STAT 659 Spring 2016
Homework 2 Solution

1.6

(a) $n_3 = n - n_1 - n_2$.

(b) The possible observations are:

$$(3,0,0),(0,3,0),(0,0,3),(2,1,0),(2,0,1),(1,2,0),(0,2,1),(1,0,2),(0,1,2),(1,1,1)$$

(c) The probability is $\frac{3!}{1!2!0!}(1/4)^1(1/2)^2(1/4)^0 = 3/16 = 0.1875$.

(d) n_1 has a binomial distribution with sample size $n = 3$ and probability of success $p = 0.25$.

1.20

(a) The sample mean and sample standard deviation for treatment A is $\bar{x}_A = 5, \mathrm{SD}_A = 2.055$ while the sample mean and sample standard deviation for treatment B is $\bar{x}_B = 9, \mathrm{SD}_B = 2.906$. The Wald interval for mean number of imperfections for group A is $\bar{x}_A \pm 1.96\sqrt{\bar{x}_A/n_A} = (3.614, 6.386)$ and the Wald interval for treatment B is $\bar{x}_B \pm 1.96\sqrt{\bar{x}_B/n_B} = (7.141, 10.859)$; the score interval is $\hat{\lambda} + \frac{1.96^2}{2n} \pm \sqrt{\hat{\lambda}\frac{1.96^2}{n} + \frac{1.96^4}{4n^2}}$. By plugging $\hat{\lambda}_A = 5, \hat{\lambda}_B = 9$, we can obtain the score interval for treatment A is $(3.793, 6.591)$ and the score interval for treatment B is $(7.323, 11.061)$.

(b) For poisson distribution, the mean equals to the variance. For treatment A, the sample mean is 5 which is very close to the sample variance 4.22 and also for treatment B, its sample mean is 9 which is close to sample variance 8.44. For the goodness test, $Z_A = -0.33$, $Z_B = -0.13$. There is not sufficient evidence to conclude that the data from either treatment is not from a Poisson distribution.

1.21

If all the numbers have the same probability, the expected observations are 15. The test statistics is $TS = \sum_{i=0}^{9} \frac{(n_i - \mu_i)^2}{\mu_i} = 11.2 < \chi^2_{0.95,9} = 16.92$. So we cannot reject the null hypothesis.

1.22

The sample size $n = 1301$ and if the probabilities for each type of plants are true, then the expected observations are $\mu_1 = 731.81, \mu_2 = 243.94, \mu_3 = 243.94, \mu_4 = 81.31$.The test statistics is $TS = \sum_{i=1}^{4} \frac{(n_i - \mu_i)^2}{\mu_i} = 6.48 < \chi^2_{0.95,3} = 7.81$. Thus the theoretical proportions are consistent with the data.

1.23

The sample size $n = 3839$ and if the probabilities for each type of plants are true, then the expected observations are $\mu_1 = 1953.76, \mu_2 = 925.49, \mu_3 = 925.49, \mu_4 = 34.26$. So the test statistics is $TS = \sum_{i=1}^{4} \frac{(n_i - \mu_i)^2}{\mu_i} = 2.016 < \chi^2_{0.95,2} = 5.9915$. Thus the theoretical proportions are consistent with the data.

2.1

(a) $, P(-|C) = \frac{1}{4}$ and $, P(+|\bar{C}) = \frac{2}{3}$ are correct.

(b) Sensitivity is $P(+|C)$. Hence, $, P(+|C) = 1 - P(-|C) = \frac{3}{4}$.

(c) Since $, P(C) = 0.01$, $, P(+ \cap C) = P(C)P(+|C) = 0.0075$, $, P(- \cap C) = P(C)P(-|C) = 0.0025$, $, P(+ \cap \bar{C}) = (1 - P(c))P(+|\bar{C}) = 0.66$ and $, P(- \cap \bar{C}) = (1 - P(C))P(-|\bar{C}) = 0.33$.

(d) $, P(+) = P(+ \cap C) + P(+ \cap \bar{C}) = 0.6675$ and $, P(-) = P(- \cap C) + P(- \cap \bar{C}) = 0.3325$.

(e) $, P(C|+) = \frac{P(C \cap +)}{P(+)} = \frac{P(+|C)P(C)}{P(+)} = 0.0112$,which is the probability that men diagnosed with prostate cancer given that the test result is positive is 0.01124.

2.2

(a) $\pi_1 = P(Y = 1|X = 1)$, which is the probability the diagnosis test is positive given a subject has disease, so it is the sensitivity according to the definition. $1 - \pi_2 = 1 - P(Y = 1|X = 2) = P(Y = 2|X = 2)$, which is the probability the diagnosis test is negative, given a subject does not have disease, so it is the specificity due to its definition.

(b) $\gamma = P(X = 1)$. So due to the Bayes's theorem, $P(X = 1|Y = 1) = \frac{P(Y=1|X=1)P(X=1)}{P(Y=1)} = \frac{P(Y=1|X=1)P(X=1)}{P(Y=1|X=1)P(X=1)+P(Y=1|X=2)P(X=2)}$ which is $\frac{\pi_1\gamma}{\pi_1\gamma+\pi_2(1-\gamma)}$.

(c) Plug in $\gamma = 0.01, \pi_1 = 0.86, \pi_2 = 1 - 0.88 = 0.12$, we have $P(X = 1|Y = 1) = \frac{0.01*0.86}{0.01*0.86+0.99*0.12} = 0.0675$.

