DATA 580

Modeling and Simulation I

Models for More than One Measurement

- Joint, Marginal and Conditional pdfs
- Independence, mathematically and graphically
- Introduction to Predictive Modeling - Simple Linear Regression

Models for More than One Measurement

- **A single measurement is modelled using a continuous random variable X with probability density function $f(x)$, e.g. normal, exponential, etc.**

- **What if there are 2 or more measurements?**

- **Model each measurement with a random variable:** $X_1, X_2, \ldots,$

- **How do we evaluate** $P(a < X_1 < b, c < X_2 < d)$**?**

A Model for Two Independent Measurements

- **A basic probability result says that if random events A and B are independent, then $P(A \textbf{ occurs and } B \textbf{ occurs}) = P(A \textbf{ occurs})P(B \textbf{ occurs})$.**

- **This allows us to say that if the events $\{a < X_1 < b\}$ and $\{c < X_2 < d\}$ are independent, then we can write**

$$P(a < X_1 < b, c < X_2 < d) =$$

$$P(a < X_1 < b)P(c < X_2 < d)$$

$$= \int_a^b f_1(y_1)dy_1 \int_c^d f_2(y_2)dy_2$$

$$= \int_a^b \int_c^d f_1(y_1)f_2(y_2)dy_1dy_2$$

where $f_1(y_1)$ and $f_2(y_2)$ are the density functions for X_1 and X_2.

A Model for Two Independent Measurements

- Set

$$f(y_1, y_2) = f_1(y_1)f_2(y_2).$$

- $f(y_1, y_2)$ is called the joint density function for X_1 and X_2.

$$P(a < X_1 < b, c < X_2 < d) =$$

$$\int_a^b \int_c^d f(y_1, y_2) dy_1 dy_2$$

- If the measurements X_1 and X_2 are not independent, we can still define $f(y_1, y_2)$, the joint density function for X_1 and X_2.

Joint Probability Density Function

- **Properties:**

  1. $f(y_1, y_2) \geq 0$

  2. $\int \int f(y_1, y_2) dy_1 dy_2 = 1$

- **Computation of probabilities:**

$$P(a < X_1 < b, c < X_2 < d) =$$

$$\int_a^b \int_c^d f(y_1, y_2) dy_1 dy_2$$

- **Joint density for more than two measurements ($X_1, X_2, \ldots, X_k$):**

$$f(y_1, y_2, \ldots, y_k)$$

Example.

A machine is used to automatically fill cylinders with propane gas.

* X **= amount of propane in a randomly selected cylinder (moles)**
* T **= temperature (C) at the time of filling**
* **model for X and T – joint density:**

$$f(x,t) = \frac{x + \frac{t}{5} - 13}{5}, \; 10 \le x \le 11, \; 15 \le t \le 20$$

$f(x,t) = 0$ **for other values of x and t**

Verify that $f(x,t)$ is a valid joint density function.

Joint Probability Density Function

Example (cont'd).

We verify the two properties of the joint density function:

1. $f(x,t) \geq 0$ **for all** x, t

2. $\int \int f(x,t)dtdx = 1$:

$$\int_{10}^{11} \int_{15}^{20} \frac{x + t/5 - 13}{5} dtdx =$$

$$\int_{10}^{11} \frac{xt + t^2/10 - 13t}{5} \Big|_{15}^{20} dx =$$

$$\int_{10}^{11} (x - 9.5)dx = 1$$

Joint Probability Density Function

Example (cont'd).

Calculate the probability that the amount of propane is between 10.4 and 10.6 moles, and that the temperature is between 16 and 17 degrees.

$$P(10.4 < X < 10.6, 16 < T < 17) =$$

$$\int_{10.4}^{10.6} \int_{16}^{17} \frac{x + t/5 - 13}{5} dtdx = \int_{10.4}^{10.6} \frac{x - 9.7}{5} dx = .032$$

Joint Probability Density Function

Example 2.

* **Suppose the reliability of an electric motor depends upon two critical components, having lifetimes X and Y which can be modelled with the joint probability density function**

$$f(x,y) = \begin{cases} xe^{-x(1+y)}, & x \geq 0, y \geq 0 \\ 0, & \textbf{otherwise} \end{cases}$$

Find the probability that both components last more than 1 unit of time.

Example 2 (cont'd).

The question asks for the probability that X is greater than 1 and Y is greater than 1:

$$P(X > 1, Y > 1) =$$

$$\int_1^\infty \int_1^\infty x e^{-x(1+y)} dy dx = e^{-2}/2$$

Marginal Probability Density Functions

When we have more than one measurement, the joint density function summarizes the overall model. Each individual random variable still has a probability density function (as discussed earlier), but when in this larger context, the pdf is referred to as a marginal pdf.

- If X_1 and X_2 have joint density function $f(y_1, y_2)$,
  * the density function for X_1 can be determined as

$$f_1(y_1) = \int_{-\infty}^{\infty} f(y_1, y_2) dy_2$$

That is, the pdf of X_1 is obtained by integrating over all possible values of the other variable.

  * the density function for X_2 can be determined as

$$f_2(y_2) = \int_{-\infty}^{\infty} f(y_1, y_2) dy_1$$

That is, the pdf of X_2 is obtained by integrating over all possible values of the other variable.

Marginal Probability Density Functions

Example.

Find the marginal densities of X and T for the propane example:

$$f_X(x) = \int_{15}^{20} \frac{x + \frac{t}{5} - 13}{5} dt = x - 9.5, \quad (x \in [10, 11])$$

$$f_T(t) = \int_{10}^{11} \frac{x + \frac{t}{5} - 13}{5} dx = \frac{t}{25} - \frac{1}{2}, \quad (t \in [15, 20])$$

Marginal Probability Density Functions

Example. **Find the marginal density of X for the reliability example:**

$$f_1(x) = \int_0^\infty x e^{-x(1+y)} dy = e^{-x}, \quad x \geq 0$$

Find the marginal density function for Y:

$$f_2(y) = \int_0^\infty x e^{-x(1+y)} dx = \frac{1}{1+y^2} \quad y \geq 0$$

by integrating by parts.

Observe that X does not have the same density function as Y, and we say that X and Y are not identically distributed.

If X and Y had the same marginal distributions, we would say they are identically distributed.

