Engineering Computation

Semester 2, 2021

Dr Michael Dallaston

Tarang Janawalkar

Contents

1   MATLAB Functions

Function Syntax	Function Output
y = **sin**(x)	Sine with x in radians.
y = **sind**(x)	Sine with x in degrees.
y = **asin**(x)	Arcsine with y in radians.
y = **exp**(x)	e^x.
y = **log**(x)	$\ln(x)$.

Table 1: Common mathematical functions in MATLAB.

All the above functions are element-wise.

Function Syntax	Function Output(s)
A = **zeros**(m, n)	Creates an $m \times n$ matrix containing zeros.
A = **ones**(m, n)	Creates an $m \times n$ matrix containing ones.
I = **eye**(m)	Creates an $m \times m$ identity matrix.
a = **linspace**(a, b, x)	Creates an evenly spaced vector with bounds $[a, b]$.
y = **length**(A)	The largest dimension of A.
[m, n] = **size**(A)	The dimensions of A.
y = **min**(a)	The minimum value in the vector a.
y = **max**(a)	The maximum value in the vector a.

Table 2: Matrices and arrays in MATLAB.

When manipulating matrices, *****, **^**, perform matrix operations, while prepending an operator with a dot (**.**) performs an element-wise operation.

1.1   Plotting

Function Syntax	Function Output(s)
plot(x, y)	Plots given x and y coordinate vectors.
fplot(@f, [a, b])	Plots the anonymous function over the domain $[a, b]$.
title('string')	Adds title to current plot.
xlabel('string')	Adds x-axis label to current plot.
ylabel('string')	Adds y-axis label to current plot.
legend('string1',...)	Adds legend to plot.
figure	Creates a new figure.

Table 3: Plotting in MATLAB.

1.2   Operations in MATLAB

1.2.1   Conditional Operations

3

```
if expression
    statements
else if expression
    statements
else
    statements
end
```

Code inside an `if` statement only executes if the expression is true. Note that only one branch will execute depending on which expression is true.

1.2.2 Iterative Operations

```
while expression
    statements
end
```

Statements inside a `while` loop repeat while the expression is true.

```
for index = values
    statements
end
```

Statements inside a `for` loop execute a specific number of times, based on the length of `values`.

2 Differential Equations

Definition 2.0.1. A differential equation is an equation that involves the derivatives of a function as well as the function itself. An ordinary differential equation (ODE) is a differential equation of a function with only one independent variable.

2.1 Electrical Systems

Theorem 2.1.1 (VI Relationship between Resistors).

$$v = iR$$

Theorem 2.1.2 (VI Relationship between Inductors).

$$v = L\frac{\mathrm{d}i}{\mathrm{d}t}$$

Theorem 2.1.3 (VI Relationship between Capacitors).

$$i = C\frac{\mathrm{d}v}{\mathrm{d}t}$$

Theorem 2.1.4 (Kirchhoff's Voltage Law). *The sum of all voltages around a loop equals zero.*

$$\sum v_{\mathrm{loop}} = 0$$

Theorem 2.1.5 (Kirchhoff's Current Law). *The sum of all currents into a node equals zero.*

$$\sum i_{\mathrm{node}} = 0$$

2.2   Mechanical Systems

Theorem 2.2.1 (Newton's Second Law)**.**

$$\sum F = \frac{\mathrm{d}p}{\mathrm{d}t}$$

where $p = mv$*.*

Theorem 2.2.2 (Thrust Force)**.**

$$F_T = (c - v)d_f$$

Theorem 2.2.3 (Force of Gravity)**.**

$$F_g = mg$$

Theorem 2.2.4 (Force of a Spring)**.**

$$F = -kx$$

3   First-Order Ordinary Differential Equations

3.1   Separable ODEs

$$\frac{\mathrm{d}y}{\mathrm{d}t} = F(y,\,t)$$

1. Rewrite the equation in the form: $f(y)\,\mathrm{d}y = g(t)\,\mathrm{d}t$.

2. Integrate both sides: $\int f(y)\,\mathrm{d}y = \int g(t)\,\mathrm{d}t$.

3. Rearrange for the explicit form of $y(t)$.

3.2   Linear ODEs

Let $P = P(t)$, $Q = Q(t)$ and $\mu = \mu(t)$

$$\frac{\mathrm{d}y}{\mathrm{d}t} + Py = Q$$

1. Determine the integrating factor: $\mu = \exp\left(\int P\,\mathrm{d}t \right)$.

2. Solve:
$$y = \frac{1}{\mu}\left(\int Q\mu\,\mathrm{d}t + C \right)$$

Proof. To solve a first-order linear differential equation, determine an integrating factor $\mu = \mu(t)$ such that

$$P\mu = \frac{\mathrm{d}\mu}{\mathrm{d}t} \tag{1}$$

5

Multiplying the equation by μ gives

$$\mu\frac{\mathrm{d}y}{\mathrm{d}t} + P\mu y = Q\mu$$

$$\mu\frac{\mathrm{d}y}{\mathrm{d}t} + \frac{\mathrm{d}\mu}{\mathrm{d}t}y = Q\mu$$

$$\frac{\mathrm{d}}{\mathrm{d}t}(\mu y) = Q\mu$$

$$\int \frac{\mathrm{d}}{\mathrm{d}t}(\mu y)\,\mathrm{d}t = \int Q\mu\,\mathrm{d}t$$

$$\mu y = \int Q\mu\,\mathrm{d}t$$

$$y = \frac{1}{\mu}\left(\int Q\mu\,\mathrm{d}t + C\right)$$

To determine μ we can rearrange Equation 1 into

$$P = \frac{1}{\mu}\frac{\mathrm{d}\mu}{\mathrm{d}t}$$

By recognition, this is the derivative of the natural logarithm of μ with respect to t.

$$P = \frac{\mathrm{d}}{\mathrm{d}t}(\ln(\mu))$$

$$\int P\,\mathrm{d}t = \int \frac{\mathrm{d}}{\mathrm{d}t}(\ln(\mu))\,\mathrm{d}t$$

$$\int P\,\mathrm{d}t = \ln(\mu)$$

$$\mu = \exp\left(\int P\,\mathrm{d}t\right)$$

$\square$

3.3   Solution using Linearisation

A function can be linearised by using its 1st degree Taylor polynomial near a.

$$f(x) \approx f(a) + f'(a)(x-a) + \mathcal{O}(x^2)$$

This new polynomial can be substituted to form a linear ODE, which can be solved using an integrating factor.

4 Second-Order Ordinary Differential Equations

4.1 Constant Coefficient Linear ODEs

$$a\frac{\mathrm{d}^2 y}{\mathrm{d}t^2} + b\frac{\mathrm{d}y}{\mathrm{d}t} + cy = Q(t)$$

where a, b, c are constants.

4.2 Linearity of Solutions

Theorem 4.2.1 (Principle of Superposition)**.** *As the given ODE is linear, if $y_1(t)$ is a solution to the equation*

$$a\frac{\mathrm{d}^2 y_1}{\mathrm{d}t^2} + b\frac{\mathrm{d}y_1}{\mathrm{d}t} + cy_1 = Q_1(t)$$

and $y_2(t)$ is a solution to

$$a\frac{\mathrm{d}^2 y_2}{\mathrm{d}t^2} + b\frac{\mathrm{d}y_2}{\mathrm{d}t} + cy_2 = Q_2(t)$$

then for the function $y = c_1 y_1 + c_2 y_2$

$$a\frac{\mathrm{d}^2 y}{\mathrm{d}t^2} + b\frac{\mathrm{d}y}{\mathrm{d}t} + cy = c_1 Q_1(t) + c_2 Q_2(t)$$

where c_1 and c_2 are constants.

