Limited Dependent Variables

Jos Elkink

February, 2008

Outline

Notation

- Independent variables X

Notation

- Independent variables X
- Dependent variable Y

Notation

- Independent variables X
- Dependent variable Y
- Y_i is a random variable (not directly observed)

Notation

- Independent variables X
- Dependent variable Y
- Y_i is a random variable (not directly observed)
- y_i is a realised value of Y_i (observed)

Notation

- Independent variables X
- Dependent variable Y
- Y_i is a random variable (not directly observed)
- y_i is a realised value of Y_i (observed)
- So dependent variable sometimes refers to a set of numbers in your dataset (y) and sometimes to a random variable at each i (Y_i).

Linear model

The linear model can be written as:

$$Y_i = X_i\beta + \varepsilon_i$$
$$\varepsilon_i \sim N(0, \sigma^2)$$

Linear model

The linear model can be written as:

$$Y_i = X_i\beta + \varepsilon_i$$
$$\varepsilon_i \sim N(0, \sigma^2)$$

Or, alternatively, as:

$$Y_i \sim N(\mu_i, \sigma^2)$$
$$\mu_i = X_i\beta$$

Linear model

Two components of the model:

$Y_i \sim N(\mu_i, \sigma^2)$	Stochastic
$\mu_i = X_i\beta$	Systematic

Linear model

Two components of the model:

$$
\left.\begin{array}{l} Y_i \sim N(\mu_i, \sigma^2) \\ \mu_i = X_i\beta \end{array}\right| \begin{array}{l} \text{Stochastic} \\ \text{Systematic} \end{array}
$$

Generalised version:

$$
\left.\begin{array}{l} Y_i \sim f(\theta_i, \alpha) \\ \theta_i = g(X_i, \beta) \end{array}\right| \begin{array}{l} \text{Stochastic} \\ \text{Systematic} \end{array}
$$

Model

$$\begin{array}{c|l} Y_i \sim f(\theta_i, \alpha) & \text{Stochastic} \\ \theta_i = g(X_i, \beta) & \text{Systematic} \end{array}$$

Stochastic component: varies over repeated (hypothetical) observations on the same unit.

Model

$$\begin{array}{r|l} Y_i \sim f(\theta_i, \alpha) & \text{Stochastic} \\ \theta_i = g(X_i, \beta) & \text{Systematic} \end{array}$$

Stochastic component: varies over repeated (hypothetical) observations on the same unit.

Systematic component: varies across units, but constant given X.

Model

$$\begin{array}{c|l} Y_i \sim f(\theta_i, \alpha) & \text{Stochastic} \\ \theta_i = g(X_i, \beta) & \text{Systematic} \end{array}$$

Two types of uncertainty:

Estimation uncertainty: lack of knowledge about α and β; can be reduced by increasing N.

Model

$$
\begin{array}{r|l}
Y_i \sim f(\theta_i, \alpha) & \text{Stochastic} \\
\theta_i = g(X_i, \beta) & \text{Systematic}
\end{array}
$$

Two types of uncertainty:

Estimation uncertainty: lack of knowledge about α and β; can be reduced by increasing N.
Fundamental uncertainty: represented by stochastic component and exists independent of researcher.

Outline

Probability: definition

Frequentist approach: a probability is the proportion of times a particular event will occur given a long run of repeated observations / experiments.

Probability: definition

Frequentist approach: a probability is the proportion of times a particular event will occur given a long run of repeated observations / experiments.

Subjective (Bayesian) approach: a probability is the formal quantification of the subjective belief about how likely a certain event will occur when an observation / experiment is performed.

Probability: properties

- For any event A, $0 \leq Pr(A) \leq 1$

Probability: properties

- For any event A, $0 \leq Pr(A) \leq 1$
- If S contains all possible events, $Pr(S) = 1$

Probability: properties

- For any event A, $0 \leq Pr(A) \leq 1$
- If S contains all possible events, $Pr(S) = 1$
- If B is "not A", $Pr(B) = 1 - Pr(A)$

Probability: properties

- For any event A, $0 \leq Pr(A) \leq 1$
- If S contains all possible events, $Pr(S) = 1$
- If B is "not A", $Pr(B) = 1 - Pr(A)$
- If A and C are independent events, $Pr(AC) = Pr(A)Pr(C)$

Probability: example

Imagine a coin flip. The probabilities are:

- $Pr(head) = \frac{1}{2}$

Probability: example

Imagine a coin flip. The probabilities are:

- $Pr(head) = \frac{1}{2}$
- $Pr(tail) = \frac{1}{2}$

Probability: example

Imagine a coin flip. The probabilities are:

- $Pr(head) = \frac{1}{2}$
- $Pr(tail) = \frac{1}{2}$
- Sum of all possible events: $Pr(head) + Pr(tail) = \frac{1}{2} + \frac{1}{2} = 1$

Probability: example

Imagine a coin flip. The probabilities are:

- $Pr(head) = \frac{1}{2}$
- $Pr(tail) = \frac{1}{2}$
- Sum of all possible events: $Pr(head) + Pr(tail) = \frac{1}{2} + \frac{1}{2} = 1$
- Probability of head and then tail: $Pr(head, thentail) = Pr(head)Pr(tail) = \frac{1}{4}$

Outline

Probability distributions

For nominal or ordinal variables, the probability distribution looks like a relative frequency graph (histogram); "probability mass function" (p.m.f.).

Probability distributions

For nominal or ordinal variables, the probability distribution looks like a relative frequency graph (histogram); "probability mass function" (p.m.f.).

For interval and ratio variables, the probability distribution is a continuous function, the "probability density function" (p.d.f.).

Probability distributions: properties

Probability mass function: all probabilities together sum to 1.

Probability distributions: properties

Probability mass function: all probabilities together sum to 1.

Probability density function: the total area under the graph is 1.

Probability distributions: properties

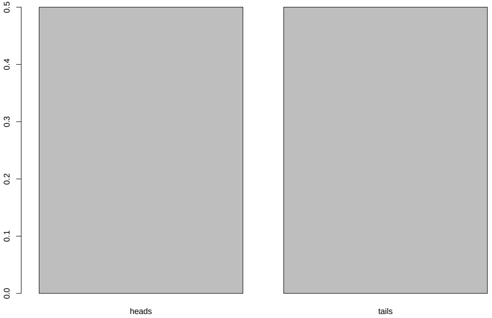

Figure: Example of probability distribution

Probability distributions: properties

For a continuous variables, you can only speak of the probability that the observation will be within a certain range of values, not of the probability of a specific value.

Probability distributions: properties

For a continuous variables, you can only speak of the probability that the observation will be within a certain range of values, not of the probability of a specific value.

The probability is represented by the area under the graph over that range.

Probability distributions: properties

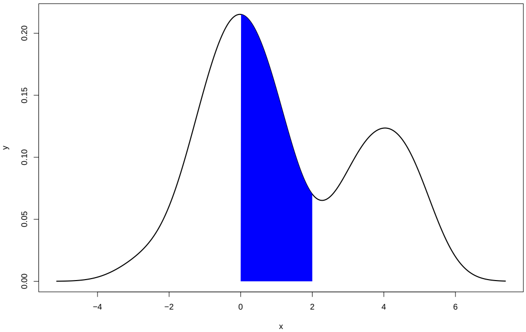

Figure: Example of probability distribution, $Pr(0 \leq X \leq 2)$

Probability distributions: properties

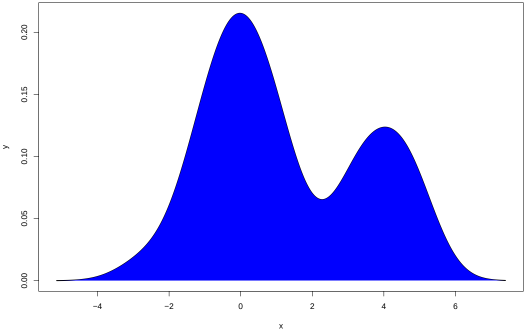

Figure: Example of probability distribution

Normal distribution

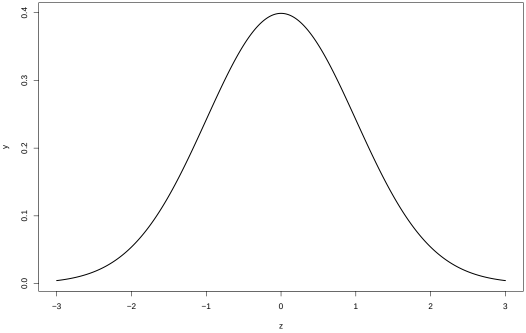

Figure: Normal distribution

Normal distribution

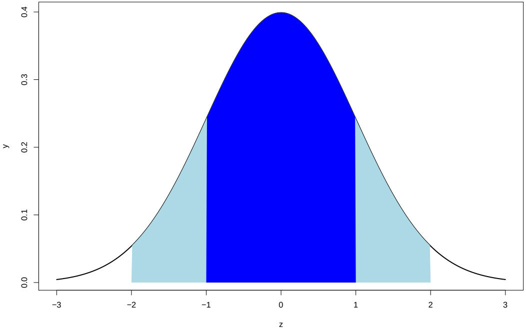

Figure: Normal distribution, with 68% and 95% zones.

