

Analysis of Covariance (ANCOVA)

รองศาสตราจารย์ ดร. มนตรี พิริยะกุล

Email: mpiriyakul@yahoo.com, 2012

ภาควิชาสถิติ คณะวิทยาศาสตร์ มหาวิทยาลัยรามคำแหง

ANCOVA เป็นเทคนิคการศึกษาเปรียบเทียบค่าเฉลี่ยโดยควบคุมตัวแปรอื่นที่มีอิทธิพลต่อตัวแปรตาม เพราะหากไม่ควบคุมเอาไว้ผลการศึกษาอาจไม่ชัดเจนว่าการที่ตัวแปรตามมีความแตกต่างกันในระหว่างกลุ่มของตัวแปรอิสระ (เช่น เพศ อายุ หรือ Treatment ที่ใช้ในการทดลอง) เป็นเพราะผลของ Treatment เท่านั้นหรือว่า เพราะมีเหตุแทรกซ้อนอย่างอื่นคือมีตัวแปรอื่นมาทำให้แตกต่างกัน เช่น

1. ถ้าเราต้องการเปรียบเทียบผลผลิตผักที่อาจเติบโตไม่เท่ากันเมื่อให้ปุ๋ยต่างสูตรกัน เหตุแทรกซ้อนอาจเป็น ค่า Ph ของดิน ปริมาณน้ำฝนที่ตกลงในแปลงทดลอง นอกเหนือจากปริมาณน้ำที่ใส่รดแปลงผักตามปกติ หรือความชื้นแสง
2. ถ้าเราต้องการเปรียบเทียบผลการฝึกอบรมโดยใช้วิธีฝึกอบรมที่ต่างกันจะมีผลต่อการเรียนรู้ต่างกันหรือไม่ ตัวแปรร่วมอาจเป็นระดับความรู้เดิมในหัวข้อฝึกอบรม
3. ถ้าต้องการเปรียบเทียบความภักดีต่อองค์กรในระหว่างพนักงานที่มีระดับการศึกษาต่างกัน ตัวแปรร่วมอาจเป็น ความยุติธรรมในองค์กร การสนับสนุนขององค์กร หรือการรับรู้ภาวะผู้นำของผู้บริหาร

จากตัวอย่างข้างต้นจะเห็นว่าเรามีความจำเป็นต้องควบคุมอิทธิพลของตัวแปรร่วมเอาไว้จึงจะทราบผลการเปรียบเทียบที่ถูกต้อง คำว่าควบคุมก็คือบังคับไม่ให้เข้ามาเกี่ยวข้อง ซึ่งคงทำไม่ได้เพราะตัวแปรร่วมย่อมส่งผลอยู่แล้ว วิธีควบคุมจึงอาจทำได้ 2 ทางคือ 1) หักอิทธิพลของตัวแปรร่วมออกจากค่าตัวแปรตาม ซึ่งอาจหักมาน้อยเพียงใด ขึ้นอยู่กับความสัมพันธ์ระหว่างตัวแปรตามกับตัวแปรร่วม ถ้าเกี่ยวข้องมากก็หักออกมาก

ถ้าเกี่ยวข้องกับน้อยก็หักออกน้อย 2) นำตัวแปรพร้อมมารวมเป็นสภาพแวดล้อมการทดลอง (ถ้าเป็นงานทดลอง) เช่นควบคุมความชื้นเป็นระดับต่าง ๆ ควบคุมความเข้มแสงเป็นระดับต่าง ๆ ควบคุมค่าของ Ph เป็นระดับต่าง ๆ แล้วทำการทดลองตาม combination ของระดับเหล่านั้น

สำหรับกรณี One way ANCOVA ตัวแบบจะเป็นดังนี้

$$y_{ij} = \mu_y + \alpha_j + \beta_w(x_{ij} - \mu_x) + \epsilon_{ij} ; j = 1, 2, \dots, n$$

นั่นคือ $y_{ij} - \beta_w(x_{ij} - \mu_x) = \mu_y + \alpha_j + \epsilon_{ij} ; j = 1, 2, \dots, n$

$$\text{หรือ } \bar{y}_j = \mu_y + \alpha_j + \beta_w(\bar{x}_j - \bar{x}) + \bar{e} ; j = 1, 2, \dots, n$$