4.3 Homogeneous ODEs

Definition 4.3.1. A homogeneous ODE has $Q(t) = 0$, which gives

$$a\frac{\mathrm{d}^2 y}{\mathrm{d}t^2} + b\frac{\mathrm{d}y}{\mathrm{d}t} + cy = 0$$

This differential equation has a solution of the form:

$$y_H = e^{rt}$$

4.4 Characteristic Equation

By making the substitution $y = e^{rt}$, we get

$$\begin{aligned}
a\frac{\mathrm{d}^2 y_H}{\mathrm{d}t^2} + b\frac{\mathrm{d}y_H}{\mathrm{d}t} + cy_H &= 0 \\
ar^2 e^{rt} + bre^{rt} + ce^{rt} &= 0 \\
(ar^2 + br + c)e^{rt} &= 0 \\
ar^2 + br + c &= 0
\end{aligned}$$

This is known as the characteristic or *auxiliary* equation. The next step is to calculate the roots of the equation.

Real Distinct Roots. If $b^2 > 4ac$.

Real Repeated Roots. If $b^2 = 4ac$.

Complex Conjugate Roots. If $b^2 < 4ac$.

4.4.1   Real Distinct Roots

Given r_1 and r_2 are real and distinct:

$$y_1(t) = e^{r_1 t} \qquad\qquad y_2(t) = e^{r_2 t}$$

Hence the solution to the homogeneous equation is given by:

$$y_H(t) = c_1 e^{r_1 t} + c_2 e^{r_2 t}$$

4.4.2   Real Repeated Roots

Given r is a repeated root:

$$y_1(t) = e^{rt} \qquad\qquad y_2(t) = te^{rt}$$

Hence the solution to the homogeneous equation is given by:

$$y_H(t) = c_1 e^{rt} + c_2 te^{rt}$$

4.4.3   Complex Conjugate Roots

Given $r_1 = \alpha + \beta i$ and $r_2 = \alpha - \beta i$ are complex conjugates:

$$y_1(t) = e^{r_1 t} \qquad\qquad y_2(t) = e^{r_2 t}$$

Hence the solution to the homogeneous equation is given by:

$$y_H(t) = c_1 e^{\alpha t} \cos(\beta t) + c_2 e^{\alpha t} \sin(\beta t)$$

4.5   Nonhomogeneous ODE

A nonhomogeneous differential equation is of the form

$$a\frac{\mathrm{d}^2 y}{\mathrm{d}t^2} + b\frac{\mathrm{d}y}{\mathrm{d}t} + cy = Q(t)$$

where $Q(t) \neq 0$.

4.6   General Solution of a Nonhomogeneous ODE

Recall that the solutions to any linear ODE are additive, so that if a solution y_P satisfies the nonhomogeneous ODE, and y_H satisfies the homogeneous ODE,

$$y = y_H + y_P$$

must also satisfy the ODE.

4.7   Undetermined Coefficients

To solve for y_P, we substitute a guess, and determined the coefficients from the ODE itself.
    The particular solution will depend on what $Q(t)$ looks like.

$Q(t)$	y_P
a constant	A
nth degree polynomial	$A_0 + A_1 t + \cdots + A_{n-1} t^{n-1} + A_n t^n$
$e^{\alpha t}$	$A e^{\alpha t}$
$\cos{(\alpha t)}$	$A \cos{(\alpha t)} + B \sin{(\alpha t)}$
$\sin{(\alpha t)}$	$A \cos{(\alpha t)} + B \sin{(\alpha t)}$
$\cos{(\alpha t)} + \sin{(\alpha t)}$	$A \cos{(\alpha t)} + B \sin{(\alpha t)}$

Table 4: Particular solutions for undetermined coefficients.

4.8   Special Forms

4.8.1   Product of Forms

If $Q(t)$ is a product of the functions shown above, then the particular solutions are also multiplied
together and any coefficients are simplified.

4.8.2   Sum of Forms

If $Q(t)$ is a sum of the functions shown above, then the particular solutions are also added together.

4.8.3   Linearly Dependent Case

If $Q(t)$ is similar to any homogenous solution, then by *definition* of a homogeneous solution, the
solution will be 0. Hence, y_P must be multiplied by t to ensure that the particular solution is
linearly independent to the homogeneous solutions, in order to form a *fundamental set of solutions*.

4.9   Solving the Particular Solution

1. Solve y_H

2. Find an appropriate form for y_P

3. Ensure that y_P is linearly independent to the homogeneous solutions

4. Substitute y_P into the nonhomogeneous ODE and solve for the undetermined coefficients

5. Find the general solution $y = y_H + y_P$

6. Apply initial conditions

5   Systems of Ordinary Differential Equations

A first-order system of differential equations has the form

$$
\begin{cases}
x_1' = a_{11}x_1 + a_{12}x_2 + \cdots + a_{1n}x_n \\
x_2' = a_{21}x_1 + a_{22}x_2 + \cdots + a_{2n}x_n \\
\vdots \qquad \vdots \qquad \vdots \qquad\qquad \vdots \\
x_n' = a_{n1}x_1 + a_{n2}x_2 + \cdots + a_{nn}x_n
\end{cases}
$$

where $x_1 = x_1(t)$, $x_2 = x_2(t)$, ..., $x_n = x_n(t)$ are the functions to be determined. In matrix form, the system can be written as

$$
\frac{\mathrm{d}}{\mathrm{d}t}
\begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{bmatrix}
=
\begin{bmatrix}
a_{11} & a_{12} & \cdots & a_{1n} \\
a_{21} & a_{22} & \cdots & a_{2n} \\
\vdots & \vdots & & \vdots \\
a_{n1} & a_{n2} & \cdots & a_{nn}
\end{bmatrix}
\begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{bmatrix}
$$

$$
\mathbf{x}' = \boldsymbol{A}\mathbf{x}
$$

5.1   Higher-Order ODEs

A higher-order linear differential equation can be solved by first converting it to a first-order linear system. Consider the nth-order homogeneous differential equation

$$
y^{(n)} + a_1 y^{(n-1)} + \cdots + a_{n-1}y' + a_n y = 0
$$

Let

$$
x_1 = y
$$
$$
x_2 = y'
$$
$$
\vdots
$$
$$
x_n = y^{(n-1)}
$$

so that $\boldsymbol{x} = \begin{bmatrix} x_1 & x_2 & \cdots & x_n \end{bmatrix}^\top$. Then the differential equation can be expressed as the following first-order linear system of differential equations

$$
\frac{\mathrm{d}}{\mathrm{d}t}
\begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{bmatrix}
=
\begin{bmatrix}
0 & 1 & 0 & \cdots & 0 \\
0 & 0 & 1 & \cdots & 0 \\
\vdots & \vdots & \vdots & \ddots & \vdots \\
0 & 0 & 0 & \cdots & 1 \\
-a_n & -a_{n-1} & -a_{n-2} & \cdots & -a_1
\end{bmatrix}
\begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{bmatrix}
$$

5.2   Solution Form

Like the homogeneous case, we will guess a solution of the form

$$
\mathbf{x} = \mathbf{q}e^{\lambda t}
$$

which allows for the following substitution

$$\lambda \mathbf{q} e^{\lambda t} = \boldsymbol{A} \mathbf{q} e^{\lambda t}$$
$$\boldsymbol{A} \mathbf{q} e^{\lambda t} - \lambda \mathbf{q} e^{\lambda t} = \mathbf{0}$$
$$\left(\boldsymbol{A} - \lambda \boldsymbol{I} \right) \mathbf{q} e^{\lambda t} = \mathbf{0}$$
$$\left(\boldsymbol{A} - \lambda \boldsymbol{I} \right) \mathbf{q} = \mathbf{0}$$

This equation has the trivial solution $\mathbf{q} = \mathbf{0}$, however for a fundamental set of solutions, we must let $\boldsymbol{A} - \lambda \boldsymbol{I}$ be singular.

