Worksheets

Stephanie Ranft S2459825

October 25, 2019

Statistics 1A
PSBE1-08

1 Week 4

1. You throw a pair of dice and sum the values shown face up, e.g. if you throw and 2 and a 4, you have a sum of $X = 6$.

   (a) Is this a discrete or continuous probability distribution? Explain your answer, and give the name of it if you know it.

   (b) What is the probability that your sum $X = 2$ (expressed as a fraction)?

   (c) What is the probability that your sum $X = 5$ (expressed as a fraction)?

   (d) What is the probability that at least one of the dice shows a 5 (expressed as a fraction)?

   (e) What is the probability that your sum $X = 12$ (expressed as a fraction)?

   (f) Complete the table below.

x	2	3	4	5	6	7	8	9	10	11	12
$\mathbb{P}(X = x)$											

Table 1

   (g) What is the probability of having at least 10 (expressed as a fraction)?

   (h) You roll the first die and get a 5. After you roll the second die, what is the probability of having at least 10 (expressed as a fraction)?

   (i) What is the expected value and variance?

   (j) Let $Y = X - 7$. What is the expected value and variance of Y?

2. In a Child Health Survey conducted in a school, 1567 children answered "yes" and 433 children answered "no" in response to the question: "Do you drink milk?"

   (a) Estimate the probability that a randomly selected child drinks milk.

   (b) Estimate the probability that a child does not drink milk.

   (c) Of those who drink milk, 40% like it with chocolate. Estimate the probability that a randomly chosen child likes to drink chocolate milk.

3. In a town of 10,000 people, 400 have beards (all men), 4000 are adult men, 5 of the townspeople are murderers. All 5 murderers are men and 4 of the murderers have beards.

   Suppose you go to this town and select a towns person at random.

   - Let A be the event that the person turns out to be one of the five murderers.
   - Let B be the event the person is bearded.
   - Let C be the event the person is an adult male.

   Find $\mathbb{P}(A)$, $\mathbb{P}(A \mid B)$, $\mathbb{P}(B \mid A)$, $\mathbb{P}(A \mid B^c)$, $\mathbb{P}(A \mid C)$, $\mathbb{P}(A \mid C^c)$.

4. Disease X is a disease affecting about 1 percent of the population. A test for Disease X will test positive on all afflicted with the disease and will also test positive for 5% of the population who do not have the disease.

   (a) Draw a tree diagram to display this information.

   (b) Fill in the table.

	Positive	Negative
Disease		
No disease		

Table 2

(c) What is the probability a randomly chosen person has disease X?

(d) What is the probability a randomly chosen person will test positive for the disease?

(e) Suppose you test positive for the disease. What is the probability you don't actually have the disease? (This is the conditional probability that you don't have the disease given you tested positive for it).

(f) What would be the probability that you test positive for the disease twice given that you don't have disease X?

5. A die is thrown 3 times. What is the probability of

(a) No 5's showing up?

(b) One 5 showing up?

(c) Three 5's showing up?

6. Hospital records show that of patients suffering from a certain disease, 75% die of it. What is the probability that of 6 randomly selected patients, 4 will recover?

7. In the old days, there was a probability of 0.8 of success in any attempt to make a telephone call. (This often depended on the importance of the person making the call, or the operator's curiosity!)

(a) Calculate the probability of having 7 successes in 10 attempts.

(b) Can we use the normal distribution to approximate this? Why or why not? If not, how can we modify our distribution so that we can, and what is the resulting mean and standard deviation for this normal approximation?

8. A (blindfolded) marksman finds that on the average he hits the target 4 times out of 9. If he fires 45 shots, what is the probability of

(a) more than 2 hits?

(b) at least 3 misses?

(c) Can we use the normal distribution to approximate this? Why or why not? If not, how can we modify our distribution so that we can, and what is the resulting mean and standard deviation for this normal approximation?

9. The ratio of boys to girls at birth in Singapore is quite high at 1.09:1. What proportion of Singaporian families with exactly 6 children will have at least 3 boys (ignoring the probability of twins, or multiple births).

10. A manufacturer of metal pistons finds that on the average, 12% of his pistons are rejected because they are either oversize or undersize. What is the probability that a batch of 10 pistons will contain no more than 2 rejects? What about the probability that in this batch of 10, there will be at least 2 rejects?

11. Victor is waiting his turn to spin the prize wheel below. Each of the 6 sectors of the prize wheel are equal sizes. If Victor lands on the 6 (orange) he wins $1,000 dollars. If Victor lands on any other sector, he receives nothing.

If Victor has 2 chances to spin the prize wheel, what is the probability that he will win the $1,000 prize?

12. A box contains 5 red balls and 8 violet. A ball is removed and replaced by two of the other colour and then a second ball is drawn.

(a) Draw a tree diagram. Why is this method of displaying the information better than others, for example Venn diagram or tabulated.

(b) Calculate the probability that the second ball is violet.

(c) Calculate the probability that both balls drawn from the box are the same colour.

13. In a class in which all students practice at least one sport, 60% of students play soccer or basketball and 10% practice both sports. There is also 60% that do not play soccer.

(a) What is the best method to represent this information? Draw a diagram.

For the following questions, calculate the probability that a student chosen at random from the class:

(b) Plays soccer only.

(c) Plays basketball only.

(d) Plays only one of the sports.

(e) Plays neither soccer nor basketball.

14. In a city, 40% of the population have brown hair, 25% have brown eyes and 15% have both brown hair and eyes. A person is chosen at random.

    (a) Draw this information in a Venn diagram. Can you also tabulate the information?

    (b) If they have brown hair, what is the probability that they also have brown eyes?

    (c) If they have brown eyes, what is the probability of them not having brown hair?

    (d) What is the probability of them having neither brown hair nor brown eyes?

    (e) Are these events (statistically) independent? Yes/no, why?

15. There are two boxes. Box A contains 6 red balls and 4 blue balls and Box B contains 4 red balls and 8 blue balls. A die is rolled, if the number is less than 3, a ball is selected from box A. If the result is 3 or more, a ball is selected from Box B.

    (a) Decide on an appropriate method to represent this information.

    (b) Calculate the probability that the ball will be red and selected from Box B.

    (c) Calculate the probability that the ball will be blue.

    (d) Are these events (statistically) independent? Yes/no, why?

16. In order to write an exam, a student needs an alarm clock to wake up, which has proven to successfully wake him 80% of the time. If he hears the alarm in the morning the probability of writing the test is 0.9 and if he doesn't hear the alarm the probability is 0.5.

