บทที่ 7
การวิเคราะห์สมการถดถอยพหุคูณ
(Multiple Regression Analysis)

วัตถุประสงค์ : เพื่อศึกษาถึงลักษณะของปัญหาต่าง ๆ ในสมการถดถอยพหุคูณและแนวทางแก้ไข ได้แก่ปัญหา autocorrelation Heteroscedasticity Multicolinearity และการตรวจสอบอำนาจการพยากรณ์ โดยนักศึกษาจะได้ศึกษาจากตัวอย่าง ตลอดจนตัวอย่างการอ่านผลการวิเคราะห์

บทที่ 7

การวิเคราะห์สมการถดถอยพหุคูณ

(Multiple Regression Analysis)

7.1 บทนำ

เมื่อนักศึกษาสร้างแบบจำลองที่มีตัวแปรอิสระตั้งแต่สองตัวขึ้นไปแล้วใช้วิธีการทางเศรษฐมิติเพื่อหาค่าพารามิเตอร์ต่าง ๆ นั้น นักศึกษาจะใช้วิธีการวิเคราะห์ความถดถอยพหุคูณในรูปสมการเส้นตรง (Multiple Linear Regression Analysis) โดยมีรูปแบบของสมการถดถอยดังนี้

$$Y = \alpha+\beta_1X_1+\beta_2X_2+\beta_3X_3+\ldots+\beta_kX_k+u$$

โดยที่ Y = ค่าของตัวแปรตาม

$X_1X_2X_3\ldots X_k$ = ค่าของตัวแปรอิสระมีค่าตั้งแต่ 1,2,...,k ตัว

α = ค่าคงที่หรือค่า Intercept ของสมการถดถอย

β_1,β_2,β_k = ค่าพารามิเตอร์หรือค่าสัมประสิทธิ์การถดถอยของตัวแปรอิสระตัวที่ 1,2,...,k ตามลำดับ

u = ค่าความคลาดเคลื่อน(Error or Residual)

7.2 ข้อกำหนดของการวิเคราะห์ความถดถอยพหุคูณ

ในการวิเคราะห์ความถดถอยพหุคูณนั้น เพื่อเป็นการควบคุมความผิดพลาดที่อาจจะเกิดขึ้นจากตัวแปรทั้งหลาย ที่จะส่งผลถึงค่าประมาณของตัวแปรตามอันอาจจะส่งผลเสียต่อการพยากรณ์ในโอกาสต่อไปได้ จึงต้องมีข้อตกลงหรือข้อกำหนดของการวิเคราะห์ดังนี้

(1) Zero mean value of u or

$$E(u_i \mid X_{1i},X_{2i},\ldots,X_{ki}) = 0 \qquad \text{for each i}$$

(2) No autocorrelation (serial correlation) or

$$Cov(u_i,u_j) = 0 \qquad \text{for } i\#j$$

(3) Homoscedasticity or

$$Var(u_i) = \delta^2$$

(4) Zero covariance between u_i and each X variable or

$$Cov(u_i,X_{1i})=Cov(u_i,X_{2i})=\ldots..=0$$

(5) No specification bias or

The model is correctly specified

(6) No exact collinearity (no multicollinearity) between the X variable or

No exact linear relationship between Xl and X2 or and X3 . . . .

7.3 การประมาณค่า

ในการประมาณค่า β ของสมการถดถอยพหุคูณนั้น มีวิธีการประมาณ 2 วิธีคือ

(1) การประมาณค่าด้วยวิธี OLS (Ordinary Least Square)

(2) การประมาณค่าด้วยวิธี MLE (Maximum Likelihood Estimator)

ทั้งสองวิธีนี้ถึงแม้ว่า จะมีวิธีการดำเนินการประมาณค่าที่แตกต่างกันแต่สุดท้ายแล้วจะได้ผลลัพธ์ออกมาเท่ากัน คือ จะได้ค่า β ที่เท่ากันโดยค่า β ต่าง ๆ จะหาได้จาก

$$\hat{\beta} = (X'X)^{-1}X'Y$$

แต่ทั้งสองวิธีนี้จะมีข้อแตกต่างกันเล็กน้อยคือค่า δ^2 ที่หาได้จากวิธี MLE จะมีค่า เท่ากับ $\sum \frac{u_i^2}{n}$ ซึ่งเป็นค่าประมาณที่ biased ส่วนค่า δ^2 ที่หาได้จากวิธี OLS จะเท่ากับ $\sum \frac{u_i^2}{n-2}$ จะเป็นค่าที่ unbias แต่อย่างไรก็ตาม ถ้าค่า n มีค่ามากขึ้นมาก ๆ จะทำให้ค่า δ^2 ที่ได้จากการคำนวณทั้ง 2 วิธี มีค่าเกือบเท่ากัน หรือใกล้เคียงกันนั่นเอง

7.4 การตรวจสอบปัญหาต่าง ๆ ตามข้อกำหนด

7.4.1 การตรวจสอบปัญหา Normality

จากข้อกำหนดข้อ 1 ที่ระบุว่าค่า u หรือค่าคลาดเคลื่อน (residual) มีค่าเฉลี่ยเท่ากับ 0 และจากข้อ 3 ทำให้เราทราบว่าค่า u มีความแปรปรวนคงที่ variance = δ^2 นั่นคือค่า u จะมีการแจกแจงแบบปกติโดยมีค่าเฉลี่ย = 0 และค่าความแปรปรวน = δ^2 และถ้าเรากำหนดให้มีการแจกแจงแบบปกติ แล้วเราสามารถเขียนได้ว่า

$$u_i \sim N(0, \delta^2)$$

ถ้าค่า u ไม่มีการแจกแจงแบบปกติจะส่งผลให้ Y_j และ β จะไม่มีการแจกแจงแบบปกติไปด้วย ทำให้การใช้ t-test และ F-test จะไม่มีประสิทธิภาพในกรณี Finite Sample

7.4.1.1 วิธีที่ใช้ในการตรวจสอบข้อกำหนดว่า u_i มีการแจกแจงแบบปกติหรือไม่นั้นสามารถกระทำได้หลายวิธี เช่น

(1) วิธีการวัดค่า Skewness และ Kurtosis

(2) วิธี Jarque – Bera Test (JB)

(3) วิธี Kolmogorv – Smirnov Test (K-S)

(4) วิธี Sharpiro – Walk Test

วิธี Skewness และ Kurtosis เป็นการเปรียบเทียบอย่างคร่าว ๆ

โดยปกติแล้วถ้าตัวแปรสุ่มใดมีการแจกแจงแบบปกติจะมีค่าความเบ้ (Skewness) = 0 แสดงว่าเป็นโค้งสมมาตร ถ้า มากกว่า 0 แสดงว่าโค้งเบ้ขวา และ ถ้าน้อยกว่า 0 แสดงว่าโค้งเบ้ซ้าย และมีค่าความโด่ง (Kurtosis) = 3 ถือว่าเป็นโค้งที่มีความลาดชันเป็นปกติถ้ามากกว่า 3 แสดงว่าโค้งมีความโด่ง โด่งกว่าโค้งปกติ และถ้าน้อยกว่า 3 แสดงว่าเป็นโค้งที่มีความโด่งแบนราบกว่าโค้งปกติ

วิธี Jarque – Bera Test (JB)

สูตรที่ใช้ในการคำนวณคือ $$JB = n\left[\frac{S^2}{6} + \frac{(K-3)^2}{24}\right]$$

โดย JB = ค่าสถิติ JB

n = จำนวนตัวอย่าง

S = ค่าสถิติ Skewness

K = ค่าสถิติ kurtosis

เมื่อทำการคำนวณค่า JB ได้แล้วให้ทำการทดสอบสมมติฐานโดยตั้งสมมติฐานหลัก H_o: u~N หมายถึงค่า u มีการแจกแจงแบบปกติ (normal distribution) และนำค่าสถิติ JB ที่คำนวณได้ไปเปรียบเทียบกับค่าตารางของตัวสถิติ χ^2 ด้วย degree of freedom = N-k-1 โดย N = จำนวนตัวอย่าง k = จำนวนพารามิเตอร์

สำหรับวิธี Kolmogorv – Smirnov Test (K-S) และวิธี Sharpiro - Wilk Test มีความซับซ้อนพอสมควรจึงจะไม่แสดงวิธีให้ดูในที่นี้