5.2.1   Characteristic Equation

To determine the eigenvalues λ of the matrix $\mathbf{A}$, we must solve the characteristic equation associated with the system of ODEs. Namely,
$$\det \left(\boldsymbol{A} - \lambda \boldsymbol{I} \right) = 0$$

These eigenvalues can then be used to solve the eigenvectors of $\boldsymbol{A}$

5.3   Solving a System of ODEs

1. Model the system of ODEs in the form $\mathbf{x}' = \boldsymbol{A} \mathbf{x}$

2. Solve the characteristic equation for the eigenvalues of $\boldsymbol{A}$

3. Solve the corresponding eigenvectors of $\boldsymbol{A}$ by solving $\left(\boldsymbol{A} - \lambda \boldsymbol{I} \right) \mathbf{q} = \mathbf{0}$

4. Write the general solution: $\mathbf{x} = c_1 \mathbf{q}_1 e^{\lambda_1 t} + c_2 \mathbf{q}_2 e^{\lambda_2 t}$

5. Apply initial conditions to solve c_1 and c_2

6   Probability

Definition 6.0.1 (Random Variables)**.** A random variable X is a measurable variable that doesn't hold a definitive value.

Definition 6.0.2 (Discrete Random Variables)**.** A discrete random variable X has a countable number of possible values.

Definition 6.0.3 (Continuous Random Variables)**.** A continuous random variable X can take all values in a given interval.

Definition 6.0.4 (Probability)**.** Probability is used to mean the chance that a particular event (or set of events) will occur, expressed on a linear scale from 0 to 1. The probability of the random variable X taking the value x is denoted

$$\Pr \left(X = x \right)$$

Definition 6.0.5 (Sample Space)**.** Let the set of all possible outcomes of a random variables be called the sample space, denoted Ω, of that random variable.

Let X be a random variable that can take on values $x \in \Omega$. Then for all $x \in \Omega$ there is an associated probability $p(x)$, such that

$$\forall x \in \Omega : 0 < p(x) \leq 1$$

$$\sum p(x) = 1.$$

6.1   Events

Definition 6.1.1 (Events)**.** An event A is a set of individual outcomes within Ω. Then for some event $A \subset \Omega$, the probability is given by

$$\Pr(A) = \sum_{x \in A} p(x).$$

The complementary event, denoted A^C (also $\overline{A}$) is the set of all outcomes within the sample space that are not within A.

$$\Pr(A^C) = 1 - \Pr(A).$$

Definition 6.1.2 (Combination of Events)**.** Events can be combined with the two logical connectors AND and OR, which are equivalent to the intersection ($\cap$) and union ($\cup$) of set.

Definition 6.1.3 (Mutually Exclusive Events)**.** If two events have no possible outcomes in common, they are mutually exclusive or disjoint events.

$$A \cap B = \varnothing.$$

It follows that

$$\Pr(A \cap B) = 0.$$

Theorem 6.1.1 (Probability of Union)**.**

$$\Pr(A \cup B) = \Pr(A) + \Pr(B) - \Pr(A \cap B)$$
$$= 1 - \Pr(A^C \cap B^C)$$

Definition 6.1.4 (Independent Events)**.** Two events are independent if the outcome of one event has no influence on the outcome of the other. For these cases, the joint probability is given by

$$\Pr(A \cap B) = \Pr(A)\Pr(B).$$

6.2   Dependent Events

Two events are dependent if the outcome of one event influences the outcome of the other.

Definition 6.2.1 (Conditional Probability)**.** In the case of dependent events, we must use conditional probability concepts in calculating joint probabilities. The probability of A given that the event B has occurred is

$$\Pr(A \mid B) = \frac{\Pr(A \cap B)}{\Pr(B)}$$

Theorem 6.2.1 (Total Probability). *If B is a sample space of disjoint events B_1, B_2, ..., B_n, then $A \cap B_1$, $A \cap B_2$, ..., $A \cap B_N$ are also disjoint, and*

$$A = (A \cap B_1) \cup (A \cap B_2) \cup \cdots \cup (A \cap B_n)$$

This gives

$$\Pr(A) = \sum_{i=1}^{n} \Pr(A \cap B_i)$$
$$= \sum_{i=1}^{n} \Pr(A \mid B_i)\Pr(B_i)$$

Theorem 6.2.2 (Bayes' Theorem). *Using the commutativity of intersections, the rule for conditional probability gives*

$$\Pr(A \cap B) = \Pr(A \mid B)\Pr(B) = \Pr(B \mid A)\Pr(A)$$

Therefore

$$\Pr(B \mid A) = \frac{\Pr(A \mid B)\Pr(B)}{\Pr(A)}$$

7   Probability Distributions

Definition 7.0.1 (Discrete Random Variables). A discrete random variable has countably many outcomes.

Definition 7.0.2 (Continuous Random Variables). A continuous random variable can take an infinite number of individual outcomes.

Definition 7.0.3 (Probability Distributions). The probabilities of random variables make up a probability distribution.

For discrete random variables, the distribution is described with a Probability Mass Function (PMF)

$$p(x) = \Pr(X = x)$$

For continuous variables, the distribution is described with a Probability Density Function (PDF) and the associated Cumulative Distribution Function (CDF).

Here, probabilities are represented by areas under the PDF:

$$\Pr(x_1 \leq X \leq x_2) = \int_{x_1}^{x_2} f(u)\,\mathrm{d}u$$

and the CDF is defined as

$$F(x) = \Pr(X \leq x) = \int_{-\infty}^{x} f(u)\,\mathrm{d}u\,.$$

Note that $f(x)$ is a valid PDF provided

$$f(x) \geq 0 : \forall x \quad \text{and} \quad \int_{-\infty}^{\infty} f(u)\,\mathrm{d}u = 1$$

while $F(x)$ is a valid CDF if:

1. F is a non-decreasing right continuous function

2. $\lim_{x \to -\infty} F(x) = 0$ and $\lim_{x \to \infty} F(x) = 1$

7.1 Summary Statistics

Definition 7.1.1 (Expectation)**.** The expected value $\mathrm{E}\,(X)$, of a random variable is the average outcome that could be expected from an infinite number of observations of that variable. This is also known as the mean of the variable, denoted μ.

$$\mathrm{E}\,(X) = \begin{cases} \sum_{\Omega} x p(x) & \text{for discrete variables} \\ \int_{\Omega} x f(x)\,\mathrm{d}x & \text{for continuous variables} \end{cases}$$

Definition 7.1.2 (Variance)**.** The variance $\mathrm{Var}\,(X)$, of a random variable is a measure of spread of the distribution (defined as the average squared distance of each value from the mean). $\mathrm{Var}\,(X)$ is also denoted as σ^2.

$$\mathrm{Var}\,(X) = \begin{cases} \sum_{\Omega} (x - \mu)^2 p(x) & \text{for discrete variables} \\ \int_{\Omega} (x - \mu)^2 p(x) f(x)\,\mathrm{d}x & \text{for continuous variables} \end{cases}$$
$$= \mathrm{E}\,(X^2) - \mathrm{E}\,(X)^2$$

