2101

**Basic Probability, Including Contingency Tables[[1]](#footnote-1)©**

**Probability and Probability Experiments**

 A **probability experiment** is a well-defined act or process that leads to a single well defined outcome. Example: toss a coin (H or T), roll a die (1,2,3,4,5,6), measure your height (X cm where X is greater than 0).

 The **probability** of an event, **P(A)** is the fraction of times that event will occur in an indefinitely long series of trials of the experiment. This may be estimated:

 1. **Empirically**: conduct the experiment many times and compute , the sample relative frequency of A. Roll die 1000 times, even numbers appear 510 times, P(even) = 510/1000 = .51 or 51%.

 2. **Rationally or Analytically**: make certain assumptions about the probabilities of the elementary events included in outcome A and compute probability by rules of probability. Assume each event 1,2,3,4,5,6 on die is equally likely. The sum of the probabilities of all possible events must equal one. Then P(1) = P(2) = P(3) = P(4) = P(5) = P(6) = 1/6. P (even) = P(2) + P(4) + P(6) = 1/6 + 1/6 + 1/6 = 1/2 (addition rule) or 50%.

 3. **Subjectively**: a measure of an individual’s degree of belief assigned to a given event in whatever manner. I think that the probability that ECU will win its opening game of the season is 1/3 or 33%. This means I would accept 2:1 **ODDS** against ECU as a fair bet (if I bet $1 on ECU and they win, I get $2 in winnings).

**Independence, Mutual Exclusion, and Mutual Exhaustion**

 Two events are **independent** iff (if and only if) the occurrence or non-occurrence of the one has no effect on the occurrence or non-occurrence of the other.

 Two events are **mutually exclusive** iff the occurrence of the one precludes occurrence of the other (both cannot occur simultaneously on any one trial).

 Two (or more) events are **mutually exhaustive** iff they include all possible outcomes.

**Marginal, Conditional, and Joint Probabilities**

 The **marginal probability** of event A, **P(A)**, is the probability of A ignoring whether or not any other event has also occurred.

 The **conditional probability of A given B, P(A|B)**, is the probability that A will occur given that B has occurred. **If A and B are independent, then P(A) = P(A|B).**

 A **joint probability** is the probability that both of two events will occur simultaneously.

**The Multiplication Rule**

 If two probabilities are **independent**, their joint probability is the product of their marginal probabilities,

**P(A ∩ B) = P(A) \* P(B)**.

 Regardless of the independence or nonindependence of A and B, the joint probability of A and B is: **P(A ∩ B) = P(A) \* P(B|A) = P(B) \* P(A|B).**

**The Addition Rule**

 If A and B are **mutually exclusive**, the probability that either A or B will occur is the sum of their marginal probabilities, **P(A ∪ B) = P(A) + P(B)**.

Regardless of their exclusivity, **P(A ∪ B) = P(A) + P(B) - P(A ∩ B)**

**Probabilities and Odds**

**Odds to Probabilities**

 If the odds of some event occurring are A to B, then the probability of that event occurring is A divided by (A + B). For example, if you think the odds of ECU winning a particular football game are 3 to 1. That means that the probability of ECU winning is three times the probability of ECU not winning. To convert the odds of 3 to 1 to a probability divide the 3 by (3 + 1) to get 3/4 or 75%. The probability of ECU not winning would then be 25%. 75% is three times 25%. Suppose a fan of the opposing team opines that the odds of ECU winning are 2 to 9 (2/9 = 22.2%). The probability would then be 2 divided by (2+9) = 18.2%. The probability of ECU not winning would be 81.8. 18.2 is only (18.2/81.8) = 22% of 81.8%.

**Probabilities to Odds**

 If the probability of an event happening is A%, then the odds of A happening are A/(100-A). For example, if you think the probability of precipitation tomorrow is 65%, then the odds of precipitation are 65/35 = 13 to 7 = 1.857

**Working with Contingency Tables**

* See [Contingency Tables](http://core.ecu.edu/psyc/wuenschk/PP/Contingency-2101.ppt) -- slide show

 To determine whether two categorical variables are correlated with one another, we can use a two-dimensional table of frequencies, often called a contingency table. For example, suppose that we have asked each of 150 female college students two questions: 1. Do you smoke (yes/no), and, 2. Do you have sleep disturbances (yes/no). Suppose that we obtain the following data (these are totally contrived, not real):

# **Data From Independent Variables**

|  |  |  |
| --- | --- | --- |
|  | Sleep? |  |
| Smoke? | No | Yes |  |
| No | 20 | 30 | 50 |
| Yes | 40 | 60 | 100 |
|  | 60 | 90 | 150 |

# **Marginal Probabilities**



# **Conditional Probabilities**



 Notice that the conditional probability that the student has sleeping disturbances is the same if she is a smoker as it is if she is not a smoker. Knowing the student’s smoking status does not alter our estimate of the probability that she has sleeping disturbances. That is, for these contrived data, smoking and sleeping are independent, not correlated.