Normal distribution

$$f(x|\mu,\sigma) = \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{1}{2\sigma^2}(x-\mu)^2}$$

Normal distribution

$$f(x|\mu,\sigma) = \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{1}{2\sigma^2}(x-\mu)^2}$$

Thus entirely determined by the mean μ and the standard deviation σ.

Normal distribution: notation

$$y_i \sim N(\mu, \sigma^2)$$

Normal distribution: notation

$$y_i \sim N(\mu, \sigma^2)$$

Y is normally distributed with mean μ and variance σ^2.

Normal distribution: z-scores

If a variable follows a normal distribution, we can standardize it, converting the values of the variable to z-scores.

Normal distribution: z-scores

If a variable follows a normal distribution, we can standardize it, converting the values of the variable to z-scores.

Definition: the z-score of a value y_i for variable Y is the number of standard deviations between y_i and the mean of Y, μ.

Normal distribution: z-scores

If a variable follows a normal distribution, we can standardize it, converting the values of the variable to z-scores.

Definition: the z-score of a value y_i for variable Y is the number of standard deviations between y_i and the mean of Y, μ.

$$z_i = \frac{y_i - \mu}{\sigma}$$

Normal distribution: z-scores

If a variable follows a normal distribution, we can standardize it, converting the values of the variable to z-scores.

Definition: the z-score of a value y_i for variable Y is the number of standard deviations between y_i and the mean of Y, μ.

$$z_i = \frac{y_i - \mu}{\sigma}$$

$$z_i \sim N(0, 1)$$

Normal distribution

Say, the average height is $\mu = 1.70$ meters and the standard deviation is $\sigma = .2$ meters. What percentage of people is shorter than 1.6?

Normal distribution

Say, the average height is $\mu = 1.70$ meters and the standard deviation is $\sigma = .2$ meters. What percentage of people is shorter than 1.6?

```
> pnorm(1.6, 1.70, .2)
[1] 0.3085375
```

Normal distribution

Say, the average height is $\mu = 1.70$ meters and the standard deviation is $\sigma = .2$ meters. What percentage of people is shorter than 1.6?

```
> pnorm(1.6, 1.70, .2)
[1] 0.3085375
```

$$z = \frac{x - \mu}{\sigma} = \frac{1.6 - 1.7}{.2} = -\frac{1}{2}$$

Normal distribution

Say, the average height is $\mu = 1.70$ meters and the standard deviation is $\sigma = .2$ meters. What percentage of people is shorter than 1.6?

```
> pnorm(1.6, 1.70, .2)
[1] 0.3085375
```

$$z = \frac{x - \mu}{\sigma} = \frac{1.6 - 1.7}{.2} = -\frac{1}{2}$$

```
> pnorm(-.5)
[1] 0.3085375
```

Normal distribution

Say, the average height is $\mu = 1.70$ meters and the standard deviation is $\sigma = .2$ meters. What percentage of people is taller than 1.6?

Normal distribution

Say, the average height is $\mu = 1.70$ meters and the standard deviation is $\sigma = .2$ meters. What percentage of people is taller than 1.6?

```
> 1 - pnorm(1.6, 1.70, .2)
[1] 0.6914625
```

Normal distribution

Say, the average speed on a motorway is $\mu = 110$ meters and the standard deviation is $\sigma = 7$ meters. What percentage of people drives faster than the maximum of 120?

