Problem 1

Fill in the blank with the *capital* letter associated with the word from the following list that *best* illustrates the given scenario.

A. Boxplot	B. Simple Random Sample	C. Stratified Sample	D. Convenience Sample
E. Cluster Sample	F. Probability Sample	G. Time series	H. Cross-sectional
I. Nominal	J. Ordinal	K. Experimental	L. Observational

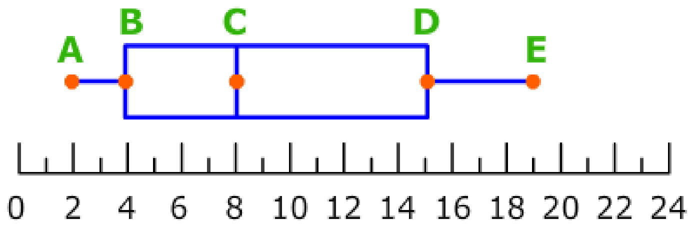
(a) Likert scale questions typically have answers like very dissatisfied, dissatisfied, neutral, satisfied, and very satisfied. This is a J, qualitative variable.

(b) Daily temperatures in Clemson from 1950-2015 is an G, data set.

(c) Noah divided the animal kingdom into species and gender, then randomly picked 1 from each combination. What sampling technique did he use? C

(d) A scientist tries his weight loss drug on a group of monkeys with identical diets. 60 monkeys are randomly assigned to either get the drug or not get the drug (30 in each group). The weight gained/lost was recorded for each monkey. This is an K study.

Problem 2



Use the boxplot above to answer the following questions.

(a) What does the labeled point C represent on the boxplot plot?

A : Mean

B : Median

C : Mode

D : Range

(b) What do the labeled points B and D represent on the boxplot?

A : Mean and Mode

B : Median and Mode

C : Least and Greatest value

D : Lower and Upper quartile

(c) What is the maximum value of the data set?

A : 8

B : 11

C : 15

D : 19

(d) If we replace the maximum value by 38, which of the following statistics WILL NOT change?

A : Range

B : Mean

C : Variance

D : IQR

Problem 3

Use the data pertaining to marital status and gender to calculate the following probabilities.

	Married	Single	Divorced/Widowed	Total
Men	55	100	45	200
Women	90	35	25	150
Total	145	135	70	350

(a) What percent of the individuals were male?

$\frac{200}{350} = 57.14\%$

(b) What percent of the individuals were male and married?

$\frac{55}{350} = 15.71\%$

(c) What percent of the men were single?

$\frac{100}{200} = 50\%$

(d) What percent of the married individuals were male?

$\frac{55}{145} = 37.93\%$

Problem 4

Event A: Rolling at least one six in 4 throws of a die
Event B: Rolling at least one double six in 24 throws of a pair of dice

(a) Let X be the number of six in 4 throws of a die and Y be the number of double six in 24 throws of a pair of dice, State the distribution and parameters for X and Y, respectively.

$X \sim \text{Bin}(n = 4, p = 1/6)$
$Y \sim \text{Bin}(n = 24, p = 1/36)$

(b) Compute $\mathbb{E}[X]$ and $\mathbb{E}[Y]$.

$\mathbb{E}[X] = 4 \times 1/6 = 2/3 = 0.6667$
$\mathbb{E}[Y] = 24 \times 1/36 = 2/3 = 0.6667$

(c) Compute the probability of event A and event B. Which event is more likely to occur?

$\mathbb{P}(A) = 1 - \mathbb{P}(A^c) = 1 - \binom{4}{0}(\frac{1}{6})^0(\frac{5}{6})^4 = 1 - 0.4823 = 0.5177$

$\mathbb{P}(B) = 1 - \mathbb{P}(B^c) = 1 - \binom{24}{0}(\frac{1}{36})^0(\frac{35}{36})^{24} = 1 - 0.5086 = 0.4914$

$\Rightarrow$ Event A is more likely to occur.

Problem 5

Denver Downs has a large pumpkin patch, where the weight of the pumpkins follows a normal distribution with an average of 14 pounds, and a standard deviation of 4 pounds. Each pumpkin's weight is independent of all other pumpkins.

(a) What is the probability that a randomly selected pumpkin weighs over 16 pounds?