7.4.2 การตรวจสอบปัญหา Autocorrelation หรือ Serial correlation

Autocorrelation คือเหตุการณ์ที่ค่าความคลาดเคลื่อน u มีความสัมพันธ์ต่อกัน หรือมี covariation ระหว่าง u_i กับ u_j หรือ $E(u_i, u_j)$ ไม่เท่ากับ 0 เมื่อ $i \neq j$ ซึ่งเป็นเหตุการณ์ที่ขัดกับสมมติฐานของ สมการถดถอยที่กำหนดว่า $E(u_i, u_j)$ ต้องเท่ากับ 0 เมื่อ $i \neq j$

ปัญหา autocorrelation นี้จะทำให้ระดับนัยสำคัญทางสถิติมีค่าสูงกว่าที่ควรจะเป็น ค่า R^2 และค่า F statistic จะไม่น่าเชื่อถือในความถูกต้อง ปัญหานี้อาจเกิดจากการที่ตัวแปรทางเศรษฐ-

ศาสตร์มีลักษณะเป็นแนวโน้ม (trends) หรือวัฏจักร(cycles) หรือเกิดจากการที่ไม่รวมตัวแปรอิสระไว้ในแบบจำลอง

7.4.2.1 วิธีตรวจสอบปัญหา autocorrelation นั้นสามารถตรวจสอบได้หลายวิธีแต่วิธีที่ง่ายและนิยมใช้กันก็คือ พิจารณาจากค่า Durbin-Watson

โดยค่า durbin-watson หรือ D_w หรือ d หาได้จาก

$$d = \frac{\sum (u_t - u_{t-1})^2}{\sum u_t^2}$$

ในการทดสอบปัญหา autocorrelation อาจจะตั้งสมมติฐานดังนี้

(1) ตั้ง H_o : ไม่เกิด autocorrelation

H_a : เกิด autocorrelation ทางบวก

ถ้า $d < d_L$ จะปฏิเสธสมมติฐานหลัก ณ ระดับนัยสำคัญ α คือ เกิด auto correlation ทางบวก

ถ้า $d > d_u$ จะยอมรับสมมติฐานหลัก ณ ระดับนัยสำคัญ α คือไม่เกิด auto correlation

ถ้า $d_L < d < d_u$ ไม่อาจจะตัดสินใจได้ ต้องใช้วิธีการทดสอบ วิธีอื่นต่อไป

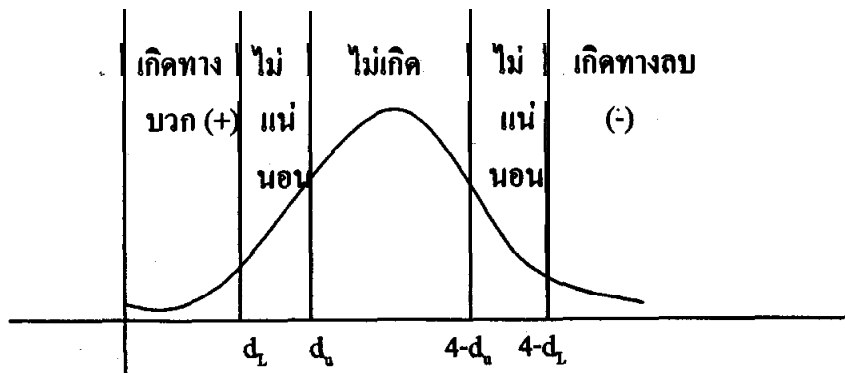
(2) ตั้ง H_o : ไม่เกิด autocorrelation

H_a : เกิด autocorrelation ทางลบ

ถ้า $d > 4-d_L$ จะปฏิเสธสมมติฐานหลัก ณ ระดับนัยสำคัญ α คือ เกิด autocorrelation ทางลบ

ถ้า $d < 4-d_u$ จะยอมรับสมมติฐานหลัก ณ ระดับนับสำคัญ α คือ ไม่เกิด autocorrelation

ถ้า $4-d_u < d < 4-d_L$ ไม่สามารถตัดสินใจได้ ต้องใช้วิธีการทดสอบวิธีอื่น



ภาพที่ 7.1 แสดงการทดสอบปัญหา autocorrelation

7.4.2.2 การแก้ปัญหา autocorrelation หรือ serial correlation อาจจะทำได้โดยการเพิ่มตัวแปรเวลาเข้าไปในแบบจำลอง โดยอาจจะเพิ่มในลักษณะกำลังหนึ่ง กำลังสอง หรือรูปแบบอื่น ๆขึ้นอยู่กับลักษณะของแนวโน้มว่า จะเป็นเส้นตรง เส้นโค้ง หรืออื่น ๆ

7.4.3 การตรวจสอบปัญหา Heteroscedasticity

ปัญหา Heteroscedasticity คือการที่ข้อมูลเกิดปัญหาการขัดแย้งกับข้อกำหนดที่ว่า Variance ของ u_i จะต้องมีค่าคงที่ทุกค่าของ i = 1,2, … , n ปัญหาดังกล่าวนี้มักจะเกิดขึ้นในกรณีที่ใช้ข้อมูลภาคตัดขวาง (cross-section data) ซึ่งมีผลทำให้การใช้วิธี OLS ในการประมาณค่าขาดลักษณะที่พึงประสงค์ของตัวประมาณค่าที่ดี

7.4.3.1 วิธีที่ใช้ในการทดสอบปัญหา Heteroscedasticity มีหลายวิธีแต่ในที่นี้จะเสนอวิธีทดสอบของ Goldfeld – Quandt test

Goldfeld – Quandt test โดยมีขั้นตอนการทดสอบดังนี้

(1) ทำการจัดเรียงค่าสังเกตของตัวแปรทุกตัวให้มีค่าจากน้อยไปมาก หรือมากไปน้อยก็ได้ โดยยึดตัวแปรอิสระในสมการตัวที่มีความสำคัญมากที่สุด (ตัวที่คิดว่ามีความสำคัญมากที่สุดถ้าไม่ทราบ) อาจจะพิจารณาจากค่า Standardized Coefficients (Beta) ที่มีค่ามากที่สุด

(2) ทำการแบ่งค่าสังเกตออกเป็น 3 ส่วน และตัดส่วนกลางทิ้งไปจำนวน c ตัวอย่าง โดยให้ส่วนที่เหลือ หัว-ท้าย มีจำนวนข้อมูลเหลืออยู่เท่ากันเป็นจำนวนเท่ากับ (n-c)/2

(3) ทำการวิเคราะห์สมการถดถอยด้วยวิธี OLS จากค่าตัวอย่างทั้งสองกลุ่ม แล้วทำการหาค่า residual sums of squares RSS_1 กับ RSS_2 โดย RSS_1 คือค่าจากชุดที่ค่า X มีค่าน้อย และ RSS_2 คือค่าจากชุดที่ค่า X มีค่ามาก

(4) ทำการคำนวณหาค่า $F^* = RSS_2 / RSS_1$

(5) ทำการทดสอบสมมติฐาน $H_o : \delta_1^2 = \delta_2^2$ ณ ระดับนัยสำคัญ α

ถ้า $F^* > F_{\{(n-c-2k)/2,(n-c-2k)/2,\alpha\}}$ แสดงว่าเราปฏิเสธ H_o หมายความว่าเกิดปัญหา Heteroscedasticity ขึ้นแล้ว แต่ถ้า $F^* < F_{\{(n-c-2k)/2,(n-c-2k)/2,\alpha\}}$ แสดงว่าเราไม่สามารถปฏิเสธ H_o หมายความว่าไม่เกิดปัญหา Heteroscedasticity แต่อย่างไร

7.4.3.2 วิธีแก้ปัญหา Heteroscedasticity

การแก้ปัญหาดังกล่าวนี้ อาจจะทำได้โดยการแปลงตัวแปรตามและตัวแปรอิสระ ให้เป็นอัตราส่วน (ratio) ด้วยการนำค่าของตัวแปรอิสระตัวที่ก่อให้เกิดปัญหา Heteroscedasticity ไปหาร

