Unit 19: Probability Models

PREREQUISITES

Students should be introduced to the concept of probability before working through this unit. Unit 18, Introduction to Probability, will provide that needed background.

ADDITIONAL TOPIC COVERAGE

Additional coverage of probability models can be found in *The Basic Practice of Statistics*, Chapter 10, Introducing Probability.

ACTIVITY DESCRIPTION

In this activity, students collect data from rolling dice to see how closely the probability models in Tables 19.1 and 19.2 capture the patterns of real data. Students can work individually or in pairs. Collecting the data – outcomes from 100 rolls of one die and sums from rolling two dice – could be done as a homework assignment to be completed before beginning the activity.

MATERIALS

Two dice for each student (or each pair).

The activity is in two parts. In Part I, students roll a single (fair) die 100 times and organize their results into a relative frequency table. Then they compare the relative frequencies from their data with the probability model in Table 19.2. They represent their results with a histogram and compare it to Figure 19.2, the probability histogram for rolling a single die. Students may find that their results do not match the probability model as closely as they expect. This is where you can remind them, for example, that the probability of rolling a 3 is the relative frequency (or proportion) of 3's over many, many rolls – 100 rolls is not enough.

To get more data, individuals or pairs combine their results to form the class data. The relative frequency tables and histograms for the class data should come closer to the probability model and probability histogram for rolling a single die.

Part II is a repeat of Part I, only this time students roll two dice and collect data on the sums.

THE VIDEO SOLUTIONS

1. A probability model is the set of all possible outcomes together with the probabilities associated with those outcomes.

2. $S = \{2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12\}$

3. $P(7) = 6/36 = 1/6$

4. $P(\text{not } A) = 1 - P(A)$

5. If events *A* and *B* are mutually exclusive, you can use the Addition Rule to calculate $P(A \text{ or } B)$, the probability that either *A* or *B* occurs.

6. If events *A* and *B* are independent, you can use the Multiplication Rule to calculate $P(A \text{ and } B)$, the probability that both *A* and *B* occur.

UNIT ACTIVITY:
PROBABILITY MODELS AND DATA SOLUTIONS

1.a. Sample answer: Here are the 100 outcomes when we rolled the die. (See solution to (b) for the frequencies.)

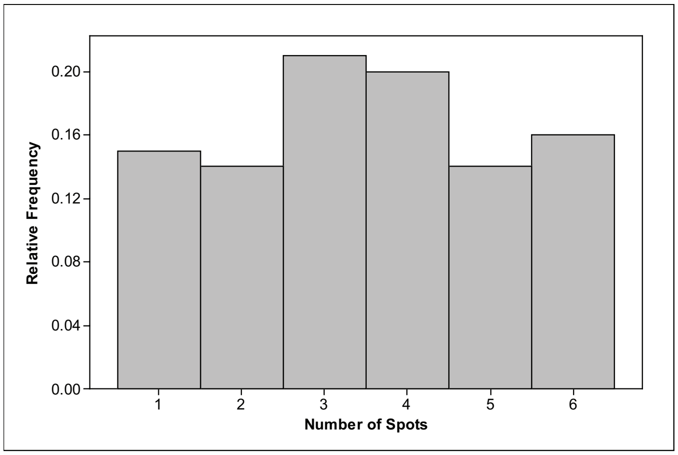
2	6	3	3	5	5	6	4	3	6	2	1	1	6	4	1	3	6	4	1
3	3	5	3	3	4	5	6	3	5	5	6	3	3	1	2	2	4	1	4
2	4	4	6	3	4	3	1	3	6	5	5	2	3	4	4	5	4	6	3
1	2	1	2	5	5	6	6	6	5	4	4	4	4	3	2	2	5	1	4
2	3	4	1	3	4	2	4	6	6	1	1	6	2	1	2	1	3	5	3

b. Sample answer (based on data from 1(a)):

Number of Spots	Frequency	Relative Frequency
1	15	0.15
2	14	0.14
3	21	0.21
4	20	0.20
5	14	0.14
6	16	0.16

The relative frequencies range in value from 0.14 to 0.21, so they are somewhat close to $1/6 \approx 0.167$ but they are certainly not equal.