    (a) What is the best way to draw this?

    (b) If he writes the test what is the probability that he heard the alarm clock?

    (c) If he doesn't write the test: what is the probability that he didn't hear the alarm clock?

    (d) Are these events (statistically) independent? Yes/no, why?

17. It is determined that 25 of every 100 men and 600 of every 1,000 women wear glasses. If the number of women in a particular room is four times more than that of men, calculate the probability of:

    (a) woman with glasses being randomly selected.

    (b) person without glasses being randomly selected.

18. There are three sets of key rings A, B and C, for a house. The first set has five keys, the second has seven and the third has eight, of which only one key in each set opens the door to the store room. A key-chain is chose at random followed by a key from the set.

    (a) What is the probability of the selected key being able to open the store room?

    (b) What is the probability that the chosen key chain is from the third set and the key does not open the door?

    (c) What is the probability that the chosen key opens the door and it came from the first key chain?

19. For this problem assume (though it's not quite true) that any child is male or female with probability 0.5.

    (a) You know only that your colleague has three children. What is the probability that they are all girls?

    (b) You know only that your colleague has three children. You ask him if the oldest one is a girl and he says yes. Given this information, what is the probability that all three children are girls?

    (c) You know only that your colleague has three children. You ask him if he is lucky enough to have at least one daughter and he says yes. What now is the probability that all three children are girls?

20. An agent sells life insurance policies to five equally aged, healthy people. According to recent data, the probability of a person living in these conditions for 30 years or more is 2/3. Calculate the probability that after 30 years:

    (a) All five people are still living.

    (b) At least three people are still living.

    (c) Exactly two people are still living.

21. There are 10 red and 20 blue balls in a box. A ball is chosen at random and it is noted whether it is red. The process repeats, returning the ball 20 times.

    (a) Calculate the expected value and the standard deviation of this game.

    (b) Is this a Binomial distribution? Yes/no, why? If yes, can we approximate using the Normal distribution?

22. It has been determined that 5% of drivers checked at a road stop show traces of alcohol and 10% of drivers checked do not wear seat belts. In addition, it has been observed that the two infractions are independent from one another. If an officer stops five drivers at random:

    (a) Calculate the probability that exactly three of the drivers have committed any one of the two offences.

    (b) Calculate the probability that at least one of the drivers checked has committed at least one of the two offences.

    (c) Is this a Binomial distribution? Yes/no, why? If yes, can we approximate using the Normal distribution?

23. A pharmaceutical lab states that a drug causes negative side effects in 3 of every 100 patients. To confirm this affirmation, another laboratory chooses 5 people at random who have consumed the drug. What is the probability of the following events?

    (a) None of the five patients experience side effects.

    (b) At least two experience side effects.

    (c) What is the average number of patients that the laboratory should expect to experience side effects if they choose 100 patients at random?

    (d) Is this a Binomial distribution? Yes/no, why? If yes, can we approximate using the Normal distribution? (Answer this question for the choice of 5 and 100 people.)

1.1 Solutions

1. (a) Discrete. Not Binomial distribution.

   (b) $X = 2$ if you roll two 1's, i.e.

$$\mathbb{P}(\text{die one shows } 1 \cap \text{die two shows } 1) = \frac{1}{6} \cdot \frac{1}{6} = \frac{1}{36}.$$

   This is because rolling of the die are **independent** of each other and the probability with a 6-sided die is $1/6$.

   (c) The sum $X = 5$ when I roll a 1 and a 4, a 2 and a 3, a 3 and a 2, or a 4 and a 1

$$\Longrightarrow \mathbb{P}(X = 5) = \underbrace{\frac{1}{6} \cdot \frac{1}{6}}_{\text{a 1 and a 4}} + \underbrace{\frac{1}{6} \cdot \frac{1}{6}}_{\text{a 2 and a 3}} + \underbrace{\frac{1}{6} \cdot \frac{1}{6}}_{\text{a 3 and a 2}} + \underbrace{\frac{1}{6} \cdot \frac{1}{6}}_{\text{a 4 and a 1}} = \frac{4}{36} = \frac{1}{9}.$$

   (d) The probability that one die shows a 5 (and the other shows any other number) is $2 \cdot 1/6 \cdot 5/6 = 10/36$, and the the probability that both dice show a 5 is $1/6 \cdot 1/6 = 1/36$. Then, the probability that at least one of the dice shows a 5 is the probability that one or both show a 5, i.e. $10/36 + 1/36 = 11/36$.
   Another thing to thing about:

$$\mathbb{P}(\text{at least one}) = 1 - \mathbb{P}(\text{none}).$$

   This intuitively makes sense! Then, the probability that neither of the dice shows a 5 is $5/6 \cdot 5/6 = 25/36$, and the probability of at least one is $1 - 25/36 = 11/36$.

   (e) The sum equals 12 when both dice show a 6, i.e. $1/6 \cdot 1/6 = 36$.

   (f)

x	2	3	4	5	6	7	8	9	10	11	12
$\mathbb{P}(X = x)$	$1/36$	$1/18$	$1/12$	$1/9$	$5/36$	$1/6$	$5/36$	$1/9$	$1/12$	$1/18$	$1/36$

Table 3

   (g) The probability of having at least 10 is the sum of the probabilities of having 10, 11, or 12.

$$\mathbb{P}(X \geq 10) = \mathbb{P}(X = 10) + \mathbb{P}(X = 11) + \mathbb{P}(X = 12) = \frac{1}{12} + \frac{1}{18} + \frac{1}{36} = \frac{1}{6}.$$

   The probabilities are added in this way as the distribution is discrete!

