Logit and Loglinear

목 차

**1. 이항 반응변수**

1. **데이터 내용 ------------------------------------------------------------3**
2. **분석 내용**
3. **분석 목적**
4. **분석 방법**

**a. 가설**

**b. 검정통계량**

1. **결론 ---------------------------------------------------------------------4**
2. **코드**
3. **SAS**

**2. 다항 반응변수-순서형**

1. **데이터 내용 ------------------------------------------------------------7**
2. **분석 내용 ---------------------------------------------------------------8**
3. **분석 목적**
4. **분석 방법**

**a. 가설**

**b. 검정통계량**

1. **결론 ---------------------------------------------------------------------8**
2. **코드 ---------------------------------------------------------------------9**
3. **SAS**

**3. 다항 반응변수 – 명목형**

가**) 데이터 내용 ----------------------------------------------------------10**

**나) 분석 내용**

1. **분석 목적**
2. **분석 방법**
3. **가설**
4. **검정통계량**

**다) 결론 ------------------------------------------------------------------11**

**라) 코드 ------------------------------------------------------------------12**

**ㄱ. SAS**

**4**. **Loglinear**

**가) 데이터 내용**

**나) 분석 내용**

1. **분석 목적**
2. **분석 방법**
3. **가설**
4. **검정통계량**

**다) 결론 ------------------------------------------------------------------11**

**라) 코드 ------------------------------------------------------------------12**

**ㄱ. SAS**

**5. Best Loglinear Model**

**가) 데이터 내용**

**나) 분석 내용**

**ㄱ. 분석 목적**

**ㄴ. 분석 방법**

**a. 가설**

**b. 검정통계량**

**다) 결론**

**라) 코드**

**ㄱ. SAS**

1. 이항 반응변수

가) 데이터 내용

총 5001명을 대상으로 남 녀 간의 신체적 특성을 분류한 자료이다. 항목은 각각 남녀. 코의 길이, 코의 너비로 나누었다.

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Gender  (Z,k) | Nose\_Length  (X,i) | Nose\_Width(Y,j) | | Total |
| Wide | Narrow |
| Male | Long | 1932 | 268 | 2200 |
| Short | 251 | 49 | 300 |
| Female | Long | 29 | 311 | 340 |
| Short | 258 | 1903 | 2161 |

<출처><https://www.kaggle.com/datasets/elakiricoder/gender-classification-dataset>

나) 분석 내용

ㄱ. 분석 목적

코의 너비가 성별과 코의 길이에 따라 달라지는지 알아보는 것

ㄴ. 분석 방법

a. 모델

ln(pi(X)/(1-pi(X)) = alpha + beta(1)gender + beta(2)length

반응변수 – 이항 반응(코가 넓은 지, 좁은 지)

설명변수 – 성별과 코 길이

b. goodness of fit test

H0 : 자료가 모형에 적합하다.

Deviance(LRT) = 7.4471, p-value = 0.0064

Pearson X2 = 7.3898, p-value = 0.0066

로짓 모형이 자료에 적합하지 않다.

H0 : beta(1) = beta(2) = 0

p-value < .0001 from LRT, score, and Wald test.

c. fitted logistic regression model :

ln(pi(X)/(1-pi(X)) = 0.0569 + 1.9804gender – 0.00795length

M: 1 L:1

F : -1 S: -1

1) gender : M vs F

Length가 일정할 때,

Gender = M, gender =1, beta1(M)(hat) = 1.9804

Gender = F, gender = -1, beta1(F)(hat) = -1.9804

Odds ratio :

M vs F

Log odds ratio = beta1(hat) –(-beta1(hat)) = 2beta1(hat) = 3.961

Odds ratio = exp(3.961) = 52.51

Gender = M 인 경우 코가 넓을 odds는 Gender = F인 경우의 52.51이다.

2) length : L vs S

Gender 가 일정할 때,

Length = L -> length = 1 -> beta2(hat) = -0.00795

Length = S -> length = -1 -> beta2(hat) = 0.00795

Odds ratio:

L vs S:

Log odds ratio = 2\*beta2(hat) = -0.016

Odds ratio = exp(-0.016) = 0.984

다) 결론

3종 오류로부터 gender 와 코 길이는 모두 코 너비에 유의한 영향을 미친다.

라) 코드

ㄱ. SAS

-input

**data** nose;

input gender $ length $ width $ count @@;

cards;

m l w 1932 m l n 268

m s w 251 m s n 49

f l w 29 f l n 311

f s w 258 f s n 1903

;

**proc** **logistic**;

freq count;

class gender length;

model width = gender length / scale=none aggregate;

**run**;

-output

The LOGISTIC Procedure

| **Model Information** | |
| --- | --- |
| **Data Set** | WORK.NOSE |
| **Response Variable** | width |
| **Number of Response Levels** | 2 |
| **Frequency Variable** | count |
| **Model** | binary logit |
| **Optimization Technique** | Fisher's scoring |

|  |  |
| --- | --- |
| **Number of Observations Read** | 8 |
| **Number of Observations Used** | 8 |
| **Sum of Frequencies Read** | 5001 |
| **Sum of Frequencies Used** | 5001 |

| **Response Profile** | | |
| --- | --- | --- |
| **Ordered Value** | **width** | **Total Frequency** |
| **1** | n | 2531 |
| **2** | w | 2470 |

|  |
| --- |
| **Probability modeled is width='n'.** |

| **Class Level Information** | | |
| --- | --- | --- |
| **Class** | **Value** | **Design Variables** |
| **gender** | **f** | 1 |
|  | **m** | -1 |
| **length** | **l** | 1 |
|  | **s** | -1 |

| **Model Convergence Status** |
| --- |
| Convergence criterion (GCONV=1E-8) satisfied. |

| **Deviance and Pearson Goodness-of-Fit Statistics** | | | | |
| --- | --- | --- | --- | --- |
| **Criterion** | **Value** | **DF** | **Value/DF** | **Pr > ChiSq** |
| **Deviance** | 7.4471 | 1 | 7.4471 | 0.0064 |
| **Pearson** | 7.3898 | 1 | 7.3898 | 0.0066 |

|  |
| --- |
| **Number of unique profiles: 4** |

| **Model Fit Statistics** | | |
| --- | --- | --- |
| **Criterion** | **Intercept Only** | **Intercept and Covariates** |
| **AIC** | 6934.114 | 3689.693 |
| **SC** | 6940.631 | 3709.245 |
| **-2 Log L** | 6932.114 | 3683.693 |

| **Testing Global Null Hypothesis: BETA=0** | | | |
| --- | --- | --- | --- |
| **Test** | **Chi-Square** | **DF** | **Pr > ChiSq** |
| **Likelihood Ratio** | 3248.4215 | 2 | <.0001 |
| **Score** | 2877.2105 | 2 | <.0001 |
| **Wald** | 2090.6596 | 2 | <.0001 |

| **Type 3 Analysis of Effects** | | | |
| --- | --- | --- | --- |
| **Effect** | **DF** | **Wald Chi-Square** | **Pr > ChiSq** |
| **gender** | 1 | 935.5085 | <.0001 |
| **length** | 1 | 0.0151 | 0.9023 |

