Quantiles

popper 12 : choose d for 1-5

Let F be a given cumulative distribution and let p be any real number between 0 and 1. The **(100p)th percentile** of the distribution of a continuous random variable X is defined as

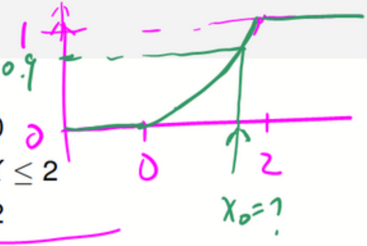
$$F^{-1}(p) = min\{x | F(x) \geq p\}.$$

For continuous distributions, $F^{-1}(p)$ is the smallest number x such that $F(x) = p$.

Determine the Percentiles

$P(X \le c) = P(X < c)$

Given a cdf,



$$F(x) = \begin{cases} 0 & X < 0 \\ \frac{1}{8}x^3 & 0 \le X \le 2 \\ 1 & X > 2 \end{cases}$$

1. Determine the 90^{th} percentile.

set $F(X_0) = 0.9$, solve for X_0

set $\frac{1}{8}X_0^3 = 0.90$

$\Rightarrow X_0 = (8 * 0.90)^{\frac{1}{3}}$

2. Determine the 50^{th} percentile.

set $\frac{1}{8}X_0^3 = 0.5$, solve for X_0.

$X_0 = (8 * 0.5)^{\frac{1}{3}}$

3. Find the value of c such that $P(X \le c) = 0.75$.

$P(X \le c) = F(c)$ set $\frac{1}{8}c^3 = 0.75$, solve for c.

MATH 3339
Statistics for the Sciences
Sec 5.4; 5.5

Wendy Wang
wwang60@central.uh.edu

Lecture 10 - 3339

UNIVERSITY of HOUSTON
DEPARTMENT OF MATHEMATICS

Outline

for discrete,

$$E(X) = \sum x P(X=x)$$

1. Expected Values
2. Exponential Distribution
3. Gamma Distribution
4. Normal Distribution
5. The Empirical Rule

UNIVERSITY of HOUSTON
DEPARTMENT OF MATHEMATICS

Expected Values for Continuous Random Variables

The **expected** or **mean value** of a continuous random variable X with pdf $f(x)$ is

$$E(X) = \int_{-\infty}^{\infty} xf(x)dx.$$

More generally, if h is a function defined on the range of X,

$h(x)$

$$E(h(X)) = \int_{-\infty}^{\infty} h(x)f(x)dx.$$

$sd(X) = \sqrt{Var(X)}$

for example, $h(x) = x^2$

$E(x^2) = \int_{-\infty}^{\infty} (x^2) f(x) dx$

$Var(X) = E((X-\mu)^2) = E(X^2) - (E(X))^2$

Example From Quiz 8

$c' = 0$

x

Let X be the amount of time (in hours) the wait is to get a table at a restaurant. Suppose the cdf is represented by

$$F(X) = \begin{cases} 0 & x < 0 \\ \frac{x^2}{9} & 0 \leq x \leq 3 \\ 1 & x > 3 \end{cases}$$

Use the cdf to determine E[X]. $\rightarrow$

$$f(x) = F'(x) = \begin{cases} 0, & x<0, \text{ or } x>3 \\ \frac{2x}{9}, & 0 \leq x \leq 3 \end{cases}$$

$$E(x) = \int_{-\infty}^{\infty} x f(x)\,dx$$

$$= \cancel{\int_{-\infty}^{0} (x \cdot 0)\,dx} + \int_0^3 \left(x \cdot \frac{2x}{9}\right) dx + \cancel{\int_3^{\infty} (x \cdot 0)\,dx}$$

$$= \int_0^3 \frac{2x^2}{9}\,dx = \frac{2}{27} x^3 \Big|_0^3 = \frac{2}{27}(3^3 - 0) = 2.$$

UNIVERSITY of HOUSTON
DEPARTMENT OF MATHEMATICS

The Exponential Distribution

binomial(n, p)

