

# Applied Review Prof Rashid 7/16/19

1 a) 1.5

b) 4 or more,  $X \sim \text{Poisson}(1.5)$   $\Pr(X \geq 4) = 0.0656$

Among 453 expect 29.7

Test if observed proportion is = to theoretical  $\alpha$ ?

$H_0: \pi = 0.0656$  vs  $H_1: \pi \neq 0.0656$

↪ 1 sample test of 1 proportion.

Under  $H_0$ :  $Z = \frac{\bar{Y} - n\pi}{\sqrt{n\pi(1-\pi)}} \sim N(0, 1)$   $Z = -5.6$

$-5.6 < -1.96$ ,

reject null, observed

number (zero) is sig less than the expected #

c)  $Y$  denote GA measured by ultrasound

→ Shapiro Wilk is acceptable, Kolmogorov-Smirnov

$H_0: Y \sim N(\mu, \sigma^2)$   $H_1: Y$  not normally distributed

d) evaluate agreement  $\rightarrow \text{IRR} \in$  package in R

- Cohen's Kappa  $\rightarrow$  chisq

$K = 0.69$

$\alpha: (63, 75)$

- moderate agreement:  $0.4 - 0.7$

- strong agreement:  $> 0.7$

— fmsb package also

e)  $\chi^2$  test of trend,  $P_i = \text{prob of preterm delivery}$

operative word: monotonically  $\rightarrow$  implying trend

$H_0: P_0 = P_1 = P_2 = P_3$

$H_A: P_0 < P_1 < P_2 < P_3$  or  $P_0 > P_1 > P_2 > P_3$

Cochran-Armitage trend test

don't reject  $H_0 \rightarrow$  not much evidence for a

monotonic trend in risk for preterm delivery w/ number of previous pregnancies,

SAS output  
 1 e)  $\chi^2_{\text{trend}} = -0.0076^2 = 0.0001, p=0.9939$

(continuous version  $\rightarrow$  setup as a cell-means

1-way ANOVA  $\rightarrow$  similar  $H_0, H_A$

- if you want to impose requirement of linearity,

it's linear regression (predictor is 1 column, diff values)

$\hookrightarrow$  tree heights at alpine levels

full model:  $H_0: \mu_1 = \mu_2 = \mu_3 = \mu_4 \rightarrow$  "reduced model"  
 $H_A: \text{otherwise} \rightarrow$  "full model"

Corresponding linear model:  $\rightarrow Y_{ij} = \mu + \text{grand mean} \quad \downarrow$   
 $\rightarrow Y_{ij} = \mu_j + \epsilon_{ij} \quad \xrightarrow{i=1 \dots n} N(0, \sigma^2) \quad \xrightarrow{j=0, \dots 3}$  F-test

Test for trend  $\quad \quad \quad$  mean for zero group

$$H_0: Y_{ij} = \mu_0 + \beta_1 x_i \quad (\beta_1 \text{ is what mean } \mu_{0j})$$

$$H_A: \mu_j + \epsilon_{ij} = Y_{ij} \quad \leftarrow \text{means can be different}$$

whatever

General linear hypothesis testing framework

1)  $\chi^2$  test of association (or whether proportion is equal in 2 groups), Don't reject null

$H_0:$  preterm  $\perp$  of treatment group  $\quad p\text{value} = 0.4303$

$H_A:$  "  $\quad \quad \quad$  values: 0.6221

$\rightarrow$  no evidence treatment reduces the risk of preterm delivery

g) can't use 2 sample test

- calculate difference, test if mean is zero  
large sample approximation

$$z = (0.0034 - 0) / \sqrt{0.2493/176} = 0.0897 > 0.96$$

h) Is there a difference in the difference?  
sample sizes are large  $\rightarrow$  CLT

Shtskiy's Theorem to test Z  $H_0: \mu_1 = \mu_2$

$$Z = \frac{(\bar{Y}_1 - \bar{Y}_2) - \delta}{\sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}}}$$

reject null  $\rightarrow$  there's sig diff

- i)  
g) didn't improve  
h) sig diff in change

Therapy doesn't seem to improve status,  
appear to prevent deterioration + may be  
potential to have an effect on birth outcomes

2) a) i)  $\rightarrow$  simulating ANOVA  
read.table

suppresses  
int.

