

Economics 103 – Statistics for Economists

Francis J. DiTraglia

University of Pennsylvania

Lecture # 12

Continuous RVs II: The Normal RV



Figure: Standard Normal RV (PDF)

Standard Normal Random Variable: $N(0, 1)$

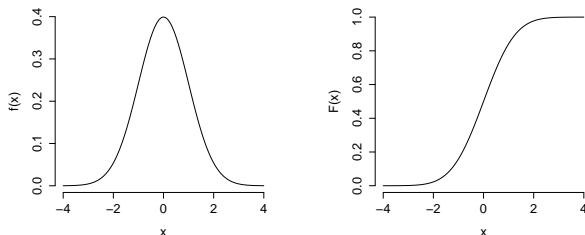
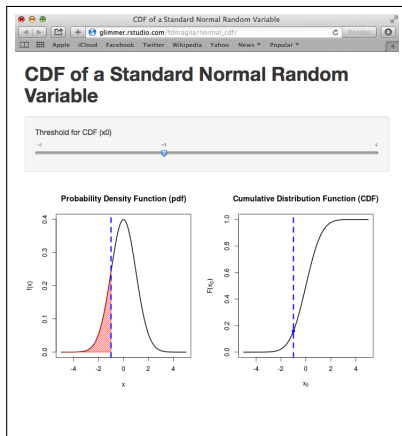


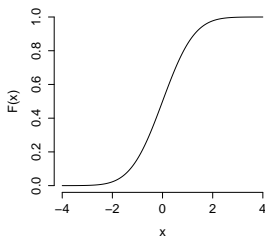
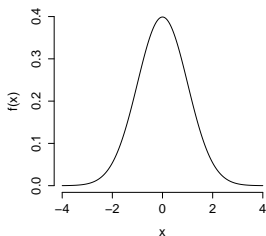
Figure: Standard Normal PDF (left) and CDF (Right)

- ▶ Notation: $X \sim N(0, 1)$
- ▶ Symmetric, Bell-shaped, $E[X] = 0$, $Var[X] = 1$
- ▶ Support Set = $(-\infty, \infty)$

https://fditraglia.shinyapps.io/normal_cdf/



Standard Normal Random Variable: $N(0, 1)$



- ▶ There is no closed-form expression for the $N(0, 1)$ CDF.
- ▶ For Econ 103, don't need to know formula for $N(0, 1)$ PDF.
- ▶ You *do need* to know the R commands. . .

R Commands for the Standard Normal RV

`dnorm` – Standard Normal PDF

- ▶ Mnemonic: `d` = density, `norm` = normal
- ▶ Example: `dnorm(0)` gives height of $N(0, 1)$ PDF at zero.

R Commands for the Standard Normal RV

`dnorm` – Standard Normal PDF

- ▶ Mnemonic: `d` = density, `norm` = normal
- ▶ Example: `dnorm(0)` gives height of $N(0, 1)$ PDF at zero.

`pnorm` – Standard Normal CDF

- ▶ Mnemonic: `p` = probability, `norm` = normal
- ▶ Example: `pnorm(1)` = $P(X \leq 1)$ if $X \sim N(0, 1)$.

R Commands for the Standard Normal RV

`dnorm` – Standard Normal PDF

- ▶ Mnemonic: `d` = density, `norm` = normal
- ▶ Example: `dnorm(0)` gives height of $N(0, 1)$ PDF at zero.

`pnorm` – Standard Normal CDF

- ▶ Mnemonic: `p` = probability, `norm` = normal
- ▶ Example: `pnorm(1)` = $P(X \leq 1)$ if $X \sim N(0, 1)$.

`rnorm` – Simulate Standard Normal Draws

- ▶ Mnemonic: `r` = random, `norm` = normal.
- ▶ Example: `rnorm(10)` makes ten iid $N(0, 1)$ draws.

$\Phi(x_0)$ Denotes the $N(0, 1)$ CDF

You will sometimes encounter the notation $\Phi(x_0)$. It means the same thing as `pnorm(x0)` but it's not an R command.

The $N(\mu, \sigma^2)$ Random Variable

Idea

Take a linear function of the $N(0, 1)$ RV.

Formal Definition

$N(\mu, \sigma^2) \equiv \mu + \sigma X$ where $X \sim N(0, 1)$ and μ, σ are constants.

