Math 485 HW 2

$\pi_i = P(Y=1 \mid X=i), \quad i=1,2$

1) 2.2

a) By definition, we have

$$\text{sensitivity} = P(Y=1 \mid X=1) = \pi_1$$

$$\text{specificity} = P(Y=2 \mid X=2) = 1 - P(Y=1 \mid X=2) = 1 - \pi_2$$

Notic that: $P(A^c) = 1 - P(A)$.

b) let $\gamma = P(\text{subject has the disease}) = P(X=1)$

$P(\text{subject truly has the disease} \mid \text{the diagnosis is positive})$

$$= P(X=1 \mid Y=1) = \frac{P(X=1, Y=1)}{P(Y=1)}$$

$$= \frac{P(Y=1 \mid X=1)\, P(X=1)}{P(Y=1 \mid X=1) P(X=1) + P(Y=1 \mid X=2) P(X=2)}$$

$$= \frac{\pi_1 \gamma_1}{\pi_1 \gamma_1 + \pi_2 (1-\gamma_1)}$$

Notice that, from Bayes's theorem, for two event A & B, we have

$$P(A \mid B) = \frac{P(A \cap B)}{P(B)} = \frac{P(B \mid A) P(A)}{P(B \mid A) P(A) + P(B \mid A^c) P(A^c)}$$

1) continue

c) $\gamma = 0.01$, Sensitivity $= \pi_1 = .86$

Specificity $= 1-\pi_2 = .88 \Rightarrow \pi_2 = .12$

P(the woman truly has breast cancer | the test result positive)

$$P(X=1 \mid Y=1) = \frac{\pi_1 \gamma_1}{\pi_1 \gamma_1 + \pi_2(1-\gamma)}$$

$$= \frac{.86 \times (0.01)}{.86(.01) + .12(.99)} = 0.0675$$

d) $P(X=1, Y=1) = P(Y=1 \mid X=1)\,P(X=1)$

$= \pi_1 \gamma = .86(.01) = .086$

$P(X=1, Y=2) = P(Y=2 \mid X=1)\,P(X=1)$

$= (1 - P(Y=1 \mid X=1))\,P(X=1)$

$= (1-.86)(.01) = .0014$

$P(X=2, Y=1) = P(Y=1 \mid X=2)\,P(X=2) = .12(.99) = .1188$

$P(X=2, Y=2) = P(Y=2 \mid X=2)\,P(X=2)$

$= (1 - P(Y=1 \mid X=2))\,P(X=2) = (1-.12)(.99) = .8712$

	X \ Y	(+) 1	(−) 2	
disease	1	.086	.0014	.01
No disease	2	.1188	.8712	.99
				1

Almost all (99%) do not have the disease. In the column of the positive test, a higher proportion are in the "No disease" category.

2.3

2)

a) I $0.0000624 - .0000013 = 0.0000611$

II $\frac{62.4}{1.3} = 48$

Therefore, the estimated probability of a death by gun in US was 48 times higher than in Britain.

b) Relative risk is more useful. Because difference of proportions make It misleading.

2.5

3)

a) Relative risk.

b) I $\pi_1 = .55\,\pi_2 \Rightarrow \frac{\pi_1}{\pi_2} = .55$

II $\frac{\pi_2}{\pi_1} = \frac{1}{.55} = 1.82$.

2.6

4)

a) I $0.001304 - 0.000121 = 0.0012$

II $\frac{0.001304}{0.000121} = 10.78$

based on the The relative risk is more informative. Becase, difference of propotions seems there is no association.

2.6

4) continue

b) $$\frac{\frac{\pi_1}{1-\pi_1}}{\frac{\pi_2}{1-\pi_2}} = \frac{\frac{0.001304}{1-.001304}}{\frac{0.000121}{1-.000121}} = 10.79.$$

when the probability of success is close to zero for each group, then the relative risk and the odds ratio are very close to each other.

2.7

5) a) Notice that probability is always between 0 and 1. This is the interpretation of the relative risk not the probability. Therefore "odds" should be substituted for "probability".

b) recall: $$P(S) = \frac{odds(S)}{1+odds(S)}$$

for female: $$p(S) = \frac{2.9}{1+2.9} = 0.744$$

fo male: $$P(S) = \frac{.254}{1+.254} = 0.203$$

c) $$R = \frac{.744}{.203} = 3.7$$

Notice that we have $$\text{odds ratio} = \frac{\text{odds of Femal}}{\text{odds of Male}} = 11.4$$

So, $$\text{odds of Male} = \frac{\text{odds of Female}}{11.4} = \frac{2.9}{11.4} = .254$$

2.8

6) a) $$\frac{\text{odds of black}}{\text{odds of white}} = \frac{\frac{\pi_1}{1-\pi_1}}{\frac{\pi_2}{1-\pi_2}} = \frac{\frac{.847}{1-.847}}{\frac{.906}{1-.906}} = 0.574$$

b) This is the interpretation for relative risk, not the odds ratio. The relative risk here is $\frac{.847}{.906} = 0.935$.

Therefore, 60% should be replaced with 93.5%.

---

7) 2.12

a)

Treatment	Heart attack Yes	No	Total
Placebo	193	19749	19942
Aspirin	198	19736	19934

b) $$\hat{\theta} = \frac{n_{11} n_{22}}{n_{12} n_{21}} = \frac{193 \times 19736}{198 \times 19749} = 0.974$$

the estimated odds of a heart attack were a bit less for the placebo group.

