# Outline

## Exercise:

- Assume $X = (x = 1, x = 3, x = 5, x = 7)$ and $Y = (y = 1, y = 2, y = 3, y = 4)$. We have the following joint probability distribution:

| | X | | | |
|---|---|---|---|---|
| Y | 1 | 3 | 5 | 7 |
| 1 | 0.066 | 0.044 | 0.132 | 0.066 |
| 2 | 0.044 | 0.044 | 0.033 | 0.044 |
| 3 | 0.121 | 0.099 | 0.084 | 0.073 |
| 4 | 0.026 | 0.053 | 0.042 | 0.024 |

- Recreate this table in R
- Calculate the marginal probabilities of $X$ and $Y$
- Calculate the expected value of $X$ and $Y$, i.e., $E[X]$ and $E[Y]$
- Given the condition that we have $x = 3$, what is the probability of getting y=2, i.e., Calculate $P(y = 2|x = 3)$?
- Calculate the conditional mean and variance of $X$ when $y = 2$?
- Calculate the variance of X and Y
- Calculate the covariance between $X$ and $Y$?

# Outline

## Solution in R:

- **Q1: Recreate this table in R**
  - First create the matrix of probabilities and convert it into a table in R:

```
data <- matrix( c (0.066, 0.044, 0.132, 0.066,
                   0.044, 0.044, 0.033, 0.044,
                   0.121, 0.099, 0.084, 0.073,
                   0.026, 0.053, 0.042, 0.024),
                  ncol = 4, byrow = TRUE)
# Define column names
colnames(data) <- c("1", "3", "5", "7")
# Define row names
rownames(data) <- c("1", "2", "3", "4")
data <- as.table(data)

data<- prop.table(data)
data
```

```
##        1      3      5      7
## 1 0.0663 0.0442 0.1327 0.0663
## 2 0.0442 0.0442 0.0332 0.0442
## 3 0.1216 0.0995 0.0844 0.0734
## 4 0.0261 0.0533 0.0422 0.0241
```

## Solution in R:

- **Q2: Calculate the marginal probabilities of $X$ and $Y$**
  - Calculate the marginal probabilities in R:

```
table <- addmargins(data)
table
```

```
##           1      3      5      7    Sum
## 1    0.0663 0.0442 0.1327 0.0663 0.3095
## 2    0.0442 0.0442 0.0332 0.0442 0.1658
## 3    0.1216 0.0995 0.0844 0.0734 0.3789
## 4    0.0261 0.0533 0.0422 0.0241 0.1457
## Sum 0.2583 0.2412 0.2925 0.2080 1.0000
```

## Solution in R:

- **Q3: Calculate the expected value of $X$ and $Y$, i.e., $E[X]$ and $E[Y]$**
  - **Mean of X:**
    - What are the possible values taken by X?

```
x=c(1,3,5,7)
```

    - What are the marginal probabilities of X?

```
px=c(0.2583, 0.2412, 0.2925, 0.2080)
```

    - Now calculate the mean of X

```
mean_x= sum(x*px);        mean_x

## [1] 3.9
```

  - **Mean of Y:**
    - Same strategy again!

```
y=c(1,2,3,4);                    py=c(0.3095, 0.1658, 0.3789, 0.1457)
mean_y = sum(y*px);              mean_y

## [1] 2.45
```

## Solution in R:

- **Q3: Calculate the expected value of $X$ and $Y$, i.e., $E[X]$ and $E[Y]$**
  - **Mean of X:**
    - What are the possible values taken by X?

```
x=c(1,3,5,7)
```

    - What are the marginal probabilities of X?

```
px=c(0.2583, 0.2412, 0.2925, 0.2080)
```

    - Now calculate the mean of X

```
mean_x= sum(x*px);        mean_x

## [1] 3.9
```

