Overview

- Joint Probability Mass Function
- Covariance
- Correlation
- Dependence and Correlation
Pairs of Random Variables

Example: exam scores

![Scatter plot of mid-term score vs. final exam score with random data points.](image-url)
Pairs of Random Variables

dashed – 45° line, green – least squares fit

Regression to the mean?
Joint Probability Mass Function

Suppose we have two discrete random variables $X$ and $Y$ on same sample space $S$.

- $P(X = x \text{ and } Y = y)$ is called their joint probability mass function
- Let’s go back to sample space $S$. Remember RV $X$ is really a function mapping from $S$ to a real value i.e. should really be written $X(\omega)$. Ditto $Y$.
- Let $E_x = \{\omega \in S : X(\omega) = x\}$ be set of outcomes for which $X = x$
- Let $E_y = \{\omega \in S : Y(\omega) = y\}$ be set of outcomes for which $Y = y$
- $P(X = x) = P(E_x), \ P(Y = y) = P(E_y)$
- Probability of both is $P(E_x \cap E_y)$ and $P(X = x \text{ and } Y = y) = P(E_x \cap E_y)$.
Joint Probability Mass Functions

Example: operating system loyalty. Person buys one computer, then another. \( X = 1 \) if first computer runs windows, else 0. \( Y = 1 \) is second computer runs windows, else 0.

- Joint probability mass function:

<table>
<thead>
<tr>
<th></th>
<th>( x=0 )</th>
<th>( x=1 )</th>
<th>( P(Y=y) )</th>
</tr>
</thead>
<tbody>
<tr>
<td>( y=0 )</td>
<td>0.2</td>
<td>0.3</td>
<td>0.5</td>
</tr>
<tr>
<td>( y=1 )</td>
<td>0.1</td>
<td>0.4</td>
<td>0.5</td>
</tr>
<tr>
<td>( P(X=x) )</td>
<td>0.3</td>
<td>0.7</td>
<td>1</td>
</tr>
</tbody>
</table>

- \( P(X = 0 \text{ and } Y = 0) = 0.2, \ P(X = 0 \text{ and } Y = 1) = 0.3 \) etc.
Covariance

Say $X$ and $Y$ are random variables with expected values $\mu_X$ and $\mu_Y$. The covariance of $X$ and $Y$ is defined as:

$$\text{Cov}(X, Y) = E[(X - \mu_X)(Y - \mu_Y)]$$

Equivalently

$$\text{Cov}(X, Y) = E[XY] - E[X]\mu_Y - E[Y]\mu_X + \mu_X\mu_Y$$

$$= E[XY] - E[E[X]Y] - E[X]\mu_Y + \mu_XE[Y]$$

$$= E[XY] - \mu_X\mu_Y - \mu_Y\mu_X + \mu_X\mu_Y$$

$$= E[XY] - \mu_X\mu_Y = E[XY] - E[X]E[Y]$$

- $\text{Cov}(X, X) = \text{Var}(X)$.
- Recall when $X$ and $Y$ are independent then $E[XY] = E[X]E[Y]$, so $\text{Cov}(X, Y) = 0$. But $\text{Cov}(X, Y) = 0$ does not imply that $X$ and $Y$ are independent – more on this shortly.
Correlation

- Example 1: Suppose $X = Y$, then

- Example 2: Suppose $X = -Y$, then

- In English: $\text{Cov}(X, Y)$ is positive if $X$ and $Y$ tend to increase together, and negative if and increase in one tends to correspond to a decrease in the other.

- But the magnitude of the covariance can be hard to understand

- The correlation between $X$ and $Y$ is defined as:

  $$\text{Corr}(X, Y) = \frac{\text{Cov}(X, Y)}{\sqrt{\text{Var}(X)\text{Var}(Y)}}$$

- Also use $\rho_{X,Y}$ instead of $\text{Corr}(X, Y)$, similarly to the way we use $\mu_X$ as shorthand for expected value $E[X]$ and $\sigma_X$ for standard deviation $\sqrt{\text{Var}(X)}$ (so $\sigma_X^2 = \text{Var}(X)$)

- Sometimes also called the Pearson correlation coefficient.
Correlation

- Correlation varies between -1 and 1.
- If $X = Y$ then $\text{corr}(X, Y) = 1$. If $X = -Y$ then $\text{corr}(X, Y) = -1$.
- Example: Suppose $Y = X + aN$, where $N$ is -1 with probability 0.5 and +1 with probability 0.5. Plot$^1$ of $\text{corr}(X, Y)$ vs parameter $a$:

```
rho = []; for a=[0:0.1:10], x=[0:0.001:1]; y=x+a*(2*(rand(1,length(x))>0.5)-1); rho=[rho;a,corr(x',y')]; end; plot(rho(:,1),rho(:,2))
```
Example: Correlation Between Height and Weight

