

Introductory Statistics – Day 6

Probability and Pivot Tables

Definition (Probability)

*The **probability** of an outcome is the proportion of times the outcome would occur if we observed the random process an infinite number of times.*

Definition (Probability)

*Another definition: The **probability** of an event A is the ratio of the cases that are part of event A out of all possible cases in the sample space.*

Example: With a deck of 52 cards, $\text{Probability}(\text{picking a facecard}) = 12 / 52$.

Example: With a fair 6-sided die, $\text{Probability}(\text{rolling a 2}) = 1/6$

Definition (Independent Events)

When the probability of an event does not change just because another event occurs, then that event is **independent** of the other.

If events A and B are independent then

$$P(A \text{ and } B) = P(A) \times P(B).$$

What kinds of events are independent? What kinds of events are not?

Activity 1: With the person sitting next to you, come up with some independent events and some dependent events.

Activity 2: Pivot table introduction

Open the `videoGamesXBoxPS3.xls` data set from Moodle. Suppose we want to know how many of the games for each console type have a rating of M for Mature, T for teen, E for everyone, or E+10.

Fill in the following chart with counts from Excel.

	M for mature	T for teen	E+10	E for everyone	Total
PS3					
Xbox					
Total					

- 1 If a game is selected at random, what's the chance it is rated E?
- 2 If a game is selected at random, what's the chance it is rated E and it's an XBox game?
- 3 If a game is selected at random, what's the chance it is rated E and it's a PS3 game?
- 4 If a game is selected at random, what's the chance it is rated M?
- 5 What proportion of M rated games are for XBox? (This time, limit your focus to only M-rated games)

Activity 3: Pivot table task

Open the *NCBabySmoke* data set from Moodle.

This data set is from the OpenIntro textbook. It represents a random sample of 1000 mothers and their newborns in North Carolina. Here is a codebook for the data set.

column name	description and units
fage	father's age
mage	mother's age
mature	under 35 vs. 35 or older
weeks	length of pregnancy
premie	premie or full term
visits	number of doctor visits
marital	married or not married
gained	weight gained by mom (lbs)
weight	weight of baby (lbs)
lowbirthweight	low is ≤ 5.5 lbs
gender	baby's gender
habit	smoking habit of mom
whitemom	white or not white

Activity 3, continued.

Use pivot tables to answer the following probability questions.

- 1 If a baby is randomly selected from this data set, what's the chance it is a premie?
- 2 If a baby is randomly selected from this data set, what's the chance it is male?
- 3 If a baby is randomly selected from this data set, what's the chance it is a male premie?
- 4 What's the chance a randomly selected baby weighs less than 5.5 lbs?
- 5 If a mom is randomly selected from this data set, what's the chance she smokes?
- 6 If a mom is randomly selected from this data set, what's the chance she is 35 or older?
- 7 If a mom is randomly selected from this data set, what's the chance she is married?
- 8 What's the chance of randomly selecting a mom who is 28 years old?

Activity 4:

The following table provides information on housing units in some part of the U.S. The top row indicates how many rooms the housing unit has. The second row indicates how many thousands of units correspond with that number.

Rooms	1	2	3	4	5	6	7	8+
Thousands of units	47	140	1170	2350	2450	2130	1370	1560

Find the probability of each of the events A through E and then answer the follow-up questions.

A = the unit has at most 4 rooms

B = the unit has at least 2 rooms

C = the unit has between 5 and 7 rooms (inclusive)

D = the unit has more than 7 rooms

E = the unit has less than 3 rooms

1 Find: $P(A \text{ or } D)$

4 Find: $P(B \text{ or } E)$

2 Find: $P(A \text{ and } D)$

5 Find: $P(\text{not } B)$

3 Find: $P(A \text{ or } C)$

Activity 5:

A consumer organization estimates that over a 1-year period 17% of cars will need to be repaired once, 7% will need repairs twice, and 4% will require three or more repairs.

What is the probability that a car chosen at random will need exactly 1 repair?
(A) 0.17 (B) 0.07 (C) 0.89 (D) 0.28 (E) 0.72

Activity 5:

A consumer organization estimates that over a 1-year period 17% of cars will need to be repaired once, 7% will need repairs twice, and 4% will require three or more repairs.

What is the probability that a car chosen at random will need no repairs?

(A) 0.17

(B) 0.07

(C) 0.89

(D) 0.28

(E) 0.72

Activity 5:

A consumer organization estimates that over a 1-year period 17% of cars will need to be repaired once, 7% will need repairs twice, and 4% will require three or more repairs.

What is the probability that a car chosen at random will need at least 1 repair?
(A) 0.17 (B) 0.07 (C) 0.89 (D) 0.28 (E) 0.72

Activity 5:

A consumer organization estimates that over a 1-year period 17% of cars will need to be repaired once, 7% will need repairs twice, and 4% will require three or more repairs.

What is the probability that a car chosen at random will need no more than 1 repair?

(A) 0.17

(B) 0.07

(C) 0.89

(D) 0.28

(E) 0.72

Activity 5:

A consumer organization estimates that over a 1-year period 17% of cars will need to be repaired once, 7% will need repairs twice, and 4% will require three or more repairs.

If you own two cars, what is the probability that both will need exactly one repair?

- (A) 0.0289 (B) 0.0049 (C) 0.7921 (D) 0.0784
(E) 0.5184

Activity 5:

A consumer organization estimates that over a 1-year period 17% of cars will need to be repaired once, 7% will need repairs twice, and 4% will require three or more repairs.

If you own two cars, what is the probability that neither care will need a repair?

(A) 0.0289

(B) 0.0049

(C) 0.7921

(D) 0.0784

(E) 0.5184