

**Bài 1:** Các yếu tố ảnh hưởng tới mức độ lo lắng của giám đốc

Nguồn gốc của sự biến thiên	Tổng bình phương	Bậc tự do
Hồi quy trên X1	981.326	1
Hồi quy trên X2 X1	190.232	1
Hồi quy trên X3 X1, X2	129.431	1
Sai số	442.292	18
Tổng quát	1743.281	21

1. Tổng bình phương hồi quy trên các biến X1, X2 và X3 là:

$$\begin{aligned} SSR &= SSR_{X1} + SSR_{X2|X1} + SSR_{X3|X1,X2} \\ &= 981.326 + 190.232 + 129.431 \\ &= 1300.989 \end{aligned}$$

2. Tỷ lệ phần trăm sự biến thiên của mức độ lo lắng được giải thích bởi các biến độc lập:

$$R^2 = \frac{SSR}{SST} = \frac{1300.989}{1743.281} = 74.63\%$$

Vậy có 74.63% sự biến thiên trong mức độ lo lắng của các giám đốc được giải thích bởi ba biến độc lập X1, X2 và X3: áp lực công việc, kỹ năng quản lý, mức độ hài lòng với chức vụ của mình.

3. Ta không có đủ thông tin để có thể kết luận rằng: tất cả ba biến giải thích đều có ảnh hưởng đáng kể tới mức độ lo lắng.

Tuy nhiên ta có thể khẳng định rằng ít nhất một trong ba biến X có ảnh hưởng đáng kể tới Y- mức độ lo lắng, qua kiểm định F:

$$H_0: \beta_1 = \beta_2 = \beta_3 = 0 \text{ cho mô hình } Y = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \beta_3 X_3 + \epsilon$$

Với giả thuyết đối:

$$H_1: \text{Ít nhất một } \beta_j \neq 0 \text{ (với } j = 1, 2, 3)$$

$$\begin{aligned} F_{obs} &= \frac{SSR/(p-1)}{SSE/(n-p)} = \frac{(SSR_{X1} + SSR_{X2|X1} + SSR_{X3|X1,X2})/(p-1)}{SSE/(n-p)} \\ &= \frac{1300.989/(4-1)}{442.292/(22-4)} = 17.649 > F_{3,18}^{\alpha=0.05} = 3.1599 \end{aligned}$$

Do Fobs lớn hơn F value nên ta bác bỏ Ho với mức ý nghĩa 5%.

Kết luận: có ít nhất một tham số beta khác 0, do đó mô hình hồi quy tồn tại, tức là trong các biến độc lập có ít nhất một biến ảnh hưởng đáng kể tới biến phụ thuộc.

4. Nếu chỉ xét biến  $X_1$ , ta có bảng ANOVA sau:

Biến thiên	SS	df	MS	
Reg	$SSR_{X_1} = 981.326$	1	981.326	$F = \frac{MSR}{MSE}$ $= 25.758$
E	$SST - SSR_{X_1}$ $= 1743.281 - 981.326$ $= 761.955$	$n-2 = 20$	38.098	
T	$SST = 1743.281$	$n-1 = 21$		

5. Kiểm định giả thuyết với mức ý nghĩa 5% cho các mô hình:

a.  $H_0: \beta_1 = 0$  cho mô hình  $Y = \beta_0 + \beta_1 X_1 + \epsilon$

Với giả thuyết đối:  $H_1: \beta_1 \neq 0$

Từ bảng ANOVA cho một biến  $X_1$  trên, ta có:

$$F_{obs} = \frac{MSR}{MSE} = 25.758 > F_{1,20}^{\alpha=0.05} = 4.3512$$

Do đó ta bác bỏ giả thuyết  $H_0$  với mức ý nghĩa 5%, nghĩa là  $X_1$ : áp lực công việc có ảnh hưởng tới mức độ lo lắng, và có tồn tại trong mô hình hồi quy đơn.

b.  $H_0: \beta_2 = 0$  cho mô hình  $Y = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \epsilon$

Với giả thuyết đối:  $H_1: \beta_2 \neq 0$

Ta sử dụng Partial F-test với số biến bỏ đi là  $r = 1$

$$F_{obs} = \frac{[SSE(H_0) - SSE(H_1)]/r}{[SSE(H_1)]/(n-p)}$$

Trong đó:

$$SSE(H_0) = SSE + SSR_{X_2|X_1} + SSR_{X_3|X_1, X_2} = 761.955$$

$$SSE(H_1) = SSE + SSR_{X_3|X_1, X_2} = 571.723$$

$$\text{Suy ra } F_{obs} = 6.322 > F_{1,19}^{\alpha=0.05} = 4.3807$$

Vậy ta bác bỏ giả thuyết  $H_0$  với mức ý nghĩa 5%, nghĩa là biến  $X_2$ : kỹ năng quản lý có ảnh hưởng tới mức độ lo lắng, khi áp lực công việc đã có tồn tại trong mô hình.

