X is continuous, has support $\mathbb{R} = (-\infty, \infty)$.

STAT 515 Lec 07 slides

The Normal distribution

(Gaussian)

Karl Gregory

University of South Carolina

These slides are an instructional aid; their sole purpose is to display, during the lecture, definitions, plots, results, etc. which take too much time to write by hand on the blackboard. They are not intended to explain or expound on any material.

Normal or Gaussian probability distribution

A continuous rv X with pdf given by

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

has the *Normal distribution* with mean μ and $\boxed{\text{variance } \sigma^2}$

Carl Friedrich

Gauß

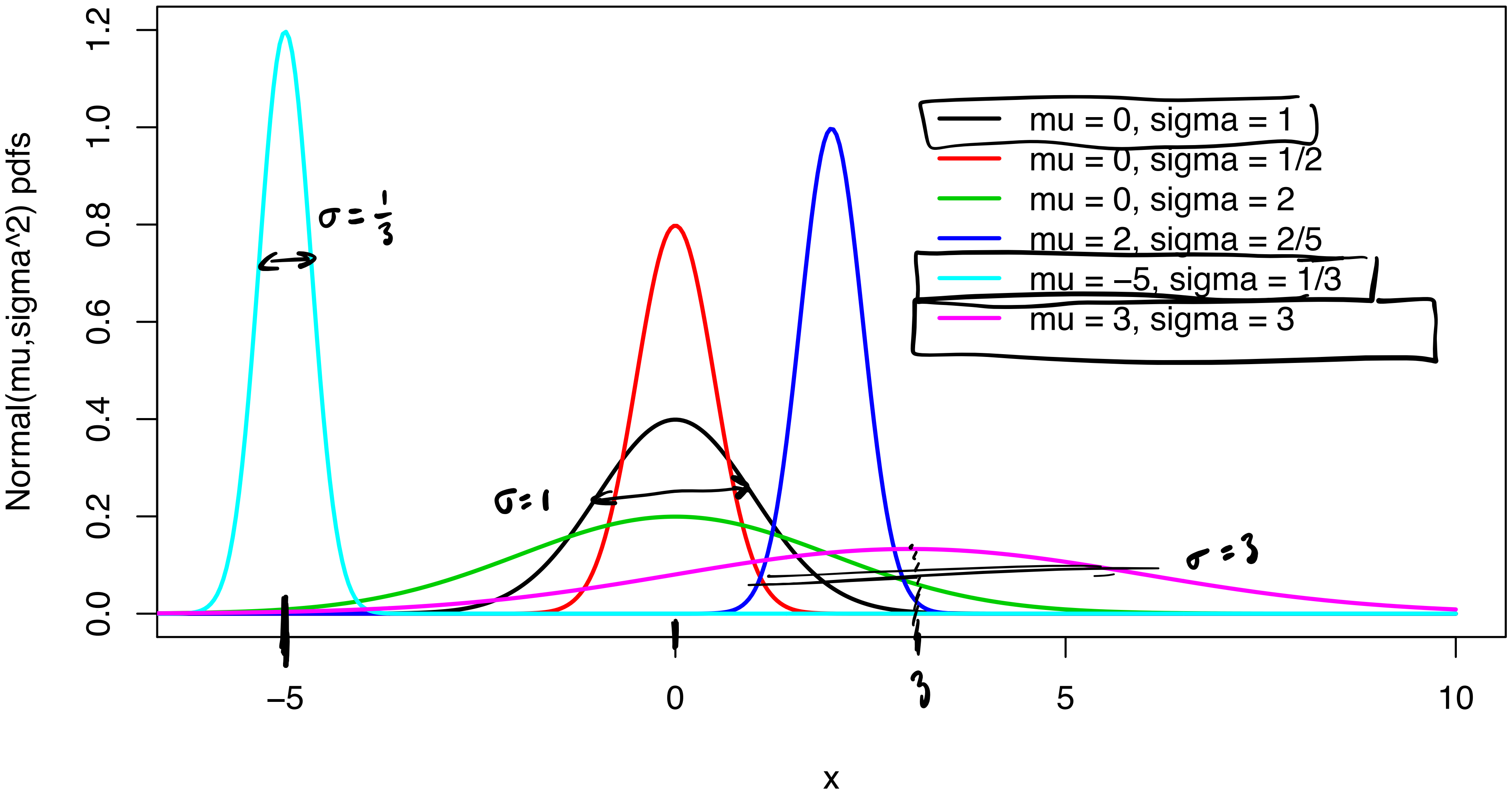
We write $\boxed{X \sim \mathcal{N}(\mu, \sigma^2).}$

