~~AAA~~ STA303. 2014. Final

1.

a)

sampling

Binomial

$H_0: \pi_1 = \pi_2$

$H_a: \pi_1 \neq \pi_2$

$$Z = \frac{(\hat{\pi}_1 - \hat{\pi}_2) - 0}{\sqrt{\hat{\pi}_p (1-\hat{\pi}_p)(\frac{1}{n_1}+\frac{1}{n_2})}} \sim N(0,1) \qquad \hat{\pi}_p = \frac{188+104}{188+104+108+1433}$$

$$p = 2P(Z > z)$$

b)

Independent observation

c) Pearson:

$$X^2 = \sum_{i=1}^{2}\sum_{j=1}^{2} \frac{(y_{ij} - \hat{\mu}_{ij})^2}{\hat{\mu}_{ij}} \sim \chi^2_{(2-1)(2-1)}$$

$H_0: \pi_{ij} = \pi_{i\cdot}\pi_{\cdot j}$

$H_a: \pi_{ij} \neq \pi_{i\cdot}\pi_{\cdot j}$

$$p\text{-value} = P(\chi^2 > \quad)$$

d).

(test statistic in a$)^2$ = test statistic in c.

e) $H_0 = \pi_{ij} = \pi_{i\cdot}\pi_{\cdot j}$ $\qquad H_a: \pi_{ij} \neq \pi_{i\cdot}\pi_{\cdot j}$

$$G^2 = -2\log\left(\frac{L_r}{L_f}\right)$$

$$= 23.37^2 \sim \chi^2_{(2-1)(2-1)}$$

d)

e)

2 a) For saturated model $\hat{u}_{ijk} = y_{ijk}$

The deviance: $2\log(L_{sat}) - 2\log(L_{fitted}) = 2\sum_i\sum_j\sum_k \big(-y_{ijk}$

$y_{ijk}\log(y_{ijk}) - \log(y_{ijk}!) + \hat{u}_{ijk} - y_{ijk}\log(\hat{u}_{ijk}) + \log(y_{ijk}!)$

and

$$\sum_i\sum_j\sum_k \hat{u}_{ijk} = \sum_i\sum_j\sum_k y_{ijk}.$$

Thus the deviance is $2\sum_i\sum_j\sum_k y_{ijk}\log\left(\frac{y_{ijk}}{\hat{u}_{ijk}}\right)$

ation

b) $\hat{\lambda}_{1,2,4} = n\frac{y_{1\cdot\cdot}}{n}\frac{y_{\cdot 2\cdot}}{n}\frac{y_{\cdot\cdot k}}{n}$

where $y_{\cdot\cdot} = \sum_{j=1}^{J}\sum_{k=1}^{K} y_{ijk}$. etc

and $n = \sum_i\sum_j\sum_k y_{ijk}$.

c) $\log(\hat{u}) = \beta_0 + \beta_1 I_I + \beta_2 I_J + \beta_3 I_K$

$+ \beta_4 (I_I * I_J) + \beta_5 (I_J * I_K) + \beta_6 (I_I * I_K)$

log-linear model

d)

$IJK - 7$

~~Sat~~ Degree of (saturated model - Fitted model)

e) Every pair of variation (~~a~~ variable 1, 2, 3) of all is associated. but we don't have an interaction.

e) $H_0: \pi_{ij} = \pi_{i\cdot}\pi_{\cdot j}$ $H_a: \pi_{ij} \neq \pi_{i\cdot}\pi_{\cdot j}$

$$G^2 = -2\log\left(\frac{L_r}{L_f}\right)$$
$$= 25.372 \sim \chi^2_{(2-1)(2-1)}$$

p-value $= 4.727 \times 10^{-7}$

$\Rightarrow$ Reject.

f) Independ observation
large sample size
no large outliers

g) $Y_{ij} \sim \text{Poisson}(\lambda_{ij})$

~~$\log(\lambda_{ij}) = \beta_0 + \beta_1$~~

M1: $\log(\lambda) = 1.2992 - 0.08 I_p - 4.615 I_m + 0.605438 (I_p \times I_m)$

M2: $\log(\lambda) = 1.299 - 0.0002710 I_p - 4.3084 I_m$

Use LRT or Deviance Test

$$G^2 = -2\log\left(\frac{L_r}{L_f}\right) = 25.372 \sim \chi^2_1$$

$$P = P(\chi^2_1 > 25.372)$$

d) $e^{-0.4885 - 1.1318(x)}$

e) no. we can't.
because 30mg exceed our scope range

f) [redacted]

95% CI: $\beta_1 \pm 1.\cancel{96} \times 0.04786.$

$$\Rightarrow \frac{e^{-0.64238-0.3481\times 5 \pm \cancel{1.96}\times 1.96\times}}{e^{-0.64238-0.3481\times 10 \pm 1.96\times 0.04786}}$$

g) No. For log-linear model.
we need sample size ≥ 5
since it's already $16 > 5$.
Assumption is satisfied.

