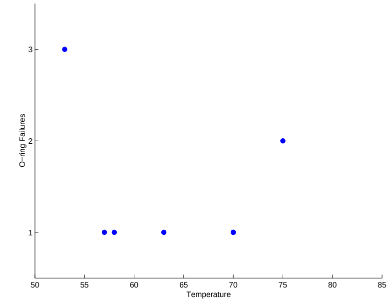
Could the Challenger disaster have been prevented with data analysis?

On January 28, 1986, the space shuttle Challenger exploded in the early stages of its flight. The cause of the explosion was a failure of O-rings, seals on the booster rockets that prevent hot gases from escaping the rockets. When the O-rings failed, one booster rocket caught on fire. The fire eventually burned a hole in the main fuel tank rocket, allowing elements to mix together to produce the catastrophic explosion.

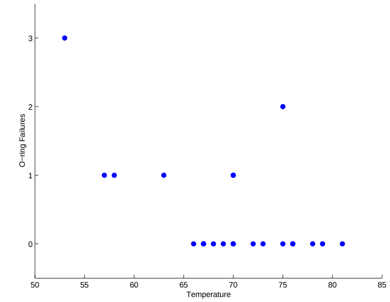
The ambient temperature was 36 degrees on the morning of the launch. Some of the NASA's scientists worried that the O-rings may not perform well in low temperatures; they had no experience with launches in such temperatures.

To aid their decision-making process, the NASA scientists examined data from previous shuttle flights. The graphed the number of O-rings failures versus the ambient temperature for all flights where there was at least one O-ring failure. They obtain these data by examining the booster rockets that fall to the Earth after being jettisoned during a flight.

Plot of number of O-ring failures vs temperature, for flights with at least one O-ring failure



Plot of number of O-ring failures vs temperature, using all shuttle flights



Concepts similar to Univariate r.v.

- Joint distribution
- Expectation

New concepts related with Bivariate r.v.

- Marginal distribution
- Conditional distribution
- Independent r.v.
- Covariance, correlation coefficient

Titanic: Class by Survival

- Joint probability distribution

	1st	2nd	3rd	Crew	
Dead	122 (5%)	167 (8%)	528 (24%)	696 (31%)	**1513** (68%)
Alive	203 (9%)	118 (5%)	178 (8%)	212 (10%)	**711** (32%)
	325	**285**	**706**	**908**	**2224**

- Marginal probabilities

  $\text{P(Alive)} = 9\% + 5\% + 8\% + 10\%$

- Conditional probabilities

  $\text{P(Alive} \mid \text{1st)} = \frac{203}{325} = \frac{9\%}{5\%+9\%}$

Bivariate Distribution

Discrete Random Variables

- The **joint probability distribution** for two discrete random variables, x and y, is a function $p(x, y)$, if
  - $p(x_0, y_0)$ is equal to the probability of the event $\{x = x_0 \text{ AND } y = y_0\}$
  - $0 \leq p(x, y) \leq 1$
  - $\sum_x \sum_y p(x, y) = 1$.
- The **marginal distributions** of x and y are equal to

$$\begin{aligned} p_x(x_0) &= \sum_y p(x_0, y) \\ p_y(y_0) &= \sum_x p(x, y_0) \end{aligned}$$

- The **conditional distribution** for x given that $y = y_0$ is

$$p(x \mid y_0) = \frac{p(x, y_0)}{p_y(y_0)} = \frac{p(x, y_0)}{\sum_x p(x, y_0)}$$

Continuous Random Variables

- The **joint probability density** function for two continuous r.v. x and y is a function $f(x, y)$, if
  - $f(x, y) \geq 0$ for all values of x and y
  - $\int_{-\infty}^{\infty} \int_{-\infty}^{\infty} f(x, y) dx dy = 1$,
  - $P((x, y) \in R) = \iint_R f(x, y) dx dy$ where R is any region in the (x, y) plane. Specially

$$P(a \leq x \leq b, c \leq y \leq d) = \int_c^d \int_a^b f(x, y) dx dy$$

for all constants a, b, c and d.

- **Marginal Densities :**

$$f_x(x) = \int_{-\infty}^{\infty} f(x,y)dy$$
$$f_y(y) = \int_{-\infty}^{\infty} f(x,y)dx$$

- **Conditional Densities :**

$$f_{x|y}(x \mid y) = \frac{f(x,y)}{f_y(y)} = \frac{f(x,y)}{\int_{-\infty}^{\infty} f(x,y)dx}$$
$$f_{y|x}(y \mid x) = \frac{f(x,y)}{f_x(x)} = \frac{f(x,y)}{\int_{-\infty}^{\infty} f(x,y)dy}$$

Expectation

The expected value of a function of random variable x and y, $g(x,y)$, is defined to be

$$\mathbb{E}[g(x,y)] = \sum_y \sum_x g(x,y)p(x,y)$$

$$\mathbb{E}[g(x,y)] = \iint g(x,y)f(x,y)\ dxdy$$

- $\mathbb{E}(c) = c$

- $\mathbb{E}[cg(x,y)] = c\mathbb{E}[g(x,y)]$

- $\mathbb{E}[g_1(x,y)+g_2(x,y)] = \mathbb{E}[g_1(x,y)]+\mathbb{E}[g_2(x,y)]$

Example 11-1: A bank operates both a drive-up facility and a walk-up window. On a randomly selected day, let x = the proportion of time that the drive-up facility is in use and y = the proportion of them that the walk-up window is in use. Suppose the joint pdf for (x, y) is given by

$$f(x,y) = \begin{cases} \frac{6}{5}(x+y^2) & 0 \le x, y \le 1 \\ 0 & \text{otherwise} \end{cases}$$

- Verify that this is a legitimate pdf.

- Find the probability that neither facility is busy more than one-quarter of the time. (.0109)

- Find the marginal distributions for x and y.

- Find the conditional pdf of y given that $x = .8$. $((24 + 30y^2)/34, 0 < y < 1)$

- Find the probability that the walk-in facility is busy at most half the time given that $x = .8$. (.39)

- Find the expected proportion of time that the walk-in facility is busy given that $x = .8$ (**conditional expectation**). (.574)

Example 11-2: Let x denote the time until a computer server connects to your machine (in milliseconds) and let y denote the time until the server authorizes you as a valid user (in milliseconds). Both x and y measures the wait from a common starting time and $x < y$. Assume that the joint pdf for x and y is

$$f(x, y) = 6 \times 10^{-6} exp(-0.001x - 0.002y), \quad x < y$$

- Verify that it is a legitimate pdf.

- Find the probability that $x < 1000$ and $y < 2000$. (.915)