16875.00 | 1 | 16875.00 | 7.620 | $0.033^{(S)}$ |
| Interaction AB | 8.33 | 1 | 8.33 | 0.004 | $0.953^{(NS)}$ |
| Error | 13295.80 | 6 | 2215.97 | | |
| Total | 52916.70 | 11 | | | |

A comparison between the ANCOVA (Table 7), and the two-factor ANOVA RCBD (Table 8) summary tables reveals notable differences in the interpretation of factor effects. In the ANCOVA, each of the Factor A (P = 0.030), Block (P = 0.041), and the covariate X (P = 0.037) all have significant effect on the dependent variable. where as Factor B (P = 0.183) and the interaction AB (P = 0.906) are not significant. This indicates that adjusting for the covariate allows a more accurate estimation of Factor A and Block effects.

On the contrary, the ANOVA (Table 8) clearly indicates that only Factor B (P = 0.033) is significant, while Factor A (P = 0.059) and Block (P = 0.170) are not significant, and the interaction AB (P = 0.953)remains non-significant. Compared to ANCOVA, the ANOVA fails to detect the significance of (Factor A) and (Block), which may lead to erroneous conclusions regarding their effects on the dependent variable(DV). From this, it becomes clear to us that the results of ANCOVA show more precision and accuracy.

This comparison underscores that ANCOVA, by controlling for covariates, provides more reliable and precise results, especially in experimental designs where nuisance factors or covariates can influence the response variable.

## 5th: Discussion

A. The type of chemotherapy (Factor A) affects IgM levels.
B. Puberty (Factor B) does not affect IgM levels.
C. There is no interaction between the type of chemotherapy and pubertal age (interaction AB) on IgM levels.
D. Age group or blocking effect significantly resolves changes in IgM responses.
E. Initial IgM levels before chemotherapy(covariate-X) regulate or explain the changes.
F. Compared to ANOVA, ANCOVA controlling for the covariate detected significant effects of Factor A and Block, while ANOVA missed these, identifying only Factor B as significant. This demonstrates that ANCOVA provides more accurate and reliable results by accounting for covariates.

## 6th: Conclusion

The study revealed that the type of chemotherapy had a significant effect on IgM levels in breast cancer patients, whereas age at puberty had no measurable impact. This indicates that the immune response is more sensitive to chemotherapy rather than to age at puberty. The comparison highlighted the superiority of ANCOVA by including the covariate (pre-treatment IgM) as prior information, and patients' baseline immune status. ANCOVA reduced the error variance and improved the precision of the estimated treatment effects. However, in the ANOVA, the effect of chemotherapy was not fully apparent. Age at puberty emerged as significant, which is the opposite of what was identified in the ANCOVA. Moreover, using the RCBD design with ANCOVA strengthened the analysis by controlling for age-related variability across the blocks. Our findings emphasized the importance of adjusting for covariates in clinical studies. ANCOVA demonstrated that it can identify treatment effects that might otherwise have remained obscure.

## References

1- DÍAZ-ZARAGOZA, M., HERNÁNDEZ-ÁVILA, R., VIEDMA-RODRÍGUEZ, R., ARENAS-ARANDA, D. & OSTOA-SALOMA, P. 2015. Natural and adaptive IgM antibodies in the recognition of tumor-associated antigens of breast cancer. Oncology reports, 34, 1106-1114.
2- DOWNS, T. L., WHITESIDE, E. J., FOOT, G., MILLS, D. E. & BLISS, E. S. 2023. Differences in total cognition and cerebrovascular function in female breast cancer survivors and cancer-free women. Breast, 69, 358-365.
3- GOMEZ, K. A. & GOMEZ, A. A. 1984. Statistical procedures for agricultural research, John wiley & sons.
4- HUITEMA, B. 2011. The analysis of covariance and alternatives: Statistical methods for experiments, quasi-experiments, and single-case studies, John Wiley & Sons.
5- JUNG, W., LEE, K., KIM, H. H. & LIM, C. 2025. How to construct analysis of covariance in clinical trials: ANCOVA with one covariate in a completely randomized design structure. Korean J Anesthesiol, 78, 315-320.
6- KARWAN, M., ABDULLAH, O. S., AMIN, A. M., MOHAMED, Z. A., BESTOON, B., SHEKHA, M., NAJMULDEEN, H. H., RAHMAN, F. M., HOUSEIN, Z. & SALIH, A. M. 2022. Cancer incidence in the Kurdistan region of Iraq: Results of a seven-year cancer registration in Erbil and Duhok Governorates. Asian Pacific journal of cancer prevention: APJCP, 23, 601.
7- LEON, A. C., PORTERA, L., LOWELL, K. & RHEHEIMER, D. 1998. A strategy to evaluate a covariate by group interaction in an analysis of covariance. Psychopharmacology Bulletin, 34, 805.
8- MCLOUTH, C. J. 2018. Analysis of Covariance with Heterogeneity of Regression and a Random Covariate.
9- MISTRY, E. A., YEATTS, S. D., KHATRI, P., MISTRY, A. M., DETRY, M. A., VIELE, K., HARRELL, F. E. & LEWIS, R. J. 2022. National Institutes of Health Stroke Scale as an Outcome in Stroke Research: Value of ANCOVA Over Analyzing Change From Baseline. Stroke, 53, e150 - e155.
10- MONTGOMERY, D. C. 2017. Design and analysis of experiments, John wiley & sons.
11- RUTHERFORD, A. 2000. Introducing ANOVA and ANCOVA: a GLM approach.
12- RUTHERFORD, A. 2011. ANOVA and ANCOVA: a GLM approach, John Wiley & Sons.
13- SHIEH, G. 2017. On tests of treatment-covariate interactions: An illustration of appropriate power and sample size calculations. PloS one, 12, e0177682.
14- WANG, B., OGBURN, E. L. & ROSENBLUM, M. 2019. Analysis of covariance in randomized trials: More precision and valid confidence intervals, without model assumptions. Biometrics, 75, 1391-1400.
15- WINER, B. J., BROWN, D. R. & MICHELS, K. M. 1971. Statistical principles in experimental design, Mcgraw-hill New York.