Definition 7.1.3 (Standard Deviation)**.** The standard deviation is defined as

$$\sigma = \sqrt{\mathrm{Var}\,(X)}$$

7.1.1 General Linear Combinations

For a simple linear function of a random variable

$$\mathrm{E}\,(aX \pm b) = a\,\mathrm{E}\,(X) \pm b$$
$$\mathrm{Var}\,(aX \pm b) = a^2\,\mathrm{Var}\,(X)$$

For a general linear combination of two random variables,

$$\mathrm{E}\,(aX \pm bY) = a\,\mathrm{E}\,(X) \pm b\,\mathrm{E}\,(Y)$$
$$\mathrm{Var}\,(aX \pm bY) = a^2\,\mathrm{Var}\,(X) + b^2\,\mathrm{Var}\,(Y) \pm 2ab\,\mathrm{Cov}\,(X,\,Y)$$

Definition 7.1.4 (Covariance)**.** Covariance is the joint variability of two random variables.

$$\mathrm{Cov}\,(X,\,Y) = \mathrm{E}\,(XY) - \mathrm{E}\,(X)\,\mathrm{E}\,(Y)$$
$$= \mathrm{Corr}\,(X,\,Y)\sqrt{\mathrm{Var}\,(X)\,\mathrm{Var}\,(Y)}$$

Definition 7.1.5 (Correlation)**.** The correlation of two random variables is any statistical relationship between those two variables. The correlation $\mathrm{Corr}\,(X,\,Y)$ is usually denoted ρ_{XY} or ρ, and it always satisfies $-1 \leq \rho \leq 1$.

8   Common Probability Distributions

8.1   Bernoulli Random Variables

A Bernoulli random variable takes values according to the probability distribution:

$$X = \begin{cases} 1 & \text{with probability } p \in [0,\ 1] \\ 0 & \text{with probability } 1 - p \end{cases}$$

This distribution has the following mean and variance

$$\mathrm{E}\left(X\right) = p$$
$$\mathrm{Var}\left(X\right) = p\left(1 - p\right)$$

8.2   Binomial Distribution

The Binomial distribution describes a discrete random variable that is the sum of the outcomes of n independent and identically distributed Bernoulli random variables. The PMF for this distribution is given by

$$p(x) = \binom{n}{x} p^x \left(1 - p\right)^{n - x}$$

where

$$\binom{n}{x} = \frac{n!}{x!\left(n - x\right)!}$$

This distribution has the following mean and variance

$$\mathrm{E}\left(X\right) = np$$
$$\mathrm{Var}\left(X\right) = np\left(1 - p\right)$$

8.3   Poisson Distribution

The Poisson distribution describes discrete events that occur randomly at an average rate λ. The PMF for this distribution is given by

$$p(x) = \frac{\mu^x e^{-\mu}}{x!}$$

where $\mu = \lambda t$ is the average number of events occurring in any interval of length t. This distribution has the following mean and variance

$$\mathrm{E}\left(X\right) = \mu$$
$$\mathrm{Var}\left(X\right) = \mu$$

8.4   Uniform Distribution

The uniform distribution describes a continuous random variable which is only non-zero over a finite interval $(a,\ b)$, with every value within this interval being equally likely to occur. The PDF

and CDF for this distribution are given by

$$f(x) = \frac{1}{b-a}$$
$$F(x) = \frac{x-a}{b-a}$$

This distribution has the following mean and variance

$$\mathrm{E}\left(X\right) = \frac{a+b}{2}$$
$$\mathrm{Var}\left(X\right) = \frac{(b-a)^2}{12}$$

8.5   Exponential Distribution

The exponential distribution describes the time between successive events that are occurring under the same conditions as Poisson distributions. The PDF and CDF for this distribution are given by

$$f(x) = \lambda \exp\left(-\lambda x\right)$$
$$F(x) = 1 - \exp\left(-\lambda x\right)$$

This distribution has the following mean and variance

$$\mathrm{E}\left(X\right) = \frac{1}{\lambda}$$
$$\mathrm{Var}\left(X\right) = \frac{1}{\lambda^2}$$

8.6   Normal Distribution

9   Statistical Inference

The aim of statistical inference is to infer the true mean of a numerical variable from sample data.

In order to carry out inference, we must use data that is properly representative of the situation we wish to draw conclusions about.

9.1   Sample Statistics

Let x be a set of sample data, with a sample size n.

Theorem 9.1.1 (Sample Mean)**.** *The sample mean $\overline{X}$ is given by*

$$\overline{X} = \frac{1}{n}\sum_{i=1}^{n} x_i$$

Theorem 9.1.2 (Sample Variance). *The sample variance s^2 is given by*

$$s^2 = \frac{1}{n-1} \sum_{i=1}^{n} \left(x_i - \overline{X} \right)^2$$

$$= \frac{1}{n-1} \left(\sum_{i=1}^{n} x_i^2 - n\overline{X}^2 \right)$$

9.2   Sampling Distribution of $\overline{X}$

The Central Limit theorem states that the distribution of $\overline{X}$ will approach a Normal distribution as $n \to \infty$. In most cases, $n = 30$ is sufficiently large for a Normal distribution to accurately reflect the distribution of $\overline{X}$

Theorem 9.2.1 (Expectation of a Sample Mean).

$$\mathrm{E} \left(\overline{X} \right) = \mu$$

Theorem 9.2.2 (Variance of a Sample Mean).

$$\mathrm{Var} \left(\overline{X} \right) = \frac{\sigma^2}{n}$$

By standardising $\overline{X}$ we get

$$Z = \frac{\overline{X} - \mu}{\sigma/\sqrt{n}} \sim N(0,\ 1)$$

as $n \to \infty$. But as σ is often unknown, and can vary between samples, the Normal distribution may no long apply.

9.3   Student's t-distribution

Definition 9.3.1. When using sampled data, we can use the t-distribution for $\overline{X}$.

$$T = \frac{\overline{X} - \mu}{s/\sqrt{n}} \sim t_{n-1}$$

where t_{n-1} is Student's t-distribution with $n-1$ degrees of freedom.
    The t-distribution converges to the standard Normal distribution as $n \to \infty$.

Values along the horizontal axis of the t-distribution can be denoted by $t_{n-1,\ 1-\alpha/2}$.

$n-1$ represents the degrees of freedom.

$1-\alpha/2$ represents the CDF probability:

$$\mathrm{Pr} \left(T \leq t_{n-1,\ 1-\alpha/2} \right) = 1 - \alpha/2$$

9.4 Confidence Intervals

Definition 9.4.1 (Confidence Intervals). As $\overline{X}$ varies between samples, and is never equal to μ, we can determine a range of values (based on the observation of $\overline{X}$) that is likely to contain the true mean. This range is called an interval estimate or confidence interval for μ.

The $(1 - \alpha)$ confidence interval for μ is

$$CI_{(1-\alpha)} = \overline{X} \pm t_{n-1,\,1-\alpha/2} \frac{s}{\sqrt{n}}$$

Typically we construct $1 - \alpha = 95\,\%$ confidence intervals.

9.5 Hypothesis Testing

Definition 9.5.1 (Hypothesis Tests). In statistics, a hypothesis test allows us to assess how consistent a particular assumption (hypothesis) is with our observed data.

The process of hypothesis test starts with an assumption about the data. This assumption is known as the "null hypothesis", denoted H_0.

We then assess to what extent the sample is inconsistent with the assumption H_0, by determining how likely we are to observe the sample if the null hypothesis is true.