   (h) This is a conditional probability question given by $\mathbb{P}(X = 10 \mid X \geq 6)$; as you have already rolled a 5 and the values on the dice start at 1, you know that your sum is at least 6.

$$\mathbb{P}(X = 10 \mid X \geq 6) = \frac{\mathbb{P}(X = 10 \cap X \geq 6)}{\mathbb{P}(X \geq 6)} = \frac{\mathbb{P}(X = 10)}{\mathbb{P}(X \geq 6)}.$$

   The probability $\mathbb{P}(X \geq 6)$ is calculated as

$$\mathbb{P}(X \geq 6) = \mathbb{P}(X = 6) + \mathbb{P}(X = 7) + \mathbb{P}(X = 8) + \mathbb{P}(X = 9) + \mathbb{P}(X \geq 10) = \frac{5}{36} + \frac{1}{6} + \frac{5}{36} + \frac{1}{9} + \frac{1}{6} = \frac{26}{36}.$$

   Therefore,

$$\mathbb{P}(X = 10 \mid X \geq 6) = \frac{1/12}{26/36} = \frac{3}{26}.$$

   (i) The expected value is the sum of the probabilities times the values, i.e.

$$\mu = \sum_{x=2}^{12} x \cdot \mathbb{P}(X = x) = \frac{2}{36} + \frac{1}{6} + \frac{1}{3} + \frac{5}{9} + \frac{5}{6} + \frac{7}{6} + \frac{40}{36} + 1 + \frac{5}{6} + \frac{11}{18} + \frac{1}{4} = 7.$$

   We compute the variance by look at the sum of product of the squared differences between the values and the mean, and the probabilities, i.e.

$$\sigma^2 = \sum_{x=2}^{1} 2 (x - \mu)^2 \cdot \mathbb{P}(X = x) = \frac{25}{36} + \frac{8}{9} + \frac{3}{4} + \frac{4}{9} + \frac{5}{36} + 0 + \frac{5}{36} + \frac{4}{9} + \frac{3}{4} + \frac{8}{9} + \frac{25}{36} = 5 + \frac{5}{6} \approx 5.83.$$

   By taking the square root of the variance, we conclude that a person playing this game could expect a payout of $7 \pm 2.415 = (4.585, 9.415)$.

(j) The mean of Y is equal to the mean of X minus 7, i.e. zero. The variance of Y is the same as X.

2. (a) The proportion of milk drinkers $p = 1567/(1567 + 433) = 0.7835$.

   (b) The proportion of non- milk drinkers is $1 - p = 0.2165$.

   (c) This is a conditional probability, i.e. $\mathbb{P}(\text{likes chocolate} \mid \text{milk=“yes”}) = 0.40$, and we are looking for the probability that they drink milk **and** like chocolate, i.e.

   $$\mathbb{P}(\text{likes chocolate} \cap \text{milk=“yes”}) = \mathbb{P}(\text{likes chocolate} \mid \text{milk=“yes”}) \cdot \mathbb{P}(\text{milk=“yes”}) = 0.4 \cdot 0.7835 = 0.3134.$$

3.

$$\mathbb{P}(A) = \frac{5}{10000} = 0.0005 \qquad \mathbb{P}(A \mid B) = \frac{4}{400} = 0.01 \qquad \mathbb{P}(B \mid A) = \frac{4}{5} = 0.8$$
$$\mathbb{P}(A \mid B^c) = \frac{5-4}{10000-400} = 0.000104167 \qquad \mathbb{P}(A \mid C) = \frac{5}{4000} = 0.00125 \qquad \mathbb{P}(A \mid C^c) = 0$$

4. (a)

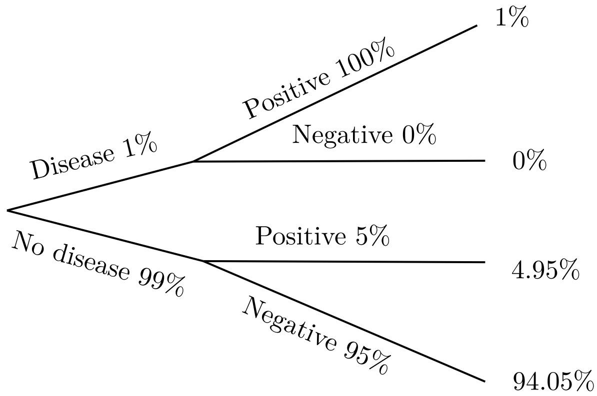

Figure 1

   (b)

	Positive	Negative
Disease	1%	0%
No disease	4.95%	94.05%

Table 4

   (c) $\mathbb{P}(D) = 0.01$.

   (d) $\mathbb{P}(P) = \mathbb{P}(P \mid D) \cdot \mathbb{P}(D) + \mathbb{P}(P \mid ND) \cdot \mathbb{P}(ND) = 1 \cdot 0.01 + 0.05 \cdot 0.99 = 0.0595$.

   (e) $\mathbb{P}(ND \mid P) = \frac{\mathbb{P}(ND \cap P)}{\mathbb{P}(P)} = \frac{\mathbb{P}(P|ND) \cdot \mathbb{P}(ND)}{\mathbb{P}(P)} = \frac{0.05 \cdot 0.99}{0.0595} \approx 0.832$.

   (f) We assume that taking the test are independent events, so $\mathbb{P}(P \text{ twice} \mid ND) = [\mathbb{P}(P \mid ND)]^2 = 0.025$.

5. This is a Binomial distribution where X is the number of 5's shown in three throws of the die.

$$\mathbb{P}(X = x) = \binom{3}{x}\left(\frac{1}{6}\right)^x \left(\frac{5}{6}\right)^{3-x} = \frac{3!}{x!(3-x)!}\frac{5^{3-x}}{6^3}; \quad x = 0, 1, 2, 3.$$

   (a) Here, $x = 0$ so

$$\mathbb{P}(X = 0) = \frac{3!}{0!(3-0)!}\frac{5^{3-0}}{6^3} = \frac{125}{216}.$$

   (b) Here, $x = 1$ so

$$\mathbb{P}(X = 1) = \frac{3!}{1!(3-1)!}\frac{5^{3-1}}{6^3} = 3 \cdot \frac{25}{216} = \frac{75}{216}.$$

   (c) Here, $x = 3$ so

$$\mathbb{P}(X = 3) = \frac{3!}{3!(3-3)!}\frac{5^{3-3}}{6^3} = \frac{1}{216}.$$

6. This is a binomial distribution where X is the number of people (out of 6) who recover.

$$\mathbb{P}(X=x)=\binom{6}{x}\left(\frac{1}{4}\right)^x\left(\frac{3}{4}\right)^{6-x}=\frac{6!}{x!(6-x)!}\frac{3^{6-x}}{4^6}\implies\mathbb{P}(X=4)=\frac{6!}{4!(6-4)!}\frac{3^2}{4^6}\approx 0.03296.$$

7. (a)

$$\mathbb{P}(X=7)=\binom{10}{7}(0.8)^7(0.2)^{10-7}\approx 0.20133.$$

(b) No, we require that np and $n(1-p)$ are both larger than 5. In order to achieve this, we need

$$0.8n\geq 5\quad\text{and}\quad 0.2n\geq 5\implies n\geq 25.$$

If the sample size is 25, we can approximate using a normal distribution with mean $25\cdot 0.8=20$ and variance $25\cdot 0.8\cdot 0.2=4$ (standard deviation of 2).