(d) $P(X=1,Y=1)=P(Y=1|X=1)P(X=1)=\pi_1\gamma=0.0086, P(X=1,Y=2)=P(Y=2|X=1)P(X=1)=(1-\pi_1)\gamma=0.0014, P(X=2,Y=1)=P(Y=1|X=2)P(X=2)=\pi_2(1-\gamma)=0.1188, P(X=2,Y=2)=P(Y=2|X=2)P(X=2)=(1-\pi_2)(1-\gamma)=0.8712$. We can see $P(X=2,Y=1)$ is a bit greater than $P(X=1,Y=1)$, this is because although the sensitivity is large, the probability of the disease is very small which leads to smaller joint probability $P(X=1,Y=1)$.

2.7

(a) The relative risk of survival for females was 11.4 times that for males.

(b) Denote the survival proportion for female by p_1 and the survival proportion for male by p_2, then $\frac{p_1/(1-p_1)}{p_2/(1-p_2)}=11.4$. Also, $p_1/(1-p_1)=2.9$. Thus by solving the above two equations for p_1 and p_2, we can obtain $p_1=2.9/3.9=0.7436, p_2=0.2028$.

(c) R is the relative risk with value $0.7436/0.2028=3.667$.

2.8

(a) $\frac{0.847/(1-0.847)}{0.906/(1-0.906)}=0.574$.

(b) This is interpretation for relative risk, not odds ratio. The relative risk is $0.847/0.906=0.9384$. Hence, 60% should be changed to 93.84%.

2.11

(a) The difference of proportions of lung cancer for smokers and non-smokers is 0.0013 while the difference of proportions of heart disease is 0.00256. The relative risk of lung cancer for smokers and non-smokers is 14 while the relative risk of heart disease is 1.6199. The odd ratio of lung cancer for smokers and non-smokers is 14.018 while the odd ratio of heart disease is 1.624.

(b) In terms of reduction of deaths with an absence of smoking, the reduction for lung caner is $0.0013N$ and that for heart disease is $0.00256N$, where N is the sample size. We can see for this criterion, heart disease is more strongly related to cigarette smoking.

2.12

(a) The table is as follows:

	yes	no
aspirin	198	19736
placebo	193	19749

(b) The estimate of odd ration is $\hat{\theta}=\frac{198/19736}{193/19749}=1.0266$, which means the relative risk of heart attach for women taking aspirin is 1.0266 times of that for women taking placebo.

(c) The estimated standard error of $log(\hat{\theta}) = \sqrt{1/198 + 1/19736 + 1/193 + 1/19749} = 0.10165$. So the ninety five percent CI for $\log(\theta)$ is $\log(\hat{\theta}) \pm 1.96 * 0.10165$, which is $(-0.17298, 0.22549)$. So the CI for θ is $(0.8412, 1.2529)$ which contains value 1. So we can not conclude that the relative risk of heart attack for women taking aspirin is different from that for women taking placebo.

2.15

(a) By plugging in $p_1 = 0.0171, p_2 = 0.0094, n_1 = 11034, n_2 = 11037$, we can obtain 95 percent CI for $log(p_1/p_2)$, which is $(0.3603, 0.8365)$. So 95 percent CI for p_1/p_2 is $(1.434, 2.308)$.

(b) To construct the Newcombe interval, we need to obtain the score interval for placebo group and aspirin group. The 95 percent score interval for placebo group is

$$\frac{2p_1 + z_0^2/n_1 \pm \sqrt{z_0^4/n_1^2 + 4z_0^2 p_1(1-p_1)/n_1}}{2(z_0^2/n_1 + 1)}$$

After plugging in p_1, n_1, the lower bound $l_1 = 0.01484$ and the upper bound is $u_1 = 0.01969$. Similarly, we can obtain the lower bound for aspirin group, which is $l_2 = 0.0078$ and the upper bound $u_2 = 0.0114$. Then the lower bound for Newcombe interval is

$$0.0171 - 0.0094 - 1.96 * \sqrt{l_1(1-l_1)/n_1 + u_2(1-u_2)/n_2} = 0.0047$$

and the upper bound for Newcombe interval is

$$0.0171 - 0.0094 + 1.96 * \sqrt{u_1(1-u_1)/n_1 + l_2(1-l_2)/n_2} = 0.0108$$

. The Agresti-Caffo interval is (0.0047, 0.0107). We can see they are very close to the result in section 2.2.2. which is $(0.005, 0.011)$.

(c) For the data from the aspirin and heart attack study, $NNT = \frac{1}{p_1 - p_2} = \frac{1}{0.0077} = 129.68$. Thus, 130 is needed to treat with aspirin to prevent one heart attack.

The remaining problems are only for students who have taken STAT 414, 610 or STAT 630.

1.23b

The likelihood function $l(\theta) = [0.25(2+\theta)]^{1997}[0.25(1-\theta)]^{906}[0.25(1-\theta)]^{904}[0.25\theta)]^{32}$. Then the log-likelihood function is $L(\theta) = logl(\theta) = 1997log[0.25(2+\theta)] + 906[0.25(1-\theta)] + 904[0.25(1-\theta)] + 32[0.25\theta)]$. Thus $\frac{d}{d\theta}L(\theta) = \frac{1997}{2+\theta} - \frac{906}{1-\theta} - \frac{904}{1-\theta} + \frac{32}{\theta}$. Let this equal to zero and solve for θ , then we can get the mle of θ : $\hat{\theta} = 0.0357$.

2.2b

solution is provided above.