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

pdfs of several Normal distributions

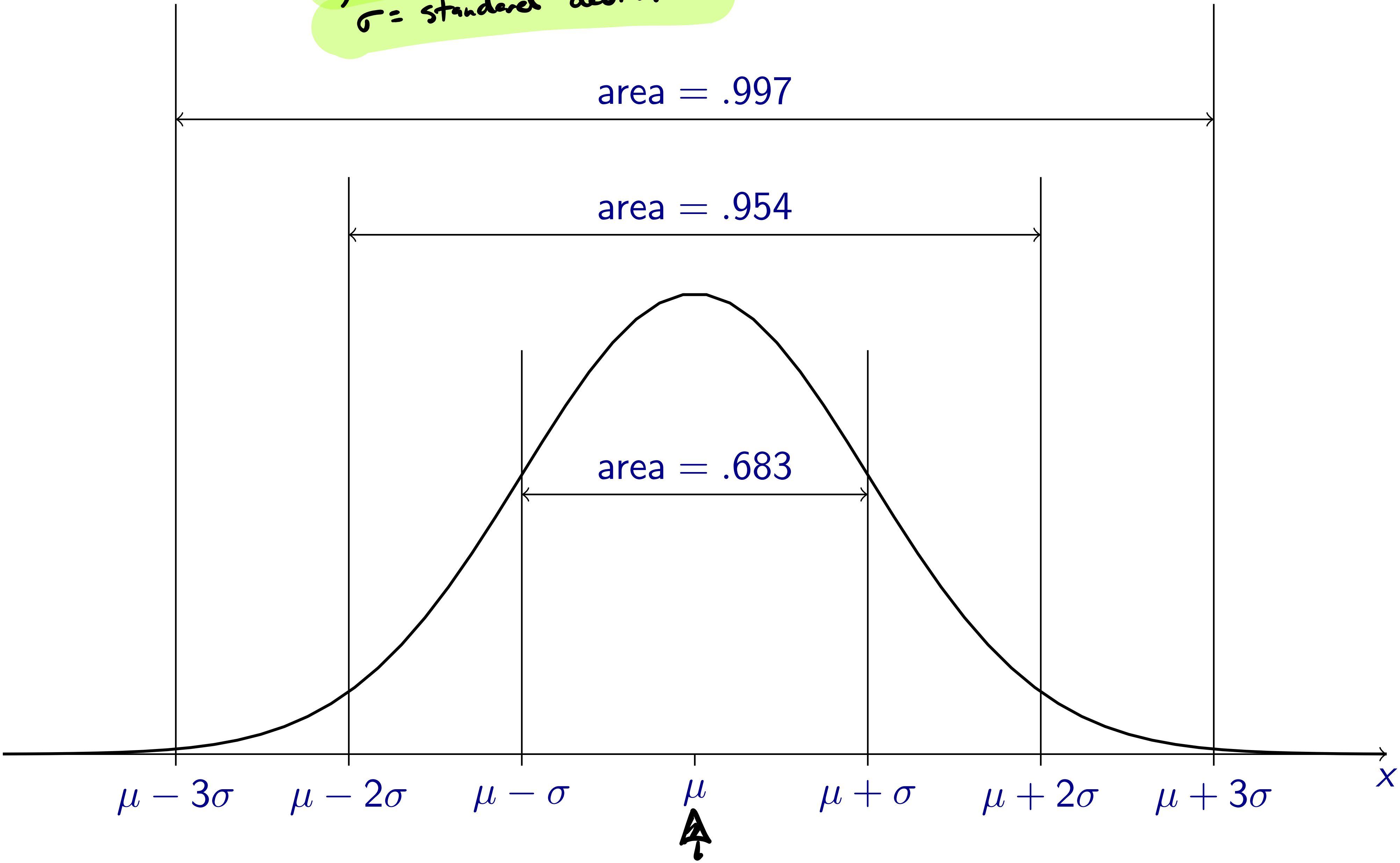
$\exp(z) = e^{z}$

$$\frac{1}{\sqrt{2\pi}} \frac{1}{\sigma} \exp\left[-\frac{1}{2} \frac{(x-\mu)^2}{\sigma^2} \right]$$



The pdf of the $\mathcal{N}(\mu, \sigma^2)$ distribution:

μ = mean
σ = standard deviation



Mean and variance of Normal distribution

If $X \sim \mathcal{N}(\mu, \sigma^2)$, then

- $\mathbb{E}X = \mu$
- $\text{Var}\, X = \sigma^2$

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

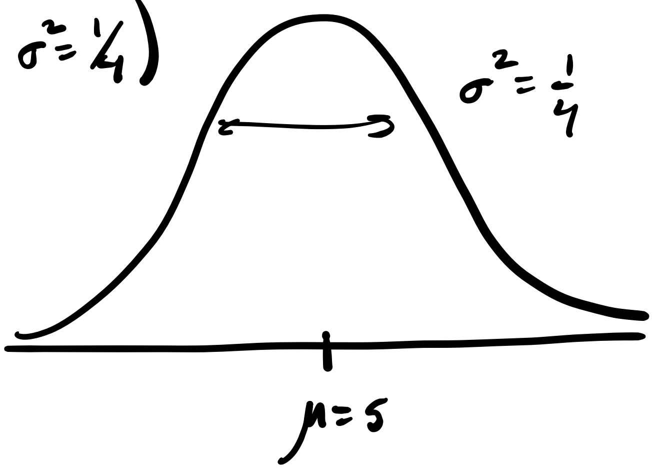
$$\text{Var}\, X = \int_{-\infty}^{\infty} (x-\mu)^2 f(x)\, dx = \cdots = \sigma^2$$

Exercise: Suppose growth in height (ft) of Loblolly pines from age three to five is $\mathcal{N}(\mu = 5, \sigma^2 = 1/4)$. Give the probability that the growth of a randomly selected Loblolly pine is

$X \sim \mathcal{N}(\mu = 5, \sigma^2 = 1/4)$

1. between 4.5 and 5.5 feet.
2. more than 7 feet.
3. less than 5.5 feet.
4. between 3.5 feet and 5.5 feet.

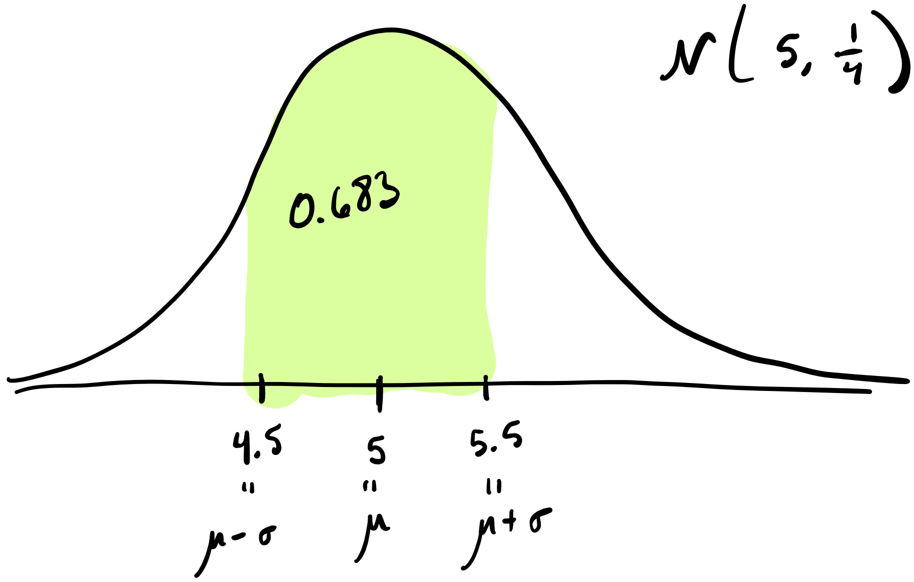
Use the picture on the previous slide.