| **Analysis of Maximum Likelihood Estimates** | | | | | | |
| --- | --- | --- | --- | --- | --- | --- |
| **Parameter** |  | **DF** | **Estimate** | **Standard Error** | **Wald Chi-Square** | **Pr > ChiSq** |
| **Intercept** |  | 1 | 0.0569 | 0.0435 | 1.7138 | 0.1905 |
| **gender** | **f** | 1 | 1.9804 | 0.0647 | 935.5085 | <.0001 |
| **length** | **l** | 1 | -0.00795 | 0.0648 | 0.0151 | 0.9023 |

| **Odds Ratio Estimates** | | | |
| --- | --- | --- | --- |
| **Effect** | **Point Estimate** | **95% Wald Confidence Limits** | |
| **gender f vs m** | 52.500 | 40.732 | 67.669 |
| **length l vs s** | 0.984 | 0.764 | 1.269 |

| **Association of Predicted Probabilities and Observed Responses** | | | |
| --- | --- | --- | --- |
| **Percent Concordant** | 79.7 | **Somers' D** | 0.759 |
| **Percent Discordant** | 3.8 | **Gamma** | 0.909 |
| **Percent Tied** | 16.5 | **Tau-a** | 0.379 |
| **Pairs** | 6251570 | **c** | 0.879 |

2. 다항 반응변수 – 순서형

가) 데이터 내용

아래 표는 e-commerce를 통한 소비에 관한 자료이다. 가족 구성원의 수에 따라 실제 관측한 결과와 예상된 결과를 할당하여 e-commerce를 통해 구입한 마지막 시간대를 ‘never before’,’1 to 3 months’,’3 to 6 months’의 3가지 순서형 반응으로 측정되었다. 구매 기간이 관측 여부와 가족 구성원의 수에 따라 영향을 받는 지 알아보고자 한다.

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| Size of family | obv | Never before(n) | 1to3months  (o) | 3to6months  (t) | total |
| 2 or less  (s) | Observed  (y) | 10 | 19 | 3 | 34 |
|  | Expected  (e) | 11.4 | 11.2 | 5.6 | 34 |
| 3people  (m) | Observed  (y) | 17 | 17 | 13 | 57 |
|  | Expected  (e) | 19.2 | 18.8 | 9.3 | 57 |

<출처: Validating Intrinsic Factors Informing E-Commerce: Categorical Data Analysis Demo**,** Anthony Joe Turkson\*, John Awuah Addor, Douglas Yenwon Kharib>

3. 분석 내용

ㄱ. 분석 목적

구매 기간이 관측 여부와 가족 구성원의 수에 따라 영향을 받는 지 알아보고자 한다.

ㄴ. 분석 방법

a. 로짓 모델

마지막 구매 기간을 나타내는 변수인 ‘period’를 반응변수로 하고, 설명변수인 가족 구성원의 수(size)와 observed(obv)를 각각 x1,x2라고 했을 때 비례오즈( proportional odds)로짓 모형은 다음과 같다.

Logit[Fij] =αj + β1 xi1 + β2 xi2  , j=1,2

b. 검정통계량

-score test for the proportional odds assumption

X2 = 2.9731, p-value = 0.2262 > 0.05

Proportional odds assumption is satistied

-goodness of fit test

Deviance(lrt) = 5.2353, p-value = 0.2640

Pearson X2 = 4.9911, p-value = 0.2882

로짓 모형이 자료에 적합하다

-H0 : beta(1) = beta(2) = 0

LRT p-value = 0.6523

SC p-value = 0.6608

Wald p-value = 0.6458

-model

Lj|i(x)(hat) = log(F(x)(hat)/(1-F(x)(hat))

= αj(hat) + β1 size(hat) + β2 obv(hat)

= αj(hat) + (-0.0793)size + 0.1206obv

S :-1 Y : -1

M : 1 E : 1

Estimated cutpoint parameters: α1(hat)=-0.5104 , α2(hat)=1.4488

* Size : s vs m

Obv가 일정할 때

Size = “s”, size = -1 β1 (s)(hat)=0.0793 , p-value = 0.6093

Size = “m”, size = 1 β1 (m)(hat)=-0.0793

* Obv : Y vs E

Size가 일정할 때

Obv = “y”, obv = -1 β2 (y)(hat) = -0.1206, p-value = 0.4251

Obv = “E”, obv = 1 β2 (y)(hat) = 0.1206

* β1 (m)(hat), β2 (y)(hat) < 0 : 구매기간이 늘어나는 확률은 가족 구성원의 수가 많고, 관측된 값일수록 더 늘어난다. 즉, chance of longer period = “t” is higher for 3 people than 1 or 2 people, and also is higher for observed = “Y” than observed = “E”.
* at a given value of whether it’s observed, the odds that period is below any fixed level is estimated to be exp(-0.0793-0.0793) = exp(-0.1586) = 0.8533
* at a given value of family size, the odds that period is below any fixed level is estimated to be exp(0.1206-(-0.1206)) = exp(0.2412) = 1.2728

-estimated cumulative probability Fj(x)(hat)

첫 2줄은 첫 번째 개체에 대해 반응변수인 period의 수준이 n인 경우와 o이하인 경우의 추정된 누적 확률이다.

즉, P(Y<=n) = 0.36548로 구매 기간이 없는 경우일 확률이 0.36548이고 구매 기간이 1-3개월 미만일 확률은 P(Y<=o)=0.80338로 나타난다. 따라서 구매 기간이 1-3개월일 확률은 P(Y<=n)- P(Y<=o)= 0.4397 임을 알 수 있다.

| **OBS** | **size** | **obv** | **period** | **count** | **\_LEVEL\_** | **p** |
| --- | --- | --- | --- | --- | --- | --- |
| **1** | s | y | n | 10.0 | n | 0.36548 |
| **2** | s | y | n | 10.0 | o | 0.80338 |
| **3** | s | y | o | 19.0 | n | 0.36548 |
| **4** | s | y | o | 19.0 | o | 0.80338 |
| **5** | s | y | t | 3.0 | n | 0.36548 |
| **6** | s | y | t | 3.0 | o | 0.80338 |
| **7** | s | e | n | 11.4 | n | 0.42301 |
| **8** | s | e | n | 11.4 | o | 0.83873 |
| **9** | s | e | o | 11.2 | n | 0.42301 |
| **10** | s | e | o | 11.2 | o | 0.83873 |
| **11** | s | e | t | 5.6 | n | 0.42301 |
| **12** | s | e | t | 5.6 | o | 0.83873 |
| **13** | m | y | n | 17.0 | n | 0.32953 |
| **14** | m | y | n | 17.0 | o | 0.77710 |
| **15** | m | y | o | 19.0 | n | 0.32953 |
| **16** | m | y | o | 19.0 | o | 0.77710 |
| **17** | m | y | t | 13.0 | n | 0.32953 |
| **18** | m | y | t | 13.0 | o | 0.77710 |
| **19** | m | e | n | 19.2 | n | 0.38483 |
| **20** | m | e | n | 19.2 | o | 0.81609 |
| **21** | m | e | o | 18.8 | n | 0.38483 |
| **22** | m | e | o | 18.8 | o | 0.81609 |
| **23** | m | e | t | 9.3 | n | 0.38483 |
| **24** | m | e | t | 9.3 | o | 0.81609 |

다) 결론

만약 관측여부가 고정이라면, P(Y=t) 는 2명 이하의 구성원 수가 3인가족에서보다 더 높게 나타난다. 만약 가족 구성원 수가 고정이라면, P(Y=t)는 관측값의 경우보다 예상값의 경우에서 더 높게 나타난다.

