spncrfx@gmail.com

@foxandtheflu

Introduction to Probability and Likelihood in R

Spencer Fox
14 October 2016

Week 6 goals

Week 6 goals

- Working with probability distributions

Week 6 goals

- Working with probability distributions
- Likelihood
  - Calculating likelihoods
  - Interpreting likelihood results

R probability distribution functions are extremely consistent

R probability distribution functions are extremely consistent

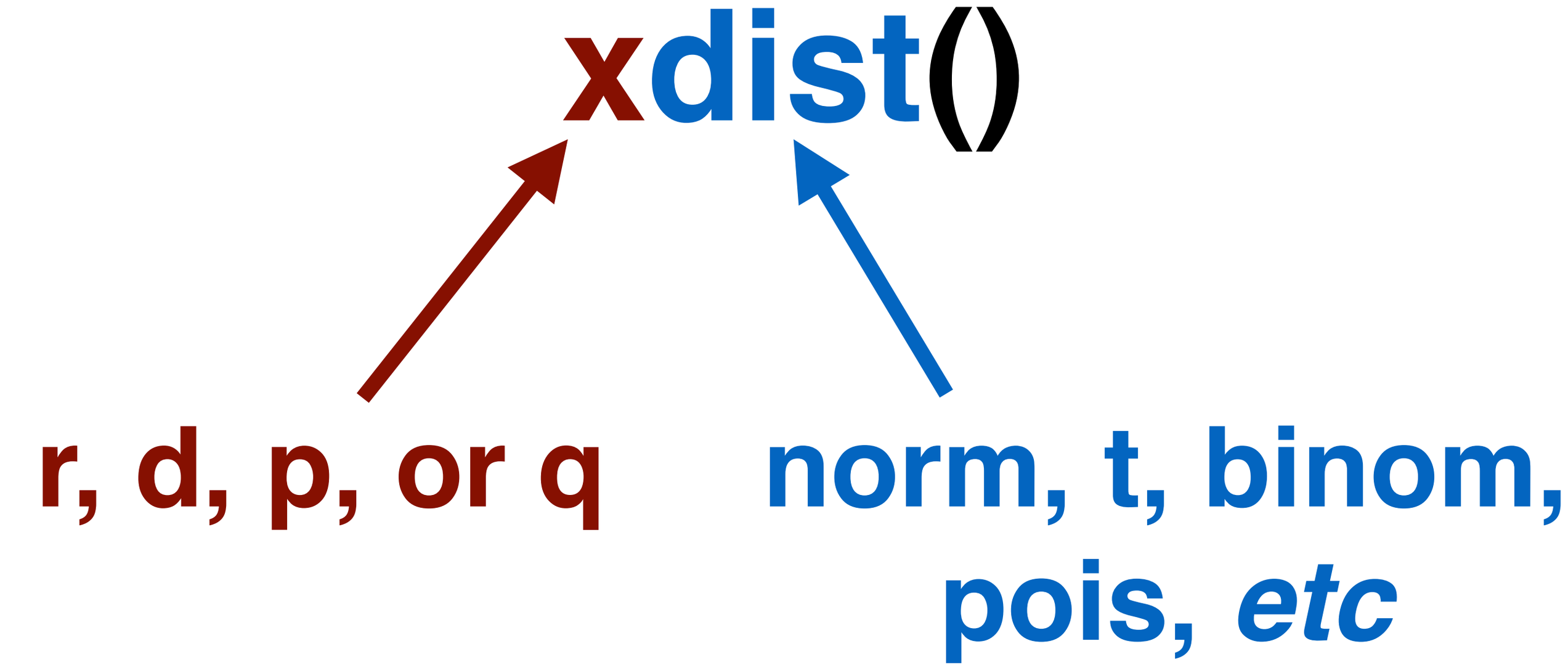
xdist()

R probability distribution functions are extremely consistent

xdist()

r, d, p, or q

R probability distribution functions are extremely consistent



Drawing (r)andom (norm)al samples

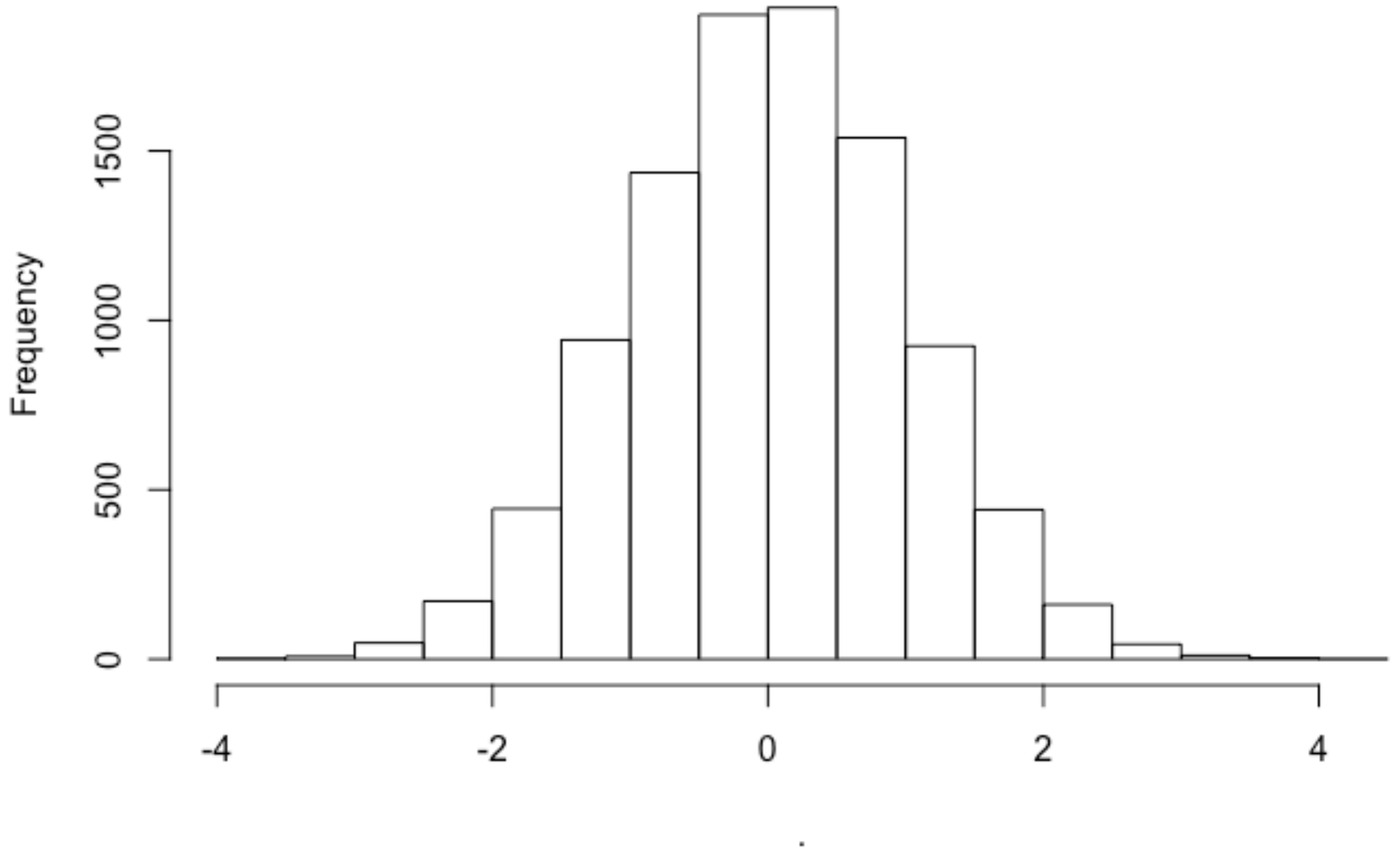
```
rnorm(n = 10000, mean = 0, sd = 1)
```

Drawing (r)andom (norm)al samples

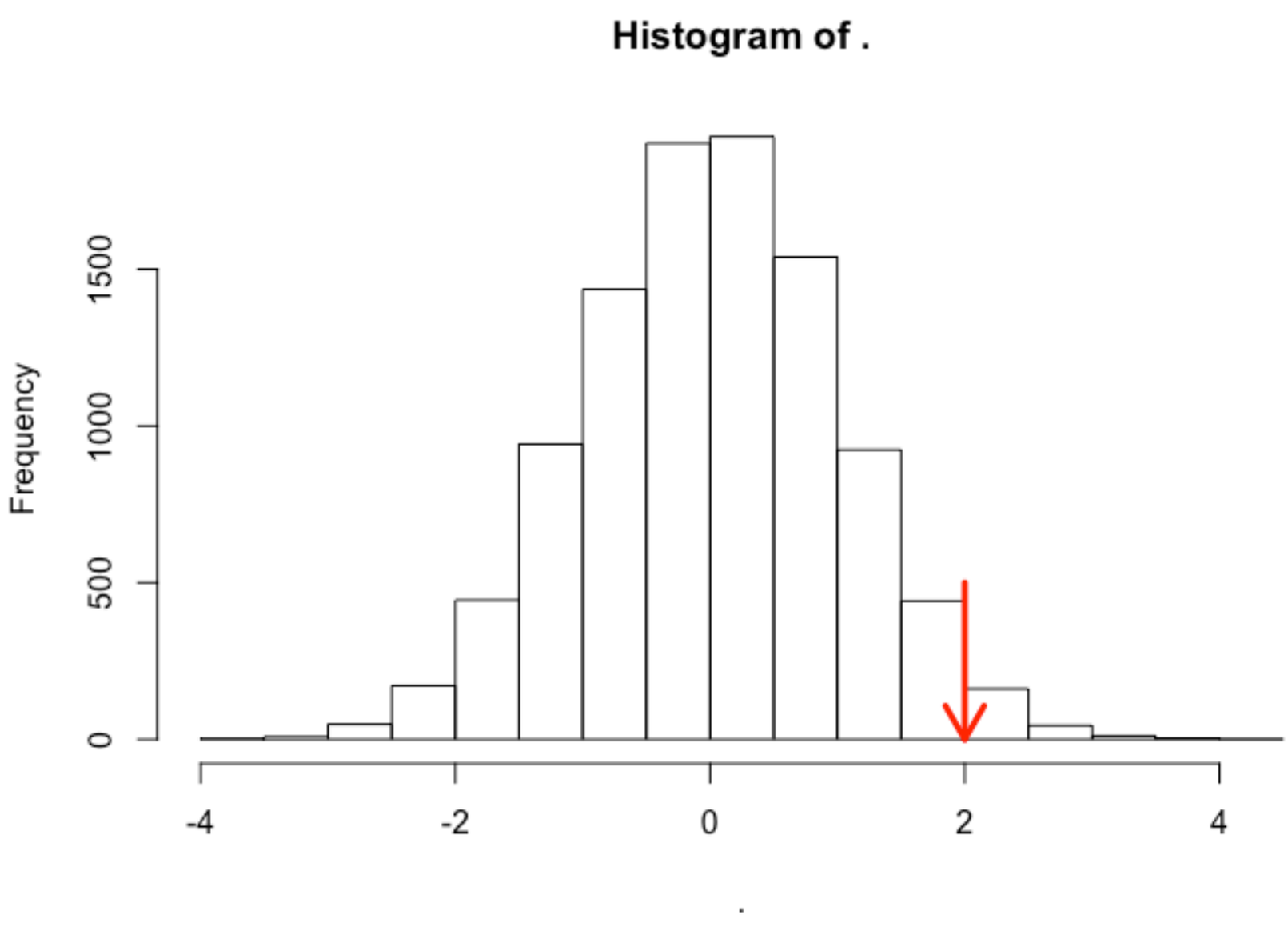
```
rnorm(n = 10000, mean = 0, sd = 1) %>%
  hist()
```

Drawing (r)andom (norm)al samples

```
rnorm(n = 10000, mean = 0, sd = 1) %>%
  hist()
```



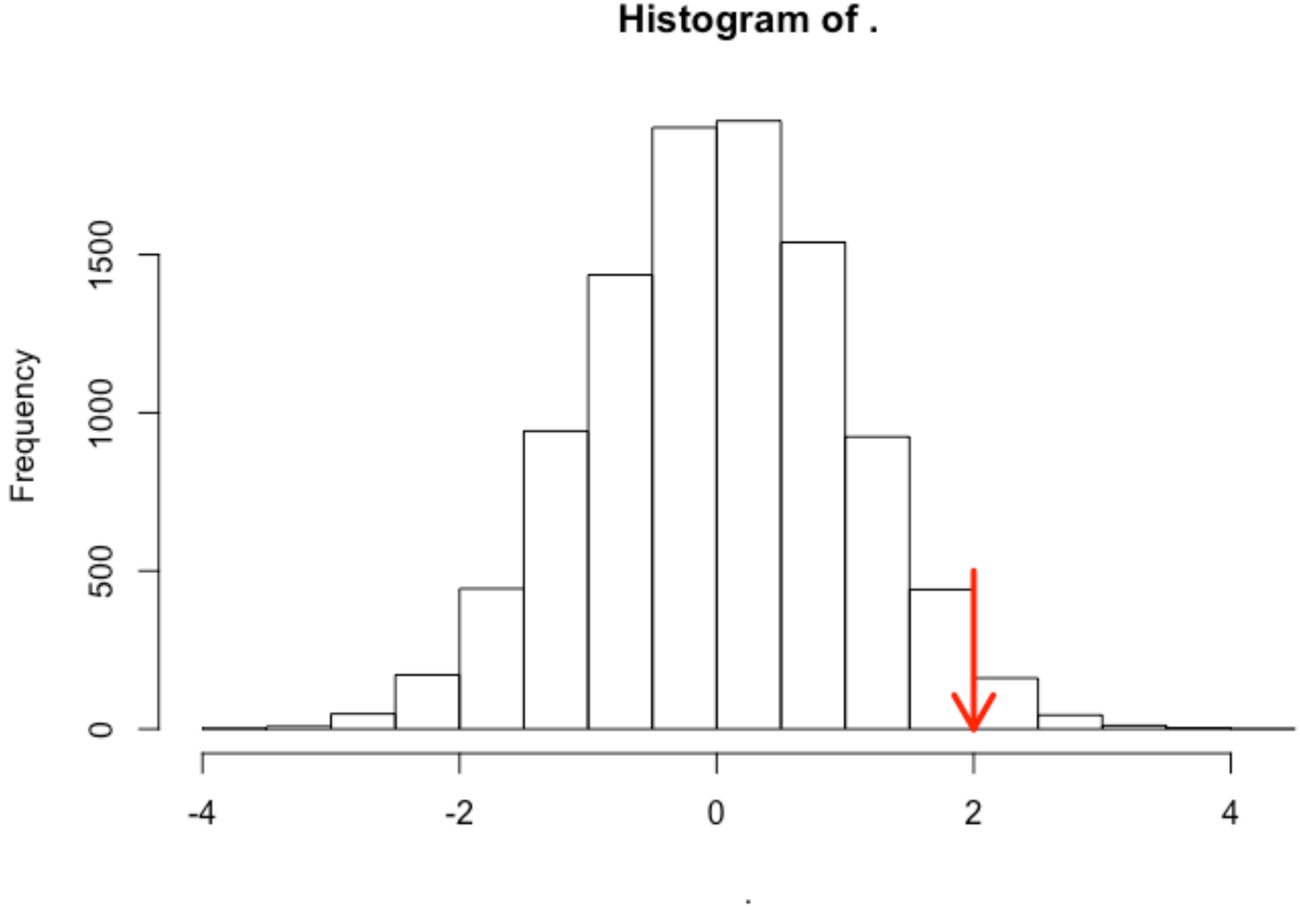
Getting the (d)ensity or likelihood of a data point



