Lecture 11

Nancy Pfenning Stats 1000

Chapter 6: Relationships Between Categorical Variables

Some variables are categorical by nature (eg. sex, race, major, political party); others are created by grouping quantitative variables into classes (eg. age as child, adolescent, young adult, etc.). We analyze categorical data by recording counts or percents of cases occurring in each category. In this course, we mainly consider just two categorical variables at a time.

Example

We can construct a two-way table (also called a contingency table) showing the relationship between gender (row variable) and lenswear (column variable) of statistics class members during a previous semester.

	None	Glasses	Contacts	Total
Male	65	36	37	138
Female	110	32	91	233
Total	175	68	128	371

One natural question to ask is, which group tends to wear glasses more, males or females? The *counts* are quite close for males and females (36 vs. 32) but since there are relatively few males in the class, the *percentage* is actually much higher for the males. Thus, to compare lenswear of males and females, since there are way more females in the classes than males, it is better to report percentages instead of counts, concentrating on one gender category at a time. This tells the **conditional distribution** of lenswear, given gender. It suggests we are thinking of gender as the explanatory variable.

	NONE	GLASSES	CONTACTS	TOTAL
Cond.dist.of lens, given male	$\frac{65}{138} = 47\%$	$\frac{36}{138} = 26\%$	$\frac{37}{138} = 27\%$	100%
Cond.dist.of lens, given female	$\frac{110}{233} = 47\%$	$\frac{32}{233} = 14\%$	$\frac{91}{233} = 39\%$	100%

To use a bar graph to display a conditional distribution, label the horizontal axis with the explanatory variable—in this case, males and females. Over the male label would be bars of height 47%, 26%, and 27%, for percentages wearing none, contacts, or glasses. Over the female label would be bars of height 49%, 14%, and 39%, for percentages wearing none, contacts, or glasses. This impresses on us visually the difference between males and females: males have a tendency to wear glasses, females to wear contacts.

Alternatively, (but perhaps less intuitively) we may choose to think of lenswear as the explanatory variable and consider the distribution of gender separately for each lens category:

Lenswear	Cond.Dist.of Gender, given none	Cond.Dist.of Gender, given glasses	Cond.Dist.of Gender, given contacts
Male	$\frac{65}{175} = 37\%$	$\frac{36}{68} = 53\%$	$\frac{37}{128} = 29\%$
Female	$\frac{110}{175} = 63\%$	$\frac{32}{68} = 47\%$	$\frac{91}{128} = 71\%$
TOTAL	100%	100%	100%

We could display this distribution by labeling three lenswear categories on the horizontal axis, and drawing bars with heights to represent the percentages of males and females in each.

Example

Of the male students in the classes, what percentage wear glasses? The *condition* of being male is given; we look at the conditional distribution of lenswear, given that a person is male, and find the percentage wearing glasses to be 26%.

Example

Of the students who wear glasses, what percentage are males? We look at the conditional distribution of gender, given that a student wears glasses, and find the percentage to be 53%.

Example

Who has a higher percentage wearing glasses, males or females? The males: 26% vs. 14%.

Example

Suppose a student left behind a pair of glasses on a desk; was the student more likely to be male or female? In other words, of the students who wear glasses, are a higher percentage male or female? A slightly higher percentage are male: 53% male vs. 47% female.

Example

Of all the students, what percentage are males wearing glasses? For this, we need our original two-way table of counts: 36 out of 371 $\approx$ 10%.