Marginal Probability Density Functions

Exercise.

Suppose the joint density function of X_1 and X_2 is

$$f(x_1, x_2) = \frac{\lambda}{x_1} e^{-\lambda x_1 - x_2/x_1}, \qquad x_1, x_2 \geq 0$$

Find the marginal density function for X_1 by integrating over all possible values of x_2 (i.e. $x_2 > 0$).

Ans.

$$f_{X_1}(x_1) = \int_0^\infty \frac{\lambda}{x_1} e^{-\lambda x_1 - x_2/x_1} dx_2 = \lambda e^{-\lambda x_1}, \quad x_1 \geq 0$$

and 0, otherwise. We recognize this as the exponential density function.

Marginal Probability Density Functions

Exercise 2.

Try to obtain the marginal density function for X_2.

Ans.

$$f_{X_2}(x_2) = \int_0^\infty \frac{\lambda}{x_1} e^{-\lambda x_1 - x_2/x_1} dx_1$$

which can only be evaluated numerically. Monte Carlo integration, anyone?

To summarize:

When in the context of several random variables, the marginal probability density function of a single one of the random variables can be obtained by integrating the joint density function over all possible values of all of the random variables apart from the one of interest.

Conditional Density Functions

Suppose we know the value of X_1, and we would like to predict the value of X_2, using this information.

The joint probability density function tells us how X_1 and X_2 are related, so we might think that $f(x_1, x_2)$ is the probability density function of X_2 for each given value of X_1, but this would be an incorrect interpretation.

This is because

$$\int_{-\infty}^{\infty} f(x_1, x_2) dx_2 = f_{X_1}(x_1)$$

If $f(x_1, x_2)$ were a probability density function for X_2, given X_1, the above integral should evaluate to 1.

Conditional Density Functions

However, if we divide $f(x_1, x_2)$ by $f_{X_1}(x_1)$, we obtain a function of x_2 which integrates to 1:

$$\int_{-\infty}^{\infty} \frac{f(x_1, x_2)}{f_{X_1}(x_1)} dx_2 = \frac{f_{X_1}(x_1)}{f_{X_1}(x_1)} = 1.$$

It turns out that this gives a very useful predictive density function for X_2, given knowledge of X_1.

We write

$$f_{X_2|X_1}(x_1, x_2) = \frac{f(x_1, x_2)}{f_{X_1}(x_1)}$$

as the *conditional density function* of X_2 given X_1. This density function predicts the probability density of X_2, when we know the value of X_1.

Conditional Density Functions

Similar reasoning tells us that the predictive probability density of X_1 or its conditional density function is given by

$$f_{X_1|X_2}(x_1, x_2) = \frac{f(x_1, x_2)}{f_{X_2}(x_2)}$$

This completely summarizes how knowledge of X_2 can help us make predictions about X_1.

Conditional Density Functions

Example. **Suppose X_1 and X_2 have the joint pdf**

$$f(x_1, x_2) = \frac{\lambda}{x_1} e^{-\lambda x_1 - x_2/x_1}, \qquad x_1, x_2 \geq 0$$

and we know the value of X_1 as x_1, say 10, or 3, etc. Earlier, we found that

$$f_{X_1}(x_1) = \lambda e^{-\lambda x_1}, \quad x_1 \geq 0.$$

Then the conditional density for X_2, given X_1

$$f_{X_2|X_1}(x_1, x_2) = \frac{1}{x_1} e^{-x_2/x_1}, \quad x_2 \geq 0.$$

This is an exponential density function, but now the rate is $1/x_1$. e.g. If we know that $x_1 = 10$, then the expected value of X_2 would be 10, and if $x_1 = 3$, the expected value of X_2 is 3.

The conditional density function of X_2, given X_1, is not the same as the marginal density of X_2. Thus, X_1 gives predictive information about X_2. The two random variables are not independent.

Independent Random Variables

We have spoken earlier of cases where X_1 and X_2 are independent. In such cases, X_1 provided no predictive information about X_2; technically, this means that the conditional density function of X_2 given X_1 is identical to the marginal density function of X_2:

$$f_{X_2|X_1}(x_1, x_2) = f_{X_2}(x_2).$$

Multiplying both sides of this by $f_{X_1}(x_1)$ gives the joint density function on the left and the product of the marginal density function on the right. Random variables are independent if their joint distribution can be factored in this way. That is, X_1 and X_2 are independent, if their joint density is

$$f(x_1, x_2) = f_{X_1}(x_1) f_{X_2}(x_2) \tag{1}$$

where the two functions are the respective marginal densities of X_1 and X_2.

Example.

Consider the propane example.

$$f(x,t) = \frac{x + \frac{t}{5} - 13}{5}$$

$$f_X(x)f_T(t) = (x - 9.5)\left(\frac{t}{25} - \frac{1}{2}\right)$$

$\Rightarrow$

$$f(x,t) \neq f_X(x)f_T(t)$$

so X and T are not independent.

Exercise. **For the reliability problem, show that X and Y are not independent.**

Independent Random Variables

Example. **An electronic surveillance system has one of each of two types of components in joint operation. The joint density function of the lifelengths X_1 and X_2 of the two components is**

$$f(y_1, y_2) = (1/8)y_1 e^{-\frac{1}{2}(y_1+y_2)}$$

for $y_1 > 0$ and $y_2 > 0$, and it is 0, otherwise.

Are X_1 and X_2 independent?

$$f_1(y_1) = \int_0^\infty \frac{1}{8} y_1 e^{-\frac{1}{2}(y_1+y_2)} dy_2 = \frac{1}{4} y_1 e^{-\frac{1}{2}y_1}, \qquad y_1 > 0$$

and

$$f_2(y_2) = \frac{1}{2} e^{-\frac{1}{2}y_2}, \qquad y_2 > 0$$

$\Rightarrow$

$$f(y_1, y_2) = f_1(y_1) f_2(y_2)$$

so X_1 and X_2 are independent.

Independent Random Variables

Example.

