Statistical Analysis

JOHNS HOPKINS
CAREY BUSINESS SCHOOL

Probability

1. The Concept of Probability
2. Sample Spaces and Events
3. Some Elementary Probability Rules
4. Conditional Probability and Independence
5. Bayes' Theorem
6. Counting Rules

Definitions

Probability- The chance that something will happen expressed in fractions, decimals or percents

Event- One or more of the possible outcomes of doing something

Experiment- The activity that produces an event

Sample Space- A set of possible outcomes of an experiment

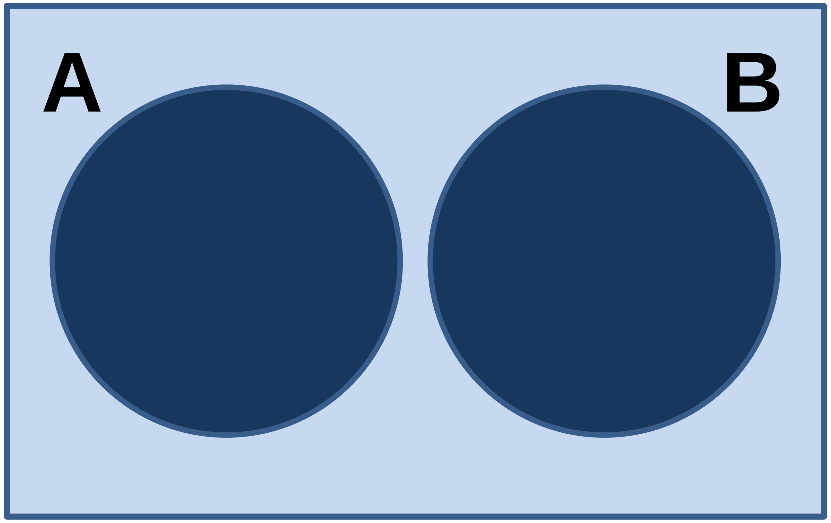
Mutually Exclusive- Events are said to be mutually exclusive if one and only one can take place at a given time

Collectively Exhaustive- A list of all the possible events that can result from an experiment

Mutually Exclusive

A and B are mutually exclusive if they have no sample space outcomes in common

In other words:
$P(A \cap B) = 0$



Three Types of Probabilty

a) Classical

Statements of probability based on logical reasoning before any experiment takes place $P(A) = \frac{(\# of \ occurences \ of \ A)}{(total \ \# \ of \ \ possible \ outcomes)}$

b) Relative Frequency of Occurrence

Proportion of times that an event occurs in the long run when conditions are stable – observed relative frequency of an event in a very large number of trials

c) Subjective

The probability assigned to an event on the basis of whatever evidence is available – the decision maker can assign biases to manipulate evidence

Rules

- Probabilities can be expressed as percentages, decimals or fractions

  Example: probability of tossing heads with an unbias coin 50%, ½ or 0.50

- With classical probability, all outcomes are known and therefore, must total 1.00
- Classical probability is also known as a priori probability and objective probability

Three Types of Classical Probability and Notation

Marginal Probability – the probability of a single event $P(A)$

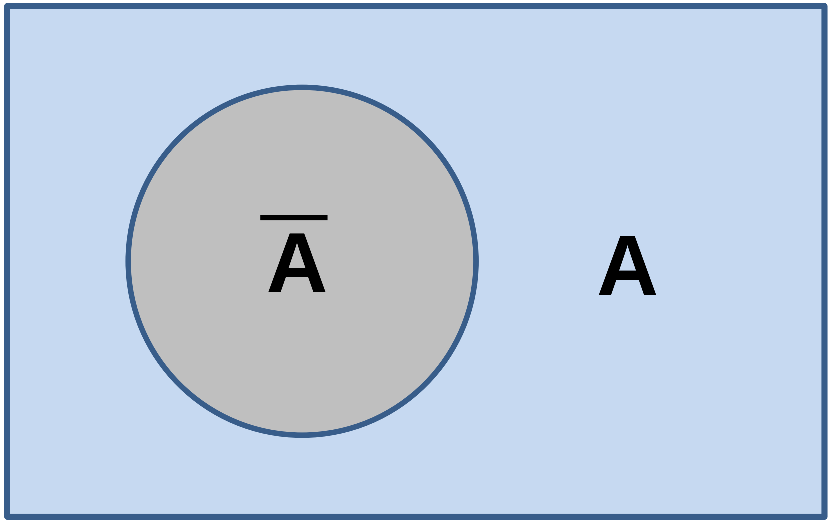
Joint Probability – probability of two or more events occurring simultaneously or in sequence $P(AB)$