ตัวแปรตามและตัวแปรอิสระทุกตัว หรือใช้วิธีประมาณสมการถดถอยแบบถ่วงน้ำหนัก (weighted least square regression methods)

7.4.4 การตรวจสอบปัญหา Multicolinearity

ปัญหา Multicolinearity คือ ปัญหาที่ตัวแปรอิสระมีความสัมพันธ์กันเองเกินกว่าระดับที่ยอมรับได้ เพราะถ้าตัวแปรอิสระมีความสัมพันธ์กันเองมากจะไปละเมิด สมมติฐานข้อที่ว่า ตัวแปรอิสระต้องไม่มีความสัมพันธ์เชิงเส้นต่อกัน การที่เกิดปัญหานี้จะทำให้ค่าพารามิเตอร์ของตัวแปรอิสระเกิดความผิดพลาด และไม่มีนัยสำคัญทางสถิติ ทำให้ข้อสรุปเกี่ยวกับ marginal effect ของตัวแปรตามที่เกิดจากตัวแปรอิสระตัวนั้น ๆ เกิดความผิดพลาดได้

สาเหตุที่ตัวแปรอิสระมีความสัมพันธ์กันอาจเนื่องมาจากตัวแปรทางเศรษฐศาสตร์มีแนวโน้มเคลื่อนไหวไปด้วยกันในช่วงเวลาใดเวลาหนึ่ง และอีกสาเหตุหนึ่งอาจมาจากการใช้ตัวแปรในอดีตมาเป็นตัวแปรอิสระในแบบจำลองร่วมกับตัวแปรปัจจุบันซึ่งการใช้ตัวแปรในอดีตนี้มักจะได้ผลเป็นที่น่าพอใจแต่ในความเป็นจริงแล้วตัวแปรในอดีตมักจะมีความสัมพันธ์กับตัวแปรในปัจจุบัน เช่นตัวแปรรายได้ในช่วงเวลาที่แล้วมักจะมีความสัมพันธ์กับตัวแปรรายได้ในปัจจุบัน

7.4.4.1 วิธีที่ใช้ในการตรวจสอบปัญหา Multicolinearity นั้นมีวิธีตรวจสอบง่าย ๆ อาจจะพิจารณาจากค่า partial correlation coefficient ระหว่างตัวแปรอิสระด้วยกัน ถ้าค่า Pearson Correlation มีค่ามากกว่า 0.7 ก็อาจจะถือว่า ตัวแปรอิสระคู่นั้นๆ มีความสัมพันธ์กันมาก ควรได้รับการแก้ไข

7.4.4.2 ในการแก้ไขปัญหา Multicolinearity นั้นอาจจะแก้ได้โดย

(1) ทำการเพิ่มขนาดของจำนวนตัวอย่าง (sample size) หรือชุดข้อมูล (observations) ให้มากขึ้น อาจจะสามารถแก้ปัญหาได้

(2) ละทิ้งตัวแปรใดตัวแปรหนึ่ง โดยอาจจะพิจารณาจากความสำคัญของตัวแปรด้วยทฤษฎีว่าตามทฤษฎีแล้วตัวแปรตัวใดควรจะมีความสำคัญมากกว่า ก็เลือกตัวแปรนั้นหรือ ถ้าคิดว่ามีความสำคัญเท่ากัน แต่จำเป็นต้องตัดออก 1 ตัว ก็อาจจะพิจารณาจากค่า Pearson correlation ดูว่าตัวแปรอิสระคู่ที่มีปัญหา ตัวแปรใดมีความสัมพันธ์กับตัวแปรตาม มากกว่าก็จะเก็บตัวแปรอิสระตัวนั้นไว้ และตัดตัวแปรอีกตัวหนึ่งออกไป

(3) ถ้าไม่อยากละทิ้งตัวแปรใดตัวแปรหนึ่งไปเลยก็อาจจะหาความสัมพันธ์ของทั้ง 2 ตัวแปร แล้วนำส่วนที่ไม่มีความสัมพันธ์กันมาใช้ เช่นตัวแปรการศึกษา (EDUC) กับ ตัวแปรรายได้เริ่มต้น (BEGINSAL) อาจจะหาความสัมพันธ์โดย

$$BEGINSAL = a + b\,EDUC$$

ทำการประมาณค่าได้

$$\widehat{BEGINSAL} = \hat{a} + \hat{b}\,EDUC$$

นำค่า $\widehat{BEGINSAL}$ ที่ได้จากการประมาณค่าลบออก จาก BEGINSAL ก็จะได้ ส่วนที่ไม่มีความสัมพันธ์กับEDUC สมมติได้เป็นSALEDU

มาจาก $SALEDU = BEGINSAL - \widehat{BEGINSAL}$

ใช้ตัวแปร SALEDU ที่คำนวณได้ไป เป็นตัวแปรอิสระแทน BEGINSAL

7.4.5 การตรวจสอบอำนาจการพยากรณ์

การทดสอบว่าสมการประมาณค่า $Y = a+b_1x_1+b_2x_2+\ldots+b_kx_k$ มีอำนาจในการพยากรณ์ สูง-ต่ำ เพียงใด นั้น เราอาจจะทดสอบได้ง่าย ๆ คือ

(1) ทำการแทนค่าตัวแปรอิสระทุกตัวจะได้ค่า ณ เวลาต่าง ๆ $t = 1,2,3,\ldots,n$

(2) นำค่า $\hat{Y}_t$ ที่ได้เป็นตัวแปรอิสระ ดูความสัมพันธ์กับตัวแปรตาม Y_t โดย

$$Y_t = a + b\hat{Y}_t$$

ทำการวิเคราะห์โดยวิธี OLS ได้ค่า a และ b

ถ้าทดสอบได้ว่า a = 0 และ b = 1 แสดงว่า $\hat{Y}_t$ จะเป็นตัวพยากรณ์ที่มีประสิทธิภาพ (Efficient Predictor) คือค่า $\hat{Y}_t$ หรือสมการ $\hat{Y}_t = \hat{a} + \hat{b}_1X_1 + \hat{b}_2X_2 + \ldots$ เป็นสมการที่สามารถนำไปใช้ในการทำนายผลของตัวแปรตาม Y ได้เป็นอย่างดี

7.5 ตัวอย่าง

แบบจำลองที่มีตัวแปรตามคือ รายได้ในปัจจุบัน (Current Salary) โดยมีตัวแปรอิสระคือ เพศ (Gender) ระดับการศึกษา (Educational Level) โดยวัดระดับการศึกษาด้วยระยะเวลาที่ใช้ในการศึกษาเป็นปี (Year) รายได้เมื่อแรกจ้าง (Beginning Salary) อายุการทำงาน (Job time month since hire) และ ประสบการณ์ (Previous Experience) โดยใช้ข้อมูลจากการสำรวจชาวอเมริกันจำนวน 474 คน ในการวิเคราะห์จะใช้วิธีวิเคราะห์ถดถอยแบบพหุคูณ(Multiple Linear Regression Analysis) ด้วยวิธี ENTER