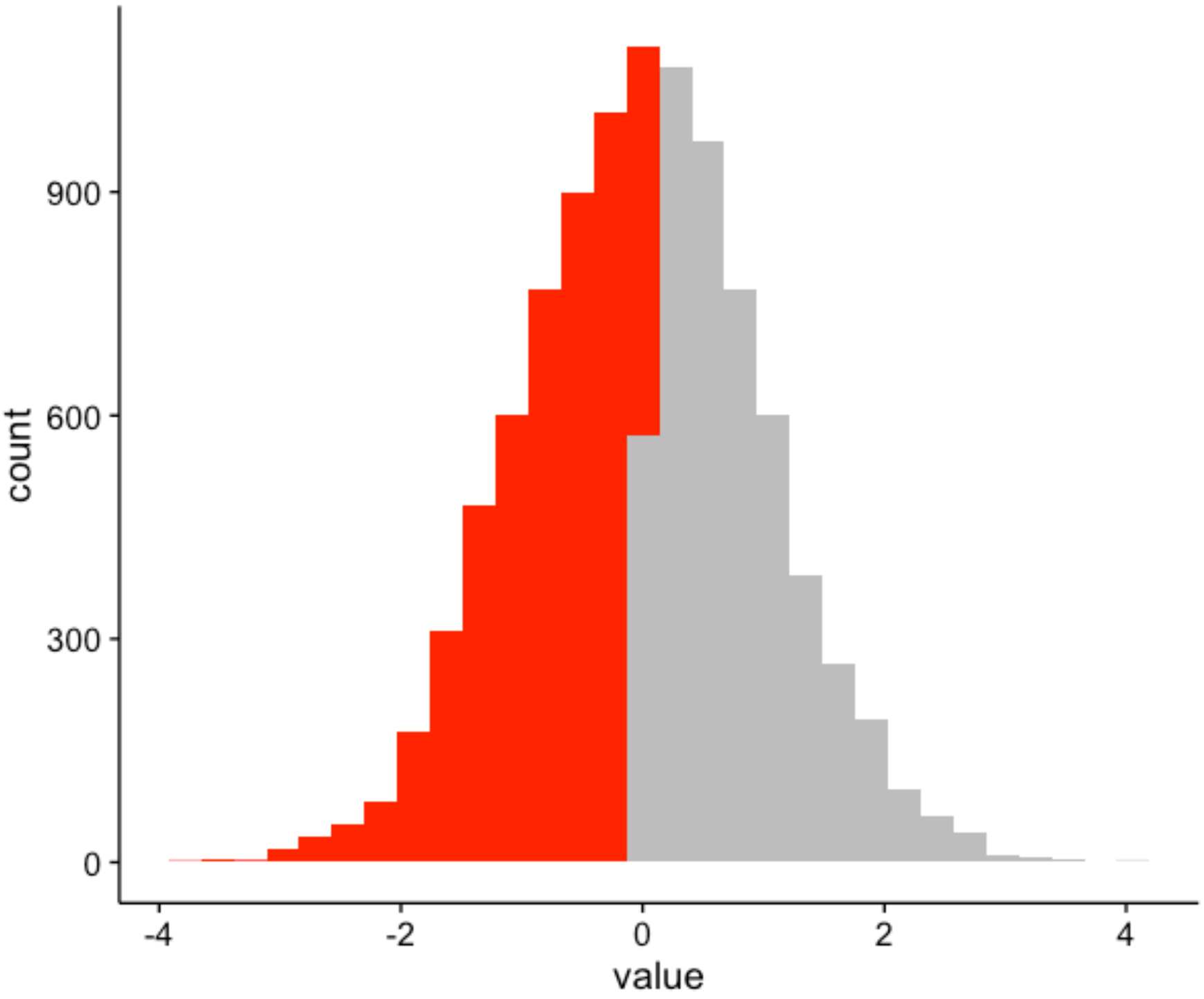
Getting the (d)ensity or likelihood of a data point

```
dnorm(x = 2)
```