Simpson's Paradox

Example

(Hypothetical) Suppose sugar/hyperactivity observational studies have been conducted separately for boys and girls, and then combined:

Boys	Normal	Hyper	Rate of Hyperactivity
Low Sugar	25	50	$\frac{50}{75} = .67$
High Sugar	50	100	$\frac{100}{150} = .67$

Girls	Normal	Hyper	Rate of Hyperactivity
Low Sugar	75	25	$\frac{25}{100} = .25$
High Sugar	25	8	$\frac{8}{33} = .25$

Combined	Normal	Hyper	Rate of Hyperactivity
Low Sugar	100	75	$\frac{75}{175} = .43$
High Sugar	75	108	$\frac{108}{183} = .59$

Note that the rate of hyperactivity for boys is the same, .67, whether their sugar intake is high or low: what we observe is exactly what we'd expect if there were no relationship between sugar intake and activity level. Likewise for girls, the rate of hyperactivity is .25, regardless of sugar intake. However, strong evidence of a relationship appears to exist when the two groups are combined: the hyperactive rates are .43 for low sugar intake vs. .59 for high sugar intake. This phenomenon is known as **Simpson's Paradox**, which describes the apparent change in a relationship in a two-way table when groups are combined. In our hypothetical example, boys tended to consume more sugar than girls, and also tended to be more hyperactive than girls. This results in the apparent relationship in the combined table. The confounding variable, gender, should be controlled for by studying boys and girls separately, not ignored by combining.

In general, relationships between two categorical variables should always be scrutinized for possible confounding variables.

Example

An article entitled **Virginity pledge works** reports that "teens who promised to refrain until marriage delayed having sex about 18 months longer than others". Is there a better way to explain this than to say that the "virginity pledge works"?

Lecture 12

Risk and Odds

For any situation with two categorical variables, we can construct a two-way table and compare rates for the various values of the explanatory variable. When the response variable is for the occurrence of something like illness or death, each rate may be referred to as a **risk**, and the ratio of rates is the **relative risk**. **Increased risk** represents a relative change. **Odds** give a ratio of occurrence vs. non-occurrence using whole numbers.

Example

We have already looked at a two-way table for crimes in Maryland, classified according to race of the victim and whether or not the defendent was given the death penalty:

	Death Penalty	No Death Penalty	Total	Death Penalty Rate
Black Victim	15	675	690	$\frac{15}{690} = 2.2\%$
White Victim	61	560	621	$\frac{61}{621} = 9.8\%$

1. The **percentage** sentenced to death is 2.2% in the case of a black victim, vs. 9.8% in the case of a white victim.
2. The **proportion** sentenced to death in the case of a white victim is .098.
3. The **probability** of being sentenced to death in the case of a white victim is .098.
4. The **risk** of being sentenced to death was .022 when the victim was black and .098 when the victim was white.
5. The **relative risk** of incurring the death penalty for white victims over black victims is $\frac{.098}{.022} = 4.5$: The death penalty is more than four times likelier when the victim was white.
6. **odds** of being sentenced to death when the victim was white are 61 to 560, or about 1 to 9.

Since

$$\text{increased risk} = (\text{relative risk} - 1) * 100\%$$

the percent increase in risk from black victim to white is $(4.5 - 1) * 100\% = 350\%$.

The following example explores the *statistical significance* of the difference observed for black vs. white victims.

Example

The difference in death penalty rates for black vs. white victims, 2.2% vs. 9.8%, does seem substantial, but someone may claim that it could easily enough have come about by chance. What if the difference were 4% vs. 8%? 5% vs. 7%? How large does the difference have to be in order to convince us that it's not just chance variation? [Something else to think about: would a difference of 2.2% vs. 9.8% out of just 45 crimes (1 vs. 4 death penalties) be as convincing as in the case of 1311 crimes (15 vs. 61 death penalties)?] Based on the information in the two-way table above, can we produce statistical evidence of a relationship between race of the victim and whether or not the death penalty was imposed?

To answer this question, we compute a number which measures overall how far the *observed* numbers in the 2×2 table are from the numbers *expected* if there were no relationship.

