PERSILA Workshop:
Introduction to Analysis of Covariance (ANCOVA)

Wun Thiam Yew
Computer Lab A,
School of Educational Studies,
Universiti Sains Malaysia
21 September 2013

Introduction to Analysis of Covariance (ANCOVA)

- Analysis of covariance (ANCOVA) is an extension of analysis of variance (ANOVA).

- In ANCOVA, “main effects and interaction effects of IVs are assessed after DV scores are adjusted for differences associated with one or more covariates (CVs).

One-Way ANCOVA

- One-way ANCOVA is designed to assess group differences on a single DV after the effects of one or more covariates are statistically removed.

- E.g., “age and degree of reading disability are usually related to outcome of a program of educational therapy (DV). If groups are formed by randomly assigning children to different types of educational therapies (IV), it is useful to remove differences in age and degree of reading disability before examining the relationship between outcome and type of therapy. Prior differences among children in age and degree of reading disability are used as covariates.” (Tabachnick & Fidell, 2007, p. 20)

One-Way ANCOVA

- The ANCOVA question is: “Are there mean differences in outcome associated with type of educational therapy after adjusting for differences in age and degree of reading disability?” (Tabachnick & Fidell, 2007, p. 20)

- “ANCOVA gives a more powerful look at the IV-DV relationship by minimizing error variance. The stronger the relationship between the DV and the covariate(s), the greater the power of ANCOVA over ANOVA.” (Tabachnick & Fidell, 2007, p. 20)

- “ANCOVA is also used to adjust for differences among groups when groups are naturally occurring and random assignment to them is not possible.” (Tabachnick & Fidell, 2007, p. 20)

One-Way ANCOVA

- “When there are more than two groups, planned or post hoc comparisons are available in ANCOVA just as in ANOVA. With ANCOVA, selected and/or pooled (totalled) group means are adjusted for differences on covariates before differences in means on the DV are assessed.” (Tabachnick & Fidell, 2007, p. 20)

Factorial ANCOVA

- What is the different between one-way ANCOVA and factorial ANCOVA ?

- One-way ANCOVA has one IV. Factorial ANCOVA has more than one IV.

- E.g., “in the educational therapy example, another interesting IV might be gender of the child. The effects of gender, type of educational therapy and their interaction on outcome are assessed after adjusting for age and prior degree of reading disability. The interaction of gender with type of therapy asks if boys and girls differ as to which type of educational therapy is more effective after adjustment for covariates.” (Tabachnick & Fidell, 2007, p. 20)

What is ANCOVA?

- “Variables that are not part of the main experimental manipulation but have an influence on the DV are known as covariates.” (Field, 2009, p. 396)

- When we measure covariates and include them in an ANOVA we call it ANCOVA.

- Could you think about other variables that might influence students’ mathematics performance besides teaching methods?

  (a) tuition

  (b) IQ

  (c) …

What is ANCOVA?

- If covariates are measured, then it is possible to control for the influence they have on the DV.

- “If we enter the covariate into the regression model first, and then IV, we can see what effect an IV has after the effect of the covariate. As such, we partial out the effect of the covariate.” (Field, 2009, p. 396)

What is ANCOVA?

- There are two reasons for including covariates in ANOVA (Field, 2009, p. 396):

  (a) To reduce within-group error variance,

  (b) Elimination of confounds.

To Reduce Within-group Error Variance

- "We assess the effect of an experiment by comparing the amount of variability in the data that the experiment can explain against the variability that it cannot explain. If we can explain some of this 'unexplained' variance (SS_R) in terms of other variables (covariates), then we reduce the error variance, allowing us to more accurately assess the effect of the IV (SS_M)." (Field, 2009, p. 396)

  SS_M : The model sum of squares (i.e., the variability explained by the model fitted to the data)

  SS_R: The residual sum of squares (i.e., the variability that the model can't explain – the error in the model)

Elimination of Confounds

- “In any experiment, there may be unmeasured variables that confound the results (i.e., variables that vary systematically with the IV). If any variables are known to influence the DV being measured , then ANCOVA is ideally suited to remove the bias of these variables. Once a possible confounding variable [is an extraneous variable] has been identified, it can be measured and entered into the analysis as a covariate.” (Field, 2009, p. 397)

Assumptions and Issues in ANCOVA

- ANCOVA has the same assumptions as ANOVA except that there are two important additional considerations (Field, 2009, p. 397):

  (1) independence of the covariate and IV, and

  (2) homogeneity of regression slopes.

Independence of the Covariate and IV

- ANCOVA is able to reduce the within-group error variance by allowing the covariate to explain some of this error variance if the covariate is independent from the IV.

- Figures 2 through 4 show three different scenarios. Figure 2 shows a basic ANOVA. It shows that the DV (e.g., students' mathematics performance) can be partitioned into two parts that represents the IV (e.g., teaching methods) and the error or unexplained variance (i.e., factors that affects students' mathematics performance that we have not measured).