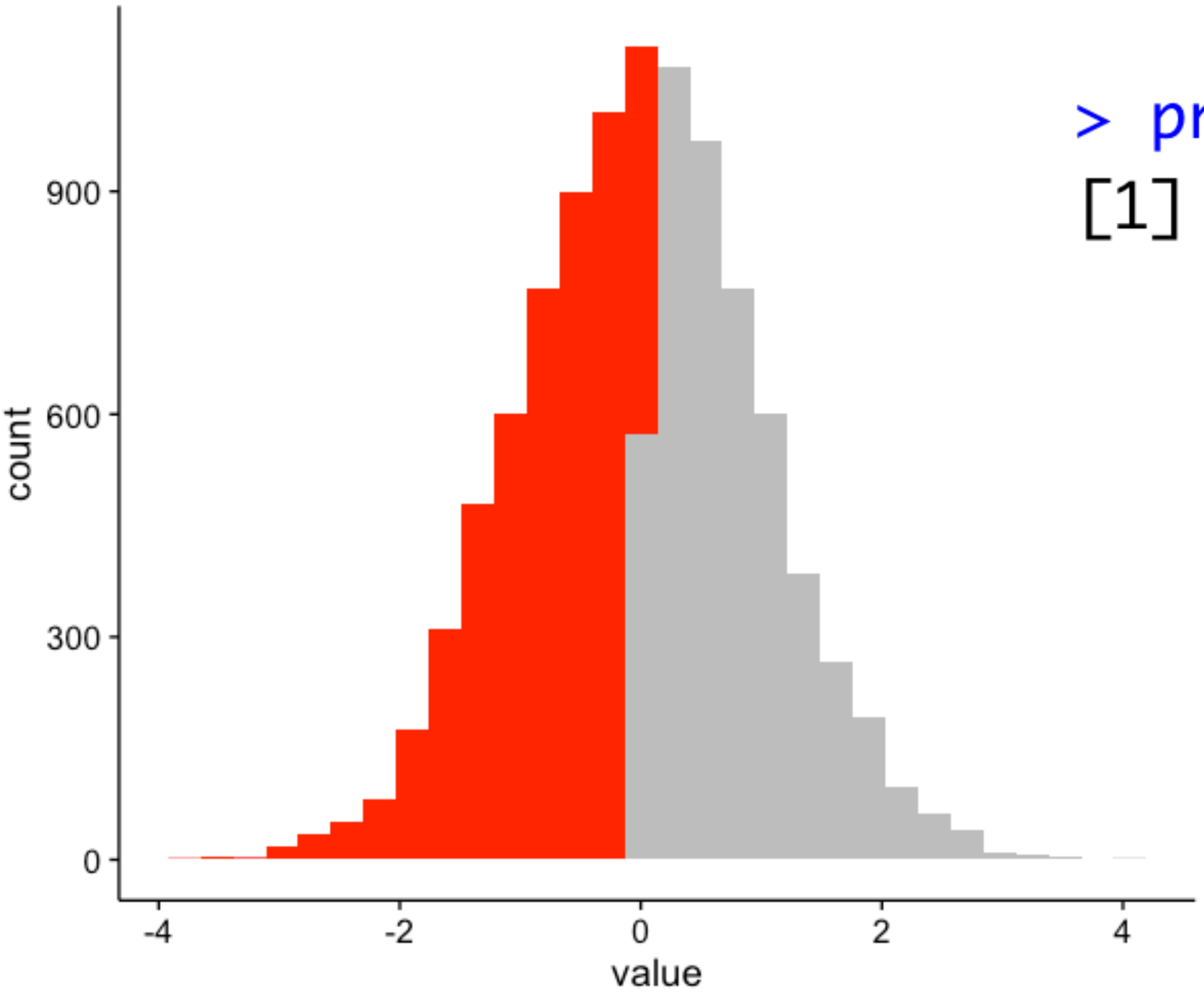
```
[1] 0.05399097
```



Getting cumulative (p)robability and the inverse (q)umulative probability

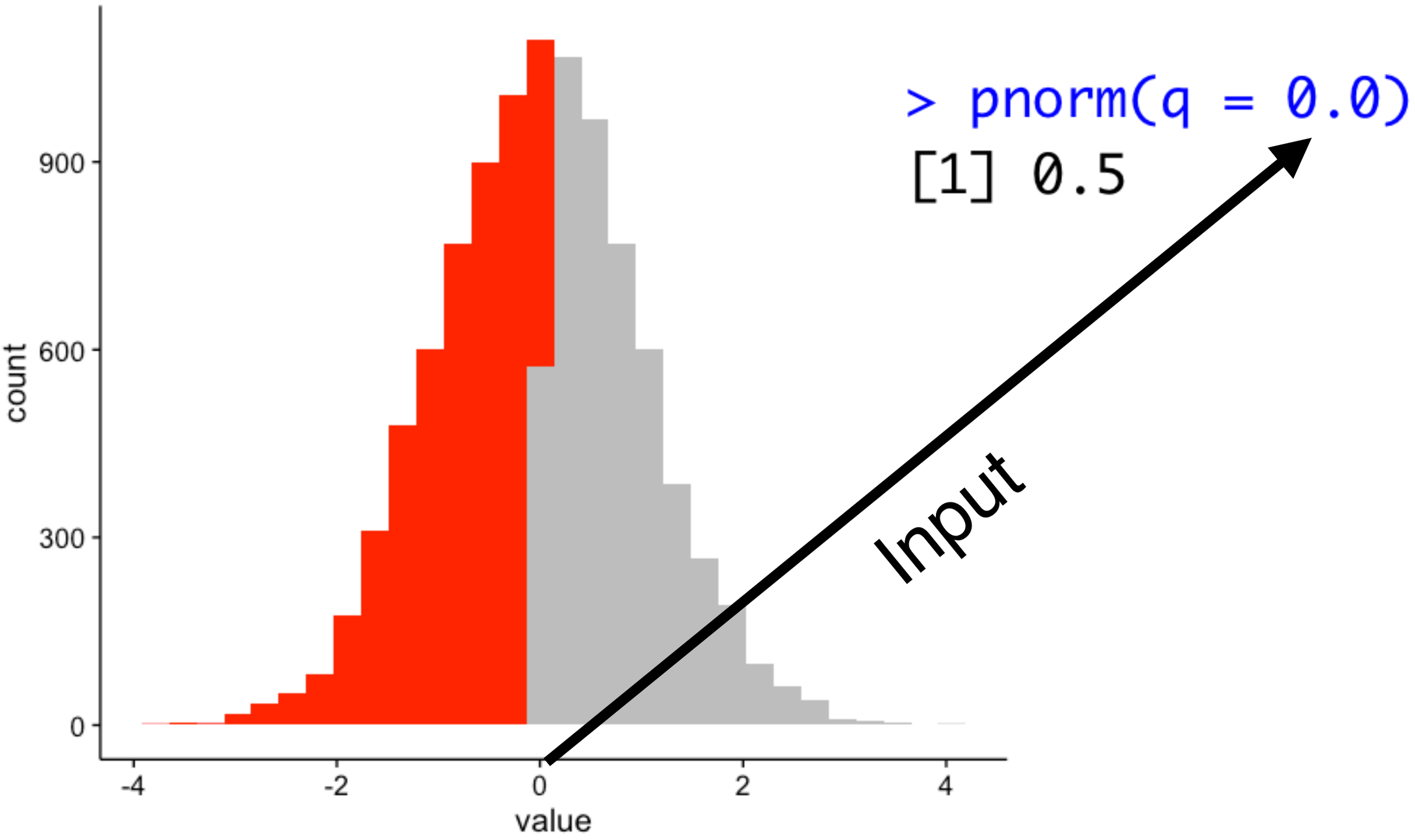


Getting cumulative (p)robability and the inverse (q)umulative probability

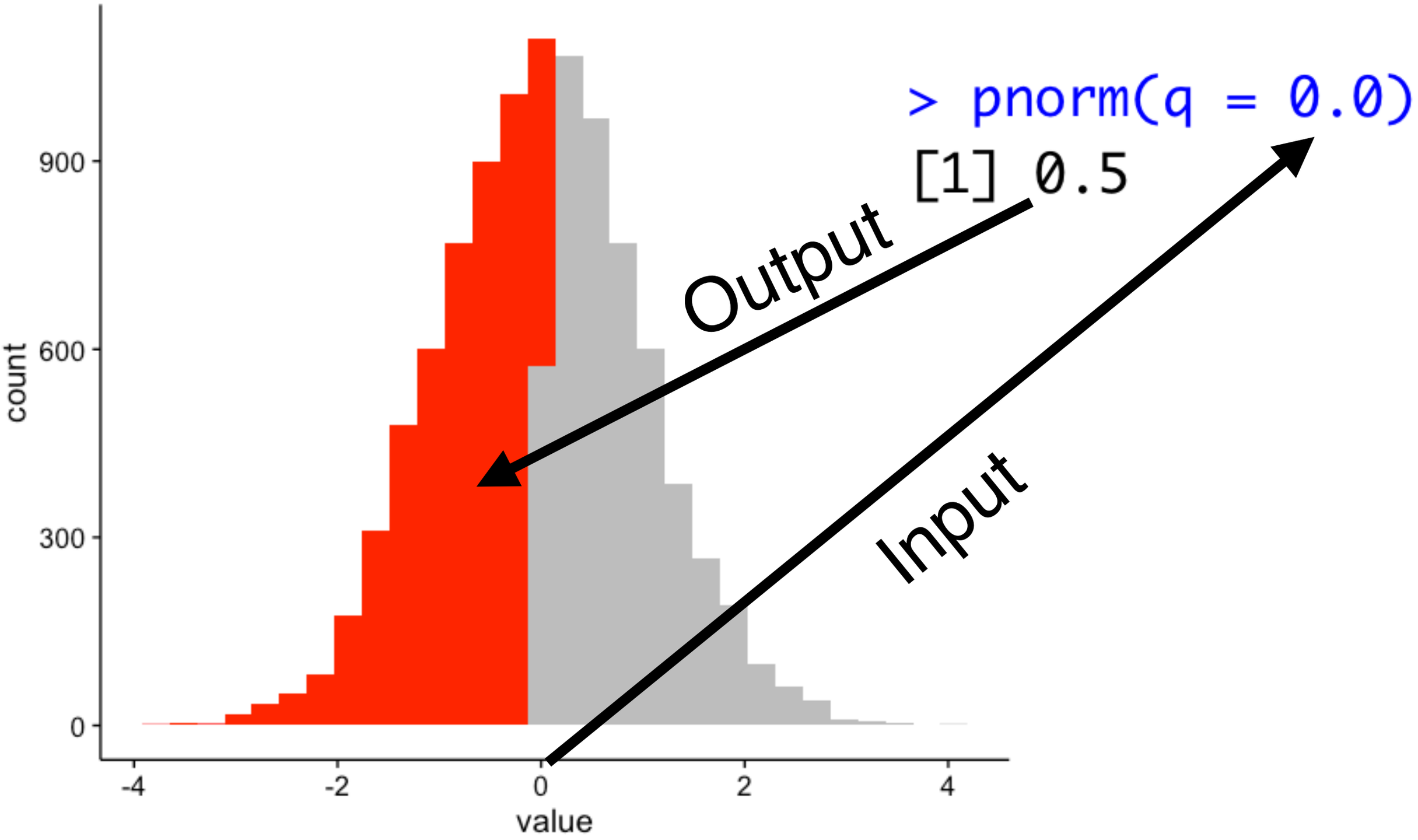



```
> pnorm(q = 0.0)
[1] 0.5
```

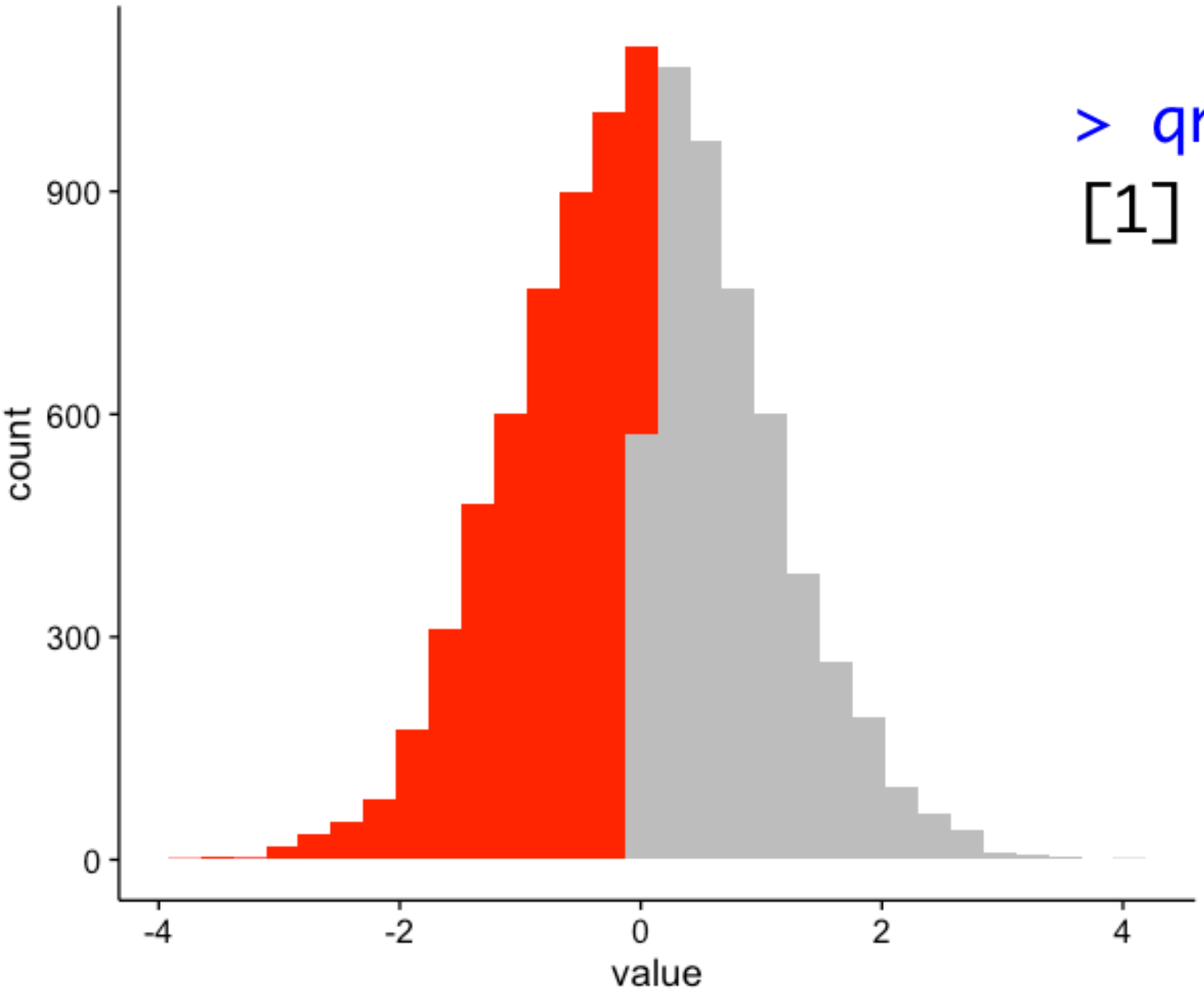
Getting cumulative (p)robability and the inverse (q)umulative probability



Getting cumulative (p)robability and the inverse (q)umulative probability

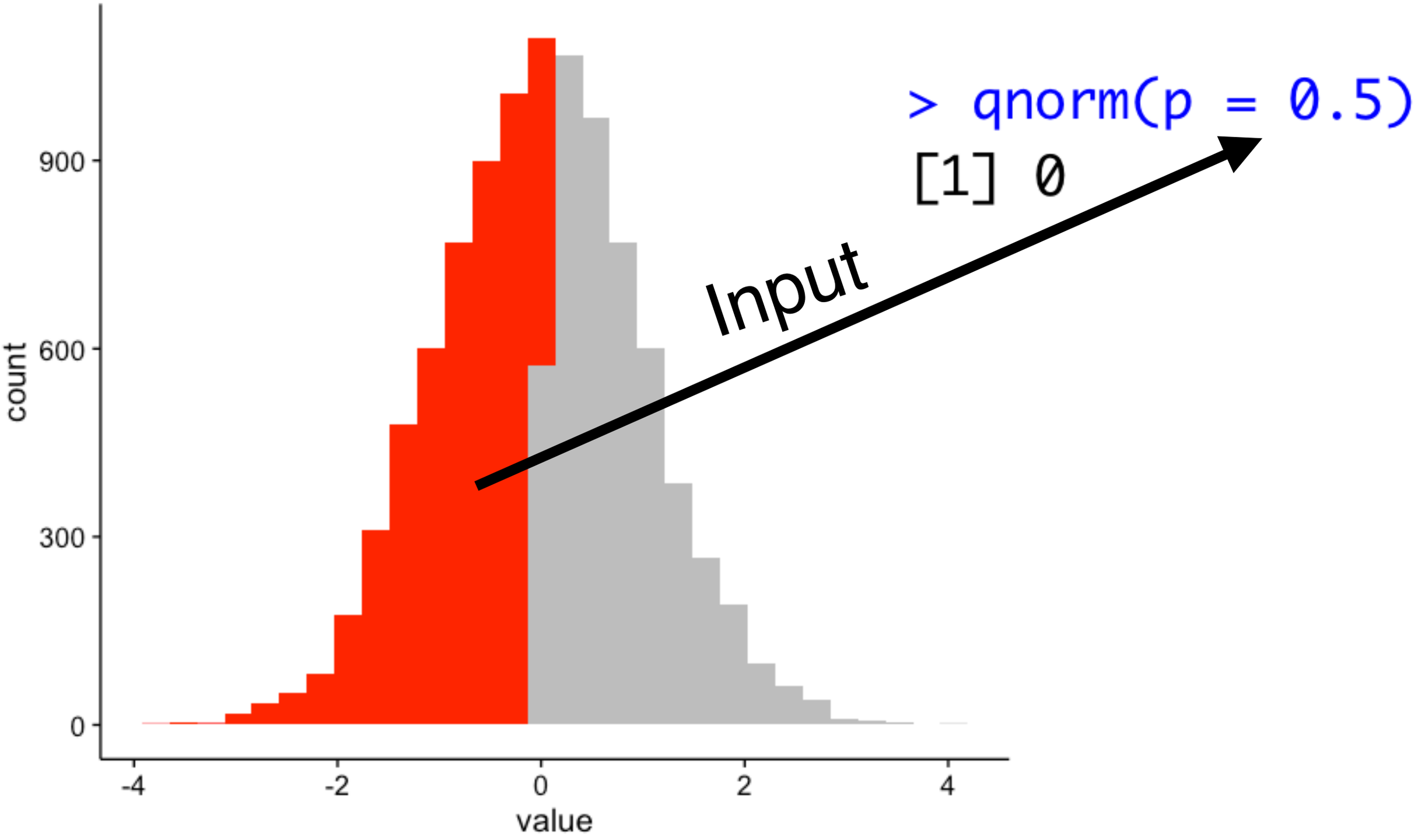


Getting cumulative (p)robability and the inverse (q)umulative probability

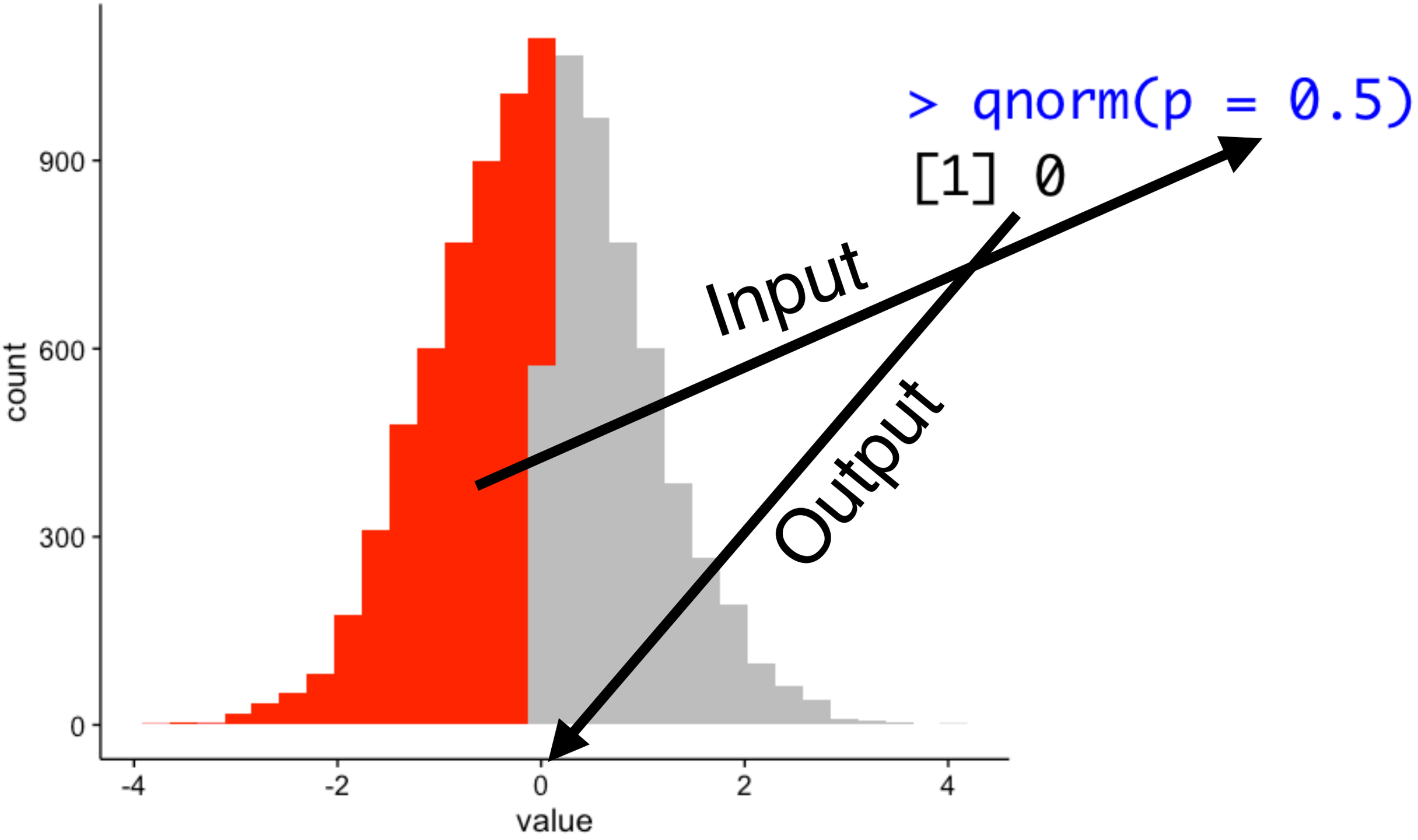



```
> qnorm(p = 0.5)
[1] 0
```

Getting cumulative (p)robability and the inverse (q)umulative probability



Getting cumulative (p)robability and the inverse (q)umulative probability



1st Programming Exercise

Likelihood is the probability of your data under a given model

$$P(Y \mid \theta)$$

Likelihood is the probability of your data under a given model

$$P(Y \mid \theta)$$

Data

Likelihood is the probability of your data under a given model

$$P(Y \mid \theta)$$

Data

Model

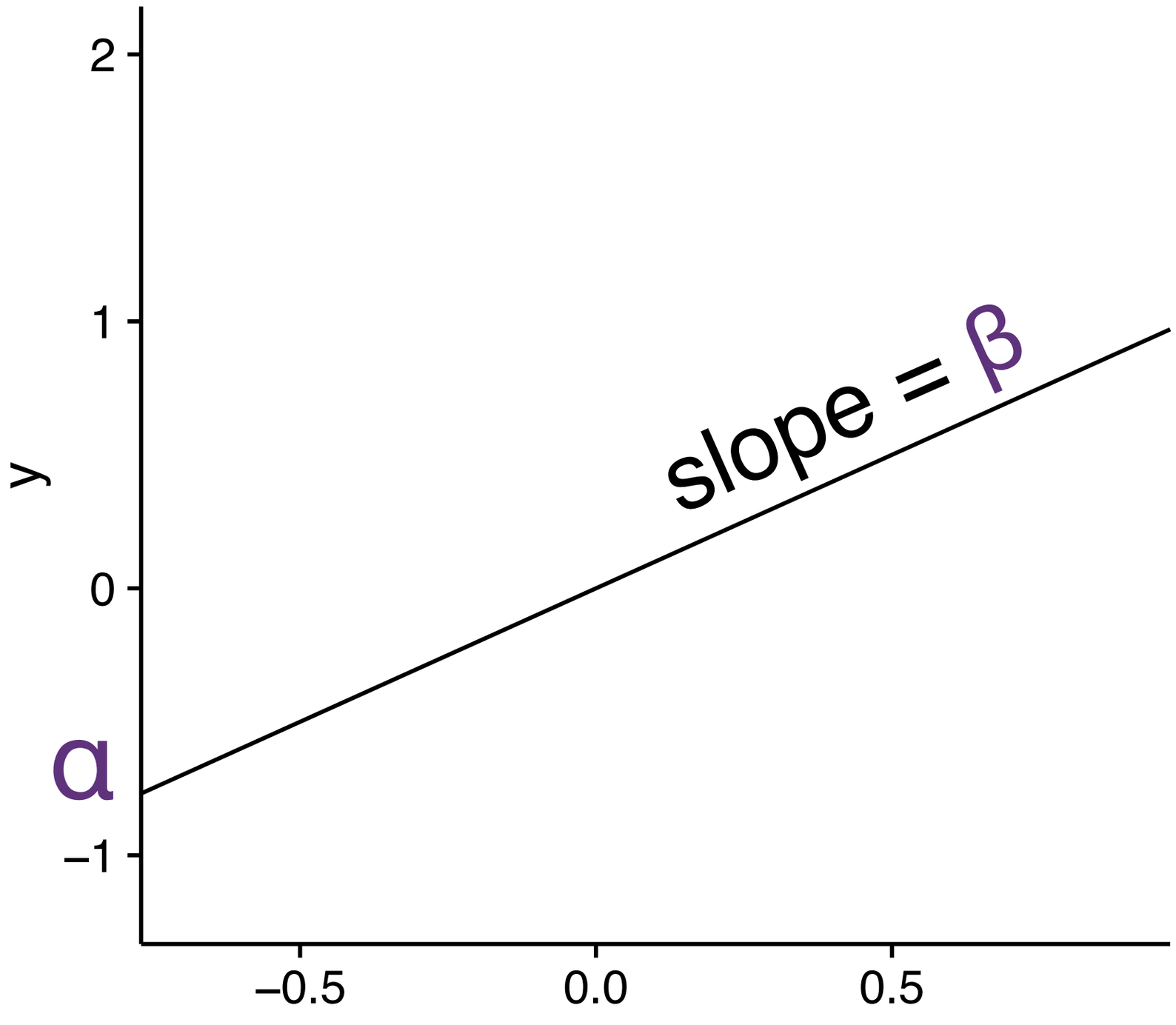
We often denote the parameters and not the model itself