* **The time X_1 until failure of a fuel pump in an internal combustion engine can be modelled as a normal random variable with expected value 2000 hours and standard deviation 400 hours.**
* **The lifetime X_2 of a timing belt can be modelled as an exponential random variable with expected value 2800 hours.**
* **Supposing that these parts operate independently, find the probability that both fail before 1000 hours of operation.**

$$P(X_1 < 1000, X_2 < 1000) = P(X_1 < 1000)P(X_2 < 1000)$$

$$= P(Z < -2.5)(1 - e^{-.357}) = .0062(.300) = .0019$$

Here $Z = \frac{X_1 - 2000}{400}$ is a standard normal random variable. The `pnorm` function can be used to determine that $P(Z < -2.5) = .0062$.

Graphical Views of Independence and Dependence

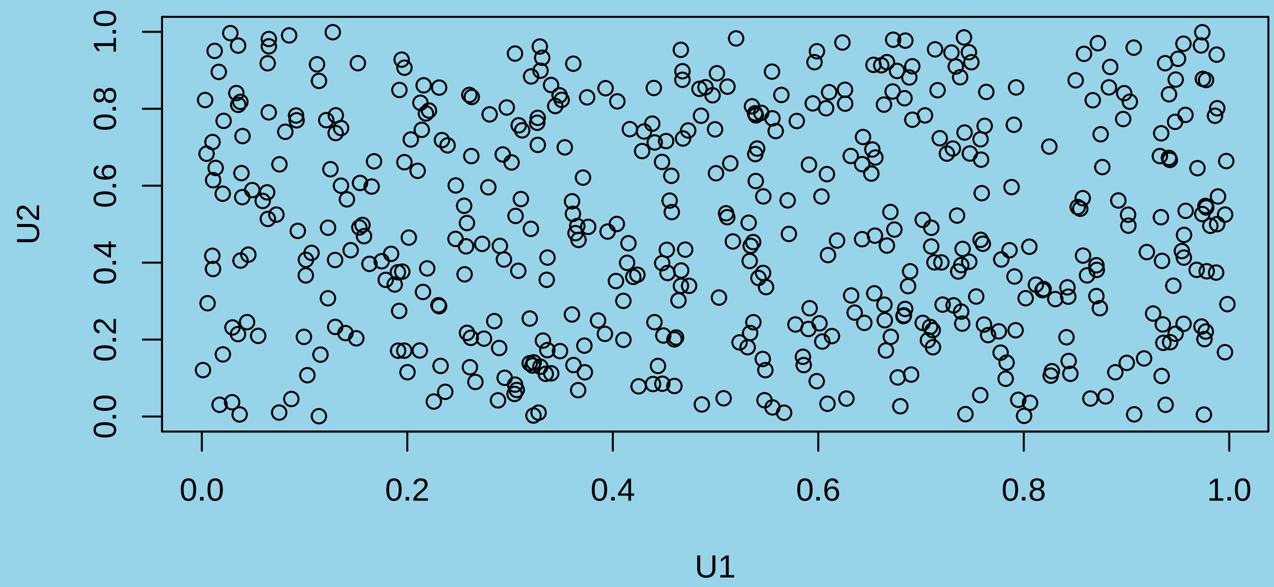
We can get an intuitive feel for the independence modelling assumption (1) by graphing some simulated random variables, both independent and for some forms of dependence.

First, let's consider two independent uniform random variables which take values in the interval $[0, 1]$.

The next figure displays a scatterplot of 500 values taken from the distributions of U_2 and U_1, where are both uniformly distributed.

Graphical Views of Independence and Dependence

```
U1 <- runif(500)
U2 <- runif(500)
plot(U2 ~ U1)
```



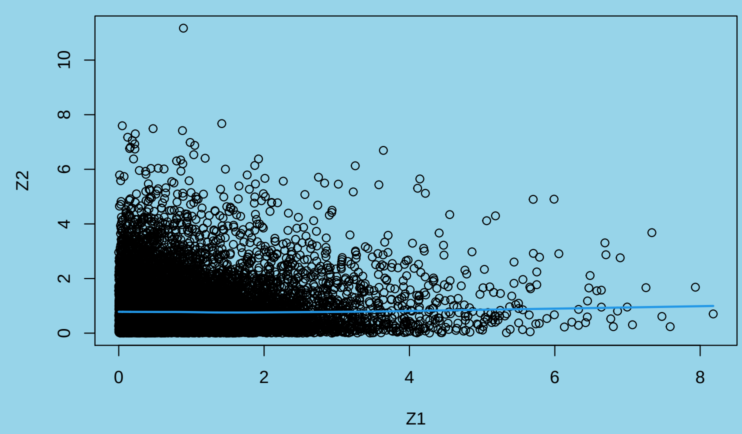
There is no structure to the patterns that might be discerned from this picture. This is the clearest possible illustration of variables which have no relation.

Graphical Views of Independence and Dependence

Independence manifests itself in other ways.

Next, we consider exponential random variables, Z_1 and Z_2. In both cases, we will sample 10000 points from their respective distributions and look at a scatterplot of the corresponding pairs of data points.

```r
Z1 <- rexp(10000)
Z2 <- rexp(10000)
plot(Z2 ~ Z1)
lines(lowess(Z1, Z2),
      col=4, lwd=2)
```



What you should observe in the figure is that it is not possible to predict the value of Z_2 from knowledge of Z_1.

This is a random collection of points, even though it might appear that there is a pattern (points are bunched up towards the lower left corner of the plot).

What characterizes independence is that neither variable gives predictive information about the other variable.