นั่นคือ $\bar{y}_j - \beta_w(\bar{x}_j - \bar{x}) = \mu_y + \alpha_j + \bar{e} ; j = 1, 2, \dots, n$

จะเห็นได้ว่าการควบคุมในที่นี้คือการหักอิทธิพลของ x (ตัวแปรร่วม) ออกจากค่า y ค่าของ x ที่นำไปหักคือ $\beta_w(x_{ij} - \mu_x)$ หรือในรูปค่าเฉลี่ยคือ $\beta_w(\bar{x}_j - \bar{x})$ โดยที่ β_w คือสัมประสิทธิ์การถดถอยที่ใช้เป็นระดับอิทธิพลของ x ที่มีตัว y ในกลุ่มย่อยของตัวแปรอิสระ ซึ่งถ้า β_w มีค่ามากเราจะหักอิทธิพลของ x ออกไปมาก ในที่นี้

y_{ij} คือค่าที่ i ของ y ในกลุ่มที่ j

x_{ij} คือค่าที่ i ของ x ในกลุ่มที่ j

μ_y คือค่าเฉลี่ยรวมของ y

μ_x คือค่าเฉลี่ยรวมของ x

β_w คือค่าสัมประสิทธิ์การถดถอยของสมการในกลุ่มที่ j

α_j คืออิทธิพลของ Treatment ที่ j ที่มีต่อ y

ϵ_{ij} คือค่าความคลาดเคลื่อน

จากสมการ $\bar{y}_j = \mu_y + \alpha_j + \beta_w(\bar{x}_j - \bar{x}) + \bar{e}$ จะพบว่า

$$\begin{aligned} \bar{y}_k - \bar{y}_p &= (\mu_y - \mu_y) + (\alpha_k - \alpha_p) + \beta_w\{(\bar{x}_k - \bar{x}) - (\bar{x}_p - \bar{x})\} \\ &= (\alpha_k - \alpha_p) + \beta_w(\bar{x}_k - \bar{x}_p) \end{aligned}$$

จะเห็นได้ว่าความแตกต่างระหว่างค่าเฉลี่ย \bar{y}_k กับ \bar{y}_p มิได้มีสาเหตุมาจากเฉพาะอิทธิพลของ Treatment คือ $(\alpha_k - \alpha_p)$ แต่เพราะส่วนหนึ่งมาจากอิทธิพลตัวแปรร่วม ซึ่งมากน้อยขึ้นอยู่กับค่า β ถ้า $\beta = 0$ ก็แปลว่าตัวแปรร่วมไม่มีอิทธิพล ซึ่งหากเราควบคุมอิทธิพลของตัวแปรร่วมโดยหักออกจากค่าของตัวแปรตาม จะพบว่า (ต่อไปนี้จะใช้สัญลักษณ์ β แทน β_w เพื่อแสดงให้เห็นว่าเราต้องการให้ β เป็นค่าคงที่)

$$(\bar{y}_k - \bar{y}_p) - \beta(\bar{x}_k - \bar{x}_p) = (\alpha_k - \alpha_p)$$

คราวนี้เราจะเห็นได้ว่ามีเฉพาะอิทธิพลของ Treatment เท่านั้นที่ส่งผลต่อความแตกต่างระหว่างค่าเฉลี่ยของตัวแปรตาม และหากเราหักอิทธิพลของตัวแปรร่วมออกจากตัวแปรตามจะพบว่า

$$y'_{ij} = y_{ij} - \beta(x_{ij} - \bar{x}) = \mu_y + \alpha_j + \epsilon_{ij}; i = 1, 2, \dots, n_j; j = 1, 2, \dots, m$$

หรือ $\bar{y}'_j = \bar{y}_j - \beta(\bar{x}_j - \bar{x}) = \bar{y}_j + \alpha_j; j = 1, 2, \dots, m$

ซึ่งแสดงให้เห็นว่าการควบคุมอิทธิพลของตัวแปรร่วมหมายถึงการหักอิทธิพลของตัวแปรร่วมออกไปและสิ่งที่เหลือคืออิทธิพลของ Treatment

