0917

7.28 Let p be the proportion of deaths from lung cancer in this plant.

$H_0: p = 0.12$ $\quad$ $H_1: p \neq 0.12$

7.29 two-sided

7.30 Let X be the number of deaths due to lung cancer.

$$P(X \geq 5 \mid H_0 \text{ is true}) = \sum_{i=5}^{20} C_{20}^{i}\, 0.12^{i} (1-0.12)^{20-i}$$
$$= 0.082719$$

binom.test (5, 20, 0.12, alt = "two.sided)

7.31 $Z = \frac{\hat{p} - p_0}{\sqrt{p_0(1-p_0)/n}}$ where $\hat{p} = 19/90$, $p_0 = 0.12$, $n = 90$, $Z_0 = 2.6599$

when H_0 is true, $Z \sim N(0,1)$

$P(|Z| > Z_0 \mid Z \sim N(0,1)) = 2 \times 0.00391 = 0.00782$

7.32 the ~~ratt~~ proportion of deaths from lung cancer when deaths caused by IHD excluded is p_0

$$p_0 = \frac{0.12}{1-0.4} = 0.2$$

$$\hat{p} = \frac{19}{90-18} = \frac{19}{72}, \quad n = 90 - 18 = 72$$

$$z_0 = \frac{\hat{p} - p_0}{\sqrt{p_0(1-p_0)/n}} = 1.355288$$

$$P(|Z| > z_0 \mid Z \sim N(0,1)) = 0.174576$$

7.51 one-sample t test

7.52 let μ be the mean of stomach-cancer cases.

$H_0: \mu = 2.88 \quad H_1: \mu \neq 2.88$

$$T = \frac{\bar{X} - \mu}{s/\sqrt{n}} \overset{H_0}{\sim} t(19)$$

when $\bar{x} = 2.65$, $s/\sqrt{n} = se = 0.11$

$T = -2.09091$

while $-t_{0.975}(19) = -2.093024 < T$

thus ~~do~~ we reject H_0 at 95% confidence level.

$$p = P(|T| > 2.09091 \mid T \sim t(19)) = 0.05021$$

7.53 pwr.t.test(d = 0.2/0.11/sqrt(20), power = 0.8,
type = "one.sample", alt = "two.sided)

$$d = \frac{\text{difference between the means}}{\text{pooled standard deviation (here just } se\sqrt{n} = s)}$$

8.25 μ_1: mean concentration of drug A
μ_2: mean concentration of drug B

$H_0: \mu_1 = \mu_2 \quad H_1: \mu_1 \neq \mu_2$

8.26 use paired t test

8.27 d_i: the difference between drug A and drug B

$$T = \frac{\bar{d}}{S_d/\sqrt{n}} \overset{H_0}{\sim} t(n-1)$$

when $S_d = \sqrt{\frac{\sum(d_i-\bar{d})^2}{n-1}} = 3.0984$

$\bar{d} = 3.6$

$T = 3.6742$

$$P(|T| > 3.6742 \mid T \sim t(n-1)) = 0.0051 < 0.05$$

thus we reject H_0 at 95% confidence level.

8.28. $\bar{d} = 3.6$

8.29

$$\frac{|d-\bar{d}|}{S_d/\sqrt{n}} < t_{0.975}(9)$$

$-1.3835 < d < 5.8164$

0924

9.7 wilcoxn rank-sum test

$F(x)$: the p.d.f. of birthweight in treatment group

$G(x)$: the p.d.f of birthweight in control group

H_0: $F(x) = G(x)$ H_1: $F(x) \neq G(x)$

具体计算过程参照课件

$W = 170.5$, $p = 0.01692$

9.8 since the sample sice is small and the distribution of birthweight is ~~ule~~ unknown, parametric methods are not preferable.

9.15 m_1: median duration of effussion in breastfed body.

m_2: median duration of effussion in bottle-fed body.

H_0: $m_1 = m_2$ H_1: $m_1 < m_2$

9.17 wilcoxon signed-rank test

9.18 H_0: $m_1 = m_2$ H_1: $m_1 \neq m_2$

具体计算过程参照课件

$W^+ = 61$, $E(W^+) = 138$, $Var(W^+) = 1079$

$P = 0.01986$

1008

$H_0: p_1 = p_2 \qquad H_1: p_1 \neq p_2$

$$Z = \frac{\hat{p}_1 - \hat{p}_2}{\sqrt{\bar{p}(1-\bar{p})(\frac{1}{n_1}+\frac{1}{n_2})}} \overset{H_0}{\sim} N(0,1) \qquad \bar{p} = \frac{p_1 + k p_2}{1+k}$$

$$Power = P\left(Z > Z_{1-\frac{\alpha}{2}} \,\middle|\, |p_1 - p_2| = \Delta \neq 0 \right)$$