$$P(Y \mid \beta, \alpha)$$

Data **Parameters**

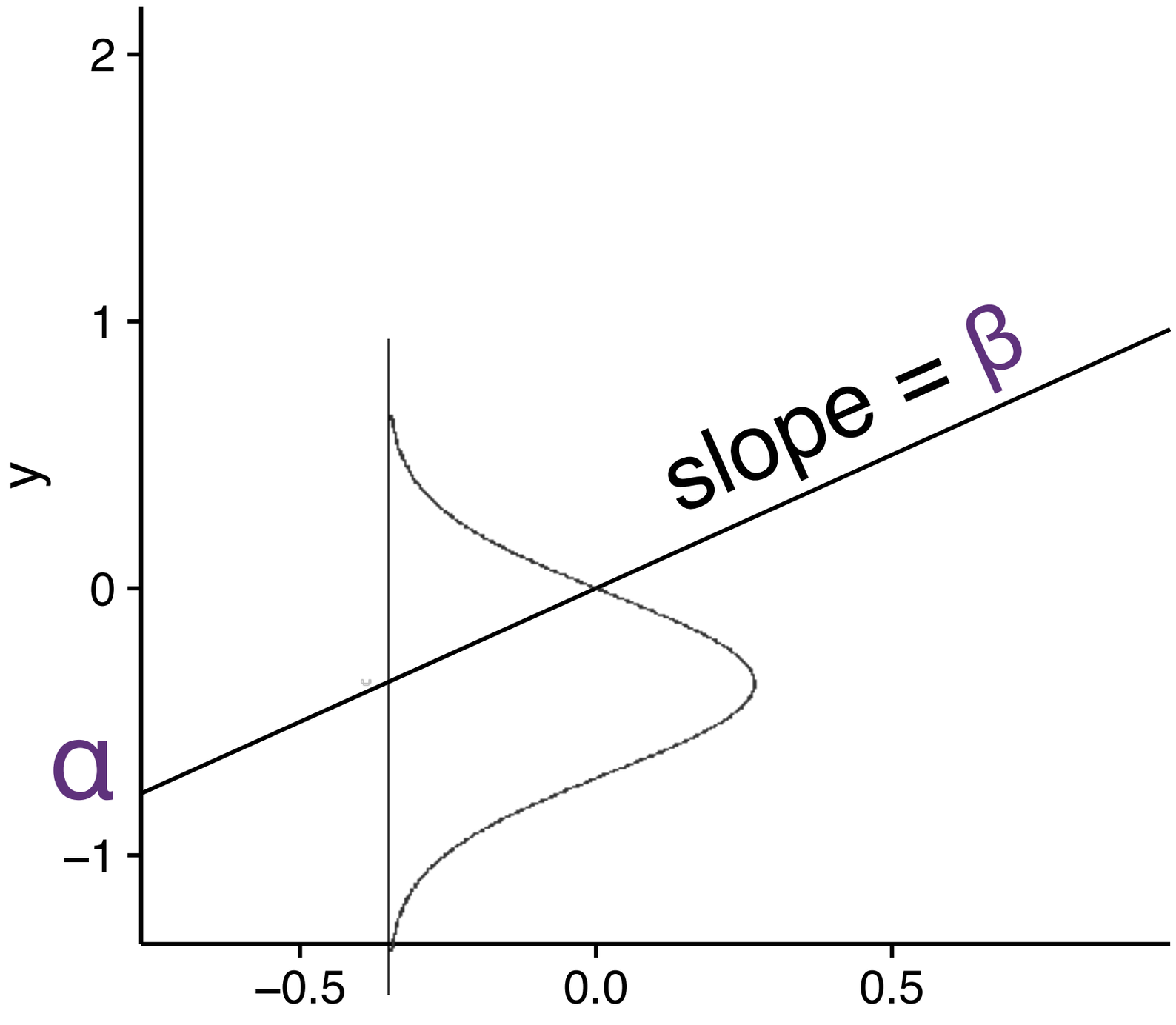
Our first model: linear regression

$$Y = \beta X + \alpha$$



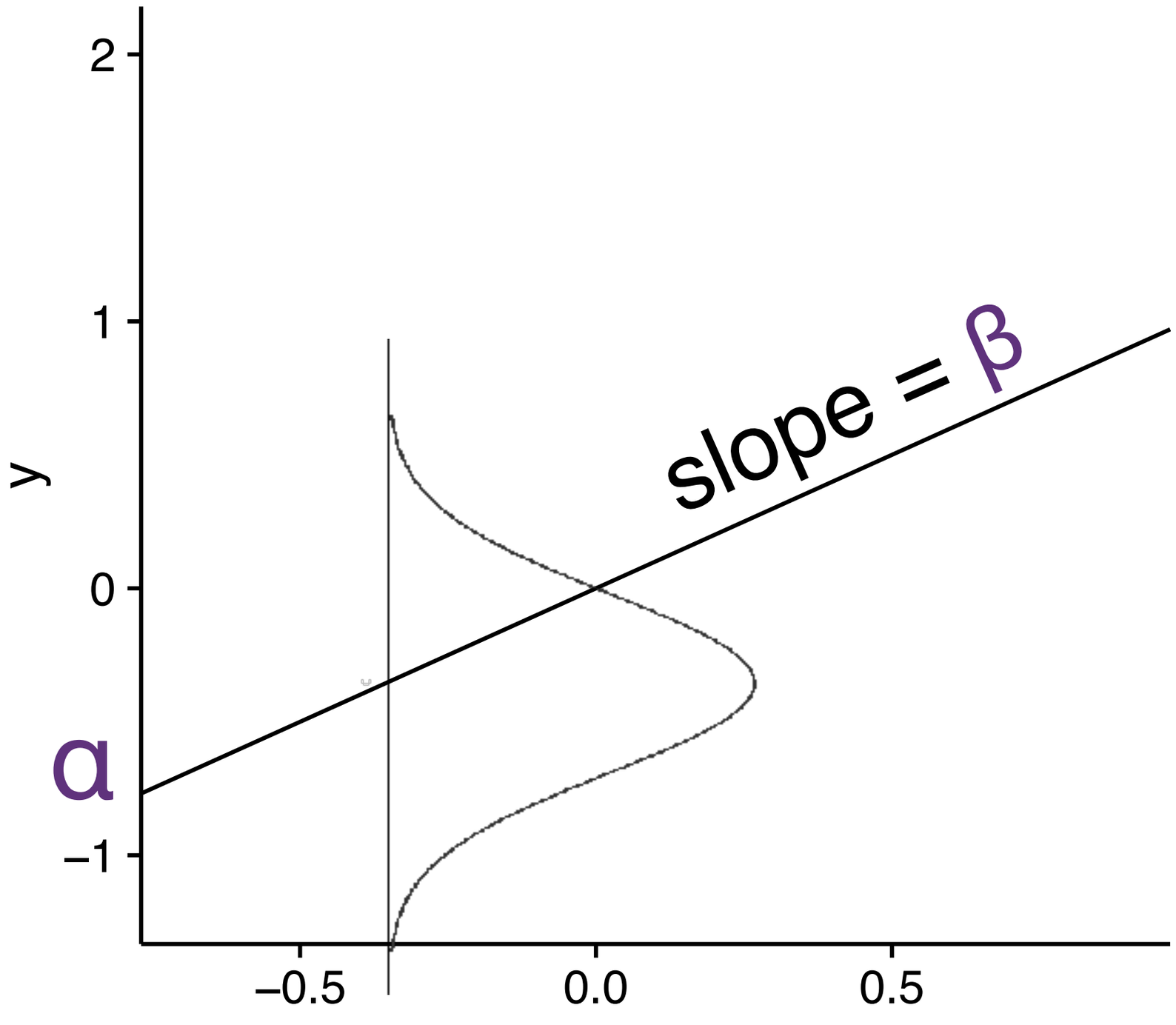
Our first model: linear regression

$$Y = \beta X + \alpha$$



Our first model: linear regression

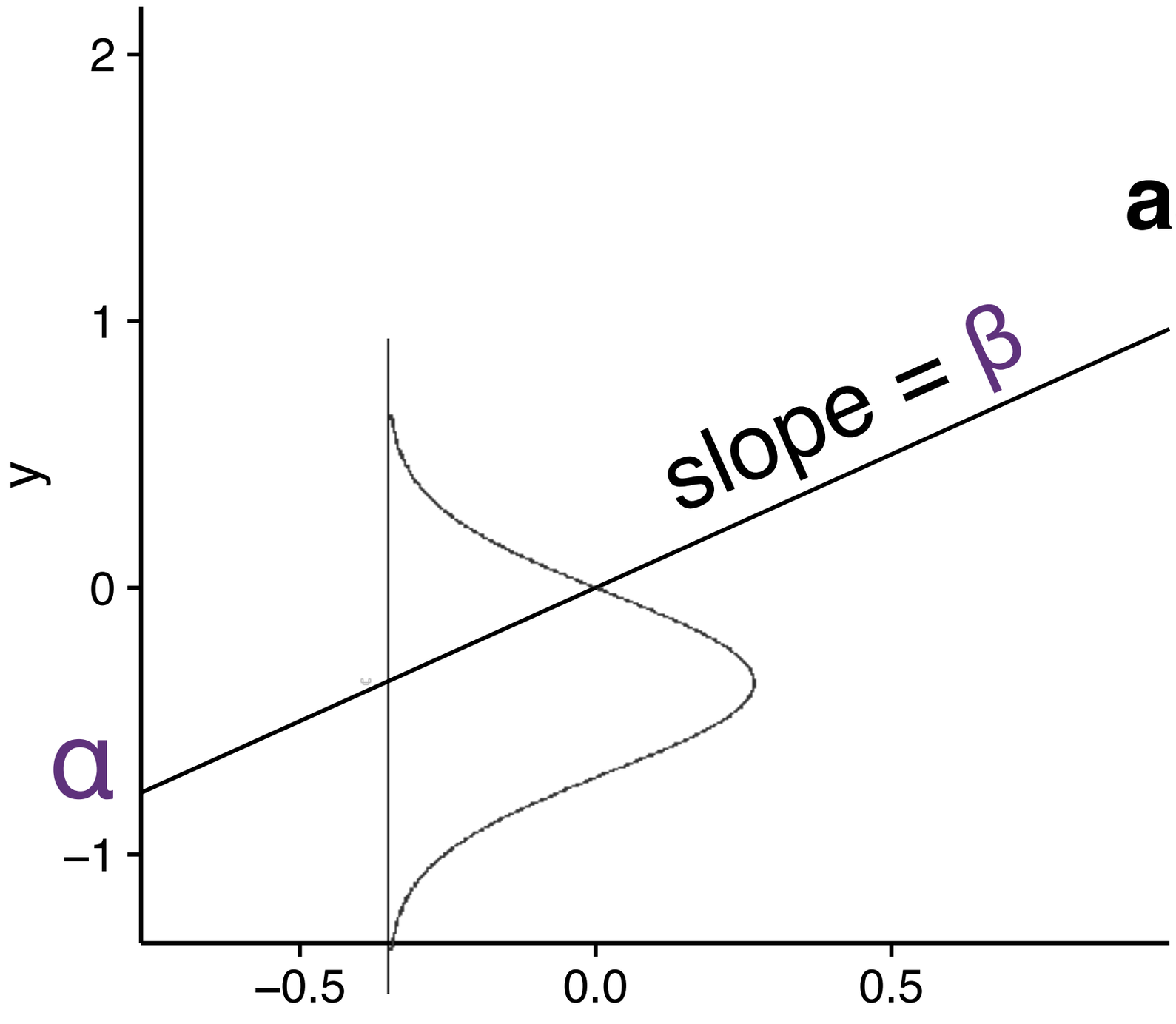
$$Y_i \sim N(\beta X_i + \alpha, \sigma)$$



