Assignment 5

STAT 292 Applied Statistics 2A

20/06/2020

1.

(a)

Hypothesis:

$\mathcal{H}_0$: The model provides a good fit.

$\mathcal{H}_\mathcal{A}$: The model does not provide a good fit.

Likelihood ratio goodness-of-fit test:

$G^2 = 2\sum_{i=1}^{I}\sum_{j=1}^{J} n_{ij} log(\frac{n_{ij}}{\hat{\mu}_{ij}})$

$= 2\sum\ Observed \times log(\frac{Observed}{Expected})$

$\approx 2 \times [5log(\frac{5}{3.76}) + 70log(\frac{70}{71.24}) + ... + 35log(\frac{35}{36.53})]$

≈ 3.7326,

which follows a χ^2_5 distribution.

p-value $= P(\chi^2_5 > G^2) \approx P(\chi^2_5 > 3.7326) \approx 0.5885$

Pearson chi-square goodness-of-fit test:

$X^2 = \sum_{i=1}^{I}\sum_{j=1}^{J} \frac{(n_{ij}-\hat{\mu}_{ij})^2}{\hat{\mu}_{ij}}$

$X^2 = \sum \frac{(Observed-Expected)^2}{Expected}$

$\approx \frac{(5-3.76)^2}{3.76} + \frac{(70-71.24)^2}{71.24} + ... + \frac{(35-36.53)^2}{36.53}$

≈ 3.5477,

also following a χ^2_5 distribution.

p-value $= P(\chi^2_5 > X^2) \approx P(\chi^2_5 > 3.5477) \approx 0.6162$

At the 5% significance level we fail to reject the null hypothesis with a p-value of 0.5885 for the liklihood ratio goodness of fit test and 0.6162 for the Pearson chi-square test. Thus we don't have sufficient evidence to suggest that the model is not a good fit for the data.

(b)

$\hat{\beta}_0 \approx -1.7022$

$\hat{\beta}_1 \approx -0.00667$

(c)

$exp(\hat{\beta}_1) \approx exp(-0.00667) \approx 0.993$

An increase in area by $1km^2$ results in a multiplicative change of 0.993 (0.988, 0.999), the numbers in the brackets is the 95% confidence interval for this odds ratio. Since our odds ratio is very close to 1, this means that the size of the area has very little effect on the odds of extinction (increase in area by $1km$ results in a very small decrease in odds of extinction by 0.7%)

(d)

$\hat{P}(X) \approx \frac{exp(\hat{\beta}_0+\hat{\beta}_1 x)}{1+exp(\hat{\beta}_0+\hat{\beta}_1 x)}$

$\hat{P}(50) \approx \frac{exp(-1.7022-0.00667\times 50)}{1+exp(-1.7022-0.00667\times 50)} \approx 0.11550531074$ or 11.55%

(e)

Extinct bird species on the island of Ulkokrunni:

$(5+70) \times \frac{exp(\hat{\beta}_0+\hat{\beta}_1 x)}{1+exp(\hat{\beta}_0+\hat{\beta}_x)}$

$75 \times \frac{exp(-1.7022-0.00667\times 185.80)}{1+exp(-1.7022-0.00667\times 185.80)} \approx 75 \times 0.05014045304 \approx 3.76$

Non-extinct bird species on the island of Ulkokrunni:

$75 - 3.76 \approx 71.24$

(f)

Test statistic: $Z^2 \approx 5.2804$,

which follows a χ^2_1 distribution.

p**-value** $\approx P(\chi^2_1 > 5.2804) \approx 0.0216$

At the $\alpha = 0.05$ significance level we reject the null hypothesis with a p-value of 0.0216. There is evidence that there is an association between area of the island and a species' survival.

2.

(a)

$log(\frac{p_{ijk}}{1-p_{ijk}}) = \beta_0 + \beta_i^W + \beta_j^X + \beta_k^Y + \beta_{ij}^{WX} + \beta_{ik}^{WY} + \beta_{jk}^{XY} + \beta_{ijk}^{WXY}$

A model is considered saturated if it's residual df is 0.

residual df = no. of logits - no. of non-redundant parameters

no. of logits $= I \times J \times K = 2 \times 2 \times 2 = 8$

no. of non-redundant parameters $= 1+(I-1)+(J-1)+(K-1)+(I-1)(J-1)+(I-1)(K-1)+(J-1)(K-1)+(I-1)(J-1)(K-1) = 1+(2-1)+(2-1)+(2-1)+(2-1)(2-1)+(2-1)(2-1)+(2-1)(2-1)+(2-1)(2-1)(2-1) = 1+1+1+1+1+1+1+1 = 8$

residual df $= 8 - 8 = 0$

Since our residual df is 0, M_2 is a saturated model.

(b)

Full model (M_2): $log(\frac{p_{ijk}}{1-p_{ijk}}) = \beta_0 + \beta_i^W + \beta_j^X + \beta_k^Y + \beta_{ij}^{WX} + \beta_{ik}^{WY} + \beta_{jk}^{XY} + \beta_{ijk}^{WXY}$

Reduced model (M_1): $log(\frac{p_{ijk}}{1-p_{ijk}}) = \beta_0 + \beta_i^W + \beta_j^X + \beta_k^Y + \beta_{ij}^{WX} + \beta_{ik}^{WY} + \beta_{jk}^{XY}$

Hypothesis:

$\mathcal{H}_0$: The additional terms in M_2 can be deleted.

$\mathcal{H}_\mathcal{A}$: The additional terms in M_2 cannot be deleted.

Test statistic: $G^2 = 0.1472$

which follows a χ_1^2 distribution.

p**-value** $\approx P(\chi_1^2 > 0.1472) \approx 0.7012$

Under the significance level $\alpha = 0.05$, we fail to reject the null hypothesis with a p-value of 0.7012. Meaning that the ineraction term between all three of the parameters (Gender, Pre-marital Sex, Extra-marital Sex) β_{ijk}^{WXY} can be removed.

(c)

Final model: $log(\frac{p_{ijk}}{1-p_{ijk}}) = \beta_0 + \beta_i^W + \beta_j^X + \beta_k^Y + \beta_{jk}^{XY}$

(d)

Hypothesis:

$\mathcal{H}_0$: The model provides a good fit.

$\mathcal{H}_\mathcal{A}$: The model does not provide a good fit.

Likelihood ratio goodness-of-fit test ≈ 0.6978,

which follows a χ_3^2 distribution.

p-value $= P(\chi_3^2 > G^2) \approx P(\chi_3^2 > 0.6978) \approx 0.8737$

Pearson chi-square goodness-of-fit test ≈ 0.7013,

also following a χ_3^2 distribution.

p-value $= P(\chi_3^2 > X^2) \approx P(\chi_3^2 > 0.7013) \approx 0.8729$

At the 5% significance level we fail to reject the null hypothesis with a p-value of 0.8737 for the liklihood ratio goodness of fit test and 0.8729 for the Pearson chi-square test. Thus we don't have sufficient evidence to suggest that the model is not a good fit for the data.

(e)

$exp(\hat{\beta}_1^W) \approx exp(-0.3089) \approx 0.734$, 0.734 (0.552, 0.977)

Men are 0.734 times more likely than women to be divorced with a 95% confidence interval between 0.552 and 0.977. In other words women are 26.6% more likely to be divorced than men.