c.  $H_0: \beta_3 = 0$  cho mô hình  $Y = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \beta_3 X_3 + \epsilon$

Với giả thuyết đối:  $H_1: \beta_3 \neq 0$

Ta sử dụng Partial F-test với số biến bỏ đi là  $r = 1$

$$F_{obs} = \frac{[SSE(H_0) - SSE(H_1)]/r}{[SSE(H_1)]/(n-p)}$$

Trong đó:

$$SSE(H_0) = SSE + SSR_{X_3|X_1, X_2}$$

$$SSE(H_1) = SSE$$

$$\Rightarrow F_{obs} = 5.2675 > F_{1,18}^{\alpha=0.05} = 4.4139$$

Vậy ta bác bỏ giả thuyết  $H_0$  với mức ý nghĩa 5%, nghĩa là biến  $X_3$ : mức độ hài lòng với chức vụ của mình có ảnh hưởng tới mức độ lo lắng, khi áp lực công việc và kỹ năng quản lý đã tồn tại trong mô hình.

6. Hệ số xác định cho 5a, b, c:

5a: mô hình  $Y = \beta_0 + \beta_1 X_1 + \epsilon$

$$R^2 = \frac{SSR_{X_1}}{SST} = \frac{981.326}{1743.281} = 56.29\%$$

5b: mô hình  $Y = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \epsilon$

$$R^2 = \frac{SSR_{X_1} + SSR_{X_2|X_1}}{SST} = \frac{981.326 + 190.232}{1743.281} = 67.20\%$$

5c: mô hình  $Y = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \beta_3 X_3 + \epsilon$

$$\begin{aligned} R^2 &= \frac{SSR_{X_1} + SSR_{X_2|X_1} + SSR_{X_3|X_1, X_2}}{SST} \\ &= \frac{981.326 + 190.232 + 129.431}{1743.281} = 74.63\% \end{aligned}$$

7. Do các mô hình trên có số biến độc lập khác nhau nên ta xét thêm hệ số xác định hiệu chỉnh: (adjusted R squared) để đưa ra lựa chọn mô hình.

5a: mô hình  $Y = \beta_0 + \beta_1 X_1 + \epsilon$

$$\begin{aligned} Adj R^2 &= 1 - \frac{SSE/(n-p)}{SST/(n-1)} = 1 - \frac{(SST - SSR_{X_1})/(n-p)}{SST/(n-1)} \\ &= 1 - \frac{(1743.281 - 981.326)/(22-2)}{1743.281/(22-1)} = 54.11\% \end{aligned}$$

5b: mô hình  $Y = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \epsilon$

$$\begin{aligned} Adj R^2 &= 1 - \frac{SSE/(n-p)}{SST/(n-1)} = 1 - \frac{(SST - SSR_{X_1} - SSR_{X_2|X_1})/(n-p)}{SST/(n-1)} \\ &= 1 - \frac{(1743.281 - 981.326 - 190.232)/(22-3)}{1743.281/(22-1)} = 63.75\% \end{aligned}$$

5c: mô hình  $Y = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \beta_3 X_3 + \epsilon$

$$\begin{aligned} Adj R^2 &= 1 - \frac{SSE/(n-p)}{SST/(n-1)} \\ &= 1 - \frac{442.292/(22-4)}{1743.281/(22-1)} = 70.40\% \end{aligned}$$

Qua đó ta có bảng tóm tắt sau:

Mô hình	R squared	Adjusted R squared
mô hình $Y = \beta_0 + \beta_1 X_1 + \epsilon$	56.29%	54.11%
mô hình $Y = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \epsilon$	67.20%	63.75%
mô hình $Y = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \beta_3 X_3 + \epsilon$	74.63%	70.40%

Dựa vào các kết quả về hệ số xác định cho 3 mô hình, đặc biệt là hệ số xác định hiệu chỉnh (rất hiệu quả khi so sánh các mô hình có số biến độc lập khác nhau), ta có thể chọn mô hình có hệ số xác định hiệu chỉnh cao nhất là 70.40% cho bài toán, với ba biến  $X_1$ ,  $X_2$ ,  $X_3$  độc lập.

Ngoài ra, các kết quả kiểm định tham số hồi quy  $\beta_1$ ,  $\beta_2$ ,  $\beta_3$  ở câu 5 cũng cho thấy  $X_1$ ,  $X_2$ ,  $X_3$  có ý nghĩa trong mô hình hồi quy.

Cuối cùng, do số lượng mẫu khá ít:  $n = 22$  nên cần phải thu thập nhiều dữ liệu hơn để có thể đưa ra được lựa chọn chính xác hơn nữa về mô hình.