$$= P\left(\frac{\hat{p}_1 - \hat{p}_2}{\sqrt{\bar{p}\bar{q}(\frac{1}{n_1}+\frac{1}{n_2})}} > Z_{1-\frac{\alpha}{2}} \,\middle|\, |p_1 - p_2| = \Delta \neq 0 \right)$$

$$= P\left(\frac{\hat{p}_1 - \hat{p}_2 - \Delta}{\sqrt{p_1 q_1/n_1 + p_2 q_2/n_2}} > Z_{1-\frac{\alpha}{2}} \frac{\sqrt{\bar{p}\bar{q}(\frac{1}{n_1}+\frac{1}{n_2})}}{\sqrt{p_1 q_1/n_1 + p_2 q_2/n_2}} - \frac{\Delta}{\sqrt{p_1 q_1/n_1 + p_2 q_2/n_2}} \right)$$

$$\frac{\hat{p}_1 - \hat{p}_2 - \Delta}{\sqrt{p_1 q_1/n_1 + p_2 q_2/n_2}} \overset{H_1}{\sim} N(0,1)$$

$$\Rightarrow \quad 1-\beta = \Phi\left(\frac{\Delta}{\sqrt{p_1 q_1/n_1 + p_2 q_2/n_2}} - Z_{1-\frac{\alpha}{2}} \frac{\sqrt{\bar{p}\bar{q}(\frac{1}{n_1}+\frac{1}{n_2})}}{\sqrt{p_1 q_1/n_1 + p_2 q_2/n_2}} \right)$$

$$Z_{1-\beta} = \frac{\Delta}{\sqrt{p_1 q_1/n_1 + p_2 q_2/n_2}} - Z_{1-\frac{\alpha}{2}} \frac{\sqrt{\bar{p}\bar{q}(\frac{1}{n_1}+\frac{1}{n_2})}}{\sqrt{p_1 q_1/n_1 + p_2 q_2/n_2}}$$

$n_2 = k n_1$ 代入

得 $n_1 = \left[\sqrt{\bar{p}\bar{q}(1+\frac{1}{k})}\, Z_{1-\frac{\alpha}{2}} + \sqrt{p_1 q_1 + \frac{p_2 q_2}{k}}\, Z_{1-\beta} \right]^2 / \Delta^2$

$n_2 = k n_1$

11.12 推导 E(MSW) 和 E(MSB) 两个表达式

$y_{ij}=\mu+\alpha_i+e_{ij}$， $\alpha_i \sim N(0,\sigma_A^2)$ ， $e_{ij}\sim N(0,\sigma^2)$ $i=1,\cdots,k, j=1,\cdots n_i$

$MSW=\frac{1}{n-k}\sum_i^k\sum_j^{n_i}(y_{ij}-\bar{y}_i)^2$

$\bar{y}_i=\frac{1}{n_i}\sum_j^{n_i}y_{ij}=\mu+\alpha_i+\frac{1}{n_i}\sum_j^{n_i}e_{ij}$

$\therefore E(y_{ij}-\bar{y}_i)=E(e_{ij}-\frac{1}{n_i}\sum_j^{n_i}e_{ij})=0$

$Var(y_{ij}-\bar{y}_i)=Var(e_{ij}-\frac{1}{n_i}\sum_j^{n_i}e_{ij})$

$=Var(e_{ij})+\frac{1}{n_i^2}Var(\sum_j^{n_i}e_{ij})-2cov(e_{ij},\frac{1}{n_i}\sum_j^{n_i}e_{ij})$

$=\sigma^2+\frac{\sigma^2}{n_i}-\frac{2}{n_i}\sigma^2$

$=(1-\frac{1}{n_i})\sigma^2$

$E(MSW)=\frac{1}{n-k}\sum_i^k\sum_j^{n_i}E(y_{ij}-\bar{y}_i)^2$

$=\frac{1}{n-k}\sum_i^k\sum_j^{n_i}[E^2(y_{ij}-\bar{y}_i)+Var(y_{ij}-\bar{y}_i)]$

$=\frac{1}{n-k}\sum_i^k\sum_j^{n_i}(1-\frac{1}{n_i})\sigma^2$

$=\sigma^2$

$MSB=\frac{1}{k-1}\sum_i^k\sum_j^{n_i}(\bar{y}_i-\bar{y})^2$

$\bar{y}=\frac{1}{n}\sum_i^k\sum_j^{n_i}y_{ij}=\frac{1}{n}\sum_i^k\sum_j^{n_i}(\mu+\alpha_i+e_{ij})$

