

Các yếu tố ảnh hưởng đến

# QUYẾT ĐỊNH MUA QUẦN ÁO

của **sinh viên**

**Giáo viên hướng dẫn: Thầy Phạm Trí Cao**

**Nhóm 4 - Sáng Thứ 7**

1. Nguyễn Thị Duyên
2. Trương Thị Hồng Mai
3. Doãn Phương Hà My
4. Hoàng Nhật
5. Huỳnh Phương Thảo



@kinhteluongshop

@kinhteluongshop



# Mô tả các biến



STT	Tên biến	Mô tả
0	<b>slquanao</b>	Số lượng quần áo mua (cái/năm)
1	<b>phucap</b>	Phụ cấp từ người thân (nghìn đồng/tháng)
2	<b>thunhap</b>	Thu nhập từ việc làm thêm (nghìn đồng/tháng)
3	<b>dichoi</b>	Số lần đi chơi trong 1 tháng
4	<b>dihoc</b>	Số ngày đi học trong 1 tuần
5	<b>sanlong</b>	Số tiền sẵn lòng chi trả cho 1 sản phẩm (nghìn đồng)
6	<b>nam</b>	Giới tính (nam, nữ)
7	<b>kaki, jeans, cotton</b>	Chất liệu quần áo thu hút
8	<b>thuongmai</b>	Các cửa hàng quần áo thường mua có sử dụng nền tảng thương mại điện tử không
9	<b>mxh, ngoaiduong, nguoiathan</b>	Đa phần thấy các quảng cáo về quần áo ở đâu
10	<b>tot, binhthuong</b>	Đánh giá dịch vụ chăm sóc khách hàng ở các cửa hàng thường mua

# Chuẩn bị data



```
# Import data  
library(readxl)  
data <- read_excel('data.xlsx')  
is.data.frame(data)
```



@kinhteluongshop



@kinhteluongshop



01

# HỒI QUY CÁC BIẾN ĐỊNH LƯỢNG

# 1.1 Hồi quy Số lượng quần áo theo Phụ cấp



```
> caul <- lm(slquanao~phucap,data=data)
> summary(caul)
```

Call:

```
lm(formula = slquanao ~ phucap, data = data)
```

Residuals:

Min	1Q	Median	3Q	Max
-13.697	-9.002	-5.002	4.303	86.040

Coefficients:

	Estimate	Std. Error	t value	Pr(> t )
(Intercept)	1.396e+01	2.330e+00	5.991	1.13e-08 ***
phucap	3.474e-04	5.740e-04	0.605	0.546

---

Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 15.54 on 178 degrees of freedom

Multiple R-squared: 0.002054, Adjusted R-squared: -0.003553

F-statistic: 0.3663 on 1 and 178 DF, p-value: 0.5458

Trong điều kiện các yếu tố khác không đổi, nếu phụ cấp tăng 1 (nghìn đồng) thì số lượng quần áo mua (cái/năm) tăng  $3.474 \times 10^{-04}$ .

## 1.2 Hồi quy Số lượng quần áo theo Thu nhập



```
> cau2 <- lm(slquanao~thunhap,data=data)
> summary(cau2)

Call:
lm(formula = slquanao ~ thunhap, data = data)

Residuals:
    Min       1Q   Median       3Q      Max
-33.379  -8.395  -4.096   4.627  85.053

Coefficients:
            Estimate Std. Error t value Pr(>|t|)
(Intercept) 1.410e+01  1.290e+00  10.929  <2e-16 ***
thunhap      8.513e-04  4.593e-04   1.854   0.0654 .
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 15.41 on 178 degrees of freedom
Multiple R-squared:  0.01894,    Adjusted R-squared:  0.01343
F-statistic: 3.436 on 1 and 178 DF,  p-value: 0.06544
```

Trong điều kiện các yếu tố khác không đổi, nếu thu nhập tăng 1 (nghìn đồng) thì số lượng quần áo mua (cái/năm) tăng  $8.513 \times 10^{-04}$ .

## 1.3 Hồi quy Số lượng quần áo theo Đi chơi



```
> cau3 <- lm(slquanao~dichoi,data=data)
> summary(cau3)

Call:
lm(formula = slquanao ~ dichoi, data = data)

Residuals:
    Min       1Q   Median       3Q      Max
-37.067  -6.786  -2.797   2.149  90.236

Coefficients:
              Estimate Std. Error t value Pr(>|t|)
(Intercept)   7.7427     1.5546   4.980 1.49e-06 ***
dichoi        1.0108     0.1565   6.459 9.75e-10 ***
---
Signif. codes:
  0 '***' 0.001 '**' 0.01 '*' 0.05 '.'
  0.1 ' ' 1

Residual standard error: 14 on 178 degrees of freedom
Multiple R-squared:  0.1899, Adjusted R-squared:  0.1853
F-statistic: 41.71 on 1 and 178 DF,  p-value: 9.751e-10
```

Trong điều kiện các yếu tố khác không đổi, nếu số lần đi chơi tăng 1 (lần/tháng) thì số lượng quần áo mua (cái/năm) tăng 1.0108.

## 1.4 Hồi quy Số lượng quần áo theo Đi học



```
> cau4 <- lm(slquanao~dihoc,data=data)
> summary(cau4)

Call:
lm(formula = slquanao ~ dihoc, data = data)

Residuals:
    Min       1Q   Median       3Q      Max
-14.740  -8.490  -5.708   4.268  86.970

Coefficients:
            Estimate Std. Error t value Pr(>|t|)
(Intercept)  17.7728     3.2049   5.545 1.05e-07 ***
dihoc        -0.6775     0.7819  -0.866   0.387
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 15.53 on 178 degrees of freedom
Multiple R-squared:  0.0042, Adjusted R-squared:  -0.001395
F-statistic: 0.7507 on 1 and 178 DF,  p-value: 0.3874
```

Trong điều kiện các yếu tố khác không đổi, nếu số ngày đi học tăng 1 (ngày/tuần) thì số lượng quần áo mua (cái/năm) giảm 0.6775.



## 1.5 Hồi quy Số lượng quần áo theo Săn lòng



```
> cau5 <- lm(slquanao~sanlong,data=data)
> summary(cau5)
```

Call:

```
lm(formula = slquanao ~ sanlong, data = data)
```

Residuals:

Min	1Q	Median	3Q	Max
-29.809	-8.638	-4.594	5.005	85.362

Coefficients:

	Estimate	Std. Error	t value	Pr(> t )
(Intercept)	1.410e+01	1.206e+00	11.693	< 2e-16 ***
sanlong	1.785e-03	6.649e-04	2.685	0.00794 **

---

Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 15.25 on 178 degrees of freedom

Multiple R-squared: 0.03892, Adjusted R-squared: 0.03352

F-statistic: 7.209 on 1 and 178 DF, p-value: 0.007941

Trong điều kiện các yếu tố khác không đổi, nếu số tiền săn lòng chi trả cho một sản phẩm tăng 1 (nghìn đồng) thì số lượng quần áo mua (cái/năm) tăng  $1.785 \times 10^{-03}$ .