Properties of $N(\mu, \sigma^2)$ RV

- ▶ Parameters: Expected Value = μ , Variance = σ^2
- ▶ Symmetric and bell-shaped.
- ▶ Support Set = $(-\infty, \infty)$
- ▶ $N(0, 1)$ is the special case where $\mu = 0$ and $\sigma^2 = 1$.

Expected Value: μ shifts PDF

all of these have $\sigma = 1$

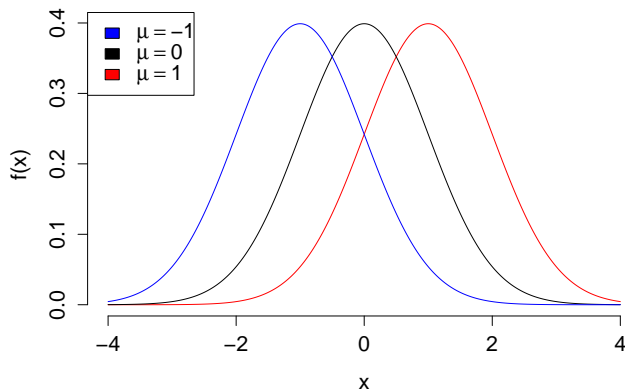


Figure: Blue $\mu = -1$, Black $\mu = 0$, Red $\mu = 1$

Standard Deviation: σ scales PDF

all of these have $\mu = 0$

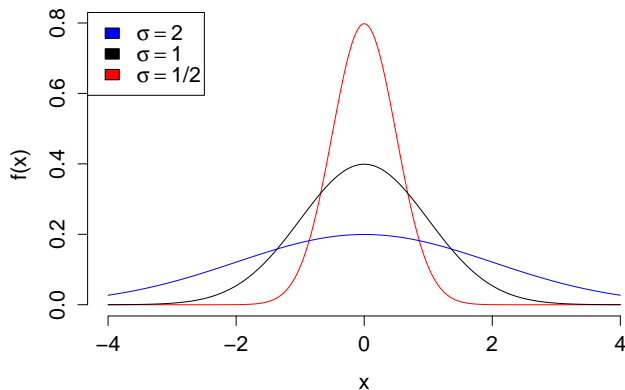


Figure: Blue $\sigma^2 = 4$, Black $\sigma^2 = 1$, Red $\sigma^2 = 1/4$

Linear Function of Normal RV is a Normal RV

Suppose that $X \sim N(\mu, \sigma^2)$. Then if a and b constants,

$$a + bX \sim N(a + b\mu, b^2\sigma^2)$$

Important

- ▶ For *any* RV X , $E[a + bX] = a + bE[X]$ and $Var(a + bX) = b^2 Var(X)$.
- ▶ Key point: linear transformation of normal is still normal!
- ▶ Linear transformation of Binomial is *not* Binomial!

Example



Suppose $X \sim N(\mu, \sigma^2)$ and let $Z = (X - \mu)/\sigma$. What is the distribution of Z ?

- (a) $N(\mu, \sigma^2)$
- (b) $N(\mu, \sigma)$
- (c) $N(0, \sigma^2)$
- (d) $N(0, \sigma)$
- (e) $N(0, 1)$

Linear Combinations of *Multiple Independent* Normals

Let $X \sim N(\mu_x, \sigma_x^2)$ independent of $Y \sim N(\mu_y, \sigma_y^2)$. Then if a, b, c are constants:

$$aX + bY + c \sim N(a\mu_x + b\mu_y + c, a^2\sigma_x^2 + b^2\sigma_y^2)$$

Important

- ▶ Result assumes independence
- ▶ Particular to Normal RV
- ▶ Extends to more than two Normal RVs

Suppose $X_1, X_2, \sim \text{iid } N(\mu, \sigma^2)$



Let $\bar{X} = (X_1 + X_2)/2$. What is the distribution of \bar{X} ?

- (a) $N(\mu, \sigma^2/2)$
- (b) $N(0, 1)$
- (c) $N(\mu, \sigma^2)$
- (d) $N(\mu, 2\sigma^2)$
- (e) $N(2\mu, 2\sigma^2)$

Where does the Empirical Rule come from?