h)

$H_0: \beta_1 = 0 \quad H_1: \beta_1 \neq 0$

$$G^2 = -2\log\left(\frac{L_r}{L_f}\right) = -2\log(L_r) + (-2\log$$

$H_0: \beta_1 = 0 \quad H_1: \beta_1 \neq 0$

$$G^2 = -2\log\left(\frac{L_r}{L_f}\right) = -2\log(L_r) + (-2\log L_f)$$

$$= 23.811$$

$$\sim \chi^2_{2-1}$$

p-value $= 1.063 \times 10^{-6}$.

p is small

→ Reject H_0

→ Has effects on odds of survival.

i) we can use Wald's Test

$H_0: \beta_1 = 0 \quad H_A: \beta_1 \neq 0$

$$Z = \frac{\hat{\beta}_1 - 0}{sd} \sim N(0,1)$$

$$p = 2.36 \times 10^{-6}$$

We reach same conclusion.

g) No. For log-linear model.
we need sample size ≥ 5
since it's already $16 > 5$.
Assumption is satisfied.

h)

$H_0: \beta_1 = 0 \quad H_1: \beta_1 \neq 0$

$$G^2 = -2\log\left(\frac{L_r}{L_f}\right) = -2\log(L_r) - (-2\log L_f)$$
$$= 23.811$$
$$\sim \chi^2_{27}$$

p-value $= 1.063 \times 10^{-6}$.

p is small
→ Reject H_0
→ Has effects on odds of survival.

i) We can use Wald's Test
$H_0: \beta_1 = 0 \quad H_A: \beta_1 \neq 0$
$$Z = \frac{\hat{\beta}_1 - 0}{sd} \sim N(0,1)$$
$p = 2.36 \times 10^{-6}$

We reach same conclusion.

b)
a)
i)
ii)
ii)
ii)
b)

(since we need $\lambda \geq 5$)

scatterplot matrix:

we have a ~~nearly linear relationship between~~ a correlation between dose = 10 and numbers of termites (almost a linear relationship).

but for dose = 5mg. the correlation is weak and scatterplots random distributed.

$$\hat{\beta}_g = \frac{\sum x_i y_i - n\bar{x}\bar{y}}{\sum x_i^2 - n\bar{x}^2} = \frac{\sum_{n_g} y_{gi} - n\bar{y}}{n_g - n\left(\frac{n_g}{n}\right)^2}$$

since $\sum_{n=1}^{n} x_i y_i = \sum_{g} y_{gi}$, $\sum x_i^2 = n_g$

$$= \frac{\left(\sum_{n_g} y_{gi} - n\bar{y}\right) n}{n \cdot n_g - n_g^2} = \frac{n \cdot \sum_{n_g} y_{gi} - n_g \sum_{n_g} y_{gi}}{n_g(n - n_g)}$$

$$= \frac{(n - n_g)\sum_{g} y_{gi}}{(n - n_g) n_g} = \frac{1}{n_g}\sum_{g} y_{gi} = \bar{y}_g$$

io: for β_0: it means the level of Y when no factor level involved

for others. the ~~xx~~ interpretations are the same

H_0: ~~$\beta_1 = \beta_2 = \dots = \beta_g$~~ $\mu_A = \mu_B = \mu_g = \dots$

H_a: at least one mean μ_g differs from others

4. I will use Poisson Regression Test

Because dose and days are considered categorical variables

and sample size is large enough

(since we need $\lambda \geq 5$)

Scatterplot Matrix:

We have a ~~nearly linear relationship between~~ a correlation between dose=10 and numbers of termites (almost a linear relationship).

but for dose = 5mg. the correlation is weak and scatter plots random distributed.

$$\hat{\beta}_g = \frac{\sum x_i y_i - n\bar{x}\bar{y}}{\sum x_i^2 - n\bar{x}^2} = \frac{\sum_{n_g} y_{gi} - n\bar{y}}{n_g - n\left(\frac{n_g}{n}\right)^2}$$

since $\sum_{n=1}^{n} x_i y_i = \sum_g y$

$\sum x_i^2 = n_g$

$$= \frac{\left(\sum_{n_g} y_{gi} - n\bar{y}\right) n}{n \cdot n_g - n_g^2} = \frac{n \cdot \sum_{n_g} y_{gi} - n_g \sum_{n_g} y_{gi}}{n_g(n - n_g)}$$

$e^{-0.4815 - 1.1317 x}$

e) no. we can't.
because 30mg exceed our scope range.

f) ~~$\frac{e^{-0.64238 - 0.3481 \times 5}}{e^{-0.64238 - 0.3481 \times 10}}$~~

95% CI: $\beta_1 \pm 1.96 \times 0.04786$.