라) 코드

-sas input

**data** period;

input size $ obv $ period $ count @@;

cards;

s y n 10 s y o 19 s y t 3

s e n 11.4 s e o 11.2 s e t 5.6

m y n 17 m y o 19 m y t 13

m e n 19.2 m e o 18.8 m e t 9.3

;

**proc** **logistic** order = data;

freq count;

class size obv;

model period = size obv / scale = none aggregate;

output out = prob pred = p;

**run**;

**proc** **print** data = prob (obs=**24**);

**run**;

-sas output

|  |
| --- |
| SAS 시스템 |

The LOGISTIC Procedure

| **Model Information** | |
| --- | --- |
| **Data Set** | WORK.PERIOD |
| **Response Variable** | period |
| **Number of Response Levels** | 3 |
| **Frequency Variable** | count |
| **Model** | cumulative logit |
| **Optimization Technique** | Fisher's scoring |

|  |  |
| --- | --- |
| **Number of Observations Read** | 12 |
| **Number of Observations Used** | 12 |
| **Sum of Frequencies Read** | 154 |
| **Sum of Frequencies Used** | 154 |

| **Response Profile** | | |
| --- | --- | --- |
| **Ordered Value** | **period** | **Total Frequency** |
| **1** | n | 57 |
| **2** | o | 67 |
| **3** | t | 30 |

|  |
| --- |
| **Probabilities modeled are cumulated over the lower Ordered Values.** |

| **Class Level Information** | | |
| --- | --- | --- |
| **Class** | **Value** | **Design Variables** |
| **size** | **m** | 1 |
|  | **s** | -1 |
| **obv** | **e** | 1 |
|  | **y** | -1 |

| **Model Convergence Status** |
| --- |
| Convergence criterion (GCONV=1E-8) satisfied. |

| **Score Test for the Proportional Odds Assumption** | | |
| --- | --- | --- |
| **Chi-Square** | **DF** | **Pr > ChiSq** |
| 2.9731 | 2 | 0.2262 |

| **Deviance and Pearson Goodness-of-Fit Statistics** | | | | |
| --- | --- | --- | --- | --- |
| **Criterion** | **Value** | **DF** | **Value/DF** | **Pr > ChiSq** |
| **Deviance** | 5.2353 | 4 | 1.3088 | 0.2640 |
| **Pearson** | 4.9911 | 4 | 1.2478 | 0.2882 |

|  |
| --- |
| **Number of unique profiles: 4** |

| **Model Fit Statistics** | | |
| --- | --- | --- |
| **Criterion** | **Intercept Only** | **Intercept and Covariates** |
| **AIC** | 326.973 | 330.118 |
| **SC** | 333.047 | 342.266 |
| **-2 Log L** | 322.973 | 322.118 |

| **Testing Global Null Hypothesis: BETA=0** | | | |
| --- | --- | --- | --- |
| **Test** | **Chi-Square** | **DF** | **Pr > ChiSq** |
| **Likelihood Ratio** | 0.8545 | 2 | 0.6523 |
| **Score** | 0.8285 | 2 | 0.6608 |
| **Wald** | 0.8747 | 2 | 0.6458 |

| **Type 3 Analysis of Effects** | | | |
| --- | --- | --- | --- |
| **Effect** | **DF** | **Wald Chi-Square** | **Pr > ChiSq** |
| **size** | 1 | 0.2612 | 0.6093 |
| **obv** | 1 | 0.6363 | 0.4251 |

| **Analysis of Maximum Likelihood Estimates** | | | | | | |
| --- | --- | --- | --- | --- | --- | --- |
| **Parameter** |  | **DF** | **Estimate** | **Standard Error** | **Wald Chi-Square** | **Pr > ChiSq** |
| **Intercept** | **n** | 1 | -0.5104 | 0.1708 | 8.9242 | 0.0028 |
| **Intercept** | **o** | 1 | 1.4488 | 0.2085 | 48.2935 | <.0001 |
| **size** | **m** | 1 | -0.0793 | 0.1552 | 0.2612 | 0.6093 |
| **obv** | **e** | 1 | 0.1206 | 0.1512 | 0.6363 | 0.4251 |

| **Odds Ratio Estimates** | | | |
| --- | --- | --- | --- |
| **Effect** | **Point Estimate** | **95% Wald Confidence Limits** | |
| **size m vs s** | 0.853 | 0.464 | 1.568 |
| **obv e vs y** | 1.273 | 0.704 | 2.302 |

| **Association of Predicted Probabilities and Observed Responses** | | | |
| --- | --- | --- | --- |
| **Percent Concordant** | 40.4 | **Somers' D** | 0.070 |
| **Percent Discordant** | 33.4 | **Gamma** | 0.095 |
| **Percent Tied** | 26.2 | **Tau-a** | 0.045 |
| **Pairs** | 7539 | **c** | 0.535 |

3. 다항 반응변수 – 명목형

가) 데이터 내용

버섯의 특성에 따라 분류한 2x2x2 분할표이다. Grill-size, class, 그리고 cap surface로 분류하였다.

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| Grill-Size | Class | Cap surface | | | total |
| Scaly(y) | Smooth(s) | Other(o) |
| Broad(b) | Edible(e) | 1432 | 0 | 2488 | 3920 |
| Poisonous(p) | 720 | 324 | 648 | 1692 |
| Narrow(n) | Edible(e) | 72 | 48 | 168 | 288 |
| Poisonous(p) | 1020 | 1088 | 116 | 2224 |

<출처>https://www.kaggle.com/datasets/uciml/mushroom-classification

나) 분석 내용

ㄱ. 분석 목적

Cap surface 가 grill-size와 class에 따라 상이한 상태를 가지는 지 알아보고자 한다.

ㄴ. 분석 방법

a. 가설

b. 검정통계량

-in the response profiles, the highest value, in this example “o”, becomes the reference category parameterizing the model.

-size effect is significant (P-value <.0001)

-class effect is significant (P-value < .0001).

-likelihood ratio(Deviance) indicated that the model does not fit well (P-value<.0001).

다) 결론

Xi1 : size , xi2 : class 라 할 때 로짓 모형은 다음과 같다.

j = 1 :

(Scaly)

Log(piY / pio) = 1.7478 + 0.6205xi1 + 1.3284xi2

1 : broad 1 : Edible

-1 : narrow -1 : Poisonous

j = 2 :

log(piS/piO) = 1.2785 + 1.2401xi1 + 1.833xi2

범주 Y와 S의 비교는 위의 두 식의 차를 구하면 되므로 다음과 같다.

Log(piY/piS) = 0.469 + (-0.638)xi1 + (-0.505)xi2

첫 번째 식에서 smooth surface대신 Scaly 추정오즈:

Broad는 narrow의 exp(2\*0.6205) = 3.459배가 되므로 broad 가 narrow보다 scaly surface를 가질 비율이 높다. 또한 edible은 poisonous의 exp(2\*1.3284) = 14.251 배가 되어 edible의 smooth surface 대비 scaly surface를 가질 비율이 더 높다.

두 번째 식에서 smooth 대신other surface 추정오즈:

Broad가 narrow 보다 오즈가 exp(2\*1.2401) = 11.944 배로 높아서 smooth 대신 other surface를 취할 비율이 더 높다. Edible의 추정오즈는 poisonous에 비해 exp(2\*1.833) = 39.095배로 smooth surface에 대비 other surface를 가질 비율이 더 높다.