obj = lm(response ~ factor(Klin-sub) - 1,  
data = data[data\$treatment == 1, ]

summary(obj)

$$\vec{Y} = \vec{X} \vec{b} + \varepsilon$$

$$\begin{bmatrix} Y_1 \\ \vdots \\ Y_{168} \end{bmatrix}_{168 \times 1} = \begin{bmatrix} 1 & 0 \\ 1 & 0 \\ 0 & 1 \\ \vdots & \vdots \end{bmatrix}_{168 \times 2} \begin{bmatrix} \mu_1 \\ \mu_2 \end{bmatrix}_{2 \times 1}$$

$\mu_1$  = mean response for group 1

$\mu_2$  = mean response for group 2

$$F: H_0: \mu_1 = \mu_2 = \mu_0$$

$\hookrightarrow$  then intercept only model!

$$\begin{bmatrix} 1 \\ 1 \end{bmatrix} \mu_0 = \mu_1 = \mu_2$$

$$\Theta = CB$$

contrast,  $1 \times 2 \rightarrow [1, -1]$

Ask Ann about the reduced model method

"obj0"  $\leftarrow$   $H_0: \mu_1 = \mu_2 = \mu_0$   $\leftarrow$  reduced model (intercept only model)  
"obj"  $\leftarrow$   $H_A: \mu_1 \neq \mu_2$   $\leftarrow$  full model

ANOVA(obj, obj0)

in R: ~ 1  
 $\hookrightarrow$  "obj0"

2a) ii) if this was ref cell  $\rightarrow$  then test is if  $\beta_1 = 0$

iv)  $H_0: \mu_1 = 2\mu_2$  vs  $H_A: \mu_1 \neq 2\mu_2$

$c: [2 -1]$  (might be flipped)

$\rightarrow$  This Q is about setting up C properly  
reject null

means  $\begin{matrix} 63 \\ 90 \end{matrix}$

v)  $\rightarrow \hat{\sigma}^2 (X'X)^{-1}$  (in SAS, procreg, covb)

\* If you can be specific, be specific \*

2b) i) interaction here  $\rightarrow$  main effects for each + interaction

m1 is full model  $\rightarrow$  ref cell

✓ mean of ref group  $\rightarrow$  chemo + 1st clinical subtype

$\beta_0 + \beta_1$  clin-subtype +  $\beta_2$  treatment,  $\beta_3$  clinxtreat

$H_0: \beta_3 = 0, H_A: \beta_3 \neq 0 \rightarrow$  Wald test

addresses this

$\rightarrow$  default test in R

equivalent contrast:  $([0 0 0 1])$

3)

[56	14
24	6

$$O-E = 4$$

$$\chi^2 = 4^2 (1/56 + 1/14 + 1/24 + 1/6) = 4.762$$

$\chi^2$ , p-value  $\approx 0.291$

2nd group:

b) Odds ratio, Kappa, pearson correlation

Odds ratio

non-HT: 3, CI: (1.09, 8.25)

HT: 3.273 CI: (1.37

c) McNemar's test, p-value from  $\chi^2$

don't: compare independent proportions  
or test of independence

$$(20-10)^2/(10+10), p = 0.06789$$

d) 2x4 table:

60	10	20	10
48	12	22	18

expected: [54 11 21 14]

$$|O-E| = 6 \quad 1 \quad 1 \quad 4$$

$$\chi^2 =$$

$$\chi^2_3$$

e) assumptions: binary outcome, not means  
logit( $P$ )(low birth weight)

$\beta_1$  = log odds of reference group

$\beta_2$  = change in log odds when  $x=1$

$$\hat{\beta}_1 = \log(30/70), \hat{se} = \sqrt{Y_{30} + Y_{70}}$$

$$\hat{\beta}_2 = \log(2800, 1800) \quad \hat{se} = \sqrt{Y_{30} + Y_{70} + Y_{40} + Y_{60}}$$

f)  $\gamma_3$ : log ratio of odds of LBW in 1<sup>st</sup> child  
to odds of LBW in 2<sup>nd</sup> child, non-hypertensive  
women,  
interaction model

$\gamma_4$ : how much that log odds ratio in non-hyp  
women is higher than comparable log odds  
ratio in hyp. women (contrast between  
hyp and non-hyp women)

g) No, all samples are not independent (each  
unit has 2 observations  $\rightarrow$  correlated) not  $\perp$

proc logistic assumes  $\perp$

point estimates ok, but standard errors, p-values,  
CIs not

2) ANCOVA model

b) i) none of variables pertaining to diabetes (main effect,  
interaction term) = 0  $\rightarrow$  two rows in C matrix