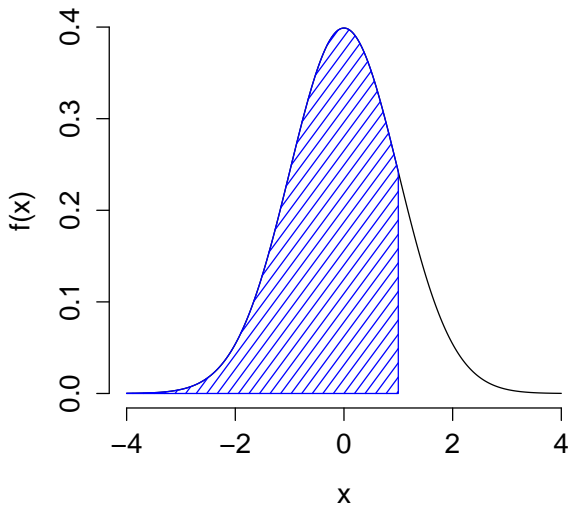
Empirical Rule

Approximately 68% of observations within $\mu \pm \sigma$

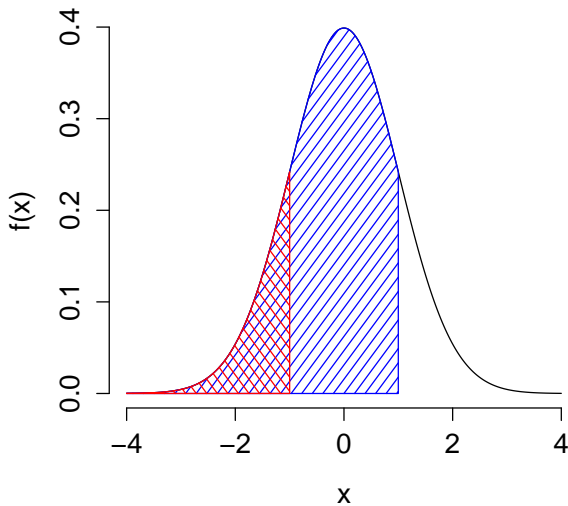
Approximately 95% of observations within $\mu \pm 2\sigma$

Nearly all observations within $\mu \pm 3\sigma$

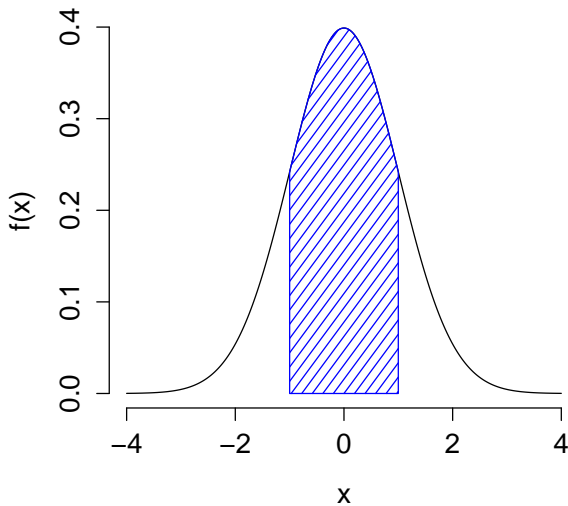
$$\text{pnorm}(1) \approx 0.84$$



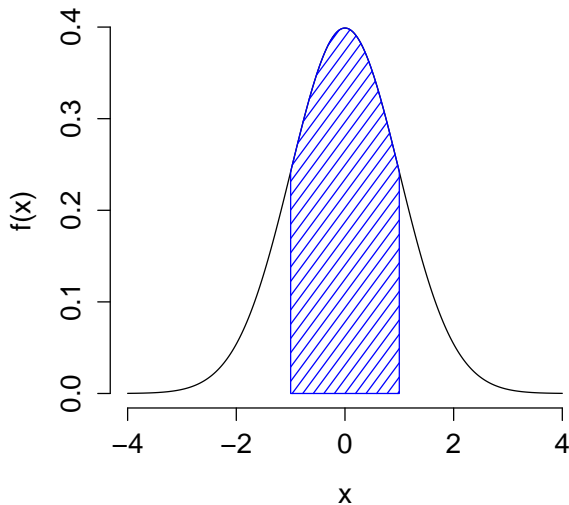
$$\text{pnorm}(1) - \text{pnorm}(-1) \approx 0.84 - 0.16$$



$$\text{pnorm}(1) - \text{pnorm}(-1) \approx 0.68$$



Middle 68% of $N(0, 1) \Rightarrow$ approx. $(-1, 1)$



Suppose $X \sim N(0, 1)$

$$\begin{aligned} P(-1 \leq X \leq 1) &= \text{pnorm}(1) - \text{pnorm}(-1) \\ &\approx 0.683 \end{aligned}$$

$$\begin{aligned} P(-2 \leq X \leq 2) &= \text{pnorm}(2) - \text{pnorm}(-2) \\ &\approx 0.954 \end{aligned}$$

$$\begin{aligned} P(-3 \leq X \leq 3) &= \text{pnorm}(3) - \text{pnorm}(-3) \\ &\approx 0.997 \end{aligned}$$