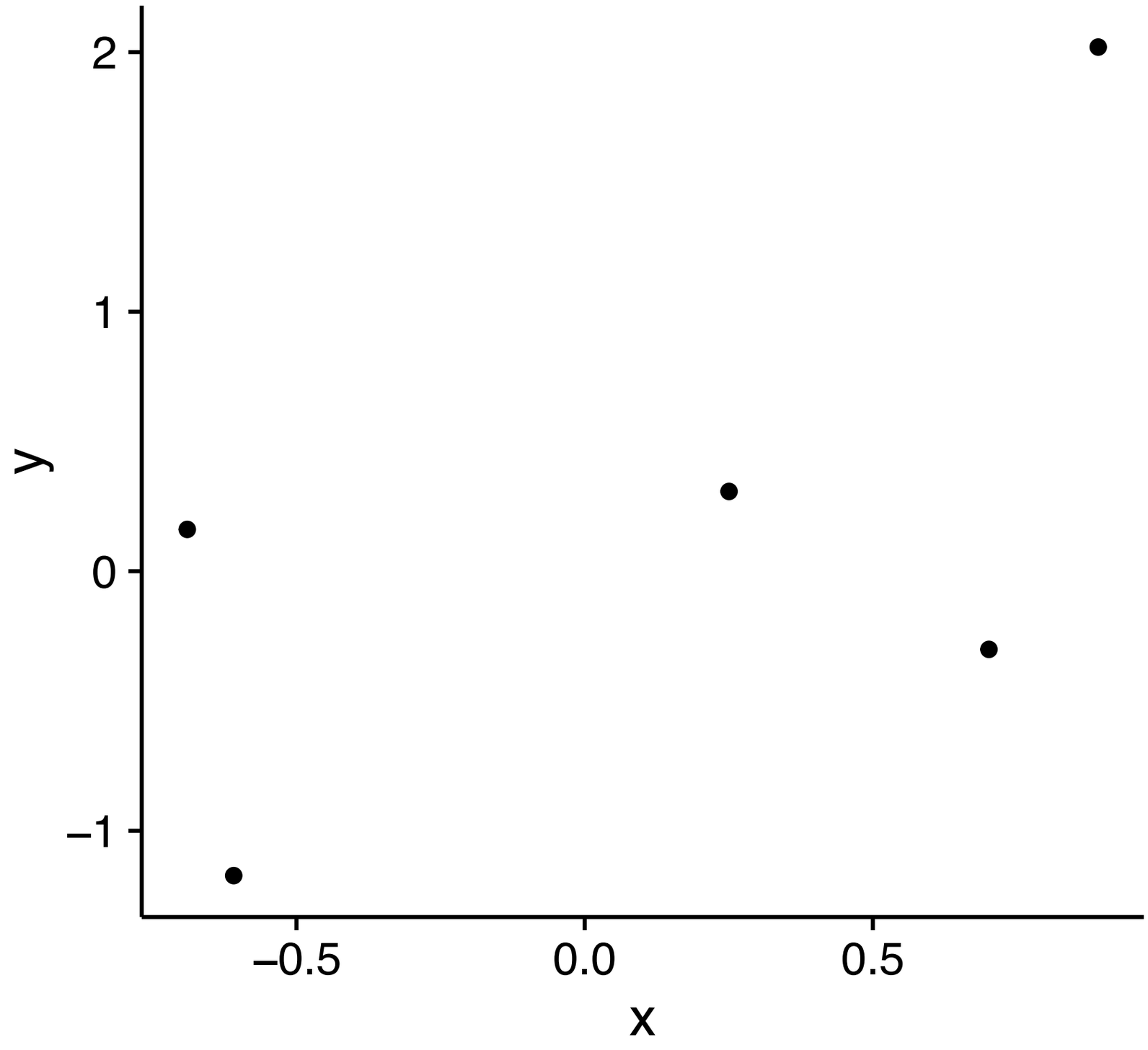
Our first model: linear regression

$$Y_i \sim N(\beta X_i + \alpha, \sigma)$$

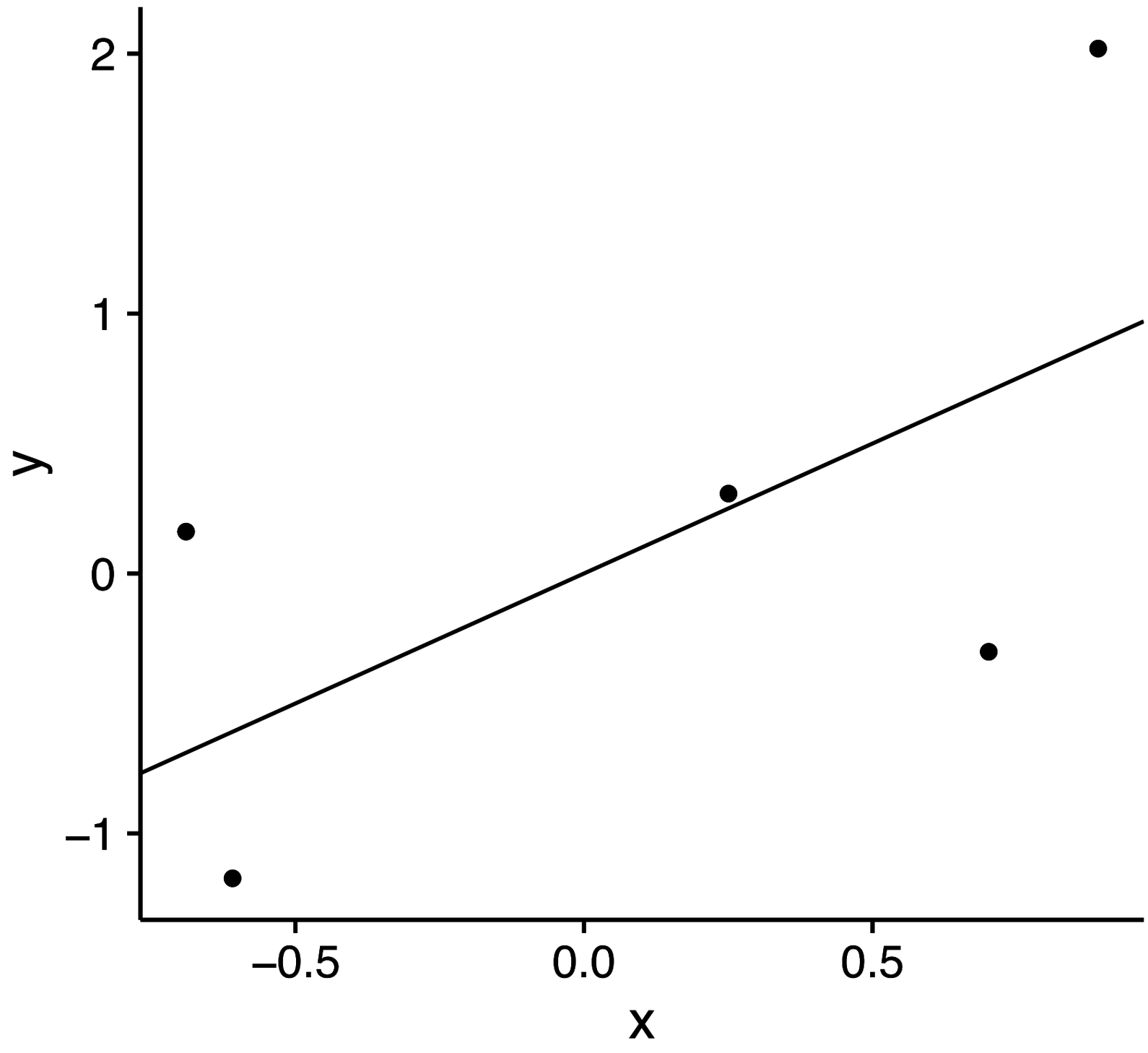
assumed sd



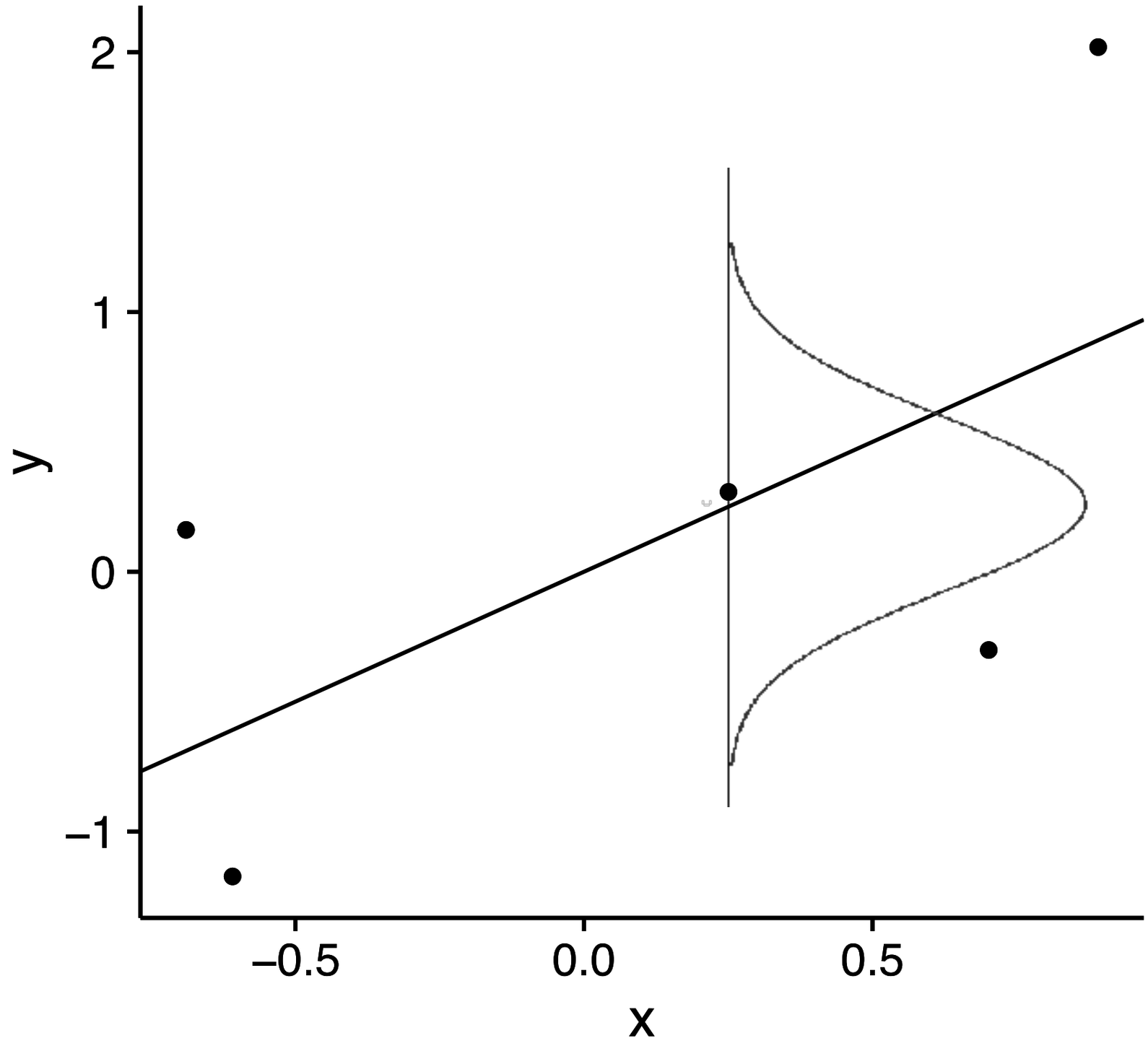
$$Y_i \sim N(\beta X_i + \alpha, \sigma)$$



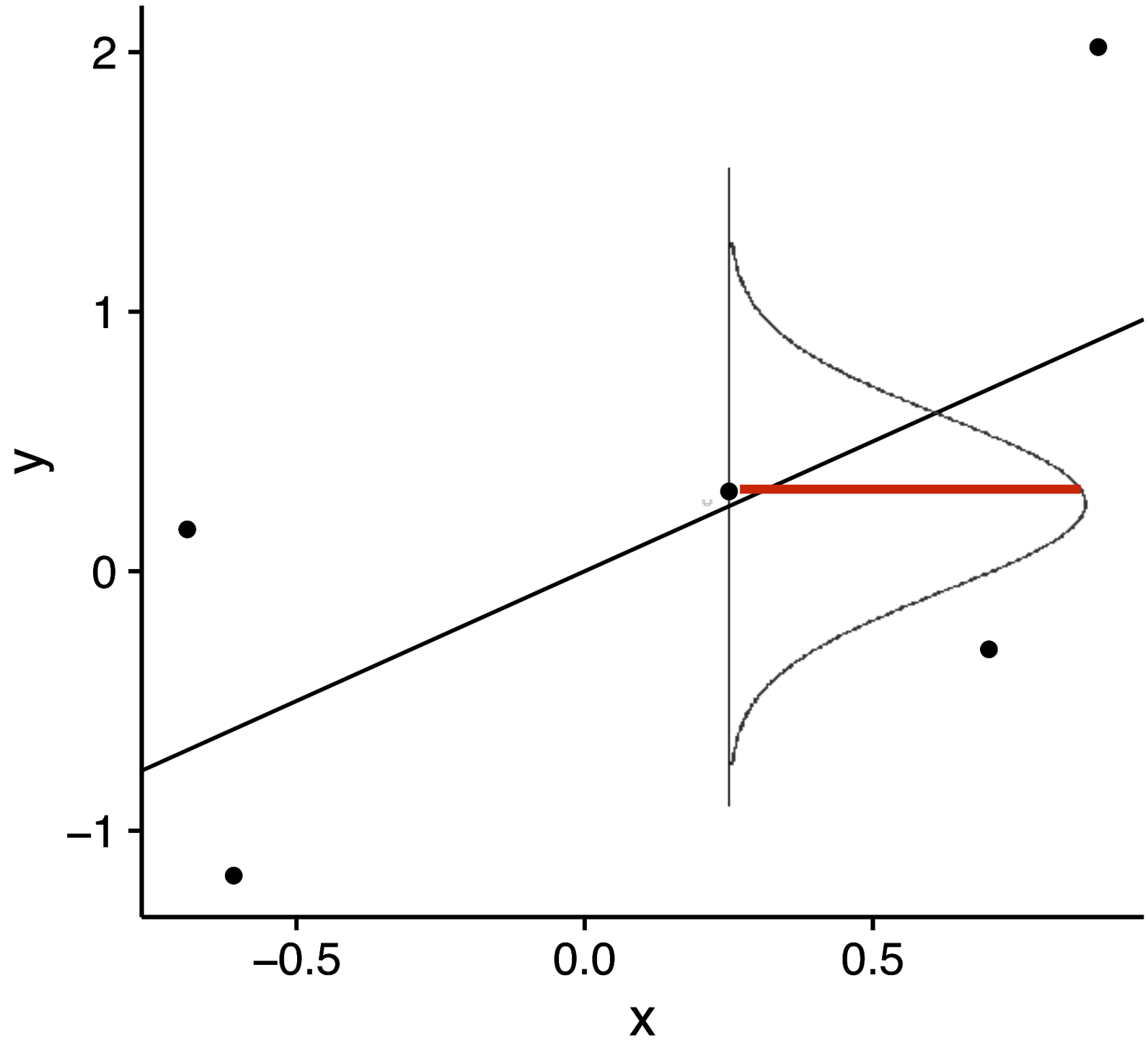
$$Y_i \sim N(\beta X_i + \alpha, \sigma)$$



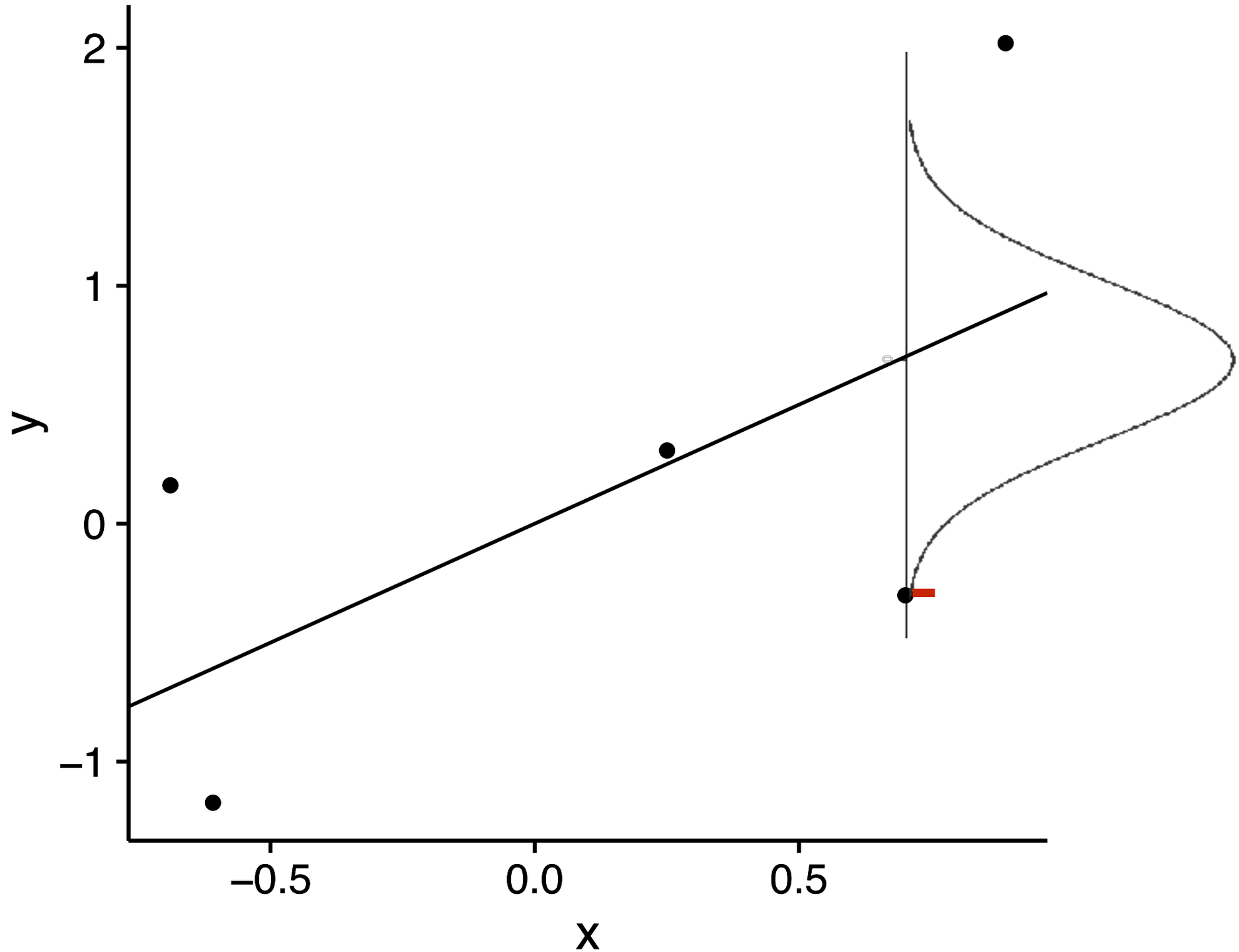
$$Y_i \sim N(\beta X_i + \alpha, \sigma)$$



