STAT 205
Fall 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 eight of the following nine questions. If you complete more than eight, I will grade only the first eight. Five points each.

1) **Fill in the blank.** A researcher counts the number of dendritic branches emanating from a nerve cell taken from the brains of 36 newborn guinea pigs.

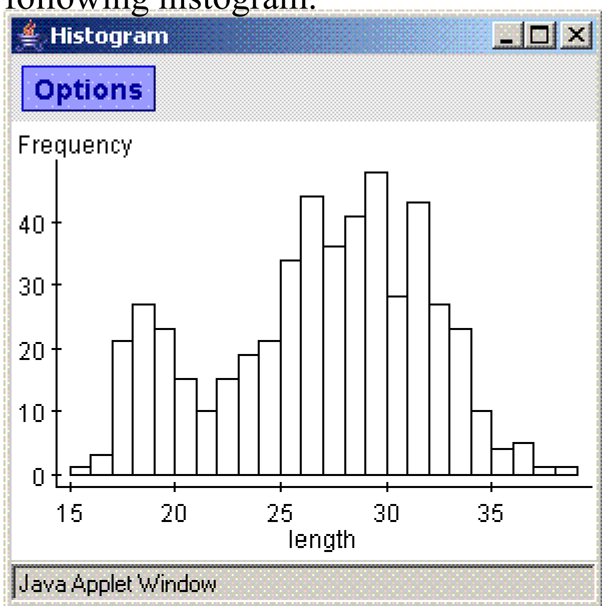
Variable __number of dendritic braches________________________

Type of Variable__quantitative - discrete______________________

Observational Unit_nerve cell from brain of newborn guinea pig_

Sample Size__36____

2) Trypanosomes are parasites which cause disease in humans and animals. In an early study of trypanosome morphology, researchers measured the lengths (in µm) of 500 individual trypanosomes taken from the blood of a rat. These data were used to create the following histogram.



Using the histogram, describe the shape of the distribution of these trypanosome lengths.

Asymmetric and bimodal

3) Recall the empirical rule which allows us to estimate the percent of observations falling within 1 SD (68%), 2 SDs (95%), and 3 SDs (99.7%) of the mean. Could the empirical rule be used for the trypanosome data from question (2)? Why or why not?

The empirical rule requires that the distribution be unimodal and not too skewed. The histogram of trypanosome length shows these requirements are not met.

4) Dopamine is a chemical that plays a role in the transmission of signals in the brain. A pharmacologist measured the amount of dopamine in the brain of each of seven rats. The dopamine levels (nmol/g) were as follows:

```
6.8   5.3   6.0   5.9   6.8   7.4   6.2
```

Calculate the mean of this data.

$$\frac{6.8+5.3+6.0+5.9+6.8+7.4+6.2}{7} = 6.343\ nmol/g$$

Calculate the sample standard deviation.

$S^2 =$

$$\frac{(6.8-6.343)^2+(5.3-6.343)^2+(6.0-6.343)^2+(5.9-6.343)^2+(6.8-6.343)^2+(7.4-6.343)^2+6.2-6.343)^2}{(7-1)}$$