Partitioning Variance for ANOVA

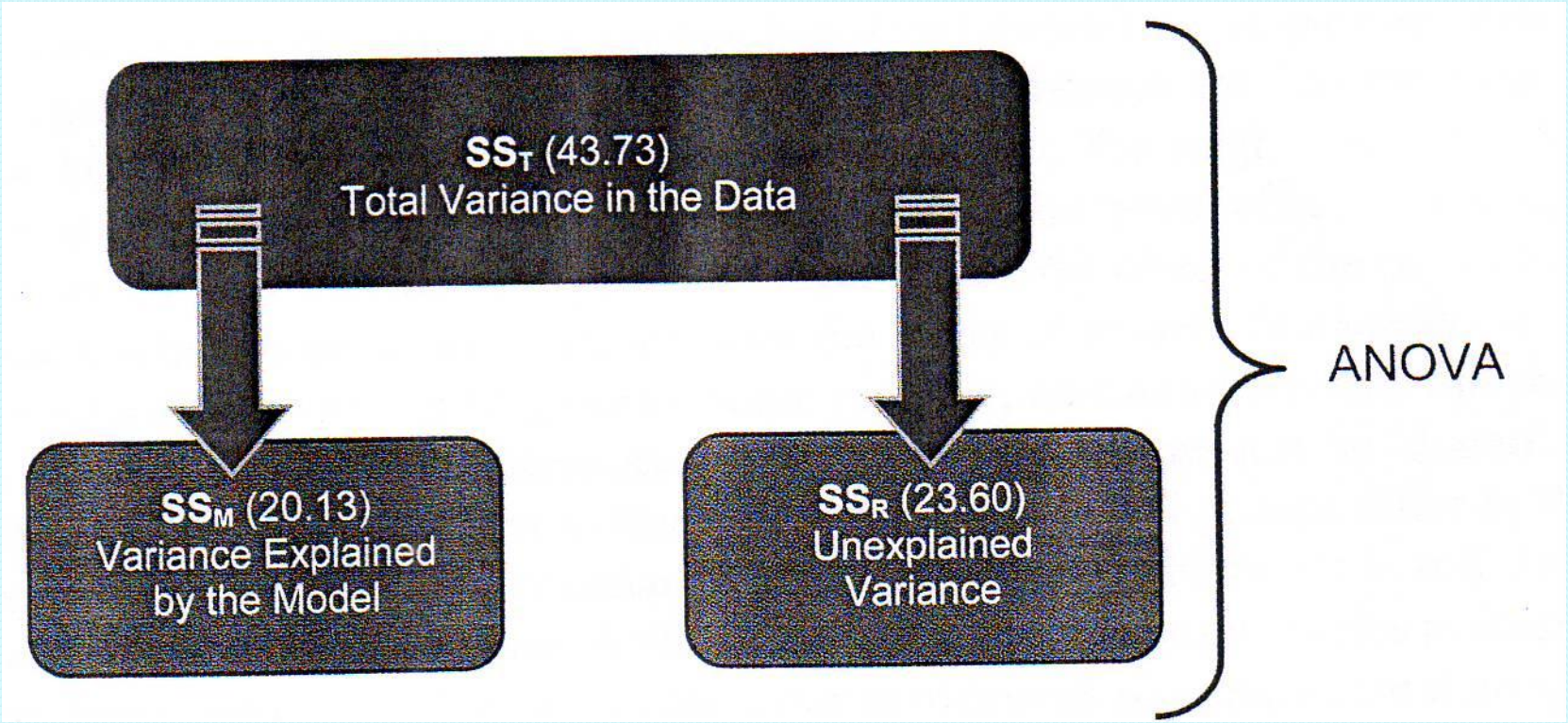


Figure 1. Partitioning variance for ANOVA.

Source: Fig. 10.3 from Field, 2009, p. 361.