Conditional Probability – the probability of event A occurring, given that B has already occurred $P(A|B)$

Complement

The complement $(\overline{A})$ of an event A is the set of all sample space outcomes not in A

$$P(\overline{A}) = 1 - P(A)$$



Statistical Independence or Statistical Dependent

Events are *statistically independent* when the occurrence of one event makes it neither more nor less probable

For example, tossing an unbiased coin has the same probability of landing on tails each time, no matter what happened in the previous toss. Therefore, this is an example of a statistically independent condition.

Events are *statistically dependent* when a previous event affects the probability of a subsequent event or when two or more conditions occur simultaneously.

On the other hand, example of a statistically dependent condition is if you first toss a coin, if it lands on head, you select a ticket from a basket that contains 10 winning tickets from a total of 100 tickets; and if you toss tails, you select one ticket from a basket containing one winning ticket from a total of 100 tickets. Therefore, you can see your probability of selecting a winning ticket depends on what happens when you toss the coin.

Conditional Probability and Independence

The probability of an event A, given that the event B has occurred, is called the conditional probability of A given B

Denoted as $P(A|B)$

Further, $P(A|B) = P(A \cap B) / P(B)$

$$P(B) \neq 0$$

Bowerman, B. L., O'Connell, R. T., & Murphree, E. S., (2010). Business Statistics in Practice (6th Ed.), Copyright © McGraw-Hill Education

Examples

Which of the following appears most likely, secondly and least likely?

a. Drawing a red marble from a bag containing 50% red marbles and 50% white marbles.

Answer: $P(R)$ *= 50% This a marginal probability*

b. Drawing a red marble seven times in succession with replacement (a selected marble is put back in the bag before the next marble is selected), from the bag containing 90% red marbles and 10% white marbles.

Answer: $P(RRRRRRR)$ *= (.90)* ^{7} *= .48 This is joint probability*

c. Drawing a red marble from a bag containing 50% red marbles and 50% white marbles given last drawing you selected a red marble with replacement.

Answer: $P(R|R) = P(R)$ *= 50% This conditional probability under statistically independent conditions*

Independence of Events

Two events A and B are said to be independent if and only if:

$$P(A|B) = P(A)$$

This is equivalent to

$$P(B|A) = P(B)$$

Assumes $P(A)$ and $P(B)$ greater than zero

The Multiplication Rule

The joint probability that A and B (the intersection of A and B) will occur is:

$$P(A \cap B) = P(A) \cdot P(B|A) = P(B) \cdot P(A|B)$$

If A and B are **independent**, then the probability that A and B will occur is:

$$P(A \cap B) = P(A) \cdot P(B|A) = P(B) \cdot P(A)$$

Bowerman, B. L., O'Connell, R. T., & Murphree, E. S., (2010). Business Statistics in Practice (6th Ed.), Copyright © McGraw-Hill Education

Three Types of Probability and Notation

Types of Probability	Symbols	Conditions of Statistical Independence	Conditions of Statistical Dependence
Marginal	$P(A)$	$P(A)$	$P(A)$
Joint	$P(AB)$	$P(A) \times P(B)$	$P(A \mid B) \times P(B)$
Conditional	$P(A \mid B)$	$P(A)$	$\frac{P(AB)}{P(B)}$

Example

Frequency Table for Sales of Women's Sweaters

Size	Color			Total
	Red	Yellow	Blue	
Large	12	8	5	25
Medium	23	31	1	55
Small	6	6	8	20
Total	41	45	14	100

Statistically Dependent Example

What is the probability of the next customer at random purchasing ***a red sweater***?

Size	Color *Red*	Yellow	Blue	Total
Large	12	8	5	25
Medium	23	31	1	55
Small	6	6	8	20
Total	***41***	45	14	100

$P(R)$ Marginal probability = 41%

Statistically Dependent Example *continued*

What is the probability of the next customer at random purchasing ***size large sweater***?