ตารางที่ 1 แสดงตัวอย่างข้อมูลของตัวแปรต่าง ๆ

NUMBER	SALARY	SEX	EDUC	SALBEGIN	JOBTIME	PREVEXP	PRE_1	RES_1	RES_2
1	$57000	1	15	$27000	98	144	55206.27	1793.73	-
2	$40200	1	16	$18750	98	36	43685.46	-3485.46	1793.73
3	$21450	0	12	$12000	98	381	20746.18	703.82	-3485.46
4	$21900	0	8	$13200	98	190	24153.69	-2253.69	703.82
5	$45000	1	15	$21000	98	138	44986.2	13.8	-2253.69
6	$32100	1	15	$13500	98	67	33445.31	-1345.31	13.8
7	$36000	1	15	$18750	98	114	41576.41	-5576.41	-1345.31
8	$21900	0	12	$9,750	98	0	24275.09	-2375.09	-5576.41
9	$27900	0	15	$12750	98	115	28987.37	-1087.37	-2375.09
10	$24000	0	12	$13500	98	244	25993.1	-1993.1	-1087.37
11	$30300	0	16	$16500	98	143	35496.62	-5196.62	-1993.1
12	$28350	1	8	$12000	98	26	27506.8	843.2	-5196.62
13	$27750	1	15	$14250	98	34	35378.79	-7628.79	843.2
14	$35100	0	15	$16800	98	137	35537.04	-437.04	-7628.79
15	$27300	1	12	$13500	97	66	31531.11	-4231.11	-437.04
16	$40800	1	12	$15000	97	24	34931.6	5868.4	-4231.11
17	$46000	1	15	$14250	97	48	34952.14	11047.86	5868.4
18	103750	1	16	$27510	97	70	57961.66	45788.34	11047.86
19	$42300	1	12	$14250	97	la3	32104.06	1019594	45788.34
al	$26250	0	12	$11550	97	48	26288.63	-38.63	10195.94

```
*$$$$$$$$$$$$$$$$$$$$$$$$$$$$$$$$$$$$$$$$$$$$$$$*
DESCRIPTIVES
 VARIABLES=salary sex educ salbegin jobtime prevexp
 /STATISTICS=MEAN SUM STDDEV MIN MAX.
*$$$$$$$$$$$$$$$$$$$$$$$$$$$$$$$$$$$$$$$$$$$$$$$*
REGRESSION
 /DESCRIPTIVES MEAN STDDEV CORR SIG N
 /MISSING LISTWISE
 /STATISTICS COEFF OUTS R ANOVA COLLIN TOL CHANGE ZPP
 /CRITERIA=PIN(.05) POUT(.10)
 MOORIGIN
 /'DEPENDENT salary
 /METHOD=ENTER sex educ salbegin prevexp jobtime
 /RESIDUALS DURBIN HIST(ZRESID)
 /SAVE PRED RESID .
*$$$$$$$$$$$$$$$$$$$$$$$$$$$$$$$$$$$$$$$$$$$$$$$$*
DESCRIPTIVES
 VARIABLES=res_1
 /STATISTICS=MEAN STDDEV KURTOSIS SKEWNESS .
*$$$$$$$$$$$$$$$$$$$$$$$$$$$$$$$$$$$$$$$$$$$$$$$*
NPAR TESTS
 /K-S(NORMAL)= res_1
 /STATISTICS DESCRIPTIVES
 /MISSING ANALYSIS.
*$$$$$$$$$$$$$$$$$$$$$$$$$$$$$$$$$$$$$$$$$$$$$$$*
SORT CASES BY
 salbegin (A) .
*$$$$$$$$$$$$$$$$$$$$$$$$$$$$$$$$$$$$$$$$$$$$$$$*
```

```
*$$$$$$$$$$$$$$$$$$$$$$$$$$$$$$$$$$$$$$$$$$$$$$$$$$*
Filter off.
use 1 thru 178.
execute.
*$$$$$$$$$$$$$$$$$$$$$$$$$$$$$$$$$$$$$$$$$$$$$$$$$$*
REGRESSION
 /MISSING LISTWISE
 /STATISTICS COEFF OUTS R ANOVA
 /CRITERIA=PIN(.05) POUT(.10)
 /NOORIGIN
 /DEPENDENT salary
 /METHOD=ENTER educ jobtime prevexp salbegin sex.
*$$$$$$$$$$$$$$$$$$$$$$$$$$$$$$$$$$$$$$$$$$$$$$$$$$*
Filter off.
use all.
execute.
Filter off.
use 297 thru 474.
execute.
*$$$$$$$$$$$$$$$$$$$$$$$$$$$$$$$$$$$$$$$$$$$$$$$$$$*
REGRESSION
 /MISSING LISTWISE
 /STATISTICS COEFF OUTS R ANOVA
 /CRITERIA=PIN(.05) POUT(.10)
 /NOORIGIN
 /DEPENDENT salary
 /METHOD=ENTER educ jobtime prevexp salbegin sex.
*$$$$$$$$$$$$$$$$$$$$$$$$$$$$$$$$$$$$$$$$$$$$$$$$$$*
```

```
*$$$$$$$$$$$$$$$$$$$$$$$$$$$$$$$$$$$$$$$$$$$$$$$*
Filter off.
use all.
execute.
*$$$$$$$$$$$$$$$$$$$$$$$$$$$$$$$$$$$$$$$$$$$$$$$*
COMPUTE res_2 = LAG(res_1,1) .
EXECUTE .
*$$$$$$$$$$$$$$$$$$$$$$$$$$$$$$$$$$$$$$$$$$$$$$$*
REGRESSION
 /MISSING LISTWISE
 /STATISTICS COEFF OUTS R ANOVA
 /CRITERIA=PIN(.05) POUT(.lO)
 /NOORIGIN
 /DEPENDENT res_1
 /METHOD=ENTER res_2 .
*$$$$$$$$$$$$$$$$$$$$$$$$$$$$$$$$$$$$$$$$$$$$$$$*
REGRESSION
 /MISSING LISTWISE
 /STATISTICS COEFF OUTS R ANOVA
 /CRITERIA=PIN(.05) POUT(.10)
 /NOORIGIN
 /DEPENDENT salary
 /METHOD=ENTER pre_1
 /RESIDUALS DURBIN .
*$$$$$$$$$$$$$$$$$$$$$$$$$$$$$$$$$$$$$$$$$$$$$$$*
```

ตารางที่ 2 แสดงสถิติพรรณา(Descriptive Statistic) ของตัวแปรต่าง ๆ

Descriptive Statistics

	N	Minimum	Maximum	Sum	Mean	Std. Deviation
SALARY Current Salary	474	$15,750	$135,000	$16,314,875	$34,419.57	$17,075.66
SEX Gender	474	0	1	258	.54	.50
EDUC Educational Level (years)	474	8	21	6395	13.49	2.88
SALBEGIN Beginning Salary	474	$9,000	$79,980	$8,065,625	$17,016.09	$7,870.64
JOBTIME Months since Hire	474	63	98	38446	81.11	10.06
PREVEXP Previous Experience (months)	474	0	476	45436	95.86	104.59
Valid N (listwise)	474					

ตารางที่ 3 แสดงค่าสหสัมพันธ์ระหว่างตัวแปรต่าง ๆ(Correlation)

Correlations

		SALARY Current Salary	SEX Gender	EDUC Educational Level (years)	SALBEGIN Beginning Salary	PREVEXP Previous Experience (months)	JOBTIME Months since Hire
Pearson Correlati	SALARY Current Salary	1.000	.450	.661	.880	-.097	.084
	SEX Gender	.450	1. m	.356	.457	.165	.066
	EDUC Educational Level (years)	.661	.356	1.000	.633	-.252	.047
	SALBEGIN Beginning Sal	.880	.457	.633	1.000	.045	-.020
	Experience (months)	-.097	.165	-.252	.045	1.000	.003
	JOBTIME Months since H	.084	.066	.047	-.020	.003	1.000
Sig. (1-tailed)	SALARY Current Salary	.	.000	.000	.000	.017	.034
	SEX Gender	.000	.	.000	.000	.000	.074
	EDUC Educational Level (years)	.000	.000	.	.000	.000	.152
	SALBEGIN Beginning Sal	.000	.000	.000	.	.163	.334
	PREVEXP Previous Experience (months)	.017	.000	.000	.163	.	.474
	JOBTIME Months since H	.034	.074	.152	.334	.474	.
N	SALARY Current Salary	474	474	474	474	474	474
	SEX Gender	474	474	474	474	474	474
	EDUC Educational Level (years)	474	474	414	474	474	474
	SALBEGIN Beginning Sal	474	474	474	474	474	474
	PREVEXP Previous Experience (months)	474	474	474	474	474	474
	JOBTIME Months since H	474	474	474	474	474	474

ตารางที่ 4 แสดงตัวแปรอิสระและตัวแปรตาม

Variables Entered/Removed [b]

Model	Variables Entered	Variables Removed	Method
1	JOBTIME Months since Hire, PREVEXP Previous Experience (months)SALBEGIN Beginning Salary, SEX Gender, EDUC EducationalLevel (years) [a]		Enter

a. All requested variables entered.

b. Dependent Variable: SALARY Current Salary

ตารางที่ 5 แสดงการวิเคราะห์ความแปรปรวน(Analysis of Variance)

ANOVA [b]