8.

$$\mathbb{P}(X=x)=\binom{45}{x}\left(\frac{4}{9}\right)^x\left(\frac{5}{9}\right)^{45-x}=\frac{45!}{x!(45-x)!}\frac{4^x5^{45-x}}{9^45}.$$

(a)

$$\begin{aligned}\mathbb{P}(X>2)&=1-\mathbb{P}(X\leq 1)=1-\mathbb{P}(X=1)-\mathbb{P}(X=0)\\&=1-45\cdot\frac{4\cdot 5^{44}}{9^45}-\frac{45\cdot 44}{2}\frac{4^2\cdot 5^{43}}{9^45}\approx 1\end{aligned}$$

(b) If we have at least 3 misses, then we have at most 42 hits, i.e. $\mathbb{P}(X\leq 42)\approx 1$.

(c) As $np=20>5$ and $n(1-p)=25>5$, we can use the normal distribution to approximate this, with mean $np=20$ and standard deviation $\sqrt{np(1-p)}\approx 3.333$. The z-score for $x=2$ is $(2-20)/3.33=-5.4$ and for $x=42$ is $(42-20)/3.33=6.60$, so this helps us make sense of the results for parts (a) and (b).

9. If the ratio of boys to girls is 1.09:1, then the proportion of boys born is $1.09/(1+1.09)\approx 0.5215311$. We can now structure this a binomial distribution where $n=6$ and X is the number of boys:

$$\mathbb{P}(X\geq 3)=1-\mathbb{P}(X\leq 2)=1-\binom{6}{0}0.522^0(1-0.522)^6-\binom{6}{1}0.522^1(1-0.522)^5-\binom{6}{2}0.522^2(1-0.522)^4\approx 0.6957.$$

10. The batch size $n=10$ and X is the number of rejected pistons, with probability 0.12:

$$\mathbb{P}(X=x)=\binom{10}{x}0.12^x0.88^{10-x}.$$

If there are no more than 2 rejects, then

$$\mathbb{P}(X\leq 2)=\mathbb{P}(X=0)+\mathbb{P}(X=1)+\mathbb{P}(X=2)\approx 0.89132.$$

If there are at least two rejects, then

$$\mathbb{P}(X\geq 2)=1-\mathbb{P}(X\leq 1)\approx 0.3417.$$

11. If Victor wins the \$1000 on the first spin, then he obviously doesn't get a second spin. As the game does not allow a spin to land on any two colours at the same time, the events (landing on some colour) are disjoint (mutually exclusive). Therefore any **or** probabilities become additive: $\mathbb{P}(\text{orange or blue})=\mathbb{P}(\text{orange})+\mathbb{P}(\text{blue})-\mathbb{P}(\text{orange and blue})$, but $\mathbb{P}(\text{orange and blue})=0$. We can safely assume that the game is not rigged, and that the spins are independent of each other, so the second spin doesn't affect the outcome of the first.

This is the probability that Victor's first spin lands on orange, **or** his first spin lands on one of the five other colours **and** on the second spin he lands on orange, i.e.

$$\mathbb{P}(\text{orange})=\mathbb{P}(\text{orange 1st spin})\cup\mathbb{P}(\text{orange 2nd spin})=\frac{1}{6}+\frac{1}{6}\cdot\frac{5}{6}=\frac{11}{36}.$$

12. In the original setting, the probability of removing 1 red ball is $\mathbb{P}(B_1=R)={}^5/_{13}$ and the probability of removing 1 violet ball is $\mathbb{P}(B_1=V)={}^8/_{13}$. If we remove 1 red ball, then we add 2 violet balls, i.e. 4 red and 10 violet balls. If we remove 1 violet ball, then we add 2 red balls, i.e. 7 red and 7 violet balls.

(a)

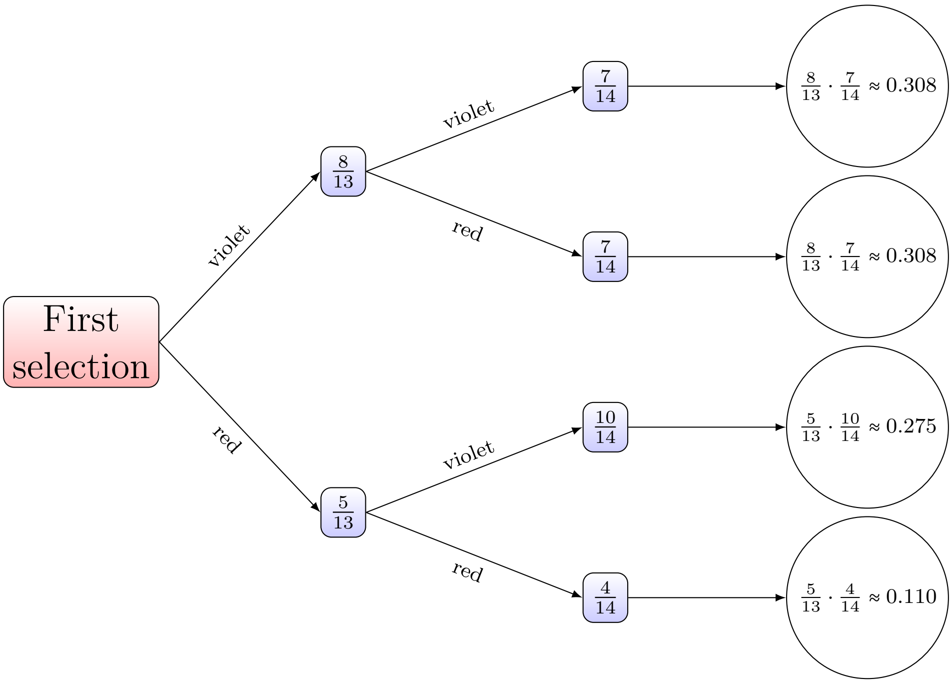

Figure 2

(b) The question is asking us to find the probability that the second ball drawn is violet, but this is conditional on whether the first ball was red or violet.

$$\mathbb{P}(B_2 = V) = \mathbb{P}(\{B_2 = V \cap B_1 = R\} \cup \{B_2 = V \cap B_1 = V\})$$

As we cannot have a ball which is half red and half violet, the events $B_1 = R$ and $B_1 = V$ are mutually exclusive (disjoint), so $\mathbb{P}(\{B_2 = V \cap B_1 = R\} \cap \{B_2 = V \cap B_1 = V\}) = 0$.