สิ่งที่เป็นความจำเป็นต้องทราบคือเราต้องทราบอย่างแน่ชัดว่า x มีความสัมพันธ์กับ y จริงๆ (มี linear relationship) ซึ่งทราบได้จากผลการทบทวนวรรณกรรม และระดับอิทธิพลของ x ที่มีต่อ y ต้องเหมือนเดิม (คือ β_w เท่ากับ β เสมอในกลุ่มย่อยของตัวแปรอิสระหรือทุกระดับของ Treatment) เช่น การศึกษาเปรียบเทียบความผูกพันต่อองค์กร (Organization Commitment--OC) แยกตามระดับการศึกษาของพนักงาน (ต่ำกว่าปริญญาตรี ปริญญาตรี สูงกว่าปริญญาตรี) โดยมีการรับรู้ความยุติธรรมในองค์กร (perceived organization justice--POJ) เป็นตัวแปรร่วม กรณีนี้สมการความสัมพันธ์เชิงเส้นระหว่างความผูกพันต่อองค์กรและการรับรู้ความยุติธรรมในองค์กรจะต้องเป็นสมการที่มีความชัน (β_w) เท่ากันเท่ากับ β เสมอในทุกสมการคือ

กลุ่มพนักงานที่จบการศึกษาระดับต่ำกว่าปริญญาตรี $OC_{1j} = \beta_0 + \beta POJ_{1j} + u_{1j}$

กลุ่มพนักงานที่จบการศึกษาระดับปริญญาตรี $OC_{2j} = \beta_0 + \beta POJ_{2j} + u_{2j}$

กลุ่มพนักงานที่จบการศึกษาระดับสูงกว่าปริญญาตรี $OC_{3j} = \beta_0 + \beta POJ_{3j} + u_{3j}$

เส้นสมการเหล่านี้จะต้องขนานกัน ที่ต้องมีเงื่อนไขนี้เพราะเมื่อเราจะหักอิทธิพลของ POJ ออกจาก OC ก็สามารถกระทำได้โดยเสมอภาค มิใช่ว่ากลุ่มที่ 1 ถูกหักมากกว่ากลุ่มอื่น (β_1 มีค่ามากกว่า) หรือกลุ่ม 2 ถูกหักน้อยกว่ากลุ่มอื่น (β_2 มีค่าน้อยกว่า) ดังนี้ เป็นต้น

หรือในกรณีการศึกษาความภักดีต่อตราสินค้าเปรียบเทียบระหว่างลักษณะการรับรู้ข่าวสาร (สื่อมาตรฐาน, สื่อบูรณาการ, WOM) พบว่าระดับความไว้วางใจต่อตราสินค้า (brand trust--BT) มีส่วนเกี่ยวข้องกับความภักดีต่อตรา (Brand Loyalty--BL) เราต้องหักอิทธิพลของ BT ออกจาก BL มิเช่นนั้นเราจะไม่ทราบว่า BL ที่แตกต่างกันตามลักษณะการรับรู้ข่าวสารนั้นมีความแตกต่างกันจริงหรือไม่จริงโดยมิได้มีเหตุผลมาจากปัจจัยแทรกซ้อนอย่างอื่น ซึ่งการจะหักอิทธิพลของ BT ออกจาก BL เราต้องตรวจสอบดูด้วยว่า BT มีอิทธิพลต่อ BL ในทุกกลุ่มผู้บริโภค (3 กลุ่มตามลักษณะการรับรู้ข่าวสาร) เหมือนกัน (คือ $\beta_w = \beta$) หากไม่เป็นเช่นนี้ห้ามหักอิทธิพลของ BT ออกจาก BL เพราะจะไม่เป็นธรรมในระหว่างกลุ่ม และหากมีตัวแปรอื่นที่มีอิทธิพลต่อ BL อีกเช่น ความพึงพอใจและทัศนคติต่อตรา ก็ให้หักอิทธิพลของความพึงพอใจและทัศนคติต่อตราออกจาก BL โดยให้ปฏิบัติแบบเดียวกันกับกรณีของ BT

นอกจากนี้เรายังต้องตรวจสอบด้วยว่าข้อตกลงเรื่อง Stability of Variance และ Normality ยังคงเป็นจริงเช่นเดียวกับการวิเคราะห์ Analysis of variance (ANOVA) ซึ่งหากข้อตกลงนี้ไม่เป็นจริงก็ให้แปลงข้อมูล Y_{ij} เสียก่อนมิเช่นนั้นจะยังใช้ ANOVA