Size	Color Red	Yellow	Blue	Total
Large	12	8	5	***25***
Medium	23	31	1	55
Small	6	6	8	20
Total	41	45	14	100

$P(L)$ Marginal probability = 25%

Statistically Dependent Example *continued*

What is the probability of the next customer at random purchasing ***size medium in yellow***?

Size	Color			Total
	Red	***Yellow***	Blue	
Large	12	8	5	25
Medium	23	***31***	1	55
Small	6	6	8	20
Total	41	45	14	100

$P(MY)$ Joint probability = 31%

Statistically Dependent Example *continued*

What is the probability of the next customer at random purchasing ***blue in size small***?

Size	Color Red	Yellow	***Blue***	Total
Large	12	8	5	25
Medium	23	31	1	55
Small	6	6	***8***	20
Total	41	45	14	100

$P(BS)$ Joint probability = 8%

Statistically Dependent Example *continued*

What is the probability of the next customer at random purchasing ***yellow sweater given size large***?

Size	Color Red	Color ***Yellow***	Color Blue	Total
Large	12	8	5	25
Medium	23	31	1	55
Small	6	6	8	20
Total	41	45	14	100

$P(Y|L)$ Conditional prob. = 8/25 = 32%

Statistically Dependent Example *continued*

What is the probability of the next customer at random purchasing ***Given that the sweater is blue, size is small***?

Size	Color Red	Yellow	***Blue***	Total
Large	12	8	5	25
Medium	23	31	1	55
Small	6	6	8	20
Total	41	45	14	100

$P(S|B)$ Conditional prob. = 8/14 = 57%

Example

City	Output	Defects
Baltimore		
Style A	40	10
Style B	20	5
Style C	7	6
Tampa		
Style A	12	5
Style B	15	10
Style C	10	2
St. Louis		
Style A	10	5
Style B	10	2
Style C	10	4

Output levels for Ace Electronics of three different monitor styles at plants in three different cities are shown to the left, along with the defect rates for each switch type.

Example *continued*

City	Output	Defects
Baltimore		
Style A	40	10
Style B	20	5
Style C	7	6
Tampa		
Style A	12	5
Style B	15	10
Style C	10	2
St. Louis		
Style A	10	5
Style B	10	2
Style C	10	4

If a monitor is chosen at random for inspection by Ace's quality control circle, what is the probability that it is

a. From Baltimore and defective?
b. From Tampa and defective?
c. Style A and defective?
d. Style B and not defective?
e. From St. Louis given it is defective?
f. From Tampa given it is not defective?

Example *continued*

City	Output	Defects	$\overline{\text{Defects}}$
Baltimore			
Style A	40	10	*30*
Style B	20	5	*15*
Style C	7	6	*1*
Subtotal	*67*	*21*	*46*
Tampa			
Style A	12	5	*7*
Style B	15	10	*5*
Style C	10	2	*8*
Subtotal	*37*	*17*	*20*
St. Louis			
Style A	10	5	*5*
Style B	10	2	*8*
Style C	10	4	*6*
Subtotal	*30*	*11*	*19*
Total	*134*	*49*	*85*

If a monitor is chosen at random for inspection by Ace's quality control circle, what is the probability that it is

a. From Baltimore and defective? *Joint* $\frac{21}{134}$

b. From Tampa and defective? *Joint* $\frac{17}{134}$

c. Style A and defective? *Joint* $\frac{20}{134}$

d. Style B and not defective? *Joint* $\frac{28}{134}$

e. From St. Louis given it is defective? *Conditional* $\frac{11}{49}$

f. From Tampa given it is not defective? *Conditional* $\frac{20}{85}$

Some Elementary Probability Rules

- **Union**
- **Intersection**
- **Addition**
- **Multiplication**

Union and Intersection

The **union** of A and B are elementary events that belong to either A or B or both

Written as: $A \bigcup B$

The **intersection** of A and B are elementary events that belong to both A and B

Written as: $A \bigcap B$

The Addition Rule

If A and B are *mutually exclusive*, then the probability that A or B (the union of A and B) will occur is:

$$P(A \bigcup B) = P(A) + P(B)$$

If A and B are *not mutually exclusive*:

$$P(A \bigcup B) = P(A) + P(B) - P(A \bigcap B)$$

where $P(A \bigcap B)$ is the **joint** probability of A and B both occurring together