## 1.6 Chọn ra mô hình có $R^2$ lớn nhất

### Minh họa

Mô hình	$R^2$
cau1	0.002053898
cau2	0.018938372
cau3	0.189858770
cau4	0.004199629
cau5	0.038922096

Mô hình có  $R^2$  lớn nhất là mô hình hồi quy sldquanao theo dichoi

```
> # 6. Tính và chọn mô hình có  $R^2$  lớn nhất
> # Gán biến cho  $R^2$ 
> r2_cau1 = summary(cau1)$r.squared
> r2_cau2 = summary(cau2)$r.squared
> r2_cau3 = summary(cau3)$r.squared
> r2_cau4 = summary(cau4)$r.squared
> r2_cau5 = summary(cau5)$r.squared
> # Tạo bảng
> table6 <-
matrix(c(r2_cau1,r2_cau2,r2_cau3,r2_cau4,r2_cau5),ncol=1,byrow=TRUE)
> colnames(table6) <- c("R Squared")
> rownames(table6) <- c("Câu 1","Câu 2","Câu 3","Câu 4","Câu 5")
> table6
      R Squared
Câu 1 0.002053898
Câu 2 0.018938372
Câu 3 0.189858770
Câu 4 0.004199629
Câu 5 0.038922096
> max(table6)
[1] 0.1898588
```

## 1.7 Hồi quy Full Chung



```
> full_chung <- lm(slquanao ~ ., data = data)
> summary(full_chung)

Call:
lm(formula = slquanao ~ ., data = data)

Residuals:
    Min       1Q   Median       3Q      Max
-27.846  -6.137  -2.280   3.396  84.963

Coefficients:
            Estimate Std. Error t value Pr(>|t|)
(Intercept)  8.5942328  8.1999506   1.048  0.2961
nam         -5.3044700  2.3060494  -2.300  0.0227 *
phucap      -0.0007327  0.0005689  -1.288  0.1996
thunhap      0.0003083  0.0004583   0.673  0.5022
dihoc       -0.0343286  0.7220472  -0.048  0.9621
dichoi       0.9723914  0.1683232   5.777 3.73e-08 ***
sanlong      0.0007741  0.0006656   1.163  0.2465
kaki         2.1785735  4.0642855   0.536  0.5927
jeans        7.3782358  3.7699965   1.957  0.0520 .
cotton       6.0427203  3.3266293   1.816  0.0711 .
thuongmai    0.8273485  4.4862734   0.184  0.8539
mxh          -3.8618894  3.3089262  -1.167  0.2449
ngoaiduong  -1.4594681  7.0999222  -0.206  0.8374
nguoiathan  -4.5620335  5.0988137  -0.895  0.3722
tot          3.8346069  6.0168113   0.637  0.5248
binhthuong  -0.4384577  5.7344886  -0.076  0.9391
---
Signif. codes:
  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 13.66 on 164 degrees of freedom
Multiple R-squared:  0.29,    Adjusted R-squared:  0.225
F-statistic: 4.465 on 15 and 164 DF,  p-value: 4.92e-07
```

Dùng khi tính toán CP của Mallows

## 1.7 Mô hình Tuyến tính | $Y = \beta_0 + \beta_3 X$



```
> tuyentinh4 <- lm(slquanao~dichoi, data = data)
> summary(tuyentinh4)

Call:
lm(formula = slquanao ~ dichoi, data = data)

Residuals:
    Min       1Q   Median       3Q      Max
-37.067  -6.786  -2.797   2.149  90.236

Coefficients:
              Estimate Std. Error t value Pr(>|t|)
(Intercept)   7.7427      1.5546   4.980 1.49e-06 ***
dichoi        1.0108      0.1565   6.459 9.75e-10 ***
---
Signif. codes:
  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 14 on 178 degrees of freedom
Multiple R-squared:  0.1899,    Adjusted R-squared:  0.1853
F-statistic: 41.71 on 1 and 178 DF,  p-value: 9.751e-10
```

Trong điều kiện các yếu tố khác không đổi, nếu số lần đi chơi tăng 1 (lần/tháng) thì số lượng quần áo mua (cái/năm) tăng 1.0108.

## 1.7 Mô hình Tuyến tính - Logarit | $Y = \beta_0 + \beta_3 \ln(X)$



```
> linlog4 <- lm(slquanao ~ log(dichoi), data=data, dichoi>0)
> summary(linlog4)
```

Call:

```
lm(formula = slquanao ~ log(dichoi), data = data, subset = dichoi >
    0)
```

Residuals:

	Min	1Q	Median	3Q	Max
	-25.509	-7.952	-3.370	3.828	91.150

Coefficients:

	Estimate	Std. Error	t value	Pr(> t )
(Intercept)	4.330	2.440	1.774	0.0778 .
log(dichoi)	6.521	1.296	5.031	1.21e-06 ***

---

Signif. codes:

0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 14.54 on 173 degrees of freedom

Multiple R-squared: 0.1276, Adjusted R-squared: 0.1226

F-statistic: 25.31 on 1 and 173 DF, p-value: 1.215e-06

Trong điều kiện các yếu tố khác không đổi, nếu số lần đi chơi trên tháng tăng 1% thì số lượng quần áo mua (cái/năm) tăng 0.06521.

## 1.7 Mô hình hồi quy qua Gốc tọa độ | $Y=0+\beta_3X$



```
> origin4 <- lm(slquanao ~ 0 + dichoi,data=data)
> summary(origin4)
```

Call:

```
lm(formula = slquanao ~ 0 + dichoi, data = data)
```

Residuals:

Min	1Q	Median	3Q	Max
-46.653	-2.045	2.029	6.882	96.823

Coefficients:

	Estimate	Std. Error	t value	Pr(> t )
dichoi	1.5884	0.1119	14.2	<2e-16 ***

---

Signif. codes:

0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 14.91 on 179 degrees of freedom

Multiple R-squared: 0.5298, Adjusted R-squared: 0.5271

F-statistic: 201.7 on 1 and 179 DF, p-value: < 2.2e-16

( $R^2 = 0.0769632$ ,  $R^2$  hiệu chỉnh = 0.0769632)

Trong điều kiện các yếu tố khác không đổi, nếu số lần đi chơi tăng 1 (lần/tháng) thì số lượng quần áo mua (cái/năm) tăng 1.5884.