$=\mu+\frac{1}{n}\sum_i^k n_i\alpha_i+\frac{1}{n}\sum_i^k\sum_j^{n_i}e_{ij}$

$\therefore \bar{y}_i-\bar{y}=\alpha_i-\frac{1}{n}\sum_i^k n_i\alpha_i+\frac{1}{n_i}\sum_j^{n_i}e_{ij}-\frac{1}{n}\sum_i^k\sum_j^{n_i}e_{ij}$

$E(\bar{y}_i-\bar{y})=0$

$Var(\bar{y}_i-\bar{y})=Var(\alpha_i-\frac{1}{n}\sum_i^k n_i\alpha_i)+Var(\frac{1}{n_i}\sum_j^{n_i}e_{ij}-\frac{1}{n}\sum_i^k\sum_j^{n_i}e_{ij})$

$Var(\alpha_i-\frac{1}{n}\sum_i^k n_i\alpha_i)=Var(\alpha_i)+Var(\frac{1}{n}\sum_i^k n_i\alpha_i)-2cov(\alpha_i,\frac{1}{n}\sum_i^k n_i\alpha_i)$

$=\sigma_A^2+\frac{\sum_i^k n_i^2\sigma_A^2}{n^2}-\frac{2n_i\sigma_A^2}{n}$

$Var(\frac{1}{n_i}\sum_j^{n_i}e_{ij}-\frac{1}{n}\sum_i^k\sum_j^{n_i}e_{ij})=\frac{\sigma^2}{n_i}+\frac{\sigma^2}{n}-\frac{2}{n}\sigma^2=(\frac{1}{n_i}-\frac{1}{n})\sigma^2$

$\therefore E(MSB) = \frac{1}{k-1} \sum_i^k \sum_j^{n_i} E(\bar{y}_i - \bar{y})^2$

$= \frac{1}{k-1} \sum_i^k \sum_j^{n_i} [E^2(\bar{y}_i - \bar{y}) + Var(\bar{y}_i - \bar{y})]$

$= \frac{1}{k-1} \sum_i^k \sum_j^{n_i} (\sigma_A^2 + \frac{\sum^k n_i^2 \sigma_A^2}{n^2} - \frac{2 n_i \sigma_A^2}{n} + \frac{\sigma^2}{n_i} - \frac{\sigma^2}{n})$

$= (n\sigma_A^2 + \frac{\sum_i^k n_i^2 \sigma_A^2}{n} - \frac{2 \sum n_i^2 \sigma_A^2}{n} + k\sigma^2 - \sigma^2) / (k-1)$

$= [n\sigma_A^2 - \frac{\sum_i^k n_i^2 \sigma_A^2}{n} + (k-1)\sigma^2] / (k-1)$

$= \frac{n - \sum_i^k n_i^2 / n}{k-1} \sigma_A^2 + \sigma^2$

when $n_1 = n_2 = \cdots = n_k$, $\frac{n - \frac{1}{n} \sum_i^k n_i^2}{k-1} = n_1$

1126

12.37 计算前可对数据进行对数变换

random effect model: $y_{ij} = \mu + \alpha_i + e_{ij}$ $\quad k=30$

where $\alpha_i \sim N(0, \sigma_A^2)$, $e_{ij} \sim N(0, \sigma^2)$

y_{ij}: temperature at jth location on ith day

$$MSB = \frac{SSB}{k-1} = \frac{\sum_{i=1}^{k}\sum_{j=1}^{n_i}(\bar{y}_i - \bar{y})^2}{k-1} = \overset{\log}{0.02934} / 131.906$$

$$MSW = \frac{SSW}{n-k} = \frac{\sum_{i=1}^{k}\sum_{j=1}^{n_i}(y_{ij} - \bar{y}_i)^2}{n-k} = \overset{\log}{0.0006735} / 2.861$$

where $\bar{y}_i = \frac{\sum_{j=1}^{n_i} y_{ij}}{n_i}$, $\bar{y} = \sum_i^k \sum_j^{n_i} y_{ij} / \sum_i^k n_i$ $\quad n_1 = n_2 = \cdots = n_{30} = 21$

$$\hat{\sigma}_A^2 = \frac{MSB - MSW}{21} = \cdots$$

12.38. here, $k=21$, $n_1 = n_2 = \cdots = n_{21} = 30$

$y_{ij} = \mu + \alpha_j + e_{ij}$ $\quad \alpha_j \sim N(0, \sigma_B^2)$, $e_{ij} \sim N(0, \sigma^2)$

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

$$F = \frac{MSB}{MSW} \overset{H_0}{\sim} F(k-1, n-k)$$

$$MSB = \frac{\sum_i^k \sum_j^{n_i} (\bar{y}_j - \bar{y})^2}{k-1} = \overset{\log}{0.0083} / 35.65$$

$$MSW = \frac{\sum_i^k \sum_j^{n_i} (y_{ij} - \bar{y}_j)^2}{n-k} = \overset{\log}{0.00179} / 7.929$$

where $\bar{y}_j = \sum_{i=1}^{k} y_{ij} / k$

$F^* > F_{0.95}(20, 609)$, reject H_0.