The Role of the Covariate in ANCOVA

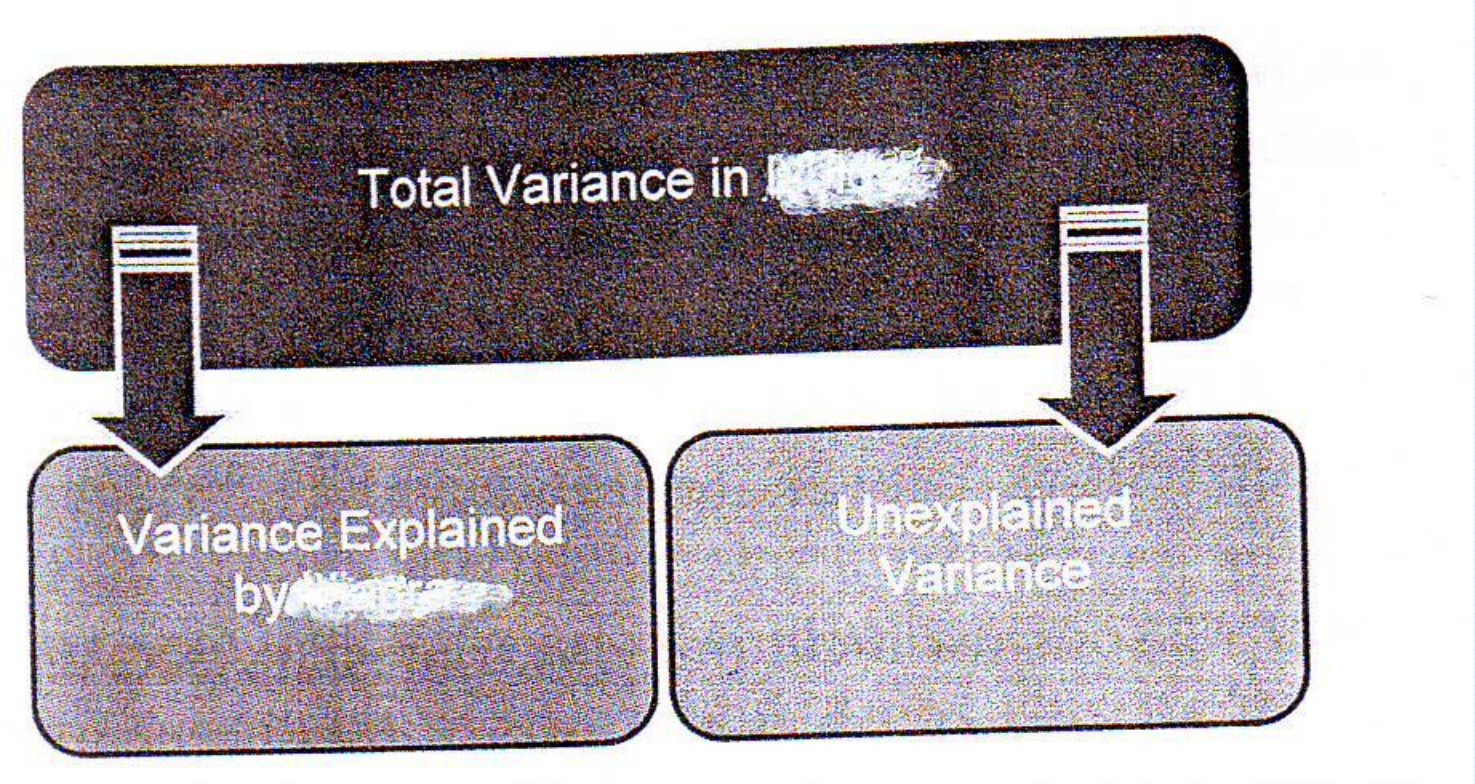


Figure 2. The role of the covariate in ANCOVA.

Source: Part A of Fig. 11.2 from Field, 2009, p. 398.

Independence of the Covariate and IV

- Figure 3 “shows the ideal scenario for ANCOVA in which the covariate shares its variance only with the bit of students’ mathematics performance that is currently unexplained. In other words, it is completely independent from the IV (it does not overlap with the effect of teaching methods at all). This scenario is the only one in which ANCOVA is appropriate.” (Field, 2009, p. 397)