## 1.7 Mô hình dạng hàm bậc hai | $Y = \beta_0 + \beta_1 X + \beta_2 X^2$



```
> power4 <- lm(slquanao ~ dichoi + I(dichoi^2), data=data)
> summary(power4)
```

Call:

```
lm(formula = slquanao ~ dichoi + I(dichoi^2), data = data)
```

Residuals:

Min	1Q	Median	3Q	Max
-41.994	-6.376	-2.815	3.255	89.192

Coefficients:

	Estimate	Std. Error	t value	Pr(> t )
(Intercept)	9.90365	2.26475	4.373	2.09e-05 ***
dichoi	0.40564	0.48771	0.832	0.407
I(dichoi^2)	0.02325	0.01775	1.310	0.192

---

Signif. codes:

0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 13.98 on 177 degrees of freedom

Multiple R-squared: 0.1976, Adjusted R-squared: 0.1886

F-statistic: 21.8 on 2 and 177 DF, p-value: 3.442e-09

Trong điều kiện các yếu tố khác không đổi, nếu số lần đi chơi tăng từ dichoi lên dichoi+1 (lần/tháng) thì số lượng quần áo mua (cái/năm) tăng  $0.40564 + (0.02325 \times \text{dichoi})$ .

## 1.7 Chọn mô hình bằng các tiêu chí

```
> # Tuyển tính
> r2a_tuyentinh4 = summary(tuyentinh4)$adj.r.squared
> aic_tuyentinh4 = AIC(tuyentinh4)
> bic_tuyentinh4 = BIC(tuyentinh4)
> tuyentinh4_cp=ols_mallows_cp(tuyentinh4, full_chung)
> # Lin - Log
> r2a_linlog4 = summary(linlog4)$adj.r.squared
> aic_linlog4 = AIC(linlog4)
> bic_linlog4 = BIC(linlog4)
> linlog4_cp=ols_mallows_cp(linlog4, full_chung)
> # Qua Gốc Tọa Độ
> r2a_origin4 = summary(origin4)$adj.r.squared
> aic_origin4 = AIC(origin4)
> bic_origin4 = BIC(origin4)
> origin4_cp=ols_mallows_cp(origin4, full_chung)
> # Bậc Hai
> r2a_power4 = summary(power4)$adj.r.squared
> aic_power4 = AIC(power4)
> bic_power4 = BIC(power4)
> power4_cp=ols_mallows_cp(power4, full_chung)
> table7 <- matrix(c(aic_tuyentinh4,bic_tuyentinh4,tuyentinh4_cp,r2a_tuyentinh4,
                     aic_linlog4,bic_linlog4,linlog4_cp,r2a_linlog4,
                     aic_origin4,bic_origin4,origin4_cp,r2a_origin4,
                     aic_power4,bic_power4,power4_cp,r2a_power4),
                    ncol=4,byrow=TRUE)
colnames(table7) <- c("AIC","BIC","CP","R2 Adjusted")
rownames(table7) <- c("Tuyển tính","Lin Log","Gốc tọa độ","Bậc 2")
> table7
```

	AIC	BIC	CP	R2 Adjusted
Tuyển tính	1464.988	1474.567	11.12393	0.1853074
Lin Log	1437.630	1447.124	25.14523	0.1225834
Gốc tọa độ	1486.471	1492.857	37.20021	0.5271419
Bậc 2	1465.252	1478.023	11.32758	0.1885697





## 1.7 BẢNG MINH HỌA

Mô hình	AIC	BIC	CP của Mallows	R <sup>2</sup> hiệu chỉnh
Tuyến tính	1464.988	1474.567	11.12393	0.1853074
Lin Log	1437.630	1447.124	25.14523	0.1225834
Gốc tọa độ	1486.471	1492.857	37.20021	0.0769632
Bậc 2	1465.252	1478.023	11.32758	0.1885697

\*R<sup>2</sup> hiệu chỉnh của hồi quy qua gốc tọa độ đã được tính lại bằng tay

➡ Chọn mô hình Lin Log

## 1.8 Mô hình hồi quy log - log | $\ln(Y) = \beta_3 \ln(X)$



```
> # Tuyến tính log
> ttlog4 <- lm(log(slquanao)~log(dichoi),data=data,slquanao>0&dichoi>0)
> summary(ttlog4)
```

Call:  
lm(formula = log(slquanao) ~ log(dichoi), data = full, subset = slquanao > 0 & dichoi > 0)

Residuals:

Min	1Q	Median	3Q	Max
-3.07626	-0.41785	-0.06103	0.43918	2.60599

Coefficients:

	Estimate	Std. Error	t value	Pr(> t )
(Intercept)	1.72349	0.12309	14.002	< 2e-16 ***
log(dichoi)	0.39774	0.06538	6.084	7.33e-09 ***

---

Signif. codes:  
0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 0.7336 on 173 degrees of freedom  
Multiple R-squared: 0.1762, Adjusted R-squared: 0.1715  
F-statistic: 37.01 on 1 and 173 DF, p-value: 7.334e-09

Trong điều kiện các yếu tố khác không đổi, nếu số lần đi chơi tăng 1% (lần/tháng) thì số lượng quần áo mua (cái/năm) tăng 0.39774%.

## 1.8 Mô hình hồi quy log - lin | $\ln(Y)=\beta_3 X$



```
> # Log Lin
> loglin4 <- lm(log(slquanao)~dichoi,data=data,slquanao>0)
> summary(loglin4)
```

Call:  
lm(formula = log(slquanao) ~ dichoi, data = full, subset = slquanao > 0)

Residuals:

Min	1Q	Median	3Q	Max
-3.6109	-0.4538	0.0615	0.4704	2.5285

Coefficients:

	Estimate	Std. Error	t value	Pr(> t )
(Intercept)	1.967093	0.083285	23.619	< 2e-16 ***
dichoi	0.054794	0.008384	6.535	6.46e-10 ***

---

Signif. codes:  
0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 0.7503 on 178 degrees of freedom  
Multiple R-squared: 0.1935, Adjusted R-squared: 0.189  
F-statistic: 42.71 on 1 and 178 DF, p-value: 6.463e-10

Trong điều kiện các yếu tố khác không đổi, nếu số lần đi chơi tăng 1 (lần/tháng) thì số lượng quần áo mua (cái/năm) tăng 5.4794%.