Model		Sum of Squares	df	Mean Square	F	Sig.
1	Regression	112216333929.663	5	22443266785.933	408.692	.000[a]
	Residual	25700161506.677	468	54914875.014		
	Total	137916495436.340	473			

a. Predictors: (Constant), JOBTIME Months since Hire, PREVEXP Previous Experience (months), SALBEGIN Beginning Salary, SEX Gender, EDUC Educational Level (years)

b. Dependent Variable: SALARY Current Salary

ตารางที่ 6 แสดงค่าสถิติ R ต่าง ๆ และ Durbin-Watson

Model Summary

Model	R	I? Square	Adjusted R Square	Std. Error of the Estimate	Change Statistics: R Square Change	F Change	dfl	df2	Sig. F Change	Durbin-Watson
1	.902[a]	.814	.812	$7,410.46	.814	408.692	5	468	.000	1.923

a. Predictor-a: (Constant), SEX Gender, JOBTIME Months since Hire, PREVMP Previous Experience (months), SALBEGIN Beginning Salary, EDUC Educational Level (years)

b. Dependent Variable: SALARY Current Salary

ตารางที่ 7 แสดงค่าสัมประสิทธิ์สหสัมพันธ์(Coefficient)

Coefficients[a]

Model		Unstandardized Coefficients: B	Std Error	Standerdized Coefficients: Beta	t	Sig	Correlations: order	Zero-al	Parti Part	Collinearity Statistics: Tolerance	VIF
1	(Constant)	-17[illegible]2.95	3267.79		-4.524		I.000	I	I		I
	EDUC Educational Level (years)	593.031	166.630	.100	3.559	.000	.661	.162	.071	.502	1.990
	JOBTIME Months since Hire	[illegible]		[illegible]		[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible] Hire»
	PREVEXP Experience (months)	-19[illegible]436 Previous	3.583	-.119	-5[illegible]097		.000	-.24	-.108	.827	1.210
	SALBEGIN Beginning Salary	1.723	.061	.794	28.472	.000	.880	.796	.568	.512	1.953 Beginn Salary
	SEX Gender	2232.917	792.078	.065	2.819	.005	.450	.129	.056	.744	1.343

a. Dependent Variable: SALARY Current Salary

ตารางที่ 8 แสดงค่าสถิติการทดสอบ ความเป็น Normality

Descriptive Statistics

	N	Mean	Std. Deviation	Skewness		Kurtosis	
	Statistic	Statistic	Statistic	Statistic	Std. Error	Statistic	Std. Error
RES_1 Unstandardized Residual	474	1.353584E-12	7371.1857599	1.723	.112	8.520	.224
Valid N (listwise)	474						

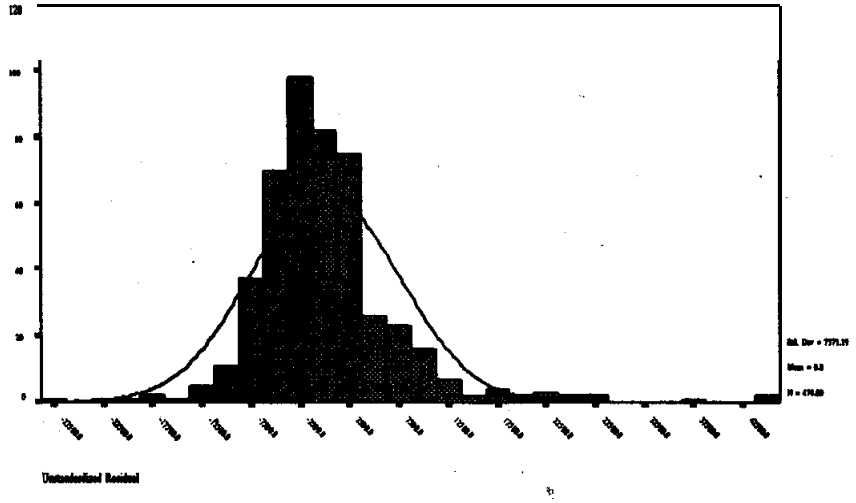
One-Sample Kolmogorov-Smirnov Test

		RES_1 Unstandardized Residual
N		474
Normal Parameters [a,b]	Mean	4.639605E-06
	Std. Deviation	7371.1855469
Most Extreme Differences	Absolute	.127
	Positive	.127
	Negative	-.091
Kolmogorov-Smirnov Z		2.767
Asymp. Sig. (2-tailed)		.000

a. Test distribution is Normal.

b. Calculated from data.

รูปที่ 1 แสดงการกระจายตัวของค่าความคลาดเคลื่อน



ตารางที่ 9 **แสดงการตรวจสอบปัญหา** Heteroscedasticity **ด้วย** Goldfeld and Quandt Test

ANOVA[b]

Mod		Sum of Squares	df	Mean Square	F	Sig.
1	Regression	1416426063.954	5	283285212.791	27.285	.000[a]
	Residual	1785790115.822	172	10382500.673		
	Total	3202216179.775	177			

a. Predictors: (Constant), SEX Gender, EDUC Educational Level (years). PREVEXP Previous Experience (months). JOBTIME Months since Hire, SALBEGIN Beginning Salary

b. Dependent Variable: SALARY Current Salary

Coefficients[a]

Model		Unstandardized Coefficients		Standardized Coefficients		
		B	Std. Error	Beta	t	Sig.
1	(Constant)	-2548.11	3518.10		-.724	.470
	EDUC Educational Level (years)	339.546	123.403	.168	2.752	.007
	JOBTIME Months since Hire	70.088	26.750	,162	2.620	.010
	PREVEXP Previous Experience (months)	-11.122	2.333	-.277	-4.766	.000
	SALBEGIN Beginning Salary	1.458	.213	.433	6.841	.000
	SEX Gender	2340.743	798.996	.185	2.930	.004

a. Dependent Variable: SALARY Current Salary

ตารางที่ 9 (ต่อ)แสดงการตรวจสอบปัญหา Heteroscedasticity **ด้วย** Goldfeld and **Quandt** Test

ANOVA[b]

Model		Sum of Squares	df	Mean Square	F	Sig.
1	Regression	57206216793.554	5	11441243358.711	118.505	.000[a]
	Residual	16605999863.889	172	96546510.837		
	Total	73812216657.444	177			

a. Predictors: (Constant), SEX Gender, PREVEXP Previous Experience (months). JOBTIME Months since Hire, SALBEGIN Beginning Salary, EDUC Educational Level (years)

b. Dependent Variable: SALARY Current Salary

Coefficients[a]

Model		Unstandardized Coefficients		Standardized Coefficients		
		B	Std. Error	Beta	t	Sig.
1	(Constant)	-19071.98	7217.318		-2.643	.009
	EDUC Educational Level (years)	687.509	407.422	.086	1.687	.093
	JOBTIME Months since Hire	249.358	72.081	.129	3.459	.001
	PREVEXP Previous Experience (months)	-53.104	10.504	-.208	-5.056	.000
	SALBEGIN Beginning Salary	1.682	.104	.779	16.122	.000
	SEX Gender	1907.742	1971.326	,037	.968	.335

a. Dependent Variable: SALARY Current Salary

ตารางที่ 10 แสดงการตรวจสอบปัญหา Autocorrelation ด้วยวิธี AR(1)

Variables Entered/Removed [b]

Model	Variables Entered	Variables Removed	Method
1	RES_2 [a]	.	Enter

a. All requested variables entered

b Dependent Variable: RES_1 Unstandardized Residual

Model Summary

Model	R	R Square	Adjusted R Square	Std. Error of the Estimate
1	.038[a]	.001	-.001	7380.9820968

a Predictors (Constant). RES_2

ANOVA [b]

Model		Sum of Squares	df	Mean Square	F	Sig.
	Regression	37376877.853	1	37316811.853	686	.408[a]
	Residual	25659560352.268	471	54418896.714		
	Total	25696937230.121	472			

a Predictors (Constant). RES_2

b Dependent Variable RES_1Unstandardized Residual

Coefficients [a]

Model		Unstandardized Coefficients		Standardized Coefficients		
		B	Std. Error	Beta	t	Sig.
1	(Constant)	-3.554	339.378		-.010	.992
	RES_2	3.814E-02	.046	.038	.828	.408

a. Dependent Variable: RES_1 Unstandardized Residual

ตารางที่ 11 แสดงการตรวจสอบอำนาจการพยากรณ์

Variables Entered/Removed [b]

