Previously, we discussed analysis of variance (ANOVA) and simple linear regression, which commonly share continuous dependent variables. While ANOVA uses categorical variables as independent variables, regression uses mainly continuous variables for them. However, we may want to include both kinds of variables in analysis. A statistical model with continuous dependent variables and both types of independent variables is called a general linear model (GLM). In this section, we discuss analysis of covariance (ANCOVA) as a type of GLM models. An ANCOVA is similar to an ANOVA model, but it includes a continuous variable as well as categorical variables as independent variables, being a mixture model of ANOVA and regression models.

RAISING A QUESTION ON IGNORING COVARIATES

An example data is composed of 3 variables, treatment effect, treatment methods (Tx; 2 groups), and age in Table 1. Our interest is on comparison of treatment effects by 2 Tx, experimental and control groups. We may consider independent t-test, ignoring age variable. Distribution of treatment effects of 2 groups is depicted in Figure 1A. We could obtain the p value, 0.048, by applying independent student’s t-test comparing treatment groups and conclude the treatment effect of treatment group is superior to that of control group.

Now let’s look into the relationship between treatment effect and age. We can notice a trend that higher age is related to higher treatment effect in Figure 1B. The positive correlation between effect and age is quantitatively measured by a Pearson correlation coefficient, 0.805 (p < 0.001). The positive correlation is further analyzed by regression analysis. We get regression equations for pooled sample of both groups as well as for each group as following:

For pooled sample: Treatment effect = 24.2 + 0.74 × Age + Error  (1)

For experimental group: Treatment effect = 47.5 + 0.33 × Age + Error  (2)

For control group: Treatment effect = 6.21 + 1.03 × Age + Error  (3)

Meanwhile, the mean age of subjects in the experimental group is 44.83 years, which is higher than that of the control group, 43.58 years. We may be suspicious that the difference of effect between two groups is partly attributed to the age difference between two groups, because age
is positively correlated with treatment effect. How can we resolve this issue? There is a clear need to consider the covariate, age, into the model to control its possible influence.

**ANCOVA MODEL: COMPARING MEANS CONSIDERING COVARIATES**

To compare 2 means, we can apply ANOVA as well, which is applicable in comparing 2 or more group means. The result shows significant difference between two groups \( p = 0.048 \), which is exactly the same with that from the independent \( t \)-test in Figure 2C. Still, the possible covariate, age, is ignored. The model including 2 groups explains the variation of effect as much as corrected model sum of squares of 737.042 among total sum of squares of 4,435.958 in Figure 2C. Figure 2B displays the proportion of errors as 0.83, which is proportion of Error sum of squares of 3,698.917 among total sum of squares of 4,435.958. The proportion of errors represents the portion of variation that the model cannot explain. Also, we find the proportion of explained variance, R-squared, is 0.166, which represents that only 16.6% of variance in the response variable is explained by this model.

The ANOVA model can be performed using GLM procedure. The result is expressed as a GLM equation in Figure 2A, as:

\[
\text{Treatment effect} = 51.25 + 11.08 \times \text{Tx} + \text{error} \quad (4)
\]

where \( \text{Tx} = 0 \) for control group and \( \text{Tx} = 1 \) for experimental group.

We can obtain the mean effects of experimental and control groups as 62.33 \( (= 51.25 + 11.08) \) and 51.25, exactly the same as which appears above.

---

**Table 1. Data of treatment groups, treatment effect (Effect), and age (Age)**

| Group          | Effect | Mean | SD |
|----------------|--------|------|----|
| Experimental   | 56     | 63   | 52 |
| Control group  | 28     | 67   | 56 |

---

**Figure 1.** Scatter plot of the example data in Table 1. (A) Distribution of treatment effects of 2 groups; (B) Scatter plot of treatment effects and age by treatment groups.
To solve the question whether different age levels influence the degree of group difference in treatment effect level, we include age into the model. We insert the covariate, age, into the previous ANOVA model, constructing an ANCOVA model. The result is shown in Figure 3.

