Stat./For./Hort. 571
Midterm I, Fall 2005
Brief Solutions

1. (a) The sample size is 13, and $13/2 = 6.5 \uparrow 7$ while $13/4 = 3.25 \uparrow 4$. The median is the 7th sorted observation, 11, while the first and third quartiles are the 4th sorted observations from the bottom and top respectively, 9 and 13.
   (b) The right boxplot matches the five number summary of the Vermillion River data.
   (c) The left boxplot has a substantially larger range of data (if one knows how to interpret the R boxplot which includes an outlier at 18) even though the interquartile ranges in the two distributions are very similar. The Black River data is more spread out and consequently has a larger standard deviation. Sample size is irrelevant for this issue.

2. (a) The 80th percentile is 0.84 standard deviations above the mean, or $52 + 0.84(4.5) \doteq 55.8$ mm.
   (b) The z-score is $(50-52)/4.5 \doteq -0.44$ and the probability is 0.67.
   (c) The z-score is $(50-52)/(4.5/\sqrt{5}) \doteq -0.99$ and the probability is 0.8389.
   (d) The number of fish longer than 50 mm is the random variable $Y \sim B(5, 0.67)$. $\mathbb{P}(Y=5) = (0.67)^5 \doteq 0.135$.

3. (a) $Y = \#$ with Rh positive blood $\sim B(15, 0.85)$. $\mathbb{P}(Y \geq 14) = 15(0.85)^{14}(0.15) + (0.85)^{15} \doteq 0.3186$.
   (b) $Y = \#$ with Rh positive blood $\sim B(150, 0.85)$. $\hat{p} = Y/150$, is approximately normal, since $150(0.85) \doteq 127.5 > 150(0.15) \doteq 22.5 > 5$. We have $\hat{p} \sim N\left(0.85, \frac{(0.85)(0.15)}{150}\right)$.

$$\begin{aligned} \mathbb{P}(\hat{p} \geq 0.9) &\approx \mathbb{P}\left(Z \geq \frac{0.9 - 0.85}{\sqrt{(0.85)(0.15)/150}}\right) \\ &\doteq \mathbb{P}(Z \geq 1.71) \\ &\doteq 0.0436 \end{aligned}$$

   (c) As the sample size n increases, $\text{Var}(\hat{p}) = p(1-p)/n$ decreases. The sampling distribution of $\hat{p}$ will be more tightly clustered around $\mu = 0.85$ when $n = 1500$ than for $n = 150$, which implies that the tail probability $a = \mathbb{P}(\hat{p} \geq 0.9)$ will be smaller when $n = 1500$.

4. Peter and Flopsy are each either in or out of the garden, a total of four possible elementary outcomes. This table is equivalent to a Venn diagram with the given information in black and the derived probabilities in blue.

	Peter In	Peter Out	total
Flopsy In	0.3	0.0	0.3
Flopsy Out	0.5	0.2	0.7
total	0.8	0.2	1

For example, $\mathbb{P}(\text{Peter out}) = 1 - \mathbb{P}(\text{Peter in}) = 0.2$, while $\mathbb{P}(\text{Peter out}) = \mathbb{P}(\text{Peter out AND Flopsy in}) + \mathbb{P}(\text{both out})$ implies $\mathbb{P}(\text{Peter out AND Flopsy in}) = 0$.

The questions are now all very straightforward.

   (a) $\mathbb{P}(\text{Peter in AND Flopsy in}) = 0.3$.
   (b) $\mathbb{P}(\text{Peter in} \mid \text{Flopsy in}) = 0.3/0.3 = 1$.
   (c) The distribution of Y is

y	0	1	2
$\mathbb{P}(Y=y)$	0.2	0.5	0.3

   and the mean is found as follows.

$$\mathbb{E}(Y) = 0(0.2) + 1(0.5) + 2(0.3) = 1.1$$

An alternative solution to (c) is to let Y_1 and Y_2 be *indicator random variables* that take values 1 or 0 to indicate if Flopsy or Peter are respectively in or out of the garden. Recalling that expected values of sums are sums of expected values (always),

$$\mathbb{E}(Y) = \mathbb{E}(Y_1) + \mathbb{E}(Y_2) = 0.3 + 0.8 = 1.1$$

5. The binomial distribution assumes a fixed sample size, dichotomous outcomes for each trial, equal success probabilities for each trial, and independence among the trials. Recall that the $np > 5$, $n(1-p) > 5$ rule is helpful to determine when a normal approximation to the binomial is accurate, but is not appropriate for determining when a random variable *is binomial*.
   (a) X_1 is not binomial because p is not the same for all 100 trials.
   (b) X_2 is not binomial because the trials are not independent.
   (c) X_3 is not approximately normal because the population is not normal (but is moderately skewed) and the sample size is very small.
   (d) X_4 is approximately normal with mean 40 and variance $8^2/1600$ despite the lack of normality in the population as a consequence of a large sample size and the Central Limit Theorem.

Grade Distribution

```
  100: 3
90-99:42
80-89:42
70-79:23        mean = 84.0, median = 87
60-69: 6
below: 5
```