DETERMINANTS OF THE PRESENCE OF BROWN FAT IN HUMAN

Yuhong Shao & Fangshu Yu & Hanghuan Yang & Shengbin Huang

2021/4/10

Casestudy Description:

This case study is to identify the factors determining the existence of brown fat in Humans. We looked at a dataset with 20 possible predictors and used R to build a predictive model that best explains the existence of brown fat in Humans.

Data Source:

Molecular Imaging Center at The University of Sherbrooke for providing the data.

Group Memember and Job Decription:

Shengbin Huang 1004740961:

1) Investigate the relationship between each of the covariates and the presence of brown fat (First half of the dataset).
2) Further analyze the model by building classification table.
3) Make presentation slides.

Yuhong Shao 1004731917:

1) Investigate the relationship between each of the covariates and the presence of brown fat (Second half of the dataset).
2) Check outliers of the model.
3) Making presentation slides.

Fangshu Yu 1002924151:

1) Build preliminary models and make simple optimization(by using the variables that selected by first group member).
2) Further analyze the model by using Hosmer-Lemeshow test.
3) Make presentation slides and add a conclusion for this report.

Hanghuan Yang 1004705768:

1) Build preliminary models and make simple optimization(by using the variables that selected by first group member).
2) Further analyze the model by using ROC curve.
3) Write this report and find additional references(For instance, some science reports relating to the brown fat).

1. Background and objective of this study:

Brown fat, also known as brown adipose tissue (BAT) is one of the two adipose tissue in human body. In contrast to white fat which stores excess energy and causes obesity, brown fat aids the process of glucose-heat transformation, leading to higher cold resistance (Brazier). There is huge potential in brown fat study, recent research show that brown fat affect cancer progression (Chu, Bos, Gill, Torriani, & Bredella, 2019), therefore knowledge on the association between cancer and brown fat may aid cancer therapy. In addition, the calorie-burning property of brown fat suggests that turning white fat into brown fat may lead to a solution to obesity (Wada et al., 2016), which requires more knowledge on the generation mechanism of brown fat. This study intends to find out the effect of factors on the existence of brown using general linear

model regression, based on data provided by Molecular Imaging Center at The University of Sherbrooke.

2. Statistical analysis of covariates in the dataset

First, we test relationship between each covariate and the presence of brown fat separately by using different tests. Then, we may have a basic understanding of which is covariate is valid when we construct the model.

(1) ***Sex:***

- Gender of each sample data. In the dataset. Females are labeled as 1, Males are labeled as 2.
- Since P-value in both Chi-square Test and Fisher's Exact Test is quite small, we may conclude that the relationship between "Sex" and "the presence of brown fat" are significant. Estimated odds-ratio is around 0.30, the odds of the presence of brown fat for females estimated to be about 0.3 times the odds for males.

(2) ***Diabetes:***

- The presence of diabetes in each sample data. Diabetes is a binary indicator of whether the observed has diabetes.
- Since P-value in both Chi-square Test and Fisher's Exact Test is very small, we may conclude that "Diabetes" is significant. The Estimated odds-ratio is around 0.14, the odds of the presence of brown fat for having diabetes is estimated to be about 0.14 times the odds for those who haven't.

(3) ***Age:***

- Age of the sample.
- Since P-value in both Wald Test and Likelihood Ratio Test is quite small, we may conclude that "Age" is significant. Unlike other covariates, "Age" and "the presence of brown fat" have negative relationship. The Estimated odds-ratio is around 0.96, the odds of the presence of brown fat is estimated to be decreased a factor of 0.96 for each 1 unit increase of age.

(4) ***Day:***

- The day that the sample data was collected. In the dataset, it ranges between 1 to 365.
- Since P-value in both Wald Test and Likelihood Ratio Test is greater than 0.05, we may conclude that the relationship between "Day" and "the presence of brown fat" are not significant, they are likely independent.

(5) ***Month:***

- The month that the sample data was collected. In the dataset, it ranges between 1 to 12.
- Since P-values for each months are greater than 0.05 and By Fisher's Exact Test, we may conclude that the relationship between "Month" and "the presence of brown fat" are not significant, they are likely independent.