Model	Variables Entered	Variables Removed	Method
1	PRE_1 Unstandardized Predicted Value [a]	.	Enter

a. All requested variables entered

b. Dependent Variable: SALARY Current Salary

Model Summary [b]

Model	R	R Square	Adjusted R Square	Std. Error of the Estimate	Durbin-Watson
1	.902[a]	.814	.813	$7378.99	1.923

a. Predictors: (Constant), PRE_1 Unstandardized Predicted Value

b. Dependent Variable: SALARY Current Salary

ANOVA [b]

Model		Sum of Squares	df	Mean Square	F	Sig.
1	Regression	112216333929.663	1	112216333929.663	2060.925	.000[a]
	Residual	25700161506.677	472	54449494.718		
	Total	137916495436.346	473			

a Predictors: (Constant), PRE_1 Unstandardized Predicted Value

b. Dependent Variable: SALARY Current Salary

Coefficients [a]

Model		Unstandardized Coefficients		Standardized Coefficients		
		B	Std. Error	Beta	t	Sig.
1	(Constant)	2.638E-11	830.491		.000	1.000
	PRE_1 Unstandardized Predicted Value	1.000	.022	.902	45.397	.000

a. Dependent Variable: SALARY Current Salary

7.6 การแปลผล

จากตัวอย่างเป็นการสร้างแบบจำลองเพื่อวิเคราะห์ความสัมพันธ์ของระดับรายได้ในปัจจุบันว่ามีสาเหตุมาจากตัวแปรอะไรบ้าง ผลจากการวิเคราะห์ข้อมูลดังปรากฏ ตามตารางต่าง ๆ ข้างต้นสามารถแปลผลได้ดังต่อไปนี้

ตารางที่ 1 แสดงตัวอย่างของข้อมูลที่ใช้ในการหาความสัมพันธ์ของตัวแปร ที่เรากำลังสนใจประกอบด้วย ตัวแปรตามคือ ตัวแปรรายได้ปัจจุบัน (Current Salary) SALARY ตัวแปรอิสระประกอบด้วยตัวแปรจำนวน 5 ตัวได้แก่ ตัวแปรเพศ (Gender) SEX ตัวแปรระดับการศึกษา (Education) EDUC ตัวแปรระดับรายได้เริ่มต้น (Beginning Salary) SALBEGIN อายุการทำงาน (Jobtime months since hire) JOBTIME และ ตัวแปรประสบการณ์การทำงาน (Previous Experience) PREVEXP

ตารางที่ 2 แสดงสถิติพรรณนา คือ ค่า Minimum Maximum Sum Mean และ Standard Deviation ของตัวแปรทุกตัว ซึ่งสามารถนำค่าต่าง ๆ เหล่านี้ไปใช้พรรณนา ให้ทราบถึงลักษณะของข้อมูลได้ จากตัวอย่างพบว่าโดยเฉลี่ยแล้วกลุ่มตัวอย่างมีรายได้เฉลี่ยคนละ $34,419.57 โดยผู้ที่มีรายได้ต่ำสุดมีรายได้เดือนละ $15,750 และผู้ที่มีรายได้สูงสุดมีรายได้เดือนละ $135,000 สำหรับสัดส่วนระหว่างเพศชาย และ เพศหญิงใกล้เคียงกัน มีระดับการศึกษาโดยเฉลี่ยอยู่ระดับสูงกว่ามัธยมศึกษาตอนปลาย ผู้ที่มีระดับการศึกษาต่ำสุดของกลุ่มตัวอย่างนี้มีการศึกษาระดับมัธยมศึกษาปีที่ 2 และผู้ที่มีการศึกษาสูงสุดมีระดับการศึกษาอยู่ในระดับจบปริญญาเอก กลุ่มตัวอย่างนี้มีประสบการณ์ในการทำงานมาแล้วเฉลี่ยไม่น้อยกว่าคนละ 95 เดือน หรือประมาณ 8 ปี และทำงานปัจจุบันมาแล้วโดยเฉลี่ยไม่น้อยกว่าคนละ 81 เดือน

การตรวจสอบปัญหา multicollinearity ด้วยการพิจารณาค่าสหสัมพันธ์ (collinearity) จากตารางที่ 3 แสดงค่าสหสัมพันธ์ระหว่างตัวแปรต่าง ๆ ถ้าคู่ของตัวแปรเป็นตัวแปรตามกับตัวแปรอิสระ มีความสัมพันธ์กันมากก็แสดงว่าตัวแปรอิสระตัวนั้น น่าจะอธิบายตัวแปรตามได้ดี แต่ถ้าคู่ของตัวแปรเป็นตัวแปรอิสระด้วยกันและมีความสัมพันธ์กันมาก(มากกว่า 0.7) ก็ให้ดูค่า Sig.< 0.05 แสดงว่าเราสามารถปฏิเสธ H_0 ที่เราตั้งว่า $H_0: \rho = 0$ และ $H_1: \rho \neq 0$ นั่นคือค่าตัวแปรคู่นั้นมีความสัมพันธ์กันจริง และน่าจะเกิดปัญหา multicollinearity ขึ้นได้ (ต้องหาวิธีทดสอบวิธีอื่นเพื่อยืนยันต่อไป) แต่ถ้า ค่า Sig. > 0.05 ก็แสดงว่าเราไม่สามารถปฏิเสธ $H_0: \rho = 0$ ได้ แสดงว่าค่าตัวแปรคู่นั้นไม่น่าจะมีความสัมพันธ์กันในระดับที่มองเห็น(ต้องทดสอบความสัมพันธ์ด้วยวิธีอื่นต่อไป)

จากตารางข้างต้น พบว่าคู่ของตัวแปรตามกับตัวแปรอิสระทุกตัวมีความสัมพันธ์กันมากบ้าง น้อยบ้าง และค่า Sig. (1-tailed) มีค่า Sig. < 0.05 ทุกตัว แสดงว่า ตัวแปรตามน่าจะอธิบายด้วยตัวแปรอิสระได้ ทุกตัว โดยที่ตัวแปรใดจะสามารถอธิบายได้มากน้อยกว่ากัน จะพิจารณาจากค่า Standardized Coefficients Beta ในตารางที่ 7 ต่อไป สำหรับค่าความสัมพันธ์ ระหว่างตัวแปรอิสระด้วยกันจากการพิจารณา พบว่ามีค่าไม่สูงกว่า 0.7 ทุกค่า แสดงว่าไม่น่าจะเกิดปัญหา multicollinearity ขึ้นถึงขั้นยอมรับไม่ได้ โดยเฉพาะค่า Sig. (1-tailed) บางค่านั้น sig. > 0.05 แสดงว่าไม่สามารถปฏิเสธ H_0 ได้ นั่นคือความสัมพันธ์ของตัวแปรคู่นั้นไม่น่าจะมีมาก เช่นค่าความสัมพันธ์ระหว่างตัวแปร Gender กับตัวแปร Jobtime month since hire มีค่าเท่ากับ0.066 และค่า Sig. = 0.074 ซึ่งมากกว่า 0.05 แสดงว่าไม่สามารถปฏิเสธ H_0 ได้ จึงสรุปว่าตัวแปร Gender กับตัวแปร Jobtime months Since hire ไม่น่าจะมีความสัมพันธ์กันเลย

ตารางที่ 4 เป็นตารางแสดงให้เห็นว่าตัวแปรตาม คือ SALARY (Current Salary) และตัวแปรอิสระที่ใช้ในสมการได้แก่ตัวแปร SEX (Gender), EDUE (Education Level (year)) , SALBEGIN (Beginning Salary) JOBTIME (Months since hire) PREVEXP(Previous Experience) โดยใช้วิธีวิเคราะห์สมการถดถอยพหุคูณ ด้วยวิธี Enter คือใส่ตัวแปรอิสระและตัวแปรตามเข้าไปคำนวณพร้อมกัน