$$\implies \mathbb{P}(\{B_2 = V \cap B_1 = R\}) + \mathbb{P}(\{B_2 = V \cap B_1 = V\}) - \underbrace{\mathbb{P}(\{B_2 = V \cap B_1 = R\} \cap \{B_2 = V \cap B_1 = V\})}_{=0}$$

As the events of removing each ball are indepedent, using Bayes' Rule $\mathbb{P}(A \cap B) = \mathbb{P}(A) \cdot \mathbb{P}(B)$, we have that

$$\mathbb{P}(B_2 = V \cap B_1 = R) = \frac{5}{13} \cdot \frac{10}{14} \approx 0.275 \quad \text{and} \quad \mathbb{P}(B_2 = V \cap B_1 = R) = \frac{8}{13} \cdot \frac{7}{14} \approx 0.308$$
$$\implies \mathbb{P}(B_2 = V) = 0.275 + 0.308 = 0.583.$$

(c) This question is asking us for the probability of selected a red ball and then a red ball, or a violet ball and then a violet ball.

$$\begin{aligned}\mathbb{P}(\{B_1 = R \cap B_2 = R\} \cup \{B_1 = V \cap B_2 = V\}) &= \mathbb{P}(B_1 = R \cap B_2 = R) + \mathbb{P}(B_1 = V \cap B_2 = V) \\ &= \mathbb{P}(B_1 = R) \cdot \mathbb{P}(B_2 = R) + \mathbb{P}(B_1 = V) \cdot \mathbb{P}(B_2 = V) \\ &= \frac{5}{13} \cdot \frac{4}{14} + \frac{8}{13} \cdot \frac{7}{14} \approx 0.110 + 0.308 = 0.418.\end{aligned}$$

13. We can calculate this as follows: let S denote the event that a randomly selected child plays soccer, and similarly B denotes basketball.

$$\begin{aligned}\mathbb{P}(S \cup B) = 0.60 = \mathbb{P}(S) + \mathbb{P}(B) - \overbrace{\mathbb{P}(S \cap B)}^{0.10} \\ \implies \mathbb{P}(S) = 0.60 - (\mathbb{P}(B) - 0.10) = 0.70 - \mathbb{P}(B). \\ \mathbb{P}(S^c) = 0.60 = 1 - \mathbb{P}(S) = 1 - (0.70 - \mathbb{P}(B)) = \mathbb{P}(B) + 0.30 \\ \implies \mathbb{P}(B) = 0.60 - 0.30 = 0.30 = \mathbb{P}(B \cap S) + \mathbb{P}(B \cap S^c) \implies \mathbb{P}(B \cap S^c) = 0.30 - 0.10 = 0.20 \\ \implies \mathbb{P}(S) = 0.70 - 0.30 = 0.40 = \mathbb{P}(B \cap S) + \mathbb{P}(B^c \cap S) \implies \mathbb{P}(B^c \cap S) = 0.40 - 0.10 = 0.30.\end{aligned}$$

Also, the area outside the circles is given by $\mathbb{P}\left(B^C \cap S^c\right) = 1 - \mathbb{P}(B \cup S) = 0.40$.

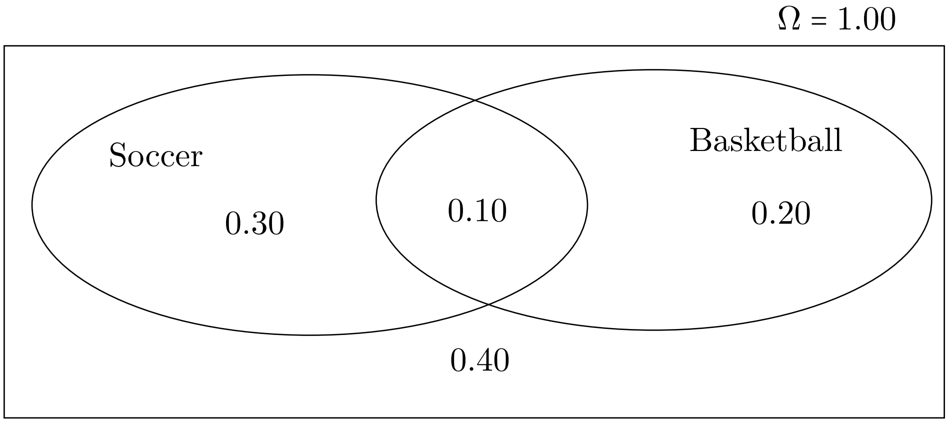

Figure 3

14.

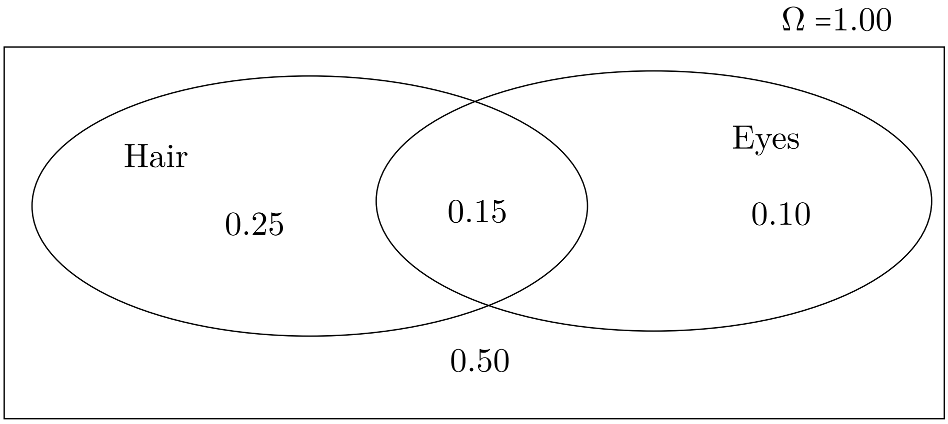

Figure 4

We can calculate this as follows: denote H to be the event that a randomly selected person has brown hair, and similarly denotes E brown eyes.

$$\left.\begin{aligned} \mathbb{P}(H) &= 0.40 \\ \mathbb{P}(E) &= 0.25 \\ \mathbb{P}(H \cap E) &= 0.15 \end{aligned}\right\} \implies \mathbb{P}(H \cup E) = 0.40 + 0.25 - 0.15 = 0.50 = 1 - \mathbb{P}(H^c \cap E^c).$$

We can also tabulate this information and we can fill in the missing (blue) numbers by using the marginal probabilities.

	H	H^c	
E	0.15	0.10	0.25
E^c	0.25	0.50	0.75
	0.40	0.60	1.00

Table 5

15.