① $N(5, \frac{1}{4})$

$\sigma^2 = \frac{1}{4}$

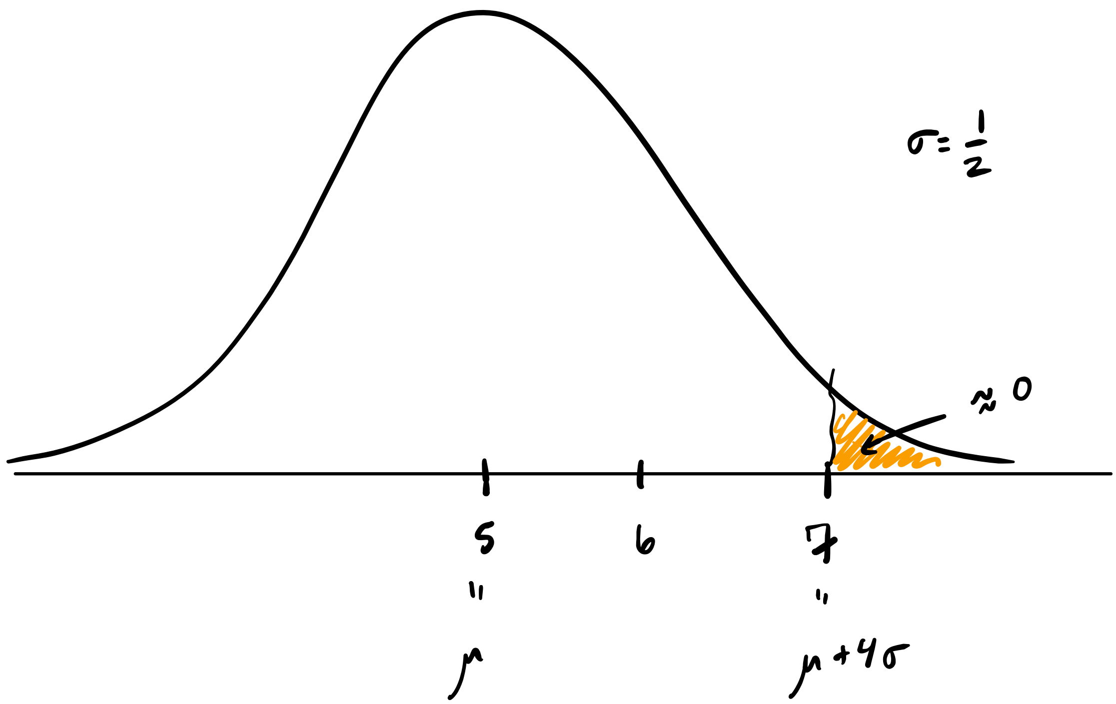
$\sigma = \frac{1}{2}$



$P(4.5 < X < 5.5) = 0.683$

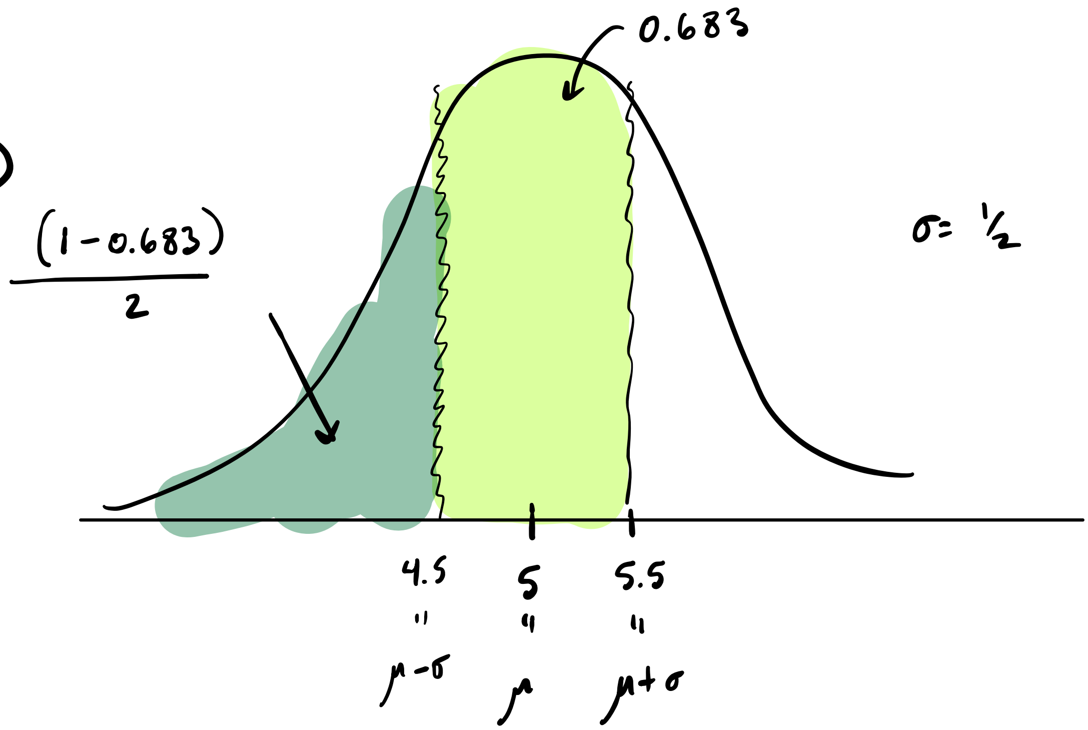
②

$\sigma = \frac{1}{2}$



$P(X > 7) \approx 0$

③

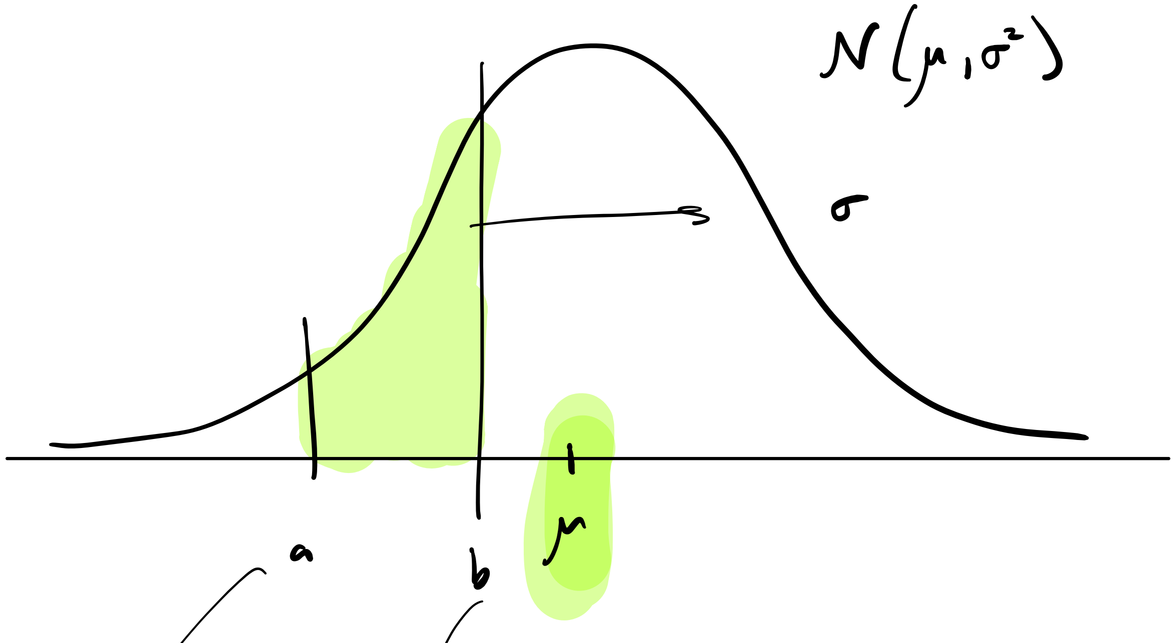


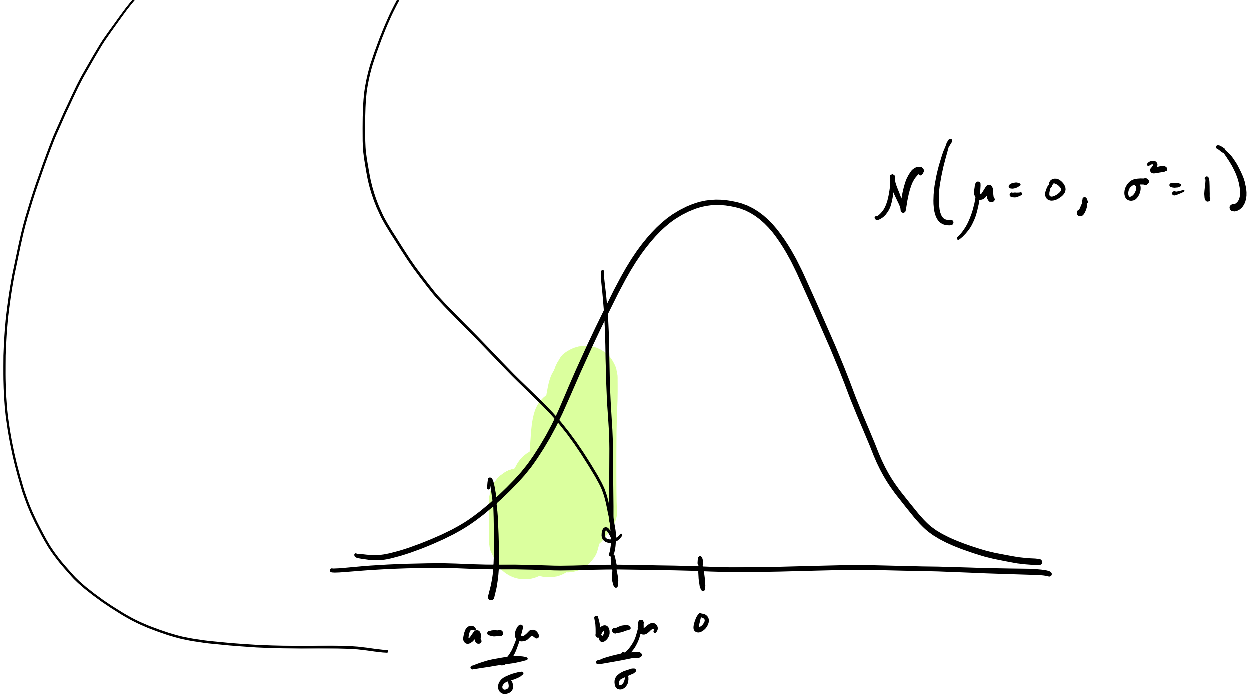
$$P(X < 5.5) = 0.683 + \frac{1}{2}(0.317) =$$

END EXAMI MATERIAL

Computing Normal probabilities

$$P(a \leq X \leq b) = \int_a^b \frac{1}{\sqrt{2\pi}} \frac{1}{\sigma} \exp\left(-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2\right) dx$$





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