라) 코드

ㄱ. sas

-input

**data** mushroom;

input size class surface $ count@@;

cards;

1 1 y 1432 1 1 s 0 1 1 o 2488

1 0 y 720 1 0 s 324 1 0 o 648

0 1 y 72 0 1 s 48 0 1 o 168

0 0 y 1020 0 0 s 1088 0 0 o 116

;

**proc** **catmod** order = data;

weight count;

model surface = size class / pred = prob;

**run**;

-output

|  |
| --- |
| SAS 시스템 |

The CATMOD Procedure

| **Data Summary** | | | |
| --- | --- | --- | --- |
| **Response** | surface | **Response Levels** | 3 |
| **Weight Variable** | count | **Populations** | 4 |
| **Data Set** | MUSHROOM | **Total Frequency** | 8124 |
| **Frequency Missing** | 0 | **Observations** | 11 |

| **Population Profiles** | | | |
| --- | --- | --- | --- |
| **Sample** | **size** | **class** | **Sample Size** |
| **1** | 1 | 1 | 3920 |
| **2** | 1 | 0 | 1692 |
| **3** | 0 | 1 | 288 |
| **4** | 0 | 0 | 2224 |

| **Response Profiles** | |
| --- | --- |
| **Response** | **surface** |
| **1** | y |
| **2** | o |
| **3** | s |

| **Maximum Likelihood Analysis** |
| --- |
| Maximum likelihood computations converged. |

| **Maximum Likelihood Analysis of Variance** | | | |
| --- | --- | --- | --- |
| **Source** | **DF** | **Chi-Square** | **Pr > ChiSq** |
| **Intercept** | 2 | 623.46 | <.0001 |
| **size** | 2 | 675.43 | <.0001 |
| **class** | 2 | 700.13 | <.0001 |
| **Likelihood Ratio** | 2 | 309.35 | <.0001 |

| **Analysis of Maximum Likelihood Estimates** | | | | | | |
| --- | --- | --- | --- | --- | --- | --- |
| **Parameter** |  | **Function Number** | **Estimate** | **Standard Error** | **Chi- Square** | **Pr > ChiSq** |
| **Intercept** |  | **1** | 1.7478 | 0.0760 | 528.99 | <.0001 |
|  |  | **2** | 1.2785 | 0.0798 | 256.64 | <.0001 |
| **size** | **1** | **1** | 0.6205 | 0.0391 | 251.99 | <.0001 |
|  | **1** | **2** | 1.2401 | 0.0477 | 675.42 | <.0001 |
| **class** | **1** | **1** | 1.3284 | 0.0776 | 293.01 | <.0001 |
|  | **1** | **2** | 1.8330 | 0.0785 | 545.91 | <.0001 |

| **Maximum Likelihood Predicted Values for Response Functions** | | | | | | | |
| --- | --- | --- | --- | --- | --- | --- | --- |
| **size** | **class** | **Function Number** | **Observed** | | **Predicted** | | **Residual** |
| **Function** | **Standard Error** | **Function** | **Standard Error** |
| **1** | **1** | **1** | . | . | 3.696617 | 0.148934 | . |
|  |  | **2** | . | . | 4.351719 | 0.148491 | . |
| **1** | **0** | **1** | 0.798508 | 0.066898 | 1.039812 | 0.066531 | -0.2413 |
|  |  | **2** | 0.693147 | 0.068041 | 0.68562 | 0.070026 | 0.007527 |
| **0** | **1** | **1** | 0.405465 | 0.186339 | 2.455693 | 0.15741 | -2.05023 |
|  |  | **2** | 1.252763 | 0.163663 | 1.871463 | 0.164763 | -0.6187 |
| **0** | **0** | **1** | -0.06454 | 0.043583 | -0.20111 | 0.043742 | 0.136573 |
|  |  | **2** | -2.23851 | 0.097672 | -1.79464 | 0.071454 | -0.44387 |

| **Maximum Likelihood Predicted Values for Probabilities** | | | | | | | |
| --- | --- | --- | --- | --- | --- | --- | --- |
| **size** | **class** | **surface** | **Observed** | | **Predicted** | | **Residual** |
| **Probability** | **Standard Error** | **Probability** | **Standard Error** |
| **1** | **1** | **y** | 0.3653 | 0.0077 | 0.339 | 0.0074 | 0.0263 |
|  |  | **o** | 0.6347 | 0.0077 | 0.6526 | 0.0074 | -0.018 |
|  |  | **s** | 0 | 0 | 0.0084 | 0.0012 | -0.008 |
| **1** | **0** | **y** | 0.4255 | 0.012 | 0.4866 | 0.0115 | -0.061 |
|  |  | **o** | 0.383 | 0.0118 | 0.3414 | 0.0109 | 0.0415 |
|  |  | **s** | 0.1915 | 0.0096 | 0.172 | 0.009 | 0.0195 |
| **0** | **1** | **y** | 0.25 | 0.0255 | 0.6085 | 0.0179 | -0.359 |
|  |  | **o** | 0.5833 | 0.0291 | 0.3393 | 0.0176 | 0.2441 |
|  |  | **s** | 0.1667 | 0.022 | 0.0522 | 0.0077 | 0.1145 |
| **0** | **0** | **y** | 0.4586 | 0.0106 | 0.4122 | 0.01 | 0.0464 |
|  |  | **o** | 0.0522 | 0.0047 | 0.0838 | 0.0052 | -0.032 |
|  |  | **s** | 0.4892 | 0.0106 | 0.504 | 0.0105 | -0.015 |

4. Loglinear

가) 데이터 내용

5001명을 대상으로 신체의 특성을 분류한 표이다. 그 중 코의 길이와 코의 너비에 대하여 분할표이다.

|  |  |  |  |
| --- | --- | --- | --- |
| Nose\_long | Nose\_wideth | | total |
| Wide | narrow |  |
| Long | 1961 | 579 | 2540 |
| short | 509 | 1952 | 2461 |
| total | 2470 | 2531 | 5001 |

<출처><https://www.kaggle.com/datasets/elakiricoder/gender-classification-dataset>

나) 분석 내용

ㄱ. 분석 목적

코의 너비와 코의 길이간의 연관성이 있는지 log-linear 모형을 통해 알아보려고 한다.

