Answer Key to First Midterm Examination, Math 225, Spring, 2001

Problem 1: TRUE or FALSE:

For strongly right skewed data, the mean will be larger than the median.

Solution: True.

Problem 2: TRUE or FALSE:

If the mean of one group of numbers is twice as high as a second group, the standard deviation of the first group must also be twice as large as the standard deviation of the second group.

Solution: False. For example, if the first set of numbers is $\{3, 5\}$ and the second is $\{1, 3\}$, then the mean of the first set is twice as high as the mean of the second set, but the standard deviations are the same.

Problem 3: TRUE or FALSE:

The standard deviation is the distance between the upper and lower quartiles.

Solution: False. The interquartile range is that distance.

Problem 4: TRUE or FALSE:

For any two events A and B, $P(A \text{ or } B) = P(A) + P(B)$.

Solution: False. This is only true for mutually exclusive events. The general statement is for any two events A and B, $P(A \text{ or } B) = P(A) + P(B) - P(A \text{ and } B)$.

Problem 5: TRUE or FALSE:

The number of ways to choose r items from k items without regarding order is $k!/r!$.

Solution: False. ${}_kC_r = k!/(r!(k-r))!$.

Problem 6: TRUE or FALSE:

If events A and B are independent, then $P(A|B) = P(B)$.

Solution: False. If A and B are independent, $P(A|B) = P(A)$.

Problem 7: TRUE or FALSE:

The specificity and sensitivity of a test must sum to one.

Solution: False. The sensitivity plus the false negative rate sum to one as do the specificity and the false positive rate.

Problem 8: TRUE or FALSE:

The Poisson distribution is continuous.

Solution: False. It is discrete.

Problem 9: TRUE or FALSE:

The 95th percentile of any normal distribution is 1.96 standard deviations above the mean.

Solution: False. That is the 0.975 quantile. The 95th percentile is 1.645 standard devaitions above the mean for a normal distribution.

Problem 10: TRUE or FALSE:

The standard deviation of the sampling distribution of the sample mean cannot be larger than the population standard deviation.

Solution: True.

Problem 11: (10 points)
A box contains fifteen balls, five of which are white and ten of which are black. Six balls are selected at random without replacement.

(a) What is the probability that the first three selected balls are white and the last three are black?

(b) What is the probability that exactly three of the six selected balls are white?

Solution:
(a) $({}_5P_3 \times {}_{10}P_3)/{}_{15}P_6 = \frac{5\times4\times3\times10\times9\times8}{15\times14\times13\times12\times11\times10} = \frac{12}{1001} = 0.0120.$

(b) $({}_5C_3 \times {}_{10}C_3)/{}_{15}C_6 = \frac{10\times120}{5005} = \frac{240}{1001} = 0.2398.$

Problem 12: (16 points)
Pap smears are often used as screening tests for cervical cancer before symptoms appear. The following table displays Pap smear test results and cancer status from a population of 1,000,000 women.

	Cancer		
Pap smear	present	absent	
test positive	70	186,384	186,454
test negative	13	813,533	813,546
Total	83	999,917	1,000,000

Find:

(a) the sensitivity of the Pap smear test

(b) the specificity of the Pap smear test

(c) the prevalence of cervical cancer in the population

(d) the probability of cervical cancer given a positive Pap smear in the population.

Solution:
(a) $P(+|\text{cancer}) = 70/83 = 0.843.$
(b) $P(-|\text{no cancer}) = 813,533/999,917 = 0.814.$
(c) $P(\text{cancer}) = 83/100,000 = 0.000083.$
(d) $P(\text{cancer}|+) = 70/186,454 = 0.000375.$

Problem 13: (12 points)
For each case, if the random variable has a binomial distribution, specify n and p. If the random variable is not binomial, explain why.

(a) The ability to process oxygen while riding a cycle is measured twice on each of twenty-five individuals, once with a nasal strip and once without. Assume that the nasal strip has no effect and each individual is equally likely to test better with or without the nasal strip. The random variable X counts the number who test better with the nasal strip.

(b) In a certain population, twenty percent of the individuals are hypertensive. An investigator adds subjects to a study one-by-one until thirty hypertensive people are sampled. The random variable X counts the number of non-hypertensive subjects in the study.