Then

$$Z = \frac{X - \mu}{\sigma} = \text{\# standard deviations from the mean (signed)}$$

$$\sim N(0, 1) \quad (\mu = 0, \ \sigma^2 = 1)$$

Get probabilities for $X \sim \mathcal{N}(\mu, \sigma^2)$ like

$$P(a < X < b) = \int_a^b \frac{1}{\sqrt{2\pi}} \frac{1}{\sigma} \exp\left[-\frac{(x-\mu)^2}{2\sigma^2}\right] dx$$

Conversion to the Standard Normal distribution

If $X \sim \mathcal{N}(\mu, \sigma^2)$, then

$$Z = \frac{X - \mu}{\sigma} \sim \mathcal{N}(0, 1).$$

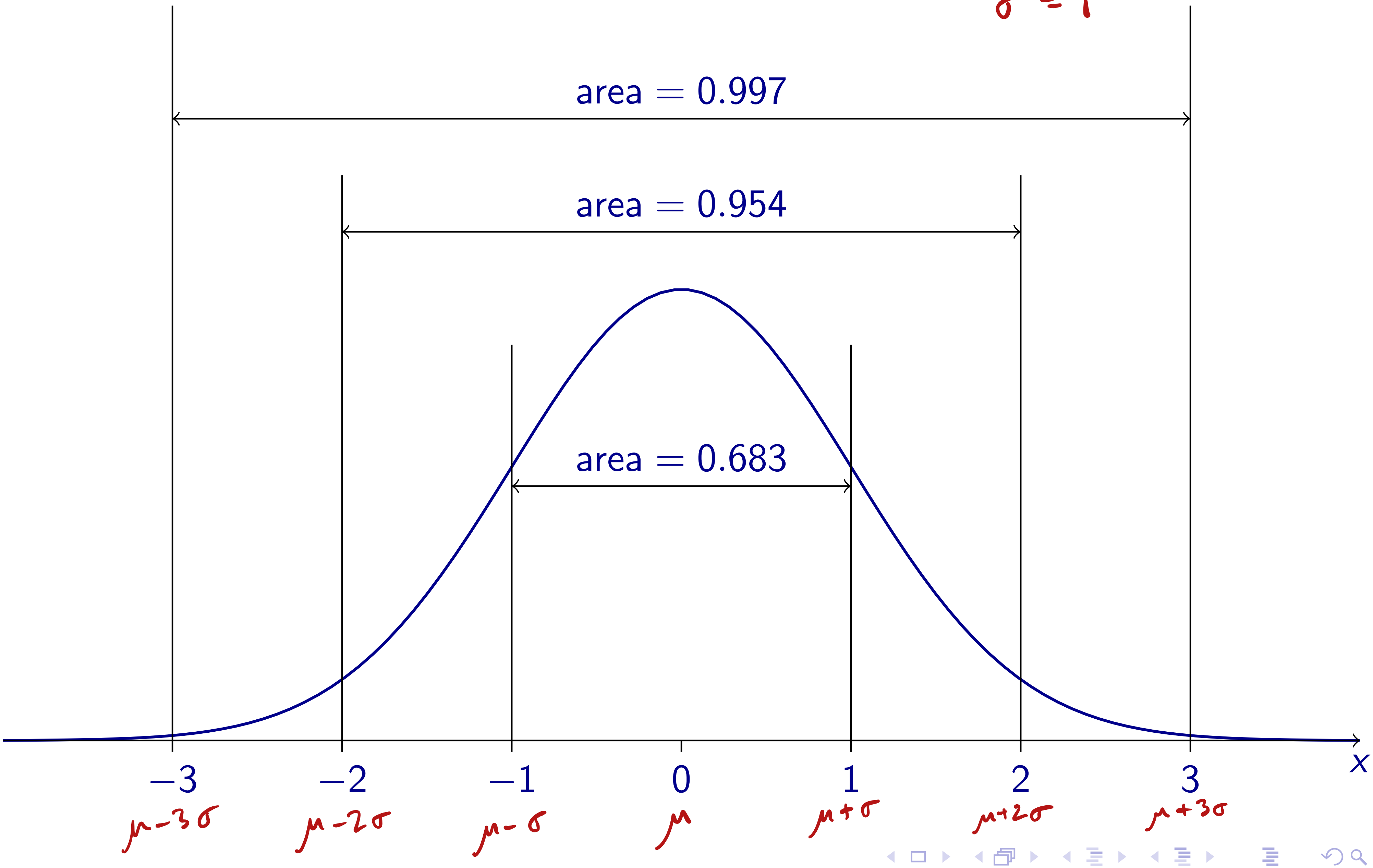
The $\mathcal{N}(0, 1)$ dist. is called the *Standard Normal distribution* and its pdf is