X is said to have an **exponential distribution** with parameter λ (*lambda* > 0) if the pdf of X is:

$$f(x) = \begin{cases} \lambda e^{-\lambda x} & x \geq 0 \\ 0 & otherwise \end{cases}$$

Where λ is a rate parameter, we write $X \sim Exp(\lambda)$. The cdf of a exponential random variable is:

$$F(x) = \begin{cases} 0 & x < 0 \\ 1 - e^{-\lambda x} & x \geq 0 \end{cases}$$

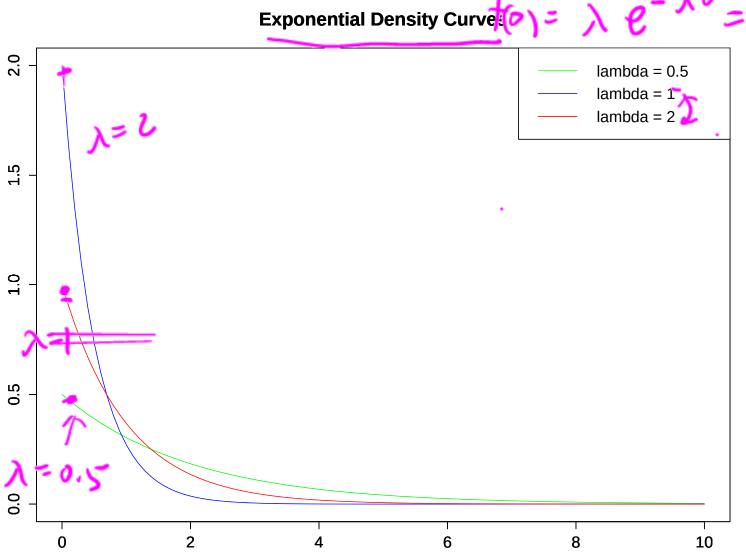
The mean of the exponential distribution is $\mu_X = E(X) = \frac{1}{\lambda}$ the standard deviation is also $\frac{1}{\lambda}$.

UNIVERSITY of HOUSTON
DEPARTMENT OF MATHEMATICS

Exponential Density Curves

$f(x) = \lambda e^{-\lambda x}$

$f(0) = \lambda e^{-\lambda 0} = \lambda$



Exponential Distribution Related to the Poisson Distribution

λ continuous.

discrete

- The exponential distribution is frequently used as a model for the distribution of times between the occurrence of successive events until the first arrival.
- Suppose that the number of events occurring in any time of length t has a Poisson distribution with parameter λt.
- Where λ, the rate of the event process, is the expected number of events occurring in 1 unit of time.
- The number of occurrences are in non overlapping intervals and are independent of one another.
- Then the distribution of elapsed time between the occurrence of two successive events is exponential with parameter λ.

UNIVERSITY of HOUSTON
DEPARTMENT OF MATHEMATICS

Example

- Suppose you usually get 3 phone calls per hour.
- 3 phone calls per hour means that we would expect one phone call every $\frac{1}{3}$ hour so $\mu = \frac{1}{3}$.
- Compute the probability that a phone call will arrive within the next hour.

Example

of phone calls/hr $\sim$ possion($\lambda = 3$/hr.)

- Suppose you usually get 3 phone calls per hour.
- 3 phone calls per hour means that we would expect one phone call every $\frac{1}{3}$ hour so $\mu = \frac{1}{3}$.
- Compute the probability that a phone call will arrive within the next hour.

X : time it takes for the next phone call

$X \sim \exp(\lambda = 3)$ $\quad E(X) = \frac{1}{\lambda} = \frac{1}{3}$

$P(X \le 1) = \text{pexp}(1, 3)$

$P(X = 1) = 0$

R code

```
> pexp(1,3)
[1] 0.9502129
```

- To find the probability of an exponetial distribution in R: pexp(x,λ).
- To find the percentile (quantile) in R: qexp(x,λ).

Examples

Applications of the exponential distribution occurs naturally when describing the waiting time in a homogeneous Poisson process. It can be used in a range of disciplines including queuing theory, physics, reliability theory, and hydrology. Examples of events that may be modeled by exponential distribution include:

- The time until a radioactive particle decays
- The time between clicks of a Geiger counter
- The time until default on payment to company debt holders
- The distance between roadkills on a given road
- The distance between mutations on a DNA strand
- The time it takes for a bank teller to serve a customer
- The height of various molecules in a gas at a fixed temperature and pressure in a uniform gravitational field
- The monthly and annual maximum values of daily rainfall and river discharge volumes

Example from Quiz 8

$\frac{1}{\lambda} = \mu = 6 \Rightarrow \lambda = \frac{1}{6}$ $X \sim \exp(\lambda = \frac{1}{6})$

1. Suppose the time a child spends waiting at for the bus as a school bus stop is exponentially distributed with mean 6 minutes. Determine the probability that the child must wait at least 9 minutes on the bus on a given morning.

$$P(X \geq 9) = 1 - P(X < 9) = 1 - P(X \leq 9)$$
$$= 1 - \text{pexp}(9, \tfrac{1}{6})$$

2. Suppose the time a child spends waiting at for the bus as a school bus stop is exponentially distributed with mean 4 minutes. Determine the probability that the child must wait between 3 and 6 minutes on the bus on a given morning.

$X \sim \exp(\lambda = \frac{1}{4})$

$$P(3 \leq X \leq 6) = P(X < 6) - P(X < 3)$$
$$= \text{pexp}(6, \tfrac{1}{4}) - \text{pexp}(3, \tfrac{1}{4})$$

UNIVERSITY of HOUSTON
DEPARTMENT OF MATHEMATICS

The "Memoryless" Property

Another application of the exponential distribution is to model the distribution of component lifetime.

- Suppose component lifetime is exponentially distributed with parameter λ.
- After putting the component into service, we leave for a period of t_0 hours and then return to find the components still working; what now is the probability that it last at least an addition t hours?
- We want to find $P(T \geq t + t_0 | T \geq t_0)$ $= P(T \geq t)$

UNIVERSITY of HOUSTON
DEPARTMENT OF MATHEMATICS

The gamma function $\Gamma(\alpha)$ is defined by:

$$\Gamma(\alpha) = \int_0^\infty x^{\alpha-1} e^{-x} dx$$

$\alpha=1$ $\Gamma(1) = \int_0^\infty x^{1-1} e^{-x} dx = \int_0^\infty e^{-x} dx$

$\rightarrow \exp(\lambda=1)$

UNIVERSITY of HOUSTON
DEPARTMENT OF MATHEMATICS

The most important properties of the gamma function are the following:

1. For any $\alpha > 1$, $\Gamma(\alpha) = \underline{(\alpha - 1)\Gamma(\alpha - 1)}$
2. For any positive integer, n, $\Gamma(n) = (n - 1)!$
3. $\Gamma(\frac{1}{2}) = \sqrt{\pi}$

$$\Gamma\left(\frac{3}{2}\right) = \left(\frac{3}{2} - 1\right)\Gamma\left(\frac{3}{2} - 1\right) = 0.5\,\Gamma(0.5) = \frac{\sqrt{\pi}}{2}$$

$\alpha = \frac{3}{2}$.

UNIVERSITY of HOUSTON
DEPARTMENT OF MATHEMATICS

The PDF of a Gamma Distribution

A continuous random variable X is said to have a **gamma distribution** if the pdf of X is

$$f(x;\alpha,\beta) = \begin{cases} \frac{1}{\beta^\alpha \Gamma(\alpha)} x^{\alpha-1} e^{-x/\beta} & x \geq 0 \\ 0 & \textit{otherwise} \end{cases}$$

where parameters α and β satisfy $\alpha > 0$, $\beta > 0$.

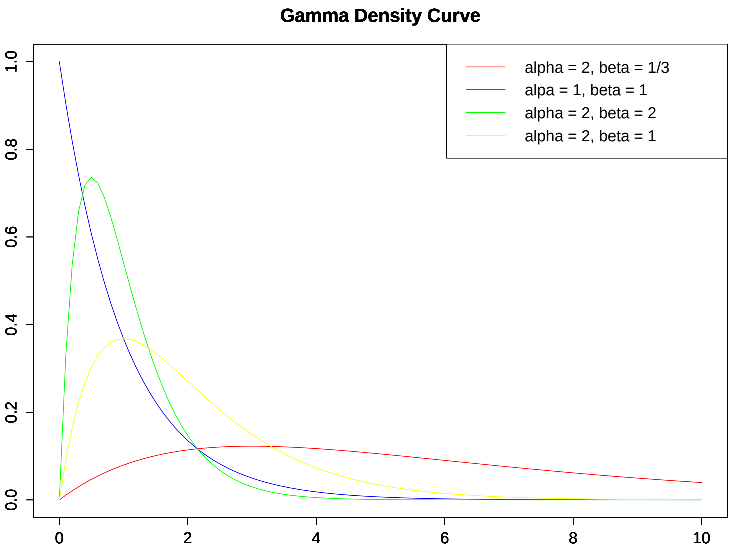
UNIVERSITY of HOUSTON
DEPARTMENT OF MATHEMATICS

Gamma Distribution Related to the Poisson

- Gamma distribution is a distribution that arises naturally in processes for which the waiting times between events are relevant.
- It can be thought of as a waiting time between Poisson distributed events, until *k* arrivals.
- Thus the scale parameter can also be thought of as the inverse of the rate parameter (λ), $\frac{1}{\lambda}$.
- Then $\alpha = k$ and $\beta = \frac{1}{\lambda}$
- In R, $P(X \leq x) = pgamma(x, \alpha, \frac{1}{\beta})$ $\rightarrow \lambda$

or, pgamma(x, α, scale = β)

Gamma Density Curve



Applications of the Gamma Distribution

The gamma distribution can be used a range of disciplines including queuing models, climatology, and financial services. Examples of events that may be modeled by gamma distribution include:

- The amount of rainfall accumulated in a reservoir
- The size of loan defaults or aggregate insurance claims
- The flow of items through manufacturing and distribution processes
- The load on web servers
- The many and varied forms of telecom exchange

Example

$\lambda = 5 / \text{min}$

Suppose that the telephone calls arriving at a prticular switchboard follow a Poisson process with an average of 5 calls coming per minute. What is the probability that up to a minute will elapse until 2 calls have come in to the switchboard?

$X \sim \text{gamma}(\alpha = 2, \beta = \frac{1}{5})$

- Average of 5 calls coming per minute means that $\beta = \frac{1}{5}$.
- Until 2 calls have come into the switchboard means that $\alpha = 2$.

$$P(X \le 1 \text{min}) = \text{pgamma}(1, 2, \underset{\uparrow}{5})$$

λ.

Mean and Variance of the Gamma Distribution

The mean and variance of a random variable X having the gamma distribution are:

$$
\begin{aligned}
E(X) &= \mu = \alpha\beta \\
Var(X) &= \sigma^2 = \alpha\beta^2
\end{aligned}
$$

$= \alpha \cdot \beta^2$

Example of Gamma Distribution

$\mu = 24$

$sd = 12$

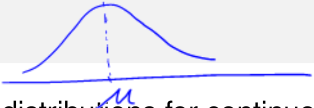
Suppose that a transistor of a certain type is subjected to an accelerated life test, the lifetime Y (in weeks) has a gamma distribution with a mean of 24 and a standard deviation of 12.

1. Find the values of α and β.

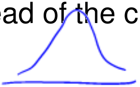
$$\left.\begin{aligned} \alpha\beta &= 24 \\ \alpha\beta^2 &= 12^2 = 144. \end{aligned}\right\} \qquad \beta = 6, \quad \alpha = 4$$

2. Find $P(Y \leq 24) =$ pgamma(24 , 4 , $\frac{1}{6}$)

The Normal distributions



- Common type of probability distributions for continuous random variables.
- The highest probability is where the values are centered around the mean. Then the probability declines the further from the mean a value gets.
- These curves are symmetric, single-peaked, and bell-shaped.
- The mean μ is located at the center of the curve and is the same as the median.
- The standard deviation σ controls the spread of the curve.
- If σ is small then the curve is tall and slim.



- If σ is large then the curve is short and fat.

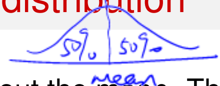
UNIVERSITY of HOUSTON
DEPARTMENT OF MATHEMATICS

Normal distributions important to statistics?

- Normal distributions are good descriptions for some distributions of real data.
- Normal distributions are good approximations to the results of many kinds of chance outcomes.
- Many statistical inference procedures based on Normal distributions work well for other roughly symmetric distributions.

UNIVERSITY of HOUSTON
DEPARTMENT OF MATHEMATICS

Facts about the Normal distribution



- The curve is symmetric about the mean. That is, 50% of the area under the curve is below the mean. 50% of the area under the curve is above the mean.

- The spread of the curve is determined by the standard deviation.

- The area under the curve is with respect to the number of standard deviations a value is from the mean.

- Total area under the curve is 1.

- Area under the curve is the same as probability within a range of values.

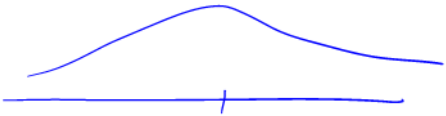
$P(-1 < X < 3) =$ Area under the curve.

- If X follows a Normal distribution with mean μ and standard deviation σ we would write it as $X \sim N(\mu, \sigma)$.

UNIVERSITY of HOUSTON
DEPARTMENT OF MATHEMATICS

$\mu = 1$ $\sigma = 0.5$

$X \sim N(\underline{1}, \underline{0.5})$



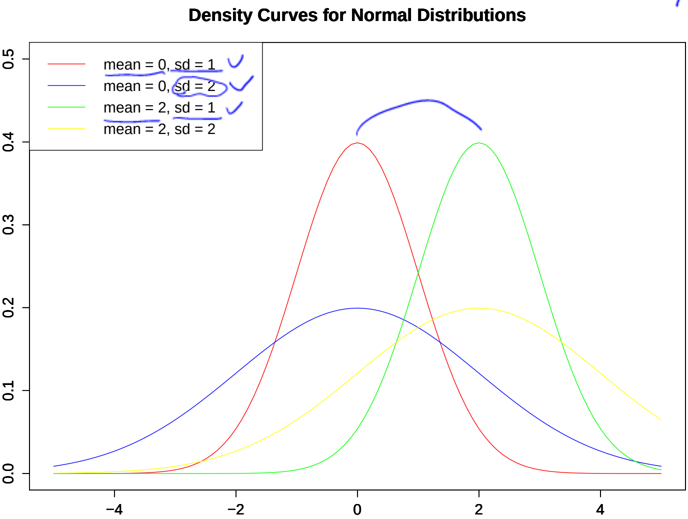
$\mu = 0$ $\sigma = 2.5$

$X \sim N(\overset{\mu}{0}, \overset{\sigma}{2.5})$

Density Function

μ: location parameter
σ: scale parameter

This is the graph of the density function.



PDF of a Normal Distribution

A continuous random variable X is said to have a **Normal distribution** with parameters μ and σ (or μ and σ^2), where $-\infty < \mu < \infty$ and $0 < \sigma$, if the pdf of X is:

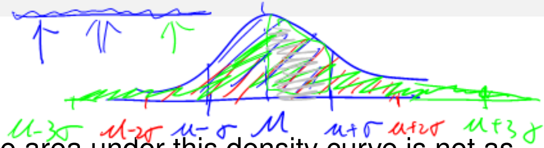
$$f(x) = \frac{1}{\sqrt{2\pi}\sigma} e^{-(x-\mu)^2/2\sigma^2}$$

For all $-\infty < x < \infty$.

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

UNIVERSITY of HOUSTON
DEPARTMENT OF MATHEMATICS

The Empirical Rule or 68-95-99.7 Rule



- Unfortunately to find the area under this density curve is not as easy to compute. Thus we can use the following approximate rule for the area under the **Normal** density curve.

- In the Normal Distribution with mean μ and standard deviation σ: $-\infty < \mu < \infty$ $\sigma > 0$
  - 68% of the observations fall within 1 standard deviation σ of the mean μ. $P(\mu-\sigma < X < \mu+\sigma) = 0.68$
  - 95% of the observations fall within 2 standard deviations 2σ of the mean μ. $P(\mu-2\sigma < X < \mu+2\sigma) = 0.95$
  - 99.7% of the observations fall within 3 standard deviations 3σ of the mean μ. $P(\mu-3\sigma < X < \mu+3\sigma) = 0.997$

$P(\mu < X < \mu+\sigma) = 0.34$

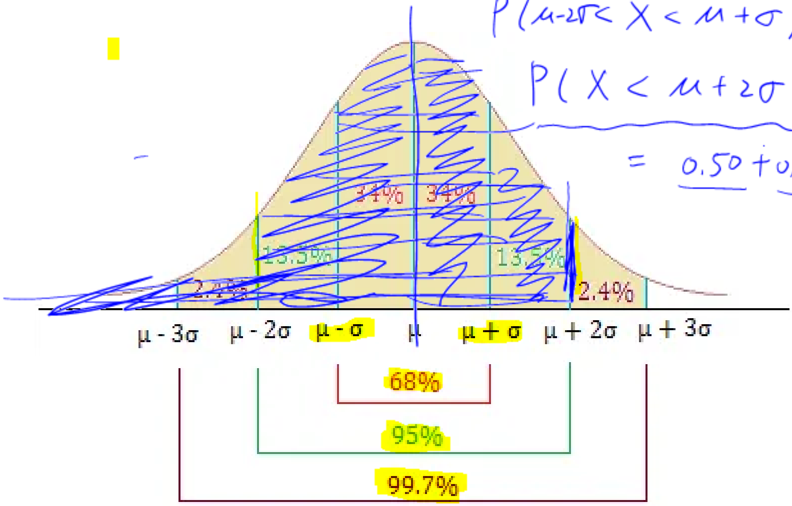
UNIVERSITY of HOUSTON
DEPARTMENT OF MATHEMATICS

The 68-95-99.7 rule for Normal distributions

$P(\mu - 2\sigma < X < \mu + \sigma) = ?$

$P(X < \mu + 2\sigma) = ?$

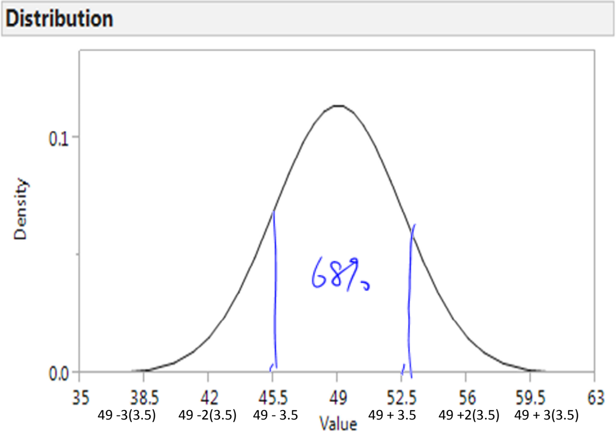
$= 0.50 + 0.34 + 0.135$



Empirical rule

MPG of Prius

The MPG of Prius has a Normal distribution with mean $\mu = 49$ mpg and standard deviation $\sigma = 3.5$ mpg.

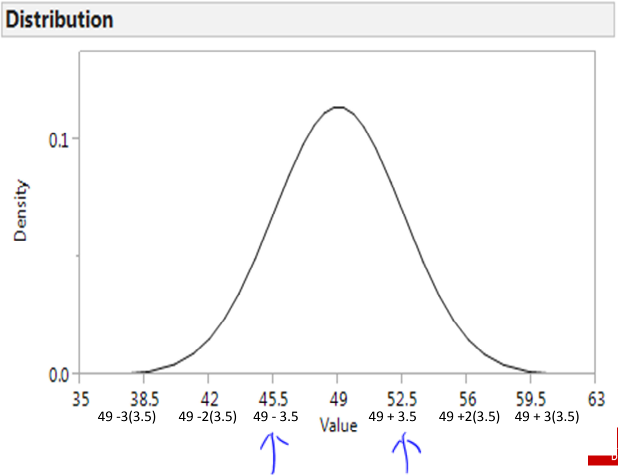


UNIVERSITY of HOUSTON
DEPARTMENT OF MATHEMATICS

MPG of Prius

About what percent have between 45.5 and 52.5 mpg?

$P(45.5 < X < 52.5) = 0.68$

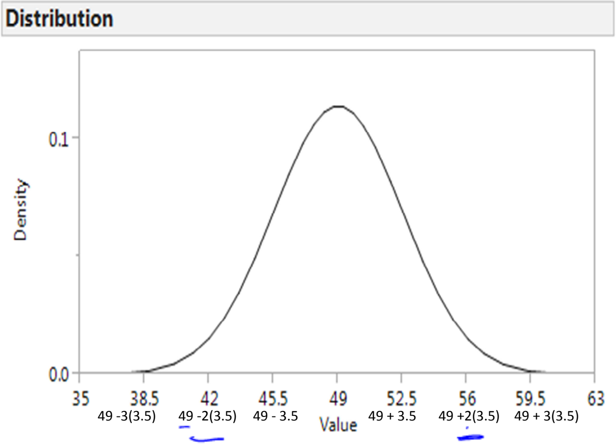


UNIVERSITY of HOUSTON
DEPARTMENT OF MATHEMATICS

MPG of Prius

About what percent have between 42 and 56 mpg?

$0.95 = P(\mu - 2\sigma < X < X + 2\sigma)$

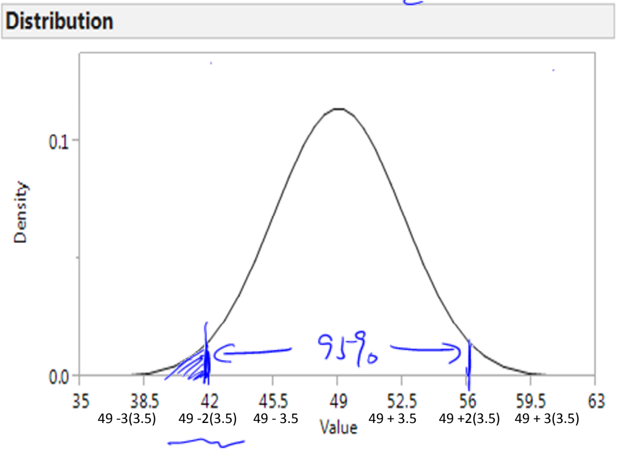


UNIVERSITY of HOUSTON
DEPARTMENT OF MATHEMATICS

MPG of Prius

About what percent have less than 42 mpg?

$P(X < 42) = \dfrac{1-95\%}{2} = 2.5\%$



UNIVERSITY of HOUSTON
DEPARTMENT OF MATHEMATICS

Orange Juice

$X \sim N(\mu = 4.7, \sigma = 0.40)$

An orange juice producer buys all his oranges from a large orange grove. The amount of juice squeezed from each of these oranges is approximately normally distributed, with a mean of 4.70 ounces and a standard deviation of 0.40 ounce. $3.9 = 4.7 - 0.4 * 2$

1. What is the probability that an orange from this orange grove has between 3.9 and 5.5 ounces of juice?

$P(3.9 < X < 5.5) = P(\mu - 2\sigma < X < \mu + 2\sigma)$

$5.5 = 4.7 + 0.4 * 2$ $\qquad = 0.95$

2. What is the probability that an orange from this orange grove has less than 4.7 ounces of juice?

$P(X < 4.7) = P(X < \mu) = 0.5$

3. Approximately ~~95%~~ of the oranges have juice between what two middle values? 99.7%

$(X - 3\sigma, \; X + 3\sigma)$

$(4.7 - 3 * 0.4, \; 4.7 + 3 * 0.4)$

UNIVERSITY of HOUSTON
DEPARTMENT OF MATHEMATICS

4. What percent of values less than 4?

$P(X < 4) = ?$

$4 = 4.7 - 0.7$