Linear Models

The word "model" is usually used in situations where we want to predict or forecast something we would like to know, based on things we do know.

Linear Models

The word "model" is usually used in situations where we want to predict or forecast something we would like to know, based on things we do know.

Example: We want to predict the highest sustained wind speed in a tropical storm at some point over the ocean, and we know the barometric pressure.

Linear Models

The word "model" is usually used in situations where we want to predict or forecast something we would like to know, based on things we do know.

Example: We want to predict the highest sustained wind speed in a tropical storm at some point over the ocean, and we know the barometric pressure.

In general, barometric pressure is related to wind speed: Lower pressure is associated with a stronger storm, and a stronger storm is associated with higher winds.

Linear Models

The word "model" is usually used in situations where we want to predict or forecast something we would like to know, based on things we do know.

Example: We want to predict the highest sustained wind speed in a tropical storm at some point over the ocean, and we know the barometric pressure.

In general, barometric pressure is related to wind speed: Lower pressure is associated with a stronger storm, and a stronger storm is associated with higher winds.

A model is desirable because barometric pressure is more stable than wind speed and can be measured more easily.

Linear Models

Let's consider a simple (if artificial) example. Suppose:

- Y represents a quantity we want to predict
- X represents a related quantity we can measure

Linear Models

Let's consider a simple (if artificial) example. Suppose:

- Y represents a quantity we want to predict
- X represents a related quantity we can measure

The simplest relationship that comes to mind is the linear equation:

$$Y = \beta X$$

where β is some coefficient we can think of as a parameter that can be estimated from a sample.

Linear Models

Let's consider a simple (if artificial) example. Suppose:

- Y represents a quantity we want to predict
- X represents a related quantity we can measure

The simplest relationship that comes to mind is the linear equation:

$$Y = \beta X$$

where β is some coefficient we can think of as a parameter that can be estimated from a sample.

There is a philosophical problem with this model though: it is **deterministic**.

Linear Models

Let's consider a simple (if artificial) example. Suppose:

- Y represents a quantity we want to predict
- X represents a related quantity we can measure

The simplest relationship that comes to mind is the linear equation:

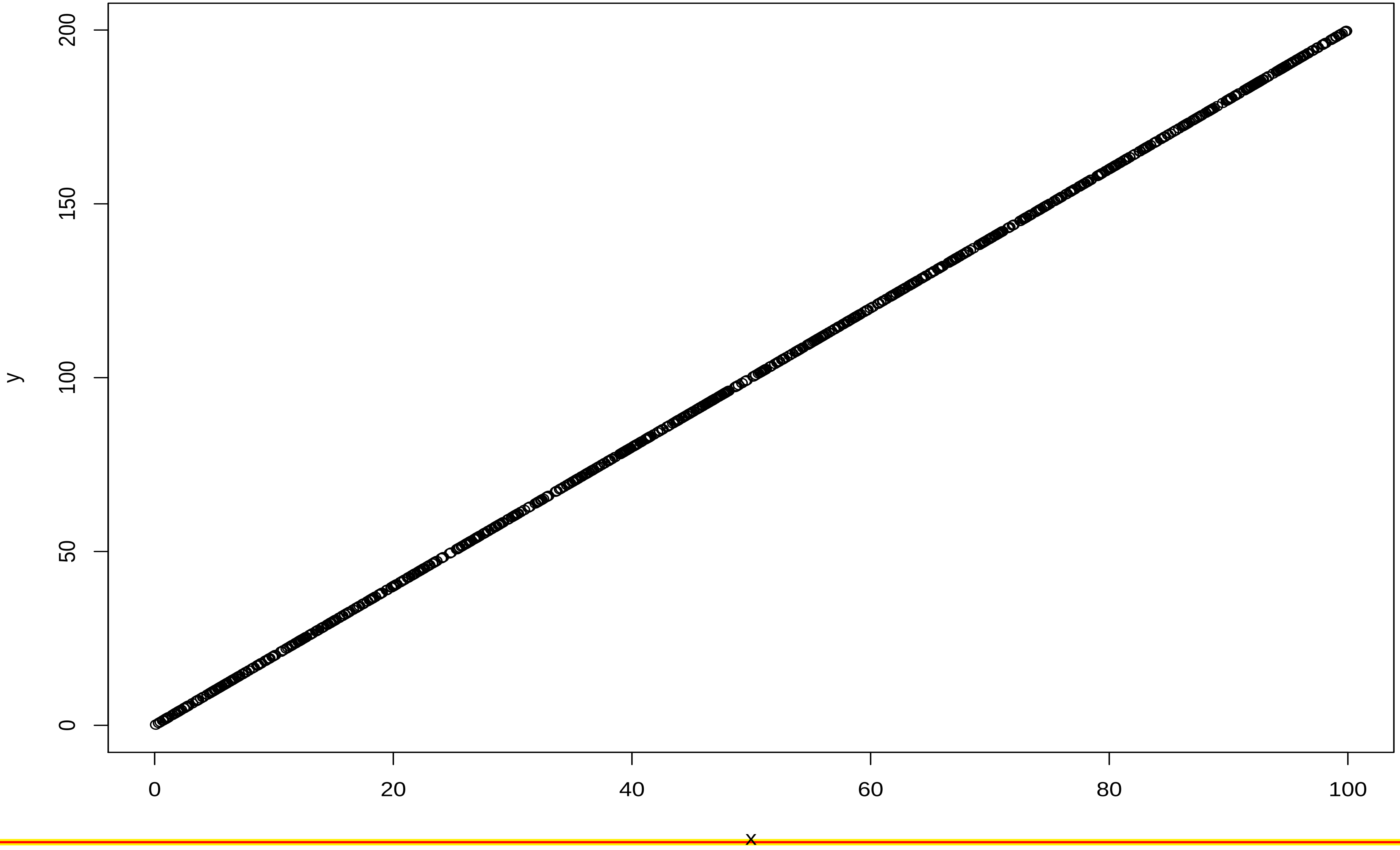
$$Y = \beta X$$

where β is some coefficient we can think of as a parameter that can be estimated from a sample.