A Case of Dependence

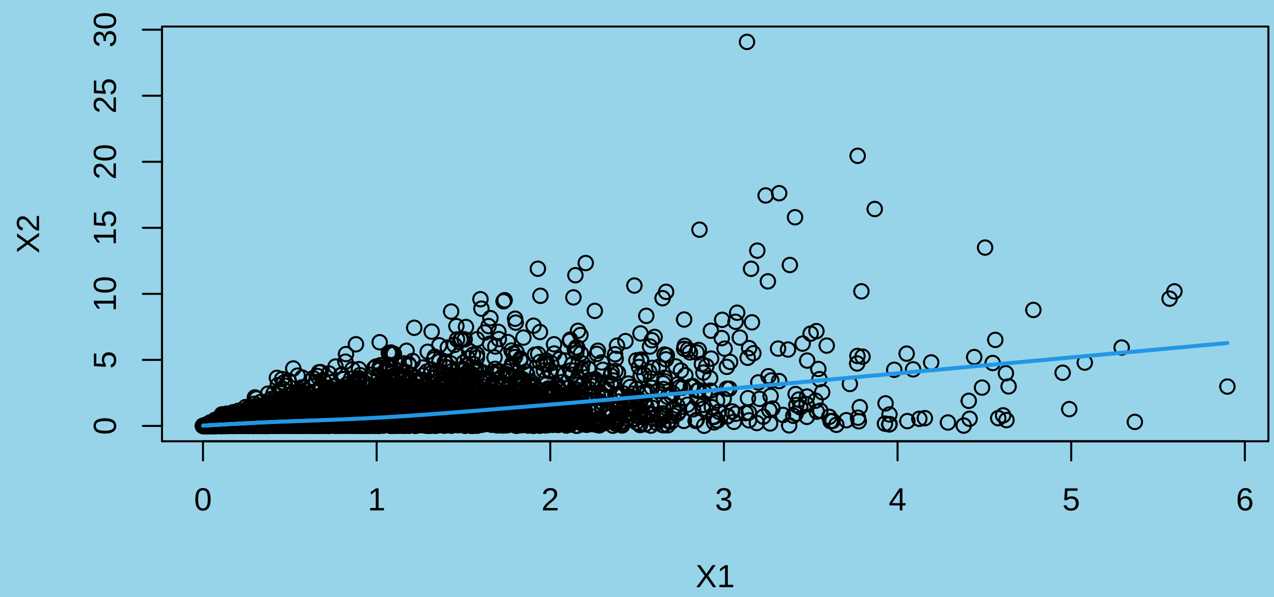
Let's change the story a bit now.

Suppose X_1 and X_2 are related. In particular, suppose X_1 is exponentially distributed with rate $\lambda = 1.5$ and X_2 is exponential with rate $1/X_1$.

Next, we will apply our simulation-based graphical analysis.

Graphical Views of Independence and Dependence

```r
X1 <- rexp(10000, rate = 1.5)
X2 <- rexp(10000, rate = 1/X1)
plot(X2 ~ X1)
lines(lowess(X1, X2), col=4, lwd=2)
```



Simple Regression

Suppose X and Y are random variables which have a joint density function given by

$$f(x,y) = \frac{\mathrm{e}^{-(y-\beta_0-\beta_1 x)^2/(2\sigma^2)-x^2/2}}{2\pi\sigma}.$$

This is an example of a bivariate normal pdf: X is normal with mean 0, and Y is normal with mean $\beta_0 + \beta_1 x$. In other words, the mean of Y is now a linear function of x.

β_0 and β_1 are unknown intercept and slope parameters.

If we want to predict Y from X, we should use the conditional density function $f_{Y|X}(x,y)$. We can obtain that density function in 2 steps:

1. Find $f_X(x)$ by integrating over all y.
2. Divide $f(x,y)/f_X(x)$.

1. $$f_X(x) = \int f(x,y)dy = \frac{\mathrm{e}^{-x^2/2}}{\sqrt{2\pi}}.$$

   We have used the fact that

   $$\frac{\mathrm{e}^{-(y-\beta_0-\beta_1 x)^2/(2\sigma^2)}}{\sqrt{2\pi}\sigma}$$

   is a normal pdf and integrates to 1.

2. Dividing $f(x,y)$ by $f_X(x)$ gives

   $$f_{Y|X}(x,y) = \frac{\mathrm{e}^{-(y-\beta_0-\beta_1 x)^2/(2\sigma^2)}}{\sqrt{2\pi}\sigma}.$$

Simple Regression

The conditional distribution that we just obtained is a normal pdf with mean

$$\beta_0 + \beta_1 x$$

and variance σ^2.

The mean is an expected value (called the *conditional expectation*) and has the notation

$$E[Y|X = x] = \beta_0 + \beta_1 x. \tag{2}$$

The conditional expectation of Y, given $X = x$ is also referred to as the regression function, a function of x.

The variance is actually a *conditional variance*:

$$\textbf{Var}(Y|X = x) = \sigma^2. \tag{3}$$

This is often referred to as the noise variance.

Simple Regression

The regression function at (2) and the variance function at (3), which is just the constant function, tell us that, given $X = x$, we could view Y as the random variable

$$Y = \beta_0 + \beta_1 x + \varepsilon$$

where ε is the noise random variable - a normal random variable with mean 0 and variance σ^2. The $\beta_0 + \beta_1 x$ terms are not random.

This is the usual form of the simple linear regression model which relates Y to x in the presence of noise.

Simple Regression

Some terminology:

The simple linear regression *model* relating a *response variable* y to a *predictor variable* x is

$$y = \beta_0 + \beta_1 x + \varepsilon$$

where β_0 is the intercept and β_1 is the slope of the regression line.

ε is a random quantity representing noise, also called error, about the line.

The noise is often assumed to be a sequence of independent normal random variables with mean 0 and constant variance σ^2. The independence assumption and 0 mean assumption are the most crucial assumptions.

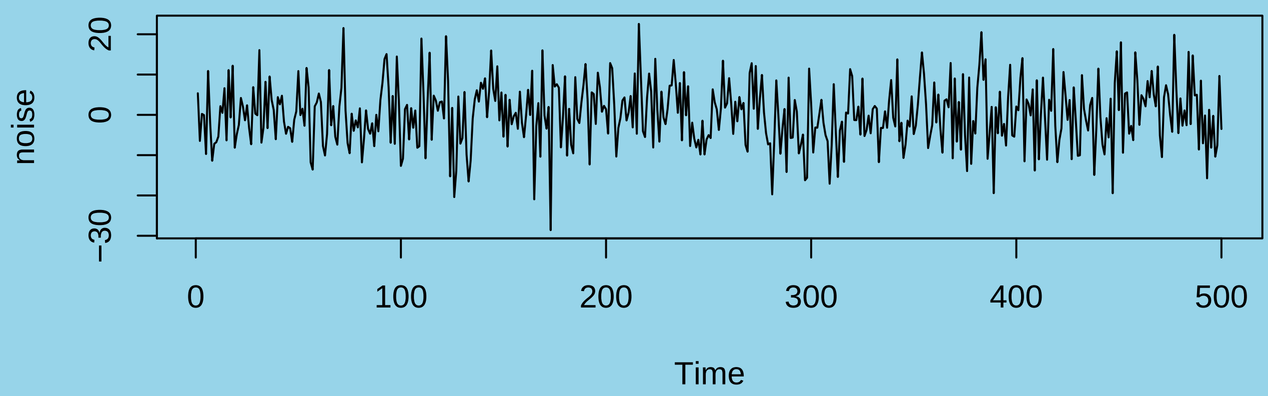
Note the switch from upper case Y to lower case y here. Typically, we refer to the random variable, before observing data, as Y and the observed data as y.