1. Thus, the first step is to compute these expected numbers, beginning with the observed totals in various categories:

Observed	Death Penalty	No Death Penalty	Total
Black Victim	15	675	690
White Victim	61	560	621
Total	76	1235	1311

   Altogether, $\frac{76}{1311}$ got the death penalty. If there were no relationship, we'd expect $\frac{76}{1311}$ of the 690 cases involving black victims to get the death penalty, i.e. $\frac{76*690}{1311} = 40$,
   and $\frac{76}{1311}$ of the 621 cases involving white victims to get the death penalty, i.e. $\frac{76*621}{1311} = 36$,
   Also, $\frac{1235}{1311}$ didn't get the death penalty. If there were no relationship, we'd expect $\frac{1235}{1311}$ of the 690 cases involving black victims not to get the death penalty, i.e. $\frac{1235*690}{1311} = 650$,
   and $\frac{1235}{1311}$ of the 621 cases involving white victims not to get the death penalty, i.e. $\frac{1235*621}{1311} = 585$.
   Notice the pattern: each

$$\text{expected number} = \frac{\text{column total} * \text{row total}}{\text{table total}}$$

Expected	Death Penalty	No Death Penalty
Black Victim	$\frac{76*690}{1311} = 40$	$\frac{1235*690}{1311} = 650$
White Victim	$\frac{76*621}{1311} = 36$	$\frac{1235*621}{1311} = 585$

2. To compare what was observed with what would be expected, we look at each

$$\frac{(observed - expected)^2}{expected}$$

Compare	Death Penalty	No Death Penalty
Black Victim	$\frac{(15-40)^2}{40} = 15.6$	$\frac{(675-650)^2}{650} = 1.0$
White Victim	$\frac{(61-36)^2}{36} = 17.4$	$\frac{(560-585)^2}{585} = 1.1$

   For both black and white victims, there is a big difference in the numbers incurring the death penalty between what we observed and what we'd expect if there were no relationship: for black victims, the large comparison number 15.6 tells us that 15 is much less than the expected 40; for white victims, the large comparison number 17.4 tells us that 61 is much more than the expected 36.

3. Taken together, how much of a difference is there between observed and expected numbers? Calculate

$$\text{chi-square} = \text{sum of all} \frac{(observed - expected)^2}{expected} = 15.6 + 17.4 + 1.0 + 1.1 = 35.1$$

4. Recall: We learned in Chapter 2 that for a bell-shaped curve, 95% of the values fall within 2 standard deviations of the mean. If we consider the z-scores for values of a variable, where $z = \frac{\text{observed value}-\text{mean}}{\text{standard deviation}}$, it followed that about 95% of values have z-scores between -2 and +2. Shifting our attention to the tails of the curve, we can say that only 5% of the time, a z-score will take a value greater than 2 in absolute value. It is common practice to think of any z-score beyond a magnitude of 2 as being unusual.

Like z-scores, the distribution of chi-square values (chi-square = sum of all $\frac{(observed-expected)^2}{expected}$) follows a regular pattern, assuming there is no relationship between the row and column variables. It is also somewhat bell-shaped like z, but unlike z it has a longer right tail (is right-skewed) and by construction only takes positive values (its curve is entirely to the right of zero instead of being centered at zero). What would be considered an "unusual" chi-square value? It depends on the number of terms in our chi-square sum, which depends on table size. For a simple 2 by 2 table, as in our capital punishment example, it turns out that only 5% of the time, a chi-square will take a value greater than +3.84. Thus, a value of 2 would not be considered unusual; a value of 3.84 would be considered unusual; a value of 10 would be even more unusual, and a value of 35.1 would be practically impossible if there were no relationship between race of victim and whether or not the death penalty is imposed. We can quantify exactly how unusual any given chi-square is by reporting the **p-value**, which is the probability of chi-square taking a value as high as the one observed if there were no relationship between row and column variables. Because 35.1 is so far out on the fringe of the chi-square curve, it's practically impossible to see this large a difference between what we observed and what we'd expect if the row and column variables were not related: the p-value is very close to zero. This constitutes strong statistical evidence against the claim that race of victim and imposition of the death penalty are not related. Thus, we have evidence of racial discrimination in the sentencing process.