ㄴ. 분석 방법

a. 모델 및 가설

-under null hypothesis

pi(ij) = pi(i+) \* pi(+j) , i=1,2 , j=1,2

m(ij) = n\*pi(ij) = n\*pi(i+)\*pi(+j)

log(m(ij)) = log(n) + log(pi(i+)) + log(pi(+j))

log(m(ij)) = mu + lamda(X,i) + lamda(Y,j) , i=1,2 , j=1,2

lamda(X,i) = log(pi(i+)) – sum(k=1to2)(log(pi(k+))/2

lamda(Y,j) = log(pi(+j)) – sum(k=1to2)(log(pi(+k))/2

-saturated log-linear model

log(mij) = mu + lamda(X,i) + lamda(Y,j) + lamda(XY,ij) , i=1,2,j=1,2

-제약식

sum(i=0to2)(lamda(X,i)) = 0, sum(j=0to2)(lamda(Y,j)) = 0

sum(i=0to2)(lamda(XY,ij))=sum(j=0to2)(lamda(XY,ij)) = 0

-가설

H0 : 코의 길이와 너비는 서로 독립이다.(lamda(XY,ij)=0) Vs H1 : not H0

b. 검정 통계량

G2 = -2logA = 2\*sum(i=1to2)sum(j=1to2)(nij\*log(nij/mij(hat)) ~ X2(1) under H0

= 1696.37 , p-value < 0.0001

A = max(likelihood in H0) / max(likelihood in H0Ha)

= (ni+n+j/n2)^(nij) / (nij/n)^(nij)

= (ni+n+j)^(nij) / nn\*(nij)^(nij)

= (ni+n+j/n)^(nij) / (nij)^(nij)

= (mij(hat) / nij)^(nij)

-2logA convergences in distribution to X2(df)

df = dim(H0Ha) – dim(H0)

= [IJ-1] – [(I-1)+(J-1)]

= (I-1)(J-1) = (2-1)\*(2-1) = 1

다) 결론

-for analogous logit model

For indep model log(mij) = mu + lamda(X,i) + lamda(Y,j)

For binary response (I\*2 table)

Log[pi(1|i)/pi(2|i)] = log[mi1/mi2] = log(mi1) – log(mi2)

= [mu+lamda(X,i)+lamda(Y,1)] – [mu+lamda(X,i)+lamda(Y,2)]

= lamda(Y,1) – lamda(Y,2)

= 2lamda(Y,1) since lamda(Y,1) + lamda(Y,2)=0 for each i

Does not depend on i

\* 로짓은 각 독립 모형 하에서 각 열에 대해 같고, exp(2lamda(Y,1)) 은 각 열의 제 1 반응의 오즈이다.

모수의 추정값들을 통해 logit과 odds를 구하면 독립모형인 경우 넓은 코를 가질(Y=1)odds는

exp(2lamda(Y,1)) = exp(2\*(-0.0122)) = 0.976 으로 모든 행에 대해서 같다.

-for saturated model for 2\*2 table

Log(theta) = log[(m11m22)/(m12m21)] = log(m1) + log(m22) – log(m12) – log(m21)

= mu + lamda(X,1) + lamda(Y,1) + lamda(XY,11) +

mu + lamda(X,2) + lamda(Y,2) + lamda(XY,22)

- (mu + lamda(X,1) + lamda(Y,2) + lamda(XY,12))

- (mu + lamda(X,2) + lamda(Y,1) + lamda(XY,21))

= lamda(XY,11) + lamda(XY,22) – lamda(XY,12) – lamda(XY,21)

= 4lamda(XY,11), since lamda(11) + lamda(12) = 0, lamda(11) + lamda(21) = 0

i.e.) sum(i)(lamda(ij)) = sum(j)(lamda(ij)) = 0

So, theta = exp(4lamda(XY,11) Note theta =1 <-> lamda(XY,11) = 0 <-> lamda(XY,ij) = 0

MLE of lamda(XY,11) = 0.6410 , p-value < 0.0001

따라서 코 길이와 코 너비는 서로 관련이 있음을 알 수 있다.

라) 코드

ㄱ. sas

-Input

**data** nose;

input length $ width $ count @@;

cards;

long wide 1961

long narrow 579

short wide 509

short narrow 1952

;

(analogous logit model)

**proc** **catmod** order = DATA;

weight count;

model length \* width = \_response\_/noreponse noiter;

loglin length width;

**run**;

(saturated logit model)

**proc** **catmod** order = DATA;

weight count;

model length\*width = \_response\_ / noresponse noiter;

loglin length|width;

**run**;

-Output

(analogous logit model)

The CATMOD Procedure

| **Data Summary** | | | |
| --- | --- | --- | --- |
| **Response** | length\*width | **Response Levels** | 4 |
| **Weight Variable** | count | **Populations** | 1 |
| **Data Set** | NOSE | **Total Frequency** | 5001 |
| **Frequency Missing** | 0 | **Observations** | 4 |

| **Population Profiles** | |
| --- | --- |
| **Sample** | **Sample Size** |
| **1** | 5001 |

| **Response Profiles** | | |
| --- | --- | --- |
| **Response** | **length** | **width** |
| **1** | long | wide |
| **2** | long | narrow |
| **3** | short | wide |
| **4** | short | narrow |

| **Maximum Likelihood Analysis** |
| --- |
| Maximum likelihood computations converged. |

| **Maximum Likelihood Analysis of Variance** | | | |
| --- | --- | --- | --- |
| **Source** | **DF** | **Chi-Square** | **Pr > ChiSq** |
| **length** | 1 | 1.25 | 0.2640 |
| **width** | 1 | 0.74 | 0.3884 |
| **Likelihood Ratio** | 1 | 1696.37 | <.0001 |

| **Analysis of Maximum Likelihood Estimates** | | | | | |
| --- | --- | --- | --- | --- | --- |
| **Parameter** |  | **Estimate** | **Standard Error** | **Chi- Square** | **Pr > ChiSq** |
| **length** | **long** | 0.0158 | 0.0141 | 1.25 | 0.2640 |
| **width** | **wide** | -0.0122 | 0.0141 | 0.74 | 0.3884 |

(saturated logit model)

|  |
| --- |
| SAS 시스템 |

The CATMOD Procedure

| **Data Summary** | | | |
| --- | --- | --- | --- |
| **Response** | length\*width | **Response Levels** | 4 |
| **Weight Variable** | count | **Populations** | 1 |
| **Data Set** | NOSE | **Total Frequency** | 5001 |
| **Frequency Missing** | 0 | **Observations** | 4 |

| **Population Profiles** | |
| --- | --- |
| **Sample** | **Sample Size** |
| **1** | 5001 |

| **Response Profiles** | | |
| --- | --- | --- |
| **Response** | **length** | **width** |
| **1** | long | wide |
| **2** | long | narrow |
| **3** | short | wide |
| **4** | short | narrow |

| **Maximum Likelihood Analysis** |
| --- |
| Maximum likelihood computations converged. |

| **Maximum Likelihood Analysis of Variance** | | | |
| --- | --- | --- | --- |
| **Source** | **DF** | **Chi-Square** | **Pr > ChiSq** |
| **length** | 1 | 3.78 | 0.0519 |
| **width** | 1 | 3.28 | 0.0703 |
| **length\*width** | 1 | 1394.69 | <.0001 |
| **Likelihood Ratio** | 0 | . | . |

| **Analysis of Maximum Likelihood Estimates** | | | | | |
| --- | --- | --- | --- | --- | --- |
| **Parameter** |  | **Estimate** | **Standard Error** | **Chi- Square** | **Pr > ChiSq** |
| **length** | **long** | 0.0334 | 0.0172 | 3.78 | 0.0519 |
| **width** | **wide** | -0.0311 | 0.0172 | 3.28 | 0.0703 |
| **length\*width** | **long wide** | 0.6410 | 0.0172 | 1394.69 | <.0001 |

5. Best Loglinear Model

가) 데이터 내용

총 5001명을 대상으로 남 녀 간의 신체적 특성을 분류한 자료이다. 항목은 각각 남녀. 코의 길이, 코의 너비로 나누었다.

