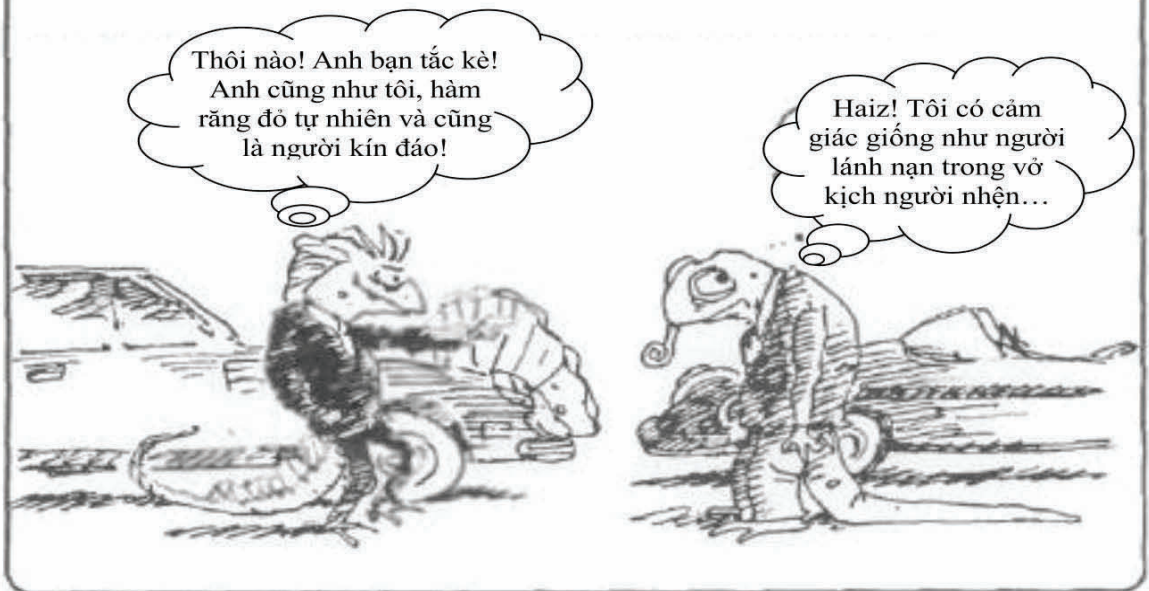


CHƯƠNG 9: SO SÁNH HAI TỔNG THỂ

(Tiếp theo)

Làm thế nào để so sánh được TRUNG BÌNH của các MẪU NHỎ?

Bạn có nhớ ví dụ về hãng xe Chameleon (con Tắc kè) không? Đối thủ của họ là hãng xe Iguana (hãng con Cự đà) cho rằng phần xốp ở mũi xe giúp bảo vệ xe tốt hơn khi bị va chạm ở phía trước, và họ đã thử nghiệm 7 lần để chứng minh điều đó!



Các kết quả của họ được so sánh với hãng xe Chameleon:

Chameleon		Iguana	
1	150\$	1	50\$
2	400\$	2	200\$
3	720\$	3	150\$
4	500\$	4	400\$
5	930\$	5	750\$
n_1	5	6	400\$
\bar{x}_1	540\$	7	150\$
s_1	299\$	n_2	7
		\bar{x}_2	300\$
		s_2	238\$



Phân bố t có thể được sử dụng nếu cả hai tổng thể có dạng hình chuông và chung độ lệch chuẩn $\sigma = \sigma_1 = \sigma_2$ chỉ khác là chúng ta phải tính trung bình các độ lệch chuẩn theo dạng ước lượng đơn của σ :

$$s_{pool}^2 = \frac{(n_1 - 1)s_1^2 + (n_2 - 1)s_2^2}{n_1 + n_2 - 2}$$



Độ lệch chuẩn tương tự như các mẫu lớn, ngoại trừ s_{pool} được thay thế bởi s_1 , và s_2

$$\begin{aligned} SE(\bar{X}_1 - \bar{X}_2) &= \sqrt{\frac{s_{pool}^2}{n_1} + \frac{s_{pool}^2}{n_2}} \\ &= s_{pool} \sqrt{\frac{1}{n_1} + \frac{1}{n_2}} \end{aligned}$$

khoảng tin cậy $(1-\alpha).100\%$ là:

$$\mu_1 - \mu_2 = \bar{x}_1 - \bar{x}_2 \pm t_{\alpha/2} SE(\bar{X}_1 - \bar{X}_2)$$