(6) ***Temperature:***

- In the dataset, we have in total 5 variables that are relevant to the temperature, which are External Temperature, External Temperature in 2 days, 3 days, 1 week and 1 month (measured in °C).
- We want to find whether the presence of brown fat is related to the temperature. After evaluating the data, we noticed the p-value for External Temperature is quite small, therefore, we can conclude that the external temperature have significant effect on the presence of brown fat. Then, we compared the p-value between the data of temperature in 2 days, 3 days, 1 week and 1 month, we found that the shorter the time span of measuring temperature, the greater the influence on the presence or absence of brown fat. Therefore, the data of External Temperature in 2 days effect the brown fat most and in 1 month least.

(7) ***Season:***

- The season that the sample data was collected. In the dataset, it ranges between 1 to 4 and represent Spring, Summer, Autumn and Winter respectively.
- Since p-value in both Wald Test and Likelihood Ratio Test is greater than 0.05, we can conclude that the relationship between "Season" and "the presence of brown fat" are not significant, they are likely independent.

(8) ***Duration of Sunshine:***

- The duration of sunshine per day (measured in minute).
- Since P-value in both Wald Test and Likelihood Ratio Test is 0.002($<$0.05), we can conclude that the relationship between "Duration Sunshine" and "the presence of brown fat" are significant. The estimated odds-ratio is round 0.9987, the odds of the presence of brown fat is estimated to be decreased a factor of 0.9987 for each 1 minute increase the sunshine duration.

(9) ***Weight:***

- The weight of each sample data (measured in kilogram).
- Since P-value in both Wald Test and Likelihood Ratio Test is small($<$0.05), we can conclude that the relationship between "Weight" and "the presence of brown fat" are significant. The estimated odds-ratio is round 0.97, the odds of the presence of brown fat is estimated to be decreased a factor of 0.97 for each 1 kg increase weight.

(10) ***Size:***

- The size of each sample data (measured in centimeter).
- Since p-value in both Wald Test and Likelihood Ratio Test is small($<$0.05), we can conclude that the relationship between "Size" and "the presence of brown fat" are significant.
- Estimated odds-ratio is round 0.97, the odds of the presence of brown fat is estimated to be decreased a factor of 0.97 for each 1 cm increase weight which is similar to the effect of weight we showed above.

(11) ***Body Mass Index:***

- BMI stands for "Body Mass index", it is a commonly used international standard to measure the degree of body fat and thinness and whether it is healthy.

- Since p-value in both Wald Test and Likelihood Ratio Test is quite small(<0.05), we can conclude that the relationship between "BMI" and "the presence of brown fat" are significant. The estimated odds-ratio is round 0.92, the odds of the presence of brown fat is estimated to be decreased a factor of 0.92 for each 1 unit increase of BMI.

(12) ***Glycemia:***

- Glycemia indicates the level of blood sugar concentration.
- Since p-value in both Wald Test and Likelihood Ratio Test is quite small(<0.05), we can conclude that the relationship between "Glycemy" and "the presence of brown fat" are significant. The estimated odds-ratio is round 0.75, the odds of the presence of brown fat is estimated to be decreased a factor of 0.75 for each 1 unit increase of Glycemia.

(13) ***Lean Body Weight:***

- LBW stands for "Lean Body Weight". LBW shows the percentage of weight without the fat.
- Since p-value in both Wald Test and Likelihood Ratio Test is quite small(<0.05), we can conclude that the relationship between "LBW" and "the presence of brown fat" are significant. The estimated odds-ratio is 0.95, the odds of the presence of brown fat is estimated to be decreased a factor of 0.95 for each 1 unit increase of LBW.

(14) ***Cancer Status:***

- The presence of cancer in each sample. Cancer status is a binary indicator of whether the sample has cancer.
- We got estimated odds ratio is 0.877 and confidence interval for odds ratio is (0.697, 1.104).Since the estimated odds ratio is closed to 1. We can conclude that there is no significant relationship between cancer status and brown fat, they are likely independent. In likelihood ratio test, p-value is greater that 0.05 which confirms this conclusion.

(15) ***Cancer Type:***

- Cancer type is a categorical variable that indicates the type of cancer the observed have. There are 17 types of cancer in this dataset, all other types of cancer fits in one category "other".
- Note that there are N/A values in cancer type, we omit them while performing independence tests and further model building.
- We fit data into poisson model, then perform Wald test on coefficients to see whether there is significant relationship between certain cancer type and existence of brown fat. There is strong evidence that digestive, breast and non-Hodgkin lymphoma cancer significantly related to brown fat. The odds of brown fat existing for people with cancers mentioned above is lower than people without cancer.

