# Model building

Level ~ chemicals\*

Chemicals\* ~ weathers\*

內容

[Model building 1](#_Toc103067612)

[level by logistic regression 2](#_Toc103067613)

[chemical by linear regression 4](#_Toc103067614)

[caffeine 4](#_Toc103067615)

[Gallic\_Acid 5](#_Toc103067616)

[Catechin 6](#_Toc103067617)

[EGC 7](#_Toc103067618)

[GCG 8](#_Toc103067619)

[EGCG 9](#_Toc103067620)

[Theanine 10](#_Toc103067621)

主要問題：模型不穩定。

# level by logistic regression

B C

51 87

B:C = 37 : 63

before log transform

forward

Step: AIC=118.71

level ~ Theanine+ Gallic\_Acid + Catechin + EGC + EGCG + GCG

backwards

Step: AIC=116.72

level ~ Theanine + Gallic\_Acid + Catechin + EGC + EGCG+ EC +

ECG

after log transform

forward

Step: AIC=106.31

level ~ caffeine + GCG + EGC + Catechin + Theanine + polyphenol

Step: AIC=101.96

level ~ Catechin + Theanine + FAA + EGC + caffeine

backward

Step: AIC=106.31

level ~ caffeine + GCG + EGC + Catechin + Theanine + polyphenol

Step: AIC=101.37

level ~ Theanine + caffeine + Catechin + EGC + GCG

問學姊農林資料的model需不需要分茶季、茶類。

Call:

glm(formula = level ~ polyphenol + Theanine + caffeine + Catechin +

EGC + GCG, family = binomial(), data = chem\_train)

Deviance Residuals:

Min 1Q Median 3Q Max

-1.8241 -0.6998 0.2369 0.6546 2.5296

Coefficients:

Estimate Std. Error z value Pr(>|z|)

(Intercept) 9.0581 4.1701 2.172 0.02984 \*

polyphenol -1.5981 1.0997 -1.453 0.14619

Theanine -0.7191 0.4049 -1.776 0.07576 .

caffeine -4.1826 1.5063 -2.777 0.00549 \*\*

Catechin -0.8424 0.3240 -2.600 0.00934 \*\*

EGC 3.1771 0.9880 3.216 0.00130 \*\*

GCG 2.0244 0.9539 2.122 0.03382 \*

---

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

(Dispersion parameter for binomial family taken to be 1)

Null deviance: 144.342 on 107 degrees of freedom

Residual deviance: 92.309 on 101 degrees of freedom

AIC: 106.31

Number of Fisher Scoring iterations: 5

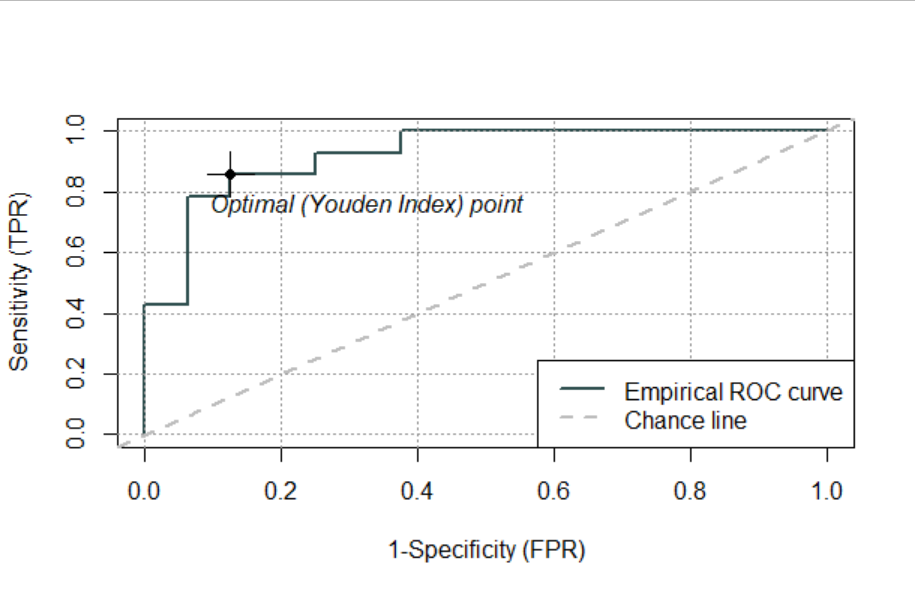
ytest

y\_pred 0 1

B 8 6

C 1 15

23/30 = 76%



Area under curve: 0.9241

# chemical by linear regression



## caffeine

forward

Start: AIC=490.25

caffeine ~ RH + Solar\_rad\_H

**lm(formula = caffeine ~ RH + Solar\_rad\_H, data = weather\_train)**

Residuals:

Min 1Q Median 3Q Max

-16.772 -7.039 -1.550 5.434 37.300

Coefficients:

Estimate Std. Error t value Pr(>|t|)

(Intercept) 151.65142 27.47178 5.520 2.46e-07 \*\*\*

RH -1.53654 0.32146 -4.780 5.73e-06 \*\*\*

Solar\_rad\_H -0.02861 0.01561 -1.833 0.0696 .

---

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

Residual standard error: 9.685 on 105 degrees of freedom

Multiple R-squared: 0.1802, Adjusted R-squared: 0.1646

F-statistic: 11.54 on 2 and 105 DF, p-value: 2.945e-05

backward

Step: AIC=496.67

Call:

**lm(formula = caffeine ~ acu\_mean\_temp + rain + temp\_differ +**

**RH + Solar\_rad\_H + Solar\_rad\_MJM2, data = weather\_train)**

Residuals:

Min 1Q Median 3Q Max

-14.405 -7.253 -1.732 5.663 33.805

Coefficients:

Estimate Std. Error t value Pr(>|t|)

(Intercept) 263.137130 49.252148 5.343 5.68e-07 \*\*\*

acu\_mean\_temp 0.033122 0.011210 2.955 0.0039 \*\*

rain -0.008324 0.003405 -2.445 0.0162 \*

temp\_differ 0.019374 0.010769 1.799 0.0750 .

RH -3.006508 0.644220 -4.667 9.41e-06 \*\*\*

Solar\_rad\_H -0.105895 0.031245 -3.389 0.0010 \*\*

Solar\_rad\_MJM2 -0.037652 0.016266 -2.315 0.0226 \*

---

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

Residual standard error: 9.308 on 101 degrees of freedom

Multiple R-squared: 0.2717, Adjusted R-squared: 0.2285

F-statistic: 6.281 on 6 and 101 DF, p-value: 1.231e-05

## Gallic\_Acid

lm(formula = Gallic\_Acid ~ Solar\_rad\_H + temp\_differ + RH, data = weather\_train)

Residuals:

Min 1Q Median 3Q Max

-0.78911 -0.24021 -0.06789 0.16864 1.91400

Coefficients:

Estimate Std. Error t value Pr(>|t|)

(Intercept) -2.1157698 1.1812792 -1.791 0.076189 .

Solar\_rad\_H 0.0030712 0.0008465 3.628 0.000444 \*\*\*

temp\_differ -0.0010001 0.0003762 -2.659 0.009082 \*\*

RH 0.0354698 0.0137374 2.582 0.011215 \*

---

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

Residual standard error: 0.4026 on 104 degrees of freedom

Multiple R-squared: 0.127, Adjusted R-squared: 0.1018

F-statistic: 5.044 on 3 and 104 DF, p-value: 0.002653

## Catechin

lm(formula = Catechin ~ RH + Solar\_rad\_H + Solar\_rad\_MJM2 + rain +

temp\_differ, data = weather\_train)

Residuals:

Min 1Q Median 3Q Max

-2.4172 -0.7220 -0.0113 0.5039 4.3023

Coefficients:

Estimate Std. Error t value Pr(>|t|)