ไม่ได้ วิธีแปลงข้อมูลที่ง่ายและมักแก้ปัญหาได้ดีคือ log transformation คือแปลง Y_{ij} เป็น $\ln(Y_{ij})$ แล้วดำเนินการต่อไปตามปกติ

ขั้นตอนการวิเคราะห์ ANCOVA มีดังนี้

1. ทบทวนวรรณกรรมว่าตัวแปรใดบ้างเป็นตัวแปรควบคุม (Controlled variable หรือ covariate) และตัวแปรตามมีค่าแตกต่างกันตามตัวแปรอิสระใด (ตัวแปรอิสระจะเป็นตัวแปรกลุ่ม) ทั้งนี้เพื่อให้ผลการศึกษามาเทียบได้กับการศึกษาในอดีตหรือในบริบทอื่น

2. ตรวจสอบข้อตกลงเรื่อง Stability of Variance และ Normality หากขัดแย้งให้แปลงข้อมูลโดยอาศัย log transformation หรือ Box-Cox transformation

3. วิเคราะห์สมการถดถอยระหว่างตัวแปรตามกับตัวแปรควบคุมแล้วทดสอบว่าตัวแปรควบคุมในภาพรวม (ยังไม่แยกตามกลุ่มตัวแปรอิสระ) มีอิทธิพลต่อตัวแปรตามหรือไม่

4. วิเคราะห์สมการถดถอยระหว่างตัวแปรตามกับตัวแปรควบคุมแยกตามกลุ่มของตัวแปรอิสระ (นี่คือ interaction ระหว่างตัวแปรอิสระกับตัวแปรควบคุม) แล้วทดสอบว่าสัมประสิทธิ์ถดถอยของกลุ่มต่าง ๆ เหล่านั้นต่างกันหรือไม่ หากพบว่าต่างก็ไม่นำตัวแปรควบคุมตัวนั้นมาร่วมวิเคราะห์เพราะหากนำอิทธิพลของตัวแปรรวมมาหักออก จากค่าของตัวแปรตามจะไม่เกิดความเป็นธรรมชาติเพราะหักมากน้อยไม่เท่ากัน

5. หักอิทธิพลของตัวแปรร่วมออกจากค่าของตัวแปรตามแล้วเปรียบเทียบค่าเฉลี่ยในระหว่างกลุ่ม

การวิเคราะห์ข้อมูลด้วยโปรแกรม SPSS ทำได้ดังนี้

1. ทำ Data entry ซึ่ง treatment อาจเป็น categorical variable ในที่นี้จะใช้ไฟล์ Employee data.sav โดยกำหนดให้ตัวแปรตามคือ current salary ตัวแปรอิสระคือ เพศ (sex) ตัวแปรร่วมคือเงินเดือนเมื่อแรกจ้าง (beginning salary)

2. Analyze > general linear model > univariate

ส่งตัวแปรตามในช่อง Dependent variable

ส่งตัวแปร Treatment ในช่อง fixed factor (s)

ส่งตัวแปร Covariate ในช่อง covariate (s)

กดปุ่ม Option ตอน dialog โดย check ที่ Homogeneity test เพื่อทดสอบ

Stability of variance ซึ่งทำได้หลายวิธี โปรแกรม SPSS ใช้ Levene's Test

ถ้า Accept Ho: $\sigma_1^2 = \sigma_2^2 = \dots = \sigma_n^2$ ให้ทำขั้นที่ 3 ถ้าไม่ accept ให้แปลงข้อมูล

โดยอาจใช้ log transformation

หมายเหตุ Levene's Test คือสถิติทดสอบ (test statistics) ที่มุ่งทดสอบ homogeneity of variance โดยที่ถ้า W ให้ค่า p-value ต่ำกว่าระดับนัยสำคัญที่กำหนด (0.10, 0.05, 0.01)

ให้สรุปว่า variance ไม่คงที่ ให้นำมาวิจัยแก้ปัญหาเสียก่อน เช่น แปลงข้อมูลหรือเปลี่ยนไปใช้ตัวสถิติเปรียบเทียบค่าเฉลี่ยสูตรอื่นที่ไม่ต้องการข้อตกลงนี้

$$W = \frac{(n-k) \sum_{i=1}^k n_i (Z_i - Z)^2}{(k-1) \sum_{i=1}^k (Z_{ij} - \bar{Z}_i)^2} \sim F_{(k-1, n-k)} \text{ และ } Z_{ij} = (Y_{ij} - \bar{Y}_i), k = \text{จำนวนกลุ่ม}$$

