

FASTFOOD NUTRITION

ที่มาความสำคัญ
อาหารfastfood เป็นที่นิยมสำหรับผู้บริโภคที่ต้องการความรวดเร็ว และอาหารfastfood ต่างให้พลังงานที่สูงมาก ผู้จัดทำจึงสนใจศึกษาปัจจัยที่มีผลต่อ แคลอรี ของอาหารfastfood โดยศึกษาจากปัจจัยเหล่านี้ โปรตีน,คาร์โบไฮเดรต,โซเดียม,คอเลสเตอรอล เพื่อเป็นประโยชน์ต่อผู้บริโภค

วัตถุประสงค์
เพื่อศึกษาการวิเคราะห์ ความเหมาะสมของตัวแบบ ความสัมพันธ์เส้นตรง การแจกแจงปกติ ความเป็นอิสระ และความสัมพันธ์เชิงเส้นแบบพหุของปัจจัยที่มีผลต่อปริมาณแคลอรี

สิ่งที่คาดว่าจะได้รับ

- ได้นำทฤษฎีจากการเรียนรู้มาประยุกต์ใช้จริงโดยการสืบค้นข้อมูลเพื่อมาใช้ในการพยากรณ์และวิเคราะห์ข้อมูลที่ต้องการทราบได้
- ได้ทราบว่า คอเลสเตอรอล โซเดียม คาร์โบไฮเดรต และโปรตีนในอาหารฟาสฟู้ดมีผลต่อปริมาณแคลอรีอย่างไร โดยอิงจากฐานข้อมูลที่ได้รับ

ค่าสถิติเบื้องต้น

X1 ปริมาณคอเลสเตอรอล X4 ปริมาณโปรตีน
X2 ปริมาณโซเดียม Y ปริมาณแคลอรี
X3 ปริมาณคาโบไฮเดรต
จำนวนข้อมูล 60 ชุด

Linearlity Test

H0 : X และ Y มีความสัมพันธ์กัน
H1 : X และ Y ไม่มีความสัมพันธ์กัน

Pearson’s product-moment correlation

data: X1 and Y
t = 12.601, df = 58, p-value < 2.2e-16
alternative hypothesis: true correlation is not equal to 0
95 percent confidence interval:
0.7690368 0.9116441
sample estimates:
cor
0.855836

data: X2 and Y
t = 13.606, df = 58, p-value < 2.2e-16
alternative hypothesis: true correlation is not equal to 0
95 percent confidence interval:
0.7947955 0.9221977
sample estimates:
cor
0.8726114

data: X3 and Y
t = 16.591, df = 58, p-value < 2.2e-16
alternative hypothesis: true correlation is not equal to 0
95 percent confidence interval:
0.8513699 0.9447293
sample estimates:
cor
0.908822

data: X4 and Y
t = 12.493, df = 58, p-value < 2.2e-16
alternative hypothesis: true correlation is not equal to 0
95 percent confidence interval:
0.7660117 0.9103922
sample estimates:
cor
0.8538539

เนื่องจาก p-value ของ X1,X2,X3,X4 น้อยกว่า 0.05(α) จึงยอมรับ H0
ดังนั้น ตัวแปร Y กับ X1,X2,X3,X4 มีความสัมพันธ์กันเชิงเส้นตรง

Multicollinearlity test

> vif(RegModel.2)

X1X2X3X4

9.33348810.1170974.03241916.559908

เนื่องจากค่า vif ของตัวแปร X แต่ละตัวมีค่าน้อยกว่า 10 ดังนั้น ตัวแปรอิสระทุกตัวไม่มีความสัมพันธ์เชิงพหุ

Normality test

Anderson - darling test

H0 : Y มีการแจกแจงปกติ
H1 : Y ไม่มีการแจกแจงปกติ

Anderson-Darling normality test

data: residuals.RegModel.2
A = 1.1429, p-value = 0.005033

เนื่องจาก p value เท่ากับ 0.005033 ซึ่งน้อยกว่า 0.05(α) จึงทำให้ยอมรับ H0
ดังนั้น Y ไม่มีการแจกแจงปกติ
จึงทำการแก้BOXCOX
โดยทำการแปลงข้อมูล Y เป็น Y^2

หลังจากแปลงข้อมูลYเป็น Y^2

Independent of error test

H0 : ส่วนเหลือไม่มีความสัมพันธ์กันเอง $\rho = 0$
H1 : ส่วนเหลือมีความสัมพันธ์กันเอง $\rho \neq 0$

Durbin-Watson test

data: y1 ~ X1 + X2 + X3 + X4
DW = 1.6256, p-value = 0.1031
alternative hypothesis: true autocorrelation is not 0