Simulated Noise.

e.g. consider 500 values of ε which have $\sigma = 8$:

```
eps <- rnorm(500, sd = 8)
```

```
ts.plot(eps, ylab="noise")
```



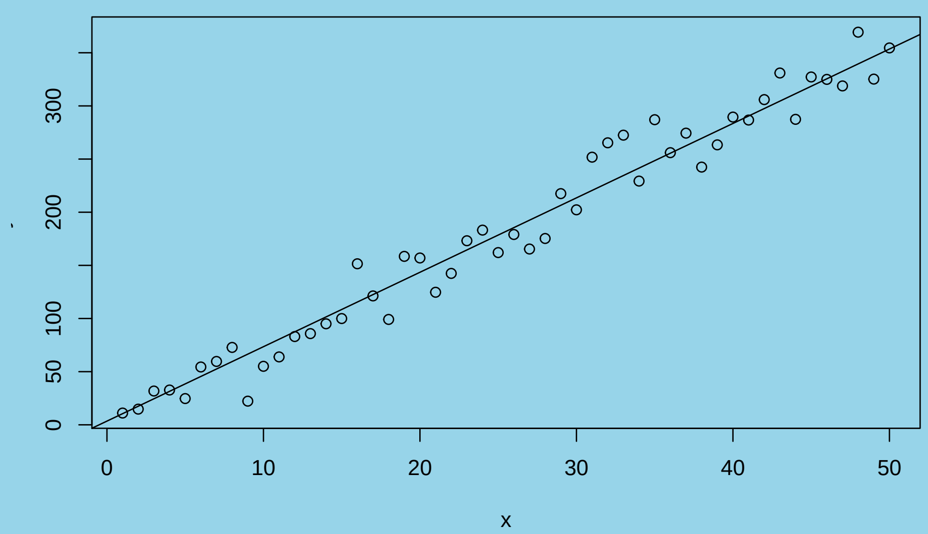
Simulating Regression Data

In the simple linear regression model, the noise is added to a line of slope β_1 and intercept β_0.

Example.

Suppose x values are taken at $\{1, 2, 3, \ldots, 50\}$. If the intercept is 3.5 and the slope is 7.0, and the noise is normal with standard deviation 16.0, we have

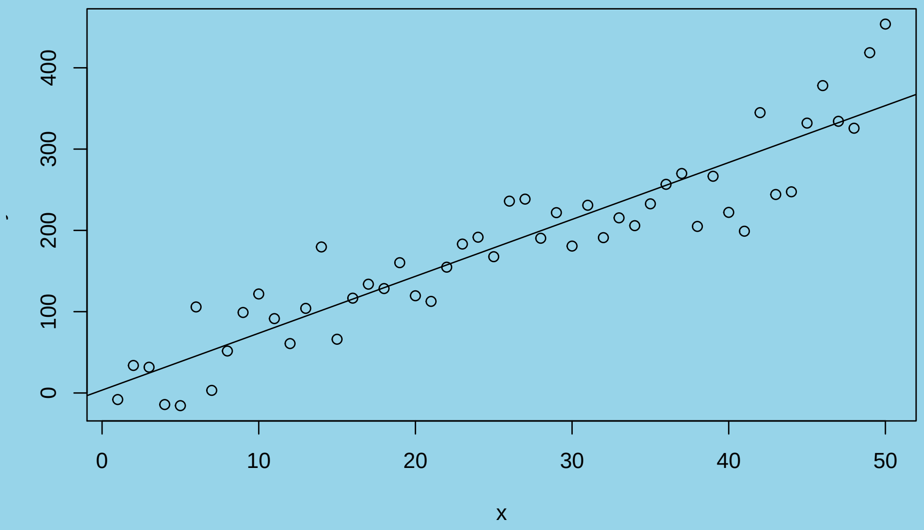
```r
x <- 1:50
eps <- rnorm(50, sd = 16)
y <- 3.5 + 7.0*x + eps
plot(y ~ x)
abline(3.5, 7)
```



Simulating Regression Data

Suppose the standard deviation is larger: 40.0, we have

```r
eps <- rnorm(50, sd = 40 )
y <- 3.5 + 7.0*x + eps
```



... larger noise standard deviation gives more variation about the true line ...

$\rightsquigarrow$ **harder to predict Y from x.**

Example: Model Car Data.

Consider the data on the model car that was released from various points on a ramp and the distance traveled was measured.

```
library(DAAG)
mcar.lm <- lm(distance.traveled ~ starting.point,
      data = modelcars)
coef(mcar.lm) # intercept and slope estimate

##    (Intercept) starting.point
##       8.083333       2.013889
```

Example: Model Car Data.