(c) Of 48 genetically modified monkey embryos, 16 contain a jellyfish gene. Twelve embryos are sampled at random. The random variable X counts the number of sampled embryos with the jellyfish gene.

(d) In mitochondrial DNA, the relative frequency of the base guanine (G) depends on codon position. In the first codon position, 22 percent are G, in the second codon position, 14 percent are G, and in the third codon position, 4 percent are G. A sequence of mitochondrial DNA contains 100 bases in each of the three codon positions. The random variable X counts the total number of G bases.

Solution:
(a) binomial, $n = 25$, $p = 0.5$.
(b) not binomial, n is not fixed.
(c) not binomial, the trials are not independent.
(d) not binomial, p is not the same for all trials.

Problem 14: (15 points)
The weight distribution of adult males in the United States is approximately normal with a mean of 172.2 pounds and a standard deviation of 29.8 pounds.

(a) What is the probability that a randomly selected adult American male weighs more than 200 pounds?

(b) What is the 90th percentile of this distribution?

(c) What is the probability that the sample mean weight of six randomly selected adult American males is larger than 200 pounds?

Solution:
(a) $z = (200 - 172.2)/29.8 = 0.93$. The probability is $1 - 0.8238 = 0.1762$.
(b) $x = 172.2 + 1.28(29.8) = 210.3$.
(c) $z = (200 - 172.2)/(29.8/\sqrt{6}) = 2.29$. The probability is $1 - 0.9887 = 0.0113$.

Problem 15: (12 points)
The most common form of color blindness in human beings is caused by a recessive gene on the X chromosome. Consider parents without this type of color blindness. If the mother is homozygous, it is certain that male offspring will not have this type of color blindness. If the mother is heterozygous, male off spring are equally likely to be color blind or not. Suppose that the mother is equally likely to be heterozygous or homozygous. Given that an only son is not color blind, what is the probability that the mother is homozygous?

Solution:

$$
\begin{aligned}
P(\text{homozygous}|\text{not colorblind}) &= \frac{P(\text{homozygous and not colorblind})}{P(\text{not colorblind})} \\
P(\text{homozygous and not colorblind}) &= P(\text{homozygous})P(\text{not colorblind}|\text{homozygous}) \\
&= 0.5 \times 1 \\
P(\text{not colorblind}) &= P(\text{homozygous})P(\text{not colorblind}|\text{homozygous}) \\
&+ P(\text{not homozygous})P(\text{not colorblind}|\text{not homozygous}) \\
&= 0.5 \times 1 + 0.5 \times 0.5 \\
P(\text{homozygous}|\text{not colorblind}) &= \frac{0.5 \times 1}{0.5 \times 1 + 0.5 \times 0.5} \\
&= \frac{2}{3}
\end{aligned}
$$

Problem 16: (15 points)
In a study in France on the effectiveness of the drug RU 486 to terminate pregnancies, 473 of 488 women successfully terminated their pregnancy after a treatment that included taking the drug.

(a) Give a justification for why statistical inference based on the normal distribution is appropriate.

(b) Calculate a 95% confidence interval for the population proportion of pregnancies that would be terminated by this treatment.

(c) If the makers of RU 486 wished to advertise that the treatment was successful at least 90% of the time, is this claim justified? Test the hypothesis that the population proportion is 0.90 versus the alternative that it is higher. State hypotheses, calculate a z test statistic, and report a p-value. Summarize your findings in the context of the problem.

Solution:
(a) There are 473 individuals in one group and $488 - 473 = 15$ in the other. These are both at least five, so a normal approximation will be okay.

(b) $0.969 \pm 1.96\sqrt{\frac{(0.969)(0.031)}{488}}$ or 0.969 ± 0.015.

(c) $H_0 : p = 0.90$, $H_a : p > 0.90$. $z = \frac{0.969 - 0.90}{\sqrt{\frac{(0.9)(0.1)}{488}}} = 5.08$.

The p-value is smaller than 0.0001. There is overwhelming evidence that the success rate of this pill is greater than 90% in the population of women from which the sample was taken.