This probability is known as the "p-value" for the test, and indicates how strong the evidence is against H_0.

p > 0.1 provides insufficient evidence to dispute H_0

p < 0.1 provides weak evidence against H_0

p < 0.05 provides some evidence against H_0

p < 0.01 provides strong evidence against H_0

p < 0.001 provides very strong evidence against H_0

The assumption when H_0 is rejected is known as the alternative hypothesis, denoted H_1 or H_A.

To test the true mean μ, we have

$$H_0 : \mu = \mu_0 \quad \text{vs} \quad H_A : \mu \neq \mu_0$$

Suppose a sample mean of $\overline{x}$ from a sample of n observations, then:

$$\Pr\left(|\overline{X} - \mu| > |\overline{x} - \mu_0|\right) = \Pr\left(\left|\frac{\overline{X} - \mu}{s/\sqrt{n}}\right| > \left|\frac{\overline{x} - \mu_0}{s/\sqrt{n}}\right|\right)$$
$$= \Pr\left(|T| > \left|\frac{\overline{x} - \mu_0}{s/\sqrt{n}}\right|\right)$$

where $T = \frac{\overline{X} - \mu}{s/\sqrt{n}} \sim t_{n-1}$.

10   Linear Regression

Definition 10.0.1 (Linear Relationship). Two variables x and y have a linear relationship if they satisfy an equation of the form

$$y = ax + b$$

where a and b are the parameters in the relationship.

To determine estimates for the values a and b, a statistical model must be formulated along with the linear relationship. To do this, we assume that the points $(x_i,\ y_i)$ have been observed with some error and formulate the following statistical model

$$y_i = \beta_0 + \beta_1 x_i + \varepsilon_i$$

where ϵ_i is the error in observing y_i, and β_0 and β_1 are the true values of the parameters.

Definition 10.0.2 (Residual). By re-expressing the equation in terms of ε_i,

$$\varepsilon_i = y_i - \beta_0 - \beta_1 x_i$$

the residual for each observation can be defined as the amount by which y_i differs from the linear model.

To make statistical inferences about various features of this model, the following assumptions must be made about the residuals:

1. The residuals are independent, identically distributed random variables.

2. $\forall i : \mathrm{E}\left(\varepsilon_i\right) = 0$.

3. The residuals are normally distributed.

In summary, we assume $\varepsilon_i \overset{\text{iid}}{\sim} \mathrm{N}\left(0,\ \sigma^2\right)$. The estimated variance of ε_i uses a similar form to the expression shown in the previous section.

$$s^2 = \frac{1}{n-2} \sum_i^n \varepsilon_i^2$$

$$= \frac{1}{n-2} \sum_i^n \left(y_i - \hat{\beta}_0 - \hat{\beta}_1 x_i\right)^2$$

where the hats indicate the estimated values for β_0 and β_1.

10.1   Sample Statistics for y_i

These assumptions also allow us to find the properties of y_i. As each observation is conducted with a known x_i, we can generally assume that it is a constant. Likewise, β_0 and β_1 are also constants, therefore the properties of means and variances tells us,

$$\mathrm{E}\left(y_i\right) = \beta_0 + \beta_1 x_i + \mathrm{E}\left(\varepsilon_i\right) = \beta_0 + \beta_1 x_i$$
$$\mathrm{Var}\left(y_i\right) = \mathrm{Var}\left(\varepsilon_i\right) = \sigma^2$$

hence, $y_i \sim \mathrm{N}\left(\beta_0 + \beta_1 x_i,\ \sigma^2\right)$.

Definition 10.1.1 (Least Squares)**.** To estimate β_0 and β_1, the intercept and slope of the model, we can use the "least squares criterion", that is, we can minimise the sum of the squares of the residuals.

$$\min_{\beta_0, \beta_1} \sum_i^n \varepsilon_i^2 \quad \text{or} \quad \min_{\beta_0, \beta_1} \sum_i^n (y_i - \beta_0 - \beta_1 x_i)^2$$

By differentiating this expression w.r.t. β_0 and β_1, and setting the partial derivatives equal to zero, we reach the following conditions:

$$\sum_i^n (y_i - \beta_0 - \beta_1 x_i) = 0$$

$$\sum_i^n x_i (y_i - \beta_0 - \beta_1 x_i) = 0$$

Proof. Differentiating w.r.t. β_0:

$$\partial_{\beta_0} \left(\sum_i^n (y_i - \beta_0 - \beta_1 x_i)^2 \right) = 0$$

$$\sum_i^n \partial_{\beta_0} (y_i - \beta_0 - \beta_1 x_i)^2 = 0$$

$$\sum_i^n 2 (y_i - \beta_0 - \beta_1 x_i) (-1) = 0$$

$$\sum_i^n (y_i - \beta_0 - \beta_1 x_i) = 0$$

Differentiating w.r.t. β_1:

$$\partial_{\beta_1} \left(\sum_i^n (y_i - \beta_0 - \beta_1 x_i)^2 \right) = 0$$

$$\sum_i^n \partial_{\beta_1} (y_i - \beta_0 - \beta_1 x_i)^2 = 0$$

$$\sum_i^n 2 (y_i - \beta_0 - \beta_1 x_i) (-x_i) = 0$$

$$\sum_i^n x_i (y_i - \beta_0 - \beta_1 x_i) = 0$$

$\square$

This gives the following estimates for β_0 and β_1.

$$\hat{\beta}_1 = \frac{\sum_i^n x_i y_i - n \overline{xy}}{\sum_i^n x_i^2 - n \overline{x}^2}$$

$$\hat{\beta}_0 = \overline{y} - \hat{\beta}_1 \overline{x}$$

where $\overline{x} = \frac{1}{n} \sum_i^n x_i$ and $\overline{y} = \frac{1}{n} \sum_i^n y_i$.

Proof. For $\hat{\beta}_0$:

$$0 = \sum_i^n \left(y_i - \hat{\beta}_0 - \hat{\beta}_1 x_i \right)$$

$$0 = \sum_i^n y_i - \hat{\beta}_0 n - \hat{\beta}_1 \sum_i^n x_i$$

$$\hat{\beta}_0 = \frac{\sum_i^n y_i - \hat{\beta}_1 \sum_i^n x_i}{n}$$

$$\hat{\beta}_0 = \overline{y} - \hat{\beta}_1 \overline{x}$$

For $\hat{\beta}_1$, using $\hat{\beta}_0$:

$$0 = \sum_i^n x_i \left(y_i - \hat{\beta}_0 - \hat{\beta}_1 x_i \right)$$

$$0 = \sum_i^n x_i y_i - \hat{\beta}_0 \sum_i^n x_i - \hat{\beta}_1 \sum_i^n x_i^2$$

$$0 = \sum_i^n x_i y_i - \left(\overline{y} - \hat{\beta}_1 \overline{x} \right) n\overline{x} - \hat{\beta}_1 \sum_i^n x_i^2$$

$$0 = \sum_i^n x_i y_i - n\overline{x}\overline{y} + \hat{\beta}_1 n\overline{x}^2 - \hat{\beta}_1 \sum_i^n x_i^2$$

$$\hat{\beta}_1 \left(\sum_i^n x_i^2 - n\overline{x}^2 \right) = \sum_i^n x_i y_i - n\overline{x}\overline{y}$$

$$\hat{\beta}_1 = \frac{\sum_i^n x_i y_i - n\overline{x}\overline{y}}{\sum_i^n x_i^2 - n\overline{x}^2}$$