The Role of the Covariate in ANCOVA

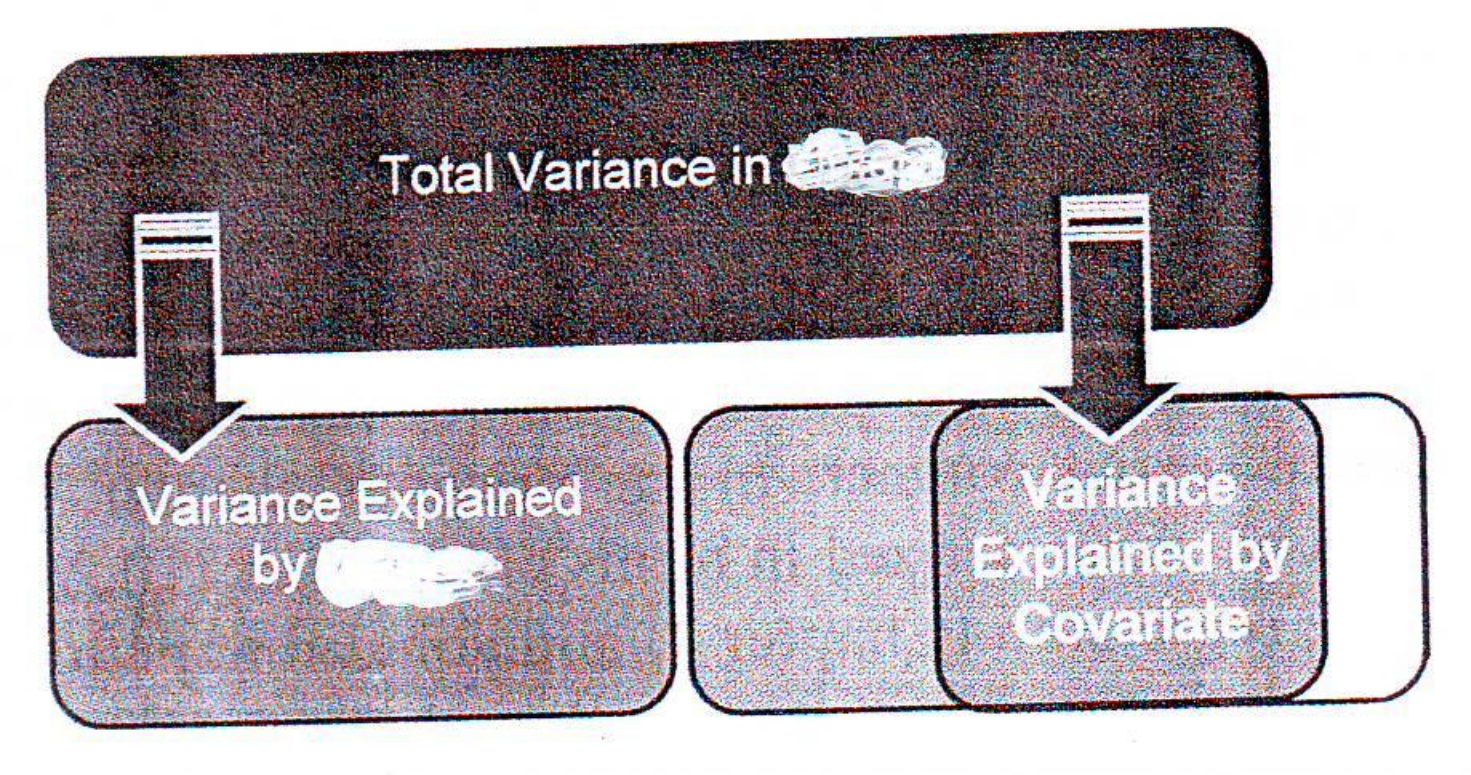


Figure 3. The role of the covariate in ANCOVA.

Source: Part B of Fig. 11.2 from Field, 2009, p. 398.

Independence of the Covariate and IV

- Figure 4 “shows a situation in which people often use ANCOVA when they should not. In this situation the effect of covariate overlaps with the IV. In other words, the covariate will reduce (statistically speaking) the IV effect because it explains some of the variance that would otherwise be attributable to the IV. When the covariate and IV are not independent, the interpretation of the ANCOVA is seriously compromised.” (Field, 2009, pp. 397-398)