Breusch-Pagan test

data: y1 ~ X1 + X2 + X3 + X4
BP = 1.3797, df = 1, p-value = 0.2402

เนื่องจาก p value เท่ากับ 0.1031 ซึ่งมากกว่า 0.05(α) จึงยอมรับ H0
ดังนั้นส่วนเหลือไม่มีความสัมพันธ์กัน

เนื่องจาก p-value เท่ากับ 0.2402 ซึ่งมากกว่า 0.05(α) จึงทำให้ยอมรับ H0 ดังนั้น ส่วนเหลือมีความแปรปรวนของค่าคลาดเคลื่อนคงที่

Normality Test

Anderson-Darling normality test

data: residuals.RegModel.4
A = 0.41258, p-value = 0.3289

H0: Y มีการแจกแจงปกติ
H1: Y ไม่มีการแจกแจงปกติ
เนื่องจาก p value เท่ากับ 0.3289 ซึ่งมากกว่า 0.05(α) จึงทำให้ยอมรับ H0
ดังนั้น Y มีการแจกแจงปกติ

การทดสอบเกณฑ์การเลือกตัวแบบที่เหมาะสมที่สุด

จากตาราง ตัวแบบประกอบด้วย X1,X2,X3,X4 เป็นตัวแบบที่ดีที่สุด เนื่องจากมีค่า Rsqr และ adjRsqr สูงที่สุดและ ค่า CP มีขนาดเล็กและมีค่าใกล้เคียงจำนวนตัวแปรอิสระทั้งหมด

> cbind(sum.all\$rsq, sum.all\$adjr2, sum.all\$cp, sum.all\$outmat)

X1X2X3X4

1(1)"0.520985712589225" "0.512726845564901" "163.004450492508" "***" " " " " " " "

1(2)"0.449494181597063" "0.440002701969426" "195.690246869141" " " " " " " " " "

1(3)"0.347897123381106" "0.336653970335953" "242.140234877871" " " " " " " " " "

1(4)"0.134245513846708" "0.119318712361306" "339.821357492896" " " " " " " " " "

2(1)"0.683526600487169" "0.672422270679701" "90.691055731221" " " " " " " " " "

2(2)"0.598827473384736" "0.584751244380692" "129.415340738521" " " " " " " " " "

2(3)"0.562370805360915" "0.547015395022701" "146.083262253976" " " " " " " " " "

2(4)"0.494852830369518" "0.47712836827722" "176.952356141102" " " " " " " " " "

3(1)"0.85568074699592" "0.84794935844213" "13.9824968279991" " " " " " " " " "

3(2)"0.736706071479589" "0.722601039594567" "68.3774994798394" " " " " " " " " "

3(3)"0.683628439284348" "0.666679962817438" "92.6444952206031" " " " " " " " " "

3(4)"0.503544731560954" "0.476948913608862" "174.97843491538" " " " " " " " " "

4(1)"0.879702052865387" "0.870953111255597" "5" " " " " " " " " "

สรุปผล

Coefficients:

EstimateStd. Errort valuePr(>|t|)

(Intercept)-220573.5351048.51-4.3216.56e-05***

X18036.71612.8313.114< 2e-16***

X2147.9344.643.3140.00163**

X38854.231095.068.0866.33e-11***

X4-23365.742467.84-9.4683.81e-13***

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

Residual standard error: 94530 on 55 degrees of freedom
Multiple R-squared: 0.8797, Adjusted R-squared: 0.871
F-statistic: 100.5 on 4 and 55 DF, p-value: < 2.2e-16

จากการวิเคราะห์จะได้สมการคือ $\hat{Y} = -220573.53 + 8036.71X1 + 147.93X2 + 8854X3 - 23365.74X4$ กล่าวคือ
เมื่อปริมาณคอเลสเตอรอลเพิ่มขึ้นหนึ่งหน่วย ทำให้ปริมาณแคลอรีเพิ่มขึ้น 8036.71 หน่วย
เมื่อปริมาณโซเดียม เพิ่มขึ้นหนึ่งหน่วย ทำให้ปริมาณแคลอรีเพิ่มขึ้น 147.93 หน่วย
เมื่อปริมาณคาโบไฮเดรตเพิ่มขึ้นหนึ่งหน่วย ทำให้ปริมาณแคลอรีเพิ่มขึ้น 8854.23หน่วย และ
เมื่อปริมาณโปรตีนเพิ่มขึ้นหนึ่งหน่วย ทำให้ปริมาณแคลอรีลดลง 23365.74 หน่วย