The Addition Rules and the Multiplication Rule

The two rules of probability

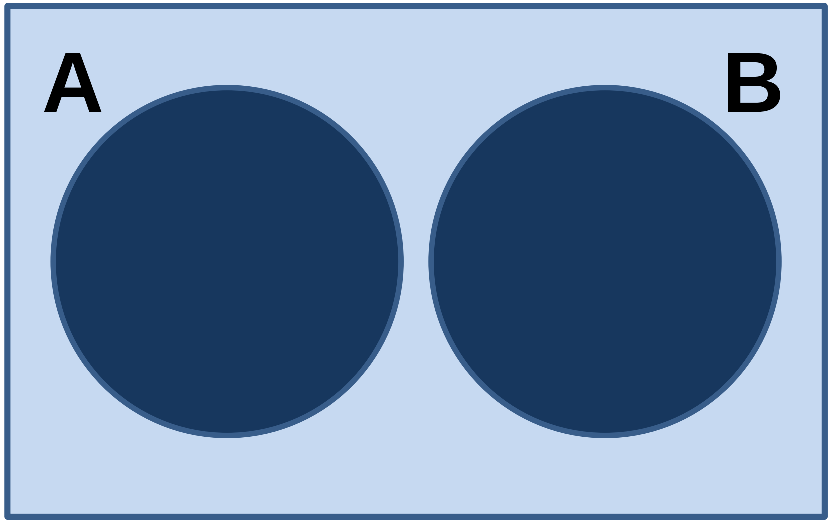
- **The addition rule $P(A \text{ or } B)$**
  - Are A and B mutually exclusive
    - No: $P(A \text{ or } B) = P(A) + P(B) - P(A \text{ and } B)$
    - Yes: $P(A \text{ or } B) = P(A) + P(B)$
- **The multiplication rule $P(A \text{ and } B)$**
  - Are A and B statistically dependent
    - No: $P(A \text{ and } B) = P(A) \times P(B)$
    - Yes: $P(A \text{ and } B) = P(A) \times P(B|A)$

Venn Diagram Mutually Exclusive Events

A and B are mutually exclusive if they have no sample space outcomes in common

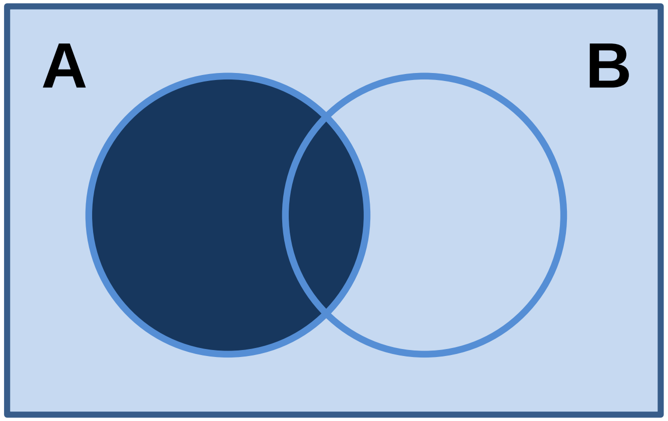
In other words:

$$P(A \bigcap B) = 0$$

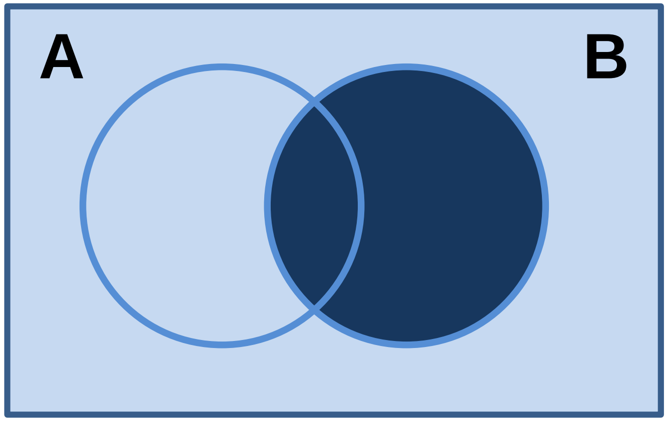


Venn Diagram Non-Mutually Exclusive Events

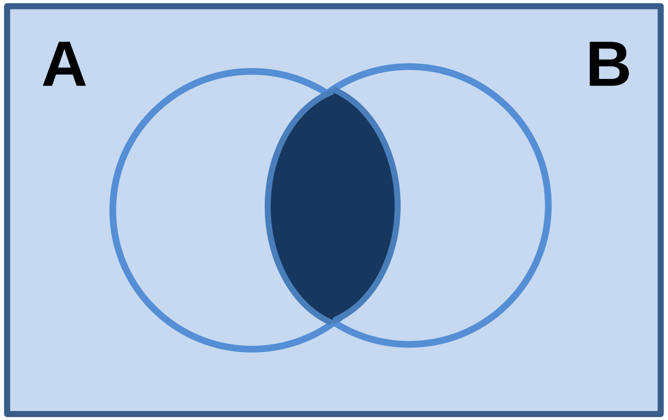
(a) The event *A* is the shaded region



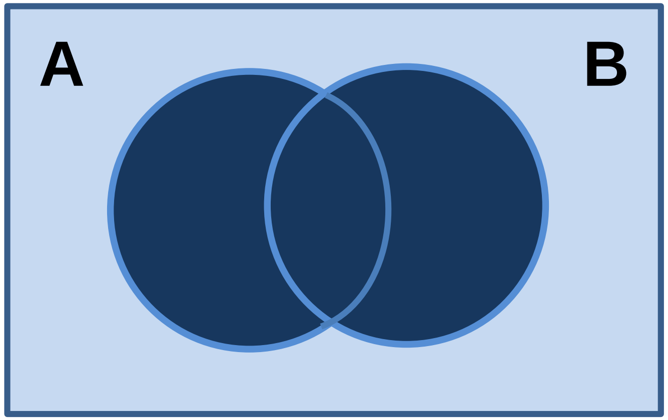
(b) The event *B* is the shaded region



(c) The event $A \cap B$ is the shaded region



(d) The event $A \cup B$ is the shaded region



Example

Dell Publishing has 75 different book titles classified by type and cost as follows:

Type	Cost		
	$10	$15	$20
Fiction	10	8	3
Biography	12	10	9
Historical	4	17	2

Example

Type	Cost		
	$10	$15	$20
Fiction	10	8	3
Biography	12	10	9
Historical	4	17	2

Find the probability that a book selected at random is

a. either fiction or costs $10
b. both historical and costs $20
c. historical and costs either $10 or $15
d. is fiction and costs less than $20
e. is biographical or costs $15
f. is historical or costs $10

Answers

Type	Cost			Total
	$10	$15	$20	
Fiction	10	8	3	21
Biography	12	10	9	31
Historical	4	17	2	23
Total	26	35	14	75

Find the probability that a book selected at random is

a. either fiction or costs $10

$$P(F\ or\ \$10) = \frac{21}{75} + \frac{26}{75} - \frac{10}{75} = \frac{37}{75}$$

b. both historical and costs $20

$$P(H\ \$20) = \frac{2}{75}$$

Answers

Type	Cost			Total
	$10	$15	$20	
Fiction	10	8	3	21
Biography	12	10	9	31
Historical	4	17	2	23
Total	26	35	14	75

Find the probability that a book selected at random is

c. historical and costs either $10 or $15

$$P(H\$10)\ or\ P(H\$15) = \frac{4}{75} + \frac{17}{75} = \frac{21}{75}$$

d. is fiction and costs less than $20

$$P(F\$10)\ or\ P(F\$15) = \frac{10}{75} + \frac{8}{75} = \frac{18}{75}$$

Answers

Type	Cost			Total
	$10	$15	$20	
Fiction	10	8	3	21
Biography	12	10	9	31
Historical	4	17	2	23
Total	26	35	14	75

Find the probability that a book selected at random is

e. biographical or costs $15

$$P(B\ or\ \$15) = \frac{31}{75} + \frac{35}{75} - \frac{10}{75} = \frac{56}{75}$$

f. historical or costs $10

$$P(H\ or\ \$10) = \frac{23}{75} + \frac{26}{75} - \frac{4}{75} = \frac{45}{75}$$

Bayes' Theorem

$$P(S_i|E) = \frac{P(S_i \cap E)}{P(E)} \qquad = \frac{P(S_i)P(E|S_i)}{P(E)}$$

$$= \frac{P(S_i)P(E|S_i)}{P(S_1)P(E|S_1) + P(S_2)P(E|S_2) + \ldots + P(S_k)P(E|S_k)}$$