The resulting ANCOVA equation is:

\[ \text{Tx effect} = 19.71 + 10.18 \times \text{Tx} + 0.72 \times \text{Age} + \text{Error} \quad (5) \]

where \( \text{Tx} = 0 \) for control group and \( \text{Tx} = 1 \) for experimental group.

The difference of effect between 2 groups has changed slightly from 11.08 in Equation 4 to 10.18 in Equation 5. The size of intercept has reduced greatly by around 31 because it has been adjusted by age. One unit increase of age is related to an increase of 0.72 unit in treatment effect. The proportion of errors has decreased greatly from 0.83 to 0.21 in Figure 3B. The reason is because age explained a big portion of variability in the response variable (gray colored segment).

As appeared in Figure 3C, the proportion explained by the ANCOVA model has improved up to 78.8%, mainly due to the contribution of age variable. The \( p \) values of Tx and age are 0.01 and < 0.001, respectively, which represent a highly significant result. The inclusion of covariate which is highly correlated with response can remove a considerable portion of errors, reducing the proportion of errors. In contrast, the explanation ability and significance of factors increase in the ANCOVA model.
It is noticeable that the slope of age is the same as 0.72 for both treatment groups in Figure 3A, which is restricted by the assumption of the ANCOVA model. The slope is similar to that of pooled sample, 0.74 as appeared in Equation 1. However, the slopes of 2 groups may actually be different because the slopes of 2 groups seem substantially different from one another, 0.33 in Equation 2 and 1.03 in Equation 3.

**ANCOVA MODEL WITH INTERACTION**

An ANCOVA model with interaction term is often called ‘a moderated regression,’ specifically [1]. Now we consider including an interaction term between group and age into the previous ANCOVA model, to assess if there is a significant difference in slopes of 2 groups. In Figure 3C, we find that the interaction term, \( Tx \times Age \), is statistically significant \( p < 0.001 \), which supports the need of interaction term. By applying the model, the proportion of errors has decreased dramatically to 7%, as a considerable portion of variance is explained by the interaction term (Figure 4B). Also, 93.2% of total variance is explained by the model \( (R^2 = 0.932, \text{ Figure 4C}) \).

The construction of ANCOVA model with interaction results in the model as follows:

\[
Tx \text{ effect} = 6.21 + 41.30 \times Tx + 1.03 \times Age - 0.70 \times (Tx \times Age) + Error \tag{6}
\]
where $Tx = 0$ for control group and $Tx = 1$ for experimental group.

Immediately, Equation 6 can create two separate models for both control and experimental groups. The resulting Equation 7 and Equation 8 is exactly the same with the results obtained by simple regression, Equation 2 and Equation 3, respectively.

For control group ($Tx = 0$): $\text{Tx effect} = 6.21 + 1.03 \times \text{Age} + \text{Error} \quad (7)$

For experimental group ($Tx = 1$): $\text{Tx effect} = (6.21 + 41.30 + (1.03 - 0.70) \times \text{Age} + \text{Error} = 47.51 + 0.33 \times \text{Age} \quad (8)$

REFERENCES

1. Leppink J. Analysis of covariance (ANCOVA) vs. moderated regression (MODREG): why the interaction matters. Health Prof Educ 2018;4:225-232.

CROSSREF

https://rde.ac

https://doi.org/10.5395/rde.2018.43.e43

5/7
Appendix 1. Procedure of analysis for analysis of covariance (ANCOVA) using IBM SPSS

The procedure of ANCOVA using IBM SPSS Statistics for Windows Version 23.0 (IBM Corp., Armonk, NY, USA) is as follows.