ย่อย n = จำนวนตัวอย่างรวม n_i = จำนวนตัวอย่างในกลุ่มย่อย

3. ไปที่ Dialog แรกกดปุ่ม model เลือก custom โดยส่งตัวแปร treatment และ covariate แยกกัน กับส่งตัวแปร treatment และ covariate พร้อมกัน

การส่ง Covariate เพื่อตรวจสอบว่า Ho: $\beta = 0$ หรือไม่ (คือตัวแปรร่วมมือมีอิทธิพลต่อตัวแปรตามหรือไม่) ถ้าไม่มีอิทธิพลก็ให้ใช้ ANOVA ทดสอบสมมุติฐานได้โดยไม่ต้องห่วงเรื่องตัวแปรควบคุม แต่ถ้ามีอิทธิพล (p-value $\leq .05$) ให้ตรวจสอบว่าตัวแปรควบคุมมีอิทธิพลสม่ำเสมอในทุกระดับของตัวแปรอิสระหรือไม่

การส่ง Treatment และ covariate พร้อมกัน คือเลือกตัวแปรอิสระและตัวแปรพร้อมกัน แล้วส่งลงบล็อก model เพื่อตรวจสอบว่า Ho: $\beta_1 = \beta_2 = \dots = \beta_k = \beta$ ถ้า

ปฏิเสธ $H_0: \beta_1 = \beta_2 = \dots = \beta_k = \beta$ (คือ $p\text{-value} \leq .05$) แสดงว่าตัวแปรพร้อมมีผลไม่สม่ำเสมอกันในทุกระดับของตัวแปรอิสระก็ให้เลิกไม่ต้องนำตัวแปรพร้อมมาพิจารณา

สรุปก็คือให้วิเคราะห์ต่อไปเฉพาะเมื่อปฏิเสธ $H_0: \beta = 0$ (คือ $p\text{-value} \leq .05$) และยอมรับสมมุติฐาน $H_0: \beta_1 = \beta_2 = \dots = \beta_k = \beta$ (คือ $p\text{-value} \geq .05$)

4. วิเคราะห์ ANOVA คือทำกระบวนการหักอิทธิพลของตัวแปรพร้อมออกจากค่าตัวแปรตาม ให้เริ่มต้นใหม่ดังนี้

1) Analyze > general linear model > univariate ตอบ dialog โดยส่งตัวแปรตามในช่อง dependent variable ส่งตัวแปร Treatment ในช่อง fixed factor ส่งตัวแปรพร้อมในช่อง covariate

2) กดปุ่ม Model เลือกใช้ full factorial (เป็น default อยู่แล้ว)

3) กดปุ่ม Option ส่งตัวแปร Treatment ลงช่อง display means for: ให้เช็คเครื่องหมายถูกในช่อง หน้า compare main effect เลือก Bonferroni และในเฟรม display ให้เลือก Descriptive statistics

หมายเหตุ Bonferroni correction ใช้สำหรับปรับระดับนัยสำคัญกรณี multiple mean comparison เพื่อให้ระดับนัยสำคัญของการเปรียบเทียบรวม (whole family of test) ยังคงเท่าเดิม โดยกำหนดให้ระดับนัยสำคัญของการเปรียบเทียบเท่ากับ $\frac{\alpha}{m}$ โดยที่ m คือจำนวนคู่เปรียบเทียบซึ่งยังคงมีผลให้ระดับนัยสำคัญรวมเท่าเดิมคือเท่ากับ α ทั้งนี้เป็นไปตาม Bonferroni inequality คือ $\Pr(U_i^m E_i) \leq \sum_i^m \Pr(E_i)$

