Analysis of covariance for completely randomized design  
(case study: the life times of cutting tools used with lathes)

Faula Arina*, Maria Ulfah 1, Dyah Lintang Trenggonowati 1
1Department of Industrial Engineering, Universitas Sultan Ageng Tirtayasa, Cilegon, Indonesia

*Email: faula.arina71@gmail.com

Abstract. Analysis of covariance (ancova) is a blend of regression and analysis of variance (anova) used for upgrade precision of an experiment. Ancova can be used for all experimental design including completely randomized design. The completely randomized design is the response are randomly intrusted to treatments. Application ancova on completely randomized design in the manufacture is given to observe the effect of types of cutting tools of lifetimes of cutting tools. In this experiment covariate variable is lathe speed. Lathe speed affects tool life meaning that the involvement of the covariate variables in the covariance analysis model is appropriate. From the covariance analysis, it is also found that tool type influences tool life. The coefficient of variance in ancova is lower than anova. This proves that the ancova has a better level of precision than anova, so that ancova is more appropriate to use than anova.

1. Introduction
The analysis of covariance (ancova) is a blend of regression model and analysis of variance (anova). Ancova is a statistical analysis used in the experiments to determine variations resulting from one or more disturbance variables [1]. Engineers usually use blocking to handle disturbance variables. Blocking mainly reduce the noise that appear from disturbance variable by salve it as a special resource variability. This technique can only be used as the disturbance variables can be controlled.

The precision in an experiment can be improved through the use of assistive observation and a technique called covariance analysis. Covariance analysis is carried out based on the consideration that in reality there are certain variables which cannot be controlled, but which affect or correlate with the observed response variables.

Much of the theory and history of ancova is included in the Biometrics series [2,3]. See Huitema [4] for a detailed theories and practice of ancova for various designs. Several introductory linear models and experimental design provide this topic [5-8].

Covariance analysis can be applied in any experimental design including Completely Randomized Design (CRD). Based on this review, in this study used covariance analysis in the CRD. To clarify understanding, the CRD case is used in life times of cutting tools data [9].

We provide this paper as a tutorial to introduce ancova theory and application, complementing the types of experiments often used in quality improvement studies. The methodology is provided using an effect model building purpose that can be applied in CRD. The software uses the minitab language.
2. Research Methodology

2.1 Ancova for Completely Randomized Design

In many experiments, the response of an experiment are influenced not only by experimental factors that can be controlled, but also by uncontrolled variations. A model can be defined that links responses to controlled and uncontrolled variability, and the effect of every variant on the response of an experiment can be explored. To determine the controlled experimental factors of an uncontrolled experimental variable, the last is called covariate. Analysis of covariance is an analysis of experimental data as covariate variables and experimental factors are provide. Ancova compared the difference between the mean treatments, as in anova, while modeling between the covariate and the response of an experiment using a regression model.

The general ancova which has interesting a single experimental factor and a single covariate

\[ b_{ij} = \mu + \alpha_i + \beta a_{ij} + \epsilon_{ij} \]

where \( b_{ij} \) is the responses, \( a_{ij} \) is the covariate variables for the \( i \)th observation, \( \alpha_i \) is the experimental factors effect, \( \beta \) is the gradient of the regression between the covariate and the response of an experiment, and \( \epsilon_{ij} \) is the random error of the \( i \)th observation [4].

| Table 1. Ancova for completely randomized design |
|-----------------|-----------------|-----------------|-----------------|
| Source          | SS              | DF              | MS              | F               |
| Covariat        | \( (S_{ab})^2 / S_{aa} \) | 1               |                 |                 |
| Faktor A        | \( SS_{Treat} = S_{bb} - (S_{ab})^2 / S_{aa} - E_{bb} - (E_{ab})^2 / E_{aa} \) | \( s - 1 \)     | \( MS_{Treat} = SS_{Treat} / s - 1 \) | \( MSE \)          |
| Error           | \( SS_{E} = E_{bb} - (E_{ab})^2 / E_{aa} \) | \( n - s - 1 \) | \( MSE = SS_{E} / n - s - 1 \) |                 |
| Total           | \( S_{bb} \)   | \( n - 1 \)     |                 |                 |

The assumptions that must be resolved in using covariance analysis are as follows:
1. Covariate variables do not correlate with the treatment experiment
2. Linier relationship between response variables and covariate variables
3. Residual are normally distributed.