$$\phi(z) = \frac{1}{\sqrt{2\pi}} e^{-z^2/2}.$$

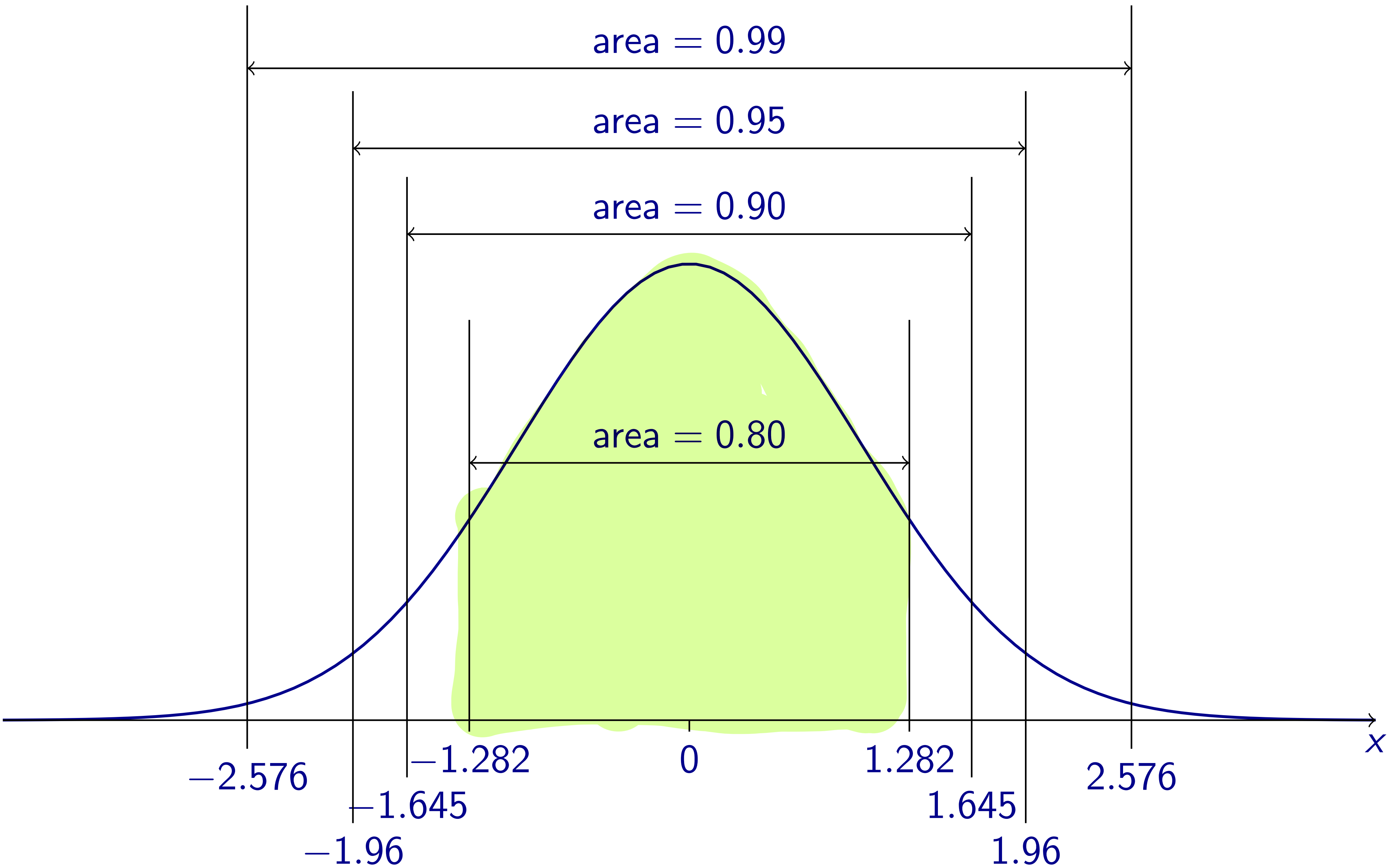
Can look up integrals over this pdf in a table.

The pdf of the $\mathcal{N}(0,1)$ distribution:

$\mu = 0$
$\sigma^2 = 1$



The pdf of the $\mathcal{N}(0,1)$ distribution:



If $X \sim \mathcal{N}(\mu, \sigma^2)$, we can find $P(a < X < b)$ in two steps:

1. Transform a and b to the Z-world (# of standard deviations world):

$$a \mapsto \frac{a-\mu}{\sigma} \quad \text{and} \quad b \mapsto \frac{b-\mu}{\sigma},$$

2. Find

$$P\left(\frac{a-\mu}{\sigma} < Z < \frac{b-\mu}{\sigma}\right)$$

by using a "Z-table"—a table of Standard Normal probabilities.