1. ANCOVA model (without interaction)

(A) Data

| Tx_Effect | Age | Tx |
|-----------|-----|----|
| 56        | 21  | 0  |
| 20        | 19  | 0  |
| 25        | 23  | 0  |
| 55        | 28  | 0  |
| 63        | 33  | 0  |
| 52        | 33  | 0  |
| 62        | 56  | 0  |
| 45        | 43  | 0  |
| 59        | 47  | 0  |
| 64        | 48  | 0  |
| 61        | 53  | 0  |
| 73        | 58  | 0  |
| 62        | 63  | 0  |
| 71        | 56  | 1  |
| 70        | 67  | 0  |
| 71        | 63  | 1  |

(B) Analyze-Regression-Linear

(B-1) Variables

(B-2) Model

(B-3) Options

(continued to the next page)
Appendix 1. (Continued) Procedure of analysis for analysis of covariance (ANCOVA) using IBM SPSS

(C-1) Descriptive statistic

| Dependent Variable: Tx_Effect | Mean  | Std. Deviation |
|------------------------------|-------|----------------|
| 0                            | 62.33 | 8.307          |
| 1                            | 51.26 | 17.189         |
| Total                        | 56.79 | 13.808         |

(C-2) Analysis of variance (ANOVA) table

| Source            | Type III Sum of Squares | df | Mean Square | F    | Sig     | Partial Eta Squared |
|-------------------|--------------------------|----|-------------|------|---------|---------------------|
| Corrected Model   | 3494.233*                | 2  | 1747.117    | 38.90 | .000    | .788                |
| Intercept         | 1499.404                 | 1  | 1499.404    | 32.23 | .000    | .613                |
| Tx                | 620.533                  | 1  | 620.533     | 13.93 | .001    | .367                |
| Age               | 2757.192                 | 1  | 2757.192    | 61.48 | .000    | .745                |
| Error             | 841.725                  | 21 | 44.844      |       |         |                     |
| Total             | 81843.000                | 24 |             |       |         |                     |
| Corrected Total   | 4435.658                 | 23 |             |       |         |                     |

a. R Squared = .788 (Adjusted R Squared = .767)

(C-3) Estimated marginal means

| Dependent Variable: Tx_Effect | Mean | Std. Error | 95% Confidence Interval |
|-------------------------------|------|------------|-------------------------|
| 0                             | 61.89 | 1.934      | 57.859 - 65.930         |
| 1                             | 51.70 | 1.934      | 47.830 - 55.724         |

a. Covariates appearing in the model are evaluated at the following values: Age = 44.21

2. ANCOVA model with interaction

(B-2) Model (interaction term included)

(C-2) ANOVA table

| Source            | Type III Sum of Squares | df | Mean Square | F    | Sig     | Partial Eta Squared |
|-------------------|--------------------------|----|-------------|------|---------|---------------------|
| Corrected Model   | 4135.293*                | 3  | 1378.432    | 91.96 | .000    | .923                |
| Intercept         | 1720.655                 | 1  | 1720.655    | 114.42 | .000 | .851 |
| Tx                | 1016.591                 | 1  | 1016.591    | 67.02 | .000 | .772 |
| Age               | 2431.784                 | 1  | 2431.784    | 160.56 | .000 | .889 |
| Tx*Age            | 641.064                  | 20 | 641.064     | 42.64 | .000 | .681 |
| Error             | 300.661                  | 20 | 15.033      |       |       |                     |
| Total             | 81843.000                | 24 |             |       |       |                     |
| Corrected Total   | 4435.958                 | 23 |             |       |       |                     |

a. R Squared = .932 (Adjusted R Squared = .923)

(C-3) Estimated marginal means

| Dependent Variable: Tx_Effect | Mean  | Std. Error | 95% Confidence Interval |
|-------------------------------|-------|------------|-------------------------|
| 0                             | 62.127 | 1.120      | 59.790 - 64.464         |
| 1                             | 51.966 | 1.120      | 49.559 - 54.332         |

a. Covariates appearing in the model are evaluated at the following values: Age = 44.21.