The Role of the Covariate in ANCOVA

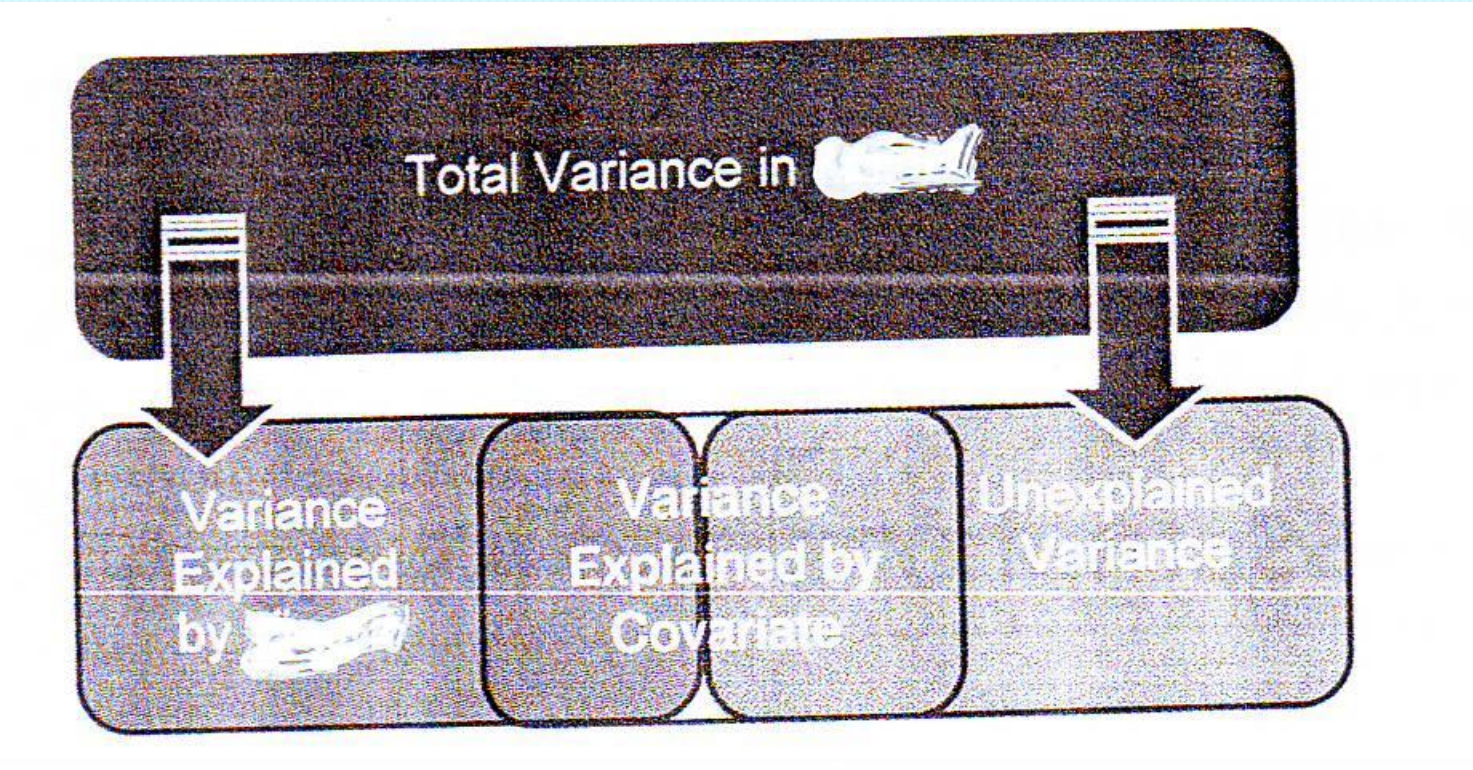


Figure 4. The role of the covariate in ANCOVA.

Source: Part C of Fig. 11.2 from Field, 2009, p. 398.

The Role of the Covariate in ANCOVA

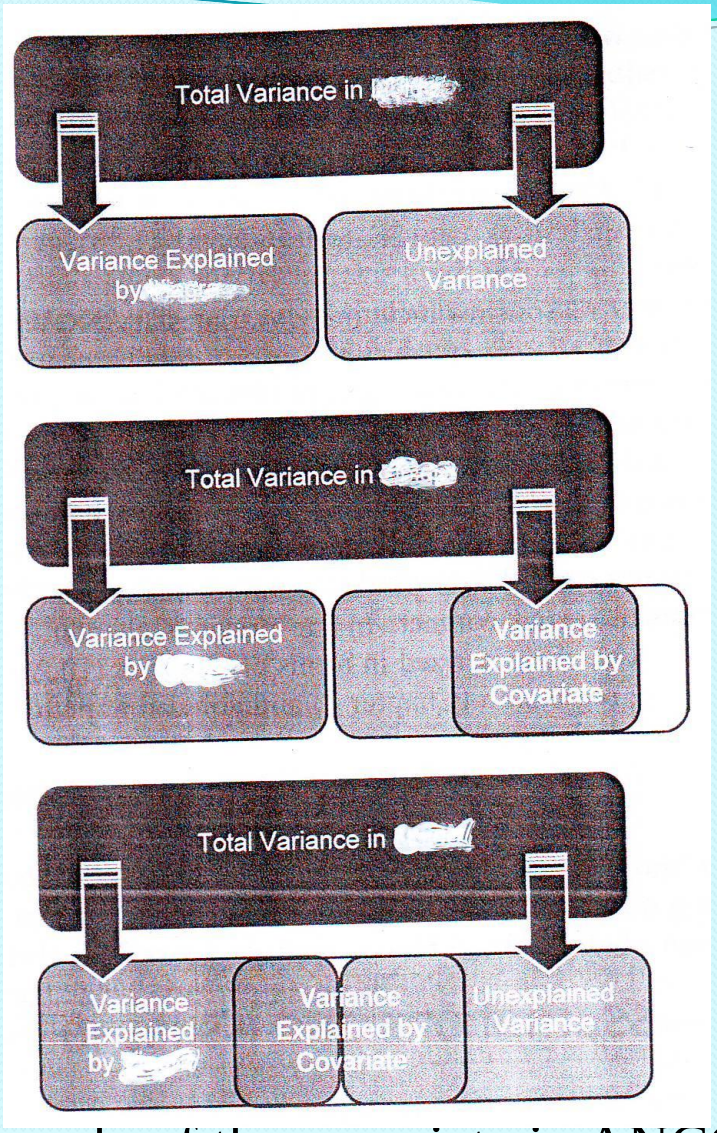


Figure 5. The role of the covariate in ANCOVA.

Source: Fig. 11.2 from Field, 2009, p. 398.

Homogeneity of Regression Slopes

- “The best way to think of this assumption is to imagine plotting a scatterplot for each experimental condition with the covariate on one axis and the outcome (DV) on the other. If you then calculated, and drew, the regression line for each of these scatterplots you should find that the regression lines look more or less the same (i.e. the values of *b* in each group should be equal).” (Field, 2009, p. 399)