2.12

7) continue

c). 95% C.I. for $\log\theta$ is (log of the odds ratio):

$$\log(\hat{\theta}) \pm Z_{.025}\, SE(\log\hat{\theta})$$

where $S.E(\log\hat{\theta}) = \sqrt{\frac{1}{n_{11}} + \frac{1}{n_{12}} + \frac{1}{n_{22}} + \frac{1}{n_{22}}}$

$$\log(.974) \pm 1.96 \sqrt{\frac{1}{193} + \frac{1}{198} + \frac{1}{19749} + \frac{1}{19736}}$$

$$-.0263 \pm 1.96\,(0.1017)$$

or $(-0.225, 0.173)$

the the C.I for θ is : $(e^{-.225}, e^{0.173})$

or $(.80, 1.19)$.

since the interval included one, therefore most likely there is no effect. or, if there is any effect, it is relatively small.

8) 2.17

	MI		
Group	Yes	No	Total
Placebo	189	10845	11034
Aspirin	104	10933	11037
	293	21778	22081

$$\hat{\mu}_{ij} = \frac{n_{i+} n_{+j}}{n}$$

Expected table.

	MI	
Group	Yes	No
Placebo	146.41	10882.59
Aspirin	146.45	10885.55

$$X^2 = \sum_{ij} \frac{(n_{ij} - \hat{\mu}_{ij})^2}{\hat{\mu}_{ij}} = \frac{(189 - 146.41)^2}{146.41} + \cdots + \frac{(10845 - 10885.55)^2}{10885.55}$$

$$= 25$$

df = 1, $\Rightarrow$ P_value < 0.0001.

$$G^2 = 2 \sum_{ij} n_{ij} \log\left(\frac{n_{ij}}{\hat{\mu}_{ij}}\right) = 2\left[189 \log\left(\frac{189}{146.41}\right) + \cdots + 10845 \log\left(\frac{10845}{10882.5}\right)\right.$$

$$= 25.4$$

df = 1.

Based on both statistics, there is a strong evidence that heart attacks depends on taking aspirin.

2.18

9) a) $\hat{\mu}_{ij} = \dfrac{n_{i+} n_{+j}}{n}$, $\mu_{11} = \dfrac{290 \times 168}{1362} = 35.8$

b) $df = (I-1)(J-1) = (3-1)(3-1) = 4$

P-value < .0001 , which shows a strong evidence of an association between happiness and income.

c) This is showing that more poeple are in the cell if the variable were independent. That is, the estimated expected value is 2.973 standard errors bigger than the number in this cell.

d) There is a strong evidence that more people in this cell than if the variables were independent.

10) 2.19 ~~H_0: X & Y are indep. v.s.~~ H_A: X & Y are Not indep.

a) $G^2 = 2 \sum_{ij} n_{ij} \log\left(\dfrac{n_{ij}}{\hat{\mu}_{ij}}\right) = \cdots = 187.6$

$X^2 = \sum_{ij} \dfrac{(n_{ij} - \hat{\mu}_{ij})^2}{\hat{\mu}_{ij}} = \cdots = 167.8$

$df = (I-1)(J-1) = (2-1)(3-1) = 2$. $\Rightarrow$ P-value < 0.0001

There is strong evidence of association

2.17

10) continue

Race		Party Identification: Democrat	Ind.	Republican	
White		871	444	873	2188
	Average	(982.21)	(438.77)	(767.01)	
	residual	−3.54	0.24	3.82	
	standardized R.	−11.85	0.692	11.77	
Black		302	80	43	425
	average	(190.78)	(85.22)	(148.98)	
	residual	8.05	−.566	−8.68	
	standardized Resi	11.85	−.692	−11.77	
		1173	524	916	2613

where $\hat{\mu}_{ij} = \dfrac{n_{i+}\,n_{+j}}{n}$ $\xRightarrow[j=1]{i=1}$ $\hat{\mu}_{11} = \dfrac{2188 \times 1173}{2613} = 982.21$

Pearson residuals : $\dfrac{n_{ij} - \hat{\mu}_{ij}}{\sqrt{\hat{\mu}_{ij}}}$ $\xRightarrow[j=1]{i=1}$ $\dfrac{871 - 982.21}{\sqrt{982.21}} = -3.54$

standardized residuals: $\dfrac{n_{ij} - \hat{\mu}_{ij}}{\sqrt{\hat{\mu}_{ij}(1-P_{i+})(1-P_{+j})}}$

$i=1, j=1$ $\Rightarrow$ $\dfrac{871 - 982.21}{\sqrt{982.21\left(1-\frac{2188}{2613}\right)\left(1-\frac{1173}{2613}\right)}} = -11.852$

$P_{1+} = \dfrac{n_{1+}}{n} = \dfrac{2188}{2613}$

$P_{+1} = \dfrac{n_{+1}}{n} = \dfrac{1173}{2613}$

2.19

10) continue

The large positive standardized residuals of -11.85 for white Democrat and -11.77 for black Republicans show strong evidence of fewer people in these cells than we would expect if X & Y were independent. . . .

c) For comparing races between "Democrat" and "Independent", we got $G^2 = 24.1$. (I)

And for comparing races among "Democrat + Independent" and "Republican", we got $G^2 = 163.5$ (II)

Therefore, there is strong evidence that white are more likely than black to be "Republicans".

(I) $$G^2 = 2\sum_{ij} n_{ij} \log\left(\frac{n_{ij}}{\hat{\mu}_{ij}}\right) = 2\left[871 \log\left(\frac{871}{908.954}\right) + 444 \log\left(\frac{444}{406.046}\right) + 302 \log\left(\frac{302}{264.046}\right) + 80 \log\left(\frac{80}{117.954}\right)\right]$$

$$= 24.045 \simeq 24.1.$$

(II) $$G^2 = 2\left[1315 \log\left(\frac{1315}{1420.98}\right) + 382 \log\left(\frac{382}{276.01}\right) + 873 \log\left(\frac{873}{767.01}\right) + 43 \log\left(\frac{43}{148.98}\right)\right] = 163.5$$