Normal distribution

Say, the average speed on a motorway is $\mu = 110$ meters and the standard deviation is $\sigma = 7$ meters. What percentage of people drives faster than the maximum of 120?

```
> 1 - pnorm(120, 110, 7)
[1] 0.07656373
```

d, p, q, and r functions

- You know x and want to know the density at that point: d

d, p, q, and r functions

- You know x and want to know the density at that point: d
- You know x and want to know the area up to that point: p

d, p, q, and r functions

- You know x and want to know the density at that point: d
- You know x and want to know the area up to that point: p
- You know x and want to know the area beyond that point: 1-p

d, p, q, and r functions

- You know x and want to know the density at that point: d
- You know x and want to know the area up to that point: p
- You know x and want to know the area beyond that point: 1-p
- You know the area and want to know the x value: q

d, p, q, and r functions

- You know x and want to know the density at that point: d
- You know x and want to know the area up to that point: p
- You know x and want to know the area beyond that point: 1-p
- You know the area and want to know the x value: q
- You want random numbers drawn from that distribution: r

d, p, q, and r functions

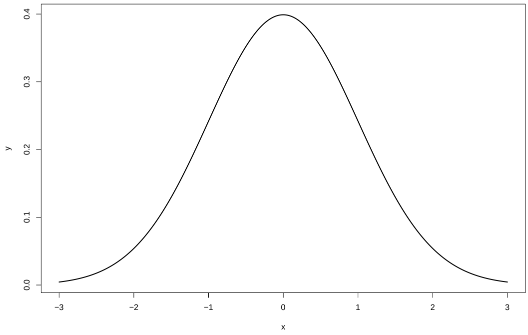

Figure: Demonstration of dnorm()

d, p, q, and r functions

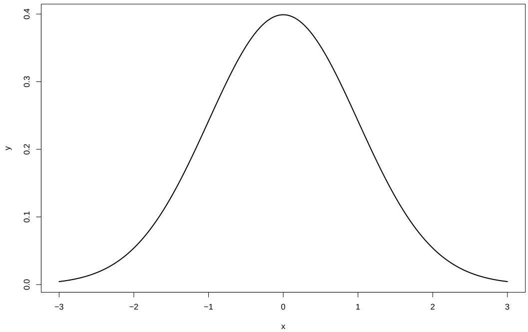

Figure: Demonstration of dnorm()

```
x <- seq(-3,3,.01)
y <- dnorm(x)
```

d, p, q, and r functions

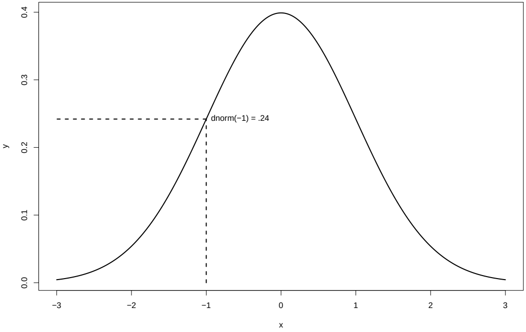

Figure: Demonstration of dnorm()

d, p, q, and r functions

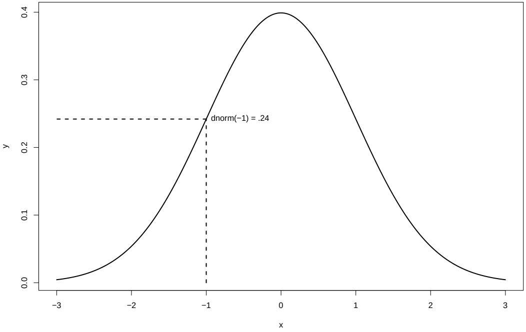

Figure: Demonstration of dnorm()

```
> dnorm(-1)
[1] 0.2419707
```

d, p, q, and r functions

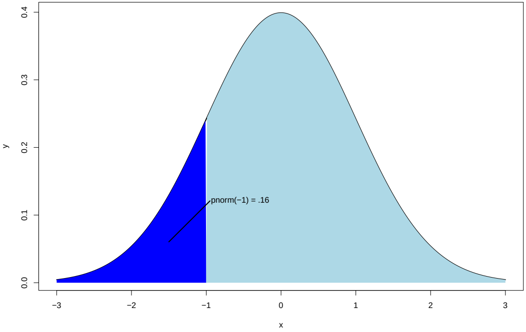

Figure: Demonstration of pnorm()