Trong đó $t_{\alpha/2}$ là giá trị tới hạn của t với bậc tự do là $(n_1 - n_2 - 2)$

Với điều kiện các độ lệch chuẩn xấp xỉ nhau và đồ thị có dạng đối xứng. Hãng xe bò sát tính được:

$$\begin{aligned} s_{pool} &= \sqrt{\frac{4 \times 299^2 + 6 \times 328^2}{10}} = 264 \\ SE(\bar{X}_1 - \bar{X}_2) &= 264 \sqrt{\frac{1}{5} + \frac{1}{7}} = 154 \end{aligned}$$

khoảng tin cậy 95% là:

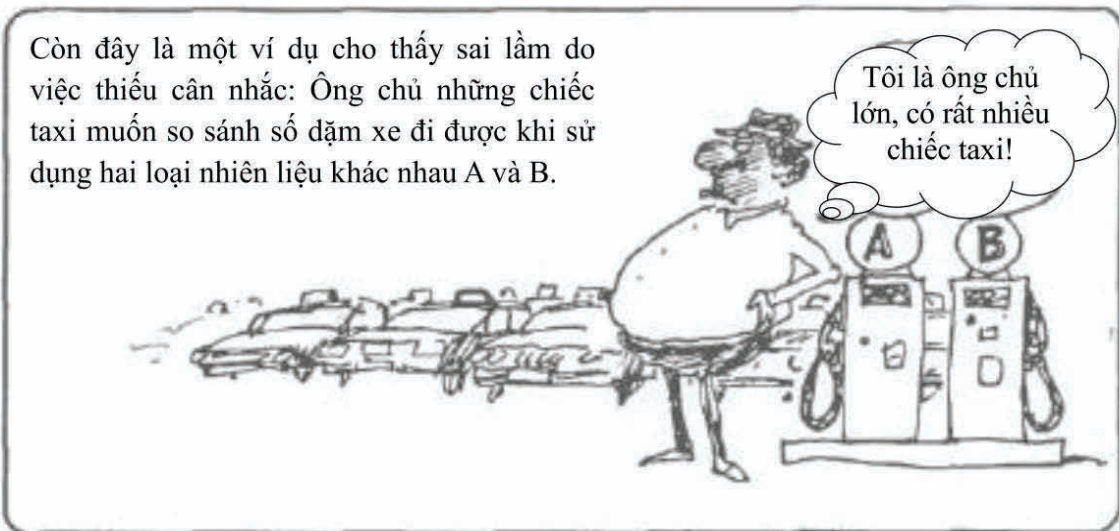
$$\begin{aligned} \mu_1 - \mu_2 &= 540 - 300 \pm t_{0.025}(154) \\ &= 240 \pm (2.23).(154) \\ &= \mathbf{240 \pm 340} \end{aligned}$$

Vì công thức này bao gồm cả giá trị 0, nên hãng xe Iguana vẫn chưa thể chỉ ra một mức ý nghĩa tốt hơn trong chi phí sửa chữa.

Ok... nếu quên mất sự an toàn ... thì anh ta cũng không thể thuyết phục người dùng cho dù kiểu dáng có đẹp đến mấy!

