

1 Playbill

Dữ liệu về doanh thu từ bán vé của các vở kịch cho hai tuần:

- October 11-17, 2004 (current week)
- và October 3-10, 2004 (last week)

Ta xem xét xem có mối quan hệ giữa doanh thu giữa current week và last week hay không.

Đầu tiên ta tải dữ liệu vào R:

```
Playbill<-read.csv(file.choose(), header = TRUE, sep = ",")
```

Dựa trên các thông tin về dữ liệu, ta nhận thấy có khả năng có mối quan hệ tuyến tính giữa current week và last week. Do đó ta fit một mô hình hồi quy tuyến tính cho hai biến trên: $Y=\beta_0+\beta_1*x+e$, trong đó:

```
y <- Playbill$CurrentWeek  
x <- Playbill$LastWeek
```

Kết quả fit: $Y = 6805 + 0.98*X$

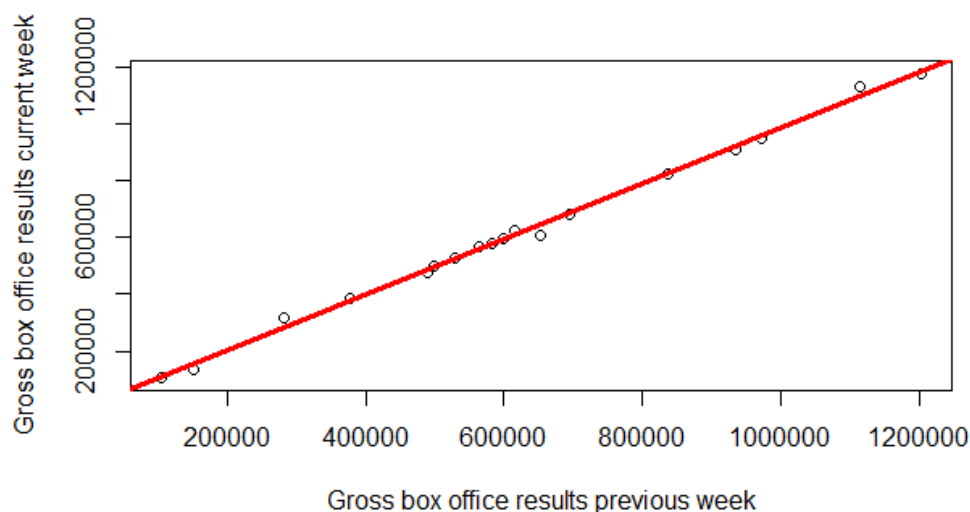
Coefficients:

	Estimate	Std. Error	t value	Pr(> t)
(Intercept)	6.805e+03	9.929e+03	0.685	0.503
x	9.821e-01	1.443e-02	68.071	<2e-16 ***

Giải thích: Ban đầu doanh thu của Current week là 6805\$. Sau đó với mỗi đô la tăng thêm của doanh thu tuần trước, thì doanh thu tuần này sẽ tăng 0.98\$.

Current week và last week có mối quan hệ tuyến tính mạnh, có thể thấy qua plot dưới:

```
plot(x, y, xlab="Gross box office results previous week",  
     ylab="Gross box office results current week")
```



a) Tìm khoảng tin cậy 95% cho β_1 ; 1 có phải là giá trị tốt cho β_1 ?

Ta tìm khoảng tin cậy qua hàm `confint()`

```
confint(M1) #với M1 là tên của mô hình hồi quy tuyến tính
              2.5 %      97.5 %
(Intercept) -1.424433e+04 27854.099443
x            9.514971e-01  1.012666
```

Ta có 95% chắc chắn rằng β_1 sẽ nằm trong khoảng từ 0.9515 tới 1.0127.

Do khoảng này chứa giá trị 1 nên 1 là một giá trị hợp lý cho β_1 .

b) Kiểm định giả thuyết hai phía cho β_0

$$H_0: \beta_0 = 10000$$

$$H_1: \beta_0 \neq 10000$$

Ta tính giá trị t của các quan sát (tobs) và so sánh nó với $t_{0.05/2}^{16}$

Tính T dựa trên công thức:

$$T = \frac{\hat{\beta}_0 - \beta_0^0}{se(\hat{\beta}_0)}$$

```
b0_head <- coef(M1)[ "(Intercept)" ]
se_b0_head = out$coefficients[1, 2]
t_val_for_b0_head <- (b0_head - 10000)/se_b0_head
t_val_for_b0_head > tval
# Kết quả: FALSE => tval > t_obs
```

Sau khi so sánh kết quả giữa tobs và tval ta thấy tval > tobs, do đó ta không thể bác bỏ giả thiết

$H_0: \beta_0 = 10000$, nghĩa là không thể bác bỏ giá trị ban đầu của doanh thu tuần này là 10000\$.

c) Dự đoán doanh thu cho current week, biết doanh thu last week là 400,000\$:

Sử dụng hàm `predict()` trong R:

```
predict(M1, data.frame(x = 400000), interval="prediction", level=0.95)
##      fit      lwr      upr
## 399637.5 359832.8 439442.2
```

Nhận xét: Nếu doanh thu tuần trước là 400,000\$ thì doanh thu tuần này được dự đoán là 399637.5\$.

Ngoài ra, 450,000\$ là dự đoán không hợp lý vì giá trị này không nằm trong khoảng dự đoán 95%: từ 359,833\$ tới 439,442\$.

d) Nhận xét về prediction rule rằng doanh thu tuần này sẽ bằng doanh thu tuần trước :

Như đã được tính từ câu a) khoảng tin cậy 95% của β_0 và β_1 là:

```
confint(M1) #với M1 là tên của mô hình hồi quy tuyến tính
```

	2.5 %	97.5 %
(Intercept)	-1.424433e+04	27854.099443
x	9.514971e-01	1.012666

Khi doanh thu của 2 tuần bằng nhau thì $\beta_0 = 0$ và $\beta_1 = 1$

Do hai giá trị này nằm trong khoảng tin cậy 95% tương ứng của β_0 , β_1 nên dự đoán này có thể chấp nhận được. Tuy nhiên trong thực tế thì doanh thu vé sẽ không giống nhau hoàn toàn, mà sẽ có xu hướng giảm dần theo thời gian vì số người xem giảm dần (ước lượng của β_1 là 0.98 cũng cho thấy mối quan hệ này – lượng tăng doanh thu tuần này nhỏ hơn tuần trước). Vậy prediction rule đã cho có thể chấp nhận nhưng chưa chính xác.