$X \sim N(\mu = 5, \sigma^2 = \frac{1}{4})$

Exercise: Suppose growth in height (ft) of Loblolly pines from age three to five is $\mathcal{N}(5, 1/4)$. Give the probability that the growth of a randomly selected Loblolly pine is

1. between 5.25 and 6.25 feet.
2. more than 7.8 feet.
3. less than 5.31 feet.
4. between 4.1 feet and 5.2 feet.

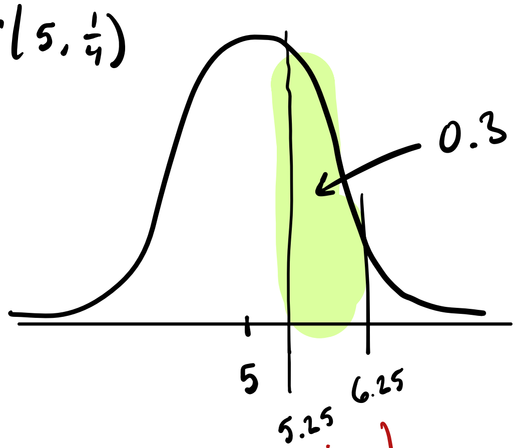
Find the probabilities in the "Z-world" using a Z-table.

1 $P(5.25 < X < 6.25)$, $X \sim N(5, \frac{1}{4})$

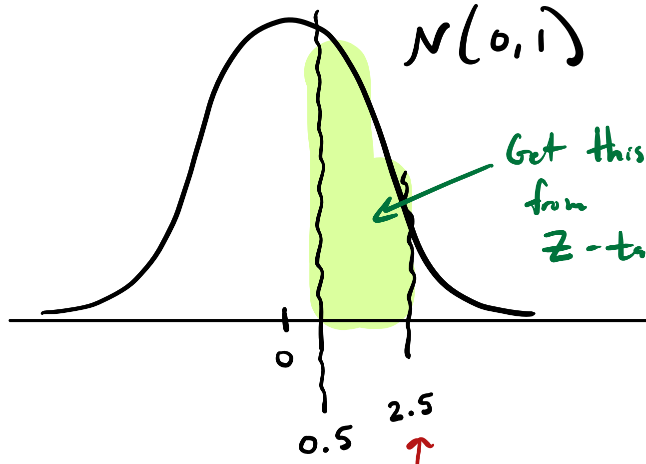
$\sigma = \frac{1}{2}$

$Z = \frac{X - \mu}{\sigma}$

$X \sim N(5, \frac{1}{4})$

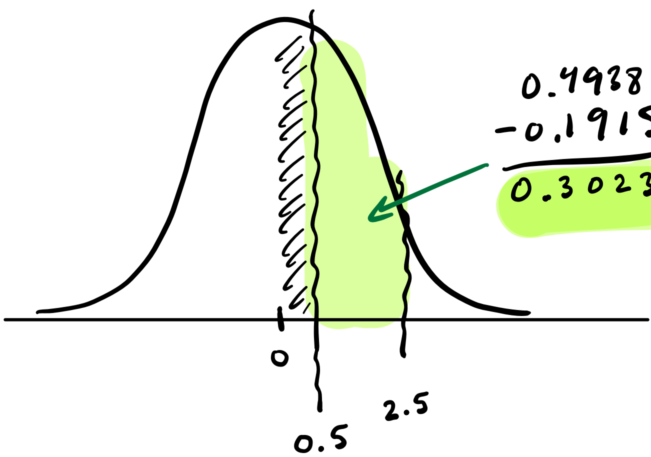


$N(0,1)$

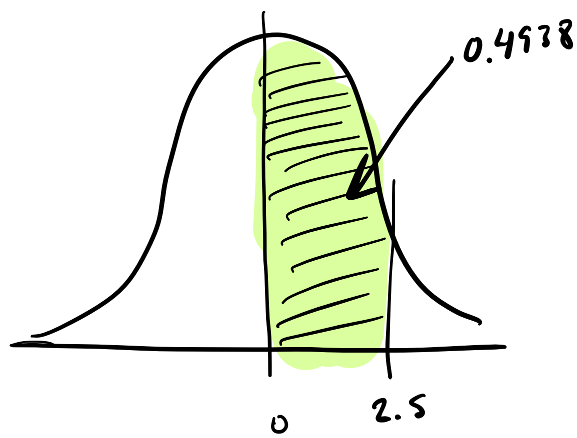


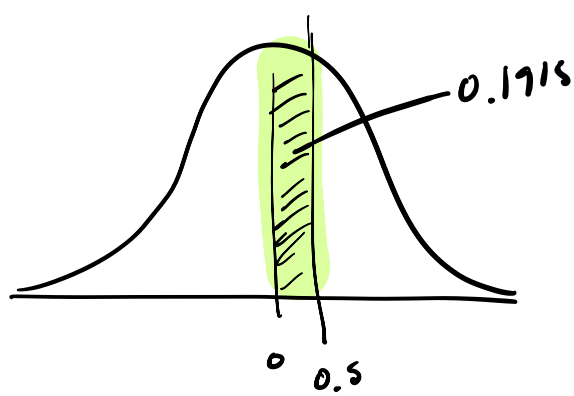
$$\frac{5.25-5}{0.5} = \frac{0.25}{0.5} = 0.5$$

$$\frac{6.25-5}{0.5} = \frac{1.25}{0.5} = 2.5$$



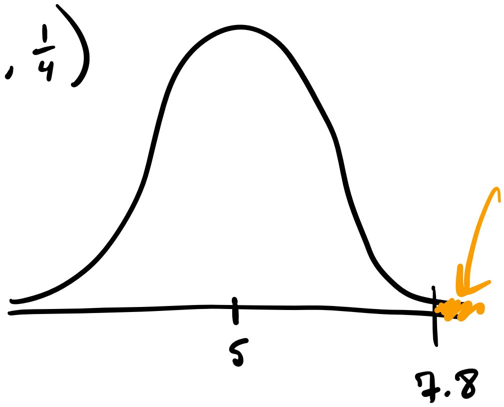
$$\begin{array}{r} 0.4938 \\ -0.1915 \\ \hline \boxed{0.3023} \end{array}$$