Example

An article entitled **Newer, costly schizophrenia drug gets low marks in study** reports: A drug that has become one of the first-line treatments for schizophrenia since the mid-1990s is not much better than older and cheaper medication, a surprising new study found.

The study was paid for by Eli Lilly & Co., maker of the newer drug, olanzapine, sold as Zyprexa.

Earlier, shorter studies showed that it was less likely to cause the tremors associated with older drugs such as haloperidol. But some previous research compared Zyprexa against only haloperidol, which typically is combined with another drug, benztropine, to reduce the risk of tremors.

The latest study, conducted for a year at 17 veterans hospitals, tested Zyprexa against the two-drug combination and found that Zyprexa patients fared only slightly better on scores of restlessness and mental function but had about the same degree of tremors.

Zyprexa costs more than $8 a day versus about 10 cents a day for the two-drug combination.

In the study, Zyprexa patients had $3,000 to $9,000 more in yearly expenses—mostly because of higher drug costs and more hospital stays—and the drug caused substantial weight gain, a known side effect.

Doctors and patients should consider those disadvantages when selecting treatment, instead of assuming Zyprexa is superior, said Dr. Robert Rosenheck, the study's lead author and director of the Veterans Affairs Department's Northeast Program Evaluation Center in West Haven, Conn.

"I had read the literature and believed that this drug would do better than haloperidol, so I was very surprised," Rosenheck said. The findings show "we may not be getting as much out of this as we thought we were."

The study appears in today's Journal of the American Medical Association. [November 2003]

Americans spend about $2 billion annually on Zyprexa.

The study involved 309 mostly male veterans with schizophrenia, a mental illness that affects about 2.2 million Americans. Patients took either daily doses of Zyprexa or the haloperidol-benztropine combination for a year.

About 6 percent of Zyprexa patients had moderate or marked restlessness compared with about 9.6 percent of the double-drug group. Zyprexa patients also had slightly better mental function, but there was no difference in tremors between the two groups.

At 12 months, nearly 25 percent of Zyprexa patients reported weight gain, compared with about 8 percent of the other group.

Dr. Alan Breier, Lilly's chief medical officer, blamed the disappointing results on the study's design. He said not enough patients were recruited and that they were sicker than typical schizophrenics.

Let's explore the relative impact of Zyprexa and the double-drug on restlessness first, and then on weight gain.

If the 309 patients were divided as evenly as possible, into groups of 154 and 155, we couldn't quite arrive at the quoted percentages having restlessness, namely 6% and 9.6%, so the actual group sizes must have been a bit uneven. 6% of 154 is about 9, and 9.6% of 155 is about 15. After establishing that treatment with Zyprexa or the double-drug plays the role of explanatory variable, across the rows, and restlessness or not plays the role of response, across the columns, we can complete a two-way table as follows:

Observed	Restless	Not Restless	Total
Zyprexa	9	145	154
Double-drug	15	140	155
Total	24	285	309

Fewer Zyprexa patients experienced restlessness, but is the difference between 6% and 9.6% statistically significant? To answer this question, we first contruct a table of counts expected if rate of restlessness were the same for both groups:

Expected	Restless	Not Restless	Total
Zyprexa	12	142	154
Double-drug	12	143	155
Total	24	285	309

Then we calculate chi-square to be $.75 + .75 + .06 + .06 = 1.62$, which is not greater than 3.84, so the difference is *not* statistically significant: there could easily enough happen by chance to be a difference in restlessness rates of this magnitude between the two groups even if it makes no difference whether patients take Zyprexa or the double-drug.