d, p, q, and r functions

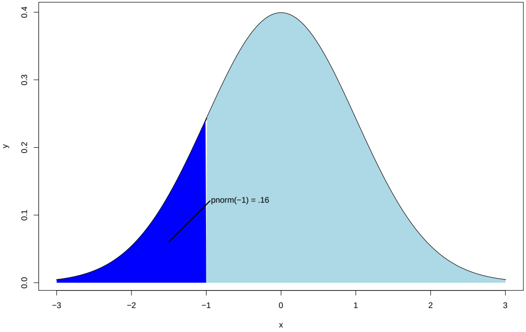

Figure: Demonstration of pnorm()

```
> pnorm(-1)
[1] 0.1586553
```

d, p, q, and r functions

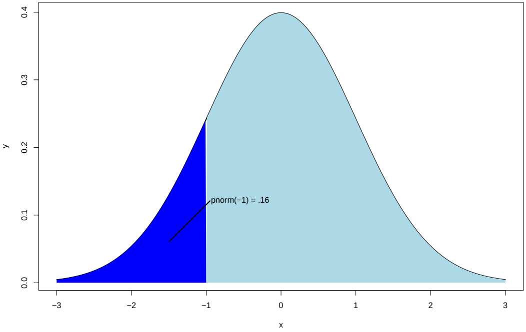

Figure: Demonstration of pnorm()

```
> 1 - pnorm(-1)
[1] 0.8413447
```

d, p, q, and r functions

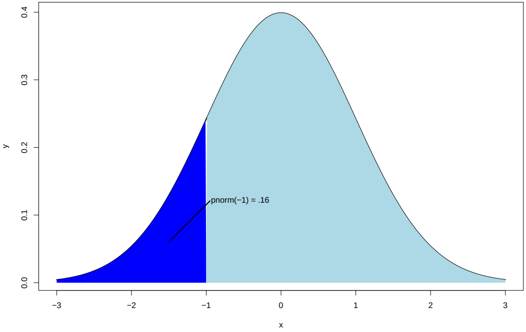

Figure: Demonstration of pnorm()

```
> qnorm(.1586553)
[1] -0.9999998
```

d, p, q, and r functions

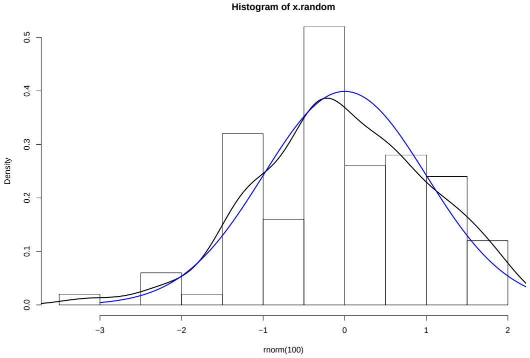

Figure: Demonstration of rnorm()

d, p, q, and r functions

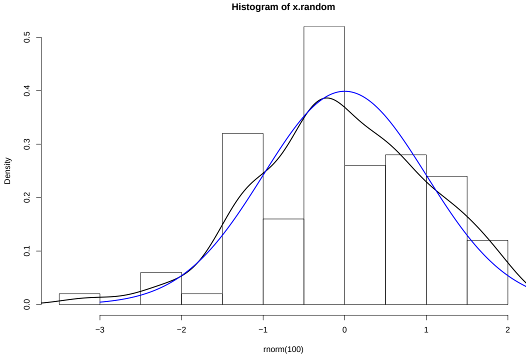

Figure: Demonstration of rnorm()

```
> x.random <- rnorm(100)
```

d, p, q, and r functions

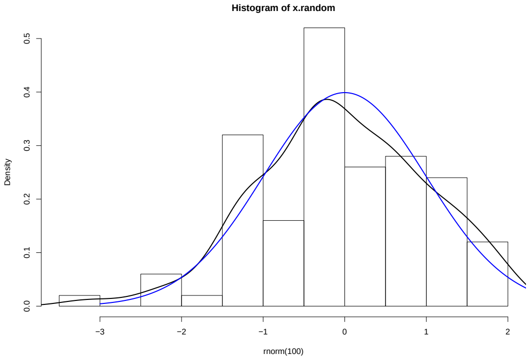

Figure: Demonstration of rnorm()

```
> hist(x.random, freq=FALSE)
```

d, p, q, and r functions

These functions work on many distributions:

- rnorm() - random from normal distribution
- pchisq() - get area under χ^2 distribution
- 1-pf() - get area under F distribution
- rbinom() - draw randomly from binomial distribution

etc.

Bernoulli distribution

Twenty throws ("Bernoulli trials") with a coin:

```
> x <- rbinom(20,1,.5)
> factor(x, labels=c("head","tails"))
 [1] head  head  head  tails head  tails tails head
 [9] tails head  head  head  tails tails tails head
[17] tails tails tails head
```