12.39

fixed effect model: $y_{ij} = \mu + \alpha_i + e_{ij}$, $e_{ij} \sim N(0, \sigma^2)$

$H_0: \alpha_i = \alpha_j$, $H_1: \alpha_i \neq \alpha_j$ $(\forall i \neq j, i < j)$

t test (unpaired)

then use Bonferroni method to get correct assessments.

12.68 & 12.69 按照课件计算公式即可

1203

1. OR 适用于 case-control study

RR. RD. OR 都可用于 prospective study

因为在 RR 和 RD 的计算中涉及 exposed 和 not exposed 中的患病率，而在回顾性研究中患病率是无法估计的，OR 的计算中涉及的是人群暴露率，这在两种研究中都可获得估计。

2.

	Y=0	Y=1
X=0	A	B
X=1	C	D

Y：疾病发生情况

X：暴露情况

$$\hat{OR} = \frac{\hat{p}_1/(1-\hat{p}_1)}{\hat{p}_2/(1-\hat{p}_2)} \qquad \hat{p}_1 = \frac{D}{C+D},\ \hat{p}_2 = \frac{B}{A+B}$$

$$= \frac{AD}{BC} \qquad n_1 = C+D,\ n_2 = A+B$$

$$Var(\ln\hat{OR}) = Var\left(\ln\frac{\hat{p}_1}{1-\hat{p}_1}\right) + Var\left(\ln\frac{\hat{p}_2}{1-\hat{p}_2}\right)$$

由 delta-method，令 $g(x) = \ln\frac{x}{1-x}$，$g'(x) = \frac{1}{x(1-x)}$

则 $Var\left(\ln\frac{\hat{p}_1}{1-\hat{p}_1}\right) \approx (g'(\hat{p}_1))^2 \frac{\hat{p}_1(1-\hat{p}_1)}{n_1}$

$$= \frac{1}{\hat{p}_1^2(1-\hat{p}_1)^2}\cdot\frac{\hat{p}_1(1-\hat{p}_1)}{n_1} = \frac{1}{n_1\hat{p}_1(1-\hat{p}_1)} = \frac{1}{C}+\frac{1}{D}$$

同理可得 $Var\left(\ln\frac{\hat{p}_2}{1-\hat{p}_2}\right) \approx \frac{1}{n_2\hat{p}_2(1-\hat{p}_2)} = \frac{1}{A}+\frac{1}{B}$

故 $Var(\ln\hat{OR}) \approx \frac{1}{A}+\frac{1}{B}+\frac{1}{C}+\frac{1}{D}$

假设 Mc Nemar's test 中列联表如下：

		Case		
		Exposed	Not exposed	Total
Control	Exposed	n_{11}	n_{12}	$n_{11}+n_{12}$
	Not exposed	n_{21}	n_{22}	$n_{21}+n_{22}$
	Total	$n_{11}+n_{21}$	$n_{12}+n_{22}$	n

共有 n 个对，在 Mantel-Haenszel test 中可对应分成 n 层 $(a_i+b_i=1, c_i+d_i=1)$

有 n_{11} 层：

	exposed	Not exposed
case	1	0
control	1	0

$O_i=a_i=1$

$E_i=\frac{(a_i+b_i)(a_i+c_i)}{n_i}=1$

$V_i=\frac{(a_i+b_i)(a_i+c_i)(b_i+d_i)(c_i+d_i)}{n_i^2(n_i-1)}$

$=0$

有 n_{12} 层

	exposed	Not exposed
case	0	1
control	1	0

$O_i=0$

$E_i=\frac{1}{2}$

$V_i=\frac{1}{4}$

有 n_{21} 层

	exposed	Not exposed
case	1	0
control	0	1

$O_i=1$

$E_i=\frac{1}{2}$

$V_i=\frac{1}{4}$

有 n_{22} 层

	exposed	Not exposed
case	0	1
control	0	1

$O_i=0$

$E_i=0$

$V_i=0$

则 $O=\sum O_i=n_{11}+n_{21}$

$E=\sum E_i=n_{11}+\frac{1}{2}n_{12}+\frac{1}{2}n_{21}$

$V=\sum V_i=\frac{1}{4}n_{12}+\frac{1}{4}n_{21}$

$$X^2_{MH}=\frac{(O-E)^2}{V}=\frac{(n_{12}-n_{21})^2}{n_{12}+n_{21}}\overset{H_0}{\sim}\chi^2(1)$$

其中 $n_{12}+n_{21}=n_D$（不一致对数），故 X^2_{MH} 亦为 Mc Nemar's test 中的统计量。