(16) ***TSH:***

- TSH is a continuous variable that measures the thyroid-stimulating hormone level in blood.
- Note that there are N/A values in TSH, we omit them while performing independence tests and further model building.
- We fit data into logistic model, then perform Wald test. The p-value of TSH is 0.876, which is greater than significant level 5%, indicating there is no significant relationship between TSH and the existence of brown fat.

Summary

After removing these insignificant vaiarables in previous pages, we are able to get these following covariants remaining:
Sex, Diabetes, Age, Temperature (External Temperature, External Temperature in 2 days, 3 days, 1 week and 1 month), Duration of Sunshine, Weight, Size, BMI, Glycemia, LBW, Cancer Status and types of Cancer(digestive, breast and non-Hodgkin lymphoma cancer with index 2, 4, 13).
We now have developed a basic understanding of what factors may affect sbp. The results above could also act as a reference when we start bulding model in Part 3.

Operations on Data

In order to build a more accurate model, we need to further reduce number of the available covariates.

1. Dealing with excessive available parameters for expressing temperature:

- We noticed that there are a total of five parameters used to express covariate temperature, so many related parameters will undoubtedly affect the accuracy of the final model construction.

- We finally decided to choose only two parameters to represent the covariate temperature, one of which is the external temperature of the day, and the other is used to represent a recent temperature trend.

- From what we got in the previous part, we know that the shorter the time span of measuring temperature, the greater the influence on the presence or absence of brown fat. As a result, we chose the parameter "External Temperature in 2 Days" to represent the recent temperature trend.

2. Dealing with covariates with high multicollinearity:
We got the following graph by using R code which indicates all the correlationships between 2 covariates. If multiple covariates share a high correlationship, we may only choose one of them when we are building the model:

From the graph, we decided to:

- Choose "Weight" among "Weight", "BMI" and "LBW"(due to high corelationship).
- Choose "Duration Sunshine" among "External Temperature", "Average Temperature" and "Duration Sunshine"(due to high corelationship).
- This step further reduce the number of covariates and it may help us constructing the model.

3. Model Selection:

To build the model, we first remove the missing data in Cancer_Status and Cancer_Type.

Positivity of Brown Fat is 6.774%, close to 5%. And the mean and variance of Borwn Fat is similar, which indicates that we should probably use poisson model.

We fitted all main effects (after multicollinearity deletion) in to Poisson model and perform forward and backward deletion. And we find the main effect model deleted with backward direction is better than the model with forward direction, so we choose the main effect deleted with backward direction.

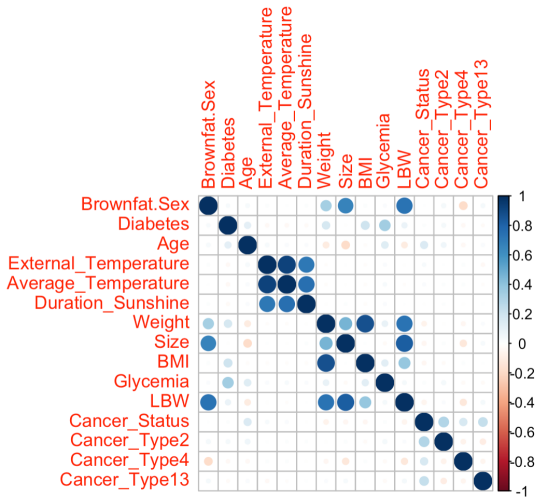


Figure 1: Correlationships between all the covariates

Then we fitted all interaction effects into the chosen poisson model and also perform forward and backward deletion. Then we got 2 models with same AIC value, but the model with backward direction has smaller Residual Deviance, so we chose that one.