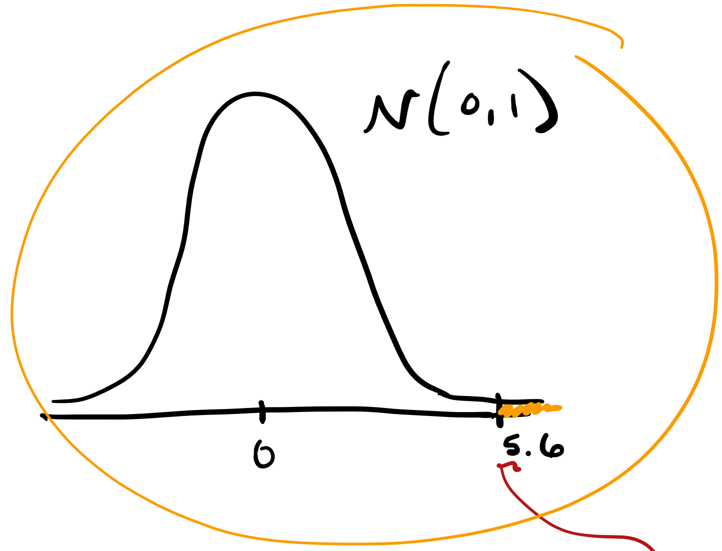




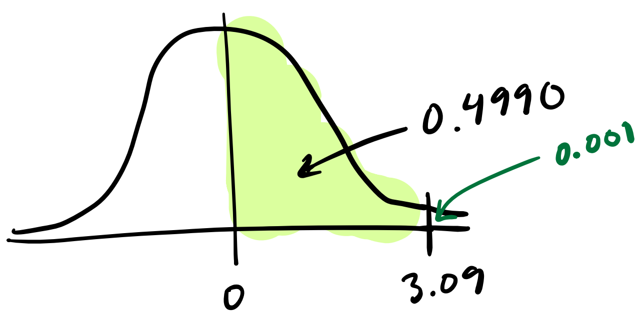
② $P(X > 7.8) \approx 0$ ← approx

$N(5, \frac{1}{4})$





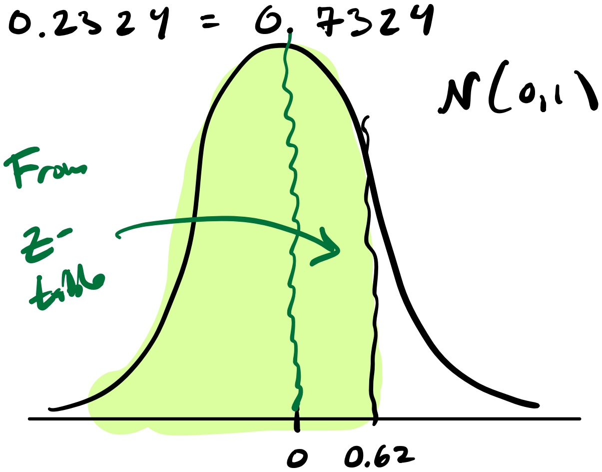
$$\frac{7.8-5}{0.5} = \frac{2.8}{0.5} = 4 + 1.6 = 5.6$$