The fitted model is

$$y = 8.0833333 + 2.0138889x + \varepsilon$$

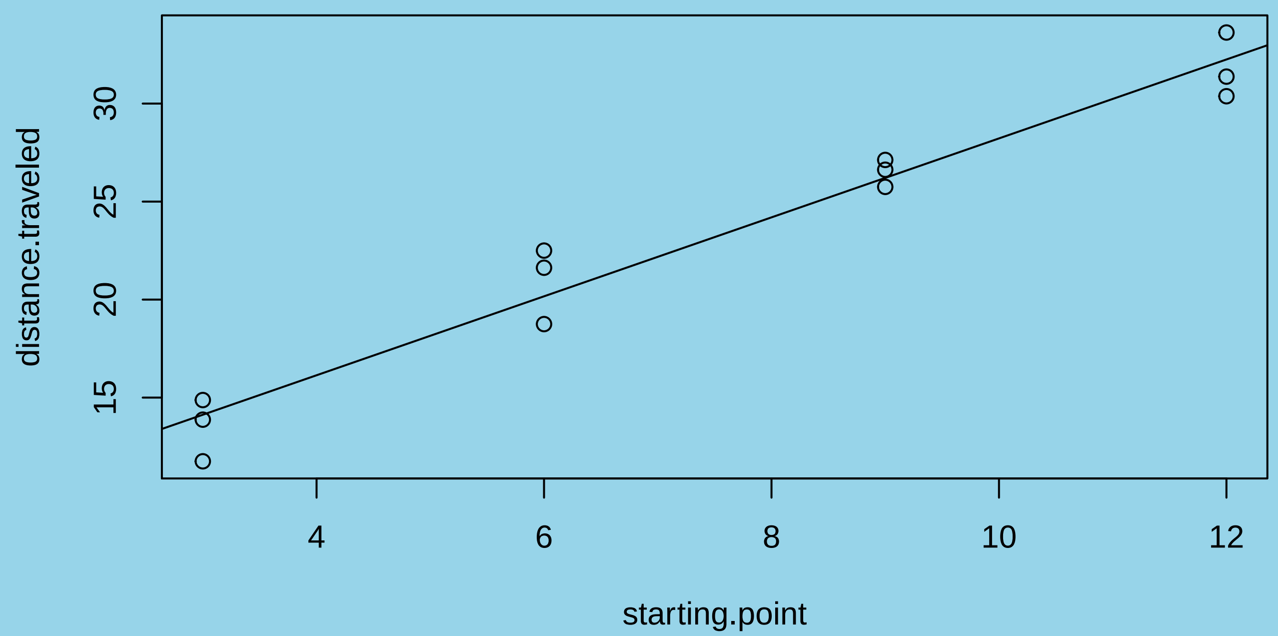
where y is distance and x is starting point. The error (ε) standard deviation is

```
summary(mcar.lm)$sigma

## [1] 1.524453
```

Simple Regression

Example: Model Car Data.

```
plot(distance.traveled ~ starting.point,
      data = modelcars)
abline(mcar.lm)
```



Using Simulation to Learn about Regression

The regression procedure is based on mathematics which would take too long to go through here – there are other courses that cover that material.

Instead, we can use simulation to gain intuition into the procedure.

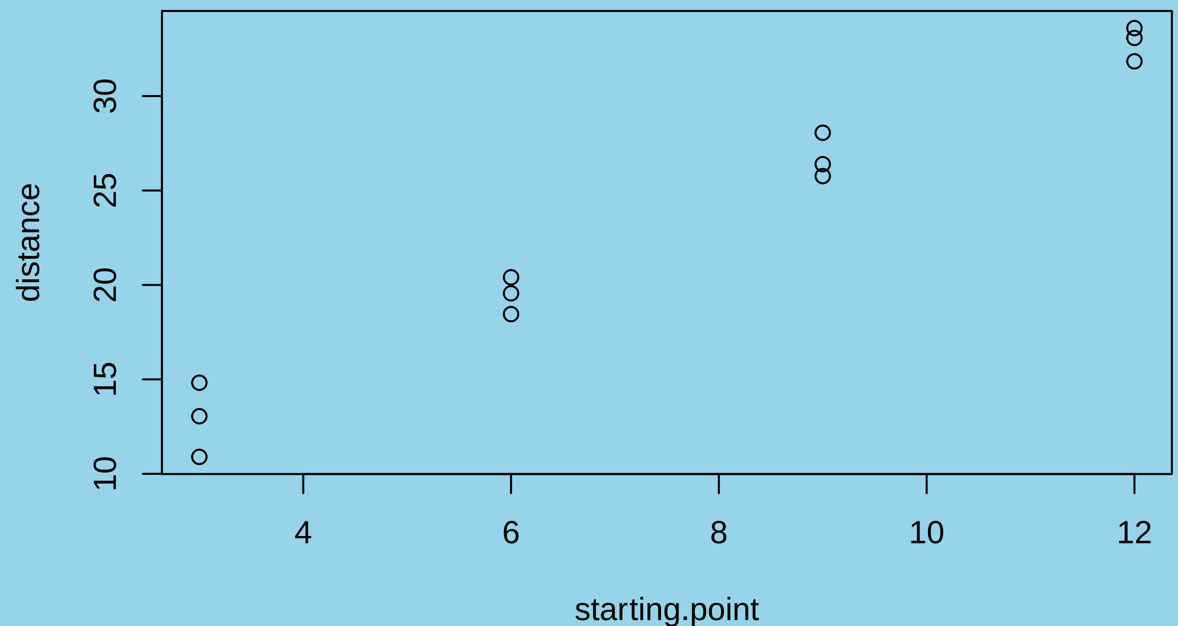
By simulating new data where we know the true coefficients and the true errors, we can see how the regression estimates differ from the truth.

We can also learn some things about the residuals and how they relate to the true errors.

Simulated Linear Regression Data

We will simulate data that "looks" like the data in `modelcars`

```
carssim <- modelcars  # carssim will contain simulated data
eps <- rnorm(n = nrow(carssim) , sd = 1.524) # simulated noise
carssim$distance <- 8.083 + 2.0138*carssim$starting.point +eps
plot(distance ~ starting.point, data = carssim)
```