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Gender  (Z,k) | Nose\_Length  (X,i) | Nose\_Width(Y,j) | | Total |
| Wide | Narrow |
| Male | Long | 1932 | 268 | 2200 |
| Short | 251 | 49 | 300 |
| Female | Long | 29 | 311 | 340 |
| Short | 258 | 1903 | 2161 |

<출처><https://www.kaggle.com/datasets/elakiricoder/gender-classification-dataset>

나) 분석 내용

ㄱ. 분석 목적

성별, 코의 길이, 코의 너비 사이에 어떤 연관성이 존재하는 가를 밝히는 것이다.

즉, 최적의 모형을 찾는 것이다.

ㄴ. 분석 방법

a.가설

H0 : 각 모형에서의 독립의 수준에 따라 독립이다.(ex. In no three-factor interaction model, H0 is lamda(XY,YZ,XZ, ijk) = 0 ) vs H1 : not H0

b. 적합 모형 찾기

|  |  |  |  |
| --- | --- | --- | --- |
| 로그-선형 모형 | G2(M)(df),p-value | G2(M2|M1)(df),p-value | X20.05(df) |
| (XY,YZ,XZ) | 7.25(1),0.0071 |  |  |
| (XZ, XY) | 7.46(2),0.024 | 0.21(1) | 3.84(1) |
| (XY, YZ) | 1559.5(2),<.0001 | 1552.25(1) |
| (XZ, YZ) | 1419.61(2),<.0001 | 1412.36(1) |
| (XY,Z) | 3255.87(3),<.0001 | 3248.62(2) | 5.991(2) |
| (YZ, X) | 4668.02(3),<.0001 | 4660.77(2) |
| (XZ, Y) | 3115.98(3),<.0001 | 3108.73(2) |
| (X,Y,Z) | 6364.38(4),<.0001 | 6357.13(3) | 7.815(3) |

모든 모형에서의 p-value(global null hypothesis : each interaction is independent

c. 최적모형선택

global null hypothesis(=each model’s independence level)

p-value in every model’s level < 0.05

따라서 global null hypothesis is not rejected.

다) 결론

성별, 코의 길이, 그리고 코의 너비는 독립적인 관계이다.

라) 코드

ㄱ. SAS

모든 가능한 모형 적합

-input

**data** nose;

input gender$ length $ width $ count @@;

cards;

M long wide 1932

M long narrow 268

M short wide 251

M short narrow 49

F long wide 29

F long narrow 311

F short wide 258

F short narrow 1903

;

/\*(XYZ)\*/

**proc** **catmod** order = data;

weight count;

model gender\*length\*width = \_response\_/noresponse noiter;

loglin gender|length|width;

**run**;

/\*(XY,YZ,XZ)\*/

**proc** **catmod** order = data;

weight count;

model gender\*length\*width = \_response\_/noresponse noiter;

loglin gender|length length|width width|gender;

**run**;

**proc** **catmod** order = data;

weight count;

model gender\*length\*width = \_response\_/noresponse noiter;

loglin gender|length gende|width;

**run**;

/\*(XY, YZ)\*/

**proc** **catmod** order = data;

weight count;

model gender\*length\*width = \_response\_/noresponse noiter;

loglin gender|length length|width;

**run**;

/\*(XZ, YZ)\*/

**proc** **catmod** order = data;

weight count;

model gender\*length\*width = \_response\_/noresponse noiter;

loglin gender|width length|width;

**run**;

/\*(XY, Z)\*/

**proc** **catmod** order = data;

weight count;

model gender\*length\*width = \_response\_/noresponse noiter;

loglin gender|length width;

**run**;

/\*(YZ, X)\*/

**proc** **catmod** order = data;

weight count;

model gender\*length\*width = \_response\_/noresponse noiter;

loglin length|width gender;

**run**;

/\*(XZ, Y)\*/

**proc** **catmod** order = data;

weight count;

model gender\*length\*width = \_response\_/noresponse noiter;

loglin gender|width length;

**run**;

/\*(X,Y,Z)\*/

**proc** **catmod** order = data;

weight count;

model gender\*length\*width = \_response\_/noresponse noiter;

loglin gender length width;