ทดลองรันโปรแกรมชุดดังต่อไปนี้ ปัญหาวิจัยคือผู้มีเงินได้มีรายได้ต่างกันตามเพศของผู้มีเงินได้หรือไม่ และถูกควบคุมด้วยรายได้เมื่อแรกจ้างหรือไม่ จากไฟล์ Employee data.sav ให้ส่งตัวแปร salary ลงในช่อง dependent ส่งตัวแปร sex ลงในช่อง Fixed Factor

salary	salbegin	jobtime	prevexp	minority	sex	var	var
\$57,000	\$27,000	98	144	0	1.00		
\$40,200	\$18,750	98	36	0	1.00		
\$21,450	\$12,000						
\$21,900	\$13,200						
\$45,000	\$21,000						
\$32,100	\$13,500						
\$36,000	\$18,750						
\$21,900	\$9,750						
\$27,900	\$12,750						
\$24,000	\$13,500						
\$30,300	\$16,500						
\$28,350	\$12,000						
\$27,750	\$14,250						
\$35,100	\$16,800						
\$27,300	\$13,500						
\$40,800	\$15,000						
\$46,000	\$14,250						
\$103,750	\$27,510						
\$42,300	\$14,250						
\$26,250	\$11,550						
\$38,850	\$15,000						

ผลการวิเคราะห์แสดงให้เห็นว่า $F = 119.798$ มีค่า $p\text{-value} = 0.000$ แสดงว่าชายและหญิงมีรายได้ต่างกัน

Between-Subjects Factors

		N
sex	1.00	258
	2.00	216

Tests of Between-Subjects Effects

Dependent Variable: Current Salary

Source	Type III Sum of Squares	df	Mean Square	F	Sig.
Corrected Model	27918533028.833 ^a	1	27918533028.833	119.798	.000
Intercept	535259299966.385	1	535259299966.385	2296.792	.000
sex	27918533028.833	1	27918533028.833	119.798	.000
Error	109997962407.507	472	233046530.524		
Total	699467436925.000	474			
Corrected Total	137916495436.340	473			

a. R Squared = .202 (Adjusted R Squared = .201)

คราวนี้ลองควบคุมเงินเดือนเมื่อแรกจ้างเอาไว้โดยทดสอบดูก่อนว่าเงินเดือนเมื่อแรกจ้างมีอิทธิพลต่อรายได้ปัจจุบันหรือไม่และมีอิทธิพลคงที่หรือไม่

	Corrected Model	1.073E11	3	3.575E10	548.023
Intercept	1.699E8	1	1.699E8	2.604	
sex	1.104E8	1	1.104E8	1.692	
salbegin	2.455E10	1	2.455E10	376.372	
sex * salbegin	22424440.21	1	22424440.21	.344	
Error	3.066E10	470	65237483.98		

Tests of Between-Subjects Effects

Dependent Variable: Current Salary

Source	Type III Sum of Squares	df	Mean Square	F	Sig.
Corrected Model	107254877963.366 ^a	3	35751625987.789	548.023	.000
Intercept	169906904.244	1	169906904.244	2.604	.107
sex	110387777.307	1	110387777.307	1.692	.194
salbegin	24553566098.339	1	24553566098.339	376.372	.000
sex * salbegin	22424440.219	1	22424440.219	.344	.558
Error	30661617472.973	470	65237483.985		
Total	699467436925.000	474			
Corrected Total	137916495436.340	473			

a. R Squared = .778 (Adjusted R Squared = .776)

พบว่าค่า $F = 376.372$ มีค่า $p\text{-value} = 0.000$ แสดงว่าเงินเดือนเมื่อแรกจ้างมีผลกระทบต่อเงินเดือนปัจจุบัน และที่บันทึก $\text{sex} * \text{salbegin}$ ซึ่งใช้ทดสอบ $H_0: \beta_1 = \beta_2 = \beta$ พบว่าตัวสถิติ F มีค่าเท่ากับ 0.344 มีค่า $p\text{-value} = 0.558$ แสดงว่ายอมรับสมมุติฐานดังกล่าว แสดงว่า salbegin มีอิทธิพลต่อ salary คงที่หรือสม่ำเสมอในทั้ง 2 เพศของผู้มีเงินได้ จากนั้นให้รันโปรแกรมใหม่อีกครั้ง โดยควบคุมตัวแปรเงินเดือนเมื่อแรกจ้างดังนี้

The image displays three SPSS dialog boxes for a univariate analysis. The background shows a data table with columns for Employee Code, Gender, Date of Birth, Educational Level, Employment Category, Months since Hire, Previous Experience, and Minority Classification, and rows for employees 22, 23, 24, and 25.

Univariate Dialog:

- Dependent Variable: Current Salary [salary]
- Fixed Factor(s): sex
- Covariate(s): Beginning Salary [s...]