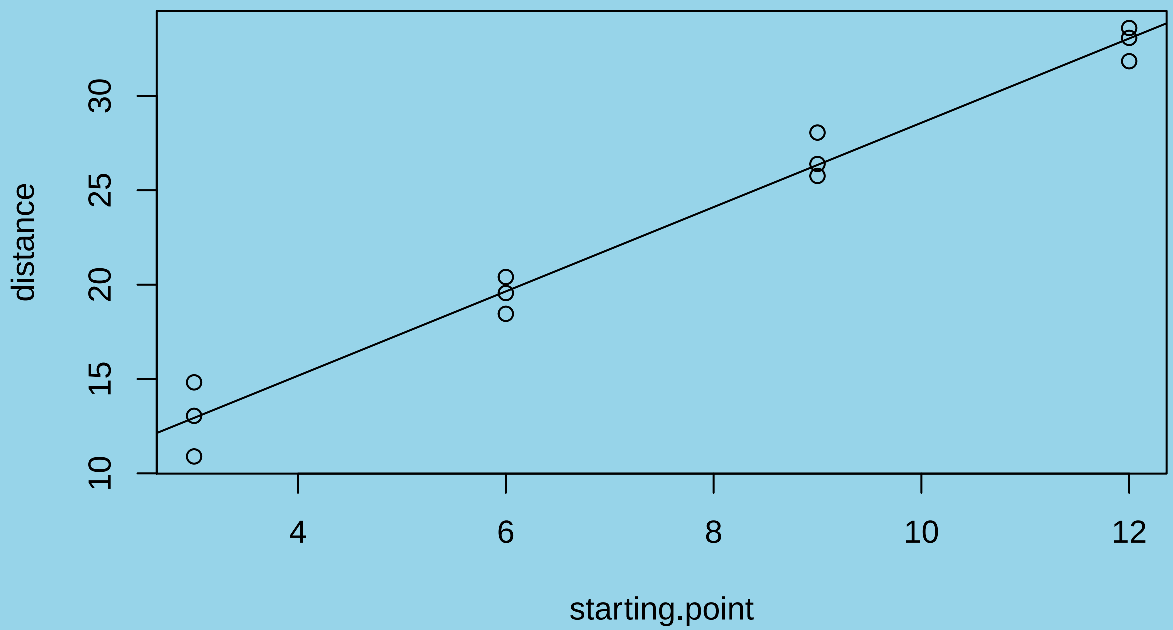
Simulated Linear Regression Data

```
carssim.lm <- lm(distance ~ starting.point, data = carssim)
#estimated intercept and slope for simulated data
coef(carssim.lm)

##    (Intercept) starting.point
##       6.239132       2.233781

summary(carssim.lm)$sigma  # sd estimate

## [1] 1.215448
```

Now we see how the estimates of the intercept, slope and estimate of σ differ from the true values.

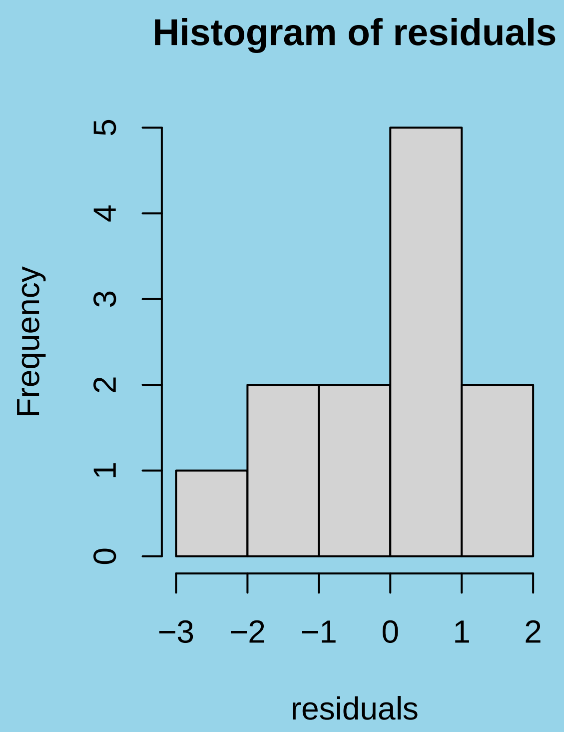
Plotting the Best Fit Line - Simulated Data

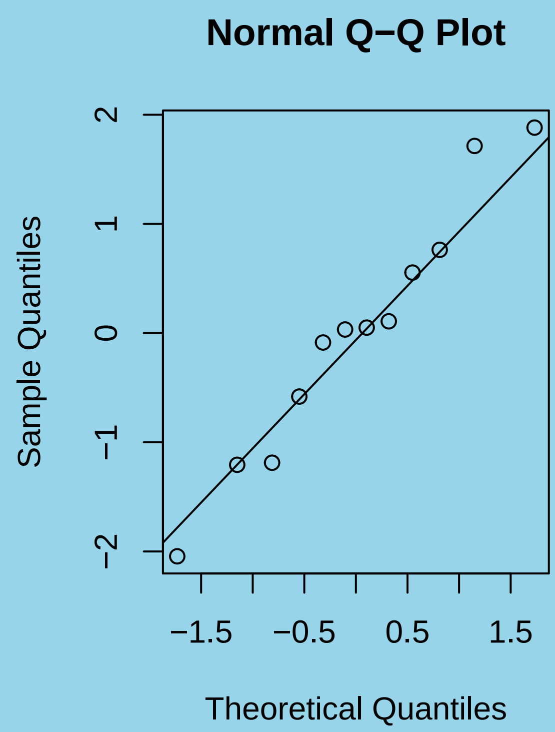
```r
plot(distance ~ starting.point, data = carssim)
abline(carssim.lm)
```



The True Errors are Normal; What about the Residuals?

```
residuals <- resid(carssim.lm)
par(mfrow=c(1,2)); hist(residuals)
qqnorm(residuals); qqline(residuals)
```