**run**;

-output

|  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| | **Data Summary** | | | | | --- | --- | --- | --- | | **Response** | length\*width | **Response Levels** | 4 | | **Weight Variable** | count | **Populations** | 1 | | **Data Set** | NOSE | **Total Frequency** | 5001 | | **Frequency Missing** | 0 | **Observations** | 4 |  | **Population Profiles** | | | --- | --- | | **Sample** | **Sample Size** | | **1** | 5001 |  | **Response Profiles** | | | | --- | --- | --- | | **Response** | **length** | **width** | | **1** | long | wide | | **2** | long | narrow | | **3** | short | wide | | **4** | short | narrow |  | **Maximum Likelihood Analysis** | | --- | | Maximum likelihood computations converged. |  | **Maximum Likelihood Analysis of Variance** | | | | | --- | --- | --- | --- | | **Source** | **DF** | **Chi-Square** | **Pr > ChiSq** | | **length** | 1 | 1.25 | 0.2640 | | **width** | 1 | 0.74 | 0.3884 | | **Likelihood Ratio** | 1 | 1696.37 | <.0001 |  | **Analysis of Maximum Likelihood Estimates** | | | | | | | --- | --- | --- | --- | --- | --- | | **Parameter** |  | **Estimate** | **Standard Error** | **Chi- Square** | **Pr > ChiSq** | | **length** | **long** | 0.0158 | 0.0141 | 1.25 | 0.2640 | | **width** | **wide** | -0.0122 | 0.0141 | 0.74 | 0.3884 | | | **Data Summary** | | | | | --- | --- | --- | --- | | **Response** | length\*width | **Response Levels** | 4 | | **Weight Variable** | count | **Populations** | 1 | | **Data Set** | NOSE | **Total Frequency** | 5001 | | **Frequency Missing** | 0 | **Observations** | 4 |  | **Population Profiles** | | | --- | --- | | **Sample** | **Sample Size** | | **1** | 5001 |  | **Response Profiles** | | | | --- | --- | --- | | **Response** | **length** | **width** | | **1** | long | wide | | **2** | long | narrow | | **3** | short | wide | | **4** | short | narrow |  | **Maximum Likelihood Analysis** | | --- | | Maximum likelihood computations converged. |  | **Maximum Likelihood Analysis of Variance** | | | | | --- | --- | --- | --- | | **Source** | **DF** | **Chi-Square** | **Pr > ChiSq** | | **length** | 1 | 3.78 | 0.0519 | | **width** | 1 | 3.28 | 0.0703 | | **length\*width** | 1 | 1394.69 | <.0001 | | **Likelihood Ratio** | 0 | . | . |  | **Analysis of Maximum Likelihood Estimates** | | | | | | | --- | --- | --- | --- | --- | --- | | **Parameter** |  | **Estimate** | **Standard Error** | **Chi- Square** | **Pr > ChiSq** | | **length** | **long** | 0.0334 | 0.0172 | 3.78 | 0.0519 | | **width** | **wide** | -0.0311 | 0.0172 | 3.28 | 0.0703 | | **length\*width** | **long wide** | 0.6410 | 0.0172 | 1394.69 | <.0001 | | The CATMOD Procedure   | **Data Summary** | | | | | --- | --- | --- | --- | | **Response** | gender\*length\*width | **Response Levels** | 8 | | **Weight Variable** | count | **Populations** | 1 | | **Data Set** | NOSE | **Total Frequency** | 5001 | | **Frequency Missing** | 0 | **Observations** | 8 |      | **Population Profiles** | | | --- | --- | | **Sample** | **Sample Size** | | **1** | 5001 |      | **Response Profiles** | | | | | --- | --- | --- | --- | | **Response** | **gender** | **length** | **width** | | **1** | M | long | wide | | **2** | M | long | narrow | | **3** | M | short | wide | | **4** | M | short | narrow | | **5** | F | long | wide | | **6** | F | long | narrow | | **7** | F | short | wide | | **8** | F | short | narrow |      | **Maximum Likelihood Analysis** | | --- | | Maximum likelihood computations converged. |  | **Maximum Likelihood Analysis of Variance** | | | | | --- | --- | --- | --- | | **Source** | **DF** | **Chi-Square** | **Pr > ChiSq** | | **gender** | 1 | 1.87 | 0.1719 | | **length** | 1 | 0.93 | 0.3338 | | **gender\*length** | 1 | 846.23 | <.0001 | | **width** | 1 | 8.20 | 0.0042 | | **gender\*width** | 1 | 900.13 | <.0001 | | **length\*width** | 1 | 0.02 | 0.9025 | | **gender\*length\*width** | 1 | 7.25 | 0.0071 | | **Likelihood Ratio** | 0 | . | . |  | **Analysis of Maximum Likelihood Estimates** | | | | | | | --- | --- | --- | --- | --- | --- | | **Parameter** |  | **Estimate** | **Standard Error** | **Chi- Square** | **Pr > ChiSq** | | **gender** | **M** | 0.0454 | 0.0332 | 1.87 | 0.1719 | | **length** | **long** | -0.0321 | 0.0332 | 0.93 | 0.3338 | | **gender\*length** | **M long** | 0.9671 | 0.0332 | 846.23 | <.0001 | | **width** | **wide** | -0.0952 | 0.0332 | 8.20 | 0.0042 | | **gender\*width** | **M wide** | 0.9975 | 0.0332 | 900.13 | <.0001 | | **length\*width** | **long wide** | -0.00407 | 0.0332 | 0.02 | 0.9025 | | **gender\*length\*width** | **M long wide** | 0.0895 | 0.0332 | 7.25 | 0.0071 | |
| The CATMOD Procedure   | **Data Summary** | | | | | --- | --- | --- | --- | | **Response** | gender\*length\*width | **Response Levels** | 8 | | **Weight Variable** | count | **Populations** | 1 | | **Data Set** | NOSE | **Total Frequency** | 5001 | | **Frequency Missing** | 0 | **Observations** | 8 |  | **Population Profiles** | | | --- | --- | | **Sample** | **Sample Size** | | **1** | 5001 |  | **Response Profiles** | | | | | --- | --- | --- | --- | | **Response** | **gender** | **length** | **width** | | **1** | M | long | wide | | **2** | M | long | narrow | | **3** | M | short | wide | | **4** | M | short | narrow | | **5** | F | long | wide | | **6** | F | long | narrow | | **7** | F | short | wide | | **8** | F | short | narrow |  | **Maximum Likelihood Analysis** | | --- | | Maximum likelihood computations converged. |  | **Maximum Likelihood Analysis of Variance** | | | | | --- | --- | --- | --- | | **Source** | **DF** | **Chi-Square** | **Pr > ChiSq** | | **gender** | 1 | 0.04 | 0.8441 | | **length** | 1 | 2.85 | 0.0912 | | **gender\*length** | 1 | 874.54 | <.0001 | | **width** | 1 | 1.71 | 0.1905 | | **length\*width** | 1 | 0.02 | 0.9023 | | **gender\*width** | 1 | 935.93 | <.0001 | | **Likelihood Ratio** | 1 | 7.45 | 0.0064 |  | **Analysis of Maximum Likelihood Estimates** | | | | | | | --- | --- | --- | --- | --- | --- | | **Parameter** |  | **Estimate** | **Standard Error** | **Chi- Square** | **Pr > ChiSq** | | **gender** | **M** | -0.00528 | 0.0268 | 0.04 | 0.8441 | | **length** | **long** | 0.0358 | 0.0212 | 2.85 | 0.0912 | | **gender\*length** | **M long** | 0.9574 | 0.0324 | 874.54 | <.0001 | | **width** | **wide** | -0.0284 | 0.0217 | 1.71 | 0.1905 | | **length\*width** | **long wide** | 0.00398 | 0.0324 | 0.02 | 0.9023 | | **gender\*width** | **M wide** | 0.9902 | 0.0324 | 935.93 | <.0001 | | The CATMOD Procedure   | **Data Summary** | | | | | --- | --- | --- | --- | | **Response** | gender\*length\*width | **Response Levels** | 8 | | **Weight Variable** | count | **Populations** | 1 | | **Data Set** | NOSE | **Total Frequency** | 5001 | | **Frequency Missing** | 0 | **Observations** | 8 |  | **Population Profiles** | | | --- | --- | | **Sample** | **Sample Size** | | **1** | 5001 |  | **Response Profiles** | | | | | --- | --- | --- | --- | | **Response** | **gender** | **length** | **width** | | **1** | M | long | wide | | **2** | M | long | narrow | | **3** | M | short | wide | | **4** | M | short | narrow | | **5** | F | long | wide | | **6** | F | long | narrow | | **7** | F | short | wide | | **8** | F | short | narrow |  | **Maximum Likelihood Analysis** | | --- | | Maximum likelihood computations converged. |      | **Maximum Likelihood Analysis of