Checking for Independence in Samples

An advertising agency in a large city is conducting a survey of adults to investigate whether there is an association between highest level of educational achievement and primary source for news. The company takes a random sample of 2,500 adults in the city. The results are shown in the table below.

Highest Level of Educational Achievement

Primary Source For News	Not a high school graduate	High school graduate but not a college graduate	College graduate	Total
Newspaper	49	205	188	442
Local Television	90	170	75	335
Cable Television	113	496	147	756
Internet	41	401	245	687
None	77	165	38	280
Total	370	1437	693	2500

Know how to find a probability, a union probability (A or B), and a conditional probability using a two way table, (11, 2a; 10b, 5ab; 03b, 2ab; M7, 18)

If an adult is selected at random from this sample, what is the probability that the selected adult obtains news primarily from the internet?

$$P(Internet) = \frac{687}{2500} = 0.2748$$

If an adult is selected at random from this sample, what is the probability that the selected adult is a college graduate or obtains news primarily from the internet?

$$P(A \cup B) = P(A) + P(B) - P(A \cap B)$$

$$P(College \cup Internet) = P(College) + P(Internet) - P(College \cap Internet)$$

$$= \frac{693}{2500} + \frac{687}{2500} - \frac{245}{2500} = \frac{1135}{2500} = 0.454$$

If an adult who is a college graduate is to be selected at random from this sample, what is the probability that the selected adult obtains news primarily from the internet?

We are only looking at college graduates

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

$$P(Internent|College) = \frac{P(Internet \cap College)}{P(College)} = \frac{\frac{245}{2500}}{\frac{693}{2500}} = \frac{245}{693} = 0.3535$$

Know the two ways to check for independence (11, 2b; 10b, 5c; 03b, 2c)

When selecting adults at random from the sample of 2500 adults are the events "is a college graduate" and "obtains news primarily from the internet" independent?

Method 1

$P(A|B) = P(A) \rightarrow P(Internet|College) = P(Internet) \rightarrow 0.3535 \neq 0.2748$

Since these are not equal, they are not independent. It appears that level of education does impact the likelihood of using the internet for news

Method 2

$$P(A \cap B) = P(A) \cdot P(B) \rightarrow P(Intenet \cap College) = P(Internet) \cdot P(College) \rightarrow \frac{245}{2500} \neq \frac{687}{2500} \cdot \frac{693}{2500}$$

Since these are not equal, they are not independent. It appears that level of education does impact the likelihood of using the internet for news

OTHER STUFF

Know how to find union probabilities of disjoint events, (09b, 2bc;08, 3cd; 04, 4a)

The ELISA tests whether a patient has contracted HIV. The ELISA is said to be positive if it indicates that HIV is present in a blood sample, and the ELISA is said to be negative if it does not indicate that HIV is present in a blood sample. Because of variability in antibody levels among human patients, the ELISA does not always indicate the correct result.

As part of a training program, the staff applied the ELISA to 500 blood samples known to not contain HIV. The ELISA was positive for 37 of those blood samples and negative for the other 463 samples.

When a blood sample yields a positive ELISA result, two more ELISAs are performed on the same blood sample. If at least one of the two additional ELISAs is positive, the blood sample is subjected to a more expensive and more accurate test to make a definitive determination of whether HIV is present in the sample. Under the assumption of independence and using the data above, what is the probability that a new blood sample that comes into the lab will be subjected to the more expensive test if that sample does not contain HIV?

$P(positive) = \frac{37}{500} = 0.074$

We will have to run the more expensive test for one of the two situations listed below

1) PP (2 positive tests in a row)
2) PNP (First positive, then a negative, but third test positive)

Since we are assuming independence, a positive result will always have 0.074 probability.

The probability of option 1, positive on first test and then another positive test is

P(positive then positive) = (0.074)(0.074) = 0.005476

The probability of option 2, positive on the first test, negative on second test, positive on third test is

P(positive then negative then positive) = (0.074)(0.926)(0.074) = 0.005070776

So, the probability of having to run the more expensive test if you are HIV negative is

= 0.005476 + 0.005070776
= 0.010546776

Know how to find a joint probability (A and B) of two or more independent events, (08, 3b; 04, 3b; 04, 4a; 03b, 5a; M7, 38; M2, 36)

In the problem above, each test is independent

$P(Positive \cap Postivie) = 0.074 \cdot 0.074$ is the joint probability of two independent events

Know how to interpret a probability in the context of a problem and deem if it is likely to occur or not, (04, 3c)

If I have a 0.0001% chance of winning a game, you would determine that this probably would not happen randomly. You would call me a cheater if I won.

If I have a 25% chance of winning a game, you would determine that this could happen randomly. You would not have strong evidence to call me a cheater if I won.

Know what it means for events to be mutually exclusive and how that affects their joint probability and their union probability, (M7, 36; M2, 23)

If two events are mutually exclusive, they cannot happen at the same time. For example, I cannot be in ninth grade and tenth grade at the same time. If two events are mutually exclusive, $P(A \cup B) = P(A) + P(B)$. You do not have to worry about $P(A \cap B)$ because $P(A \cap B) = 0$ for disjoint events. They cannot occur simultaneously.

Know how to set up a simulation and perform it using a table of random digits, (M2, 4)

If 80% of population is female and 20% male, you could use a RNG and make

0,1→Male 2, 3, 4, 5, 6, 7, 8, 9→Female