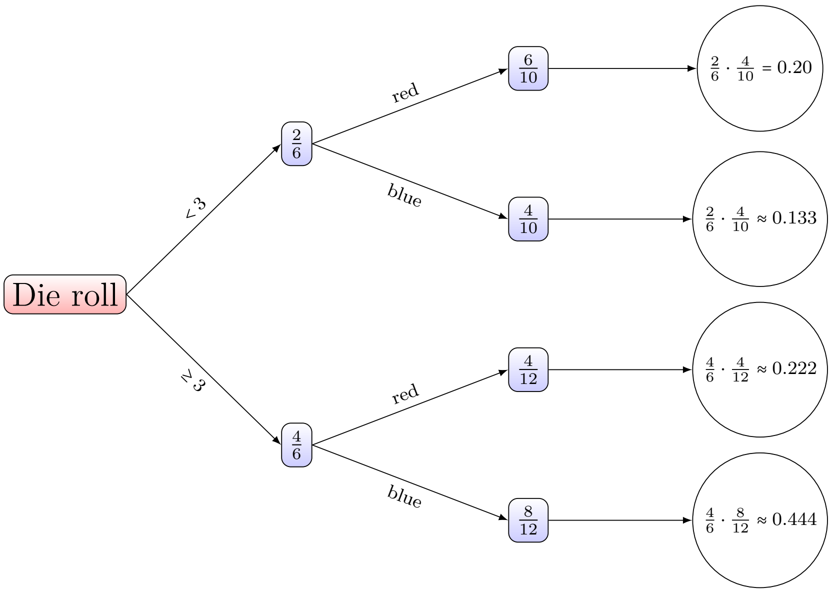

Figure 5

16.

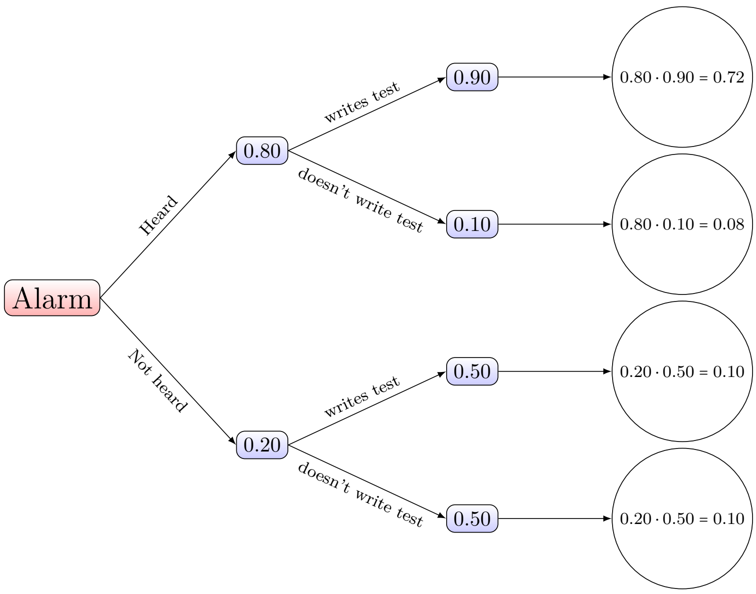

Figure 6

17. Denote M to be the event that a randomly selected person is male, and G that they were glasses. We have that $\mathbb{P}(G \mid M) = 25/100 = 0.25$, $\mathbb{P}(G \mid M^c) = 600/1000 = 0.60$, and the ratio of women to men is $4:1$, i.e. $\mathbb{P}(M) = 1/(1+4) = 0.2 = 1 - \mathbb{P}(M^c)$. Therefore $\mathbb{P}(G \cap M) = \mathbb{P}(G \mid M) \cdot \mathbb{P}(G) = 0.25 \cdot 0.20 = 0.05$.

(a) $\mathbb{P}(G \cap M^c) = \mathbb{P}(G \mid M^c) \cdot \mathbb{P}(M^c) = 0.60 \cdot 0.80 = 0.48$.

(b) $\mathbb{P}(G) = \mathbb{P}(G \cap M) + \mathbb{P}(G \cap M^c) = 0.53 \implies \mathbb{P}(G^c) = 1 - \mathbb{P}(G) = 0.47$.

18.

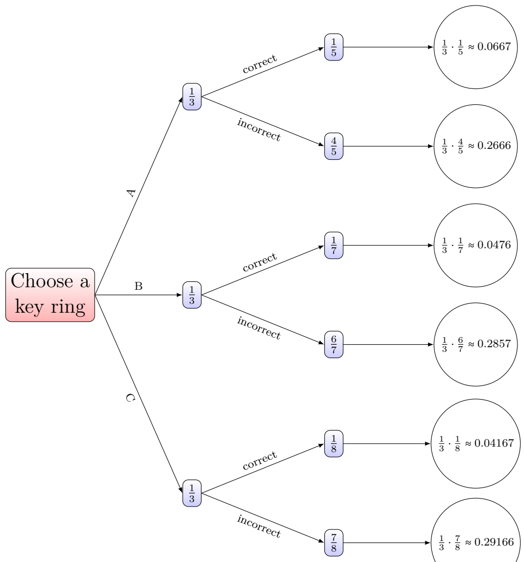

Figure 7

19. (a) There are two ways to reason. First, since the chance of a child being a girl is 1/2, the chance of all three being girls is $\frac{1}{2} \cdot \frac{1}{2} \cdot \frac{1}{2} = \frac{1}{8}$. Alternatively, make a list of all possible outcomes, in order of age: $\{GGG, GGB, GBG, BGG, GBB, BGB, BBG, BBB\}$. As each one is equally likely, the chance of GGG is again 1/8.

(b) Again, we can reason in two ways. If we know the oldest one is a girl, then the chance that all three were girls is simply the chance that the last two children were girls, and this is $\frac{1}{2} \cdot \frac{1}{2} = \frac{1}{4}$. We can also use the list of cases in part (a). Since the first born is a girl, the only cases of consideration are GGG, GGB, GBG and GBB, and only 1 of these 4 is GGG. Thus the chance is simply 1 in 4.

(c) Many people argue that the probability here is again 1/4. But it's not. The information that "there is at least one girl" only removes the outcome BBB, leaving us with the remaining 7, all equally likely. As only one of these is GGG, the probability is 1/7.

20. The event X is whether a person lives passed the 30 year mark with probability $p = 2/3$ and $n = 5$. Use Binomial probability:

$$\mathbb{P}(X = k) = \binom{5}{k}\left(\frac{2}{3}\right)^k \left(\frac{1}{3}\right)^{n-k} \quad ; \quad k = 0, 1, \ldots, 5.$$

21. The event R that we are interested in is whether the ball is removed, which occurs each time (independently) with probability $p = 10/(10 + 20) = 1/3$. The probability is the same each time as the ball is replaced, and $n = 20$.

$$X \sim \mathcal{B}in\left(20, \frac{1}{3}\right) \implies \mu_X = 20 \cdot \frac{1}{3} \approx 6.67, \quad \sigma_X \sqrt{20 \cdot \frac{1}{3} \cdot \frac{2}{3}} \approx 2.108.$$

As $np \geq 5$ **and** $nq \geq 5$, we can use the Normal Approximation to the Binomial Distribution.

22. Let X be the probability that a driver commits either offense, then $p = 0.05 + 0.10 - 0.05 \cdot 0.10 = 0.145$ and $n = 5$.

$$X \sim \mathcal{B}in(5, 0.145) \implies \mathbb{P}(X = k) = \binom{5}{k} 0.145^k 0.855^{n-k}.$$

We cannot use the normal approximation as $n = 5$ is too small. We would require that $n \geq 5/0.145 \approx 34.48$ **and** $n \geq 5/0.855 \approx 5.85$, i.e. the police officer would need to randomly stop at least 35 people.

23.

$$\mathbb{P}(X = k) = \binom{5}{k} 0.03^k 0.97^{n-k} \implies n \geq \frac{5}{0.03} \textbf{ and } n \geq \frac{5}{0.97} \implies n \geq 167.$$

2 Week 5

1. A population has a mean of 50 and a standard deviation of 6.
   (a) What are the mean and standard deviation of the sampling distribution of the mean for $n = 16$?
   (b) What are the mean and standard deviation of the sampling distribution of the mean for $n = 20$?
2. Given a test that is normally distributed with a mean of 100 and a standard deviation of 12, find:
   (a) the probability that a single score drawn at random will be greater than 110.
   (b) the probability that a sample of 25 scores will have a mean greater than 105.
   (c) the probability that a sample of 64 scores will have a mean greater than 105.
   (d) the probability that the mean of a sample of 16 scores will be either less than 95 or greater than 105.
3. What term refers to the standard deviation of the sampling distribution?
4. (a) If the standard error of the mean is 10 for $n = 12$, what is the standard error of the mean for $n = 22$?
   (b) If the standard error of the mean is 50 for $n = 25$, what is it for $n = 64$?
5. A questionnaire is developed to assess women's and men's attitudes toward using animals in research. One question asks whether animal research is wrong and is answered on a 7-point scale. Assume that in the population, the mean for women is 5, the mean for men is 4, and the standard deviation for both groups is 1.5. Assume the scores are normally distributed. If 12 women and 12 men are selected randomly, what is the probability that the mean of the women will be more than 1.5 points higher than the mean of the men?
6. A normal distribution has a mean of 20 and a standard deviation of 10. Two scores are sampled randomly from the distribution and the second score is subtracted from the first. What is the probability that the difference score will be greater than 5?
7. If you sample one number from a standard normal distribution, what is the probability it will be 0.5?
8. A variable is normally distributed with a mean of 120 and a standard deviation of 5. Four scores are randomly sampled. What is the probability that the mean of the four scores is above 127?
9. The mean GPA for students in School A is 3.0; the mean GPA for students in School B is 2.8. The standard deviation in both schools is 0.25. The GPAs of both schools are normally distributed. If 9 students are randomly sampled from each school, what is the probability that:
   (a) the sample mean for School A will exceed that of School B by 0.5 or more?
   (b) the sample mean for School B will be greater than the sample mean for School A?
10. In a city, 70% of the people prefer Candidate A. Suppose 30 people from this city were sampled.
    (a) What is the mean of the sampling distribution of p?
    (b) What is the standard error of p?
    (c) What is the probability that 80% or more of this sample will prefer Candidate A?
    (d) What is the probability that 45% or more of this sample will prefer some other candidate?
11. In the population, the mean SAT score is 1000. Would you be more likely (or equally likely) to get a sample mean of 1200 if you randomly sampled 10 students or if you randomly sampled 30 students? Explain.
12. True/false: The standard error of the mean is smaller when $n = 20$ than when $n = 10$.
13. True/false: You choose 20 students from the population and calculate the mean of their test scores. You repeat this process 100 times and plot the distribution of the means. In this case, the sample size is 100.
14. True/false: In your school, 40% of students watch TV at night. You randomly ask 5 students every day if they watch TV at night. Every day, you would find that 2 of the 5 do watch TV at night.
15. True/false: The median has a sampling distribution.

16. A certain town is served by two hospitals. In the larger hospital, about 45 babies are born each day. In the smaller one, about 15 babies are born each day. Although the overall proportion of girls is about 50%, the actual proportion at either hospital may be greater or less on any day. At the end of a year, which hospital will have the greater number of days on which more than 60% of the babies born were girls?

    (a) the large hospital

    (b) the smaller hospital

    (c) neither - the number of these days will be about the same.

17. The numerical population of grade point averages at a college has mean 2.61 and standard deviation 0.5. If a random sample of size 100 is taken from the population, what is the probability that the sample mean will be between 2.51 and 2.71?

2.1 Solutions

1. (a) $\bar{X} \sim \mathcal{N}\left(\mu = 50, \sigma = \sqrt{\frac{\sigma^2}{n}} = \frac{6}{4} = 1.5\right)$.

   (b) $\bar{X} \sim \mathcal{N}\left(\mu = 50, \sigma = \sqrt{\frac{\sigma^2}{n}} = \sqrt{1.2} \approx 1.34\right)$

2. (a) $\mathbb{P}(X > 110) = \mathbb{P}\left(Z > \frac{110-100}{12}\right) = \mathbb{P}(Z > 0.83) = 0.203269$

   (b) $\mathbb{P}\left(\bar{X}_{25} > 105\right) = \mathbb{P}\left(Z > \frac{105-100}{12/5}\right) = \mathbb{P}(Z > 2.08) = 0.018763$

   (c) $\mathbb{P}\left(\bar{X}_{64} > 105\right) = \mathbb{P}\left(Z > \frac{105-100}{12/8}\right) = \mathbb{P}(Z > 3.33) = 0.000434$

   (d) $\mathbb{P}\left(\bar{X}_{16} > 105\right) + \mathbb{P}\left(\bar{X}_{16} < 95\right) = \mathbb{P}\left(Z > \frac{105-100}{12/4}\right) + \mathbb{P}\left(Z < \frac{95-100}{12/4}\right) = \mathbb{P}(Z > 1.67) + \mathbb{P}(Z < -1.67) = 2\,\mathbb{P}(Z < -1.67) = 0.094919$

3. Standard error (of the estimate of μ)

4. (a) $\sigma/\sqrt{12} = 10 \implies \sigma = 10 * \sqrt{12} \implies \sigma/\sqrt{22} = 10 * \sqrt{12}/\sqrt{22} \approx 7.3855$

   (b) $\sigma/\sqrt{25} = 50 \implies \sigma = 5 * 50 = 250 \implies \sigma/\sqrt{64} = 250/8 = 31.25$

5. Construct a new variable $Y = W - M$, where $W \sim \mathcal{N}(\mu = 5, \sigma = 1.5)$ and $W \sim \mathcal{N}(\mu = 4, \sigma = 1.5)$. Then $W \sim \mathcal{N}\left(\mu = 5 - 4 = 1, \sigma = \sqrt{(1.5)^2 + (1.5)^2} \approx 2.12\right)$ such that $\mathbb{P}\left(\bar{W} - \bar{M} > 1.5\right) = \mathbb{P}\left(\bar{Y} > 1.5\right) = \mathbb{P}\left(Z > \frac{1.5-1}{2.12/\sqrt{12}}\right) = \mathbb{P}(Z > 0.82) = 0.206108$.

6. Construct a new variable $D = X_1 - X_2 \sim \mu_{\mathcal{D}}, \sigma_{\mathcal{D}}$, where $\mu_D = 20 - 20 = 0$ and $\sigma_D = \sqrt{10^2 + 10^2} \approx 14.14$. Then $\mathbb{P}(X_1 - X_2 > 5) = \mathbb{P}(D > 5) = \mathbb{P}\left(Z > \frac{5-0}{14.14}\right) = \mathbb{P}(Z > 0.35 = 0.363169)$.

7. $\mathbb{P}(Z = 0.5) = 0$, as this is a continuous distribution.

8. $\bar{X} \sim \mathcal{N}\left(120, 5/\sqrt{4} = 2.5\right)$. $\mathbb{P}\left(\bar{X} > 127\right) = \mathbb{P}\left(Z > \frac{127-120}{2.5}\right) = \mathbb{P}(Z > 2.8) = 0.002555$

9. $D = A - B \sim \mathcal{N}(\mu_D, \sigma_D)$, where $\mu_D = 3.0 - 2.8 = 0.2$ and $\sigma_D = \sqrt{0.25^2 + 0.25^2} \approx 0.354$. Then, $\bar{D} \sim \mathcal{N}\left(0.2, 0.354/\sqrt{9} \approx 0.118\right)$.

   (a) $\mathbb{P}\left(\bar{A} - \bar{B} > 0.5\right) = \mathbb{P}\left(\bar{D} > 0.5\right) = \mathbb{P}\left(Z > \frac{0.5-0.2}{0.118}\right) = \mathbb{P}(Z > 2.55) = 0.005386$

   (b) $\mathbb{P}\left(\bar{B} > \bar{A}\right) = \mathbb{P}\left(\bar{B} - \bar{A} > 0\right) = \mathbb{P}\left(-\bar{D} > 0\right) = \mathbb{P}\left(\bar{D} < 0\right) = \mathbb{P}\left(\bar{D} < \frac{0-0.2}{0.118}\right) = \mathbb{P}(Z < -1.67) = 0.04746$

10. $X \sim \mathcal{B}in(n = 30, p = 0.7)$; $np = 12 > 5$ and $nq = 9 > 5$ so we can use the normal approximation $X \sim \mathcal{N}(21, 2.510)$.

    (a,b) $X = np \implies p = X/n \implies p \sim \mathcal{N}\left(\frac{\mu_X}{n} = 0.7, \frac{\sigma_X}{n} \approx 0.0837\right)$.

    (c) $\mathbb{P}(p > 0.8) = \mathbb{P}\left(Z > \frac{0.8-0.7}{.0837}\right) = \mathbb{P}(Z > 1.19) = 0.117023$

    (d) $\mathbb{P}(q > 0.45) = \mathbb{P}(1 - p > 0.45) = \mathbb{P}(1 - 0.45 = 0.55 > p) = \mathbb{P}\left(Z < \frac{0.55-0.7}{0.0837}\right) = \mathbb{P}(Z < -1.79) = 0.036727$

11. As the standard error of the estimate for μ is larger in a sample size of $n = 10$ (small), than in a sample of $n = 30$ (sufficiently large), we expect greater variation (from $\mu = 1000$) in the small samples. Therefore it is more likely to observe a sample mean of 1200 in a sample of 10 students than in a sample of 30 students.

12. False; see above.

13. False; $N = 100$ number of samples (repetitions), $n = 20$ sample size.

14. False; see answer to question 11.

15. True; Median is a central location statistic, and the IQR is a measure of dispersion about the median. In a large enough sample, the data points cluster around the median in a "normal" way. Therefore the central limit theorem applies, and it is possible to assign a normal sampling distribution.

16. The answer is: the smaller hospital. The Law of Large Numbers (see the Normative Rules) says: The more people we survey, the more accurate the result. Larger samples are better.

    In this case, the number of children born each day is a random sample of the number of children born in a year. There is much more deviation in a small (daily, in this case) sample than over an entire year. Similarly, there will be more deviation in a sample of size 15 than in one of size 45. Even though both hospitals will have about 50% girls over the year, the smaller hospital will have more days in which the number is higher than 60% (and more days when it is less than 40%). (Consider the extreme case of a hospital that has one baby a day. Every day will be either more than 60% girls (one) or less than 40% girls (zero).

(Exercise for the reader: Actually the proportion of girls is about 49%. How does this affect the answer?)

People commonly and mistakenly believe that even very small samples will have the same composition as large ones. Remember this next time someone uses a small survey sample to "prove" something – the results may be invalid!

17. $X = GPA \sim \mathcal{N}(2.61, 0.5) \implies \bar{X} \sim \mathcal{N}\left(2.61, 0.5/\sqrt{100} = 0.05\right)$, then

$$\mathbb{P}\left(2.51 < \bar{X} < 2.71\right) = \mathbb{P}\left(\frac{2.51 - 2.61}{0.05} < Z < \frac{2.71 - 2.61}{0.05}\right) = \mathbb{P}(-2 < Z < 2) \approx 0.95.$$

We have used the 68-95-99.7 rule to approximate the probability, but you can also calculate it exactly as 0.9545.