There is a philosophical problem with this model though: it is **deterministic**.

The model says we know Y **exactly** if we know the value of X.

Deterministic Model



Linear Models

This is not realistic in most cases.

Linear Models

This is not realistic in most cases.

A better approach is to state the situation as a **probability** model:

Consider Y to be a random variable, and write the model as

$$E(Y) = \beta X$$

Linear Models

This is not realistic in most cases.

A better approach is to state the situation as a **probability** model:

Consider Y to be a random variable, and write the model as

$$E(Y) = \beta X$$

Now we are simply stating that the expected value or population mean μ_Y of Y given X is βX

Linear Models

This is not realistic in most cases.

A better approach is to state the situation as a **probability** model:

Consider Y to be a random variable, and write the model as

$$E(Y) = \beta X$$

Now we are simply stating that the expected value or population mean μ_Y of Y given X is βX

This avoids the requirement that every Y value exactly match βX.

Linear Models

For an individual observation Y_i with associated value X_i, we solve the problem of introducing randomness differently. We represent an individual observation as:

$$Y_i = \beta X_i + e_i$$

where:

- β is a parameter (a constant to a frequentist, a random variable to a Bayesian)
- X_i is a **known** constant
- e_i is a random variable with expected value $E(e_i) = \mu_e = 0$ and standard deviation σ_e

Linear Models

For an individual observation Y_i with associated value X_i, we solve the problem of introducing randomness differently. We represent an individual observation as:

$$Y_i = \beta X_i + e_i$$

where:

- β is a parameter (a constant to a frequentist, a random variable to a Bayesian)
- X_i is a **known** constant
- e_i is a random variable with expected value $E(e_i) = \mu_e = 0$ and standard deviation σ_e

Now Y_i is a random variable. The randomness of Y_i arises from e_i (and, in the Bayesian approach, also from β).

Linear Models - Frequentist

From the properties of expected values, recall that

$$E(Y_i) = E(\beta X_i + e_i)$$

Linear Models - Frequentist

From the properties of expected values, recall that

$$E(Y_i) = E(\beta X_i + e_i)$$

In the classical or frequentist approach, β and X_i are constants, and the expected value of a constant is just its value, so

$$E(Y_i) = \beta X_i + E(e_i)$$

Linear Models - Frequentist

From the properties of expected values, recall that

$$E(Y_i) = E(\beta X_i + e_i)$$

In the classical or frequentist approach, β and X_i are constants, and the expected value of a constant is just its value, so

$$E(Y_i) = \beta X_i + E(e_i)$$

We are assuming that $E(e_i) = 0$ for each e_i, so

$$E(Y_i) = \beta X_i + 0 = \beta X_i$$

which agrees with the earlier result.

Linear Models - Bayesian

Again recall from the properties of expected values that:

$$E(Y_i) = E(\beta X_i + e_i)$$

Linear Models - Bayesian

Again recall from the properties of expected values that:

$$E(Y_i) = E(\beta X_i + e_i)$$

In the Bayesian approach, β is treated as a random variable and we have to assume a particular probability distribution for it. This is called the **prior** distribution of β. In the expression below, $E(\beta)$ represents the expected value of β with respect to this distribution. Once again the X_i are constants and are equal to their expected values, so this time we can write

$$E(Y_i) = E(\beta) \cdot X_i + E(e_i)$$

Linear Models - Bayesian

Again recall from the properties of expected values that:

$$E(Y_i) = E(\beta X_i + e_i)$$

In the Bayesian approach, β is treated as a random variable and we have to assume a particular probability distribution for it. This is called the **prior** distribution of β. In the expression below, $E(\beta)$ represents the expected value of β with respect to this distribution. Once again the X_i are constants and are equal to their expected values, so this time we can write

$$E(Y_i) = E(\beta) \cdot X_i + E(e_i)$$

We are assuming that $E(e_i) = 0$ for each e_i, so

$$E(Y_i) = E(\beta)X_i + 0 = E(\beta)X_i$$

Linear Models

So to a frequentist,

$$E(Y_i) = \beta X_i$$

and to a Bayesian

$$E(Y_i) = E(\beta) X_i$$

Linear Models

So to a frequentist,

$$E(Y_i) = \beta X_i$$

and to a Bayesian

$$E(Y_i) = E(\beta) X_i$$

This illustrates the difference between the classical or frequentist approach and the Bayesian approach.

Linear Models

So to a frequentist,

$$E(Y_i) = \beta X_i$$

and to a Bayesian

$$E(Y_i) = E(\beta) X_i$$

This illustrates the difference between the classical or frequentist approach and the Bayesian approach.

To a frequentist, β is a constant, while a Bayesian considers it to be a random variable having the prior distribution.

Linear Models

So to a frequentist,

$$E(Y_i) = \beta X_i$$

and to a Bayesian

$$E(Y_i) = E(\beta) X_i$$

This illustrates the difference between the classical or frequentist approach and the Bayesian approach.

To a frequentist, β is a constant, while a Bayesian considers it to be a random variable having the prior distribution.

The prior distribution is subjective, and can be thought of as a mathematical model of the researcher's uncertainty about the value of β.

Linear Models

Which point of view is better, frequentist or Bayesian?

Linear Models

Which point of view is better, frequentist or Bayesian?

This question has been (and continues to be) a source of controversy and debate within the statistics community. Both have advantages and disadvantages.

Linear Models

Which point of view is better, frequentist or Bayesian?

This question has been (and continues to be) a source of controversy and debate within the statistics community. Both have advantages and disadvantages.

People in applied statistics usually try to be pragmatic and choose an approach that makes sense in the context of the data they have to work with and the question they are trying to answer.

Linear Models

The good news is that as the sample size becomes larger, the most common frequentist techniques and their Bayesian counterparts converge to the same limit.

Linear Models

The good news is that as the sample size becomes larger, the most common frequentist techniques and their Bayesian counterparts converge to the same limit.