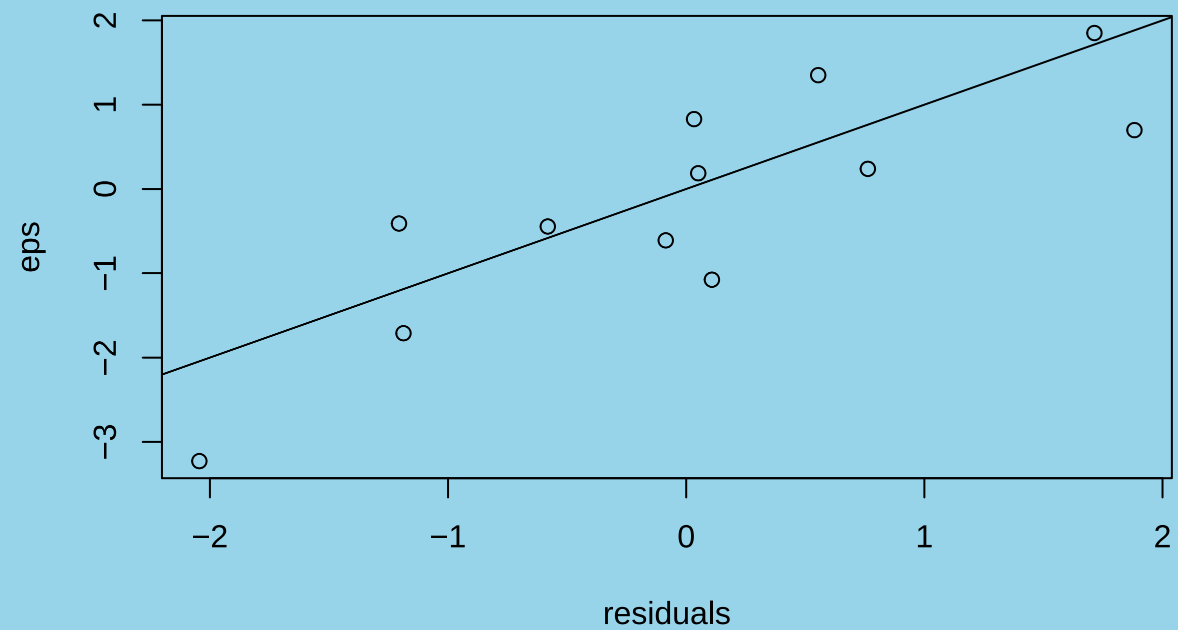




How do the Simulated Residuals Behave?

Compare the simulated residuals with the true errors:

```r
plot(eps ~ residuals)
abline(0,1)
```



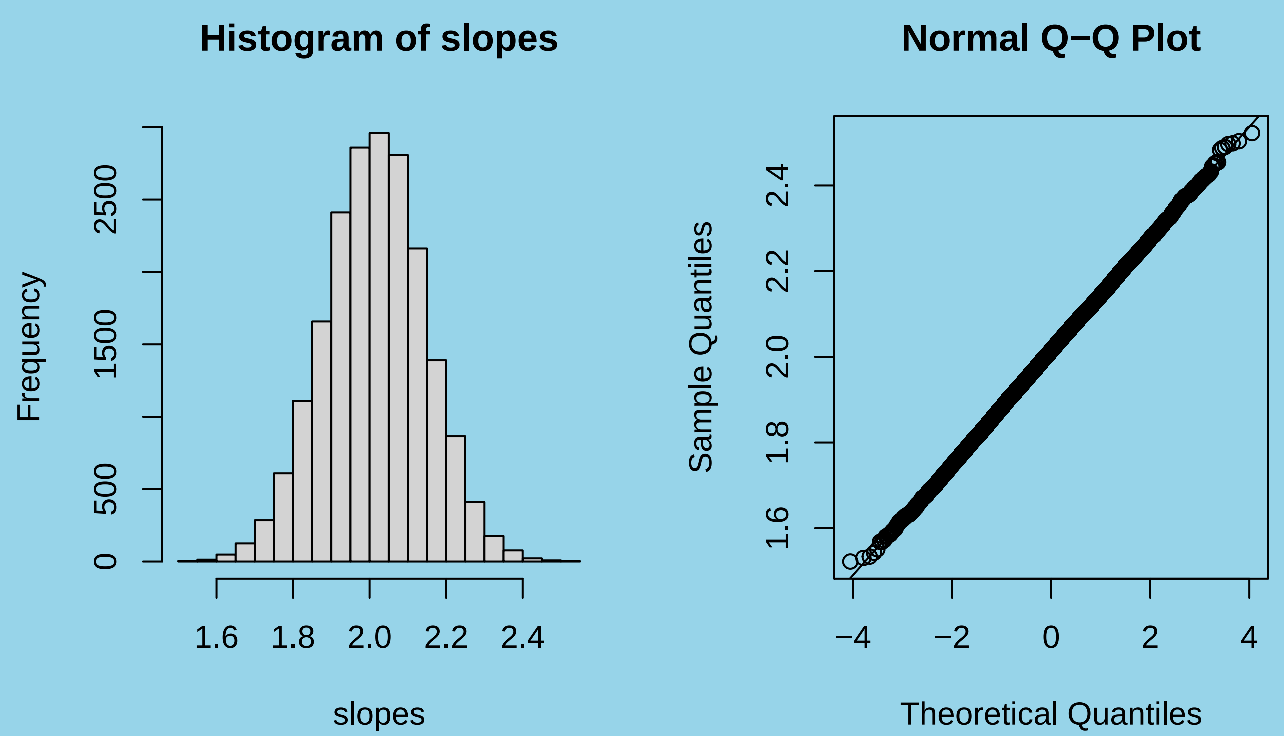
Study the Slope Estimate Distribution via Simulation

```r
b0 <- coef(mcar.lm)[1]
b1 <- coef(mcar.lm)[2]
sdCar <- summary(mcar.lm)$sigma
```

```r
Nsims <- 20000; slopes <- sderrors <- numeric(Nsims)
for (i in 1:Nsims) {# 20000 simulated data sets
    eps <- rnorm(n = nrow(modelcars) , sd = sdCar)
    modelcars$distance.traveled <-
           b0 + b1*modelcars$starting.point +eps
    mcar.lm <- lm(distance.traveled ~ starting.point,
           data = modelcars); slopes[i] <- coef(mcar.lm)[2]
    sderrors[i] <- summary(mcar.lm)$coefficients[2,2]
}
mean(slopes); sd(slopes)

## [1] 2.013438
## [1] 0.1318652
```

What is the Distribution of the Slope Estimate?

```r
par(mfrow=c(1,2)); hist(slopes); qqnorm(slopes); qqline(slopes)
```



... evidence that the slope estimate is approximately normally distributed ...

What to Take Away from this Lecture

Modelling several variables at the same time can be complicated.

The assumption of independence greatly simplifies the problem by allowing consideration of each variable separately – independently of all of the other variables.

If the independence assumption is used, it must be checked:

- graphically - familiarize yourself with what a random pattern looks like; if a nonrandom pattern appears on a scatterplot of two or three random variables, they cannot be independent.

What to Take Away from this Lecture

Dependence can also be very useful: if a random variable is dependent on other random variables, the other variables can be used for prediction.

The conditional probability density function is essentially a predictive function.

Regression modelling is based on the expected value of a random variable, using the conditional probability density function.

Simulation can sometimes be used, in place of mathematics, to gain an understanding of the outputs that come from built-in analysis functions, such as the regression function `lm()`.