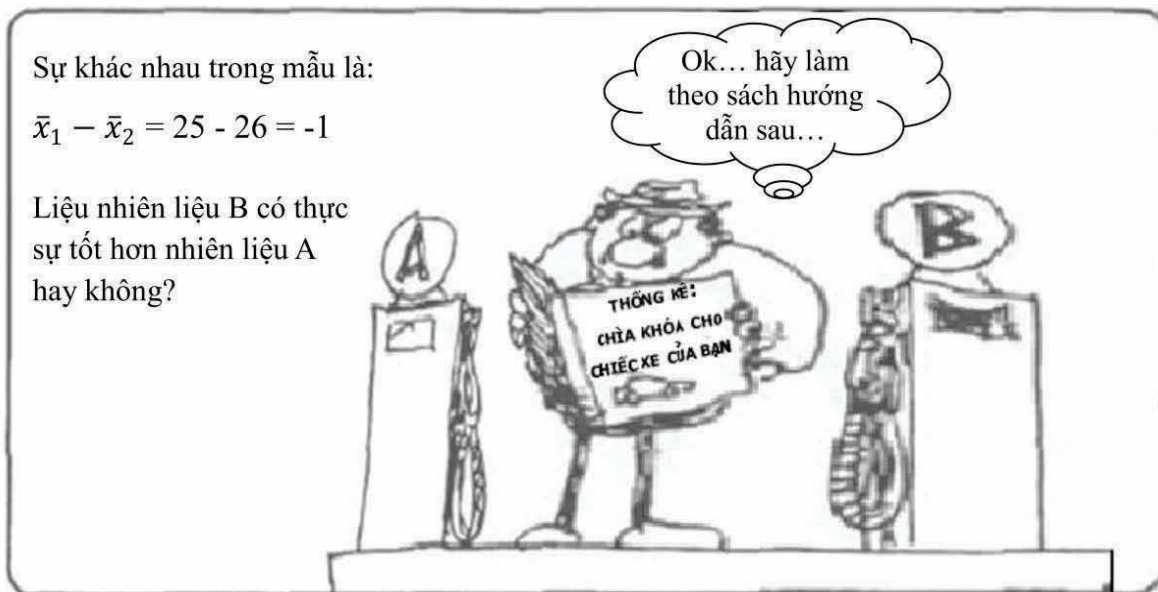


➤ ➤ ➤ HỌC THỐNG KÊ QUA TRUYỆN TRANH



Bắt đầu với 100 chiếc taxi, chọn ngẫu nhiên 50 chiếc taxi sử dụng nhiên liệu A, còn lại sử dụng nhiên liệu B, sau 1 ngày lái xe, thu được kết quả như sau:

	Cỡ mẫu	Số dặm bình quân	Độ lệch chuẩn
A	50	25	5.00
B	50	26	4.00



Vì độ lệch chuẩn lớn nên sai số chuẩn là:

$$\begin{aligned} SE(\bar{X}_1 - \bar{X}_2) &= \sqrt{\frac{S_1^2}{n_1} + \frac{S_2^2}{n_2}} \\ &= \sqrt{\frac{25}{50} + \frac{16}{50}} \\ &= \mathbf{0.905} \end{aligned}$$

Ở mức độ tin cậy 95%, chúng ta có:

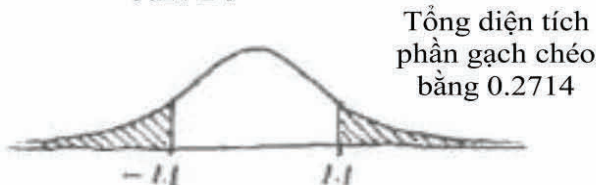
$$\begin{aligned} \mu_1 - \mu_2 &= \bar{x}_1 - \bar{x}_2 \pm z_{0.025} (0.905) \\ &= -1 \pm (1.96)(0.905) \\ &= \mathbf{-1 \pm 1.774} \end{aligned}$$

Khoảng trên bao gồm cả giá trị 0, tương ứng với $\mu_1 = \mu_2$



P-value của giả thuyết thay thế, $H_a: \mu_1 \neq \mu_2$ là:

$$\begin{aligned} \Pr(|z| \geq |z_{qs}|) &= \Pr(|z| \geq \frac{1}{0.905}) \\ &= \Pr(|z| \geq 1.1) = 2(0.1357) \\ &= \mathbf{0.2714} \end{aligned}$$



Diện tích này vượt qua cả mức ý nghĩa $\alpha=0.05$, do đó chúng ta chưa kết luận được nhiên liệu B tốt hơn nhiên liệu A.



(Còn nữa)

Biên dịch: Minh Ánh và các nghiên cứu viên, Viện Khoa học Thống kê