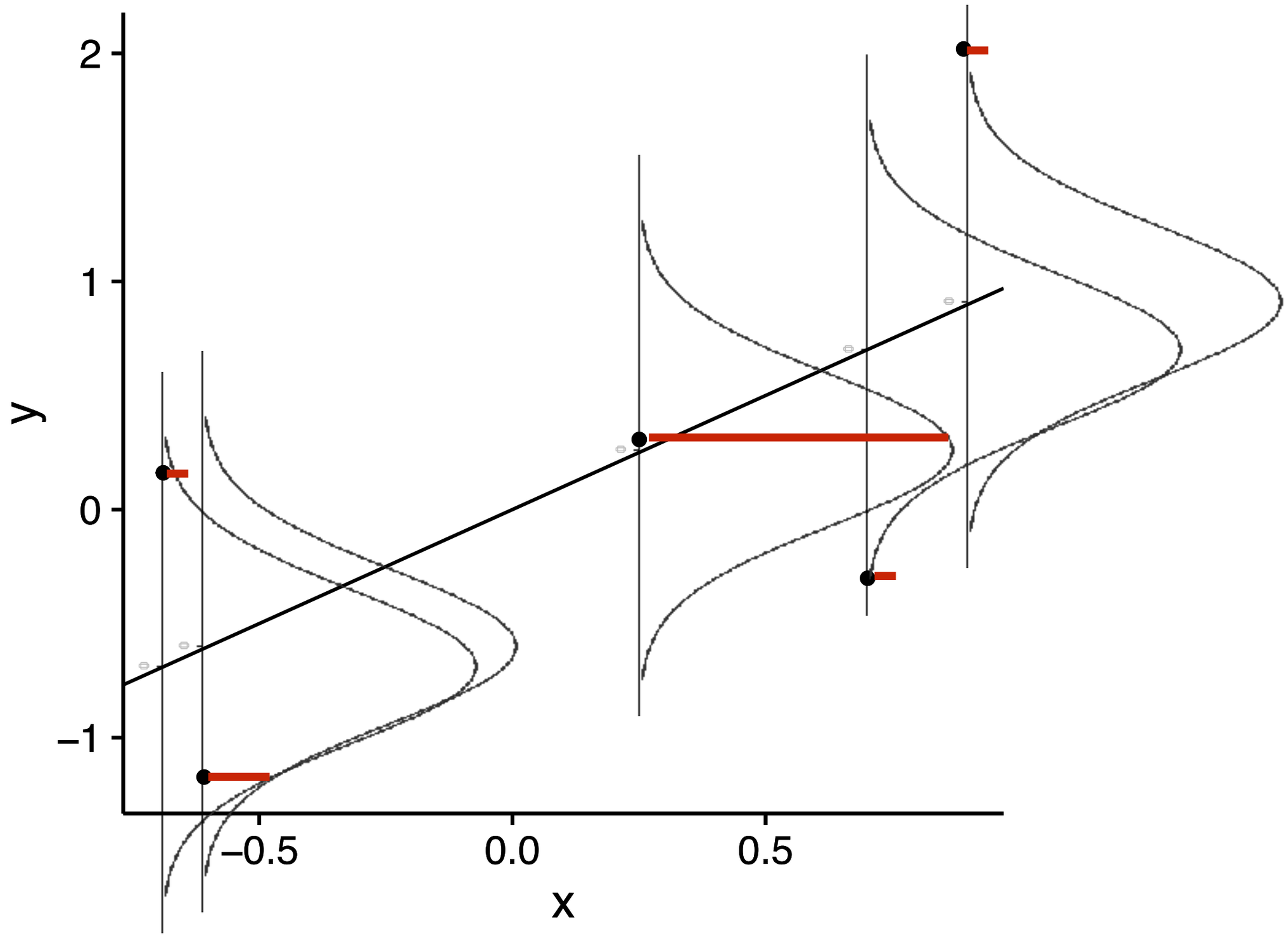
$$Y_i \sim N(\beta X_i + \alpha, \sigma)$$



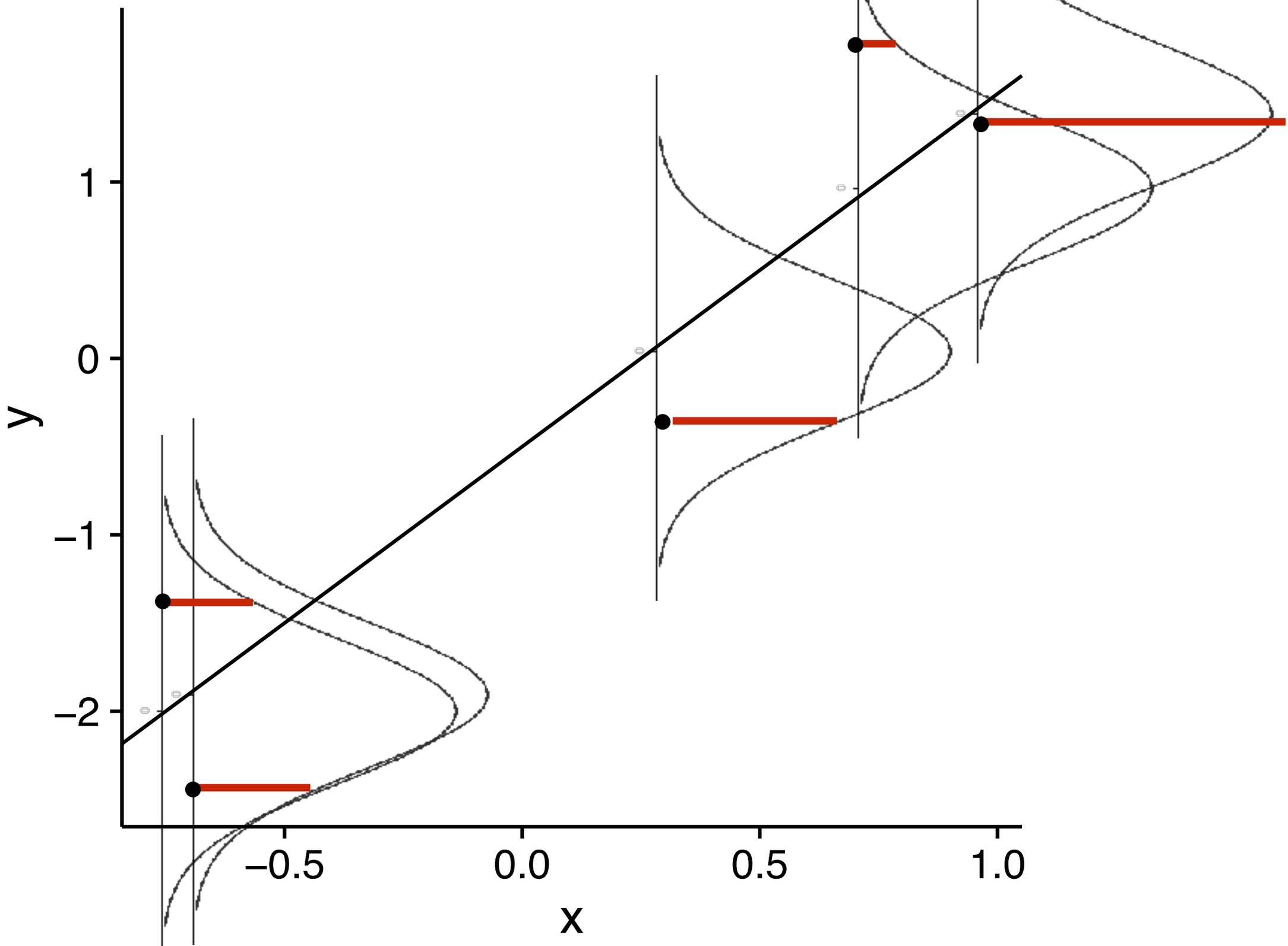
$$Y_i \sim N(\beta X_i + \alpha, \sigma)$$



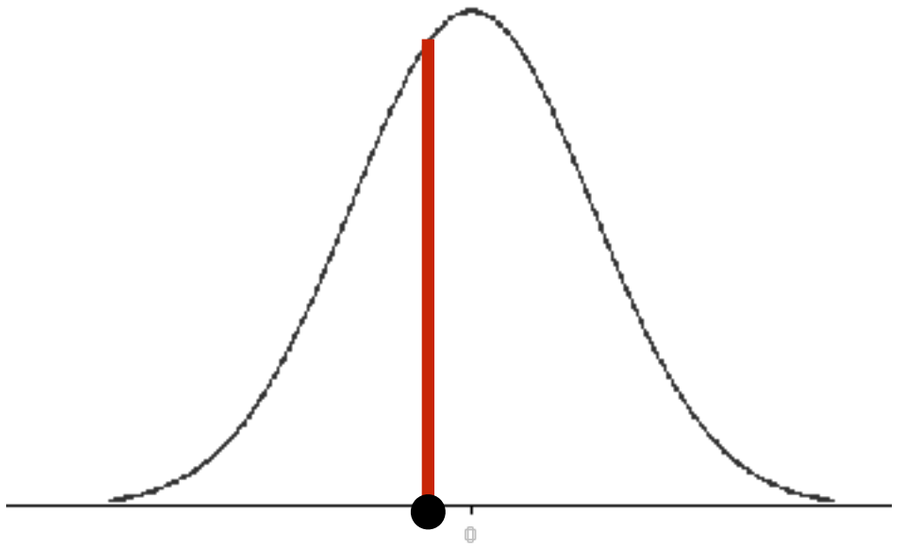
$$Y_i \sim N(\beta X_i + \alpha, \sigma)$$



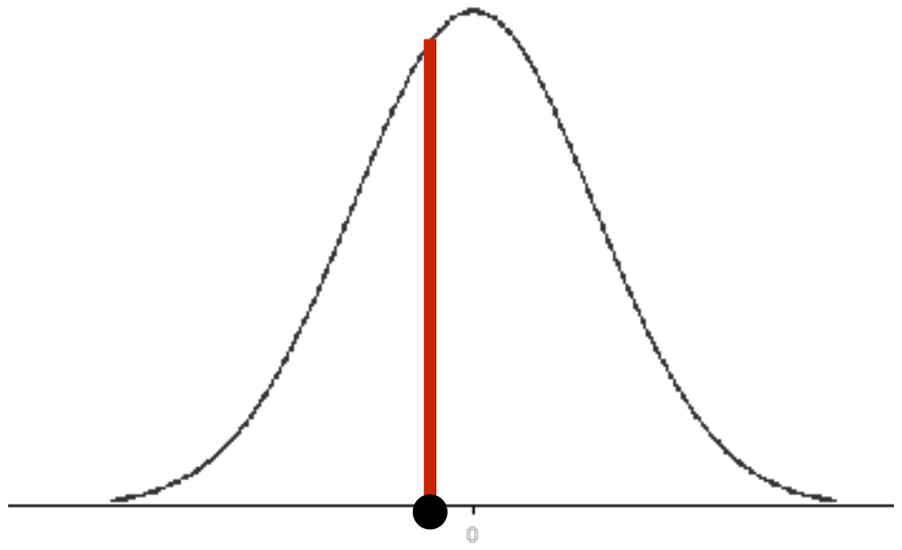
Try a different model



$$Y_i \sim N(\beta X_i + \alpha, \sigma)$$

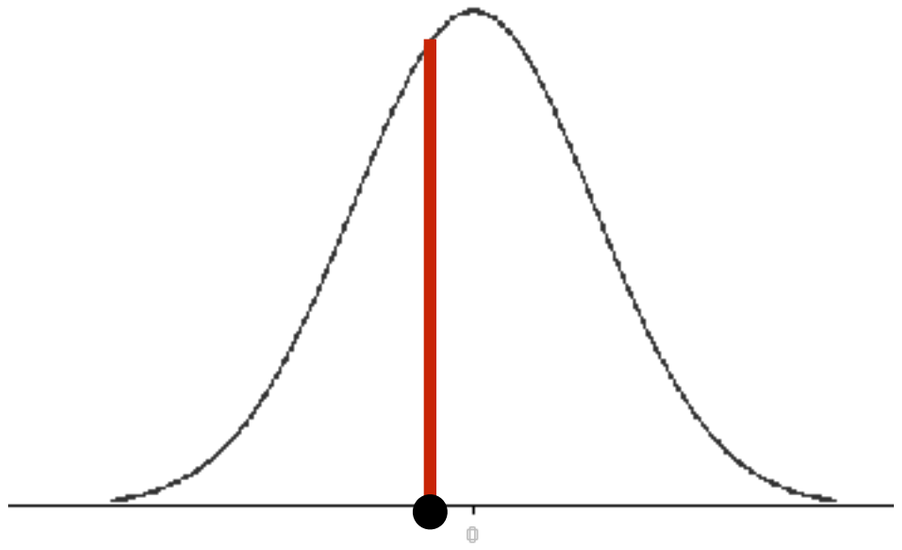


$$Y_i \sim N(\beta X_i + \alpha, \sigma)$$



$$\text{dnorm}(x=Y_i, \text{mean}=\beta X_i + \alpha, \text{sd}=\sigma)$$

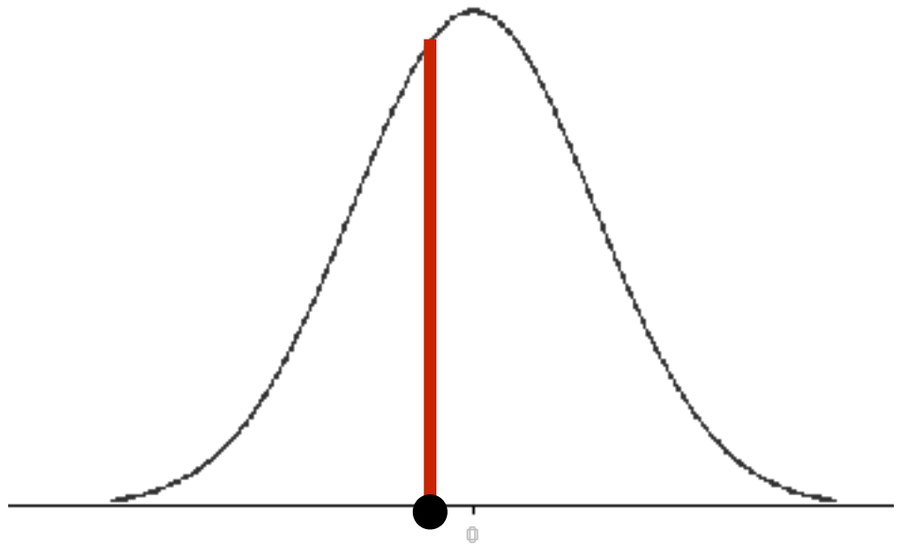
$$Y_i \sim N(\beta X_i + \alpha, \sigma)$$