$$= \frac{0.2209+1.087849+0.117649+0.196249+0.208849+1.117249+0.020449}{6}$$

$$= \frac{2.969194}{6} = 0.494866$$

$$\Rightarrow s = \sqrt{0.494866} = 0.703\ nmol/g$$

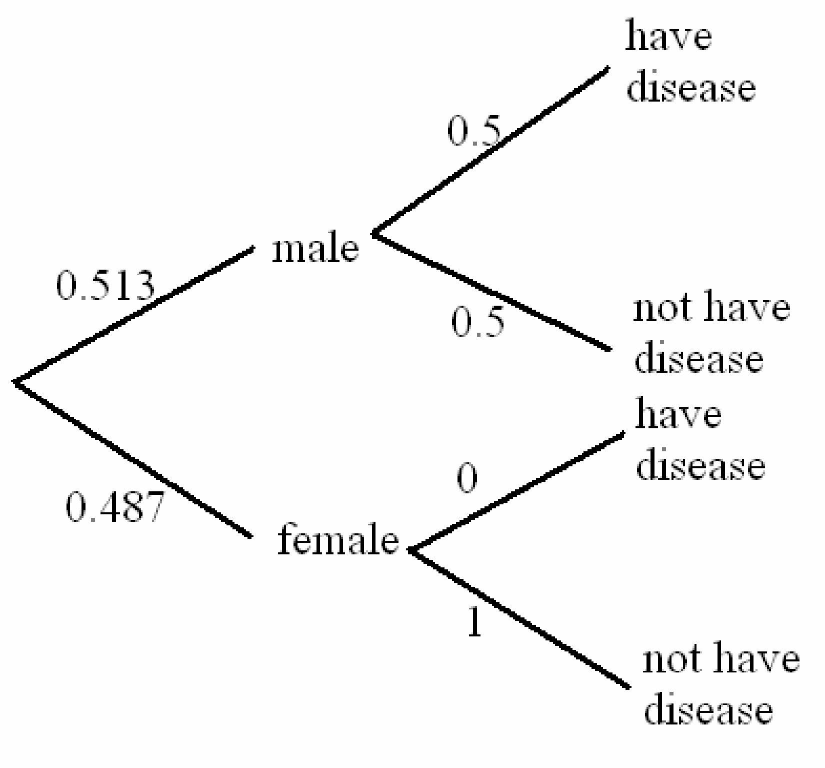
5) **Fill in the blanks.** Refer to the dopamine data of question (4). Report the five number summary.

{_<u>5.3</u>__, __<u>5.9</u>_, __<u>6.2</u>_, __<u>6.8</u>_, _<u>7.4</u>__}

6) P{A} = 0.4 P{A∩B} = 0.1 **What is the probability of B given A? That is, find P{B|A}.**

$$P\{B \mid A\} = \frac{P\{A \cap B\}}{P\{A\}} = \frac{0.1}{0.4} = 0.25$$

7) Suppose a disease is inherited via a sex-linked mode of inheritance, so that a male offspring has a 50% chance of inheriting the disease, but a female offspring has no chance of inheriting the disease. Further suppose that 51.3% of births are male. What is the probability that a randomly chosen child will have the disease? *Hint: Drawing a tree diagram might be useful here.*



Pr{having the disease} = (0.513)(0.5) + (0.487)(0) = 0.2565 or 25.65%

8) Find $Z_{0.01}$

2.33

9) **(Circle the correct answer)** Suppose the weights of USC females follow a normal distribution. A randomly selected female is told that her Z-score is -2. What information does this give her about her weight in this distribution?

Her weight is 2 pounds above the mean.

Her weight is 2 pounds below the mean.

Her weight is 2 standard deviations above the mean.

Her weight is 2 standard deviations below the mean.

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

1) (24 pts.) According to the prostate cancer foundation, prostate cancer will strike 1 in 6 American men. Then the probability that an American man will develop prostate cancer is 1/6 = 0.167.

a) If a random sample of 30 men were taken, find the probability that at most one of the men will develop prostate cancer.

By hand
P{at most one of the men}
$= P\{Y \le 1\}$
$= P\{Y = 0\} + P\{Y = 1\}$
$= {}_{30}C_0\, (0.167)^0(0.833)^{30} + {}_{30}C_1\, (0.167)^1(0.833)^{29}$
$= (0.833)^{30} + (30)(0.167)^1(0.833)^{29}$
$= 0.00416246 + 0.025034722$
$= 0.0292$

Using TI-83
Binomcdf(30, 0.167, 1) = 0.0292

b) Find the probability that exactly five men from the sample will develop prostate cancer.

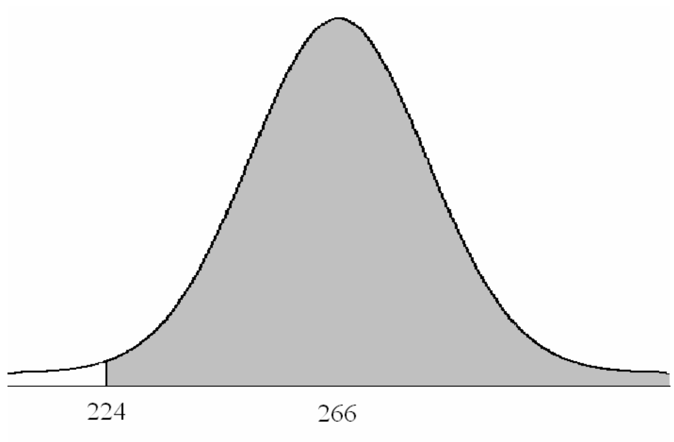
By hand
P{exactly 5}
$= P\{Y = 5\}$
$= {}_{30}C_5\, (0.167)^5(0.833)^{25}$
$= 142506\, (0.167)^5(0.833)^{25}$
$= 142506\, (0.000129892)(0.010378272)$
$= 0.192$