Univariate: Options Dialog:

- Factor(s) and Factor Interactions: (OVERALL) sex
- Display Means for: sex
- Compare main effects
- Confidence interval adjustment: Bonferroni
- Display options: Descriptive statistics, Estimates of effect size, Observed power, Parameter estimates, Contrast coefficient matrix, Homogeneity tests, Spread vs. level plot, Residual plot, Lack of fit, General estimable function
- Significance level: .05, Confidence intervals are 95.0 %

Univariate: Model Dialog:

- Specify Model: Full factorial
- Factors & Covariates: sex, salbegin
- Build Term(s): Type: Interaction
- Sum of squares: Type III
- Include intercept in model

Employee Code	Gender	Date of Birth	Educational Level	Employment Category	Months since Hire	Previous Experience	Minority Classification	Current Salary
22	m	09/24/1940	12	1	\$21,750			
23	f	03/15/1965	15	1	\$24,000			
24	f	03/27/1933	12	1	\$16,950			
25	f	07/01/1942	15	1	\$21,150			

Descriptive Statistics

Dependent Variable: Current Salary

sex	Mean	Std. Deviation	N
1.00	\$41,441.78	\$19,499.214	258
2.00	\$26,031.92	\$7,558.021	216
Total	\$34,419.57	\$17,075.661	474

Tests of Between-Subjects Effects

Dependent Variable: Current Salary

Source	Type III Sum of Squares	df	Mean Square	F	Sig.
Corrected Model	107232453523.147 ^a	2	53616226761.574	823.009	.000
Intercept	577567183.563	1	577567183.563	8.866	.003
salbegin	79313920494.315	1	79313920494.315	1217.469	.000
sex	401404773.024	1	401404773.024	6.162	.013
Error	30684041913.193	471	65146585.803		
Total	699467436925.000	474			
Corrected Total	137916495436.340	473			

a. R Squared = .778 (Adjusted R Squared = .777)

Estimated Marginal Means**sex****Estimates**

Dependent Variable: Current Salary

sex	Mean	Std. Error	95% Confidence Interval	
			Lower Bound	Upper Bound
1.00	35366.040 ^a	531.815	34321.017	36411.063
2.00	33289.058 ^a	587.251	32135.103	34443.014

a. Covariates appearing in the model are evaluated at the following values: Beginning Salary = \$17,016.09.

จากตารางจะเห็นได้ว่าผู้มีเงินได้จะมีรายได้ต่างกันระหว่างเพศโดยพบว่ารายได้เมื่อแรกจ้างส่งผลกระทบต่อรายได้ปัจจุบัน ในตอนแรกพบว่ารายได้ต่างกันมากคือชายมีรายได้เฉลี่ยเท่ากับ \$41,441 หญิงมีรายได้เฉลี่ย \$26,032 ซึ่งแตกต่างกันมาก ค่า $F = 119.798$ ซึ่งสูงมาก แสดงว่าชายและหญิงมีรายได้ที่แตกต่างกันเด่นชัด เมื่อตรวจพบว่ารายได้เมื่อแรกจ้างส่งผลให้เกิดความแตกต่างด้านรายได้โดยส่งผลกระทบค่อนข้างมากและรายได้เฉลี่ยต่างกันถึง \$15,409 นำสงสัยว่าที่จริงแล้วชายและหญิงอาจมีได้มีรายได้ต่างกันถึงขนาดนั้นหรือไม่ต่างกันเลยก็ได้เพราะประเทศที่เจริญแล้วอาจไม่คิดกันทางเพศแต่เมื่อลองหักอิทธิพลของรายได้เมื่อแรกจ้างออกก็ยังพบว่ารายได้ปัจจุบันยังคงต่างกันอยู่แต่ต่างก็น้อยลง มีค่า $F = 6.162$ ซึ่งเล็กลงมาก แปลว่าไม่ต่างกันมาก โดยพบว่าชายมีรายได้เฉลี่ย \$35,366 หญิงมีรายได้เฉลี่ย \$32,289 ซึ่งต่างกันเพียง \$3,077 ทั้งนี้รายได้เมื่อแรกจ้าง