## 1.8 Chọn mô hình bằng các tiêu chí



```
> # 8b. Chọn mô hình
> # Log Log
> r2a_ttlog4 = summary(ttlog4)$adj.r.squared
> aic_ttlog4 = AIC(ttlog4)
> bic_ttlog4 = BIC(ttlog4)
> ttlog4_cp=ols_mallows_cp(ttlog4, full_chung)
>
> # Log Lin
> r2a_loglin4 = summary(loglin4)$adj.r.squared
> aic_loglin4 = AIC(loglin4)
> bic_loglin4 = BIC(loglin4)
> loglin4_cp=ols_mallows_cp(loglin4, full_chung)
>
> table8 <-
matrix(c(aic_ttlog4,bic_ttlog4,ttlog4_cp,r2a_ttlog4,
aic_loglin4,bic_loglin4,loglin4_cp,r2a_loglin4),ncol=4,byrow=
TRUE)
> colnames(table8) <- c("AIC","BIC","CP","R2 Adjusted")
> rownames(table8) <- c("Tuyển tính Log","Log Lin")
> table8
```

	AIC	BIC	CP	R2 Adjusted
Tuyển tính Log	392.1792	401.6736	-170.5010	0.1714722
Log Lin	411.3724	420.9512	-175.4629	0.1889809



## 1.8 BẢNG MINH HỌA

Mô hình	AIC	BIC	CP của Mallows	$R^2$ hiệu chỉnh
Tuyến tính	392.1792	401.6736	-170.5010	0.1714722
Lin Log	411.3724	420.9512	-175.4629	0.1889809

➡ Chọn mô hình Lin Log

02

**Chọn 3 biến định lượng và  
2 định tính để kiểm định**

```

> hqcau2 = lm(formula = slquanao ~ phucap + thunhap + sanlong + nam + tot, data = data)
>
> print(summary(hqcau2))
Call:
lm(formula = slquanao ~ phucap + thunhap + sanlong + nam + tot,
    data = data)

Residuals:
    Min       1Q   Median       3Q      Max
-28.434  -7.990  -4.194   3.739  85.111

Coefficients:
              Estimate Std. Error t value Pr(>|t|)
(Intercept) 13.4960547   2.4722323   5.459 1.63e-07 ***
phucap       0.0002354   0.0005781   0.407  0.68430
thunhap      0.0005240   0.0004859   1.078  0.28231
sanlong      0.0011371   0.0007067   1.609  0.10941
nam          -6.2107293   2.3627393  -2.629  0.00934 **
tot          5.7789742   2.5799422   2.240  0.02636 *
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 14.82 on 174 degrees of freedom
Multiple R-squared:  0.1133, Adjusted R-squared:  0.08786
F-statistic: 4.448 on 5 and 174 DF, p-value: 0.0007727

```

Chọn biến phucap, thunhap, sanlong, nam và tot

## 2.1 Viết hàm cho kiểm định t và giá trị tới hạn

```
> # Kiểm định t
> t_test <- function(reg, coefnum, val){
  coefnum <- coefnum+1 # hệ số cần kiểm tra vị trí số 1
  co <- coef(summary(reg))
  tstat <- (co[coefnum,1]-val)/co[coefnum,2]
  p_value <- 2 * pt(abs(tstat), reg$df.residual, lower.tail = FALSE)
  result1 <- c(t_stat = round(tstat, digits=4))
  result2 = c(p_value_2p = round(p_value, digits=4))
  result3 = c(p_value_1p = round(p_value/2, digits=4))
  return(list(result1, result2, result3))
}
```

```
> # Giá trị tới hạn của Kiểm định t
> t_critical <- function(alpha, df){
+   result1 <- c(t_alpha = round(abs(qt(alpha, df)), digits=4))
+   result2 <- c(t_alpha_chia2 = round(abs(qt(alpha/2, df)), digits=4))
+   return(list(result1, result2))
+ }

> df <- hqcau2$df.residual
```



```
> t_test(hqcau2, 1, 0)
[[1]]
t_stat
0.4073

[[2]]
p_value_2p
0.6843

[[3]]
p_value_1p
0.3422
```

```
> t_critical (0.05,df)
[[1]]
t_alpha
1.6537

[[2]]
t_alpha_chia2
1.9737
```

$$H_0: \beta_1 = 0;$$

$$H_1: \beta_1 \neq 0,$$

**TẠI MỨC Ý NGHĨA 5% (KIỂM ĐỊNH t  
THEO GIÁ TRỊ TỐI HẠN, p - value)**

## Kết luận

$$H_0: \beta_1 = 0 ; H_1: \beta_1 \neq 0 \text{ (ứng với biến } phucap)$$

**Phương pháp giá trị tới hạn**, với mức ý nghĩa 5%:

$$|t| = 0.4073 < t_{0,025}(174) = 1,9737 : \text{chấp nhận } H_0$$

**Phương pháp p-value**, với mức ý nghĩa 5%:

$$p\text{-value}(2p) = 0.6843 > 0.05 = \alpha : \text{chấp nhận } H_0$$

$H_0: \beta_1 = 0;$

$H_1: \beta_1 > 0,$

**TẠI MỨC Ý NGHĨA 5% (KIỂM ĐỊNH t  
THEO GIÁ TRỊ TỐI HẠN, p - value)**

## Kết luận

$H_0: \beta_1 = 0; H_1: \beta_1 > 0$  (ứng với biến *phucap*)

**Phương pháp giá trị tới hạn**, với mức ý nghĩa 5%:

$|t| = 0.4073 < t_{0,05}(174) = 1,6537$  : chấp nhận  $H_0$

**Phương pháp p-value**, với mức ý nghĩa 5%:

$p\text{-value}(1p) = 0.3422 > 0.05 = \alpha$ : chấp nhận  $H_0$

```
> t_test(hqcau2, 1, 0)
[[1]]
t_stat
0.4073

[[2]]
p_value_2p
0.6843

[[3]]
p_value_1p
0.3422
```

```
> t_critical (0.05,df)
[[1]]
t_alpha
1.6537

[[2]]
t_alpha_chia2
1.9737
```

```
> t_test(hqcau2, 5, 0.4)
[[1]]
t_stat
2.0849

[[2]]
p_value_2p
0.0385

[[3]]
p_value_1p
0.0193
```

```
> t_critical (0.05,df )
[[1]]
t_alpha
1.6537

[[2]]
t_alpha_chia2
1.9737
```

$$H_0: \delta_2 = 0.4;$$

$$H_1: \delta_2 \neq 0.4$$

**TẠI MỨC Ý NGHĨA 5% (KIỂM ĐỊNH  
t THEO GIÁ TRỊ TỚI HẠN, p - value)**

## Kết luận

$$H_0: \delta_2 = 0,4 ; H_1: \delta_2 \neq 0,4 \text{ (ứng với biến } tot)$$

**Phương pháp giá trị tới hạn**, với mức ý nghĩa 5%:

$$|t| = 2.0849 > t_{0,025}(174) = 1.9737: \text{Bác bỏ } H_0$$

**Phương pháp p-value**, với mức ý nghĩa 5%:

$$p\text{-value}(2p) = 0.0385 < 0,05 = \alpha: \text{Bác bỏ } H_0$$

$H_0: \delta_2 = 0,3;$

$H_1: \delta_2 > 0,3$

**TẠI MỨC Ý NGHĨA 5% (KIỂM ĐỊNH t  
THEO GIÁ TRỊ TỐI HẠN, p - value)**

## Kết luận

$H_0: \delta_2 = 0,3; H_1: \delta_2 > 0,3$  (ứng với biến *tot*)

