STAT 205
Spring 2007
Exam 1

Name:__ANSWER KEY______________

$$s^2 = \frac{\sum_{i=1}^{n} (y_i - \bar{y})^2}{(n-1)}$$

$$P\{E_1 \cup E_2\} = P\{E_1\} + P(E_2\} - P\{E_1 \cap E_2\}$$

$$P\{E_1 \cap E_2\} = P\{E_1\}P\{E_2 | E_1\}$$

$$\mu_Y = \sum y_i \, P\{Y = y_i\}$$

$$\sigma_Y^2 = \sum (y_i - \mu_Y)^2 \, P\{Y = y_i\}$$
$$= E(Y^2) - (E(Y))^2$$

$$P\{Y = j\} = {}_nC_j \, p^j \, (1 - p)^{n-j}$$
$$\mu_Y = np$$
$$\sigma_Y^2 = np(1 - p)$$

$$Z = \frac{(Y - \mu)}{\sigma}$$

Part I: Answer six of the following seven questions. If you complete more than six, I will grade only the first six. Five points each.

Use this fruit fly experiment to answer any of the first three questions.

In a behavioral study of the fruitfly *Drosphila melangoster*, a biologist measured the total
time spent preening during a six-minute observation period for each of 20 flies. The following are the preening times (in seconds) for the 20 flies:

10, 16, 18, 19, 22, 24, 24, 25, 26, 29, 31, 32, 33, 34, 46, 48, 48, 52, 57, 76

1) Construct an ordered stemplot for these data.

Preening Times for 20 Drosphilia Melangoster

1 | 0 6 8 9
2 | 2 4 4 5 6 9
3 | 1 2 3 4
4 | 6 8 8
5 | 2 7
6 |
7 | 6

KEY: 3|1 means 31 seconds

2) **Fill in the blank.** Report the five number summary for the fruit fly data.

{10, 23, 30, 47, 76}

Using TI-83/84 Enter data in List 1, then STAT -> scroll over to CALC -> Enter to choose 1-Var Stats -> enter L_1 after the 1-Var Stats prompt. Scroll down to see the five number summary.

3) For the fruit fly experiment, name the

Variable	preening time
Type of variable	continuous
Observational unit	fruit fly
Sample size	20

4) **Circle the correct answer.** A random sample of USC students was selected for a study. Each person in the study was labeled as left handed, right handed, or ambidextrous. Which of the following could be a legitimate assignment of probabilities for this sample space?

0.2, 0.8, 0.1
0.3, 0.8, -0.1
0.2, 0.7, 0.1
0.1, 0.7, 0.1

5) P{A} = 0.8 P{A∩B} = 0.2 **What is the probability of B given A?**

$$\frac{0.2}{0.8}=0.25$$

6) **Circle the correct answer.** TRUE / FALSE
Let Y be a continuous random variable. Then, $P\{Y < a\} = P\{Y \leq a\}$.

7) **Find $Z_{0.05}$**
The value of Z (standard normal random variable, mean 0, standard deviation 1) that has 0.05 above it is the same value that has 0.95 below it.
invNorm(0.95, 0, 1) = 1.645

8) Let Y denote the number of female offspring the Asian Stochastic Beetle can have in her lifetime. A researcher reports the following probability distribution for Y:

Y	0	1	2
P{Y}	.2	.3	.5

Find μ_Y

$$\mu = E[Y] = \sum yP(y) = 0(0.2) + 1(0.3) + 2(0.5) = 1.3 \text{ offspring}$$

Find σ_Y

$$\sigma^2 = \sum (y - \mu)^2 P(y) = (0-1.3)^2(0.2) + (1-1.3)^2(0.3) + (2-1.3)^2(0.5) = 0.61$$

$$\Rightarrow \sigma = \sqrt{\sigma^2} = \sqrt{0.61} = 0.781 \text{ offspring}$$

Using the TI-83/84 Enter y's in List 1, P(y)'s in List 2, STAT -> scroll to CALC -> ENTER (for 1-Var Stats) -> L_1, L_2 after the 1-Var Stats prompt appears
μ_x=1.3 offspring ($\bar{x}$ is the same calculation)
σ_x = 0.61 offspring (S_x is a different calculation and would be marked incorrect)

9) **Fill in the blanks.** The Central Limit Theorem states that for any i.i.d. random sample, $Y_1, Y_2, \ldots, Y_n$ with $E[Y_i] = \mu$ and $Var(Y_i) = \sigma^2$, then as $n \to \infty$, the sample mean has approximately a normal distribution, with mean μ and variance σ^2/n.

Part II: Answer every part of the next two problems. Read each question carefully, and show your work for full credit.

1) (12 pts.) The West Nile Virus was first detected in the Western Hemisphere in 1999 and has since rapidly spread and become a serious health concern. 1%* of people who have been infected with the West Nile Virus will develop a serious illness from the virus. Suppose 50 infected adults are chosen at random.

Find the probability that none of the 50 selected individuals develop a serious illness from
the virus.
Using the TI-83/84
P(Y = 0) = binompdf(50, 0.01, 0) = 0.605

*The CDC actually reports this percentage is actually lower than 1%, but does not say how much lower - we're using 1% for example's sake.

2) The heights of a certain population of corn plants follow a normal distribution with mean $\mu = 145$ centimeters and standard deviation $\sigma = 22$ centimeters.

a) (12 pts.) What percentage of corn plants from this population have heights between 130 and 155 centimeters?

normalcdf(130, 155, 145, 22) = 0.4276 ⇒ 42.76%

a) (12 pts.) What is the 99^{th} percentile for this distribution?

invNorm(0.99, 145, 22) = 196.18 cm

a) (12 pts.) If a random sample of 16 corn plants is chosen from this population, find the probability that their average height will be between 130 and 155 centimeters.?

normalcdf(130, 155, 145, $22/\sqrt{16}$) = 0.9623

3) (12 pts.) A researcher investigated the effect of green light, in comparison to red light, on the growth rate of bean plants. The following is a QQplot for the heights (in) of plants from the soil to the first branching stem, two weeks after germination for the group grown under the green light condition.

The researcher would like to analyze this data using a statistical test that requires the assumption that the data come from a normal population. Using the QQplot, comment on whether it appears this assumption is violated and what characteristics of the plot led you to this conclusion.

There is a systematic departure from the line – the points are making a shape. The points are making an upside down "U" shape, indicating these data come from a skewed left distribution. This is a probable violation of the assumption these data come from a normal population.