<table>
<thead>
<tr>
<th>Weight</th>
<th>Height</th>
<th>$W \times H$</th>
</tr>
</thead>
<tbody>
<tr>
<td>64</td>
<td>57</td>
<td>3648</td>
</tr>
<tr>
<td>71</td>
<td>59</td>
<td>4189</td>
</tr>
<tr>
<td>53</td>
<td>49</td>
<td>2597</td>
</tr>
<tr>
<td>67</td>
<td>62</td>
<td>4154</td>
</tr>
<tr>
<td>55</td>
<td>51</td>
<td>2805</td>
</tr>
<tr>
<td>58</td>
<td>50</td>
<td>2900</td>
</tr>
<tr>
<td>77</td>
<td>55</td>
<td>4235</td>
</tr>
<tr>
<td>57</td>
<td>48</td>
<td>2736</td>
</tr>
<tr>
<td>56</td>
<td>42</td>
<td>2352</td>
</tr>
<tr>
<td>51</td>
<td>42</td>
<td>2142</td>
</tr>
<tr>
<td>76</td>
<td>61</td>
<td>4636</td>
</tr>
<tr>
<td>68</td>
<td>57</td>
<td>3876</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>$E[W]$</th>
<th>$E[H]$</th>
<th>$E[WH]$</th>
</tr>
</thead>
<tbody>
<tr>
<td>62.75</td>
<td>52.75</td>
<td>3355.83</td>
</tr>
</tbody>
</table>

$E[W^2]$  $E[H^2]$  
4011.58  2825.25

$Var(W)$  $Var(H)$  
74.02  42.69

$Cov(W, H) = E[WH] − E[W]E[H]$
$= 3355.83 − 62.75 \times 52.75$
$= 45.77$

$Corr(W, H) = \frac{Cov(W, H)}{\sqrt{Var(W)Var(H)}}$
$= \frac{45.77}{\sqrt{74.02 \times 42.69}} = 0.81$
Dice Example

Consider rolling a 6-sided die

- Indicator variable $X = 1$ if roll is 1,2,3 or 4
- Indicator variable $Y = 1$ if roll is 3,4,5 or 6

What is $\text{Cov}(X, Y)$?

- $E[X] = \frac{2}{3}$, $E[Y] = \frac{2}{3}$
- if $X = 0$ then $Y = 1$ and if $Y = 0$ then $X = 1$

\[
E[XY] = \sum_x \sum_y xyP(X = x \text{ and } Y = y)
\]

\[
= 0 \times 0 \times 0 + 0 \times 1 \times \frac{1}{3} + 1 \times 0 \times \frac{1}{3} + 1 \times 1 \times \frac{1}{3} = \frac{1}{3}
\]

- $\text{Cov}(X, Y) = E[XY] - E[X]E[Y] = \frac{1}{3} - \frac{4}{9} = -\frac{1}{9}$
- Now $P(X = 1) = \frac{2}{3}$ and $P(X = 1|Y = 1) = \frac{1}{2}$. So observing $Y = 1$ makes $X = 1$ less likely
Correlation

The correlation is another example of a summary statistic. It indicates the strength of a linear relationship between $X$ and $Y$. Great care is needed though as it can easily be misleading.

- Correlation says **nothing** about the slope of line (other than its sign).
- When relationship between $X$ and $Y$ is not roughly linear, correlation coefficient tells us almost nothing.

source: https://en.wikipedia.org/wiki/Correlation_and_dependence
Anscombe’s Quartet

- All four datasets have correlation 0.816
- Take home message: plot the data, don’t just rely on summary statistics such as mean, variance, correlation.
Dependence and Correlation

Recall when $X$ and $Y$ are independent then $E[XY] = E[X]E[Y]$, so $\text{corr}(X, Y) = 0$. But $\text{corr}(X, Y) = 0$ does not imply that $X$ and $Y$ are independent.

Example: $X$ and $Y$ are random variables with joint PMF:

<table>
<thead>
<tr>
<th></th>
<th>x=-1</th>
<th>x=0</th>
<th>x=1</th>
<th>P(Y=y)</th>
</tr>
</thead>
<tbody>
<tr>
<td>y=0</td>
<td>$\frac{1}{3}$</td>
<td>0</td>
<td>$\frac{1}{3}$</td>
<td>$\frac{2}{3}$</td>
</tr>
<tr>
<td>y=1</td>
<td>0</td>
<td>$\frac{1}{3}$</td>
<td>0</td>
<td>$\frac{1}{3}$</td>
</tr>
<tr>
<td>P(X=x)</td>
<td>$\frac{1}{3}$</td>
<td>$\frac{1}{3}$</td>
<td>$\frac{1}{3}$</td>
<td>1</td>
</tr>
</tbody>
</table>

