**Probability #1 (by Hand and in R!)**

**#1 (By Hand)** In a Stat 200 class at PSU, students were asked to rate how much they liked various kinds of music on a scale of 1 (don’t like at all) to 6 (like very much). Following are probability distributions for ratings of Top 40 music, with distributions given separately for males and females in the class. The probability for a rating of 6 by males is intentionally not given in the distributions.

**FEMALES**

Rating 1 2 3 4 5 6

Probability .04 .05 .09 .24 .32 .26

**MALES**

Rating 1 2 3 4 5 6

Probability .09 .13 .17 .28 .23 \_\_

**a.** *Explain* whether you think that the probabilities given above should be interpreted as relative frequency probabilities or as personal probabilities.

These are relative frequency probabilities. We repeated the experiment a large number of times, and counted the frequency with which each response occurred.

**b.** For males, what is the value of the missing probability for a rating of 6? Explain or show how you determined this value.

.1 🡺 because they all have to add up to 1.

**c.** Suppose that we randomly sample one male in the class. Let event A = the rating given by this person is a 6. What event is the **complement** of event A?

**d.** Refer to part c, where event A = rating is a 6.

For the females, what is the value of P(A) = probability that event A occurs? .26

For the females, what is the value of P(complement of A) = probability of the complement of A? .74

**e.** Consider randomly selecting one female from the class. Explain whether the events A = rating is a 6 and B = rating is a 5 are **mutually exclusive** events are not.

**Yes, ME because a single person cannot give both ratings.**

**f.** For each sex separately, determine the probability that the rating given by a randomly selected person is either a 5 (event A) or a 6 (event B).

For females, probability that rating is 5 or 6 is P(A or B) = .58

For males, probability that rating is 5 or 6 is P(A or B) = .33

**g.** Using the values found in part f, calculate the probability that the rating is 4 or less for each sex separately.

For females, probability that rating is 4 or less = .42

For males, probability that rating is 4 or less = .67

**i.** Using the values found in parts f and g, explain why we can say that the rating of Top 40 music and student gender are **dependent** (related) characteristics.

**If they were independent, then the rating probabilities would not differ by gender, but these do differ.**

**j.** Consider randomly selecting one female from the class. Explain whether the events A = rating is a 6 and B = rating is a 5 are **independent** events or **dependent** events.

**They are ME events for a single female, so they are dependent.**

**k.** Consider randomly selecting one female and one male from the class. Explain why we could say that the events A = the female gives a rating of 6 and B = the male gives a rating of 6 are **independent** events.

**They are independent, because a randomly chosen female will not influence a randomly chosen male.**

**#2 (R!)** At the course website, access the folder for today’s lab, where I have placed a dataset called chi2\_ex5.R, which, on a Mac, you can double-click directly to load it into R (with the load command produced in the console). On a Windows machine, you must use the load command. First, download the data (say to your Desktop), and then change the working directory of R to the location where you downloaded the data (File<Change dir…<). Then simply use

load("chi2\_ex5.R")

The data now exists in R, and to see what it is called, use the ls() command to get a list of all objects currently in R. It gives data for a few questions asked in the Spring 2006 PSU Stat 200 survey. The variable ***FavMusic*** gives responses to the question, “What is your favorite type of music (Rock, Rap/HipHop, Pop, Country, Other)?” In this activity, we’ll look at how this variable relates to gender.

**a.** For each sex, determine the sample percentages in each category of favorite music. (Percentages should add to 100% within each sex.) Give those percentages as the answer to this part.

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
|  | Country | Other | Pop | Rap/Hip Hop | Rock |
| Female | .14 | .21 | .21 | .19 | .26 |
| Male | .08 | .19 | .04 | .26 | .43 |

**b.** Using your answer to part **a**. as the basis, write a brief description of how males and females differ with regard to their favorite type of music.

Females are 5-6x as likely to list Pop and almost twice as likely to list Country as their favorite genre relative to males, whereas males are almost twice as likely to list Rock as their favorite genre.

