TA's notes

- To receive full credits, you must show the work by which you reach the answer(s).
- Interpretation of results should be phrased in the context of question(s) and as detailed as possible. Numbers themselves are not self-explanatory.

Exercise 2.1 [6 points]

The total number of concordant pairs equals

$$\begin{aligned} C &= 7(8+3+7+5+4+9+8+9+14)+7(3+7+4+9+9+14)+2(7+9+14) \\ &\quad +2(5+4+9+8+9+14)+8(4+9+9+14)+3(9+14)+1(8+9+14)+5(9+14)+4(14) \\ &= 7*67+7*46+2*30+2*49+8*36+3*23+1*31+5*23+4*14 \\ &= 1508 \end{aligned}$$

The total number of discordant pairs equals

$$\begin{aligned} D &= 3(2+8+3+1+5+4+2+8+9)+7(1+5+4+2+8+9)+9(2+8+9) \\ &\quad +2(2+8+1+5+2+8)+3(1+5+2+8)+4(2+8)+7(2+1+2)+8(1+2)+5(2) \\ &= 3*42+7*29+9*19+2*26+3*16+4*10+7*5+8*3+5*2 \\ &= 709 \end{aligned}$$

$$\hat{\gamma} = \frac{C-D}{C+D} = \frac{1508-709}{1508+709} = 0.360$$

Of the 2217 concordant and discordant pairs, 68% are concordant and 32% are discordant. The difference of the corresponding proportions gives $\hat{\gamma} = 0.360$. There is a moderately positive association in this sample for married couples to report similar levels of sexual satisfaction, i.e. For husbands who reported a higher level, their wives tended to report a higher level as well.

Exercise 2.12 [8 points]

a. [4 points]

Using the joint distribution, the odds ratio for the two binary response variables is:

π_{11}	π_{12}
π_{21}	π_{22}

$$\theta = \frac{\Omega_1}{\Omega_2} = \frac{\pi_{11}/\pi_{12}}{\pi_{21}/\pi_{22}} = \frac{\pi_{11}\pi_{22}}{\pi_{12}\pi_{21}}$$

Using the within-row conditional distributions,

$$\begin{array}{cc|c} \pi_{i=1|j=1} & \pi_{i=1|j=2} & \pi_{1+} \\ \hline \pi_{i=2|j=1} & \pi_{i=2|j=2} & \pi_{2+} \end{array} \qquad \theta = \frac{\Omega_1}{\Omega_2} = \frac{\pi_{i=1|j=1}/\pi_{i=1|j=2}}{\pi_{i=2|j=1}/\pi_{i=2|j=2}} = \frac{\frac{\pi_{11}/\pi_{1+}}{\pi_{12}/\pi_{1+}}}{\frac{\pi_{21}/\pi_{2+}}{\pi_{22}/\pi_{2+}}} = \frac{\pi_{11}\pi_{22}}{\pi_{12}\pi_{21}}$$

Using the within-column conditional distributions,

$$\begin{array}{c|c} \pi_{j=1|i=1} & \pi_{j=2|i=1} \\ \pi_{j=1|i=2} & \pi_{j=2|i=2} \\ \hline \pi_{+1} & \pi_{+2} \end{array} \qquad \theta = \frac{\Omega_1}{\Omega_2} = \frac{\pi_{j=1|i=1}/\pi_{j=1|i=2}}{\pi_{j=2|i=1}/\pi_{j=2|i=2}} = \frac{\frac{\pi_{11}/\pi_{+1}}{\pi_{21}/\pi_{+1}}}{\frac{\pi_{12}/\pi_{+2}}{\pi_{22}/\pi_{+2}}} = \frac{\pi_{11}\pi_{22}}{\pi_{12}\pi_{21}}$$

b. [4 points]

We have shown that the odds ratio for 2 binary variables are exactly the same, using the joint distribution, the within-row distributions, or the within-columns distributions. This property allows us to estimate the same characteristic (θ) in various sampling designs:

- Sampling difficulties in real life at times force us to select disproportionately large or small samples from marginal categories of a variable. θ is invariant in this situation, while difference of proportions and relative risk are not.

- - in *prospective* studies, n_{i+} are often fixed; we can use within-row conditional distributions to compare how X explains Y.
  - in *retrospective* studies, n_{+j} are often fixed; we can use within-column conditional distributions to compare how Y changes according to X.
  - in *cross-sectional* studies, the total sample size n is fixed; we can randomly sample a number of subjects, and classify them by X and Y.

  As we have shown, the odds ratio is equally valid for these three types of studies. It is thus a useful tool in comparing results from various studies regardless of their sampling designs.

