Problem 1.2

a) Because the student answers these questions of 4 answers by guessing, the chance that the correct answer is picked for each question is 1/4. In total, there are 100 such questions. Therefore, the number of correct answers follows Bin(n=100, p=.25).

b) The mean is np=100(.25)=25 and the variance is np(1-p)=100(.25)(.75)=18.75.

Yes, it would be a surprise if the student made at least 50 correct answers. Since np, n(1-p) both are larger than 10, we can use normal approximation here. The z score at 50 correct answers is

$$z = \frac{50-25}{[18.75]^{1/2}} = 5.8.$$

The probability of making 50 or more correct answers is approximately Pr[Z>5.8]=0.00. So it is almost impossible to answer 50 or more questions correctly by guessing.
c) Assuming the answers of these 100 questions are chosen independently, then (n1,n2,n3,n4) follows the multinomial distribution with n=100 and p vector= (.25,.25,.25,.25).

Problem 1.5

Given H_o: $\pi = 0.5$, H_a: $(\pi) \neq 0.5$

Score test: $z_S = \frac{u(\pi_o)}{[\iota(\pi_o)]^{1/2}} = \frac{\hat{\pi}-\pi_o}{\sqrt{\pi_o(1-\pi_o)/n}}$ =(842/1824-0.5)/[0.5(1-0.5)/1824]$^{0.5}$ = **-3.25**, which has a p-value of 0.001, which is highly significant and we reject H_o: $\pi = 0.5$.

Score Confidence Interval:

$$\hat{\pi}\left(\frac{n}{n+z_{\alpha/2}^2}\right)+\frac{1}{2}\left(\frac{z_{\alpha/2}^2}{n+z_{\alpha/2}^2}\right)\pm z_{\alpha/2}\sqrt{\frac{1}{n+z_{\alpha/2}^2}\left[\hat{\pi}(1-\hat{\pi})\right]\left(\frac{n}{n+z_{\alpha/2}^2}\right)+\left(\frac{1}{2}\right)\left(\frac{1}{2}\right)\left(\frac{z_{\alpha/2}^2}{n+z_{\alpha/2}^2}\right)}$$

$$.46\left(\frac{1824}{1824+1.96^2}\right)+\frac{1}{2}\left(\frac{1.96^2}{1824+1.96^2}\right)\pm 1.96\sqrt{\frac{1}{1824+1.96^2}\left[.46(1-.46)\right]\left(\frac{1824}{1824+1.96^2}\right)+\left(\frac{1}{2}\right)\left(\frac{1}{2}\right)\left(\frac{1.96^2}{1824+1.9}\right.}$$

Thus the 95% confidence interval is: (.439, .485)

These results show that we are 95% confident that the population proportion who would reply "yes" is between .439 and .485, shorter than .50.

Problem 2.3 Using the Florida motor vehicle accident injury data, identify: the response variable, and find and interpret the difference of proportions, relative risk and odds ratio. Explain why the relative risk and odds ratio are approximately equal.

This data was from a period prior to the widespread use of air bags in motor vehicles.

Table 2.9 - Motor Vehicle Accidents in FL

	Injury		
Seat Belt Use	Fatal	Nonfatal	Totals
None	1601	162527	164128
Yes	510	412368	412878
Totals	2111	574895	577006

The data is based on records of (presumably all) accidents in Florida in 1988. If this is the case, then we can estimate the probability of a fatal injury from the data (2111/577006). This is a retrospective study with unequal numbers of the two types of injury as well as unequal numbers of seat belt use. The total sample size is a random variable and the numbers of observations for the four combinations of seat-belt use and injury type can be treated as independent Poisson random variables with unknown means.

The type of injury is the response variable and the use of seat belts is the explanatory variable in this study.

P(I=Fatal|SB=No) = 1601/164128 = 0.00975
P(I=Fatal|SB=Yes) = 510/412878 = 0.00124

The difference of proportions P(I=Fatal|SB=No) – P(I=Fatal|SB=Yes) = 0.00852
The relative risk = 0.00975/0.00124 = 7.90
The odds ratio = (1601*412368)/(510*162527) = 7.96

All three of these measures show that there is evidence of a strong relationship between seat belt use and type of injury due to a motor vehicle accident.

The relative risk shows that the proportion of fatalities when the seat belt was not used was 7.9 times that of those who wore seat belts (remember this was before air bags!). The value is much greater than 1 so we know Injury and Seat Belt Use are not independent.

The odds of a fatality during an accident for those taking not wearing seat belts was 7.96 times the odds for those wearing seat belts.