Another observed difference between the two groups was the percentage reporting weight gain: 25% for Zyprexa vs. 8% for double-drug. Again we can construct a two-way table based on this information, and a table of expected counts:

Observed	Weight Gain	No Weight Gain	Total
Zyprexa	39	115	154
Double-drug	12	143	155
Total	51	258	309

Expected	Weight Gain	No Weight Gain	Total
Zyprexa	25	129	154
Double-drug	26	129	155
Total	51	258	309

Now the chi-square statistic is $7.84 + 7.54 + 1.52 + 1.52 = 18.42$, which is much larger than 3.84, indicating that the difference in percentages reporting weight gain *is* statistically significant. The article accurately conveys the information that Zyprexa is not convincingly better in terms of restlessness, and that it does relatively badly in terms of weight gain. Dr. Breier's criticism reported in the last paragraph, on the other hand, is ill-founded, since there were plenty enough patients to show that Zyprexa caused weight gain. It seems doubtful that this difference wouldn't arise for schizophrenics who were less sick.

Exercise: Read **Boys spur marriage** (in the packet of articles) and complete a two-way table for this study consistent with all the numbers reported. Assume that the 600 children are equally divided between boys and girls, and assume that half of the fathers of girls ended up marrying the mother. If the proportion marrying the mother is 42% higher in the case of boys, how many would that be? Use a chi-square procedure to tell whether the difference observed is statistically significant.

Note: in a situation with r possibilities for the row variable and c possibilities for the column variable, the **degrees of freedom** are $(r-1) \times (c-1)$ and we can find the chi-square value that cuts off 5% by referring to Table A.5. For example, a 2 by 2 situation has $(2-1) \times (2-1) = 1$ degree of freedom, and the chi-square value for a right-tail probability of .05 is shown to be 3.84.

Example

Is there a relationship between gender and whether a student is a vegetarian or not (or a sometime vegetarian)? Results of a survey of 444 Stats students are shown below:

	No	Some	Yes	Total
Male	145	8	9	162
Female	238	27	17	282
Total	383	35	26	444

Gender was taken to be the explanatory variable, vegetarianism the response. The percentages answering "no", "some", or "yes" for the males were 90%, 5%, and 6%; for the females they were 84%, 10%, and 6%. Thus, we observed fewer "no's" and more "some's" among the females. Could the differences have come about by chance, or do they constitute statistical evidence of a relationship between gender and vegetarianism? To answer this, we construct a table of counts expected if there were no relationship between the two variables:

Expected	No	Some	Yes	Total
Male	140	13	9	162
Female	243	22	17	282
Total	383	35	26	444

Chi-square is calculated to be $.18 + .11 + 1.92 + 1.14 + 0 + 0 = 3.35$. The degrees of freedom are $(2-1)(3-1) = 2$, and according to Table A.5 we must see if this number that measures the overall difference between observed and expected counts is greater than 5.99. Since it is not, we conclude the difference is not statistically significant. We have no evidence that gender plays a role in a student's decision to be a vegetarian or not.

Exercise: Pick two categorical variables from our survey, decide which should be explanatory and response, and discuss if and how you expect them to be related. Then analyze the relationship between them: compare conditional percentages in the response category of interest and tell whether the observed difference seems to you to be significant. Then compute a table of counts expected if the variables were not related, and compute the chi-square statistic. Use Table A.5 to tell whether there is a statistically significant relationship.

Lecture 13

Chapter 7: Probability

Last time we sought statistical evidence of a relationship between row and column variables in a two-way table for victim's race and whether or not the death penalty was imposed. The evidence we sought took the form of a **p-value**: the **probability** of observing so many death sentences in the case of white victims and so few in the case of black victims, if race of the victim really did not play a role in sentencing. Probability is the key to performing **statistical inference**, as we just performed for relationships between two categorical variables, and as we will perform for many other types of situations in the latter part of this course.

The word "probability" has two meanings: the science which is the study of random (uncertain) behavior, and the chance of happening. The latter is determined via either of two interpretations.

Example

The probability of a randomly chosen card being a spade is .25.

Example

The probability of a randomly chosen student getting an A in this course is .25.