Homogeneity of Regression Slopes

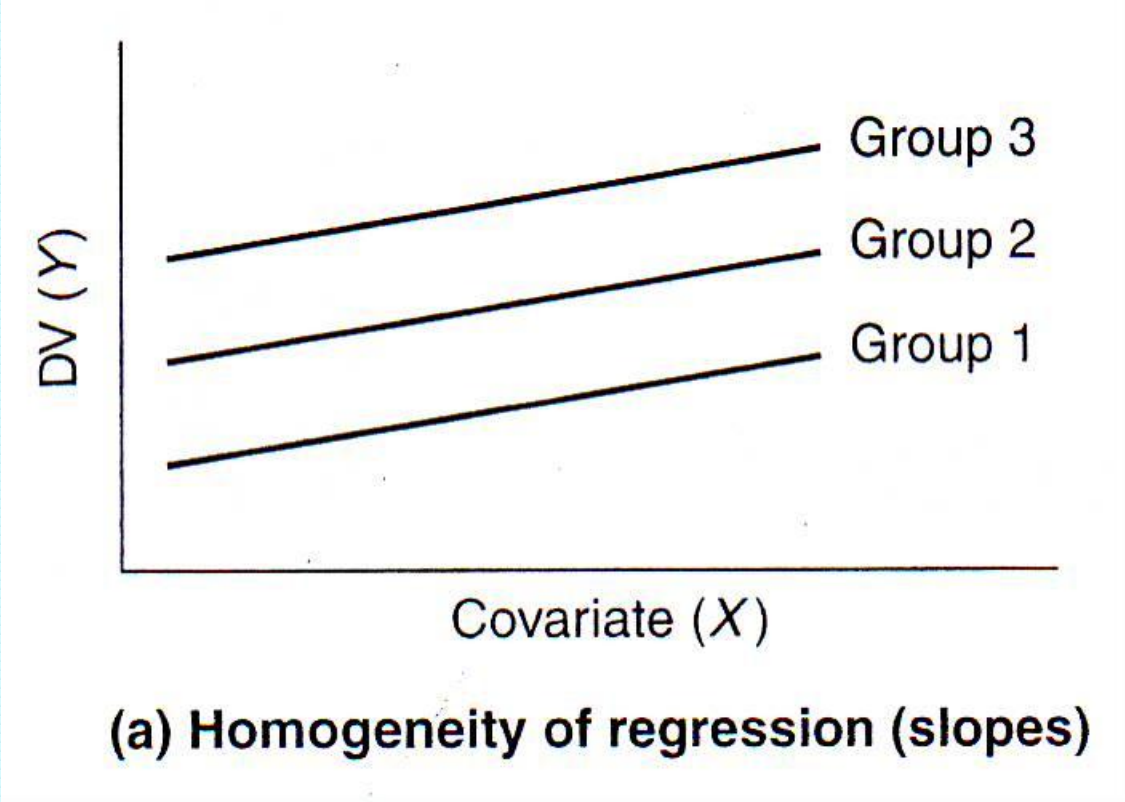


Figure 6. Homogeneity of regression slopes.

Source: Fig. 6.2a from Tabachnick & Fidell, 2007, p. 203.

Heterogeneity of Regression Slopes

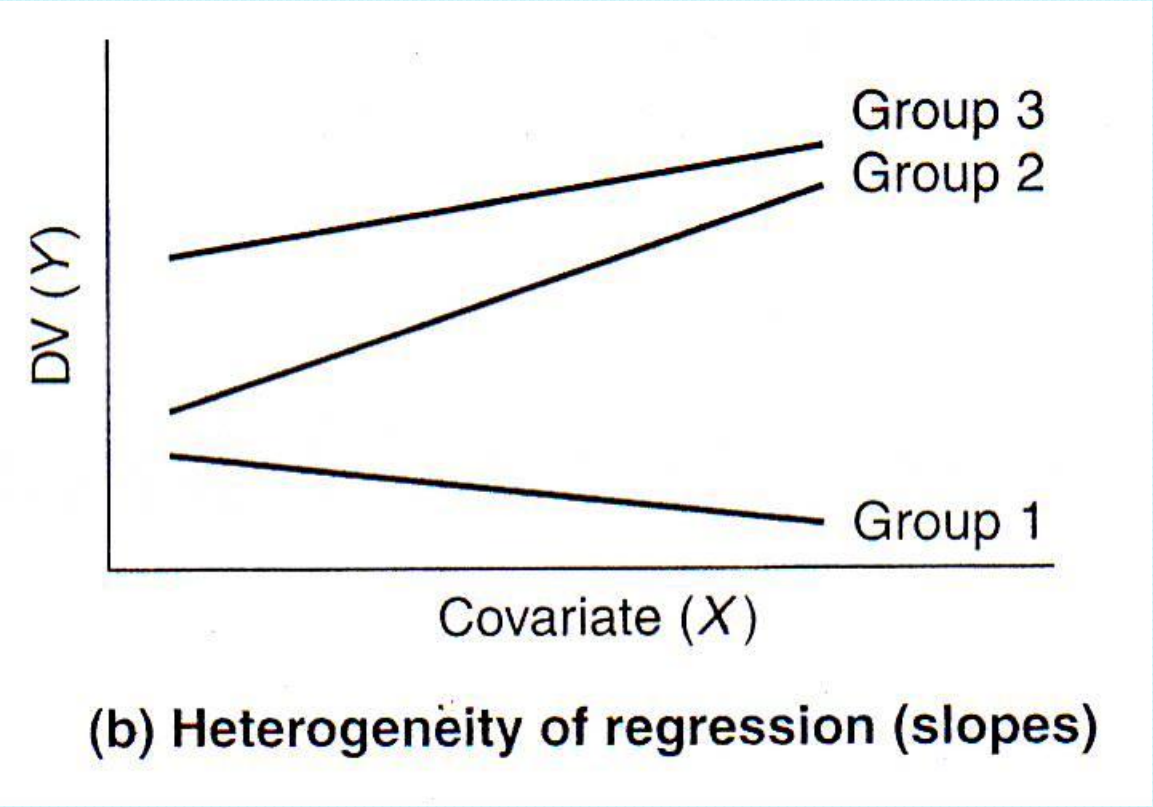
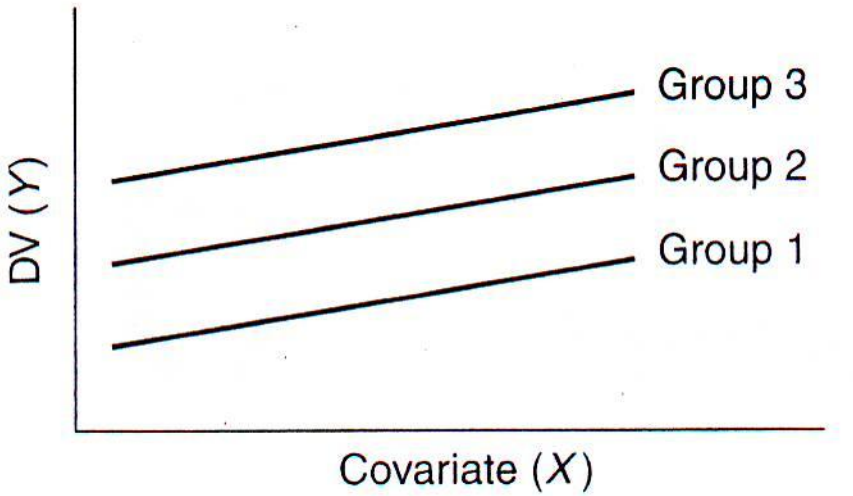