What if $X \sim N(\mu, \sigma^2)$?

$$P(X \leq a) =$$

What if $X \sim N(\mu, \sigma^2)$?

$$P(X \leq a) = P(X - \mu \leq a - \mu)$$

=

What if $X \sim N(\mu, \sigma^2)$?

$$\begin{aligned} P(X \leq a) &= P(X - \mu \leq a - \mu) \\ &= P\left(\frac{X - \mu}{\sigma} \leq \frac{a - \mu}{\sigma}\right) \\ &= \end{aligned}$$

What if $X \sim N(\mu, \sigma^2)$?

$$\begin{aligned}P(X \leq a) &= P(X - \mu \leq a - \mu) \\&= P\left(\frac{X - \mu}{\sigma} \leq \frac{a - \mu}{\sigma}\right) \\&= P\left(Z \leq \frac{a - \mu}{\sigma}\right)\end{aligned}$$

Where Z is a standard normal random variable, i.e. $N(0, 1)$.



Which of these equals $P(Z \leq (a - \mu)/\sigma)$ if $Z \sim N(0, 1)$?

- (a) `pnorm(a)`
- (b) $1 - \text{pnorm}(a)$
- (c) $\text{pnorm}(a)/\sigma - \mu$
- (d) $\text{pnorm}\left(\frac{a-\mu}{\sigma}\right)$
- (e) None of the above.

Probability Above a Threshold: $X \sim N(\mu, \sigma^2)$

$$\begin{aligned}P(X \geq b) &= 1 - P(X \leq b) = 1 - P\left(\frac{X - \mu}{\sigma} \leq \frac{b - \mu}{\sigma}\right) \\&= 1 - P\left(Z \leq \frac{b - \mu}{\sigma}\right) \\&= 1 - \text{pnorm}((b - \mu)/\sigma)\end{aligned}$$

Where Z is a standard normal random variable.

Probability of an Interval: $X \sim N(\mu, \sigma^2)$

$$\begin{aligned}P(a \leq X \leq b) &= P\left(\frac{a - \mu}{\sigma} \leq \frac{X - \mu}{\sigma} \leq \frac{b - \mu}{\sigma}\right) \\&= P\left(\frac{a - \mu}{\sigma} \leq Z \leq \frac{b - \mu}{\sigma}\right) \\&= \text{pnorm}((b - \mu)/\sigma) - \text{pnorm}((a - \mu)/\sigma)\end{aligned}$$

Where Z is a standard normal random variable.

Suppose $X \sim N(\mu, \sigma^2)$



What is $P(\mu - \sigma \leq X \leq \mu + \sigma)$?

Suppose $X \sim N(\mu, \sigma^2)$



What is $P(\mu - \sigma \leq X \leq \mu + \sigma)$?

$$\begin{aligned} P(\mu - \sigma \leq X \leq \mu + \sigma) &= P\left(-1 \leq \frac{X - \mu}{\sigma} \leq 1\right) \\ &= P(-1 \leq Z \leq 1) \\ &= \text{pnorm}(1) - \text{pnorm}(-1) \\ &\approx 0.68 \end{aligned}$$

Percentiles/Quantiles for Continuous RVs

Quantile Function $Q(p)$ is the inverse of CDF $F(x_0)$

Plug in a probability p , get out the value of x_0 such that $F(x_0) = p$

$$Q(p) = F^{-1}(p)$$

In other words:

$$Q(p) = \text{the value of } x_0 \text{ such that } \int_{-\infty}^{x_0} f(x) dx = p$$

Inverse exists as long as $F(x_0)$ is *strictly increasing*.

Example: Median

The median of a continuous random variable is $Q(0.5)$, i.e. the value of x_0 such that

$$\int_{-\infty}^{x_0} f(x) dx = 1/2$$

What is the median of a standard normal RV?