**Phương pháp giá trị tới hạn**, với mức ý nghĩa 5%

$t = 2.1237 > t_{0,05}(174) = 1.6537$ : Bác bỏ  $H_0$

**Phương pháp p-value**, với mức ý nghĩa 5%:

$p\text{-value}(1p) = 0.0176 < 0,05 = \alpha$ : Bác bỏ  $H_0$

```
> t_test(hqcau2, 5, 0.3)
[[1]]
t_stat
2.1237

[[2]]
p_value_2p
0.0351

[[3]]
p_value_1p
0.0176
```

```
> t_critical (0.05,df )
[[1]]
t_alpha
1.6537

[[2]]
t_alpha_chia2
1.9737
```

$H_0: \delta_1=0,4$  ;  $H_1: \delta_1 \neq 0,4$  , MỨC Ý NGHĨA 5% ( KIỂM ĐỊNH F THEO GIÁ TRỊ TỐI HẠN,  
p-value)

```
> library(car)
> linearHypothesis(hqcau2, c("nam = 0.4"))
```

Linear hypothesis test

Hypothesis:  
nam = 0.4

Model 1: restricted model

Model 2: slquanao ~ phucap + thunhap + sanlong + nam + tot

	Res.Df	RSS	Df	Sum of Sq	F	Pr(>F)
1	175	39928				
2	174	38209	1	1719	7.8283	0.005723 **

---

Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1

## Kết luận

$H_0: \delta_1=0,4$  ;  $H_1: \delta_1 \neq 0,4$

**Phương pháp giá trị tới hạn**, với mức ý nghĩa 5%:

$F = 7.8283 > F_{0,05}(1,174) = 3.9$  : bác bỏ  $H_0$

**Phương pháp p-value**, với mức ý nghĩa 5%:

p-value = 0,005723 < 0.05 =  $\alpha$ : bác bỏ  $H_0$

## 2.2 Thực hiện kiểm định

$H_0 : \beta_1 = 4000, \beta_2 = 2000, \delta_2 = 1; H_1: H_0 \text{ sai (dùng kiểm định F)}$

```
> library(car)
> linearHypothesis(hqcau2, c("phucap=4000", "sanlong =2000",
"tot=1"))
```

Linear hypothesis test

Hypothesis:

```
phucap = 4000
sanlong = 2000
tot = 1
```

Model 1: restricted model

Model 2: slquanao ~ phucap + thunhap + sanlong + nam + tot

	Res.Df	RSS	Df	Sum of Sq	F	Pr(>F)
1	177	1.4680e+16				
2	174	3.8209e+04	3	1.468e+16	2.2284e+13	< 2.2e-16 ***

---

Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1

## Kết luận

**Phương pháp giá trị tới hạn**, với mức ý nghĩa 5%:

$$F = 2.2284.10^{13} > F_{0,05}(3,174) = 2.66$$

$\Rightarrow$  bác bỏ  $H_0$

## 2.3 Thực hiện kiểm định

$H_0: 3\beta_1 + 2\beta_2 = 5000$ ;  $H_1: H_0$  sai (dùng kiểm định F)

```
> library(car)
> linearHypothesis(hqcau2, c("3*phucap + 2*sanlong = 5000"))
```

Linear hypothesis test

Hypothesis:

3 phucap + 2 sanlong = 5000

Model 1: restricted model

Model 2: slquanao ~ phucap + thunhap + sanlong + nam + tot

	Res.Df	RSS	Df	Sum of Sq	F	Pr(>F)
1	175	1.2639e+15				
2	174	3.8209e+04	1	1.2639e+15	5.7557e+12	< 2.2e-16 ***

---

Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1

## Kết luận

**Phương pháp giá trị tới hạn**, với mức ý nghĩa 5%:

$$F = 5.7557 \cdot 10^{12} > F_{0,05}(2,174) = 3.04$$

$\Rightarrow$  bác bỏ  $H_0$

## 2.4 Kiểm định tất cả các hệ số góc của mô hình đều bằng 0 & ý nghĩa của kiểm định

```
> library(car)
> linearHypothesis(hqcau2, c("phucap=0", "thunhap=0", "sanlong=0", "tot=0", "nam=0"))
```

Linear hypothesis test

Hypothesis:

```
phucap = 0
thunhap = 0
sanlong = 0
tot = 0
nam = 0
```

Model 1: restricted model

Model 2: slquanao ~ phucap + thunhap + sanlong + nam + tot

	Res.Df	RSS	Df	Sum of Sq	F	Pr(>F)
1	179	43093				
2	174	38209	5	4884	4.4483	0.0007727 ***

---

Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1

### Kết luận

**Phương pháp p-value**, với mức ý nghĩa 5%:

$$p\text{-value} = 0.0007727 < 0.05 = \alpha$$

$\Rightarrow$  bác bỏ  $H_0$

### Ý nghĩa

Các biến độc không tác động đồng thời đến biến phụ thuộc



## 2.5 Xác định khoảng tin cậy 94% cho $\beta_3$

```
> confint(hqcau2, level = 0.94 )
```

		3 %	97 %
(Intercept)	8.815787e+00	18.176322630	
phucap	-8.589437e-04	0.001329823	
thunhap	-3.958121e-04	0.001443793	
sanlong	-2.007313e-04	0.002474918	
nam	-1.068371e+01	-1.737746279	
tot	8.947969e-01	10.663151449	

**Kết luận**

$$\beta_3 \in (0.0002007; 0.002474918)$$

## 2.6 Tìm giá trị trung bình của các biến phucap, thunhap, sanlong

```
> phucap <- data$phucap
> thunhap <- data$thunhap
> sanlong <- data$sanlong
> print(c(
  MEAN_phucap = mean(phucap),
  MEAN_thunhap = mean(thunhap),
  MEAN_sanlong = mean(sanlong)))
```

MEAN_phucap	MEAN_thunhap	MEAN_sanlong
3522.2222	1277.0556	605.3333

## 2.7 Dự đoán giá trị trung bình của số lượng quần áo khi các biến phucap=3000, thunhap=2000, sanlong=2000, tot=1, nam=0

```
> d <- data.frame(phucap = 3000, thunhap = 2000, sanlong =  
3000, tot = 1, nam = 0)  
> predict(hqcau2, d, interval = "confidence", level = 0.95)
```

	fit	lwr	upr
1	24.44061	19.23591	29.6453

### Kết luận

Từ bảng kết quả này, ta nhận thấy khoảng tin cậy 95% cho giá trị trung bình của slquanáo là:

(19.23591; 29.6453)