Both of these probabilities arise from the **relative frequency interpretation**: the probability of any specific outcome is the proportion of time it occurs over the long run. Within this interpretation are two possible methods for determining a probability.

1. In the case of picking a card, we have made an **assumption about the physical world**—that one fourth of the cards are spades, and that all are equally likely.

2. In the case of getting a grade, I have **observed** the relative frequency of A's **over many repetitions** (over thousands of students).

Besides the relative frequency interpretation, there is the **personal probability interpretation**.

Example

Suppose a friend believes the probability of Bush being re-elected next year is .25. This is also called a **subjective** probability because it is based on an opinion. [Notice that neither relative frequency interpretation would make sense in this situation.] In some sense such a probability is quite arbitrary, but such probabilities must still be consistent and obey the basic rules of probability which we are about to discuss. For example, the friend should not say the chance of re-election is .25 and the chance of not being re-elected is .80.

Much about probability can be understood using common sense—for example, the basic conditions and rules about to be stated are very intuitive. However, there are many situations for which intuition is inadequate or even downright misleading. Systematic application of the basic rules can often accomplish what our intuition cannot. Clear definitions and notation are essential to this systematic approach.

The **sample space** is the set of all possible outcomes in a random circumstance. A **simple event** constitutes just one of these outcomes, whereas an **event** in general is a collection of one or more outcomes. We denote events with capital letters, and write $P(A)$ to denote the probability of event A. The **complement** of event A, writing A^C, denotes the event that A does *not* occur.

Example

What is the probability of rolling a nine with a single die? $P(N) = 0$.

Example

What is the probability of rolling a number less than nine? $P(L) = 1$.

Example

What do we get if we sum the probabilities of rolling a 1 or a 2 or a 3 or a 4 or a 5 or a 6? These represent all the possible outcomes, so their probabilities should sum to 1.

Rule 0:

The probability of an impossible event is 0, the probability of a certain event is 1, and all probabilities must be between 0 and 1. The sum of all possible outcomes together is 1.

Example

The probability of a student in this course getting an A is .25. What is the probability of not getting an A? Getting an A or not getting an A constitute the only possible outcomes, and so their probabilities must sum to 1. Not getting an A must have probability $1 - .25 = .75$.

Rule 1:

$P(A^C) = 1 - P(A)$.

So far the probability events considered involved a single categorical variable: getting a nine or not, getting an A or not, etc. Now we start to think about situations where two categorical events are involved, for example, a student getting an A or not *and* whether the student is male or female.

Probability starts to get more difficult—but more interesting—when we consider the interplay between events. Two events are **mutually exclusive** if they share no outcomes in common; in other words, they are **disjoint**, and they cannot both occur together. Two events are **independent** of each other if whether or not one occurs has no effect on the probability of the other occurring. Thus, two events are **dependent** if knowing whether or not one occurred tells us something about the probability of the other. The **conditional probability** of an event B, given event A, is the probability that B occurs when we know A has occurred. We write the conditional probability of B given A as $P(B|A)$. Thus, if A and B are independent, A having occurred has no impact on the probability of B, and $P(B|A) = P(B)$. If A and B are dependent, then $P(B|A)$ differs from $P(B)$.

Example

In a group of college students, most of the students with pierced ears were female, but a few were male:

	Ears Pierced	Ears Not Pierced	Total
Male	6	34	40
Female	58	2	60
Total	64	36	100

1. Are being male and being female mutually exclusive? Yes, you have to be one or the other; you can't be both.
2. Are being male and having pierced ears mutually exclusive? No, it is possible for males to have pierced ears.
3. Are being male and having pierced ears independent? No, they are dependent, because being male makes it less probable to have pierced ears.
4. If M is the event of being male, and E is the event of having ears pierced, how do $P(E)$ and $P(E|M)$ compare? $P(E) = \frac{64}{100} = .64$ is considerably higher than $P(E|M) = \frac{6}{40} = .15$.