(Intercept) 18.9948998 3.3184061 5.724 1.05e-07 \*\*\*

RH -0.2153626 0.0397867 -5.413 4.13e-07 \*\*\*

Solar\_rad\_H -0.0098165 0.0029743 -3.300 0.001331 \*\*

Solar\_rad\_MJM2 0.0052569 0.0013927 3.775 0.000269 \*\*\*

rain -0.0008650 0.0003726 -2.322 0.022224 \*

temp\_differ -0.0026752 0.0012656 -2.114 0.036967 \*

---

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

Residual standard error: 1.099 on 102 degrees of freedom

Multiple R-squared: 0.4063, Adjusted R-squared: 0.3772

F-statistic: 13.96 on 5 and 102 DF, p-value: 2.173e-10

## EGC

lm(formula = EGC ~ acu\_mean\_temp + rain + growth\_mean\_temp +

RH + Solar\_rad\_H + Solar\_rad\_MJM2, data = weather\_train)

Residuals:

Min 1Q Median 3Q Max

-17.6125 -4.3931 -0.8993 4.7922 21.0292

Coefficients:

Estimate Std. Error t value Pr(>|t|)

(Intercept) 233.155244 55.001179 4.239 4.97e-05 \*\*\*

acu\_mean\_temp 0.026307 0.010828 2.429 0.016889 \*

rain -0.007510 0.002914 -2.577 0.011401 \*

growth\_mean\_temp -2.604722 0.737735 -3.531 0.000626 \*\*\*

RH -1.868409 0.568092 -3.289 0.001385 \*\*

Solar\_rad\_H -0.145529 0.035688 -4.078 9.08e-05 \*\*\*

Solar\_rad\_MJM2 0.018433 0.011963 1.541 0.126466

## GCG

lm(formula = GCG ~ acu\_mean\_temp + RH + Solar\_rad\_MJM2, data = weather\_train)

Residuals:

Min 1Q Median 3Q Max

-4.1601 -2.1418 -0.3065 1.6212 7.4135

Coefficients:

Estimate Std. Error t value Pr(>|t|)

(Intercept) 53.575203 10.986663 4.876 3.90e-06 \*\*\*

acu\_mean\_temp 0.008746 0.002499 3.500 0.000687 \*\*\*

RH -0.579056 0.144022 -4.021 0.000110 \*\*\*

Solar\_rad\_MJM2 -0.020046 0.004091 -4.900 3.54e-06 \*\*\*

---

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

Residual standard error: 2.691 on 104 degrees of freedom

Multiple R-squared: 0.2227, Adjusted R-squared: 0.2003

F-statistic: 9.933 on 3 and 104 DF, p-value: 8.157e-06

## EGCG

lm(formula = EGCG ~ growth\_mean\_temp + RH + Solar\_rad\_H, data = weather\_train)

Residuals:

Min 1Q Median 3Q Max

-39.070 -14.316 -6.851 15.379 66.940

Coefficients:

Estimate Std. Error t value Pr(>|t|)

(Intercept) 430.81914 74.57782 5.777 8.00e-08 \*\*\*

growth\_mean\_temp -2.44028 1.21693 -2.005 0.04753 \*

RH -3.48368 0.72817 -4.784 5.69e-06 \*\*\*

Solar\_rad\_H -0.17828 0.05401 -3.301 0.00132 \*\*

---

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

Residual standard error: 20.54 on 104 degrees of freedom

Multiple R-squared: 0.2146, Adjusted R-squared: 0.1919

F-statistic: 9.472 on 3 and 104 DF, p-value: 1.376e-05

## Theanine

forward

lm(formula = Theanine ~ rain + Solar\_rad\_MJM2 + Solar\_rad\_H +

growth\_mean\_temp, data = weather\_train)

Residuals:

Min 1Q Median 3Q Max

-9.7835 -1.7656 -0.0129 1.7056 8.3451

Coefficients:

Estimate Std. Error t value Pr(>|t|)

(Intercept) 16.849159 7.988142 2.109 0.0373 \*

rain 0.001451 0.001169 1.242 0.2172

Solar\_rad\_MJM2 0.026744 0.004444 6.018 2.74e-08 \*\*\*

Solar\_rad\_H -0.057876 0.011345 -5.101 1.54e-06 \*\*\*

growth\_mean\_temp -0.508511 0.289972 -1.754 0.0825 .

---

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

Residual standard error: 3.737 on 103 degrees of freedom

Multiple R-squared: 0.4181, Adjusted R-squared: 0.3955

F-statistic: 18.5 on 4 and 103 DF, p-value: 1.751e-11