## 2.8 Xác định VIF để xem có đa cộng tuyến giữa các biến độc lập không

```
> library(car)
```

```
> vif(hqcau2)
```

phucap	thunhap	sanlong	nam	tot
1.115990	1.210565	1.196705	1.025227	1.094574

### Kết luận

Tất cả các giá trị VIF đều nhỏ hơn 10 nên ta có thể xem như mô hình không có đa cộng tuyến nghiêm trọng

## 2.9 Kiểm định xem phần dư có phân phối chuẩn không (cho mức ý nghĩa 10%)

```
> phandu <- resid(hqcau2)
> shapiro.test(phandu)

Shapiro-Wilk normality test

data:  phandu
W = 0.71942, p-value < 2.2e-16
```

### Giả thiết

$H_0$ :  $u$  có phân phối chuẩn;

$H_1$ :  $u$  không có phân phối chuẩn

### Kết luận

**Phương pháp p-value**, với mức ý nghĩa 10%:

$p\text{-value} < 2.2 \times 10^{-16} < 0,1 = \alpha$ : bác bỏ  $H_0$

## 2.10 Nêu ý nghĩa của $\beta_1$ , $\delta_1$

$\beta_1$

Trong điều kiện các yếu tố khác không đổi, hệ số  $\beta_1$  cho biết mức chênh lệch số quần áo mua trong 1 năm khi số tiền phụ cấp khác nhau 1 nghìn đồng

$\delta_1$

Trong điều kiện các yếu tố khác không đổi, hệ số  $\delta_1$  cho biết mức chênh lệch số quần áo mua trong 1 năm giữa nam và nữ

# 03

**CHẠY HỒI QUY s<sub>l</sub>quanao THEO dihoc ( $\beta_1$ ), thunhap ( $\beta_2$ ), binhthuong ( $\delta_1$ ), dihoc\*binhthuong( $\delta_2$ ), thunhap\*binhthuong ( $\delta_3$ )**

Câu 3



```

> hoiquy_cau3 <- lm(slquanao ~ dihoc + thunhap + binhthuong + dihoc:binhthuong +
thunhap:binhthuong , data = data)
> print(summary(hoiquy_cau3))
Call:
lm(formula = slquanao ~ dihoc + thunhap + binhthuong + dihoc:binhthuong +
  thunhap:binhthuong, data = data)

Residuals:
    Min       1Q   Median       3Q      Max
-22.444  -7.892  -4.161   5.289  86.139

Coefficients:
                Estimate Std. Error t value Pr(>|t|)
(Intercept)    13.2072997   5.1790112   2.550  0.01163 *
dihoc           0.5573770   1.3473857   0.414  0.67962
thunhap         0.0028122   0.0008371   3.359  0.00096 ***
binhthuong      3.4188302   6.4820324   0.527  0.59857
dihoc:binhthuong -1.4503294   1.6330796  -0.888  0.37572
thunhap:binhthuong -0.0028987  0.0009908  -2.926  0.00390 **

---
Signif. codes:
  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 14.85 on 174 degrees of freedom
Multiple R-squared:  0.1091, Adjusted R-squared:  0.08354
F-statistic: 4.264 on 5 and 174 DF,  p-value: 0.001107

```



### 3.1 Kiểm định $H_0 : \delta_1=0, \delta_2=0, \delta_3=0$ ; $H_1: H_0$ sai, tại mức ý nghĩa

```
> library(car)
> linearHypothesis(hoi quy_cau3,
c("dihoc:binhthuong=0", "thunhap:binhthuong=0", "binhthuong=0"))
Linear hypothesis test

Hypothesis:
dihoc:binhthuong = 0
thunhap:binhthuong = 0
binhthuong = 0

Model 1: restricted model
Model 2: slquanao ~ dihoc + thunhap + binhthuong + dihoc:binhthuong +
      thunhap:binhthuong

      Res.Df  RSS Df Sum of Sq      F    Pr(>F)
1       177 42032
2       174 38390   3    3642.9 5.5037 0.001235 **
---
Signif. codes:
0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

## Kết luận

Phương pháp p-value, với mức ý nghĩa 5%:

$p\text{-value} = 0.001235 < 0.05 = \alpha$   
 $\Rightarrow$  Bác bỏ  $H_0$

## 3.2 Ý nghĩa của kiểm định

Có sự khác biệt về số lượng quần áo mua trong một năm của những sinh viên có cùng thu nhập và số lần đi học trong một tuần khi mua tại cửa hàng có dịch vụ chăm sóc khách hàng bình thường hoặc không tốt lắm.

Và tác động của dịch vụ chăm sóc khách hàng lên số lượng quần áo mua trong một năm có phụ thuộc đồng thời vào thu nhập và số lần đi học trong một tuần. Hay tác động đồng thời của thu nhập và số lần đi học trong một tuần lên số lượng quần áo mua trong một năm có phụ thuộc vào dịch vụ chăm sóc khách hàng.

@kinhteluongshop



@kinhteluongshop

04

# Phương sai thay đổi

## 4.1 Dùng kiểm định Breusch - Pagan và White (đầy đủ hoặc rút gọn) để xem phương sai có thay đổi hay không

```
> > # BP test ( phương pháp LM )  
> bptest(hoi quy_cau3)
```

studentized Breusch-Pagan test

```
data: hoi quy_cau3  
BP = 11.98, df = 5, p-value =  
0.03506
```

### Giả thiết

$H_0$ : phương sai không thay đổi

$H_1$ : phương sai thay đổi

### Kết luận

**Phương pháp p-value**, với mức ý nghĩa 5%:

$p\text{-value} = 0.03506 < \alpha = 0.05$ : bác bỏ  $H_0$

$\Rightarrow$  Phương sai thay đổi

## 4.1 Dùng kiểm định Breusch - Pagan và White (đầy đủ hoặc rút gọn) để xem phương sai có thay đổi hay không

```
> # White test rut gon (phuong  
phap LM)  
> bptest(hoiquy_cau3, ~  
fitted(hoiquy_cau3) +  
I(fitted(hoiquy_cau3)^2))  
  
studentized Breusch-Pagan test  
  
data: hoiquy_cau3  
BP = 11.541, df = 2, p-value =  
0.003119
```

### Giả thiết

$H_0$ : phương sai không thay đổi

$H_1$ : phương sai thay đổi

### Kết luận

**Phương pháp p-value**, với mức ý nghĩa 5%:

$p\text{-value} = 0.003119 < \alpha = 0.05$ : bác bỏ  $H_0$

$\Rightarrow$  Phương sai thay đổi

## 4.2 Nếu có phương sai thay đổi thì dùng WLS và GLS để xử lý

### Giả thiết

$H_0$ : phương sai không thay đổi

$H_1$ : phương sai thay đổi

### Kết luận

**Phương pháp p-value**, với mức ý nghĩa 5%:

$p\text{-value} < 2.2 \times 10^{-16} < \alpha = 0.05$ : bác bỏ  $H_0$

$\Rightarrow$  Phương sai thay đổi