2.2 Coefficient of variation

The coefficient of difference is a coefficient that shows the accuracy of a conclusion or result obtained from an experiment [10].

\[ CV = \sqrt{\frac{MSE_{adj}}{\bar{b}}} \]

where

- \( CV \) = coefficient of variation
- \( MSE_{adj} \) = mean square error adjusted
- \( \bar{b} \) = Average of all experimental data

The coefficient of variation shows the level of accuracy of an experiment. In general it can be said if the value of the coefficient of variation is smaller means that the level of accuracy is higher and the validity of the conclusions obtained from these experiments the better.
2.3 Data
Montgomery [9] represented an experiment where it is believed that the lifetimes of cutting tools used with lathe ($b$) depends on type of tool (the treatment) and possibly the speed of lathe (the covariate $a$). Table 2, data presented on the cutting tool life times. This study is focused on comparison type of cutting tools, type A and B.

| Tool life times(hr) | Speed of lathe (rpm) | Type of tool | Tool life times(hr) | Speed of lathe (rpm) | Type of tool | Tool life times(hr) |
|---------------------|----------------------|--------------|---------------------|----------------------|--------------|---------------------|
| 18.73               | 610                  | A            | 30.16               | 670                  | B            | 18.73               |
| 14.52               | 950                  | A            | 27.09               | 770                  | B            | 14.52               |
| 17.43               | 720                  | A            | 25.40               | 880                  | B            | 17.43               |
| 14.54               | 840                  | A            | 26.05               | 1000                 | B            | 14.54               |
| 13.44               | 980                  | A            | 33.49               | 760                  | B            | 13.44               |
| 24.39               | 530                  | A            | 35.62               | 590                  | B            | 24.39               |
| 13.34               | 680                  | A            | 26.07               | 910                  | B            | 13.34               |
| 22.71               | 540                  | A            | 36.78               | 650                  | B            | 22.71               |

2.4 Analysis Method
The steps to be carried out in analysis of covariance in the completely randomized design are as follows:
1. Scatter plot of tool life and lathe speed
2. Conduct analysis of covariance.
3. Test the assumptions of covariance analysis
4. Coefficient of variation

3. Result and Discussion
Figure 1 shows the speed of lathe ($a$) and tool life times ($b$) linear relationship which has the same gradient for the two types of tools. The scenario would discuss modeling of covariate treatment without interactions in the model. Table 2 shows $b$ cannot be compared directly using single-factor analysis of variance or the two-sample t-test. This is because $b$ is a linear function from $a$. Although $a$ can be set at the selected value, in this experiment they did not exist. So, the speed of lathe ($a$) must be treated as a covariate. The speed of lathe ($a$) and type of tool are not influenced by each other, so $a$ can be considered a covariate and not another response.

![Scatterplot of tool life vs lathe speed](image-url)
Table 3 shows two anova tables for data of Table 2. The complete models including covariates in the first anova table. The reduced models does not include covariate in the second anova table. The anova table for the reduced models that SSE term and the degrees of freedom have been increased with the number of each for covariate in the first anova table. This shows the component error model can increase variability due to mess to enter important model terms.

**Table 3. Anova for cutting tool life times data**

| Source          | SS   | DF | MS   | F     | P_value |
|-----------------|------|----|------|-------|---------|
| Model A: Complete models, including covariates |      |    |      |       |         |
| lathe speed     | 1418.04 | 2  | 709.02 | 76.73 | 0.00    |
| type of tool    |      |    |      |       |         |
| Error           | 157.05 | 17 | 9.24  |       |         |
| Total           | 1575.09 | 19 |      |       |         |
| Model B: The Reduced models does not include covariate |      |    |      |       |         |
| type of tool    | 1097.90 | 1 | 1097.90 | 41.41 | 0.00    |
| Error           | 477.20 | 18 | 26.50 |       |         |
| Total           | 1575.10 | 19 |      |       |         |

Table 4 is the anova table on the tool life times data. Consider that the covariate is very significant (p < 0.00), Figure 1 shows a statistically significant linear trend; that is, the lathe speed and the type of tool affects the lifetimes of the tools.