The magnitude of Θ & Relative Risk are similar whenever the π_i of the outcome of interest is close to zero for both groups (see section 2.2.7 of the text). In this data, the P(I=Fatal)= 2111/577066 = 0.00366 is quite small, as are the probabilities (shown above) for fatal injuries for both of the seat belt cases of interest.

Problem 2.4 a) A Dec 3, 1998 British study reported that of smokers who get lung cancer, "women were 1.7 times more vulnerable than men to get small-cell lung cancer."

The definition of "vulnerable" is **"**capable of being physically or emotionally wounded" (Merriam-Webster Online Dictionary. 2009. http://www.merriam-webster.com/dictionary/vulnerable)

The description 1.7 times more vulnerable is referring to the Relative Risk. While, both the odds ratio and the relative risk compare the likelihood of an event between two groups, the term vulnerable implies a ratio of probabilities. "The relative risk measures events in a way that is **interpretable and consistent with the way people really think**." (Simon. Odds ratio versus relative risk. January 9, 2001).

A more clear description of the study was written by Cooper.

"The study by the British Thoracic Society (BTS), the largest British investigation into lung cancer, found nearly twice as many women as men under 65 are diagnosed with small-cell lung cancer, the most dangerous form of the disease." (Glenda Cooper, *Thursday, 3 December 1998, http://www.independent.co.uk/news/worst-cancer-hits-female-smokers-1188820.html).*

Problem 2.4 b) An April 7, 1998 National Cancer Institute study reported that "women were 45% less likely … than were women taking the placebo."

i) The word "likely" refers to probability, so this sentence is talking about RR. "45% less likely" can therefore be explained as the RR of drug group to placebo group was 1-0.45=.55.

ii) RR of drug group to placebo group = (0.55)^-1 = 1.82. The women taking the placebo were 1.82 times more likely to experience cancer compared to women taking the drug.

Problem 2.11 A 20-year cohort study of British male physicians noted that the proportion per year that died from lung cancer was 0.00140 for cigarette smokers and 0.00010 for nonsmokers. The proportion that died from coronary heart disease was 0.00669 for smokers and 0.00413 for nonsmokers.

2.11a) Describe the association of smoking with each of lung cancer and heart disease, using the difference of proportions, relative risk, and odds ratio.

Lung Cancer

P(Died=LC|S=Yes) = 0.00140
P(Died=LC|S=No) = 0.00010

The difference of proportions P(Died=LC|S=Yes) – P(Died=LC|S=No) = 0.00130

The relative risk P(Died=LC|S=Yes) / P(Died=LC|S=No) = 0.00140/0.00010 = 14

The odds ratio = OR=RR*[1- P(Died=LC|S=No)]/([1- P(Died=LC|S=Yes)]
The odds ratio = 14 * (1-0.000010)/(1-0.00140) = 14.02

The difference results show that the proportion of British male physicians who smoked died from lung cancer was .13% more than that of those who did not smoke.

The relative risk results show that the proportion of British male physicians who smoked, died from lung cancer was 14 times that of those who did not smoke.

The odds ratio shows that the odds of death due to lung cancer for who smoked was 14.02 times the odds for those who did not smoke.

Heart Disease

P(Died=HD|S=Yes) = 0.00669
P(Died=HD|S=No) = 0.00413

The difference of proportions P(Died=HD|S=Yes) – P(Died=HD|S=No) = 0.00256

The relative risk P(Died=HD|S=Yes) / P(Died=HD|S=No) = 0.00669/0.00413 = 1.62

The odds ratio = OR=RR*[1- P(Died=HD|S=No)]/([1- P(Died=HD|S=Yes)]
The odds ratio = 1.62 * (1-0.00413)/(1-0.00669) = 1.62

The difference results show that the proportion of British male physicians who smoked died from heart disease is .26% more than that of those who did not smoke.

The relative risk results show that the proportion of British male physicians who smoked, died from heart disease was 1.62 times that of those who did not smoke.

The odds ratio shows that the odds of death due to heart disease for who smoked was 1.62 times the odds for those who did not smoke.

2.11b) Describe which response is more strongly related to cigarette smoking, in terms of the reduction in number of deaths that would occur with elimination of cigarettes.

The magnitudes of the relative risk and the odds ratio were both much higher for the response of lung cancer to smoking than they were between heart disease and smoking. You would conclude that elimination of cigarettes would have a much greater impact on the reduction of death due to lung cancer than it would on the reduction of death due to heart disease.