$\square$

10.2   Sample Statistics for β_0 and β_1

As the estimators $\hat{\beta}_0$ and $\hat{\beta}_1$ are linear combinations of y_i, they are also random variables. The variance of these estimates is given by:

$$s_{\hat{\beta}_0}^2 = s^2 \frac{\sum_i^n x_i^2}{n \sum_i^n (x_i - \overline{x})^2}$$

$$s_{\hat{\beta}_1}^2 = s^2 \frac{1}{\sum_i^n (x_i - \overline{x})^2}$$

where s^2 is the estimate for $\text{Var}(y_i)$, (or $\text{Var}(\varepsilon_i)$), as shown previously. This leads to the following sampling distributions:

$$\hat{\beta}_0 \sim \text{N}\left(\beta_0,\ s^2_{\hat{\beta}_0}\right)$$

$$\hat{\beta}_1 \sim \text{N}\left(\beta_1,\ s^2_{\hat{\beta}_1}\right)$$

and it follows that

$$\frac{\hat{\beta}_0 - \beta_0}{s_{\hat{\beta}_0}} \sim t_{n-2} \quad \text{and} \quad \frac{\hat{\beta}_1 - \beta_1}{s_{\hat{\beta}_1}} \sim t_{n-2}.$$

Proof. To determine the variance for the estimators $\hat{\beta}_0$ and $\hat{\beta}_1$, we must first recognise the following identities:

$$\sum_i^n (x_i - \overline{x}) = 0 \tag{1}$$

as $\sum_i^n (x_i - \overline{x}) = \sum_i^n x_i - n\overline{x}$. Additionally,

$$\text{Var}\left(\sum_i^n a_i X_i\right) = \sum_i^n a_i^2 \text{Var}(X_i) + 2\sum_{i=1}^n \sum_{j=i+1}^n a_i a_j \text{Cov}(X_i,\ X_j) \tag{2}$$

We can now rewrite the numerator and denominator in $\hat{\beta}_1$:

$$\sum_i^n x_i y_i - n\overline{xy} = \sum_i^n y_i (x_i - \overline{x})$$

$$= \sum_i^n y_i (x_i - \overline{x}) - \sum_i^n \overline{y}(x_i - \overline{x}) \qquad \text{(Using Equation 1)}$$

$$= \sum_i^n (x_i - \overline{x})(y_i - \overline{y})$$

Similarly,

$$\sum_i^n x_i^2 - n\overline{x}^2 = \sum_i^n x_i (x_i - \overline{x})$$

$$= \sum_i^n x_i (x_i - \overline{x}) - \sum_i^n \overline{x}(x_i - \overline{x}) \qquad \text{(Using Equation 1)}$$

$$= \sum_i^n (x_i - \overline{x})^2$$

These two quantities are known as the sum of products and sum of squares, denoted S_{XY} and S_{XX} respectively. This allows us to represent $\hat{\beta}_1$ as:

$$\hat{\beta}_1 = \frac{S_{XY}}{S_{XX}} = \frac{\sum_i^n (x_i - \overline{x})(y_i - \overline{y})}{\sum_i^n (x_i - \overline{x})^2}$$

To simplify calculations regarding the variance, we will write S_{XY} as $\sum_i^n y_i \left(x_i - \overline{x}\right)$.

$$
\begin{aligned}
\operatorname{Var}\left(\hat{\beta}_1\right) = s_{\hat{\beta}_1}^2 &= \operatorname{Var}\left(\frac{\sum_i^n y_i \left(x_i - \overline{x}\right)}{\sum_i^n \left(x_i - \overline{x}\right)^2}\right) \\
&= \frac{1}{\left(\sum_i^n \left(x_i - \overline{x}\right)^2\right)^2} \operatorname{Var}\left(\sum_i^n y_i \left(x_i - \overline{x}\right)\right) \\
&= \frac{1}{\left(\sum_i^n \left(x_i - \overline{x}\right)^2\right)^2} \operatorname{Var}\left(\sum_i^n \left(\beta_0 + \beta_1 x_i + \varepsilon_i\right)\left(x_i - \overline{x}\right)\right) \\
&= \frac{1}{\left(\sum_i^n \left(x_i - \overline{x}\right)^2\right)^2} \operatorname{Var}\left(\sum_i^n \left(\beta_0 + \beta_1 x_i\right)\left(x_i - \overline{x}\right) + \sum_i^n \varepsilon_i \left(x_i - \overline{x}\right)\right) \\
&= \frac{1}{\left(\sum_i^n \left(x_i - \overline{x}\right)^2\right)^2} \operatorname{Var}\left(\sum_i^n \varepsilon_i \left(x_i - \overline{x}\right)\right)
\end{aligned}
$$

Using Equation 2,

$$
s_{\hat{\beta}_1}^2 = \frac{1}{\left(\sum_i^n \left(x_i - \overline{x}\right)^2\right)^2} \left(\sum_i^n \left(x_i - \overline{x}\right)^2 \operatorname{Var}\left(\varepsilon_i\right) + 2\sum_{i=1}^n \sum_{j=i+1}^n \left(x_i - \overline{x}\right)\left(x_j - \overline{x}\right) \operatorname{Cov}\left(\varepsilon_i, \varepsilon_j\right)\right).
$$

The independence of the residuals implies that the covariance of all residuals — $\operatorname{Cov}\left(\varepsilon_i, \varepsilon_j\right) : i \neq j$ — is equal to 0. Therefore

$$
\begin{aligned}
s_{\hat{\beta}_1}^2 &= \frac{1}{\left(\sum_i^n \left(x_i - \overline{x}\right)^2\right)^2} \sum_i^n \left(x_i - \overline{x}\right)^2 \operatorname{Var}\left(\varepsilon_i\right) \\
&= \frac{1}{\left(\sum_i^n \left(x_i - \overline{x}\right)^2\right)^2} \sum_i^n \left(x_i - \overline{x}\right)^2 s^2 \\
&= s^2 \frac{1}{\sum_i^n \left(x_i - \overline{x}\right)^2}
\end{aligned}
$$

We will use similar techniques to determine the variance of $\hat{\beta}_0$.

$$
\operatorname{Var}\left(\hat{\beta}_0\right) = s_{\hat{\beta}_0}^2 = \operatorname{Var}\left(\overline{y} - \hat{\beta}_1 \overline{x}\right)
$$

here we must assume that $\overline{y}$ and $\hat{\beta}_1$ are uncorrelated.

$$s_{\hat{\beta}_0}^2 = \text{Var}\left(\overline{y}\right) + \overline{x}^2 \,\text{Var}\left(\hat{\beta}_1\right)$$

$$= \text{Var}\left(\frac{1}{n}\sum_i^n y_i\right) + \overline{x}^2 s^2 \frac{1}{\sum_i^n (x_i - \overline{x})^2}$$

$$= \frac{1}{n^2}\text{Var}\left(\sum_i^n y_i\right) + s^2 \frac{\overline{x}^2}{\sum_i^n (x_i - \overline{x})^2}$$

$$= \frac{1}{n^2}\text{Var}\left(\sum_i^n (\beta_0 + \beta_1 x_i + \varepsilon_i)\right) + s^2 \frac{\overline{x}^2}{\sum_i^n (x_i - \overline{x})^2}$$

As the residuals are uncorrelated, and β_0, β_1, and x_i are all constants, we can write the variance of the summation as the summation of the variation.