ตารางที่ 5 จะแสดงค่าการวิเคราะห์ความแปรปรวนของการถดถอยว่าในแบบจำลองที่เราสร้างขึ้นนั้น ตัวแปรอิสระต่าง ๆ สามารถ อธิบายตัวแปรตามได้หรือไม่ โดยพิจารณาจากค่า Sig. ถ้า Sig.<0.05 แสดงว่าสามารถยอมรับแบบจำลองได้ว่า อย่างน้อยมีตัวแปรอิสระ 1 ตัว ที่สามารถอธิบายตัวแปรตามได้ โดยตัวแปรอิสระตัวใดจะสามารถอธิบายตัวแปรตามได้ จะต้องพิจารณาจากตารางที่ 7 เกี่ยวกับค่าสัมประสิทธิ์สหสัมพันธ์ต่อไป จากตัวอย่างนี้ได้ค่าSig. = 0.000 แสดงว่าสามารถใช้แบบจำลองนี้ได้

จากตารางที่ 6 แสดงค่า R ต่าง ๆ โดยค่า R = 0.902 หมายความว่าตัวแปรตาม SALARY มีความสัมพันธ์กับตัวแปรอิสระทั้ง 5 ตัว สามารถควบคุมความผันแปรของตัวแปรตามได้ 81.4% และที่เหลืออีก 18.6% ถูกควบคุมด้วยตัวแปรตัวอื่นที่ไม่ได้นำมารวมไว้ในแบบจำลอง สำหรับกรณีที่มีแบบจำลองสำหรับอธิบายตัวแปรตามหลายแบบจำลองและต้องการเปรียบเทียบว่า แบบจำลองใดให้อำนาจการอธิบายได้ดีกว่ากันต้องใช้ค่า Adjust R Square ($\overline{R^2}$) เป็นตัวเปรียบเทียบ โดยค่า Adjust R Square สามารถคำนวณได้จากสูตร ดังนี้

$$\bar{R}^2 = 1-(1-R^2)(\frac{n-1}{n-k-1})$$

โดยค่า n = จำนวนตัวอย่างทั้งหมด

k = จำนวนตัวแปรอิสระทั้งหมด

ดังนั้น

$$\bar{R}^2 = 1-(1-0.814)(\frac{474-1}{474-5-1})$$

$$\bar{R}^2 = 0.812$$

สำหรับค่า durbin-watson เป็นค่าสถิติที่ใช้สำหรับทดสอบว่ามีปัญหา autocorrelation หรือไม่คือทดสอบดูว่าค่า error term (u_t กับ u_{t-1}) ว่ามีความสัมพันธ์กันหรือไม่ มีหลักเกณฑ์การพิจารณาง่าย ๆ ว่า ถ้าค่า d $\cong$ 2 จะถือว่าไม่เกิดปัญหา autocorrelation ขึ้น โดยค่า d หาได้จากสูตร

$$d = \frac{\sum (u_t - u_{t-1})^2}{\sum u_t^2}$$

จากตัวอย่างจะได้ค่า d = 1.923 ซึ่งมีค่าใกล้เคียง 2 เราจึงสรุปว่าข้อมูลชุดนี้ไม่มีปัญหา autocorrelation

ตารางที่ 7 แสดงค่าสัมประสิทธิ์สหสัมพันธ์ โดยค่า unstandardized coefficient (B) จะแสดงถึงค่าสัมประสิทธิ์การถดถอย และค่า standard error ของสัมประสิทธิ์ การถดถอย ถ้ามีค่าต่ำแสดงว่าค่าB สามารถ ประมาณค่า B ได้แม่นยำ และค่า t เช่น t =2.819 เกิดจากค่า B หารด้วย Standard error เมื่อนำค่า t ที่คำนวณได้ไปเปรียบเทียบค่า t ที่เปิดจากตาราง ถ้าพบว่า t ที่คำนวณได้มีค่า มากกว่า t ตาราง ก็จะยอมรับว่า ตัวแปรอิสระตัวนั้นสามารถอธิบายตัวแปรตามได้โดยพิจารณาค่า Sig. ถ้า Sig. < 0.05 แสดงว่าตัวแปรอิสระตัวนั้นสามารถอธิบายตัวแปรตามได้ด้วยระดับความเชื่อมั่นมากกว่า 95 %

สำหรับค่า Beta (standardized coefficient) คือสัมประสิทธิ์ความถดถอยที่ปรับค่าเป็นค่า Standard Normal แล้วค่า Beta จึงอยู่บนพื้นฐานที่ตัวแปรทุกตัววัดค่าในสเกลเดียวกันแล้ว สามารถเปรียบเทียบความสำคัญของตัวแปรอิสระว่าตัวใดมีอิทธิพลต่อตัวแปรตามมากกว่ากัน จากตัวอย่างพิจารณาค่า Beta ของตัวแปรอิสระต่าง ๆ พบว่า ตัวแปร SALBEGIN = 0.794 รองลงมาได้

แก่ ตัวแปร PREVEXP = -0.119 ตัวแปร EDUE =0.100 ตัวแปร JOBTIME = 0.091 และตัวแปร SEX = 0.065 หมายความว่าตัวแปร SALBEGIN มีอิทธิพล ต่อตัวแปรตาม SALARY มากที่สุด รองลงมาได้แก่ตัวแปร PREVEXP EDUE JOBTIME และ SEX ตามลำดับ

Colinearity Statistics ประกอบด้วยค่าสถิติ 2 ค่าคือ Tolerance กับ VIF โดยตัวสถิติ 2 ตัวนี้จะมีความสัมพันธ์กันคือ ค่า Tolerance = 1 / VIF การพิจารณาจะใช้ค่าตัวใดตัวหนึ่งหรือทั้ง 2 ตัวพิจารณาประกอบกันก็ได้โดยถ้าค่า Tolerance มีค่าระหว่าง 0 กับ 1 ถ้าค่า Tolerance มีค่าใกล้ 0 แสดงว่าตัวแปรอิสระตัวนั้นน่าจะมีความสัมพันธ์กับตัวแปรอิสระตัวอื่นที่มีค่าใกล้ 0 เช่นกันแสดงว่าเกิดปัญหา multicollinearity ขึ้นแล้วแต่ถ้า ค่า Tolerance มีค่าใกล้ 1 แสดงว่าตัวแปรอิสระตัวนั้นไม่มีความสัมพันธ์กับตัวแปรอิสระตัวอื่น ๆ และสำหรับค่าสถิติ VIF คือค่าผกผันของค่า Tolerance จะมีค่ามากกว่า 1 ขึ้นไปถ้ามากกว่า 1 มาก ๆแสดงว่า ตัวแปรอิสระตัวนั้นน่าจะมีความสัมพันธ์กับตัวแปรอิสระตัวอื่นมากก็จะเกิดปัญหา Multicolinearity ขึ้นได้ ต้องทำการแก้ไขต่อไปสำหรับในกรณีตัวอย่าง พิจารณาค่า Tolerance พบว่ามีค่ามากกว่า 0.5 ทุกตัวแสดงว่าไม่เกิดปัญหา Multicolinearity และถ้าพิจารณาจากค่า VIF พบว่าตัวแปรอิสระทุกตัวมีค่าไม่เกิน 2 แสดงว่าตัวแปรอิสระทุกตัวจะไม่มีความสัมพันธ์กันเอง

จากตารางที่ 4-7 สามารถนำมาเสนอสมการถดถอยพร้อมค่าสถิติต่าง ๆ ที่จำเป็นได้ดังนี้

SALARY = -14782.95+2232.917 SEX+593.031 EDUC+1.723 SALBEGIN

(-4.524) (2.819) (3.559) (28.472)

-19.436 PREVEXP+154.536 JOBTIME

(-5.424) (4.534)

R^2 = 0.814

standard error of the estimate = 7410.46

durbin – watson = 1.923

จากสมการถดถอยที่ได้ สามารถอ่านค่าได้ว่าระดับรายได้ในปัจจุบัน จะเปลี่ยนแปลงมากน้อยเพียงไร ขึ้นอยู่กับปัจจัยภายนอกหรือตัวแปรอิสระต่าง ๆ ที่เราสนใจ ได้แก่ ตัวแปรเกี่ยวกับ เพศ พบว่าเพศชายจะมีโอกาสได้รับเงินเดือนหรือรายได้สูงกาว่าเพศหญิงจำนวน $2,232.917 สำหรับตัวแปรทางด้านการศึกษาจากแบบจำลองที่ได้ หรือสมการถดถอยพบว่าผู้ที่มีการศึกษามากกว่าจะได้รับเงินเดือนสูงกว่าในอัตรา $593.031 ต่อระยะเวลาที่เรียนต่างกัน 1 ปี สำหรับตัวแปรระดับรายได้ เริ่มต้นที่ได้รับ จะมีส่วนกำหนด ระดับเงินเดือนปัจจุบันในอัตรา 1.723 เท่าของที่ได้รับเริ่มต้น