  - **Mean of Y:**
    - Same strategy again!

```
y=c(1,2,3,4);                    py=c(0.3095, 0.1658, 0.3789, 0.1457)
mean_y = sum(y*py);              mean_y

## [1] 2.36
```

## Solution in R:

- **Q4: Given the condition that we have $X = 3$, what is the probability of getting Y=2, i.e., Calculate $P(Y = 2|X = 3)$?**
  - This can be manually calculated using the forumla:

$$P(Y|X) = \frac{P(y, x)}{P(x)}$$

  - So, what is the joint probability P(y,x) when Y=2 and X=3 [0.044]
  - So, what is the marginal probability of X when X=3 [0.2412]
  - Now, plug in the values in the formula:

$$P(Y|X) = \frac{0.044}{0.241} = 0.18$$

## Solution in R:

- **Q4: Given the condition that we have $X = 3$, what is the probability of getting Y=2, i.e., Calculate $P(Y = 2|X = 3)$?**
  - You can also solved this directly in R by computing the conditional probabilities (when you asume that the columns are given condition)

```
prop.table(data, margin=2)
```

```
##       1     3     5     7
## 1 0.257 0.183 0.454 0.319
## 2 0.171 0.183 0.113 0.213
## 3 0.471 0.413 0.289 0.353
## 4 0.101 0.221 0.144 0.116
```

  - After calculating condition probabilities (given the column conditions are given), we can see that when Y=2 and X=3, the answer is 0.18

- If you want to calculate conditional probabilities (given the row conditions are given), we can use the following code:

```
prop.table(data, margin=1)
```

```
##       1     3     5     7
## 1 0.214 0.143 0.429 0.214
## 2 0.267 0.267 0.200 0.267
## 3 0.321 0.263 0.223 0.194
## 4 0.179 0.366 0.290 0.166
```

## Solution in R:

- **Q5: Calculate the conditional mean and variance of $X$ when $y = 2$?**
  - For conditional mean, we also need joint probabilities between X and Y
  - But first, what are the possible values taken by X

```
x=c(1,3,5,7)
```

  - What are the joint probabilities of X when y=2?

```
px2=c(0.044, 0.044, 0.033, 0.044)
```

  - What is the marginal probability of Y when y=2? [0.1658]
  - Now calculate the mean using the formula $\mu_{X|Y} = E[X|Y] = \sum_x (x|y)P(x|y)$

```
x_cond= sum(x*px2/0.16)
```

- Now calculate the variance of X when y=2:

```
sum((x - x_cond)^2*(px2)/0.16)

## [1] 5.44
```

## Solution in R:

- **Q6: Calculate the variance of X and Y?**
  - For variance of X, I use the formula: $Var[X] = \sum(x - \mu_x)^2 P_x$
  - In Question no. 3, I have saved the value of the mean of x already:

```
mean_x
```

```
## [1] 3.9
```

  - I can proceed to calculating the variance of X:

```
var_x=sum((x - mean_x)^2*(px))
var_x
```

```
## [1] 4.72
```

- Same strategy for calculating the variance of Y

## Solution in R:

- **Q7: Calculate the covariance between $X$ and $Y$?**
  - To calculate the covariance between X and Y, I use the following formula

$$Cov(X, Y) = E(X - \mu_X)(Y - \mu_Y)$$

$$Cov(X, Y) = E(XY) - E(X)E(Y)$$

  - We can also express this as:

$$Cov(X, Y) = \sum_x \sum_y (xy)P(x, y) - \mu_X \mu_Y$$

  - 
```
cov_xy = 1*(1*0.066 + 3*0.044 + 5*0.132 + 7*0.066) +
         2*(1*0.044 + 3*0.044 + 5*0.033 + 7*0.044) +
         3*(1*0.121 + 3*0.099 + 5*0.084 + 7*0.073) +
         4*(1*0.026 + 3*0.053 + 5*0.042 + 7*0.024) -
         mean_x*mean_y
cov_xy
```

```
## [1] -0.29
```