This limit is called the **maximum likelihood** estimate of β, meaning an estimate that makes the probability of observing the sample you actually got as large as possible.

Linear Models

The good news is that as the sample size becomes larger, the most common frequentist techniques and their Bayesian counterparts converge to the same limit.

This limit is called the **maximum likelihood** estimate of β, meaning an estimate that makes the probability of observing the sample you actually got as large as possible.

We saw a similar situation when we considered the normal and t distributions for constructing confidence intervals.

Linear Models

The good news is that as the sample size becomes larger, the most common frequentist techniques and their Bayesian counterparts converge to the same limit.

This limit is called the **maximum likelihood** estimate of β, meaning an estimate that makes the probability of observing the sample you actually got as large as possible.

We saw a similar situation when we considered the normal and t distributions for constructing confidence intervals.

If you had a large enough sample, you got essentially the same answer regardless of which one you used.

Linear Models

Some prominent researchers consider the biggest disadvantage of the Bayesian approach to be the choice of the prior distribution, which is necessarily subjective.

Linear Models

Some prominent researchers consider the biggest disadvantage of the Bayesian approach to be the choice of the prior distribution, which is necessarily subjective.

How much information to incorporate in the prior, and how to go about it, remains a difficult question and is an active area of research.

Linear Models

Some prominent researchers consider the biggest disadvantage of the Bayesian approach to be the choice of the prior distribution, which is necessarily subjective.

How much information to incorporate in the prior, and how to go about it, remains a difficult question and is an active area of research.

There is some agreement that choosing a prior distribution that imposes no restrictions on the value of β is acceptable when there is outside information to be incorporated in the analysis, and this is the approach we will take.

Linear Models

We will use R to generate a model of this type.

First generate 1000 values for the X_i: Pick 1000 values between zero and 100 with the command:

*x<-100*runif(1000)*

Linear Models

We will use R to generate a model of this type.

First generate 1000 values for the X_i: Pick 1000 values between zero and 100 with the command:

```
x<-100*runif(1000)
```

Next assign a value to β. We'll use 2:

```
beta<-2
```

Linear Models

We will use R to generate a model of this type.

First generate 1000 values for the X_i: Pick 1000 values between zero and 100 with the command:

*x<-100*runif(1000)*

Next assign a value to β. We'll use 2:

beta<-2

Now generate 1,000 e_i values as normal random variables with mean zero and standard deviation 5:

e<-rnorm(1000,0,5)

Linear Models

We will use R to generate a model of this type.

First generate 1000 values for the X_i: Pick 1000 values between zero and 100 with the command:

```
x<-100*runif(1000)
```

Next assign a value to β. We'll use 2:

```
beta<-2
```

Now generate 1,000 e_i values as normal random variables with mean zero and standard deviation 5:

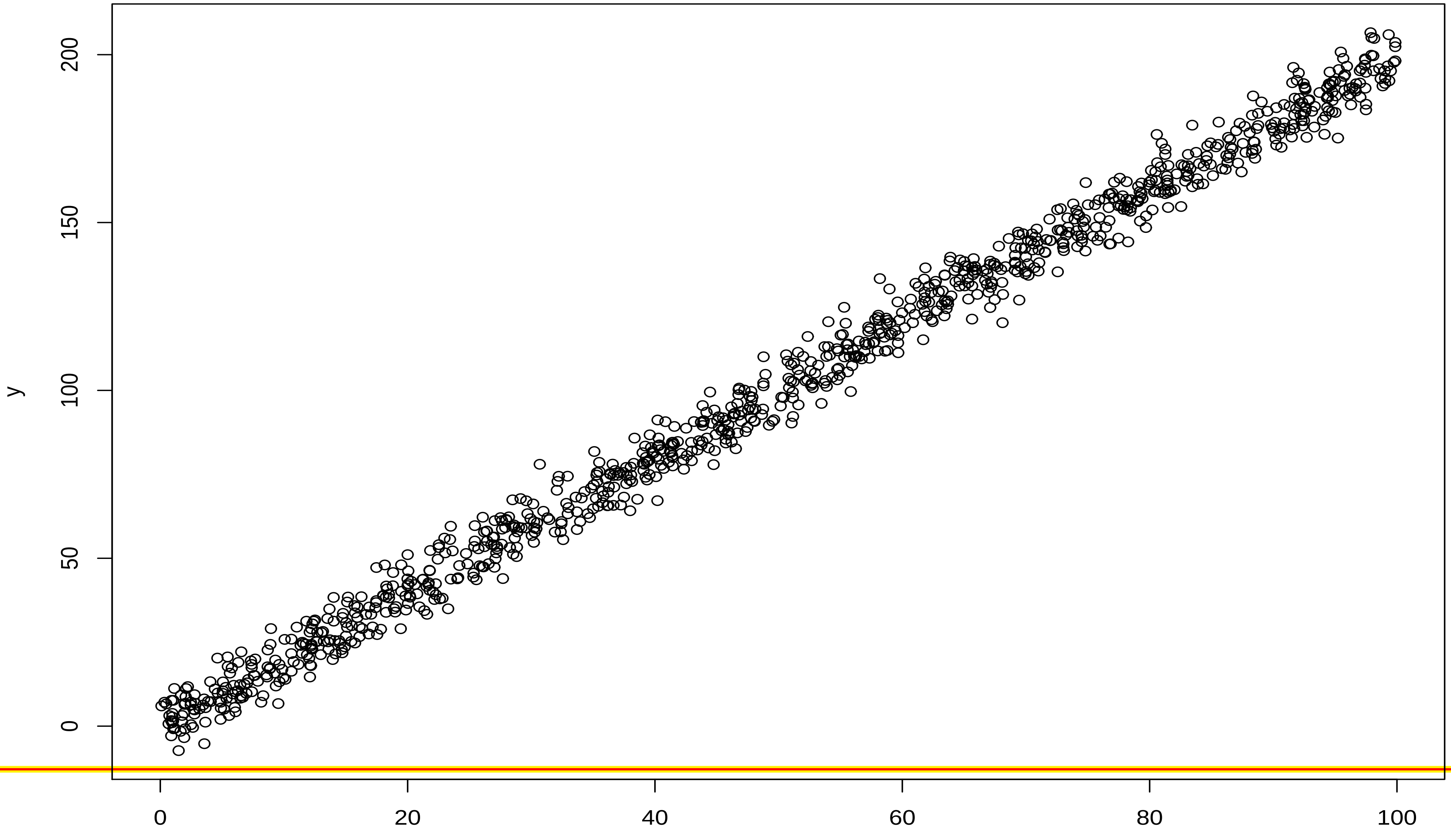
```
e<-rnorm(1000,0,5)
```

Finally generate the Y_i values as $Y_i = \beta X_i + e_i$

```
y<-beta*x+e
```

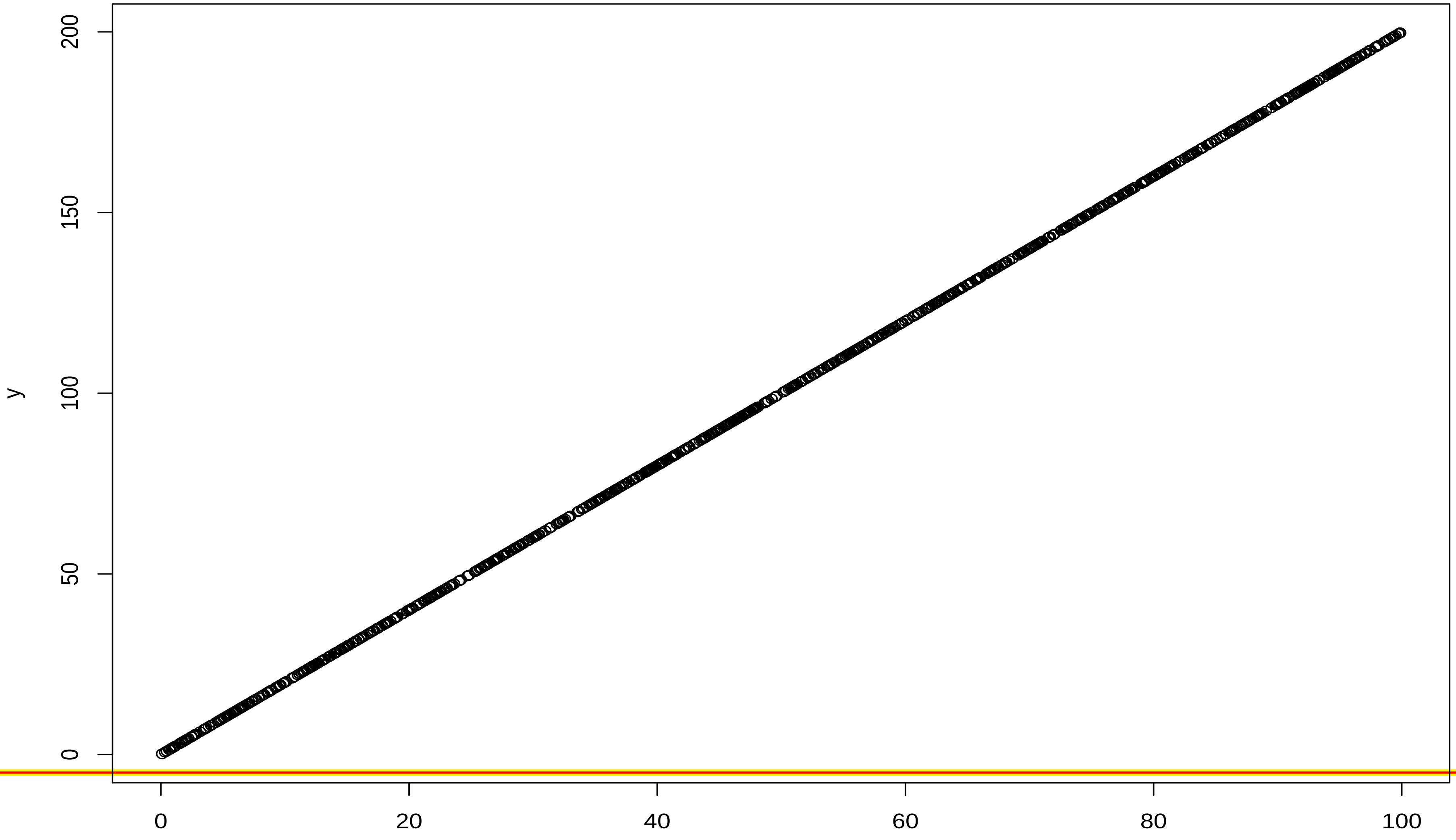
Data Plot: Beta=2 Sigma=5

plot(x,y)



Data Plot: Beta=2 Sigma=0

Compare with the deterministic model $Y = \beta X$:



Linear Models

Now we examine the effect of larger values of σ_e:

Generate 1,000 e_i values as normal random variables with mean zero and standard deviation 10:

e<-rnorm(1000,0,10)

Linear Models

Now we examine the effect of larger values of σ_e:

Generate 1,000 e_i values as normal random variables with mean zero and standard deviation 10:

e<-rnorm(1000,0,10)

Now regenerate the Y_i values as $Y_i = \beta X_i + e_i$:

*y<-beta*x+e*

Linear Models

Now we examine the effect of larger values of σ_e:

Generate 1,000 e_i values as normal random variables with mean zero and standard deviation 10:

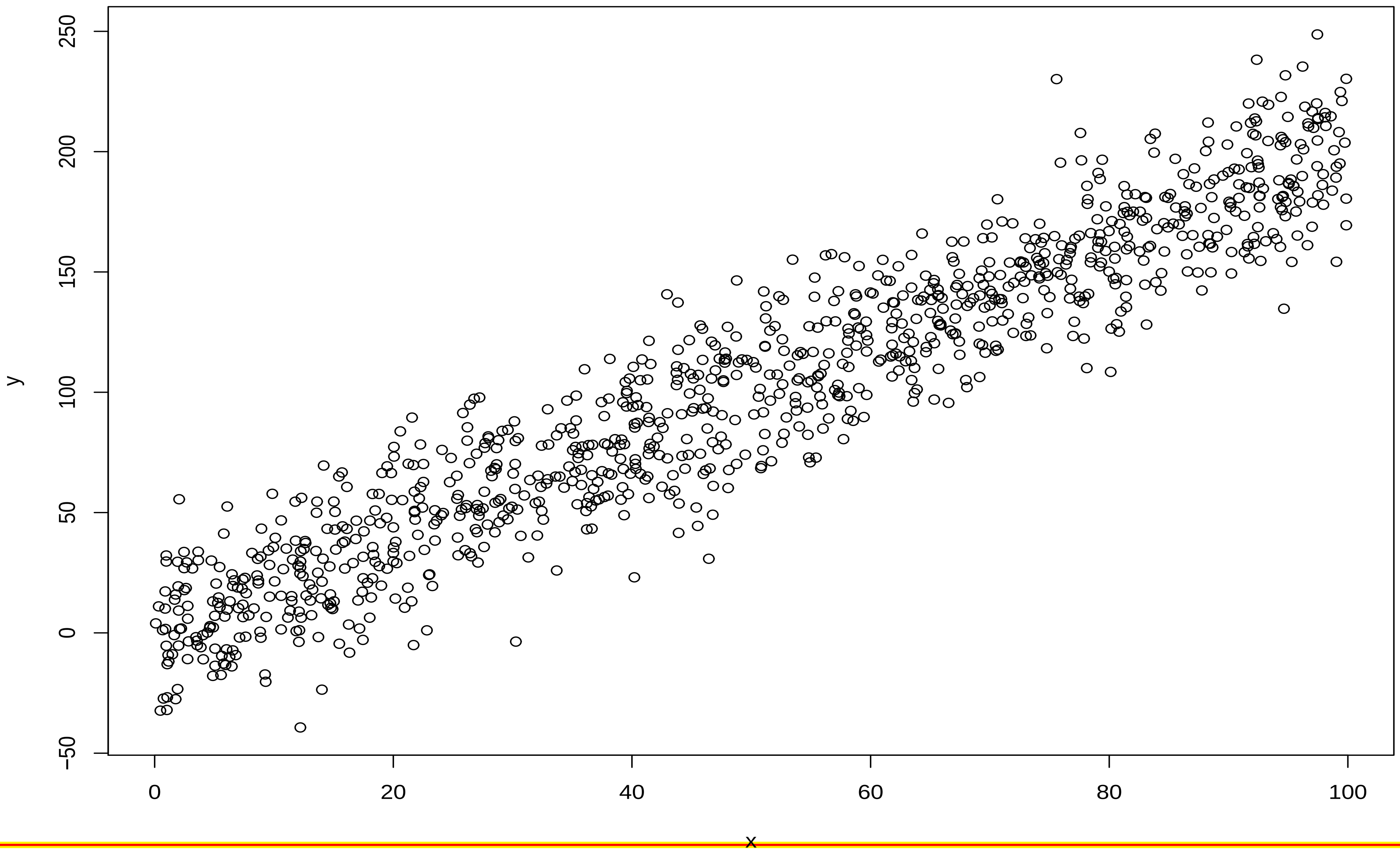
```
e<-rnorm(1000,0,10)
```

Now regenerate the Y_i values as $Y_i = \beta X_i + e_i$:

```
y<-beta*x+e
```

Now examine the plot of x and y.

Beta=2 Sigma=10



Linear Models

Repeat the process with $\sigma_e = 70$.

Generate 1,000 e_i values as normal random variables with mean zero and standard deviation 70:

e<-rnorm(1000,0,70)

Linear Models

Repeat the process with $\sigma_e = 70$.

Generate 1,000 e_i values as normal random variables with mean zero and standard deviation 70:

e<-rnorm(1000,0,70)

Now regenerate the Y_i values as $Y_i = \beta X_i + e_i$:

*y<-beta*x+e*

Linear Models

Repeat the process with $\sigma_e = 70$.

Generate 1,000 e_i values as normal random variables with mean zero and standard deviation 70:

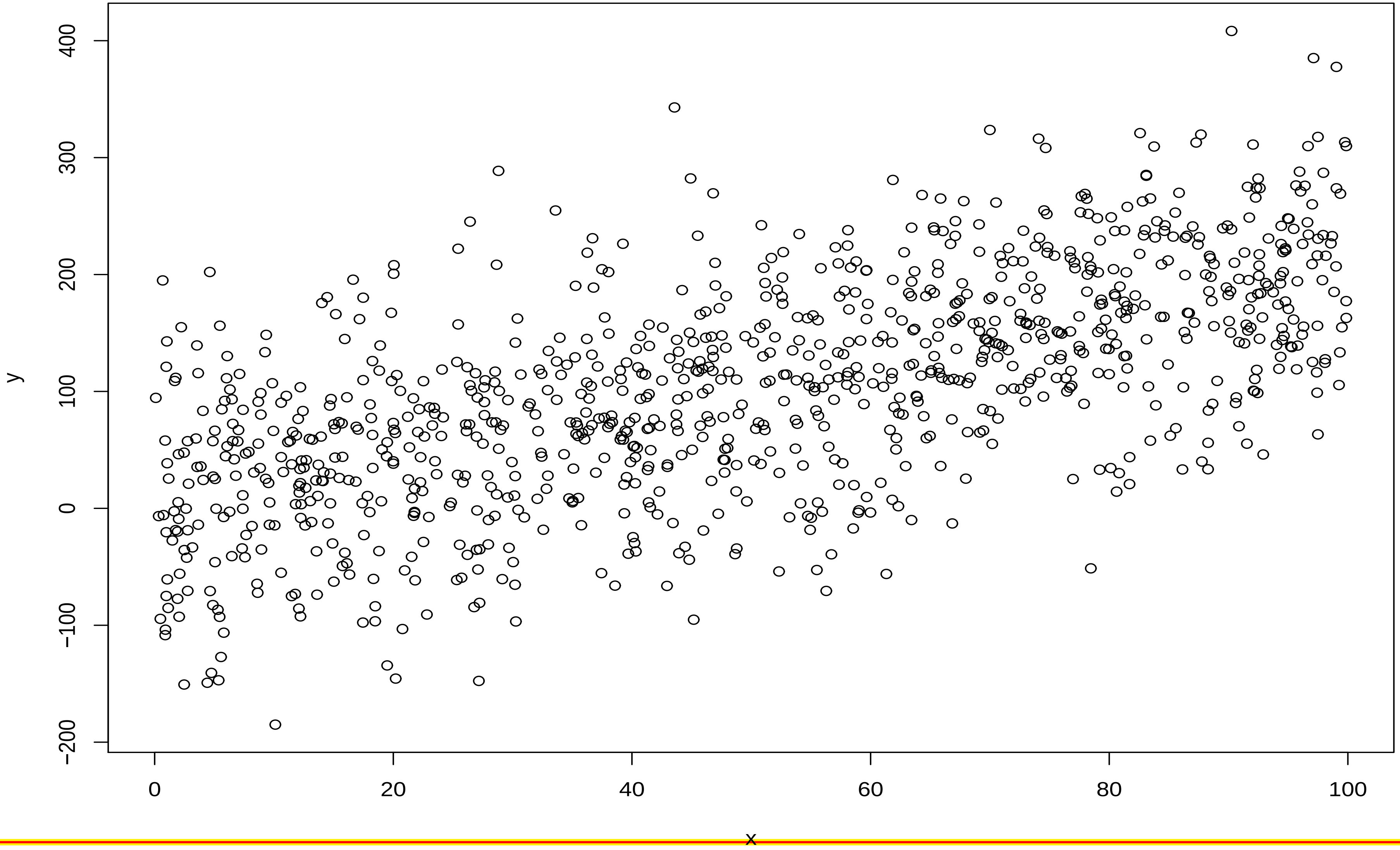
e<-rnorm(1000,0,70)

Now regenerate the Y_i values as $Y_i = \beta X_i + e_i$:

*y<-beta*x+e*

Now examine the plot of x and y.

Beta=2 Sigma=70



Linear Models

Next repeat the process with $\sigma_e = 150$.

Generate 1,000 e_i values as normal random variables with mean zero and standard deviation 150:

e<-rnorm(1000,0,150)

Linear Models

Next repeat the process with $\sigma_e = 150$.

Generate 1,000 e_i values as normal random variables with mean zero and standard deviation 150:

e<-rnorm(1000,0,150)

Now regenerate the Y_i values as $Y_i = \beta X_i + e_i$:

*y<-beta*x+e*

Linear Models

Next repeat the process with $\sigma_e = 150$.

Generate 1,000 e_i values as normal random variables with mean zero and standard deviation 150:

e<-rnorm(1000,0,150)

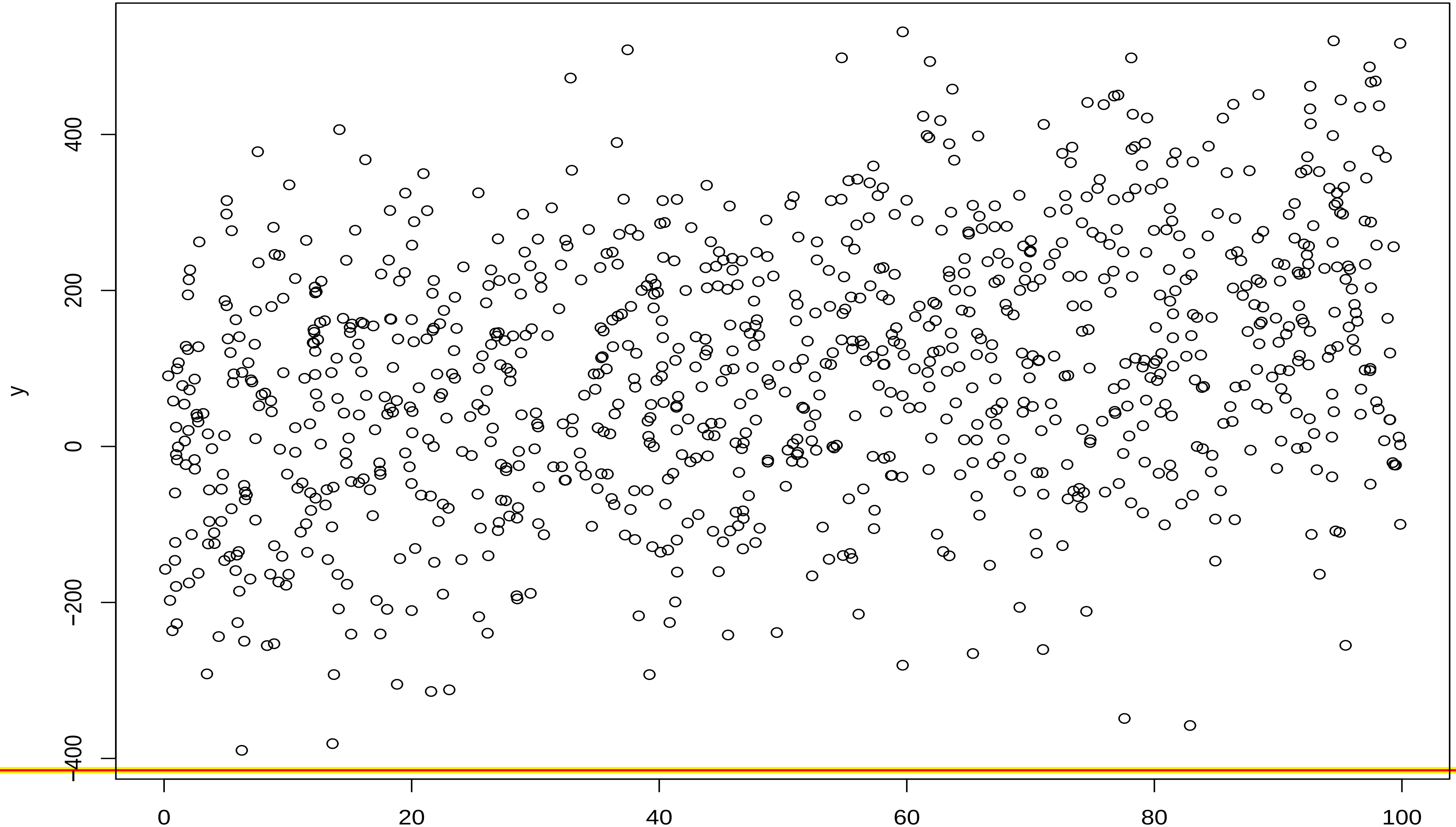
Now regenerate the Y_i values as $Y_i = \beta X_i + e_i$:

*y<-beta*x+e*

Now examine the plot of x and y.

Beta=2 Sigma=150

The trend is less obvious as the "noise" level increases:



Linear Models

Finally repeat the process with $\sigma_e = 1000$.

Generate 1,000 e_i values as normal random variables with mean zero and standard deviation 1000:

e<-rnorm(1000,0,1000)

Linear Models

Finally repeat the process with $\sigma_e = 1000$.

Generate 1,000 e_i values as normal random variables with mean zero and standard deviation 1000:

e<-rnorm(1000,0,1000)

Now regenerate the Y_i values as $Y_i = \beta X_i + e_i$:

*y<-beta*x+e*

Linear Models

Finally repeat the process with $\sigma_e = 1000$.

Generate 1,000 e_i values as normal random variables with mean zero and standard deviation 1000:

e<-rnorm(1000,0,1000)

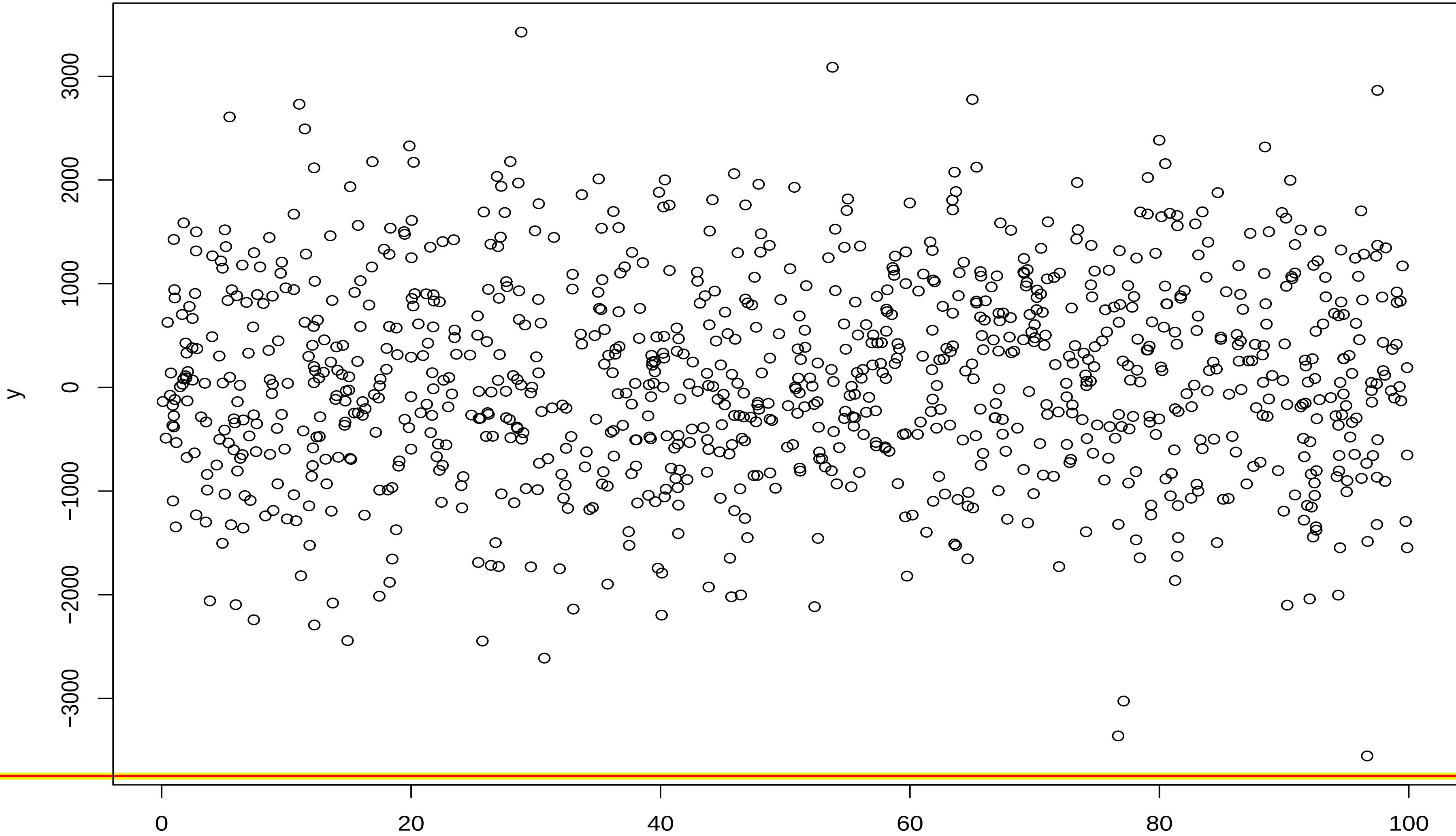
Now regenerate the Y_i values as $Y_i = \beta X_i + e_i$:

*y<-beta*x+e*

Now examine the plot of x and y.