$X$ takes values $\{-1, 0, 1\}$ with equal probability and

$$Y = \begin{cases} 1 & \text{if } X = 0 \\ 0 & \text{if } X \neq 0 \end{cases}$$

- $E[X] = -1 \times \frac{1}{3} + 0 \times \frac{1}{3} + 1 \times \frac{1}{3} = 0$, $E[Y] = 0 \times \frac{2}{3} + 1 \times \frac{1}{3} = \frac{1}{3}$
- Since $XY = 0$ then $E[XY] = 0$
- $\text{Cov}(X, Y) = E[XY] - E[X]E[Y] = 0 - 0 = 0$
- But $X$ and $Y$ are clearly dependent
Correlation and Causation

Correlation does not imply causation.

source: https://xkcd.com/552/
Correlation and Causation

- Fires and firemen
- Prices and music ...
**Conditional Expectation**

$X$ and $Y$ are jointly distributed discrete random variables.

- Recall conditional PMF of $X$ given $Y = y$ is
  \[ P(X = x | Y = y) = \frac{P((X = x \text{ and } Y = y))}{P(Y = y)} \]

- Define conditional expectation of $X$ given $Y = y$ as:
  \[ E[X | Y = y] = \sum_x x P(X = x | Y = y) \]

- This is not the same as the expectation $E[X]$ e.g. its one thing to ask what the average height of a person in Ireland it and another to ask this once we know that they are male.
Conditional Expectation

Roll two 6-sided dice. $X$ is value of the sum, $Y$ is the outcome of the first die roll.

$$E[X|Y = 6] = \sum_x xP(X = x|Y = y)$$

$$= \frac{1}{6}(7 + 8 + 9 + 10 + 11 + 12) = \frac{57}{6} = 0.5$$

- Makes sense: $6 + E[\text{value of second die roll}] = 6 + 3.5$
Properties of Conditional Expectation

Linearity:

- \[ E[\sum_i Y_i | X = x] = \sum_i E[Y_i | X = x] \]
- Proof is same as for unconditional expectation (previous lecture)

Marginalisation:

- \[ E[X] = \sum_y E[X | Y = y] P(Y = y) \]
- Proof: Recall \( E[X | Y = y] = \sum_x x P(X = x | Y = y) \) and \( P(X = x) = \sum_y P(X = x | Y = y) P(Y = y) \) so

\[
\sum_y E[X | Y = y] P(Y = y) = \sum_y \sum_x x P(X = x | Y = y) P(Y = y) \\
= \sum_x \sum_y P(X = x \text{ and } Y = y) \\
= \sum_x x P(X = x) = E[X]
\]
Example (Revisited)

A server has 32GB of memory. Suppose the memory usage of a job is 0.5GB with probability 0.5 and 1GB with probability 0.5, and that the memory usage of different jobs is independent.

- Let $X_i$ be memory usage of $i$th job.
  
  $E[X_i] = 0.5 \times 0.5 + 1 \times 0.5 = 0.75$.

- Number $N$ of jobs is random, so we need to calculate $E[\sum_{i=1}^{N} X_i]$.
  
  By marginalisation we have,

  $$E[\sum_{i=1}^{N} X_i] = E[\sum_{i=1}^{1} X_i|N = 1]P(N = 1) + E[\sum_{i=1}^{2} X_i|N = 2]P(N = 2) + \ldots$$

  $$= E[\sum_{i=1}^{1} X_i]P(N = 1) + E[\sum_{i=1}^{2} X_i]P(N = 2) + \ldots$$

  $$= \sum_{i=1}^{1} E[X_i]P(N = 1) + \sum_{i=1}^{2} E[X_i]P(N = 2) + \ldots$$

  $$= 0.75P(N = 1) + 2 \times 0.75P(N = 2) + 3 \times 0.75P(N = 3) + \ldots$$
Website Example

Say we have a website:

- Random variable $X$ is the number of visitors in one day with $E[X] = \mu_X$
- $Y_i$ is the number of minutes spent by visitor $i$, with $E[Y_i] = \mu_Y$
- $X$ and $Y_i$ are independent
- Total time spent by visitors in one day is $W = \sum_{i=1}^{X} Y_i$. What is $E[W]$?

$$E[W] = \sum_{x} E[W|X = x]P(X = x)$$

$$E[W|X = x] = E[\sum_{i=1}^{X} Y_i|X = x] = \sum_{i=1}^{X} E[Y_i|X = x]$$

$$= \sum_{i=1}^{X} E[Y_i] = x\mu_Y$$

So

$$E[W] = \sum_{x} x\mu_Y P(X = x) = \mu_Y \sum_{x} xP(X = x) = \mu_Y \mu_X$$
Making predictions

We observe random variable $X$.

- Want to make prediction about $Y$
- E.g. $X =$ stock price at 9am, $Y =$ stock price at 10am
- Use $E[Y|X]$
- More on this soon ...