Example

A group of students were classified by gender and by whether or not they received an A in their statistics class.

	A	Not A	Total
Male	10	30	40
Female	15	45	60
Total	25	75	100

1. Are being male and getting an A mutually exclusive? No, it is possible for males to get an A.
2. If M is the event of being male, and A is the event of getting an A, how do $P(A)$ and $P(A|M)$ compare? $P(A) = \frac{25}{100} = .25$ is the same as $P(A|M) = \frac{10}{40} = .25$.

3. Are being male and getting an A independent? Yes, because being male doesn't have any impact on the probability of getting an A.

The next examples illustrate the probability that *either* of two events happens.

Example

The probability of a student in this course getting an A is .25, and the probability of getting a B is .30. What is the probability of getting an A *or* a B? $.25 + .30 = .55$.

Example

The probability of a student is this course getting an A is .25; the probability of being female is .60. What is the probability of getting an A *or* being female? We could simply add probabilities in the A-or-B example because these two events are **mutually exclusive**. But getting an A and being female may both occur, so we can't just add to find the probability of one or the other—we'd be double-counting the overlap, that is, double-counting the probability of being female **and** getting an A. Thus, to find $P(A \text{ or } F)$, we add $P(A)$ and $P(F)$ and subtract the overlap $P(A \text{ and } F)$:

$$P(A \text{ or } F) = P(A) + P(F) - P(A \text{ and } F) = \frac{25}{100} + \frac{60}{100} - \frac{10}{100} = \frac{75}{100}$$

Rule 2:

In general, the probability of one event or another occurring is

$$P(A \text{ or } B) = P(A) + P(B) - P(A \text{ and } B)$$

If the events are mutually exclusive, then $P(A \text{ and } B) = 0$ and so $P(A \text{ or } B) = P(A) + P(B)$.

The next examples illustrate the probability that both of two events occur.

Example

Pick a card at random from four, consisting of one each clubs, hearts, diamonds, and spades. Replace it and pick another. What is the probability that the first is a club *and* the second is a diamond? $\frac{1}{4} * \frac{1}{4} = \frac{1}{16}$.

Example

Pick a card from those four, *don't* replace it, and pick another. What is the probability that the first is a club and the second is a diamond? All of the possible and equally likely outcomes in the sample space are

$$CD \quad CH \quad CS \quad DC \quad DH \quad DS \quad HC \quad HD \quad HS \quad SC \quad SD \quad SH$$

and so the probability of getting a club then a diamond is $\frac{1}{12}$. What probabilities must we multiply to get $\frac{1}{12}$? It is the probability of getting a club, times the probability of getting a diamond given that the first card was a club: $P(C \text{ and } D) = P(C)P(D|C)$.

Notice that in the first of these two examples, because the first card was replaced before the second card was picked, the two events were independent. In the second of these examples, because the first card was not replaced, the selections were dependent. In general, **sampling with replacement** results in **independence**, whereas **sampling without replacement** results in **dependence**.

Rule 3:

In general, the probability of one event and another is

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

If the events are independent, $P(B|A) = P(B)$ and so $P(A \text{ and } B) = P(A)P(B)$.

Note: In a previous example, we found the probability of having pierced ears, given that a student was male, by restricting our attention to the 40 male students and finding that 6 of those 40 had pierced ears. It is possible to express a conditional probability in terms of ordinary probabilities as follows:

$$P(E|M) = \frac{6}{40} = \frac{6/100}{40/100} = \frac{P(E \text{ and } M)}{P(M)}$$

In general, the conditional probability of event B, given event A, is

$$P(B|A) = \frac{P(A \text{ and } B)}{P(A)}$$

Example

The probability of being a female is .6; the probability of being a female and getting an A is .15. What is the probability of getting an A, given that a student is female? $P(A|F) = \frac{P(A \text{and} F)}{P(F)} = \frac{.15}{.6} = .25$.