$$\text{dnorm}(x=Y_i, \text{mean}=\beta X_i + \alpha, \text{sd}=\sigma)$$

$$P(Y \mid \beta, \alpha) = \text{product for all } i$$

$$Y_i \sim N(\beta X_i + \alpha, \sigma)$$



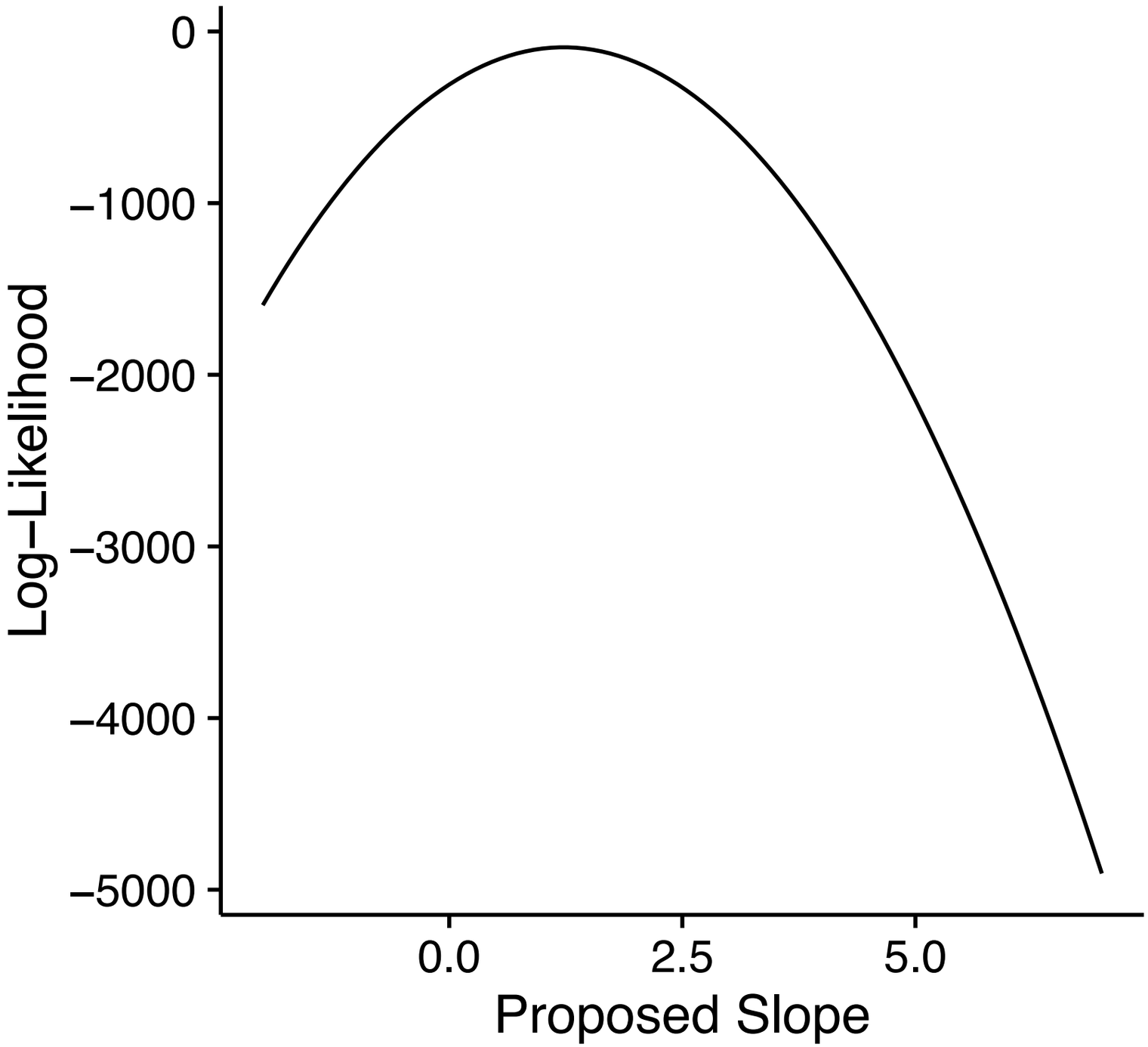
dnorm(x=Yi, mean=$\beta X_i + \alpha$, sd=σ, log=T)

$$P(Y \mid \beta, \alpha) = \text{sum for all i}$$

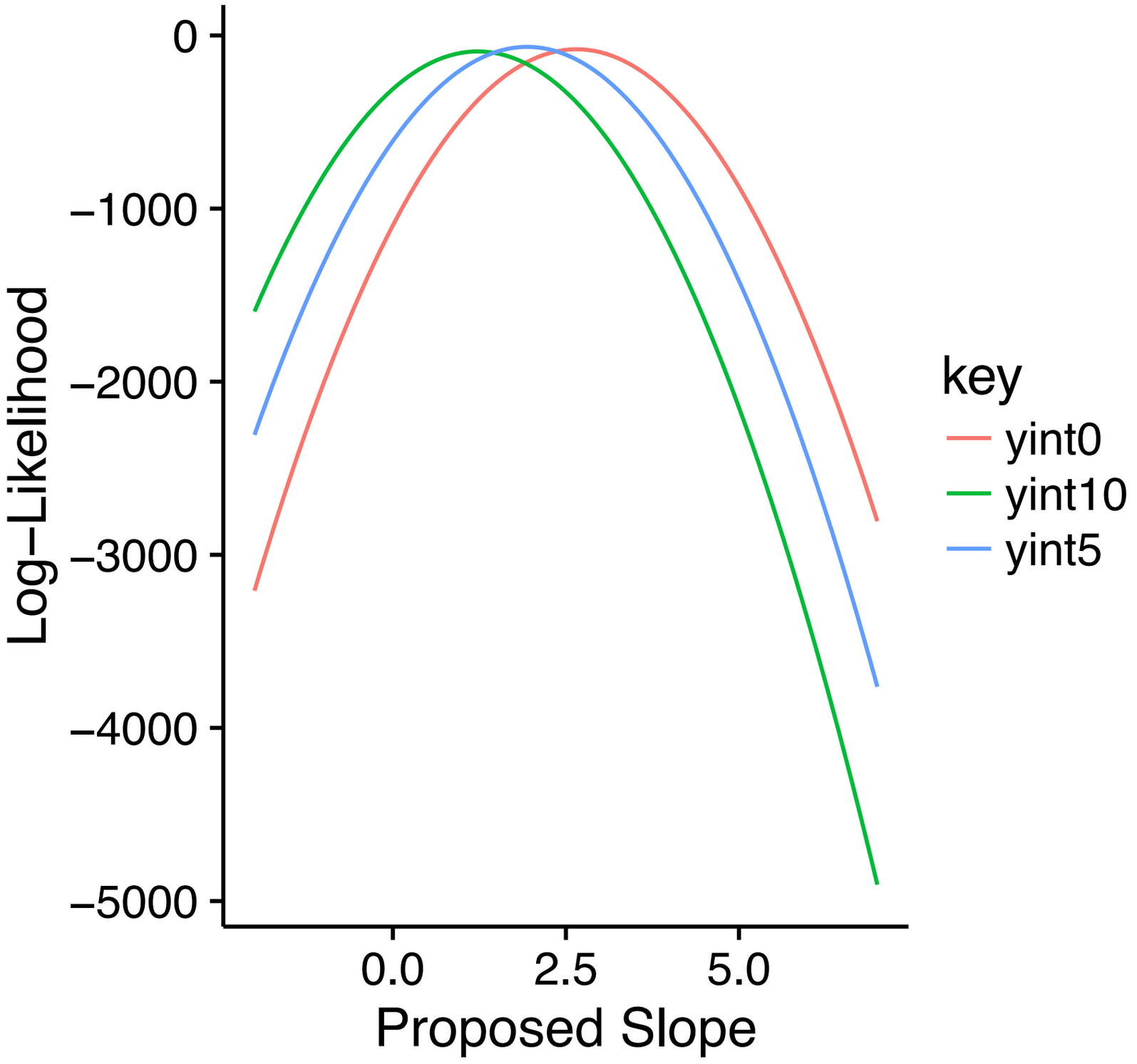
2nd Programming Exercise

```
function_name <- function(arg1, arg2, …){
    carry out function process
    output
}
```

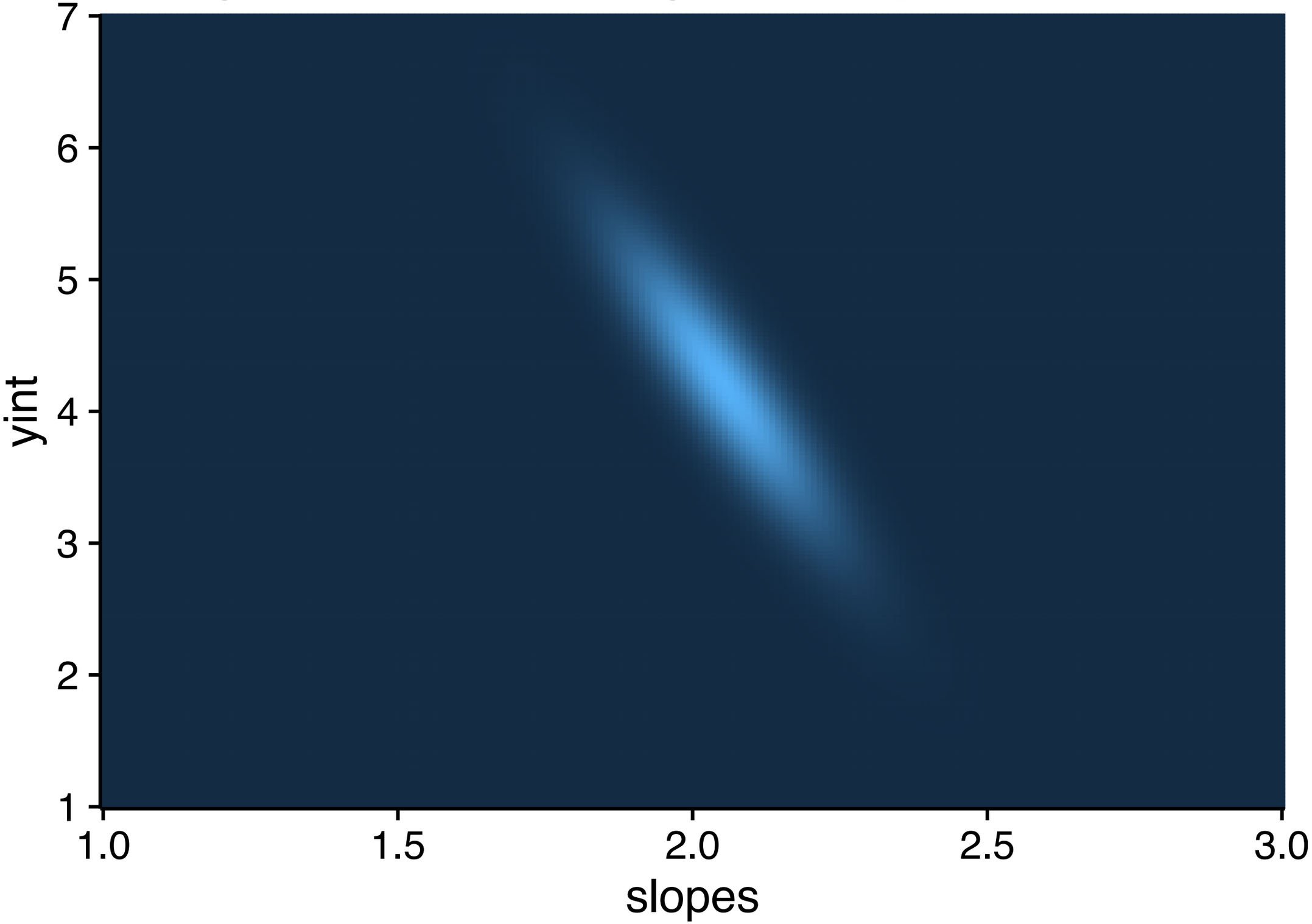
Once we have likelihood function we can systematically test parameters



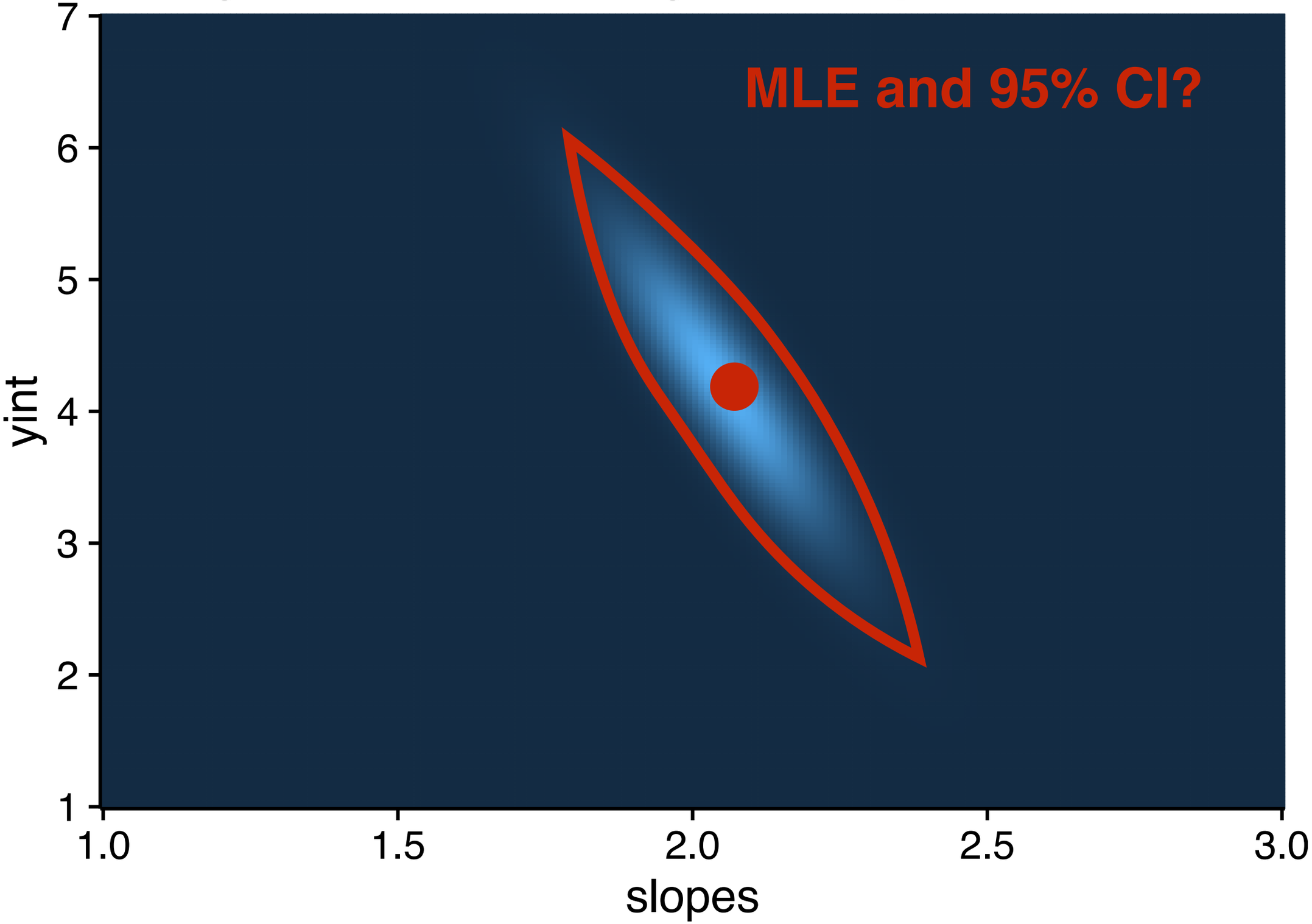
Once we have likelihood function we can systematically test parameters