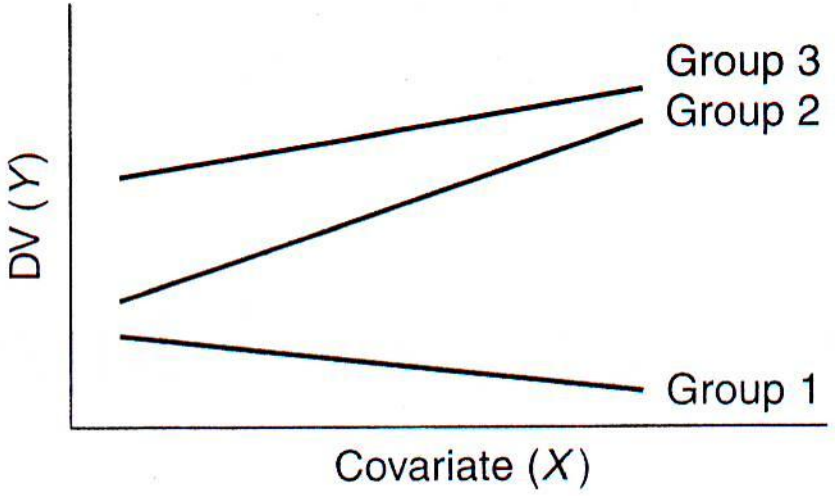
Figure 7. Heterogeneity of regression slopes.

Source: Fig. 6.2b from Tabachnick & Fidell, 2007, p. 203.

Homogeneity and Heterogeneity of Regression Slopes



(a) Homogeneity of regression (slopes)



(b) Heterogeneity of regression (slopes)

Figure 8. Homogeneity and heterogeneity of regression slopes.

Source: Fig. 6.2 from Tabachnick & Fidell, 2007, p. 203.

Sample of Research Using ANCOVA

- **Research title**: The effects of free choice learning and problem-based learning approach in a virtual science centre environment on scientific literacy among form four secondary school students.

  Note: A PhD research proposal presentation by Daniel Loy Hui Siang, 15 April 2013, Smart Lab, School of Educational Studies, Universiti Sains Malaysia.

Research Objective

- To compare the effects of free choice learning and problem-based learning approach in a virtual science centre environment on scientific literacy among form four secondary school students (Loy, 2013, p. 17).

Research Questions

1. Is there any significant difference on the result of post test of scientific literacy between students who used problem-based learning approach and students who used free choice learning approach when they are visiting the virtual science centre after the influences of pre test of scientific literacy is controlled?

- Could you formulate a RQ that involves a covariate?

IV, Covariate, and DV

- Identify the IV, covariate, and DV in the following RQs:

1. Is there any significant difference on the result of post test of scientific literacy between students who used problem-based learning approach and students who used free choice learning approach when they are visiting the virtual science centre after the influences of pre test of scientific literacy is controlled?

- Could you identify the IV, covariate, and DV in the RQs that you/your group had formulated?