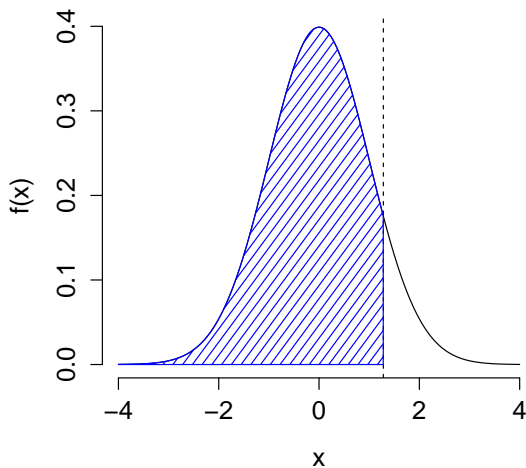
What is the median of a standard normal RV?

By symmetry, $Q(0.5) = 0$. R command: `qnorm()`



90th Percentile of a Standard Normal

$$\text{qnorm}(0.9) \approx 1.28$$



Using Quantile Function to find Symmetric Intervals

Suppose X is a standard normal RV. What is the value of c such that $P(-c \leq X \leq c) = 0.5$?



$$\text{qnorm}(0.75) \approx 0.67$$

Suppose X is a standard normal RV. What is the value of c such that $P(-c \leq X \leq c) = 0.5$?



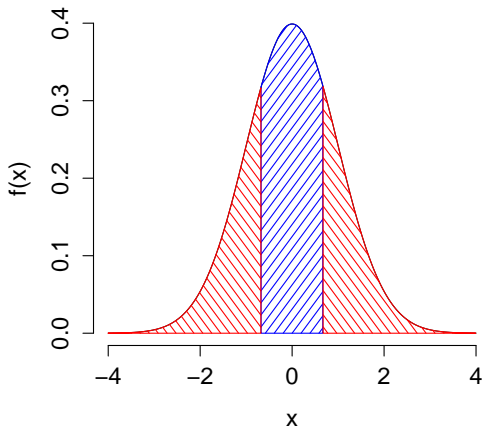
$$\text{qnorm}(0.75) \approx 0.67$$

Suppose X is a standard normal RV. What is the value of c such that $P(-c \leq X \leq c) = 0.5$?



$$\text{pnorm}(0.67) - \text{pnorm}(-0.67) \approx ?$$

Suppose X is a standard normal RV. What is the value of c such that $P(-c \leq X \leq c) = 0.5$?



$$\text{pnorm}(0.67) - \text{pnorm}(-0.67) \approx 0.5$$

Suppose X is a standard normal RV. What is the value of c such that $P(-c \leq X \leq c) = 0.5$?



95% Central Interval for Standard Normal



Suppose X is a standard normal random variable. What value of c ensures that $P(-c \leq X \leq c) \approx 0.95$?

R Commands for *Arbitrary* Normal RVs

Let $X \sim N(\mu, \sigma^2)$. Then we can use R to evaluate the CDF and Quantile function of X as follows:

CDF $F(x)$	<code>pnorm(x, mean = μ, sd = σ)</code>
Quantile Function $Q(p)$	<code>qnorm(p, mean = μ, sd = σ)</code>

Notice that this means you don't have to transform X to a standard normal in order to find areas under its pdf using R.

Example from Homework: $X \sim N(0, 16)$

One Way:

$$\begin{aligned} P(X \geq 10) &= 1 - P(X \leq 10) = 1 - P(X/4 \leq 10/4) \\ &= 1 - P(Z \leq 2.5) = 1 - \Phi(2.5) = 1 - \text{pnorm}(2.5) \\ &\approx 0.006 \end{aligned}$$

Example from Homework: $X \sim N(0, 16)$

One Way:

$$\begin{aligned}P(X \geq 10) &= 1 - P(X \leq 10) = 1 - P(X/4 \leq 10/4) \\&= 1 - P(Z \leq 2.5) = 1 - \Phi(2.5) = 1 - \text{pnorm}(2.5) \\&\approx 0.006\end{aligned}$$

An Easier Way:

$$\begin{aligned}P(X \geq 10) &= 1 - P(X \leq 10) \\&= 1 - \text{pnorm}(10, \text{mean} = 0, \text{sd} = 4) \\&\approx 0.006\end{aligned}$$