$$\Rightarrow \frac{e^{-0.64238 - 0.3481 \times 5 \pm 1.96 \times 0.0}}{e^{-0.64238 - 0.3481 \times 10 \pm 1.96 \times 0.04786}}$$

g) No. For log-linear model.
we need sample size ≥ 5
since it's already $16 > 5$.
Assumption is satisfied.

h)

$H_0: \beta_1 = 0 \quad H_1: \beta_1 \neq 0$

$G^2 = -2\log\left(\frac{L_r}{L_f}\right) = -2\log(L_r) + -2\log$

d)

$IJK - 7$

~~Sat~~ Degree of (saturated model − Fitted model)

e) Every pair of variation (~~a~~ variable 1.2.3) of all is associated. but we don't have an interaction.

Fail to reject Ho. i.e. uniform model is adequate.

3.

a) A: $Z = \frac{-1.1318 - 0}{0.2398} = \checkmark$

B: 14

C: 117.821 − 84.018

b) Model 1:

$\log\left(\frac{\pi}{1-\pi}\right) = \log(Y) = -0.4885 - 1.1318 \, I[\text{Dose Level}]10$

Model 2:

$\log\left(\frac{\pi}{1-\pi}\right) = \log(Y) = 0.1423f - 0.2263f$ Dose.

c) β_1: The effects by whether the dose level is sng / long

β_2: The effect by the amount of dose

H_a: at least one ... differs from others

b)

a) ③ Binary Model $\log\left(\frac{\pi_i}{1-\pi_i}\right) = \beta_0 + \beta_1 I_L + \beta_2 I_M + \beta_3$

i) ~~① ② ③ ⑥ One-way ANOVA with 3 levels~~

~~$Y = \beta_0 + \beta_1 I_L + \beta_2 I_M + \beta_3 I_H + e$~~

ii) ① ② ⑤ ⑥ ④

One-way ANOVA with 3 levels

$Y = \beta_0 + \beta_1 I_{L,S} + \beta_2 I_{M,S} + \beta_3 I_{H,S} + e$

iii) ① ② ④ ⑥

One-way ANOVA $Y = \beta_0 + \beta_1 I_L + \beta_2 I_M + \beta_3 ... + e$

b) i) ③ ④ ⑤

ii) ① ②

iii) ① ② ⑥

$$= \frac{(n-n_g)\sum_{g} y_{gi}}{(n-n_g)n_g} = \frac{1}{n_g}\sum_{g} y_{gi} = \bar{y}_g$$

iv. for β_0: it means the level of Y when no factor level involved

for others. the ~~...~~ interpretations are the same

H_0: $\mu_A = \mu_B = \mu_g = \cdots$ ~~$\beta \ldots \beta_g$~~

H_a: at least one mean μ_g differs from others

③ Binary Model. $\log\left(\frac{\pi_i}{1-\pi_i}\right) = \beta_0 + \beta_1 I_L + \beta_2 I_M + \beta_3 I_H$

~~②①⑥ one-way ANOVA with 3 levels~~

~~$Y = \beta_0 + \beta_1 I_L + \beta_2 I_M + \beta_3 I_H + e$~~

②⑤⑥ ④

One-way ANOVA with 3 levels

$Y = \beta_0 + \beta_1 I_{L,S} + \beta_2 I_{M,S} + \beta_3 I_{H,S} + e$

①②⑤⑥ one-way ANOVA $Y = \beta_0 + \beta_1 I_L + \beta_2 I_M + \beta_3 I_H + e$

i) ③④⑤

$$= \frac{(n-n_g)\sum_g y_{gi}}{(n-n_g)n_g} = \frac{1}{n_g}\sum_g y_{gi} = \bar{y}_g$$

iv: for β_0: it means the level of Y when no factor level included

for others: the ~~…~~ interpretations are the same

H_0: $\mu_A = \mu_B = \mu_g = \cdots$

H_a: at least one mean μ_g differs from others

③ Binary Model $\log\left(\frac{\pi_i}{1-\pi_i}\right) = \beta_0 + \beta_1 I_L + \beta_2 I_M + \beta_3 I_H$

~~②④⑤⑥ One way ANOVA with 3 levels~~

~~$Y = \beta_0 + \beta_1 I_L + \beta_2 I_M + \beta_3 I_H + e$~~

② ⑤ ⑥ ④ One-way ANOVA with 3 levels

$Y = \beta_0 + \beta_1 I_{L,S} + \beta_2 I_{M,S} + \beta_3 I_{H,S} + e$

① ② ⑤ ⑥ one-way ANOVA $Y = \beta_0 + \beta_1 I_L + \beta_2 I_M + \beta_3 I_H + e$

ii/ ③ ④ ⑤