After observing the chosen model summary, some coefficients have pretty big p-values Therefore we delete some coefficients which has p-value greater than 0.2 in Wald test.

```
# the fitted model after deleting large p-value coefficients
fit.poi <- glm(formula = BrownFat ~ Weight + Diabetes + Temperature_2Days
               + as.numeric(Cancer_Type == 13) + Size + Age:factor(Sex)
               + Age:Weight + factor(Sex):Diabetes + Weight:Size + Diabetes:Size
               + Size:External_Temperature,
               family = poisson(), data = data)
summary(fit.poi)
```

```
##
## Call:
## glm(formula = BrownFat ~ Weight + Diabetes + Temperature_2Days +
##     as.numeric(Cancer_Type == 13) + Size + Age:factor(Sex) +
##     Age:Weight + factor(Sex):Diabetes + Weight:Size + Diabetes:Size +
##     Size:External_Temperature, family = poisson(), data = data)
##
## Deviance Residuals:
##     Min       1Q   Median       3Q      Max
## -1.0331  -0.4231  -0.2736  -0.1688   3.0587
##
## Coefficients:
##                                  Estimate Std. Error z value Pr(>|z|)
## (Intercept)                    -7.351e+00  4.760e+00  -1.544  0.12249
## Weight                          8.967e-02  6.968e-02   1.287  0.19813
## Diabetes                       -1.969e+01  1.042e+01  -1.889  0.05890 .
## Temperature_2Days               1.699e-02  1.436e-02   1.183  0.23680
## as.numeric(Cancer_Type == 13) -5.572e-01  3.078e-01  -1.810  0.07027 .
## Size                            3.942e-02  2.738e-02   1.439  0.15004
## Age:factor(Sex)1                4.468e-03  1.617e-02   0.276  0.78231
```

```
## Age:factor(Sex)2             -1.207e-02  1.676e-02  -0.720  0.47155
## Weight:Age                   -4.880e-04  2.478e-04  -1.970  0.04888 *
## Diabetes:factor(Sex)2        -1.974e+00  1.373e+00  -1.437  0.15059
## Weight:Size                  -5.050e-04  3.916e-04  -1.289  0.19723
## Diabetes:Size                 1.151e-01  6.395e-02   1.799  0.07195 .
## Size:External_Temperature    -2.041e-04  7.862e-05  -2.596  0.00942 **
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## (Dispersion parameter for poisson family taken to be 1)
##
##     Null deviance: 1651.6  on 4472  degrees of freedom
## Residual deviance: 1394.0  on 4460  degrees of freedom
## AIC: 2038
##
## Number of Fisher Scoring iterations: 7
```

4. Model Diagnosis:

4.1 Goodness of fit test

Perform Hosmer-Lemeshow test with group number of 20.
p-value = 0.5594 > 5%, fail to reject H_0. Therefore, model is significant.

```
## ResourceSelection 0.3-5   2019-07-22

##
##  Hosmer and Lemeshow goodness of fit (GOF) test
##
## data:  fit.poi$y, fit.poi$fitted.values
## X-squared = 16.476, df = 18, p-value = 0.5594
```

4.2 Classification table

sensitivity = 0.802589, specificity = 0.6409702
Both values are pretty high, indicating model is a good fit.

```
##              predict
## data$BrownFat    0    1
##             0 2669 1495
##             1   61  248
```

4.3 ROC curve

Plot ROC curve. Concordance index is 0.774 greater than 0.5, we may argue that the model that we construct is relatively quite accurate.