Beta=2 Sigma=1000

By now the trend is barely discernable:



Discrete vs Continous Predictors

We just examined a model of the form

$$Y_i = \beta X_i + e_i$$

Discrete vs Continous Predictors

We just examined a model of the form

$$Y_i = \beta X_i + e_i$$

Notice that both Y_i and X_i are continuous, that is, they can assume any value.

Discrete vs Continous Predictors

We just examined a model of the form

$$Y_i = \beta X_i + e_i$$

Notice that both Y_i and X_i are continuous, that is, they can assume any value.

Now we examine a variation that allows us to use a linear model to compare the means of more than two groups.

Discrete vs Continous Predictors

We just examined a model of the form

$$Y_i = \beta X_i + e_i$$

Notice that both Y_i and X_i are continuous, that is, they can assume any value.

Now we examine a variation that allows us to use a linear model to compare the means of more than two groups.

We have developed techniques for comparing the means of two populations, but what we consider next will apply to more general types of comparisons.

Discrete Predictors

Suppose we want to compare the means of three groups, and we have samples from each group.

Discrete Predictors

Suppose we want to compare the means of three groups, and we have samples from each group.

This time the linear model looks like this:

$$Y_i = \mu + \beta_1 X_{1i} + \beta_2 X_{2i} + \beta_3 X_{3i} + e_i$$

Where:

- X_{1i} equals 1 if Y_i is in group 1, and zero otherwise
- X_{2i} equals 1 if Y_i is in group 2, and zero otherwise
- X_{3i} equals 1 if Y_i is in group 3, and zero otherwise
- μ, β_1, β_2, and β_3 are parameters (constants)

Discrete Predictors

As before, the X_{ij}s and βs are constants, and the e_is are random variables with mean $\mu_e = 0$ and standard deviation σ_e. Then because the X_{ij} values corresponding to the groups Y_i does not belong to are zero, we can write:

- $Y_i = \mu + \beta_1 X_{1i} + e_i$ if Y_i is in group 1
- $Y_i = \mu + \beta_2 X_{2i} + e_i$ if Y_i is in group 2
- $Y_i = \mu + \beta_3 X_{3i} + e_i$ if Y_i is in group 3

Discrete Predictors

As before, the X_{ij}s and βs are constants, and the e_is are random variables with mean $\mu_e = 0$ and standard deviation σ_e. Then because the X_{ij} values corresponding to the groups Y_i does not belong to are zero, we can write:

- $Y_i = \mu + \beta_1 X_{1i} + e_i$ if Y_i is in group 1
- $Y_i = \mu + \beta_2 X_{2i} + e_i$ if Y_i is in group 2
- $Y_i = \mu + \beta_3 X_{3i} + e_i$ if Y_i is in group 3

The expected values for the Y_is are:

- $E(Y_i) = \mu + \beta_1 X_{1i} \; = \; \mu + \beta_1$ if Y_i is in group 1
- $E(Y_i) = \mu + \beta_2 X_{2i} \; = \; \mu + \beta_2$ if Y_i is in group 2
- $E(Y_i) = \mu + \beta_3 X_{3i} \; = \; \mu + \beta_3$ if Y_i is in group 3

Discrete Predictors

For example, suppose there are 9 data values in the sample, 3 from each group.

Then the Y and X values are:

Y	X_{1i}	X_{2i}	X_{3i}
Y_1	1	0	0
Y_2	1	0	0
Y_3	1	0	0
Y_4	0	1	0
Y_5	0	1	0
Y_6	0	1	0
Y_7	0	0	1
Y_8	0	0	1
Y_9	0	0	1

Discrete Predictors

Now we will construct artificial data with $3,000$ observations, $1,000$ in each of three groups with the following characteristics:

- $\mu = 1$
- $\beta_1 = 1$
- $\beta_2 = 3$
- $\beta_3 = 5$
- $\sigma_e = 1$

Discrete Predictors

The expected values for the three groups are:

- $E(Y_i) = \mu + \beta_1 \;=\; 1 + 1 = 2$ if Y_i is in group 1
- $E(Y_i) = \mu + \beta_2 \;=\; 1 + 3 = 4$ if Y_i is in group 2
- $E(Y_i) = \mu + \beta_3 \;=\; 1 + 5 = 6$ if Y_i is in group 3

Generating the Data

As before we will use R to generate the model.

First generate 1000 values for each of three groups, with values 1, 3, and 5:

```
x<-c(rep(1,1000),rep(3,1000),rep(5,1000))
```

Generating the Data

As before we will use R to generate the model.

First generate 1000 values for each of three groups, with values 1, 3, and 5:

```
x<-c(rep(1,1000),rep(3,1000),rep(5,1000))
```

Next assign a value to μ. We'll use 1:

```
mu<-1
```

Generating the Data

As before we will use R to generate the model.

First generate 1000 values for each of three groups, with values 1, 3, and 5:

```
x<-c(rep(1,1000),rep(3,1000),rep(5,1000))
```

Next assign a value to μ. We'll use 1:

```
mu<-1
```

Now generate 3,000 e_i values as normal random variables with mean zero and standard deviation 1:

```
e<-rnorm(3000,0,1)
```

Generating the Data

As before we will use R to generate the model.

First generate 1000 values for each of three groups, with values 1, 3, and 5:

x<-c(rep(1,1000),rep(3,1000),rep(5,1000))

Next assign a value to μ. We'll use 1:

mu<-1

Now generate 3,000 e_i values as normal random variables with mean zero and standard deviation 1:

e<-rnorm(3000,0,1)

Now generate the Y_i values:

y<-mu+x+e

Generating the Data

Now generate the group labels:

group <- gl(3,1000,3000,
labels=c("Group1","Group2","Group3"))

Generating the Data

Now generate the group labels:

group <- gl(3,1000,3000,
labels=c("Group1","Group2","Group3"))

Finally, produce a box plot of the data:

boxplot$(y \sim group)$

mu=1 beta1=1 beta2=2 beta3=3

The following is a boxplot of the data: