Lecture 7

Normal Dist'n

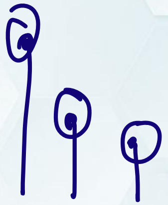
agenda

Announcements	R resources, join discord
Mini-lecture	The normal distribution
Rmd	Code along!
Wrap-up	HW 3 due Friday

Debugging tips

- Once again, DO NOT spend more than 30 minutes trying to figure something out! Post on the discussion board!
- Delete portions of your document until you locate where the error is
- Copy and paste the error message (not the whole thing) into Google
- Restart R (cmd/ctrl + shift + 0) <- zero, not “O”
- Run portions of the code at a time
- …

Normal (Gaussian) Distribution



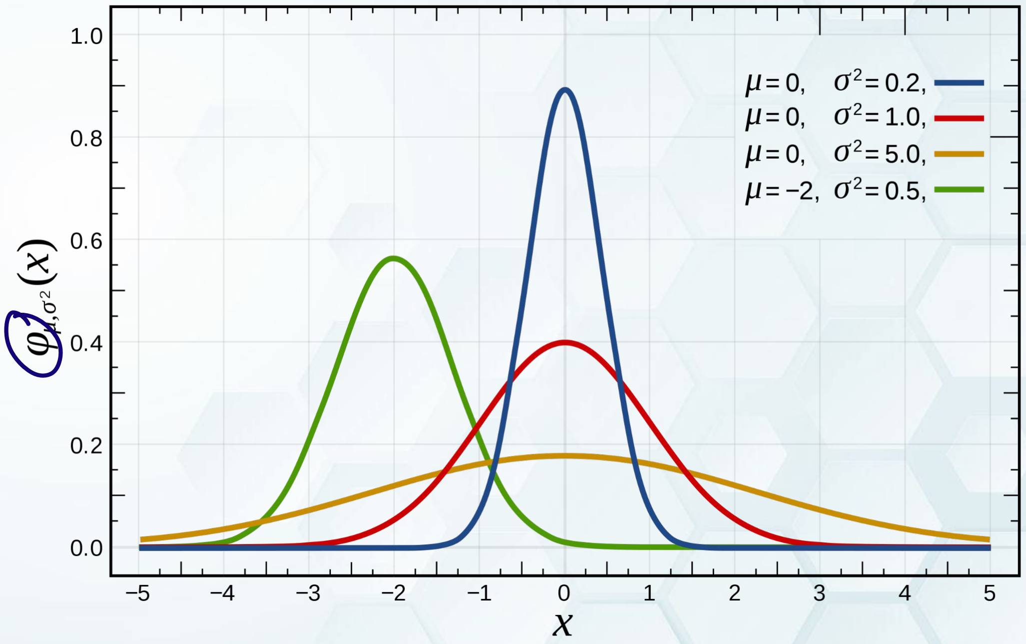
Describes events who probabilities follow a symmetric, unimodal, bell-shaped curve

$$f(x) = P(X = x)$$
$$F(x) = P(X \le x)$$

$$-\infty < X < \infty$$

$$X \sim N(\mu, \sigma)$$

$f(x)$



Wiki image

Notation and examples

01

$X \underset{\uparrow}{\sim} N(\mu, \sigma)$

X "is distributed" normal with mean μ and standard deviation σ

02

Parameters

mean $E[X] = \mu$ and

μ_X

variance $Var[X] = \sigma^2$

σ_X^2

03

pdf

$$f(x) = (2\pi\sigma^2)^{-1/2} \exp\left\{ -\frac{1}{2\sigma^2}(x-\mu)^2 \right\}$$

$$-\infty < x < \infty$$

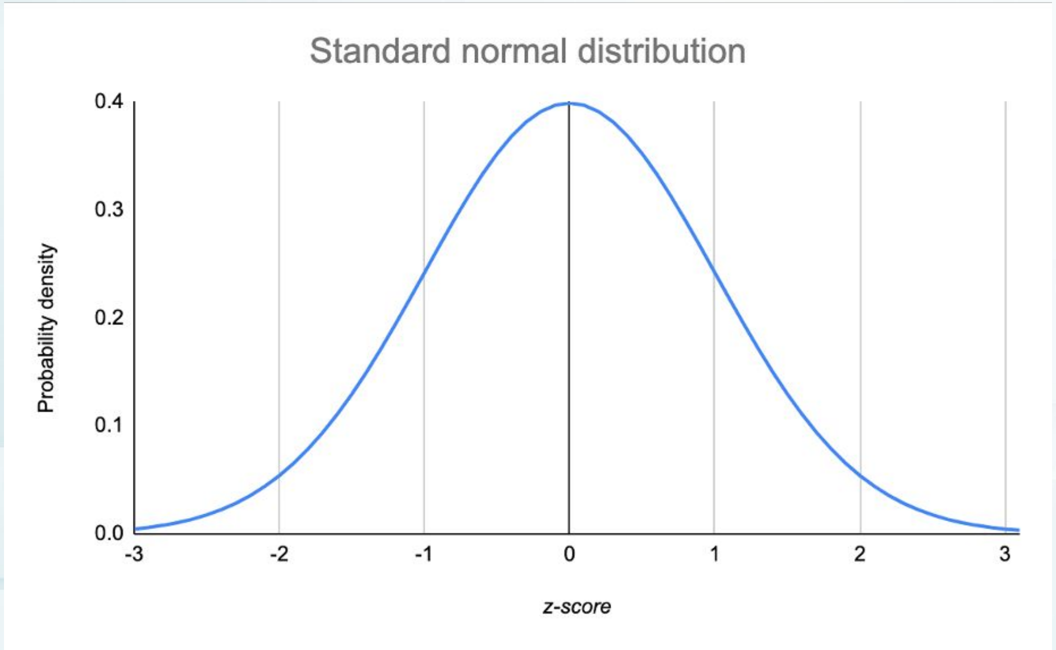
04

examples

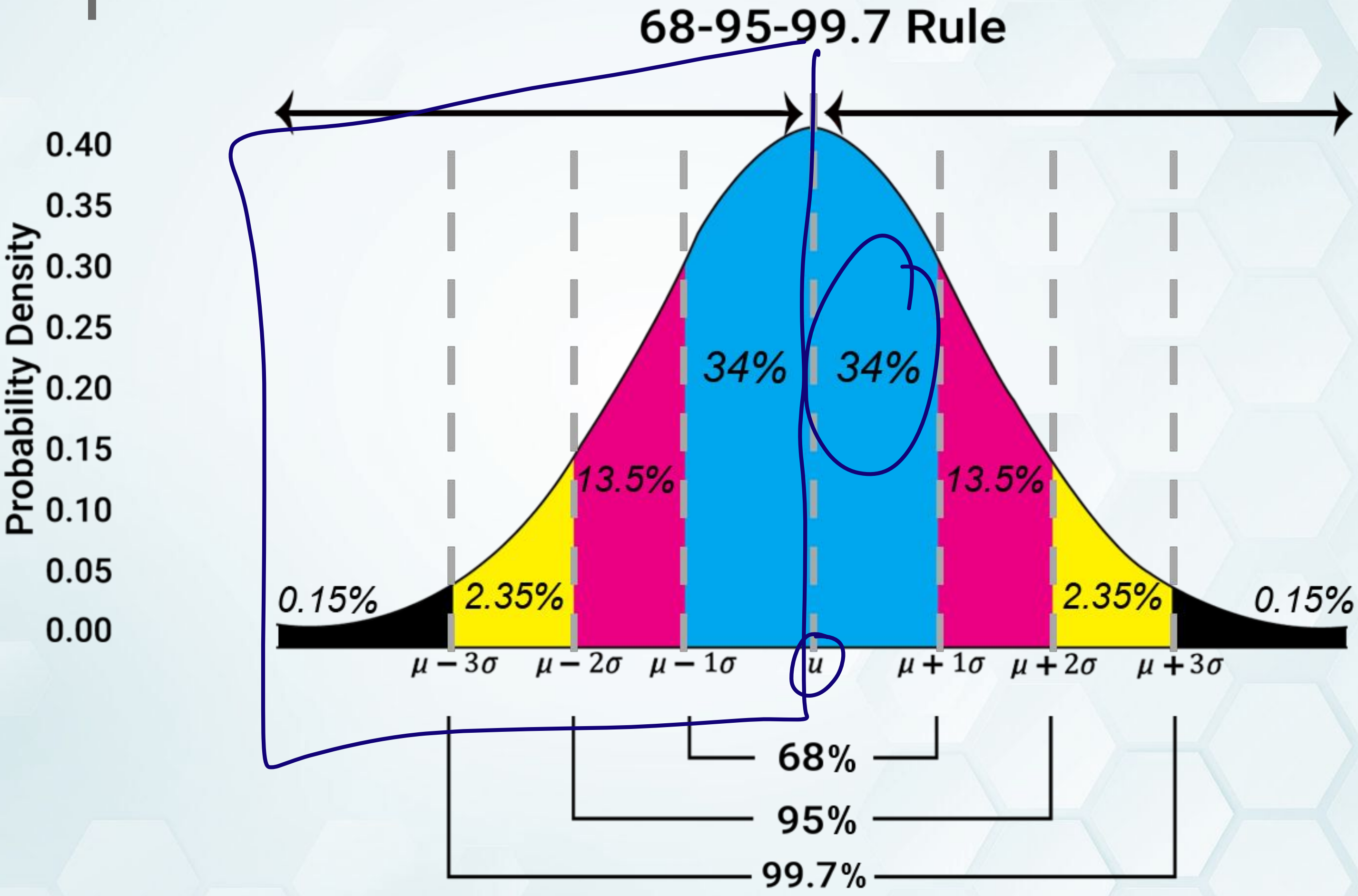
- Blood pressure
- shoe size
- job satisfaction
- stock market technical chart

Standard normal distribution

A special case of the normal distribution that occurs when a normally distributed random variable has mean 0 and standard deviation/variance 1.



The Empirical Rule



z-scores

Standardized score (or Z-score)

The number of standard deviations an observation falls above or below the mean

$$z = \frac{Observed\ value\ - mean}{standard\ deviation} = \frac{x-\mu}{\sigma}$$

- We can calculate probabilities using the standard normal distribution

Ex: Say an observation lies 1 standard deviation above the mean. Then, that observation lies at the 84th percentile (using the empirical rule)

Prove the standard normal has mean 0 and standard deviation 1

$$Z = \frac{X-\mu}{\sigma} \quad , \quad X \sim N(\mu, \sigma)$$

$$E[Z] = E\left[\frac{X-\mu}{\sigma}\right] = \frac{1}{\sigma}E[X-\mu]$$

$$= \frac{1}{\sigma}\left(E[X] - E[\mu]\right)$$

$$= \frac{1}{\sigma}(\mu - \mu)$$

$$\boxed{= 0}$$

$$\operatorname{var}[Z] = \operatorname{var}\left[\frac{X-\mu}{\sigma}\right]$$

$$= \frac{1}{\sigma^2}\operatorname{var}[X-\mu]$$

$$= \frac{1}{\sigma^2}\operatorname{var}[X] = \frac{1}{\sigma^2} \times \sigma^2 = \boxed{1}$$

Computing normal probabilities in R

$P(X = x) = 0$

pnorm(q, mean = 0, sd = 1, lower.tail = TRUE) $= P(X \le x)$ $= P(X < x)$

$P(X \le x)$

q

quantile

mean

mean of the distribution (default is set to 0)

sd

standard deviation of the distribution (default is set to 1)

lower.tail

Find the area above or below the quantile (default is TRUE)

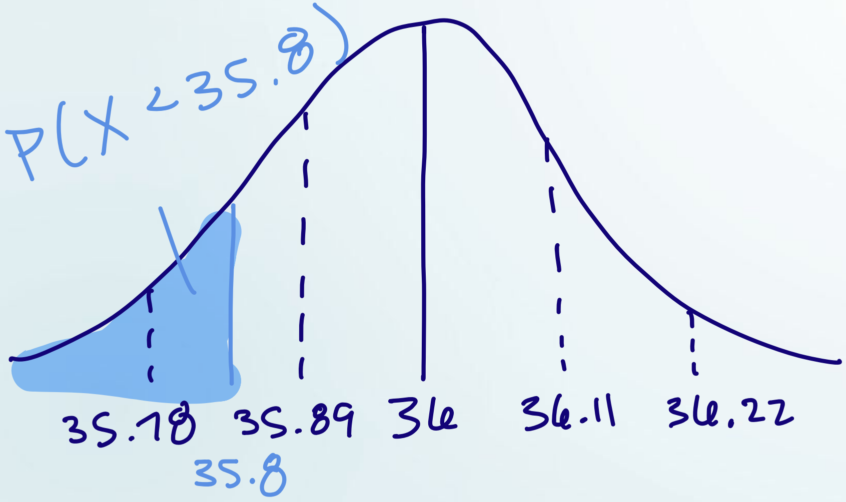
$P(X < x)$

- lower.tail = FALSE = P(X $>$ x)
- *lower.tail = TRUE = P(X $\le$ x)

At Heinz ketchup factory the amounts which go into bottles of ketchup are supposed to be normally distributed with mean 36 oz. and standard deviation 0.11 oz. Once every 30 minutes a bottle is selected from the production line, and its contents are noted precisely. If the amount of ketchup in the bottle is below 35.8 oz. or above 36.2 oz., then the bottle fails the quality control inspection. What percent of bottles have less than 35.8 ounces of ketchup?

X = amount of ketchup in a randomly selected bottle

$X \sim N(36, 0.11)$

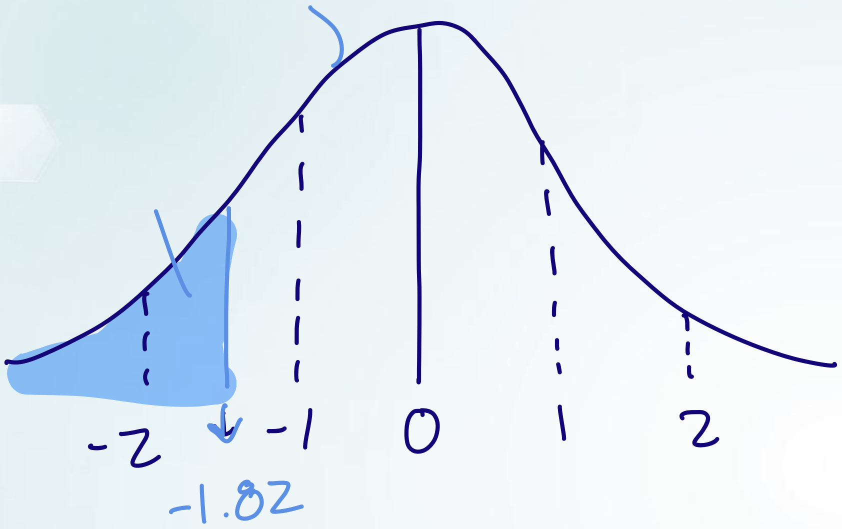


$$P(X < 35.8) = P\left(\frac{X-\mu}{\sigma} < \frac{35.8-36}{0.11}\right)$$

$$= P(Z < -1.82)$$

$$= \text{pnorm}(-1.82)$$

$$= 0.0344$$



pnorm(35.8, 36, 0.11)

pnorm(q = 35.8, mean = 36,
sd = 0.11)

$\boxed{= 0.0345}$

Getting a random sample in R

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

rnorm(50, 36, 0.11)

n

number of samples

50

mean

mean of the distribution (default is set to 0)

sd

standard deviation of the distribution (default is set to 1)

Assessing normality

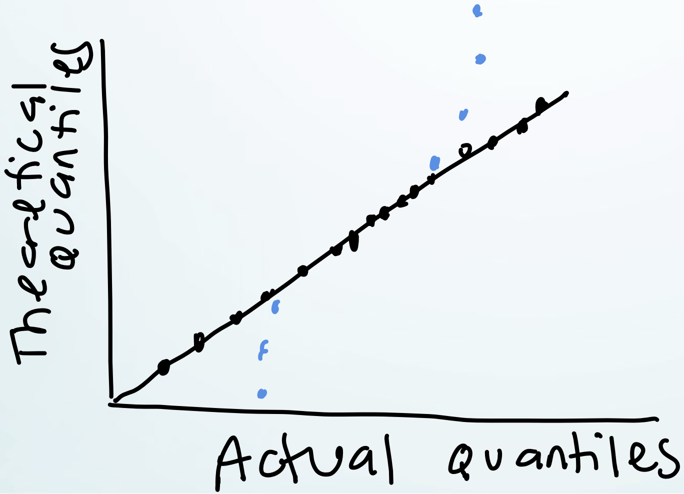
qqplot()
qqline()

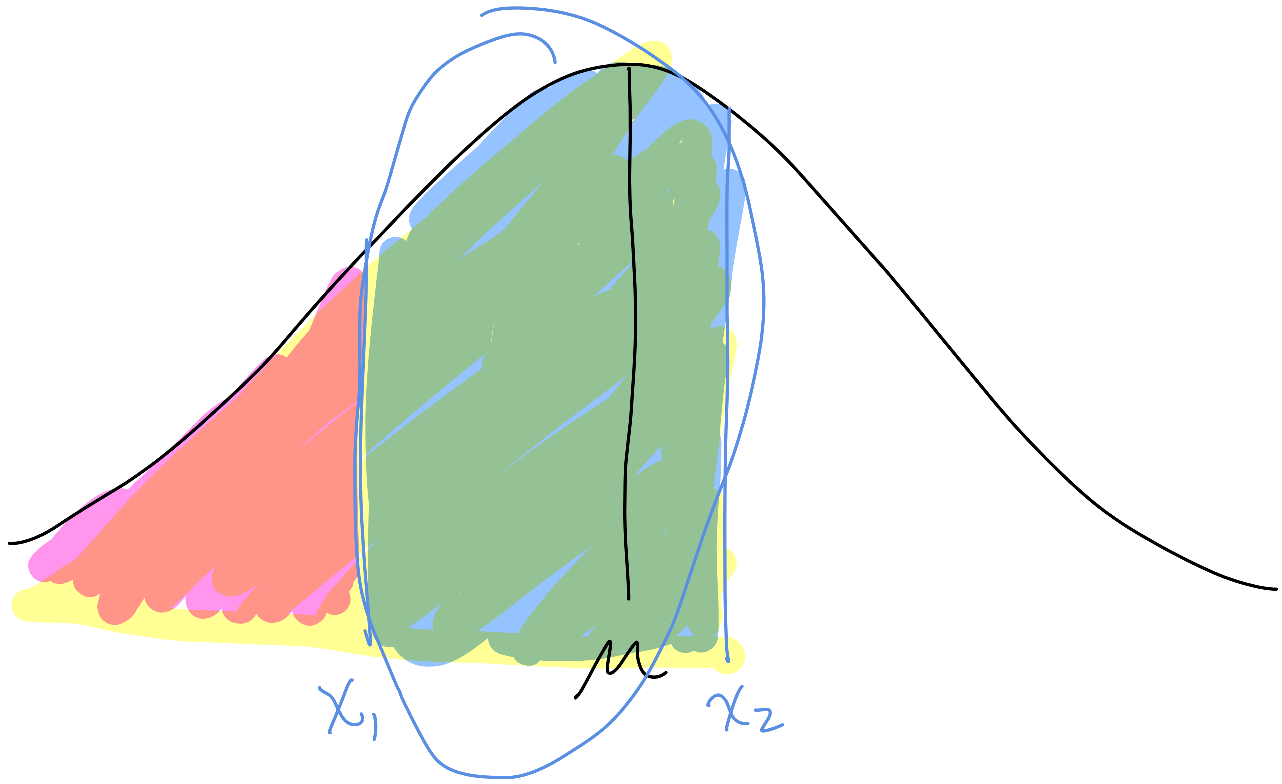
qqnorm(y)

Input: numeric vector
Output: qqplot

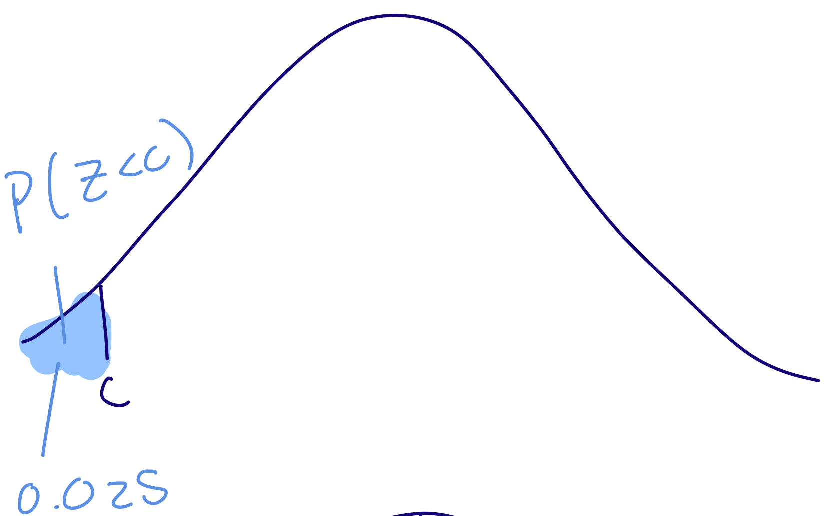
qqline(y)

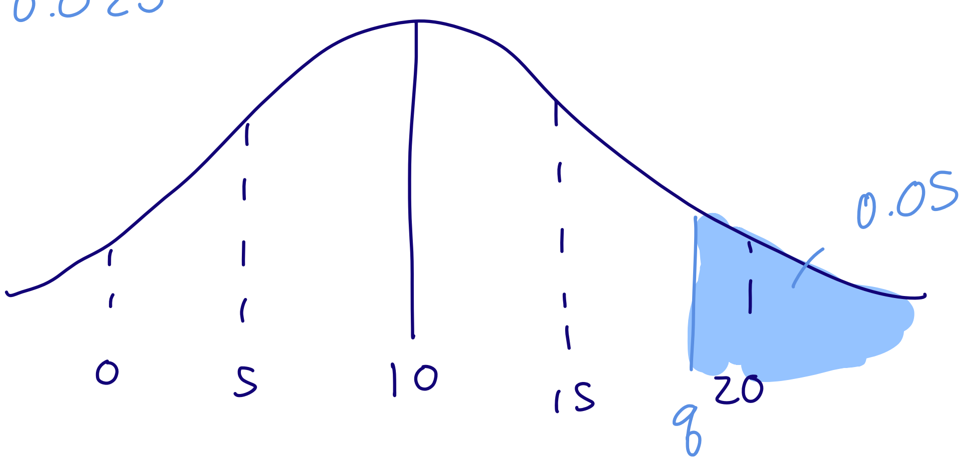
Input: numeric vector
Output: line to represent a perfect normal distribution



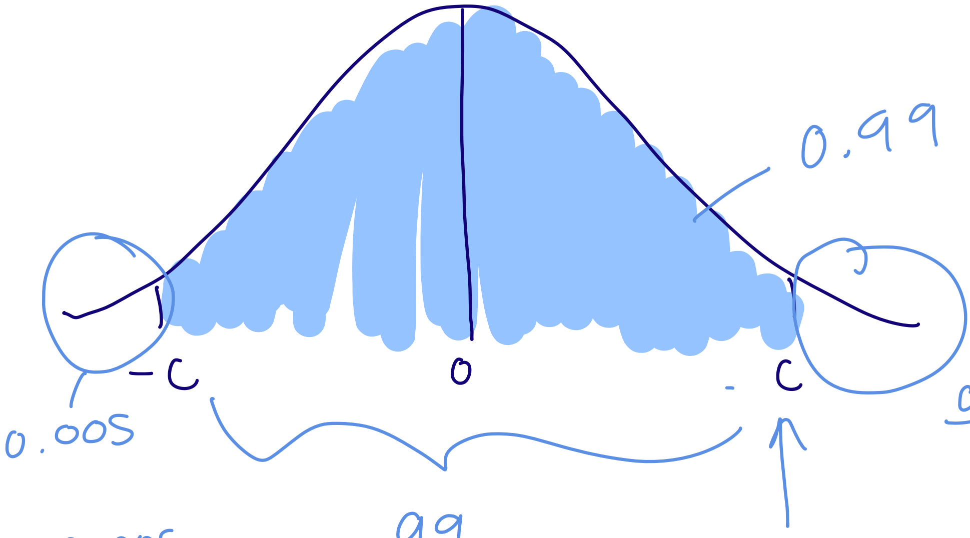


$$P(x_1 < X < x_2) = P(X < x_2) - P(X < x_1)$$





qnorm(0.95, m, sd)



$\frac{0.01}{2} = 0.005$

$0.005 + .99$

0.995