Using this principle of conditional probability as a focal point, the Reverend Thomas Bayes (1702-1761) reasoned that, given $P(A) and P(B|A)$, it was possible to determine $P(A|B)$

Example

To illustrate, Jack and Jill sell insurance in the family business. Jack sells 80 percent of the policies, and Jill sells the rest. Ten percent of the policies Jack sells have a Claim filed within one year, compared to 25 percent of those sold by Jill. To summarize:

$$P(Jack) = 0.80$$

$$P(Jill) = 0.20$$

$$P(Claim \mid Jack) = 0.10$$

$$P(Claim \mid Jill) = 0.25$$

A client announces his intention to file a claim. What is the probability Jack sold him the policy? That is, find $P(Jack \mid Claim)$. According to the formula for conditional probability,

$$P(Jack \mid Claim) = \frac{P(Jack\ and\ Claim)}{P(Claim)}$$

Answer

	Jack	Jill	
Claim	**.08**	**.05**	.13
No Claim	.72	.15	.87
	.80	**.20**	**1.0**

$$P(\text{Claim} \mid \text{Jack}) = \frac{P(\text{Jack and Claim})}{P(\text{Jack})}$$

$$P(\text{Jack and Claim}) = P(\text{Claim} \mid \text{Jack})(P(\text{Jack}))$$

joint probability *(.10)(.80)=.08*

Answer *continued*

	Jack	Jill	
Claim	**.08**	**.05**	.13
No Claim	.72	.15	.87
	.80	**.20**	**1.0**

Or

$$P(Jack|Claim) = \frac{P(Jack)\times P(Claim|Jack)}{P(Jack)\times P(Claim|Jack) + P(Jill)\times P(Claim|Jill)}$$

$$= \frac{(.80)(.10)}{(.80)(.10)+(.20)(.25)}$$

$$= \frac{.08}{.13} = .615 \approx 62\%$$

Posterior Probability/Bayes’ Theorem

The probability measure that has been revised on the condition that some known event has occurred is called posterior probability.

Posterior Probability/Bayes' Theorem

The Occupational Health and Safety Administration (OSHA) has found that 15% of all employees working at construction sites are injured on the job. OSHA has also determined that 20% of all construction workers are on jobs classified as dangerous. Accident reports filed by construction firms show that 10% of all employees are on dangerous jobs and are injured.

This is not to say that 10% of all employees on dangerous jobs are injured. It says that of all employees, 10% are on dangerous jobs and are injured.

Counting Techniques

Combinations – arrangements in which order does not make a difference and duplication is not allowed.

Formula used to calculate the possible sample combinations when employing random sampling.

$$_{n}C_{x} = \frac{n!}{x!(n-x)!}$$

Permutations – arrangement in which order makes a difference.

Use Permutations to calculate the number of letters and numbers needed to establish passwords, license numbers, addresses, etc.

$$_{n}P_{x} = \frac{n!}{(n-x)!}$$

Example *combinations*

How many ways to form a committee of 3 among 8 people (A, B, C, D, E, F, G, H)? Committees ABC, ACB, BAC, are the same.

$$_8C_3 = \frac{8!}{3!(8-3)!} = \frac{8!}{3!(5!)} = \frac{8\times 7\times 6\times 5!}{3\times 2\times 1(5!)} = 8\times 7 = 56$$

Example *permutations*

Number of ways to award 3 prizes among 8 participants?

Remember counting rule: 8* 7* 6 = 336

8* 7* 6 = (8 * 7* 6* 5* 4* 3* 2* 1*) / (5* 4* 3* 2* 1) = 8! / (8-3)!

Permutations of n items from N: $P_{N.n} = N!/(N-n)!$

Example *permutations*

In permutations ordering is important. Ways to award 3 prizes (1^{st}, 2^{nd} and 3^{rd}) among 8 participants?

(Sam, Joe, Ray) is different from (Joe, Ray, Sam).

When the ordering is not important, we use combinations

Example *permutations*

The CEO of NanoSOFT must select five people from a list of 15 young executives to serve as examples of outstanding managerial talent. Each executive is to receive a monetary reward. The first one selected will get the highest bonus, the second one the second highest, and so on.

$$ {}_{15}P_5 = \frac{15!}{(15-5)!} = \frac{15 \times 14 \times 13 \times 12 \times 11 \times 10!}{10!} = 360{,}360 $$