OI 3.2: Marginal, Joint, and Conditional Probability (Part 1)

Packages Needed To Recreate Code on Slides

```
library(tidyverse)
```

Warning: It is not expected that you understand all the R code in this presentation right now. You will go over more R code in SDS 100 in the coming weeks. However, you are welcome to try to make these plots on your own.

Two Random Processes

- We briefly introduced the concept of two random processes in OI 3.1 when we discussed independence.
- With two random processes that are dependent we often want to explore relationships between these processes.

Data Set

Run the following code to download our data set (Thanks Dr. Cook!)

```
movies <- read.csv("https://raw.githubusercontent.com/rpkgarcia/MyDataSets/main/movies.csv")
```

Data Set

- 3010 Movies. Oldest movie was made in 1970. We care about two variables
- `high_gross`: Did the movie gross twice its budget ("High"), or not ("Low")
- `content_rating`: MPAA ratings. Possible values are "Not Rated", "G", "PG", "PG-13", "R", "NC-17".

Contingency Tables

Recall contingency tables organize every possible outcome for two discrete/categorical variables and associated outcomes

	Box Office Gross		
MPAA Rating	Low	High	Total
Not Rated	4	17	21
G	25	41	66
PG	305	166	471
PG-13	303	805	1108
R	422	909	1331
NC-17	8	5	13
Total	928	2082	3010

Contingency Tables

Contingency tables can also have frequencies instead of counts. (Rounded values)

MPAA Rating	**Box Office Gross** Low	High	Total
Not Rated	.001	.006	.007
G	.008	.014	.022
PG	.055	.101	.156
PG-13	.101	.267	.368
R	.140	.302	.442
NC-17	.003	.002	.005
Total	.308	.692	1

What do you notice?

Contingency Tables

- We will use contingency tables as a tool to introduce more concepts about the relationships between two variables.
- Remember: the definitions, rules, theorems, etc regarding probability apply beyond survy data.

Marginal and Joint Probabilities

- The probability of outcomes for two or more random processes is called a *joint probability*.
- The *marginal probability* is the probability of an outcome for one random process, independent of other events.

MPAA Rating	**Box Office Gross** Low	High	Total
Not Rated	.001	.006	.007
G	.008	.014	.022
PG	.055	.101	.156
PG-13	.101	.267	.368
R	.140	.302	.442
NC-17	.003	.002	.005
Total	.308	.692	1

Marginal and Joint Probabilities

MPAA Rating	**Box Office Gross** Low	High	Total
Not Rated	.001	.006	.007
G	.008	.014	.022
PG	.055	.101	.156
PG-13	.101	.267	.368
R	.140	.302	.442
NC-17	.003	.002	.005
Total	.308	.692	1

a) What is the probability that a movie is PG-13 and high grossing?

b) What is the probability that a movie is rated G?

c) What is the probability that a movie is not high grossing?

d) Are the variables *MPAA* and *Box Office Gross* independent?

Marginal and Joint Probabilities

a) .267

b) .022

c) .308

d) No. $P(\text{Not Rated } and \text{ Low}) \neq P(\text{Not Rated})P(\text{Low})$

Practice Question

		Gender		
		Male	Female	Total
Best hamburger place	Five Guys Burgers	5	6	11
	In-N-Out Burger	162	181	343
	Fat Burger	10	12	22
	Tommy's Hamburgers	27	27	54
	Umami Burger	5	1	6
	Other	26	20	46
	Not Sure	13	5	18
	Total	248	252	500

[OI 3.17] Burger preferences from a 2010 SurveyUSA poll for 500 Los Angeles residents.

a) Are being female and liking Five Guys Burgers mutually exclusive?

b) What is the probability that a randomly chosen male likes In-N-Out the best?

c) What is the probability that a randomly chosen female likes In-N-Out the best?

d) What is the probability that a man and a woman who are dating both like In-N-Out the best? Note any assumption you make and evaluate whether you think that assumption is reasonable.

e) What is the probability that a randomly chosen person likes Umami best or that person is female?