สามารถควบคุมรายได้ปัจจุบันได้ร้อยละ 77.8 ($R^2 = 0.778$) แต่ผลการทดสอบข้อตกลงเรื่อง stability of variance ซึ่งทดสอบ $H_0: \sigma_1^2 = \sigma_2^2 = \sigma^2$ ด้วย Levene's Test พบว่า $F = 33.01$ มีค่า $p\text{-value} = 0.000$ แสดงว่าปฏิเสธ $H_0: \sigma_1^2 = \sigma_2^2 = \sigma^2$ จึงแก้ปัญหาโดยการแปลงค่าตัวแปรตามและตัวแปรร่วมด้วย log transformation คือ salary \rightarrow lnsalary และ salbeg \rightarrow lnalbeg และดำเนินการใหม่ดังภาพพบว่าผลการศึกษายังคงเหมือนเดิมแต่ไม่ขัดกับข้อตกลง การแปลงแบบนี้จะมีผลให้ค่าเฉลี่ยและ SD เป็นค่า log ให้แปลงค่าเฉลี่ยและ SD กลับสู่ค่าเดิมมิเช่นนั้นจะอ่านผลผิด

```

/MEINOU=STYPE(S)
/INTERCEPT=INCLUDE
/CRITERIA=ALPHA(0.05)
/DESIGN=sex lnalbeg lnalbeg*sex.

```

Univariate Analysis of Variance

[DataSet1] C:\Program Files\IBM\SPSS\Stat

Between-Subjects Factors	
sex	N
1.00	258
2.00	216

Dependent Variable: lnsalary

Source	Type III Sum of Squares	df	Mean Square	Sig.	Partial Eta Squared
Corrected Model	58.774 ^a	3	19.591	.000	.778
Intercept	.302	1	.302	.563	
sex	.000	1	.000	.920	
lnalbeg	25.672	1	25.672	.000	
sex * lnalbeg	.001	1	.001	.924	
Error	15.901	470	.034		
Total	50917.412	474			
Corrected Total	74.675	473			

a. R Squared = .787 (Adjusted R Squared = .786)

สรุป

ANCOVA เป็นเทคนิคการวิเคราะห์ข้อมูลเพื่อศึกษาเปรียบเทียบค่าเฉลี่ยระหว่างกลุ่มประชากรโดยผู้วิจัยต้องควบคุมปัจจัยอื่นที่สัมพันธ์กับตัวแปรเป้าหมายเอาไว้เพื่อให้เห็นอย่างชัดเจนว่าตัวแปรเป้าหมายในระหว่างกลุ่มประชากรต่างกันจริง ๆ มิใช่เพราะมีอิทธิพลอื่นแฝงอยู่ สังเกตจากตัวอย่างข้างต้นจะเห็นว่าเมื่อควบคุมตัวแปรร่วมคือรายได้เมื่อแรกจ้างเอาไว้จะพบว่าตัวแปรเป้าหมายคือรายได้ในปัจจุบันยังคงแตกต่างกันอยู่เช่นเดิมแต่ต่างกันไม่มากเท่าเมื่อยังไม่ได้ควบคุมตัวแปรร่วมเอาไว้ ผลการศึกษาด้วย ANCOVA จึงเกิดขึ้นใน 2 มิติคือทำให้ทราบถึงความแตกต่างและทราบถึงสาเหตุ จึงเป็นวิชายาสถิติที่สามารถสนองการศึกษาเชิงเปรียบเทียบและการศึกษาเชิงสาเหตุได้ในเวลาเดียวกัน ข้อที่

ควรเข้าใจก็คือ อย่าสับสนระหว่างสถิติพรรณนา กับ สถิติอนุมาน ตามตัวอย่างข้างต้น สถิติพรรณนาจะบอกให้เราทราบว่ารายได้ปัจจุบันของชายและหญิงเป็นเท่าไร สถิติอนุมานคือ ANCOVA จะบอกให้เราทราบว่าชายและหญิงมีรายได้ต่างกันจริง (significantly) หรือไม่ และมีปัจจัยใดแฝงอยู่คือแอบส่งผลให้มองเห็นว่าต่างกันทั้งที่อาจไม่ต่างกันได้

บรรณานุกรม

SPSS Inc. (2010). IBM SPSS Statistics19 Command Syntax Reference

Johnson, A. R. and Wichern, D. W. (1998). **Applied Multivariate Statistical Analysis 2**, 4th edition, Prentice-Hall, NJ.

Hair Jr., J. F., Black, W. C., Babin, B. J. and Anderson, R. F.(2010)
Multivariate Data Analysis: A Global Perspective, 7th edition, Pearson Prentice Hall, NJ.