Research Hypotheses

- Formulate a research hypothesis for the following RQs:

1. Is there any significant difference on the result of post test of scientific literacy between students who used problem-based learning approach and students who used free choice learning approach when they are visiting the virtual science centre after the influences of pre test of scientific literacy is controlled?

Sample of Research Hypothesis

1. H_{01}: There is no significant difference on the result of post test of scientific literacy between students who used problem-based learning approach and students who used free choice learning approach when they are visiting the virtual science centre after the influences of pre test of scientific literacy is controlled.

- Could you formulate a research hypothesis for each RQ that you/your group had formulated?

Research Sample 1

- A facilitator wished to determine whether three different teaching techniques (Simulation Technique, Buzz Session Technique, and Debate Technique) influence communication skills among the participants of a Corporate Communication course. The facilitator collected communication skills score from a group of the course participants (n = 48) that were randomly selected from its population. However, the facilitator aware that age might also be a factor that influenced communication skills of the participants. Thus, the facilitator decided to determine the relationship between teaching techniques and communication skills by controlling the age factor. The data is shown in the following file: BI - CYP - CK 1 (p. 143). What is the appropriate statistical test to be used to analyze the data? Why? (Source: Chua, 2009, p. 143)
- Hands-on activity 1. What is your conclusion?

Research Sample 2

- A researcher wished to determine whether thinking styles (left brain style, right brain style, and whole brain style) of a group of graphic designers (in an international advertisement company named ABZ in Kuala Lumpur) influence their creative thinking by controlling the factor of interest toward graphic arts. Findings of the study would be used as a criterion for the company to select its new staffs. In the study, 216 graphic designers were randomly selected from the population of graphic designers in ABZ Company. The respondents were requested to answer a questionnaire regarding their creative thinking, thinking styles, and interest toward graphic arts. The data (with minor modification) is shown in the following file: BI - CYP - CK 2 (p. 154). What is the appropriate statistical test to be used to analyze the data? Why? (Source: Chua, 2009, p. 154)
- Hands-on activity 2. What is your conclusion?

(Hypothetical) Research Sample 4

- A researcher wished to compare the effects of constructivist learning and traditional learning approach on mathematics performance among form two secondary school students. The hypothetical data is shown in the following file: Howell, 2013, pp. 621-622. The researcher attempted to answer the following RQ: Is there any significant difference on the result of post test of mathematics performance between students who used constructivist learning approach and students who used traditional learning approach after the influences of pre test of mathematics performance is controlled? What is the appropriate statistical test to be used to analyze the data? Why? (Source: Data modified from Howell, 2013, pp. 621-622)
- Hands-on activity 4. What is your conclusion?

Conducting ANCOVA on SPSS

- Steps:

1. Inputting data.

2. Initial considerations: Testing the assumptions of (a) the independence of the IV and covariate, and (b) homogeneity of regression slopes.

3. The main analysis.

4. Post hoc tests (if necessary).

Testing the assumptions of (a) the independence of the IV and covariate

- Hands-on activities 1-6:
  1. Research Sample 1
  2. Research Sample 2
  3. Research Sample 3
  4. Research Sample 4
  5. Research Sample 5
  6. Research Sample 6

Testing the assumptions of (b) homogeneity of regression slopes.

- Hands-on activities 1-6:
  1. Research Sample 1
  2. Research Sample 2
  3. Research Sample 3
  4. Research Sample 4
  5. Research Sample 5
  6. Research Sample 6

The main analysis

- Hands-on activities 1-6:
  1. Research Sample 1
  2. Research Sample 2
  3. Research Sample 3
  4. Research Sample 4
  5. Research Sample 5
  6. Research Sample 6

Post hoc tests

- If you had obtained a significant F-ratio for IV and your IV had more than two levels/categories, then you have to find out the source of the significance. Post hoc tests could be used for this purpose.

- Hands-on activities 1-6:
  1. Research Sample 1
  2. Research Sample 2
  3. Research Sample 3
  4. Research Sample 4
  5. Research Sample 5
  6. Research Sample 6

Interpreting the Output from ANCOVA

1. The main analysis.
2. Post hoc tests.

- Hands-on activities 1-6:
  1. Research Sample 1
  2. Research Sample 2
  3. Research Sample 3
  4. Research Sample 4
  5. Research Sample 5
  6. Research Sample 6

Reporting Results

- Hands-on activities 1-6:
  1. Research Sample 1
  2. Research Sample 2
  3. Research Sample 3
  4. Research Sample 4
  5. Research Sample 5
  6. Research Sample 6

Reporting Results of Research Sample 1

- [Report the main findings]:

  The ANCOVA test result indicates that the covariate, age, has a significant influence on the communication skills of the respondents, F (1, 44) = 42.67, p = .00 < .05. Teaching techniques has no significant influence on the communication skills of the respondents after controlling for the effect of age, F (2, 44) = 1.45, p = .25 > .05.

- [Report the Post hoc tests, if necessary]:

Conclusion of Research Sample 1

- Teaching techniques has no significant influence on the communication skills of the respondents after controlling for the effect of age.

Thank you.