Problem 3 [15 points]

a. [3 points]

Conditional probabilities for wives' heights given husbands' height ($\pi_{i|j}$):

	Ht	Hm	Hs
Wt	18/60=.30	20/99=.20	12/46=.26
Wm	28/60=.47	51/99=.52	25/46=.54
Ws	14/60=.23	28/99=.28	9/46=.20
Total	1	1	1

b. [3 points]

- If the husbands are tall, the wives are most likely to be medium (47%), then tall (30%), and least likely to be short (23%).
- If the husbands are medium, the wives are most likely to be medium (52%), then short (28%), and least likely to be tall (20%).
- If the husbands are short, the wives are most likely to be medium (54%), then tall (26%), and least likely to be short (20%).

Obviously these three within-column distributions are very similar with one another. Men of each height category are most likely to be married to medium women. It is because we have more medium-height women in the sample than either of tall or short. However, by the rank order of the conditional proportions, we can see the two conditional distributions of the tall and the short men are more similar to each other than to the medium men. (You can also do comparisons between the differences of conditional probabilities.) Tall and short men have higher percentages in marrying taller women (tall and medium) than medium men.

c. [3 points]

$\theta_{HtWtHmWm} = \frac{18*51}{20*28} = 1.639$	$\theta_{HmWtHsWm} = \frac{20*25}{12*51} = .817$
$\theta_{HtWmHmWs} = \frac{28*28}{51*14} = 1.098$	$\theta_{HmWmHsWs} = \frac{51*9}{25*28} = .656$

The interpretations of the odds ratios are as follows:

- Tall men have a higher odds (1.639 times) than medium men to marry tall women (rather than medium women).
- Tall men have a higher odds (1.098 times) than medium men to marry medium women (rather than short women).
- Short men have a higher odds ($\frac{1}{.817} = 1.22$ times) than medium men to marry tall women (rather than medium women).
- Short men have a higher odds ($\frac{1}{.656} = 1.52$ times) than medium men to marry medium women (rather than short women).

d. [6 points]

Given this particular culture in which men tend to select wives (i.e. men as the explanatory vs. women as the response variable), we hypothesize the following:

1. taller wives are generally more desirable (by the men);
2. men of medium stature are less able than tall or short men to select desirable mates (taller wives).

Using the odds ratio table in (c), you can well examine this hypothesis.
You can also construct a collapsed table, e.g.

	Wt	Wm	Ws	Total
Ht+Hs	30	53	23	116
Hm	20	51	28	99

By calculating the concordant and discordant pairs, we can estimate gamma:

$$C = 30(51+28) + 53(28) = 3854$$
$$D = 23(20+51) + 53(20) = 2693$$
$$\hat{\gamma} = \frac{C-D}{C+D} = .177$$

There is a tendency, though weak, that tall and short men to marry taller women than medium men. That is, taller women are more desirable by men; and medium men are less able to select those taller women, compared with tall and short men. We should collect more data to see if we can obtain stronger association measures for this hypothesis. However, if you set up null hypothesis (all men are equal) vs. alternative hypothesis (our claim) and used Pearson Chi-squared test or Log-likelihood Chi-squared test, you would not be able to reject the null hypothesis due to lack of significance.

Problem 4 [11 points]

a. [3 points]

If the null hypothesis is true, then the multinomial cell probabilities are:

Type	π_i
green	9/16=.5626
golden	3/16=.1875
green-striped	3/16=.1875
golden green-striped	1/16=.0625

b. [3 points]

Using this multinomial model, we can estimated the expected counts:

Type	π_i	$\hat{m}_i = n * \pi_i$
green	9/16=.5626	1301*.5625=732
golden	3/16=.1875	1301*.1875=244
green-striped	3/16=.1875	1301*.1875=244
golden green-striped	1/16=.0625	1301*.0625= 81

c. [5 points]

$$
\begin{aligned}
\chi^2 &= \Sigma \frac{n_i - \hat{m}_i}{\hat{m}_i} \\
&= \frac{(773-732)^2}{732} + \frac{(231-244)^2}{244} + \frac{(238-244)^2}{244} + \frac{(59-81)^2}{81} \\
&= 2.296 + 0.693 + 0.148 + 5.975 \\
&= 9.112
\end{aligned}
$$

The critical value $\chi^2_{\alpha=.05}$ is 7.81473 (p.506, Appendix C). We are able to reject the null hypothesis which is a simple Mendelian inheritance model. The data show strong evidence that the counts of these 4 plants would not follow a 9:3:3:1 ratio.