**Table 4 Anova for cutting tool life times data**

| Source          | Seq SS | Df | Adj SS  | Adj MS  | F     | P   |
|-----------------|--------|----|---------|---------|-------|-----|
| lathe speed     | 293.01 | 1  | 320.17  | 320.17  | 34.66 | 0.00|
| Type of tool    |        |    |         |         |       |     |
| Error           | 1125.03 | 1  | 1125.03 | 1125.03 | 121.78 | 0.00|
| Total           | 157.05 | 17 | 157.05  | 9.24    |       |     |
| R squared       | 90.03% |    |         |         |       |     |

**Figure 2. Effects model for cutting tool life data**

**Figure 3** suggests that the relationship linear between the covariate variable and the response variable, The P value is 0.00, so we assesses the covariate effect as significant. Lathe speed effect, y = 44.488 -0.02661 lathe speed. The slope for this model is -0.02661, this too shows in Figure 1 the mean slope of the two regression lines. The proportion explained by the ancova model has increased to 90.03 %, mainly due to the contribution of lathe speed variable. Figure 3 suggest that the assumption using covariance analysis is residual are normally distributed.
Figure 3. Residual plot for tool life times data

\[ \overline{y}_A = 17.11 \quad \text{and} \quad \overline{y}_B = 31.93 \]

The best tool type is tool type B because it produces a longer tool life.

\[ CV_{\text{anova}} = \sqrt{\frac{MSE}{\overline{y}}} = \sqrt{\frac{26.5}{24.519}} \times 100\% = 21\% \]

\[ CV_{\text{ancova}} = \sqrt{\frac{MSE_{\text{adj}}}{\overline{y}}} = \sqrt{\frac{9.24}{24.519}} \times 100\% = 12.4\% \]

From these calculations, the coefficient of variation after being corrected (ancova) is smaller than the coefficient of variation before being corrected (anova) which means the level of accuracy of the covariance analysis is better than the analysis of variance, so that the analysis of covariance is more appropriate to use than the analysis of variance.

4. Conclusion

We present an introductory analysis covariance for qualified engineers to help unlock the whole potency of this procedure which is often fit, but seldom used. Discussion of one factor design and examples with linear, not interacting covariates describe relationships regression analysis and anova to ancova. This initial study too begins an analysis of the apply and advantage of ancova.

Lathe speed affects tool life meaning that the involvement of the covariate variables in the covariance analysis model is appropriate. From the calculation of covariance analysis, it is also found that tool type influences tool life so that further tests are performed. From further tests that have been carried out, tool type B is the best tool type because it is able to produce a longer tool life.

The coefficient of variance in covariance analysis is lower than analysis of variance. This proves that the analysis of covariance has a better level of precision than analysis of variance, so that analysis of covariance is more appropriate to use than analysis of variance.

References

[1] Silknitter KO Wisnowski JW and Montgomery DC 1999 The Analysis Of Covariance: A Useful Technique For Analysing Quality Improvement Experiment Quality And Reliability Engineering International 15(4) 303316
[2] Cochran, WG 1957 Analysis of covariance: Its nature and uses Biometrics 13 261281
[3] ‘Special Issue on the Analysis of Covariance’ 1982 Biometrics 38 541753
[4] Huitema BE 1980 The Analysis of Covariance and Alternatives: Statistical Methods for Experiments, Quasi-Experiments, and Single-Case Studies, 2nd Edition (New York: Wiley)
[5] Myers RH and Milton JS 1999 First Course in The Theory of Linear Statistical Models.
[6] Mason RL Richard FG and James LH 2003 Statistical Design and Analysis of Experiments :With Applications to Engineering and Science (Canada : John Wiley & Sons Inc)

[7] Kutner MH Natchtsheim CJ Neter J Li W 2005 Applied Linier Statistical Models Fifth Edition (New York : McGraw-Hill Irwin)

[8] Montgomery DC 2013 Design and Analysis of Experiments (Singapore: John Wiley & Sons Inc)

[9] Montgomery D C Peck EA and Vining GG 2001 Introduction to Linear Regression Analysis, Third Edition (New York: John Wiley & Sons Inc)

[10] Farda NS Tatik W and Rita R 2017 Analisis Kovarian Pada Rancangan Bujursangkar Graeco Latin Jurnal Gaussian 6 Nomor 1 3140