# **Multiplication Rule**

 The probability of the joint occurrence of two independent events is equal to the product of the events’ marginal probabilities.



# **Addition Rule**

 Suppose that the probability distribution for final grades in PSYC 2101 were as follows:

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| Grade | A | B | C | D | F |
| Probability | .2 | .3 | .3 | .15 | .05 |

 The probability that a randomly selected student would get an A or a B, , since A and B are mutually exclusive events.

 Now, consider our contingency table, and suppose that the “sleep” question was not about having sleep disturbances, but rather about “sleeping” with men (that is, being sexually active). Suppose that a fundamentalist preacher has told you that women who smoke go to Hades, and women who “sleep” go there too. What is the probability that a randomly selected woman from our sample is headed to Hades? If we were to apply the addition rule as we did earlier, , but a probability cannot exceed 1, something is wrong here.

 The problem is that the events (sleeping and smoking) are not mutually exclusive, so we have counted the overlap between sleeping and smoking (the 60 women who do both) twice. We need to subtract out that double counting. If we look back at the cell counts, we see that 30 + 40 + 60 = 130 of the women sleep and/or smoke, so the probability we seek must be 130/150 = 13/15 = .87. Using the more general form of the addition rule, 

# **Data From Correlated Variables**

Now, suppose that the “smoke” question concerned marijuana use, and the “sleep” question concerned sexual activity, variables known to be related.

|  |  |  |
| --- | --- | --- |
|  | Sleep? |  |
| Smoke? | No | Yes |  |
| No | 30 | 20 | 50 |
| Yes | 40 | 60 | 100 |
|  | 70 | 80 | 150 |

# **Marginal Probabilities**



# **Conditional Probabilities**



 Now our estimate of the probability that a randomly selected student “sleeps” depends on what we know about her smoking behavior. If we know nothing about her smoking behavior, our estimate is about 53%. If we know she smokes, our estimate is 60%. If we know she does not smoke, our estimate is 40%. We conclude that the two variables are correlated, that female students who smoke marijuana are more likely to be sexually active than are those who do not smoke.

# **Multiplication Rule**

 If we attempt to apply the multiplication rule to obtain the probability that a randomly selected student both sleeps and smokes, using the same method we employed with independent variables, we obtain: . This answer is, however, incorrect. Sixty of 150 students are smoking sleepers, so we should have obtained a probability of 6/15 = .40. The fact that the simple form of the multiplication rule (the one which assumes independence) did not produce the correct solution shows us that the two variables are not independent.

 If we apply the more general form of the multiplication rule, the one which does not assume independence, we get the correct solution: 

## Real Data

 Finally, here is an example using data obtained by Castellow, Wuensch, and Moore (1990, *Journal of Social Behavior and Personality*, 5, 547-562). We manipulated the physical attractiveness of the plaintiff and the defendant in a mock trial. The plaintiff was a young women suing her male boss for sexual harassment. Our earlier research had indicated that physical attractiveness is an asset for defendants in criminal trials (juries treat physically attractive defendants better than physically unattractive defendants), and we expected physical attractiveness to be an asset in civil cases as well. Here are the data relevant to the effect of the attractiveness of the plaintiff.

|  |  |  |
| --- | --- | --- |
|  | Guilty? |  |
| Attractive? | No | Yes |  |
| No | 33 | 39 | 72 |
| Yes | 17 | 56 | 73 |
|  | 50 | 95 | 145 |

 Guilty verdicts (finding in favor of the plaintiff) were more likely when the plaintiff was physically attractive (56/73 = 77%) than when she was not physically attractive (39/72 = 54%). The magnitude of the effect of physical attractiveness can be obtained by computing an **odds ratio**. When the plaintiff was physically attractive, the odds of a guilty verdict were 56 to 17, that is, 56/17 = 3.29. That is, a guilty verdict was more than three times more likely than a not guilty verdict. When the plaintiff was not physically attractive the odds of a guilty verdict were much less, 39 to 33, that is, 1.18. The ratio of these two odds is  That is, the odds of a guilty verdict when the plaintiff was attractive were almost three times higher than when the plaintiff was not attractive. That is a big effect!