$$s_{\hat{\beta}_0}^2 = \frac{1}{n^2}\sum_i^n \text{Var}\left(\beta_0 + \beta_1 x_i + \varepsilon_i\right) + s^2 \frac{\overline{x}^2}{\sum_i^n (x_i - \overline{x})^2}$$

$$= \frac{1}{n^2}\sum_i^n \text{Var}\left(\varepsilon_i\right) + s^2 \frac{\overline{x}^2}{\sum_i^n (x_i - \overline{x})^2}$$

$$= \frac{1}{n^2}\sum_i^n s^2 + s^2 \frac{\overline{x}^2}{\sum_i^n (x_i - \overline{x})^2}$$

$$= \frac{1}{n^2}n s^2 + s^2 \frac{\overline{x}^2}{\sum_i^n (x_i - \overline{x})^2}$$

$$= s^2 \frac{1}{n} + s^2 \frac{\overline{x}^2}{\sum_i^n (x_i - \overline{x})^2}$$

$$= s^2 \left(\frac{1}{n} + \frac{\overline{x}^2}{\sum_i^n (x_i - \overline{x})^2}\right)$$

$$= s^2 \frac{\sum_i^n (x_i - \overline{x})^2 + n\overline{x}^2}{n\sum_i^n (x_i - \overline{x})^2}$$

$$= s^2 \frac{\sum_i^n (x_i - \overline{x})^2 + \sum_i^n \overline{x}^2}{n\sum_i^n (x_i - \overline{x})^2}$$

$$= s^2 \frac{\sum_i^n \left((x_i - \overline{x})^2 + \overline{x}^2\right)}{n\sum_i^n (x_i - \overline{x})^2}$$

$$= s^2 \frac{\sum_i^n \left(x_i^2 - 2x_i\overline{x} + \overline{x}^2 + \overline{x}^2\right)}{n\sum_i^n (x_i - \overline{x})^2}$$

$$= s^2 \frac{\sum_i^n x_i^2 - 2\overline{x}\sum_i^n (x_i - \overline{x})}{n\sum_i^n (x_i - \overline{x})^2}$$

$$= s^2 \frac{\sum_i^n x_i^2}{n \sum_i^n (x_i - \bar{x})^2} \qquad \text{(Using Equation 1)}$$

$\square$

The process described above for estimating a linear relationship between two observed variables x and y is known as linear regression.

10.3 Linear Regression using MATLAB

Linear regression can be conducted in MATLAB, using the `fitlm` function.

```
>> fitlm(x, y)

Linear regression model:
    y ~ 1 + x1

Estimated Coefficients:
                  Estimate        SE        tStat       pValue

                  --------    --------    ------    ----------

    (Intercept)    0.28667     0.24861    1.1531       0.28216
    x1             1.2024      0.040067   30.011    1.6489e-09


Number of observations: 10, Error degrees of freedom: 8
Root Mean Squared Error: 0.364
R-squared: 0.991,  Adjusted R-Squared: 0.99
F-statistic vs. constant model: 901, p-value = 1.65e-09
```

From this output:

- The `Estimate` column gives the estimates $\hat{\beta}_0$ and $\hat{\beta}_1$ for the true intercept β_0 and true slope β_1.

- The `SE` column gives the corresponding standard errors $s_{\hat{\beta}_0}$ and $s_{\hat{\beta}_1}$.

- The `tStat` column gives the test statistic for testing $H_0 : \beta_i = 0$ vs $H_A : \beta_i \neq 0$.

- The `pValue` column gives the p-value for the corresponding test statistics.

Note the rows (`Intercept`) and `x1` *refer to* $\hat{\beta}_0$ *and* $\hat{\beta}_1$ *respectively.*

10.4 Inferences about the Estimators

This information is useful to make inferences about β_0 and β_1.

10.4.1   Hypothesis Tests for the Estimators

Firstly, we can test the hypothesis

$$H_0 : \beta_1 = 0 \quad \text{vs} \quad H_A : \beta_1 \neq 0$$

The reason for this is that when $\beta_1 = 0$ the linear model collapses to $\mathrm{E}(y_i) = \beta_0$, that is, the y values are randomly distributed around a constant, hence there is no evidence that the variables are related **linearly**.

Just as we used the test statistic to determine the p-value for our hypotheses in the previous section, we can identify the strength of our hypothesis in MATLAB by using the p-value in the output of the `fitlm` function.

Similarly we can formulate a test for β_0:

$$H_0 : \beta_0 = 0 \quad \text{vs} \quad H_A : \beta_0 \neq 0$$

which gives us evidence that the true intercept for the relationship is not 0, however this test is usually not very important.

10.4.2   Confidence Intervals for the Estimators

The confidence interval for the estimators is calculated using the following formula

$$CI_c = \hat{\beta}_0 \pm t_{n-2,\,1-\alpha/2} s_{\hat{\beta}_0}$$

$$CI_c = \hat{\beta}_1 \pm t_{n-2,\,1-\alpha/2} s_{\hat{\beta}_1}$$

where $\hat{\beta}_i$ and $s_{\hat{\beta}_i}$ are provided in the MATLAB output, and $t_{n-2,\,1-\alpha/2}$ can be determined using a calculator or table.

Note that CI_c is the same as $CI_{1-\alpha}$.

10.5   Inferences about the model estimate

With estimates for β_0, β_1, and σ^2 we can also fit estimates for the expected value of each y_i

$$\hat{y}_i = \hat{\beta}_0 + x_i \hat{\beta}_1.$$

Similarly, we can determine a confidence interval for these values using the following statement:

$$CI_c = \hat{y}_i \pm t_{n-2,\,1-\alpha/2} s \sqrt{\frac{1}{n} + \frac{(x_i - \overline{x})^2}{\sum_i^n (x_i - \overline{x})^2}}$$

where $s\sqrt{\frac{1}{n} + \frac{(x_i - \overline{x})^2}{\sum_i^n (x_i - \overline{x})^2}}$ is the standard error for $\hat{y}_i$.