Variance** | | | | | --- | --- | --- | --- | | **Source** | **DF** | **Chi-Square** | **Pr > ChiSq** | | **gender** | 1 | 1.60 | 0.2059 | | **length** | 1 | 5.23 | 0.0222 | | **gender\*length** | 1 | 2052.27 | <.0001 | | **width** | 1 | 3.28 | 0.0703 | | **length\*width** | 1 | 1394.67 | <.0001 | | **Likelihood Ratio** | 2 | 1559.50 | <.0001 |      | **Analysis of Maximum Likelihood Estimates** | | | | | | | --- | --- | --- | --- | --- | --- | | **Parameter** |  | **Estimate** | **Standard Error** | **Chi- Square** | **Pr > ChiSq** | | **gender** | **M** | -0.0268 | 0.0212 | 1.60 | 0.2059 | | **length** | **long** | 0.0533 | 0.0233 | 5.23 | 0.0222 | | **gender\*length** | **M long** | 0.9605 | 0.0212 | 2052.27 | <.0001 | | **width** | **wide** | -0.0311 | 0.0172 | 3.28 | 0.0703 | | **length\*width** | **long wide** | 0.6410 | 0.0172 | 1394.67 | <.0001 | | The CATMOD Procedure   | **Data Summary** | | | | | --- | --- | --- | --- | | **Response** | gender\*length\*width | **Response Levels** | 8 | | **Weight Variable** | count | **Populations** | 1 | | **Data Set** | NOSE | **Total Frequency** | 5001 | | **Frequency Missing** | 0 | **Observations** | 8 |      | **Population Profiles** | | | --- | --- | | **Sample** | **Sample Size** | | **1** | 5001 |      | **Response Profiles** | | | | | --- | --- | --- | --- | | **Response** | **gender** | **length** | **width** | | **1** | M | long | wide | | **2** | M | long | narrow | | **3** | M | short | wide | | **4** | M | short | narrow | | **5** | F | long | wide | | **6** | F | long | narrow | | **7** | F | short | wide | | **8** | F | short | narrow |      | **Maximum Likelihood Analysis** | | --- | | Maximum likelihood computations converged. |      | **Maximum Likelihood Analysis of Variance** | | | | | --- | --- | --- | --- | | **Source** | **DF** | **Chi-Square** | **Pr > ChiSq** | | **gender** | 1 | 0.96 | 0.3261 | | **width** | 1 | 3.94 | 0.0471 | | **gender\*width** | 1 | 2090.68 | <.0001 | | **length** | 1 | 3.78 | 0.0519 | | **length\*width** | 1 | 1394.67 | <.0001 | | **Likelihood Ratio** | 2 | 1419.61 | <.0001 |      | **Analysis of Maximum Likelihood Estimates** | | | | | | | --- | --- | --- | --- | --- | --- | | **Parameter** |  | **Estimate** | **Standard Error** | **Chi- Square** | **Pr > ChiSq** | | **gender** | **M** | 0.0213 | 0.0217 | 0.96 | 0.3261 | | **width** | **wide** | -0.0472 | 0.0238 | 3.94 | 0.0471 | | **gender\*width** | **M wide** | 0.9932 | 0.0217 | 2090.68 | <.0001 | | **length** | **long** | 0.0334 | 0.0172 | 3.78 | 0.0519 | | **length\*width** | **long wide** | 0.6410 | 0.0172 | 1394.67 | <.0001 | |
| The CATMOD Procedure   | **Data Summary** | | | | | --- | --- | --- | --- | | **Response** | gender\*length\*width | **Response Levels** | 8 | | **Weight Variable** | count | **Populations** | 1 | | **Data Set** | NOSE | **Total Frequency** | 5001 | | **Frequency Missing** | 0 | **Observations** | 8 |      | **Population Profiles** | | | --- | --- | | **Sample** | **Sample Size** | | **1** | 5001 |      | **Response Profiles** | | | | | --- | --- | --- | --- | | **Response** | **gender** | **length** | **width** | | **1** | M | long | wide | | **2** | M | long | narrow | | **3** | M | short | wide | | **4** | M | short | narrow | | **5** | F | long | wide | | **6** | F | long | narrow | | **7** | F | short | wide | | **8** | F | short | narrow |      | **Maximum Likelihood Analysis** | | --- | | Maximum likelihood computations converged. |      | **Maximum Likelihood Analysis of Variance** | | | | | --- | --- | --- | --- | | **Source** | **DF** | **Chi-Square** | **Pr > ChiSq** | | **gender** | 1 | 1.60 | 0.2059 | | **length** | 1 | 2.85 | 0.0916 | | **gender\*length** | 1 | 2052.27 | <.0001 | | **width** | 1 | 0.74 | 0.3884 | | **Likelihood Ratio** | 3 | 3255.87 | <.0001 |      | **Analysis of Maximum Likelihood Estimates** | | | | | | | --- | --- | --- | --- | --- | --- | | **Parameter** |  | **Estimate** | **Standard Error** | **Chi- Square** | **Pr > ChiSq** | | **gender** | **M** | -0.0268 | 0.0212 | 1.60 | 0.2059 | | **length** | **long** | 0.0358 | 0.0212 | 2.85 | 0.0916 | | **gender\*length** | **M long** | 0.9605 | 0.0212 | 2052.27 | <.0001 | | **width** | **wide** | -0.0122 | 0.0141 | 0.74 | 0.3884 | | The CATMOD Procedure   | **Data Summary** | | | | | --- | --- | --- | --- | | **Response** | gender\*length\*width | **Response Levels** | 8 | | **Weight Variable** | count | **Populations** | 1 | | **Data Set** | NOSE | **Total Frequency** | 5001 | | **Frequency Missing** | 0 | **Observations** | 8 |      | **Population Profiles** | | | --- | --- | | **Sample** | **Sample Size** | | **1** | 5001 |      | **Response Profiles** | | | | | --- | --- | --- | --- | | **Response** | **gender** | **length** | **width** | | **1** | M | long | wide | | **2** | M | long | narrow | | **3** | M | short | wide | | **4** | M | short | narrow | | **5** | F | long | wide | | **6** | F | long | narrow | | **7** | F | short | wide | | **8** | F | short | narrow |      | **Maximum Likelihood Analysis** | | --- | | Maximum likelihood computations converged. |      | **Maximum Likelihood Analysis of Variance** | | | | | --- | --- | --- | --- | | **Source** | **DF** | **Chi-Square** | **Pr > ChiSq** | | **gender** | 1 | 0.96 | 0.3261 | | **width** | 1 | 1.71 | 0.1914 | | **gender\*width** | 1 | 2090.68 | <.0001 | | **length** | 1 | 1.25 | 0.2640 | | **Likelihood Ratio** | 3 | 3115.98 | <.0001 |      | **Analysis of Maximum Likelihood Estimates** | | | | | | | --- | --- | --- | --- | --- | --- | | **Parameter** |  | **Estimate** | **Standard Error** | **Chi- Square** | **Pr > ChiSq** | | **gender** | **M** | 0.0213 | 0.0217 | 0.96 | 0.3261 | | **width** | **wide** | -0.0284 | 0.0217 | 1.71 | 0.1914 | | **gender\*width** | **M wide** | 0.9932 | 0.0217 | 2090.68 | <.0001 | | **length** | **long** | 0.0158 | 0.0141 | 1.25 | 0.2640 | | The CATMOD Procedure   | **Data Summary** | | | | | --- | --- | --- | --- | | **Response** | gender\*length\*width | **Response Levels** | 8 | | **Weight Variable** | count | **Populations** | 1 | | **Data Set** | NOSE | **Total Frequency** | 5001 | | **Frequency Missing** | 0 | **Observations** | 8 |      | **Population Profiles** | | | --- | --- | | **Sample** | **Sample Size** | | **1** | 5001 |      | **Response Profiles** | | | | | --- | --- | --- | --- | | **Response** | **gender** | **length** | **width** | | **1** | M | long | wide | | **2** | M | long | narrow | | **3** | M | short | wide | | **4** | M | short | narrow | | **5** | F | long | wide | | **6** | F | long | narrow | | **7** | F | short | wide | | **8** | F | short | narrow |      | **Maximum Likelihood Analysis** | | --- | | Maximum likelihood computations converged. |      | **Maximum Likelihood Analysis of Variance** | | | | | --- | --- | --- | --- | | **Source** | **DF** | **Chi-Square** | **Pr > ChiSq** | | **gender** | 1 | 0.00 | 0.9887 | | **length** | 1 | 1.25 | 0.2640 | | **width** | 1 | 0.74 | 0.3884 | | **Likelihood Ratio** | 4 | 6364.38 | <.0001 |      | **Analysis of Maximum Likelihood Estimates** | | | | | | | --- | --- | --- | --- | --- | --- | | **Parameter** |  | **Estimate** | **Standard Error** | **Chi- Square** | **Pr > ChiSq** | | **gender** | **M** | -0.00020 | 0.0141 | 0.00 | 0.9887 | | **length** | **long** | 0.0158 | 0.0141 | 1.25 | 0.2640 | | **width** | **wide** | -0.0122 | 0.0141 | 0.74 | 0.3884 | |