 We also found that physical attractiveness was an asset to the defendant. Here are the data:

|  |  |  |
| --- | --- | --- |
|  | Guilty? |  |
| Attractive? | No | Yes |  |
| No | 17 | 53 | 70 |
| Yes | 33 | 42 | 75 |
|  | 50 | 95 | 145 |

 Guilty verdicts (finding in favor of the plaintiff) were less likely when the defendant was physically attractive (42/75 = 56%) than when he was not physically attractive (53/70 = 76%). The odds ratio here is  We could form a ratio of probabilities rather than odds. The ratio of [the probability of a guilty verdict given that the defendant was not physically attractive] to [the probability of a guilty verdict given that the defendant was physically attractive] is 

 We could look at these data from the perspective of the odds of a not guilty verdict. Not guilty verdicts were more likely when the defendant was physically attractive (33/75 = 44%) than when he was not physically attractive (17/70 = 24%). The odds ratio here is  Notice that the odds ratio is the same regardless of which perspective we take. This is not true of probability ratios (and is why I much prefer odds ratios over probability ratios). The ratio of [the probability of a not guilty verdict given the defendant is attractive] to [the probability of a not guilty verdict given that the defendant is not attractive] is  With probability ratios the size of the ratio depends on whether you compare the probability of [A given B] to the probability of [A given not B], or, alternatively, compare the probability of [not A given B] to the probability of [not A given not B]. See also <http://core.ecu.edu/psyc/wuenschk/StatHelp/OddsRatios.htm> .

**Probability Distributions**

 The **probability distribution** of a **discrete** variable Y is the pairing of each value of Y with one and only one probability. The pairing may be by a listing, a graph, or some other specification of a functional rule, such as a formula. Every P(Y) must be between 0 and 1 inclusive, and the sum of all the P(Y)s must equal 1.

 For a **continuous variable** we work with a probability density function, defined by a formula or a figure (such as the normal curve).

a. Imagine a relative frequency histogram with data grouped into 5 class intervals so there are 5 bars. Now increase the number of intervals to 10, then to 100, then 100,000, then an infinitely large number of intervals—now you have a smooth curve (see Howell’s *Fundamental Statistics for the Behavioral Sciences*, page 149-151, 7th ed.), the probability function. See my slide show at <http://core.ecu.edu/psyc/wuenschk/PP/Discrete2Continuous.ppt> .

b. There is an infinitely large number of points on the curve, so the probability of any one point is infinitely small.

c. We can obtain the probability of a score falling in any particular interval a to b by setting the total area under the curve equal to 1 and measuring the area under the curve between a and b. This involves finding the definite integral of the probability density function from a to b.

d. To avoid doing the calculus, we can, for some probability density functions (such as the normal), rely on tables or computer programs to compute the area.

e. Although technically a continuous variable is one where there is an infinite number of possible intermediate values between any pair of values, if a discrete variable can take on many possible values and we think it reasonable to consider the underlying dimension to be continuous, then we shall treat the variable as continuous.

**Random Sampling**

 Sampling *N* data points from a population is random if every possible different sample of size *N* was equally likely to be selected.

1. We want our samples to be representative of the population to which we are making inferences. With moderately large random samples, we can be moderately confident that our sample is representative.

2. The inferential statistics we shall use assume that random sampling is employed.

3. In fact, we rarely if ever achieve truly random sampling, but we try to get as close to it as is reasonable.

4. This definition differs from that given in Howell (page 8, 8th ed.). A sampling procedure may meet Howell’s definition but not mine. For example, sampling from a population of 4 objects (A,B,C,& D) without replacement, *N* = 2, contrast sampling procedure X with Y:

|  |  |
| --- | --- |
|  | Probability |
| Sample | X | Y |
| AB | 1/2 | 1/6 |
| AC | 0 | 1/6 |
| AD | 0 | 1/6 |
| BC | 0 | 1/6 |
| BD | 0 | 1/6 |
| CD | 1/2 | 1/6 |

 With both procedures (X and Y), it is true that every member of the population is equally likely to be sampled (*P* = .5 for each of A, B, C, and D), but for procedure X it is not true that each possible sample is equally likely to be obtained. Procedure X is not random sampling, procedure Y is.

* [Odds to Probability Calculator](https://www.calculatorsoup.com/calculators/games/odds.php)
* [Probability to Odds Calculator](https://www.calculatorsoup.com/calculators/games/odds.php) – enter the probability as a proportion rather than as a percentage – for example, .25 instead of 25%.
* [Probability FAQ](http://core.ecu.edu/psyc/WuenschK/docs01/Probability-FAQ.doc) – Answers to frequently asked questions.

# Copyright 2022, Karl L. Wuensch - All rights reserved

1. © Copyright 2022, Karl L. Wuensch - All rights reserved [↑](#footnote-ref-1)