สำหรับตัวแปรประสบการณ์ในการทำงานนั้น จากการวิเคราะห์พบว่ามีความสัมพันธ์ทางด้านตรงข้ามกับระดับรายได้ในปัจจุบัน คือ ยิ่งมีประสบการณ์ในการทำงานมาก่อน ที่จะมาทำงานปัจจุบันมากเท่าใดกลับมีผลทำให้รายได้ในปัจจุบันลดลงไป ซึ่งตามทฤษฎีหรือหลักความเป็นจริงแล้วไม่น่าจะเป็นไปได้จึงควรจะได้มีการตรวจสอบว่าเกิดจากสาเหตุอะไร ส่วนระยะเวลาในการทำงานเป็นอีกตัวแปรหนึ่งที่มีผลในทางที่ทำให้ระดับรายได้สูงขึ้นเมื่อมีอายุการทำงานมากขึ้นซึ่งเป็นไปตามทฤษฎีและหลักความเป็นจริง

ตารางที่ 8 แสดงการตรวจสอบค่าNormality คือการตรวสอบคุณสมบัติของค่าคลาดเคลื่อนว่ามีการแจกแจงแบบปกติหรือไม่ โดยพิจารณาจากค่าสถิติ Kolmogorov Smirnov ด้วยการพิจารณาค่า Sig.(2-tailed) ถ้า Sig. < 0.05 แสดงว่าการแจกแจงของค่าความคลาดเคลื่อนจะมีลักษณะเป็นแบบปกติหรืออาจจะนำค่าความคลาดเคลื่อนสะสมมาพล็อตกราฟ แล้วพิจารณาจากกราฟว่ามีลักษณะเป็นรูประฆังคว่ำหรือไม่ ถ้าเหมือนก็แสดงว่าค่าความคลาดเคลื่อนมีการแจกแจงแบบปกติ จากตัวอย่าง พิจารณาตารางที่ 8 พบว่าค่า Kolmogorov Smirnov มีค่า Sig. = 0.000 แสดงว่าค่าความคลาดเคลื่อนจะมีการแจกแจงแบบปกติ แต่เมื่อพิจารณาจากกราฟความสัมพันธ์ดังรูปที่ 1 พบว่าการแจกแจงของค่าคลาดเคลื่อนมีลักษณะโด่งกว่าปกติ และมีลักษณะเบ้ขวาเล็กน้อย ซึ่งตรงกับค่า Kurtosis ที่มีค่าเท่ากับ 8.52 ซึ่งมากกว่าค่าที่ควรเป็นคือ 3 และค่า Skewness เท่ากับ 1.723 ซึ่งมากกว่า ค่าที่ควรจะเป็นคือ 0 (ศูนย์)

การตรวจสอบปัญหา Heteroscedasticity ด้วยวิธี Goldfeld-Quandt ด้วยการตั้งสมมติฐานหลักว่า $H_0 : \delta_1^2 = \delta_2^2$ แล้วทำการคำนวณหาค่า F^* จากสูตร $F^* = SSR_2 / SSR_1$ โดยค่า SSR_2และ SSR_1 นำมาจากตารางที่ 9 ได้ $F^* = 16,605,999,863.889 / 1,785,790,115.822 = 9.299$ นำค่า F^* ที่คำนวณได้นี้ไปเปรียบเทียบกับค่า F ที่เปิดจากตารางที่ degree of freedom เท่ากับ (172,172) ด้วยความเชื่อมั่นที่ 95 % หรือที่ $\alpha = 0.05$ ได้ค่า F ตาราง = 1.26 พบว่า F คำนวณมีค่ามากกว่า F ตาราง แสดงว่าเราจะปฏิเสธ $H_0 : \delta_1^2 = \delta_2^2$หมายความว่าข้อมูลชุดนี้เกิดปัญหา Heteroscedasticity ขึ้นแล้ว

การตรวจสอบปัญหา Autocorrelation ในการตรวจสอบปัญหา Autocorrelation นั้นได้แสดงให้เห็นว่าสามารถพิจารณาได้จากค่า durbin-watson ถ้ามีค่าใกล้ 2 จะไม่เกิดปัญหาดังที่ได้แสดงให้เห็นแล้ว สำหรับตารางที่ 10 เป็นการตรวจสอบปัญหา Autocorrelation ด้วยวิธี AR(1) คือนำค่าความคลาดเคลื่อนที่ได้ ย้อนเวลาไป 1 ช่วงเวลาแล้วทำการหาความถดถอยจากสมการ

$$u_t = \rho u_{t-1} + v_t$$

แล้วทำการตรวจสอบค่า ρ ที่ได้ถ้าไม่สามารถปฏิเสธ $H_0 : \rho = 0$ ได้แสดงว่าไม่เกิดปัญหา Autocorrelation จากตัวอย่างค่าความคลาดเคลื่อน u_t คือ RES_1 และค่า u_{t-1} คือค่า RES_2 เมื่อพิจารณาค่า Sig. ของ RES_2 พบว่า Sig. = 0.408 ซึ่งมากกว่า 0.05 แสดงว่าไม่สามารถปฏิเสธ H_0 ได้ ดังนั้นจึงสรุปได้ว่ากรณีนี้ไม่มีปัญหา Autocorrelation เหมือนกับที่ได้สรุปมาแล้วก่อนหน้านี้

การตรวจสอบอำนาจการพยากรณ์ จากตารางที่ 11 ให้ตัวแปร PRE_1 คือค่าที่ได้จากการคำนวณจากสมการถดถอย ให้เป็นตัวแปรอิสระ และตัวแปร SALARY เป็นตัวแปรตามมาทำการวิเคราะห์การถดถอย พบว่าได้สมการถดถอยคือ

$$\text{SALARY} = 2.638 \times 10^{-11} + \text{PRE\_1}$$
$$(0.000) \qquad (45.397)$$

R^2 = 0.902

standard error of the estimate = 7,378.99

durbin – watson = 1.923

จากสมการพบว่า $a = 2.638 \times 10^{-11}$ มีค่าเข้าใกล้ 0 และค่า $b = 1$ แสดงให้เห็นว่าค่า PRE_1 ที่ได้ จากการคำนวณของสมการถดถอย SALARY สามารถใช้เป็นตัวแทนของ SALARY ได้คิดเป็นร้อยละ 90.20 ด้วยความเชื่อมั่นมากกว่า 95 %

สรุปว่าสมการ SALARY = -14,782.95 + 2,232.917SEX + 593.031EDUC + 1.723SALBEGIN – 19.436 PREVEXP + 154.536 JOBTIME

สามารถใช้เป็นสมการในการพยากรณ์ตัวแปรตามคือรายได้เมื่อทราบค่าตัวแปรอิสระทุกตัวซึ่งได้แก่ตัวแปร SEX, EDUC, SALBEGIN, PREVEXP และตัวแปร JOBTIME แต่ถ้าต้องการความสมบูรณ์ถูกต้องจริง ๆ แล้วต้องทำการแก้ปัญหา Heteroscedasticity เสียก่อน

คำถามท้ายบทที่ 7

1. จงอธิบายความหมาย ของแบบจำลอง สมการถดถอยพหุคูณในกรณีที่เกิดปัญหาต่อไปนี้
   ก. ความไม่คงที่ในความแปรปรวนของตัวคลาดเคลื่อน
   ข. สหสัมพันธ์ระหว่างความแปรปรวนของตัวคลาดเคลื่อน
   ค. สหสัมพันธ์ระหว่างตัวแปรอิสระ
2. ถ้าเกิดปัญหาดังกล่าวในข้อ 1 จะมีวิธีทดสอบปัญหาอย่างไร
3. เมื่อทดสอบพบปัญหาดังกล่าวในข้อ 1 จะมีวิธีแก้ปัญหาอย่างไร
4. จงอธิบายวิธีการทดสอบอำนาจการพยากรณ์