Chapter 5. Vector random variables

- A **vector random variable** $\mathbf{X} = (X_1, X_2, \ldots, X_n)$ is a collection of random numbers with probabilities assigned to outcomes.
- $\mathbf{X}$ can also be called a **multivariate random variable**.
- The case with $n = 2$ we call a **bivariate random variable**.
- Saying X and Y are **jointly distributed random variables** is equivalent to saying (X, Y) is a bivariate random variable.
- Vector random variables let us model relationships between quantities.

Example: midterm and final scores

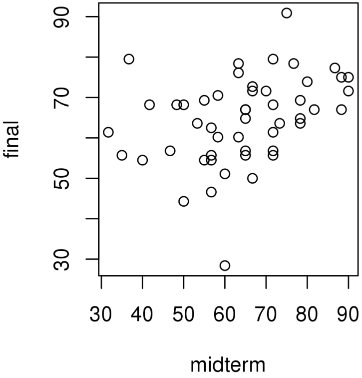
- We will look at the anonymized test scores for a previous course.

```
download.file(destfile="grades.txt",
  url="https://ionides.github.io/401f18/01/grades.txt")
```

```
# Anonymized scores for a random subset of 50 students
"total" "final" "quiz" "hw" "midterm"
"1" 71.6 56.8 95 97 71.7
"2" 70.1 60.2 85 99 58.3
```

- A probability model lets us answer a question like, "What is the probability that someone gets at least 70% in both the midterm and the final"

```
x <- read.table("grades.txt")
plot(final~midterm,data=x)
```



The bivariate normal distribution and covariance

- Let $X \sim \mathrm{normal}(\mu_X, \sigma_X)$ and $Y \sim \mathrm{normal}(\mu_Y, \sigma_Y)$.
- If X and Y are bivariate random variables we need another parameter to describe their dependence. If X is big, does Y tend to be big, or small, or does the value of X make no difference to the outcome of Y?
- This parameter is the **covariance**, defined to be

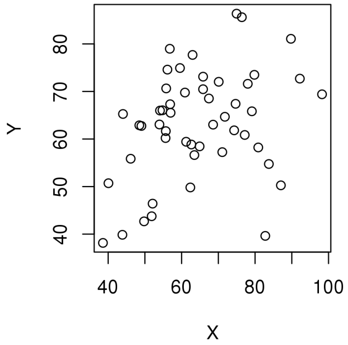
$$\mathrm{Cov}(X, Y) = \mathrm{E}\Big[\big(X - \mathrm{E}[X]\big)\big(Y - \mathrm{E}[Y]\big)\Big]$$

- The parameters of the bivariate normal distribution in matrix form are the **mean vector** $\boldsymbol{\mu} = (\mu_X, \mu_Y)$ and the **variance/covariance matrix**,

$$\mathbb{V} = \begin{bmatrix} \mathrm{Var}(X) & \mathrm{Cov}(X, Y) \\ \mathrm{Cov}(Y, X) & \mathrm{Var}(Y) \end{bmatrix}$$

- In R, the `mvtnorm` package lets us simulate the bivariate and multivariate normal distribution. It uses the vector and matrix form for the parameters.

Experimenting with the bivariate normal distribution




```
library(mvtnorm)
mvn <- rmvnorm(n=50,
  mean=c(X=65,Y=65),
  sigma=matrix(
    c(200,100,100,150),
    2,2)
)
plot(Y~X,data=mvn)
```

- We write $(X, Y) \sim \text{MVN}(\boldsymbol{\mu}, \mathbb{V})$, where MVN is read "multivariate normal".

Question 5.1. What are μ_X, μ_Y, $\text{Var}(X)$, $\text{Var}(Y)$, and $\text{Cov}(X, Y)$ for this example?

- The covariance of X and Y
- The **sample covariance** of n pairs of measurements $(x_1, y_1), \ldots, (x_n, y_n)$ is

$$\mathrm{cov}(\mathbf{x}, \mathbf{y}) = \frac{1}{n-1} \sum_{i=1}^{n} (x_i - \bar{x})(y_i - \bar{y})$$

where $\bar{x}$ and $\bar{y}$ are the sample means of $\mathbf{x} = (x_1, \ldots, x_n)$ and $\mathbf{y} = (y_1, \ldots, y_n)$.