```
> # WLS
> wls_model <- lm(I(slquanao ~ thunhap
+ dihoc + binhthuong +
thunhap:binhthuong + dihoc:binhthuong)
, data = data, weights = thunhap)
> bptest(wls_model)
```

studentized Breusch-Pagan test

```
data: wls_model
BP = 103419, df = 5, p-value < 2.2e-16
```

## 4.2 Nếu có phương sai thay đổi thì dùng WLS và GLS để xử lý

### Giả thiết

$H_0$ : phương sai không thay đổi

$H_1$ : phương sai thay đổi

### Kết luận

**Phương pháp p-value**, với mức ý nghĩa 5%:

p-value = 0.241 >  $\alpha = 0.05$ : chấp nhận  $H_0$

⇒ Phương sai không đổi

```
> # GLS
> gls_model <- lm(slquanao/sqrt(thunhap)
~ 0 + I(1/sqrt(thunhap)) +
I(thunhap/sqrt(thunhap)) +
I(dihoc/sqrt(thunhap)) +
I(binhthuong/sqrt(thunhap)) +
I(thunhap:binhthuong/sqrt(thunhap)) +
I(dihoc:binhthuong/sqrt(thunhap)), data
= data )
> bptest(gls_model)
```

studentized Breusch-Pagan test

```
data: gls_model
BP = 6.7368, df = 5, p-value = 0.241
```

**05**

# Kiểm định Reset



## Dùng mô hình câu 3, thực hiện kiểm định RESET với $y^2, y^3$

```
> # RESET
> library("lmtest")
> resettest(hoiquy_cau3, power = 2:3, type = "fitted")
RESET test

data:  hoiquy_cau3
RESET = 3.4374, df1 = 2, df2 = 172, p-value = 0.03437
```

### Giả thiết

$H_0$ : mô hình hoiquy\_cau3 có dạng hàm đúng;

$H_1$ : mô hình hoiquy\_cau3 có dạng hàm sai

### Kết luận

**Phương pháp p-value**, với mức ý nghĩa 5%:

$p\text{-value} = 0.03437 < \alpha = 0.05$ : bác bỏ  $H_0$

$\Rightarrow$  mô hình hoiquy\_cau3 có dạng hàm sai

## Dùng mô hình câu 3, thực hiện kiểm định RESET với $y^2, y^3, y^4$

```
> # RESET
> library("lmtest")
> resettest(hoiquy_cau3, power = 2:4, type = "fitted")
RESET test

data:  hoiquy_cau3
RESET = 7.398, df1 = 3, df2 = 171, p-value = 0.000109
```

### Giả thiết

$H_0$ : mô hình hoiquy\_cau3 có dạng hàm đúng;

$H_1$ : mô hình hoiquy\_cau3 có dạng hàm sai

### Kết luận

**Phương pháp p-value**, với mức ý nghĩa 5%:

$p\text{-value} = 0.000109 < \alpha = 0.05$ : bác bỏ  $H_0$

$\Rightarrow$  mô hình hoiquy\_cau3 có dạng hàm sai

**06**

# **Chọn mô hình**

## Chạy full mô hình khảo sát

```
> full_model <- lm(slquanao ~ ., data=data)
> summary(full_model)
```

Call:

```
lm(formula = slquanao ~ ., data = data)
```

Residuals:

	Min	1Q	Median	3Q	Max
	-27.846	-6.137	-2.280	3.396	84.963

Coefficients:

	Estimate	Std. Error	t value	Pr(> t )
(Intercept)	8.5942328	8.1999506	1.048	0.2961
phucap	-0.0007327	0.0005689	-1.288	0.1996
thunhap	0.0003083	0.0004583	0.673	0.5022
dichoi	0.9723914	0.1683232	5.777	3.73e-08 ***
dihoc	-0.0343286	0.7220472	-0.048	0.9621
sanlong	0.0007741	0.0006656	1.163	0.2465
nam	-5.3044700	2.3060494	-2.300	0.0227 *
kakil	2.1785735	4.0642855	0.536	0.5927
jeans1	7.3782358	3.7699965	1.957	0.0520 .
cotton1	6.0427203	3.3266293	1.816	0.0711 .
thuongmai	0.8273485	4.4862734	0.184	0.8539
mxh	-3.8618894	3.3089262	-1.167	0.2449
ngoaiduong	-1.4594681	7.0999222	-0.206	0.8374
nguoithan	-4.5620335	5.0988137	-0.895	0.3722
tot	3.8346069	6.0168113	0.637	0.5248
binhthuong	-0.4384577	5.7344886	-0.076	0.9391

---

Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 13.66 on 164 degrees of freedom

Multiple R-squared: 0.29, Adjusted R-squared: 0.225

F-statistic: 4.465 on 15 and 164 DF, p-value: 4.92e-07

## Xem xét có nên loại bỏ biến nào ra khỏi mô hình không? Tại sao?

### Kiểm định đa cộng tuyến cho mô hình

```
> vif(full_model)
```

phucap	thunhap	dichoi	dihoc	sanlong
1.272048	1.268009	1.216042	1.101876	1.249674
nam	kaki	jeans	cotton	thuongmai
1.149487	2.031952	2.239286	2.667898	1.208236
mxh	ngoaiduong	nguoithan	tot	binhthuong
1.759284	1.313424	1.560697	7.007063	6.865295

Ta thấy tất cả các biến độc lập đều có  $vif < 10$  nên mô hình không có đa cộng tuyến nghiêm trọng. Vì vậy, ở bước này ta **chưa cần phải loại bỏ các biến ra khỏi mô hình**

## Dạng hàm phù hợp với số liệu? Tại sao?

Chọn mô hình hồi  
quy tuyến tính bằng  
phương pháp  
Bayesian Model  
Average (BMA)

```
> BMA_X = data[, -1]
> BMA_Y = data$slquanao
> library(BMA)
> BMA_Search = bicreg(BMA_X, BMA_Y, strict = FALSE, OR = 20)
> summary(BMA_Search)
```

Call:

```
bicreg(x = BMA_X, y = BMA_Y, strict = FALSE, OR = 20)
```

30 models were selected

Best 5 models (cumulative posterior probability = 0.5087 ):

	p!=0	EV	SD	model 1	model 2	model 3	model 4	model 5
Intercept	100.0	9.705e+00	2.4011224	9.962e+00	9.550e+00	9.237e+00	9.332e+00	1.290e+01
phucap	3.7	-1.744e-05	0.0001351	.	.	.	.	.
thunhap	9.9	6.000e-05	0.0002238	.	.	.	6.489e-04	.
dichoi	100.0	9.799e-01	0.1583196	1.003e+00	9.579e-01	9.426e-01	9.806e-01	9.376e-01
dihoc	1.7	1.314e-03	0.0909784	.	.	.	.	.
sanlong	23.1	2.555e-04	0.0005514	.	1.141e-03	.	.	.
nam	91.2	-5.714e+00	2.7229322	-6.376e+00	-6.223e+00	-6.108e+00	-6.479e+00	-6.452e+00
kakil	6.9	-2.719e-01	1.2579482	.	.	.	.	.
jeans1	10.9	4.350e-01	1.5487567	.	.	.	.	.
cotton1	4.9	1.194e-01	0.7572898	.	.	.	.	.
thuongmai	1.8	3.543e-02	0.6235141	.	.	.	.	.
mxh	1.9	-2.745e-02	0.4036573	.	.	.	.	.
ngoaiduong	1.7	-7.762e-03	0.8126315	.	.	.	.	.
nguoithan	1.7	-1.856e-02	0.5587193	.	.	.	.	.
tot	13.7	5.314e-01	1.5972297	.	.	3.879e+00	.	.
binhthuong	10.2	-3.515e-01	1.2710093	.	.	.	.	-3.560e+00
nVar				2	3	3	3	3
r2				0.228	0.243	0.240	0.239	0.239
BIC				-3.617e+01	-3.462e+01	-3.376e+01	-3.353e+01	-3.348e+01
post prob				0.222	0.103	0.067	0.059	0.058

## Nhận xét:

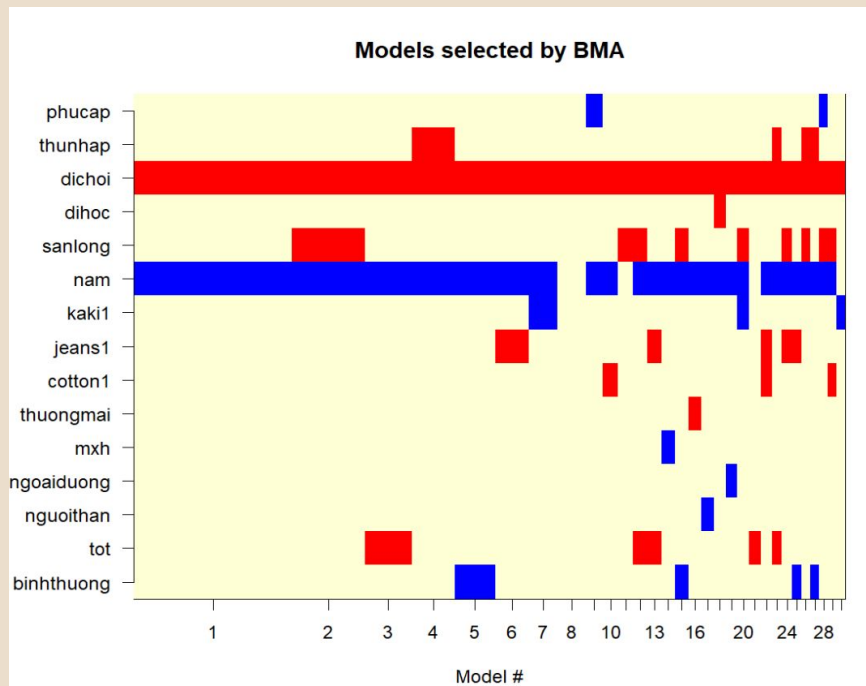
Ở kết quả trên, chúng ta có đến 5 lựa chọn. Thay vì chỉ có 1 mô hình để chúng ta lựa chọn thì bây giờ chúng ta có đến 5 lựa chọn so với các phương pháp cổ điển.

Như chúng ta thấy thì ở “Mô hình 1” (model 1) chúng thấy chỉ với 2 biến dichoi, nam chúng ta đã có thể giải thích được khoảng 22,8%.

Trong khi model 2, có 3 biến và giải thích được 24,3% nhưng tần suất xuất hiện của mô hình này chỉ có 10,3%, ít hơn một nửa so với tần suất xuất hiện của Mô hình 1 (22,2%).

Vì thế lựa chọn **Mô hình 1 là một sự lựa chọn tối ưu nhất** ở trường hợp này.

> imageplot.bma(BMA\_Search)



*Biểu đồ này thể hiện xác suất của biến đó xuất hiện trong mô hình hồi quy tuyến tính đa biến*

2 màu thể hiện dấu của hệ số hồi quy: “xanh” thể hiện cho dấu âm (-) và màu “đỏ” thể hiện cho dấu dương (+)

Vậy mô hình **tối ưu nhất** là **Mô hình 1**

$$slquanao = 9.962 - 6.376nam + 1.003dichoi$$

Các biến độc lập trong mô hình trên giải thích được **22,8%** cho biến phụ thuộc slquanao

Với tần suất xuất hiện là **22,2%**



**Kiểm định các biến còn lại (phucap, thunhap, dihoc, sanlong, kaki, jeans, cotton, thuongmai, mxh, ngoaiduong, nguoiathan, binhthuong, tot) xem có ý nghĩa đồng thời không?**

### Giả thiết

$H_0$ : các biến còn lại không có ý nghĩa đồng thời

$H_1$ : các biến còn lại có ý nghĩa đồng thời

### Kết luận

Phương pháp p-value, với mức ý nghĩa 5%:

p-value = 0.3598 >  $\alpha$  = 0.05: chấp nhận  $H_0$

⇒ Các biến còn lại không có ý nghĩa đồng thời với mức ý nghĩa 5%

Vậy mô hình cuối cùng đưa ra là:

**slquanao = 9.962 - 6.376nam + 1.003dichoi**

```
> linearHypothesis(full_model, c("phucap=0", "thunhap =0", "dihoc=0 " , "sanlong =0 " , "kaki = 0 " , " jeans =0 " , " cotton =0 " , "thuongmai =0 " , " mxh =0 " , "ngoaiduong =0 " , " nguoiathan =0 " , + "binhthuong=0","tot=0"))
```

Linear hypothesis test

Hypothesis:

```
phucap = 0  
thunhap = 0  
dihoc = 0  
sanlong = 0  
kaki = 0  
jeans = 0  
cotton = 0  
thuongmai = 0  
mxh = 0  
ngoaiduong = 0  
nguoiathan = 0  
binhthuong = 0  
tot = 0
```

Model 1: restricted model

Model 2: slquanao ~ nam + phucap + thunhap + dihoc + dichoi +  
sanlong +  
kaki + jeans + cotton + thuongmai + mxh + ngoaiduong +  
nguoiathan +  
tot + binhthuong

	Res.Df	RSS	Df	Sum of Sq	F	Pr(>F)
1	177	33272				
2	164	30597	13	2675.3	1.103	0.3598