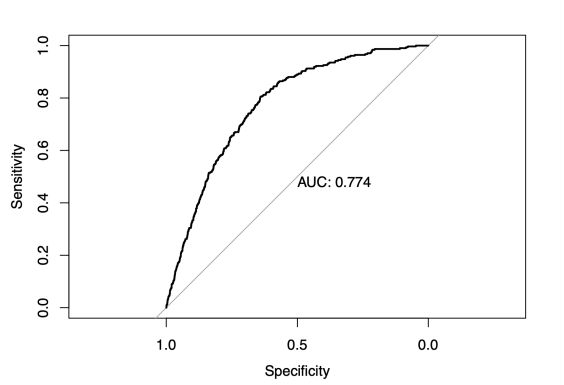


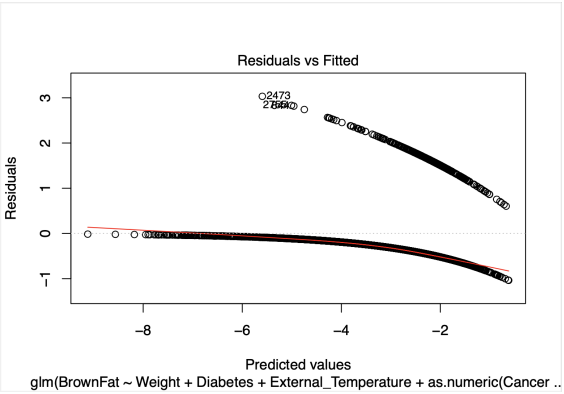
Figure 2: ROC curve

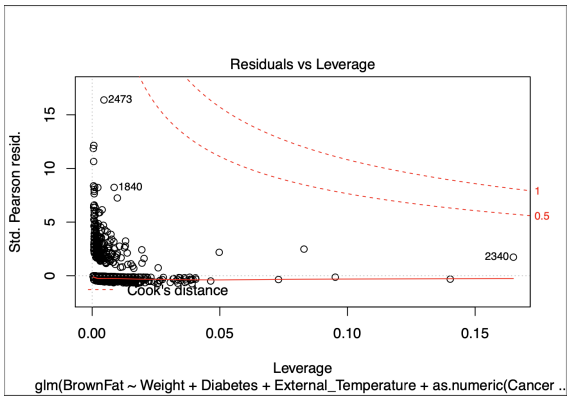
4.4 Validation

We randomly sampled 3500 observations for training, and 1342 observations for validation.
Train error = 34.46%, Validation error = 32.37% Since train error is similar to validation error, there is no sign of over-fitting.

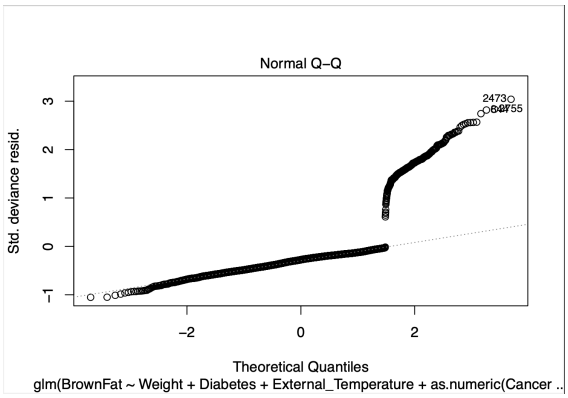
4.5 Residual Analysis

4.5.1 Outlier and Influential Points From the residual-fitted plot we observed that there are three outliers, namely observation 2473, 2755, 844. Proceed to check residual-leverage plot, non of the outliers are influential point.





5.5.2 Check Model Assumption We also checked normal Q-Q plot at the end, and we observed that the normality assumption is violated, therefore we may need to perform box-cox transformation to derive a better model.



Main effects			Interaction		
Factor	effect		Factor	effect	
Brown Fat(Y)	↓	-7.35	Age:Female	↑	0.00447
			Diabetes:Male	↓	-1.974
Cancer_Type13	↓	-0.557			
Diabetes	↓	-0.197			
			Diabetes:Size	↑	0.151
Weight	↑	0.0897			
Size	↑	0.0394			
Temp_2Days	↑	0.017			
			Age:Male	↓	-0.012
			Weight:Size	↑	0.000505
			Weight:Age	↓	-0.00049
			Size:Temp_Ext	↓	-0.0002

5. Conclusion:

The goal of this study is to derive a model based on the data from Molecular Imaging Center by identifying which factors determine the presence of brown fat in Humans. We found that the Temperature (includes duration of sunshine, external temperature, temperature in 2 days, 3 days, 7 days and 1 month), Age, Sex, Weight, Diabetes and Cancer Types have significant impact on the existence of brown fat, especially for temperature and Weight. These findings help us to learn more about how the brown fat plays a role in mitigating the deleterious effects of obesity and its benefits for cancer patients. However, as scientists mentioned, brown fat still has many unexplored benefits for humans, so our model is not completed, which may need to consider more factors in the future. Besides, we can enlarge our dataset, like using data from different hospitals or centers, as the prevalence of brown fat in current data is quite low.

Reference List

[1] Wein, H. (2019, September 17). How brown fat improves metabolism. National Institutes of Health (NIH). https://www.nih.gov/news-events/nih-research-matters/how-brown-fat-improves-metabolism

[2] Tome ME, McNabb FM, Gwazdauskas FC. Adrenal responses to chronic and acute water stress in Japanese quail Coturnix japonica. Comp Biochem Physiol A Comp Physiol. 1985;81(1):171-9. doi: 10.1016/0300-9629(85)90284-1. PMID: 2859951.

[3] Becher, T., Palanisamy, S., Kramer, D.J. et al. Brown adipose tissue is associated with cardiometabolic health. Nat Med 27, 58–65 (2021). https://doi.org/10.1038/s41591-020-1126-7

[4] Finer, N. (2012). Better measures of fat mass - beyond BMI. Clinical Obesity, 2(3–4), 65.https://doi.org/10.1111/j.1758-8111.2012.00047.x