Using TI-83
Binompdf(30, 0.167, 5) = 0.192

2) (24 pts.) The natural length of human pregnancies (that yield a live birth) varies according to a normal distribution with mean, $\mu = 266$ days, and standard deviation, $\sigma = 16$ days.

a) A baby is at increased risk of long term health problems if it is born on or before 224 days (32 weeks) of pregnancy. Find the probability that a randomly selected pregnant female gives birth after 224 days.

Using Table 3
$P\{Y > 224\}$
$= 1 - P\{Y \leq 224\}$
$= 1 - P\{Z \leq (224-266)/16\}$
$= 1 - P\{Z \leq -2.63\}$
$= 1 - 0.0043$
$= 0.9957$



Using TI-83
normcdf(224, E99, 266,16) = 0.9948

b) What is the 99th percentile for this distribution?

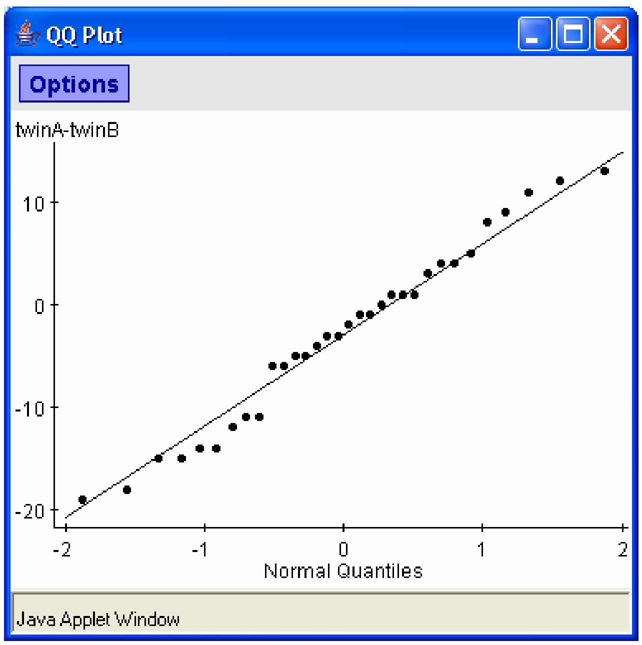
Using Table 3

$$Z_{0.01} = \frac{y^* - 266}{16} \Rightarrow 2.33 = \frac{y^* - 266}{16} \Rightarrow y^* = (2.33)(16) + 266 \Rightarrow 303.28 \text{ days}$$

Using TI-83
invNorm(0.99, 266, 16) = 303.222 days

3) (12 pts.) A classical method of researching the nature vs. nurture question is using identical twins to conduct research. A book by Susan Farber (*Identical Twins Reared Apart*, New York: Basic Books, 1981) contains a chronicle and reanalysis of identical twins reared apart. One question of interest is whether there are significant differences in the IQ scores of identical twins where one member of the pair is reared by natural parents and the other is not.

A random sample of 32 sets of twins under the above circumstances was taken. The researcher is hoping to use a statistical test which requires the assumption of the data coming from a normal population. Using the QQplot below, discuss whether this assumption appears to be met or violated and why.



When answering this question, keep in mind that I care more about how you reach your decision, than the actual decision itself. Two to four sentences should suffice…

Overall, this plot looks fairly linear. I do see that the points go above the line at the high end of the distribution (i.e. the observations are higher than what we would expect if they came from a normal distribution) and go below the line at the low end of the distribution (similarly, are lower than we'd expect from a normal). But the deviation is slight and the points end up going back to the line. Hence, I see little evidence that these points do not come from a normal distribution.