This formula can be used to create a confidence band for the true value of y at any possible x to create the following plot where the confidence level is 99%.
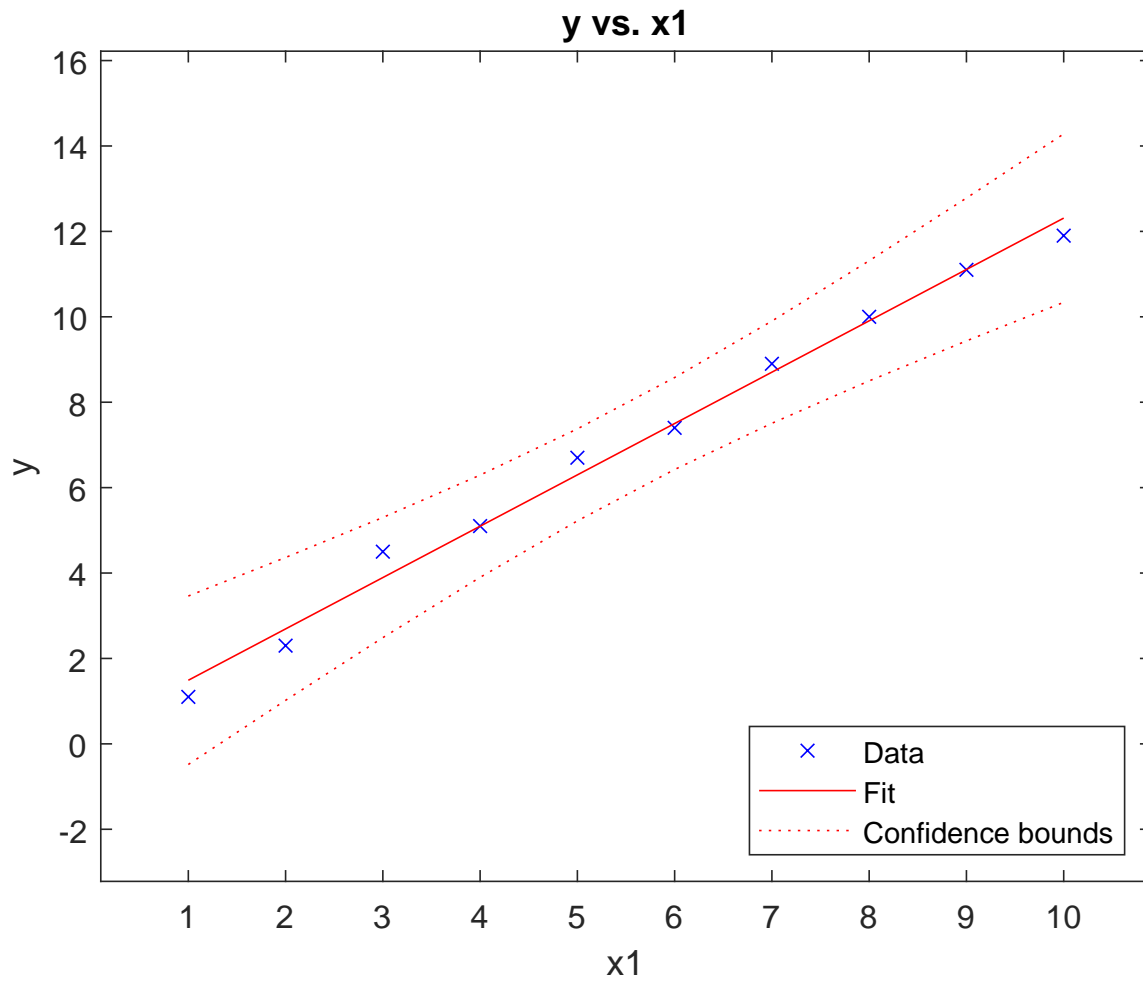
Figure 1: Graph of linear regression model.

Note that:

- The confidence band is narrowest at $x = \overline{x}$.

- More than 95% of observations fall outside the confidence band because that interval is based on the expected value not specific observations.

- As the confidence band widens further away from $\overline{x}$, extrapolation beyond the range of known x values is dangerous.

Proof. To prove the standard error $s_{\hat{y}_i}$, we notice

$$\hat{y}_i = \hat{\beta}_0 + x_i \hat{\beta}_1$$

This will allow us to represent the variance as

$$\text{Var}\left(\hat{y}_i\right) = \text{Var}\left(\hat{\beta}_0 + x_i\hat{\beta}_1\right)$$
$$= \text{Var}\left(\hat{\beta}_0\right) + x_i^2\,\text{Var}\left(\hat{\beta}_1\right) + 2x_i\,\text{Cov}\left(\hat{\beta}_0,\ \hat{\beta}_1\right)$$

Then using the rules for covariance

$$\text{Cov}\left(\hat{\beta}_0,\ \hat{\beta}_1\right) = \text{Cov}\left(\overline{y} - \overline{x}\hat{\beta}_1,\ \hat{\beta}_1\right)$$
$$= \text{Cov}\left(\overline{y},\ \hat{\beta}_1\right) - \text{Cov}\left(\overline{x}\hat{\beta}_1,\ \hat{\beta}_1\right)$$
$$= -\overline{x}\,\text{Cov}\left(\hat{\beta}_1,\ \hat{\beta}_1\right)$$
$$= -\overline{x}\,\text{Var}\left(\hat{\beta}_1\right)$$

Therefore

$$\text{Var}\left(\hat{y}_i\right) = \text{Var}\left(\hat{\beta}_0\right) + x_i^2\,\text{Var}\left(\hat{\beta}_1\right) - 2x_i\overline{x}\,\text{Var}\left(\hat{\beta}_1\right)$$
$$= \text{Var}\left(\hat{\beta}_0\right) + x_i\,\text{Var}\left(\hat{\beta}_1\right)\left(x_i - 2\overline{x}\right)$$
$$= s_{\hat{\beta}_0}^2 + x_i s_{\hat{\beta}_1}^2\left(x_i - 2\overline{x}\right)$$

We can now substitute the known values for the variance from the previous sections.

$$\text{Var}\left(\hat{y}_i\right) = s^2\frac{\sum_i^n x_i^2}{n\sum_i^n\left(x_i - \overline{x}\right)^2} + x_i s^2\frac{1}{\sum_i^n\left(x_i - \overline{x}\right)^2}\left(x_i - 2\overline{x}\right)$$
$$= s^2\left[\frac{\sum_i^n x_i^2}{n\sum_i^n\left(x_i - \overline{x}\right)^2} + \frac{x_i^2 n - 2\overline{x}x_i n}{n\sum_i^n\left(x_i - \overline{x}\right)^2}\right]$$
$$= s^2\frac{\sum_i^n x_i^2 + x_i^2 n - 2\overline{x}x_i n}{n\sum_i^n\left(x_i - \overline{x}\right)^2}$$
$$= s^2\frac{\sum_i^n\left(x_i^2 - 2\overline{x}x_i + 2\overline{x}x_i\right) + x_i^2 n - 2\overline{x}x_i n}{n\sum_i^n\left(x_i - \overline{x}\right)^2}$$
$$= s^2\frac{\sum_i^n\left(x_i^2 - 2\overline{x}x_i\right) + 2\overline{x}\sum_i^n x_i + x_i^2 n - 2\overline{x}x_i n}{n\sum_i^n\left(x_i - \overline{x}\right)^2}$$
$$= s^2\frac{\sum_i^n\left(x_i^2 - 2\overline{x}x_i\right) + 2\overline{x}^2 n + x_i^2 n - 2\overline{x}x_i n}{n\sum_i^n\left(x_i - \overline{x}\right)^2}$$
$$= s^2\frac{\sum_i^n\left(x_i^2 - 2\overline{x}x_i + \overline{x}^2\right) + \overline{x}^2 n + x_i^2 n - 2\overline{x}x_i n}{n\sum_i^n\left(x_i - \overline{x}\right)^2}$$
$$= s^2\frac{\sum_i^n\left(x_i - \overline{x}\right)^2 + n\left(x_i - \overline{x}\right)^2}{n\sum_i^n\left(x_i - \overline{x}\right)^2}$$
$$= s^2\left[\frac{1}{n} + \frac{\left(x_i - \overline{x}\right)^2}{\sum_i^n\left(x_i - \overline{x}\right)^2}\right]$$

$$\square$$

10.6 Coefficient of Determination

The linear regression output also gives us the coefficient of determination labelled R^2.

Definition 10.6.1 (Coefficient of Determination)**.** The coefficient of determination R^2 describes the percentage of the observed variance in y that is explained by the fitted model $\hat{y}$.

$$R^2 = 1 - \frac{RSS}{TSS}$$

where RSS is the regression sum of squares defined $\sum_i^n \left(\hat{y}_i - y_i \right)^2$ and TSS is the total sum of squares defined $\sum_i^n \left(y_i - \overline{y} \right)^2$.

This value is labelled `R-squared` in the MATLAB output.

11 Linear Regression Diagnostics

After fitting a linear model, it is important to check that linearity is the best assumption for the relationship between the variables.

We can do this using a residual plot.

11.1 Summary of Assumptions

During the process of linear regression, we make a number of assumptions:

1. Each observation is independent of other observations

2. The relationship between X and Y is best modelled as a linear relationship

3. The residuals are have a Normal distribution.

4. The variance of the residuals is constant for all observations.