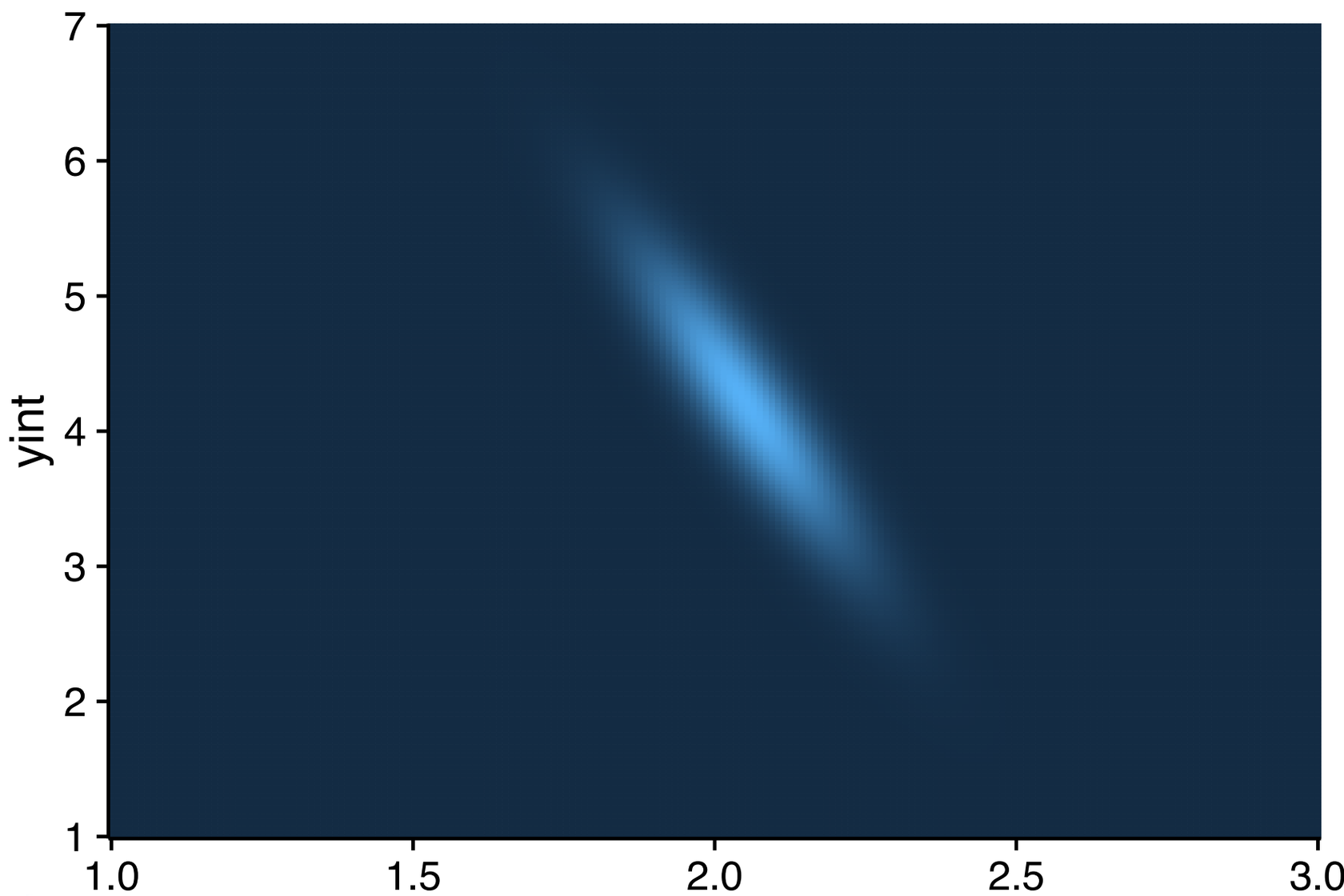


Once we have likelihood function we can systematically test parameters

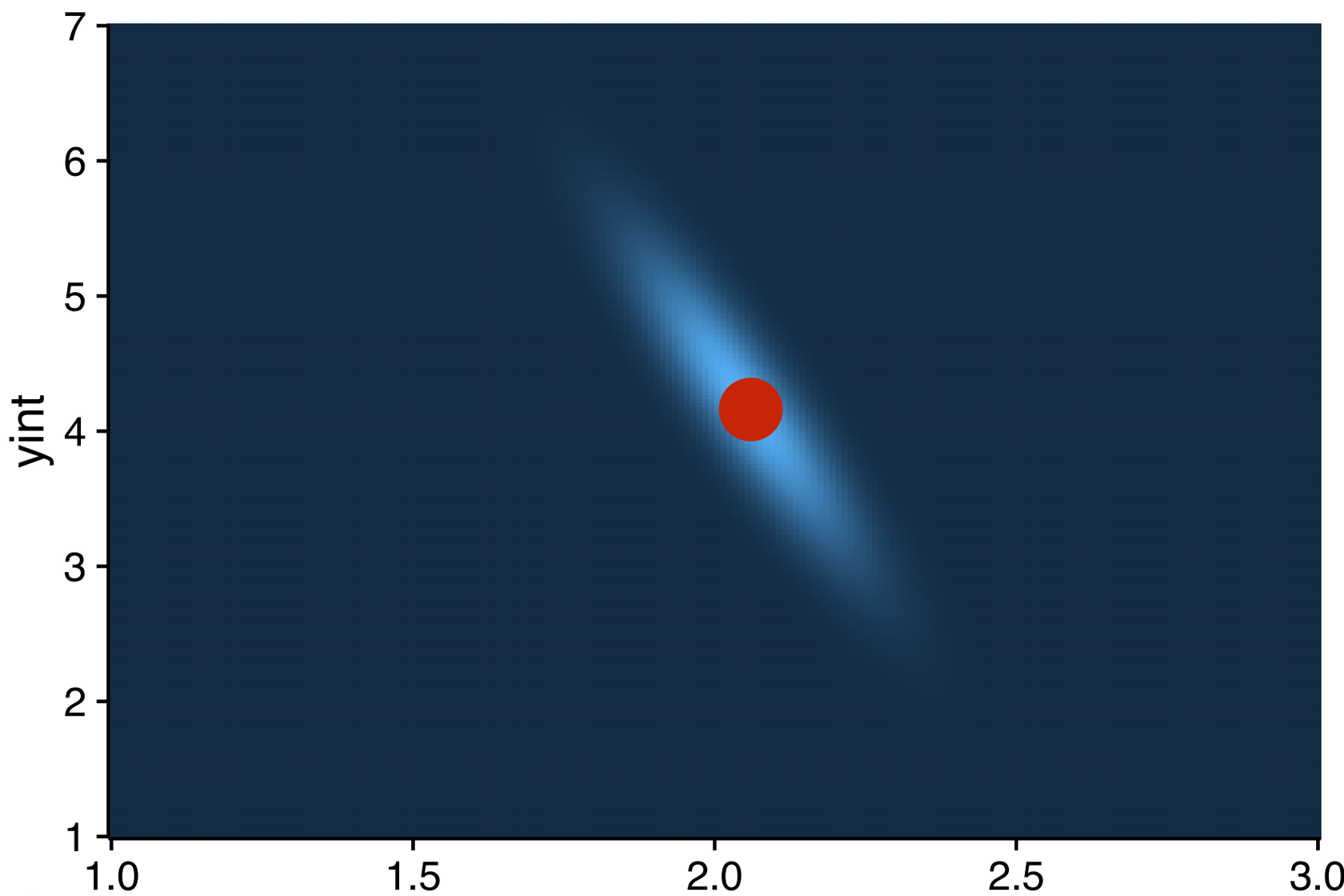


Once we have likelihood function we can systematically test parameters

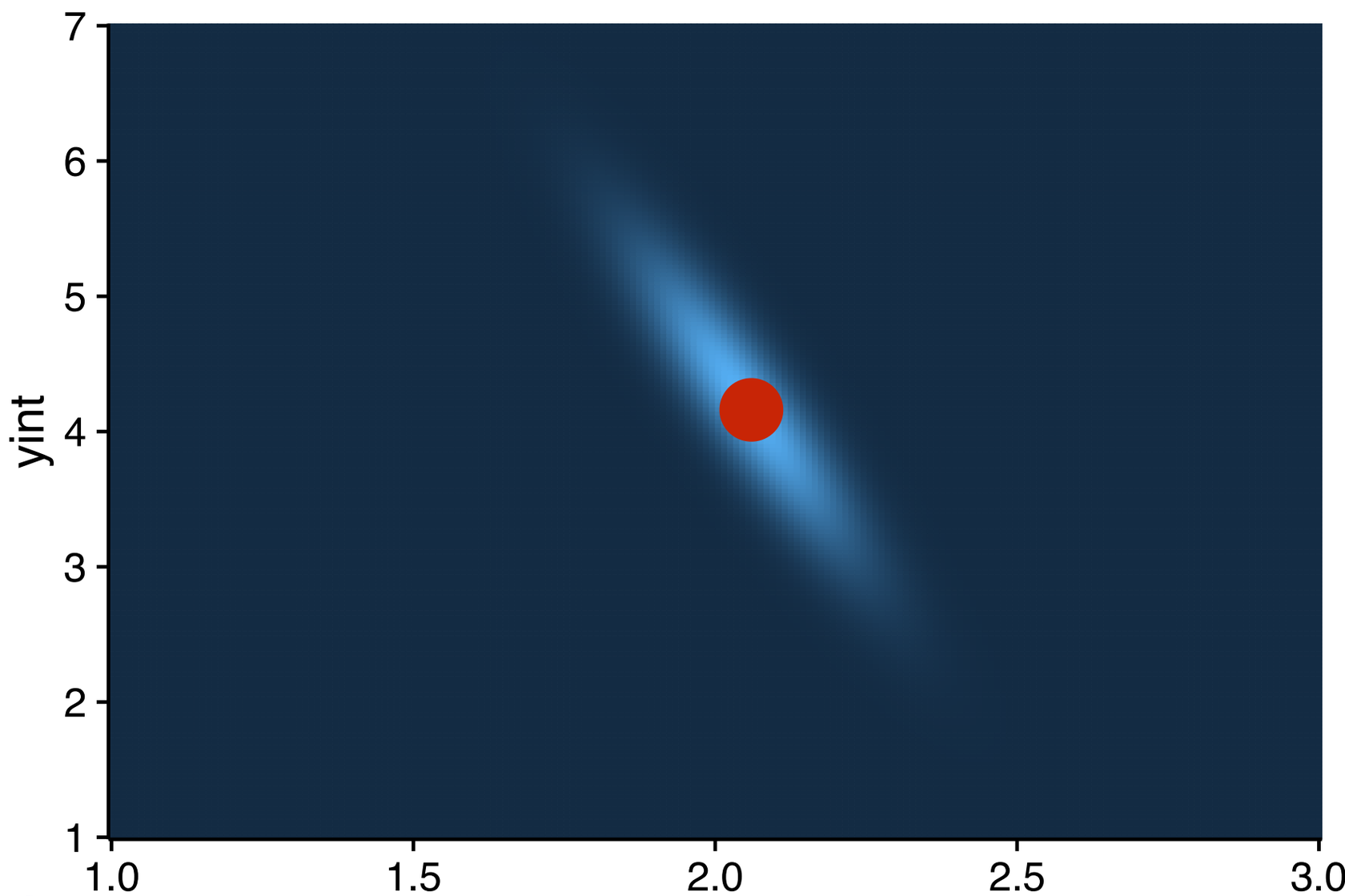





```
> lm(y~x, data =sim1)
```




```
> lm(y~x, data =sim1)
```




```
> lm(y~x, data =sim1)

Coefficients:
            Estimate Std. Error t value Pr(>|t|)
(Intercept)   4.2208     0.8688   4.858 4.09e-05 ***
x             2.0515     0.1400  14.651 1.17e-14 ***
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1
```

```
> lm(y~x, data =sim1)

Call:
lm(formula = y ~ x, data = sim1)

Residuals:
    Min      1Q  Median      3Q     Max
-4.1469 -1.5197  0.1331  1.4670  4.6516

Coefficients:
            Estimate Std. Error t value Pr(>|t|)
(Intercept)   4.2208     0.8688   4.858 4.09e-05 ***
x             2.0515     0.1400  14.651 1.17e-14 ***
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 2.203 on 28 degrees of freedom
Multiple R-squared:  0.8846,    Adjusted R-squared:  0.8805
F-statistic: 214.7 on 1 and 28 DF,  p-value: 1.173e-14
```

```
> lm(y~x, data =sim1)

Call:
lm(formula = y ~ x, data = sim1)

Residuals:
    Min      1Q  Median      3Q     Max
-4.1469 -1.5197  0.1331  1.4670  4.6516

Coefficients:
            Estimate Std. Error t value Pr(>|t|)
(Intercept)   4.2208     0.8688   4.858 4.09e-05 ***
x             2.0515     0.1400  14.651 1.17e-14 ***
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 2.203 on 28 degrees of freedom
Multiple R-squared:  0.8846,    Adjusted R-squared:  0.8805
F-statistic: 214.7 on 1 and 28 DF,  p-value: 1.173e-14
```

should be ~ normal

```
> lm(y~x, data =sim1)

Call:
lm(formula = y ~ x, data = sim1)

Residuals:
    Min      1Q  Median      3Q     Max
-4.1469 -1.5197  0.1331  1.4670  4.6516

Coefficients:
            Estimate Std. Error t value Pr(>|t|)
(Intercept)   4.2208     0.8688   4.858 4.09e-05 ***
x             2.0515     0.1400  14.651 1.17e-14 ***
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 2.203 on 28 degrees of freedom
Multiple R-squared:  0.8846,    Adjusted R-squared:  0.8805
F-statistic: 214.7 on 1 and 28 DF,  p-value: 1.173e-14
```

should be ~ normal

MLE

```
> lm(y~x, data =sim1)

Call:
lm(formula = y ~ x, data = sim1)

Residuals:
    Min      1Q  Median      3Q     Max
-4.1469 -1.5197  0.1331  1.4670  4.6516

Coefficients:
            Estimate Std. Error t value Pr(>|t|)
(Intercept)   4.2208     0.8688   4.858 4.09e-05 ***
x             2.0515     0.1400  14.651 1.17e-14 ***
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 2.203 on 28 degrees of freedom
Multiple R-squared:  0.8846,    Adjusted R-squared:  0.8805
F-statistic: 214.7 on 1 and 28 DF,  p-value: 1.173e-14
```

should be ~ normal

MLE

Standard error

```
> lm(y~x, data =sim1)

Call:
lm(formula = y ~ x, data = sim1)

Residuals:
    Min      1Q  Median      3Q     Max
-4.1469 -1.5197  0.1331  1.4670  4.6516

Coefficients:
            Estimate Std. Error t value Pr(>|t|)
(Intercept)   4.2208     0.8688   4.858 4.09e-05 ***
x             2.0515     0.1400  14.651 1.17e-14 ***
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 2.203 on 28 degrees of freedom
Multiple R-squared:  0.8846,    Adjusted R-squared:  0.8805
F-statistic: 214.7 on 1 and 28 DF,  p-value: 1.173e-14
```

should be ~ normal

MLE

Standard error

coefficient t-statistic

```
> lm(y~x, data =sim1)

Call:
lm(formula = y ~ x, data = sim1)

Residuals:
    Min      1Q  Median      3Q     Max
-4.1469 -1.5197  0.1331  1.4670  4.6516

Coefficients:
            Estimate Std. Error t value Pr(>|t|)
(Intercept)   4.2208     0.8688   4.858 4.09e-05 ***
x             2.0515     0.1400  14.651 1.17e-14 ***
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 2.203 on 28 degrees of freedom
Multiple R-squared:  0.8846,    Adjusted R-squared:  0.8805
F-statistic: 214.7 on 1 and 28 DF,  p-value: 1.173e-14
```

should be ~ normal

MLE

Standard error

coefficient t-statistic

p-val

```
> lm(y~x, data =sim1)

Call:
lm(formula = y ~ x, data = sim1)

Residuals:
    Min      1Q  Median      3Q     Max
-4.1469 -1.5197  0.1331  1.4670  4.6516

Coefficients:
            Estimate Std. Error t value Pr(>|t|)
(Intercept)   4.2208     0.8688   4.858 4.09e-05 ***
x             2.0515     0.1400  14.651 1.17e-14 ***
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 2.203 on 28 degrees of freedom
Multiple R-squared:  0.8846,    Adjusted R-squared:  0.8805
F-statistic: 214.7 on 1 and 28 DF,  p-value: 1.173e-14
```

should be ~ normal

MLE

Standard error

coefficient t-statistic

p-val

fit

3rd Programming Exercise