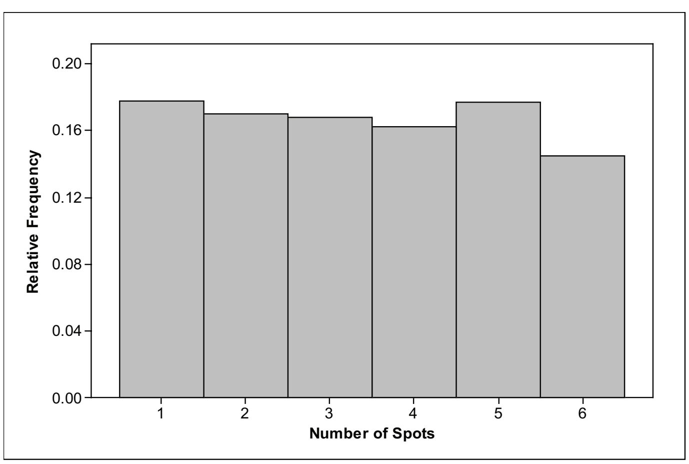
c. Sample answer: Unlike Figure 19.2, the bars are uneven in height.



2. Sample answer:

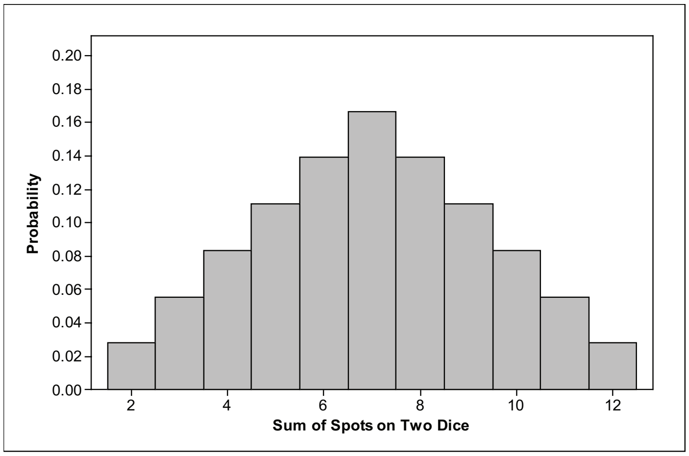
Number of Spots	Frequency	Relative Frequency
1	178	0.178
2	170	0.170
3	168	0.168
4	162	0.162
5	177	0.177
6	145	0.145

For the combined class results the relative frequencies are closer to $1/6^{th}$ than when the die was rolled 100 times. The members of the class rolled the die 1000 times.



The histogram based on the sample data of 1000 rolls has a closer resemblance to the probability histogram than the histogram based on sample data of 100 rolls.

3.

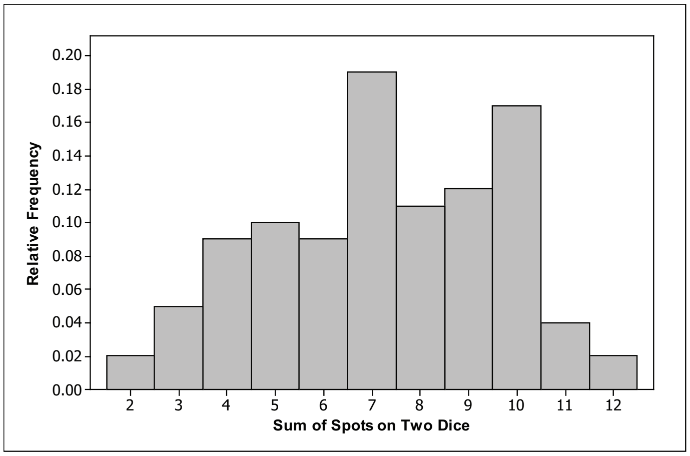


4. a. See solution to (b).

b. Sample answer: Most of the relative frequencies are close to the probabilities. The biggest difference is for the sum of 10; the relative frequency was 0.17 compared to a probability of around 0.083, a difference of 0.087 (more than double the actual probability).

Sum of Spots	Frequency	Relative Frequency	Probability
2	2	0.02	0.028
3	5	0.05	0.056
4	9	0.09	0.083
5	10	0.10	0.111
6	9	0.09	0.139
7	19	0.19	0.167
8	11	0.11	0.139
9	12	0.12	0.111
10	17	0.17	0.083
11	4	0.04	0.056
12	2	0.02	0.028

c. Sample answer (see next page): The left side of the histogram resembles the probability histogram more than the right side. (Expect a fair amount of variability in the shapes of students' histograms.)

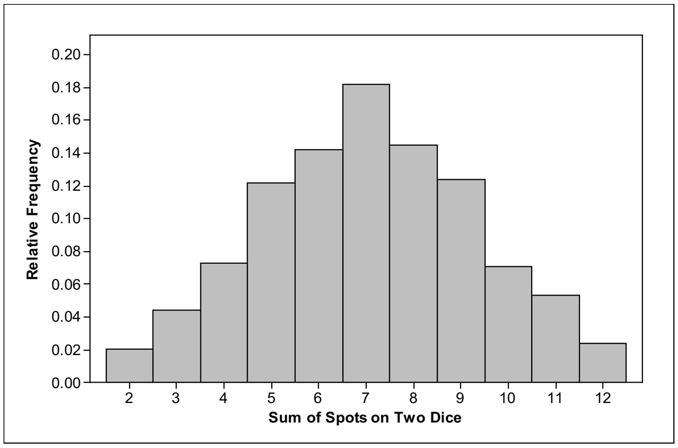


5. Sample answer: The relative frequencies are closer for the 1000 rolls than they were for the 100 rolls. The largest discrepancy was for the sum of seven where the relative frequency was off by 0.015.

Sum of Spots	Frequency	Relative Frequency	Probability
2	20	0.020	0.028
3	44	0.044	0.056
4	73	0.073	0.083
5	122	0.122	0.111
6	142	0.142	0.139
7	182	0.182	0.167
8	145	0.145	0.139
9	124	0.124	0.111
10	71	0.071	0.083
11	53	0.053	0.056
12	24	0.024	0.028

The histogram from 1000 rolls strongly resembles the probability histogram. So, the probability model appears to match what happens with the data over many rolls of the dice (certainly more than 100).

See next page...



Relative Frequency
0.20
0.18
0.16
0.14
0.12
0.10
0.08
0.06
0.04
0.02
0.00
2 3 4 5 6 7 8 9 10 11 12
Sum of Spots on Two Dice

EXERCISE SOLUTIONS

1. a. 0.01

b. *P*(not Public Transportation) = 1 – *P*(Public Transportation) = 1 – 0.05 = 0.95

c. *P*(Drives) = *P*(Drives Alone) + *P*(Carpool) = 0.76 + 0.12 = 0.88

d. *P*(not Drive) = 1 – *P*(Drive) = 1 – 0.88 = 0.12

2. a. From 1(c) *P*(Drives) = 0.88. Since the workers were chosen randomly, the fact that one drives to work does not affect that probability that the other also drives to work. So, we can use the Multiplication Rule.

P(Both drive) = *P*(Worker 1 drives and Worker 2 drives)

= *P*(Worker 1 drives)*P*(Worker 2 drives)

= (0.88)(0.88) = 0.7744

b. From 1(d) *P*(not Drives) = 0.12.

P(Neither drives) = *P*(Worker 1 does not drive and Worker 2 does not drive)

= *P*(Worker 1 does not drive)*P*(Worker 2 does not drive)

= (0.12)(0.12) = 0.0144

P(at least one drives) = 1 – *P*(neither drives) = 1 – 0.0144 = 0.9856

3. a. It would mean that the probability of having that blood type is 1/2 or 50%; it means that a person is just a likely to have that blood type as to have a different blood type.

b. *P*(Rh+) = *P*(A+, B+, AB+, O+)

= *P*(A+) + *P*(B+) + *P*(AB+) + *P*(O+)

= 0.357 + 0.085 +0.034 + 0.374 = 0.850

The chance that a person has Rh-positive blood is higher than 50-50.

c. 85%

d. Sample answer: This is not a valid reason. First, P(O+ or A+) = 0.374 + 0.357 = 0.731. Assuming that the chances of needing a transfusion do not depend on blood type, you would expect about 73% of the people who need blood transfusions to be either type O+ or A+. In order to meet the higher need, the blood banks would want approximately 73% of blood donations to be from people who are type O+ or A+. In addition, it is particularly important for people with type O+ blood to donate because their blood type could be used in 85% of all blood transfusions – for all people who have Rh-positive blood.

4. a. First, we determine the probability of having type O blood:

P(O+ or O-) = P(O+) + P(O-) = 0.374 + 0.066 = 0.440

In a random sample of size 2, the fact that Person 1 is type O does not affect the probability that Person 2 is type O. So, we can use the Multiplication Rule:

P(Person 1 is O and Person 2 is O) = P(Person 1 is O)P(Person 2 is O) = (0.440)(0.440) ≈ 0.1936

b. First, we determine the probability of not having type O blood:

P(not O) = 1 – P(O) = 1 – 0.440 = 0.560

P(exactly one is O) =

P(Person 1 is O and Person 2 is not O) + P(Person 1 is not O and Person 2 is O) =

P(Person 1 is O)P(Person 2 is not O) + P(Person 1 is not O)P(Person 2 is O) =

(0.440)(0.560) + (0.560)(0.440) = 0.4928

c. P(Person 1 not O and Person 2 not O) = P(Person 1 not O)P(Person 2 not O) = (0.560)(0.560) ≈ 0.3136

REVIEW QUESTIONS SOLUTIONS

1. a. P(C) = P(Straight or Right) = P(Straight) + P(Right) = 0.6 + 0.25 = 0.85

b. P(not Straight) = 1 - P(Straight) = 1 – 0.6 = 0.4 (Complement Rule).

Let D be the event that neither vehicle goes straight.

$P(D)$ = P(Vehicle 1 not Straight and Vehicle 2 not Straight) =

P(Vehicle 1 not Straight)P(Vehicle 2 not Straight) =

(0.4)(0.4) = 0.16. (Multiplication Rule)

c. If D is the event that neither vehicle goes straight, then not D is the event that at least one of the vehicles goes straight. P(not D) = 1 – $P(D)$ = 1 – 0.16 = 0.84. (Complement Rule)

2. a. S = {GGG, GGN, GNG, GNN, NGG, NGN, NNG, NNN}

b. A = {GNN, NGN, NNG}, B = {GGN, GNG, NGG}, C = {GGG}, and

D = {GNN, NGN, NNG, GGN, GNG, NGG, GGG}

c. A and B are mutually exclusive; A and C are mutually exclusive; B and C are mutually exclusive.

d. $P(C)$ = P(G)P(G)P(G) = (0.51)(0.51)(0.51) = 0.132651 ≈ 0.133

$P(D)$ = 1 – P(NNN) = 1 – P(N)P(N)P(N) = 1 – (0.49)(0.49)(0.49) = 0.882351 ≈ 0.882

To find P(A) we first find:

P(GNN) = P(G)P(N)P(N) = (0.51)(0.49)(0.49) = 0.122451

Note that P(NGN) = P(NNG) = P(GNN)

$P(A)$ = (3)(0.122451) = 0.367353 ≈ 0.367

To find $P(B)$ we first find:

P(GGN) = (0.51)(0.51)(0.49) = 0.127449

Since P(GGN) = P(GNG) = P(NGG), P(B) = (3)(0.127449) = 0.382347 ≈ 0.382

3. a. All the probabilities are between 0 and 1; the sum of the probabilities equals 1: 0.223 + 0.188 + 0.138 + 0.179 + 0.272 = 1

b. *P*(Less than $100,000) = 1 – *P*($100,000 or over) = 1 – 0.272 = 0.728

c. *P*($75,000 or over) = *P*($75,000 to $99,000 or $100,000 or over)

= *P*($75,000 to $99,000) + *P*($100,000 or over)

= 0.179 + 0.272 = 0.451

d. *P*(Below $75,000) = 1 – *P*($75,000 or over) = 1 – 0.451 = 0.549; 54.9% of households will have total incomes below $75,000.

4. a. Since the three households were chosen randomly, the fact that one of the households has a total income of under $25,000 does not affect the chances that either of the other two households has a total income of under $25,000. We need this for independence.

P(all three households have total incomes under $25,000) =
P(House 1 under $25,000)*P*(House 2 under $25,000)*P*(House 3 under $25,000) =
(0.223)(0.223)(0.223) = 0.0111

b. *P*($25,000 or over) = 1 – 0.223 = 0.777
P(all three households have total incomes of $25,000 or more) =
P(House 1 $25,000 or more)*P*(House 2 $25,000 or more)*P*(House 3 $25,000 or more) =
(0.777)(0.777)(0.777) = 0.469097 ≈ 0.469

c. *P*(at least one household $25,000 or over) = 1 – *P*(no household $25,000 or over) =
1 – *P*(all households under $25,000) = 1 – 0.0111 = 0.9889