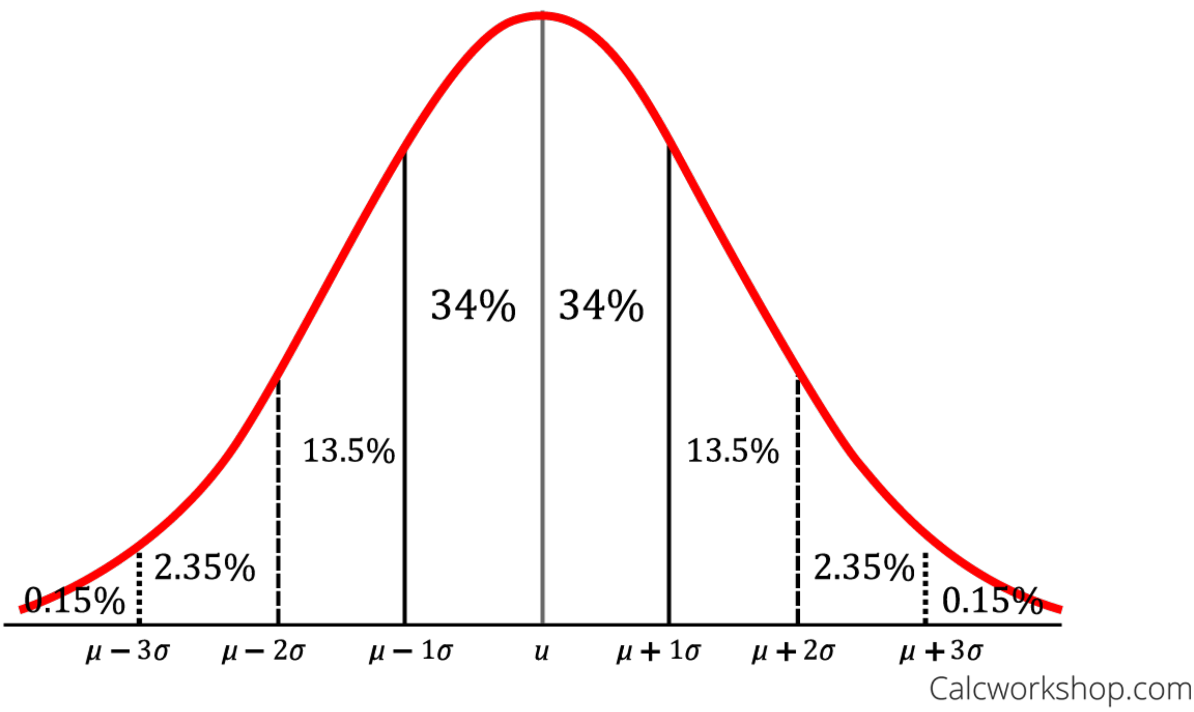
Let X be the weight of a randomly selected pumpkin. $X \sim \mathrm{N}(\mu = 14, \sigma = 4)$

$$\begin{aligned}\mathbb{P}(X > 16) &= \mathbb{P}(Z > (16-14)/4) = \mathbb{P}(Z > 0.5) \\ &= 1 - \mathbb{P}(Z \leq 0.5) = 1 - \Phi(0.5) \\ &= 1 - 0.69146 = 0.30854\end{aligned}$$

(b) What is the probability that a randomly selected pumpkin weighs over 16 pounds given that the selected pumpkin weighs over 14 pounds?

$$\begin{aligned}\mathbb{P}(X > 16 | X > 14) &= \frac{\mathbb{P}(X > 16)}{\mathbb{P}(X > 14)} \\ &= \frac{1 - \Phi(0.5)}{1 - \Phi(0)} \\ &= \frac{0.30854}{0.5} = 0.61708\end{aligned}$$

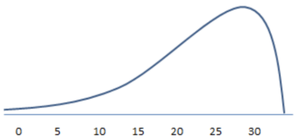
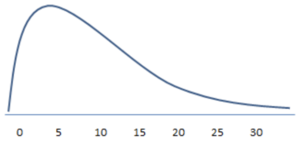
(c) Using the empirical rule to find the cutoff for the top 2.5% of pumpkin weights at Denver Downs.



Based on the empirical rule, the top 2.5% percentile is $\mu + 2\sigma = 14 + 2 \times 4 = 22$lbs.

Problem 6

Use the graphs below to answer the following questions.

A.  B. 

(a) Of graphs A and B, which graph(s) are skewed right? *Circle all that apply.*

A   B   None

(b) Of graphs A and B, which has the larger 25th percentile? *Circle one answer.*

A   B   Same

Problem 7

Given $\mathbb{P}(A) = .6$, $\mathbb{P}(B) = .4$

(a) What is the range of possible value of $\mathbb{P}(A \cap B)$?

[0, .4]

(b) What is the largest possible value of $\mathbb{P}((A \cup B)^c)$?

$\mathbb{P}((A \cup B)^c) = 1 - \mathbb{P}(A \cup B)$.
The smallest possible value of $\mathbb{P}(A \cup B) = .6$. Therefore, the largest possible value of $\mathbb{P}((A \cup B)^c)$ is .4

(c) Suppose $\mathbb{P}(A|B) = 0.8$, what is the value of $\mathbb{P}(B|A)$?

$\mathbb{P}(B|A) = \frac{\mathbb{P}(A|B)\mathbb{P}(B)}{\mathbb{P}(A)} = .53$

Problem 8

Suppose a class has 400 students (to begin with) that each student drops independently of any other student with a probability of .07. Let X be the number of students that finish this course

(a) Find the probability that X is between 370 and 373 inclusive

$X \sim \text{Bin}(n = 400, p = .93) \Rightarrow \mathbb{P}(370 \le X \le 373) = \sum_{x=370}^{373} \binom{400}{x} (.93)^x \times (.07)^{400-x} = .3010$

(b) Is an approximated distribution appropriate for X? If so, what is this distribution and what are the parameter(s)?

Yes, since $np = 372 > 5$ and $n(1-p) = 28 > 5$. The normal approximation for binomial is appropriate.
We will use $X^* \sim N(\mu = 372, \sigma^2 = 400(.93)(.07) = 26.04)$ to approximate $X \sim \text{Bin}(n = 400, p = .93)$

(c) Find the probability that X is between 370 and 373 inclusive by using the approximation in (b).

$\mathbb{P}(370 \le X \le 373) \approx \mathbb{P}(369.5 \le X^* \le 373.5) = \mathbb{P}(-.49 \le Z \le .29) = .6141 - .3121 = .3020$