**c.** Do a chi-square test of the relationship between gender and favorite type of music.

What is the *p*-value for the test? ~0

State a conclusion for the chi-square test.

We reject the null and accept the alternative. We have overwhelming evidence to state that there is a significant relationship between a person’s gender and his/her ranking of music genres.

**d.** Suppose that we were to randomly select one person from this dataset. Using relative frequency probabilities, what are the values of the following probabilities? To start, you probably should have a two-way table of counts to reference.

Probability that the person likes Rock best if the person picked is a male..43

Probability that the person likes Rock best if the person picked is a female..26

Probability that person picked is a female and also likes Pop best.107/905

Probability that person picked likes Pop best if they are a female..21

Probability that the person picked is a female if they are somebody who likes Pop best. 107/121

Probability that the person picked is a male if they are somebody who likes Pop best. 14/121

**#3 (By Hand)** For each characteristic, explain whether the random variable is continuous or discrete. If it is discrete, indicate whether it may be a binomial random variable or not.

**a.** The number of left-handed individuals in a random sample of 100 people taken from a population of 100 million.

Discrete and Binomial.

**b.** Vehicle speeds at a highway location. Continuous

**c.** The number of accidents reported last year at a highway location.

Discrete but not Binomial. There is no fixed n.

**#4 (By Hand)** For a large population, probabilities for the number of meals eaten yesterday are:

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Meals | **1** | **2** | **3** | **4** |
| Probability | 0.10 | 0.32 | 0.56 | .02 |

**a.** Calculate the expected value (mean value) of the number meals eaten in yesterday by members of this population. Formula = Sum (Value × Probability)

1\*(.1) + 2\*(.32) + 3\*(.56) + 4\*(.02) = 2.5

**b.** Write the cumulative probability distribution for the number of meals eaten yesterday are:

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Meals | **1** | **2** | **3** | **4** |
| Probability | .1 | .42 | .98 | 1.0 |

**#5 (By Hand)** The following table gives the probability distribution for *X =* classes skipped yesterday by students at a college.

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| Classes Skipped | 0 | 1 | 2 | 3 | 4 |
| Probability | .73 | .16 | .06 | .03 | .02 |

**a.** Calculate the mean of this random variable

0\*.73 + 1\*.16 + 2\*.06 + 3\*.03 + 4\*.02 = .45

**b.** Write a table that gives the cumulative distribution function for *X =* classes skipped

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| Classes Skipped | 0 | 1 | 2 | 3 | 4 |
| Cumulative Probability | .73 | .89 | .95 | .98 | 1.0 |

**#6 (By Hand)** In the casino game of roulette, a gambler can bet on which of 38 numbers will be the result when the roulette wheel is spun. On a $2 bet, a gambler gains $70 if he or she picks the right number but loses the $2 otherwise.

**a.** Let *X =* amount gained or lost on a $2 bet on a roulette number. Write out the probability distribution of *X.*

|  |  |  |
| --- | --- | --- |
| Profit | -2 | 68 |
| Probability | 37/38 | 1/38 |

**b.** Calculate expected value of X (the mean value of *X*)*.* What does this value indicate about the advantage that a casino has over roulette players? (That is, interpret the expected value in this situation.)

-2\*37/38 + 68\*1/38 = -.1578947

Everytime you play, the casino expects to earn about 16 cents of profit.

**#7 (By Hand)** In each case, explain why the random variable is discrete, but not binomial.

**a.** A woman buys a lottery ticket every week for which the probability of winning anything at all is 1/10. She continues to buy them until she has won three times. *X =* the number of tickets she buys.

The set of possible values of X is countably infinite, so X is discrete, but it is not a finite set of possible values because there is no fixed number of trials, n. So it cannot be Binomial.

**b.** A poker hand consists of 5 cards drawn from a standard deck of 52 cards. *X =* the number of aces in the hand.

There are a finite number of possible values (0,1,2,3,4) and a fixed number of successes, but these trials are not independent. Once I get a success on one trials, successes become less and less likely on subsequent trials.