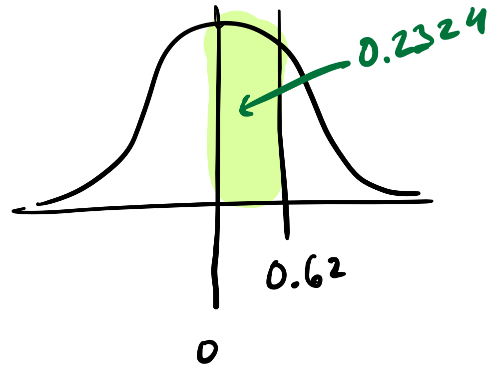
largest z in z-table

③ $P(X < 5.31) = 0.5 + 0.2324 = 0.7324$

$N(5, \frac{1}{4})$

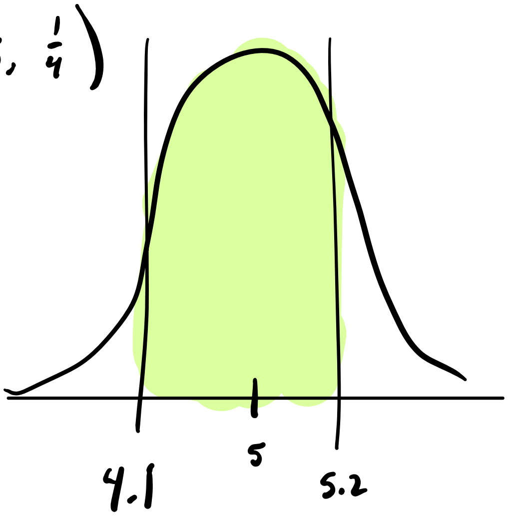


$$\frac{5.31-5}{0.5} = \frac{0.31}{0.5} = 0.62$$

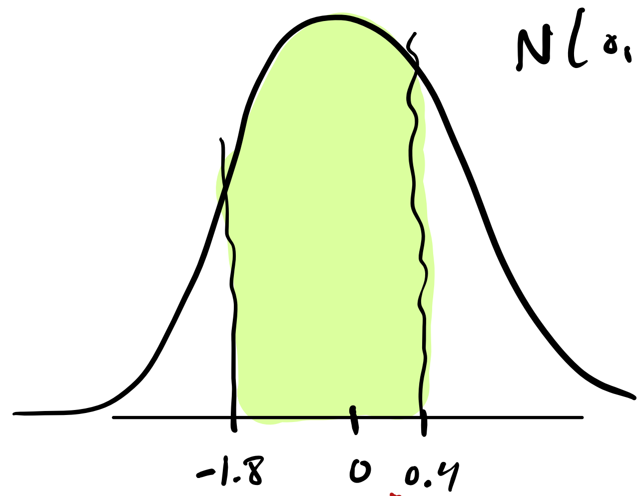


④ $P(4.1 < x < 5.2)$

$N(5, \frac{1}{4})$



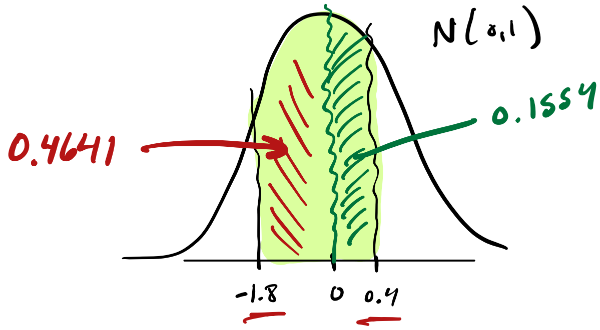
$N(0,1)$



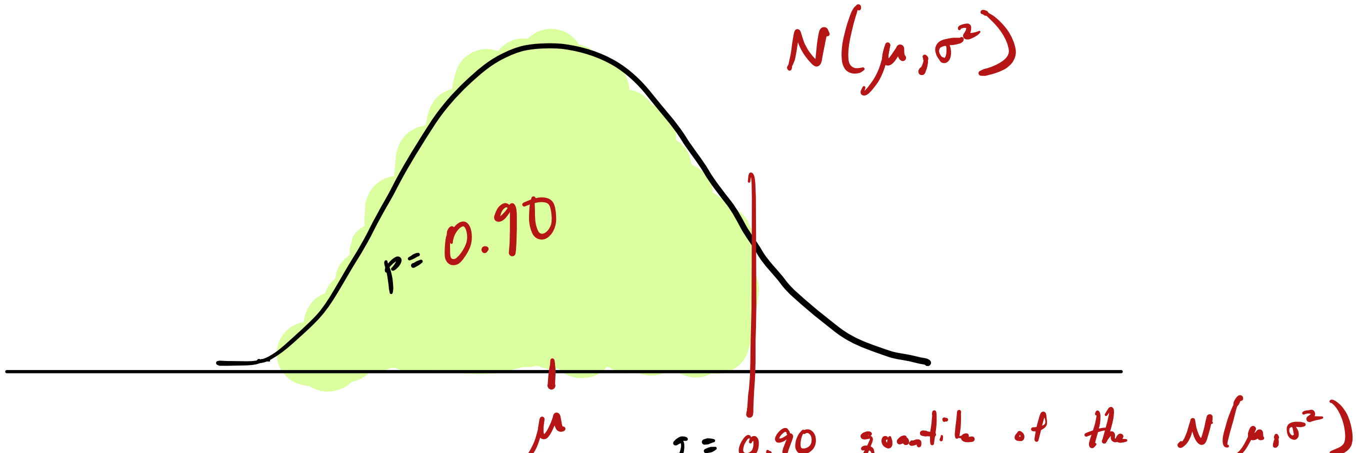
$$\frac{5.2-5}{0.5} = \frac{0.2}{0.5} = 0.4$$

$\sigma = 0.5$

$$\frac{4.1-5}{0.5} = \frac{-0.9}{0.5} = -1.8$$



$$P(4.1 < X < 5.2) = \begin{array}{r} 0.4641 \\ +\ 0.1554 \\ \hline 0.6195 \end{array} = 0.6195.$$



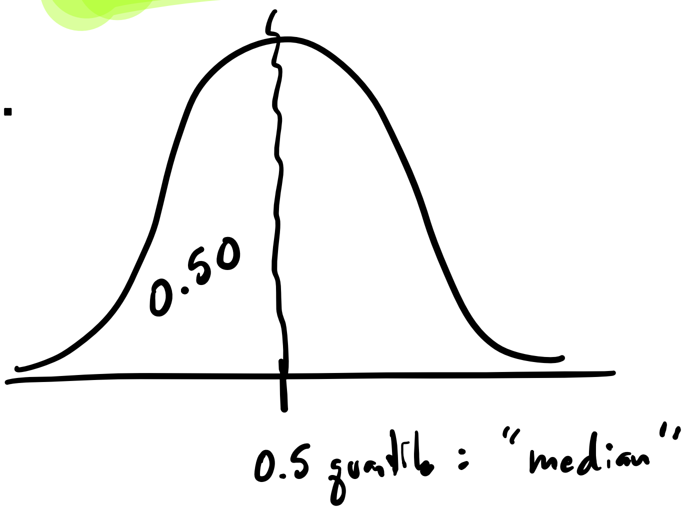
Quantiles of a continuous random variable

For a continuous rv X with a strictly increasing cdf, the p *quantile* of X is the value q which satisfies

$$P(X \leq q) = p.$$

A quantile is like a percentile, but not expressed as a percentage.

Draw pictures.



If $X \sim \mathcal{N}(\mu, \sigma^2)$, we can find q such that $P(X \leq q) = p$ in two steps:

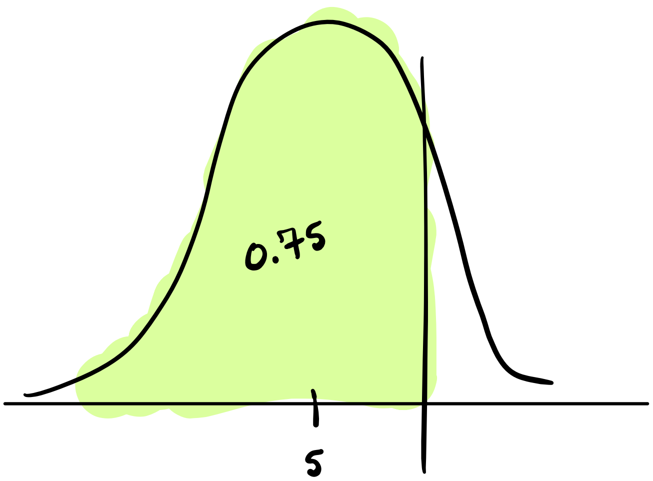
1. Find q_Z such that $P(Z < q_Z) = \theta$ using a "Z table".
2. Get the corresponding quantile in the X world as $q = \mu + \sigma q_Z$

Exercise: Suppose growth in height (ft) of Loblolly pines from age three to five is $\mathcal{N}(5, 1/4)$. Let X denote the height of a randomly selected Loblolly pine and find

1. the 75%-tile of growth. a.k.a. the 0.75 quantile
2. the median of the growths, i.e. the 50%-tile of X.
3. an interval, centered at the mean, within which X lies with probability 0.50.

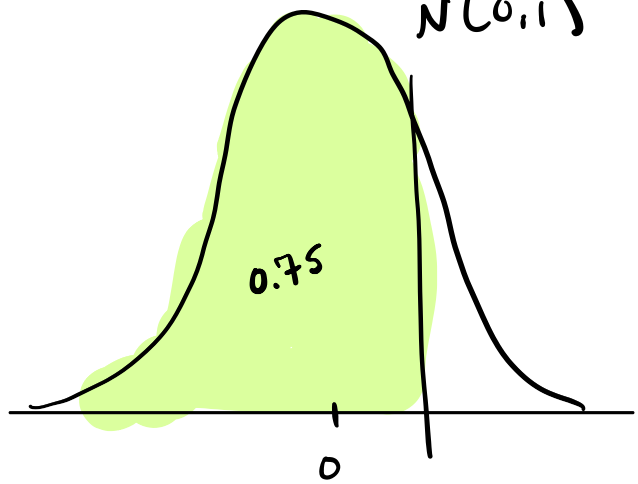
$$Z = \frac{X - \mu}{\sigma} \iff X = \mu + \sigma Z$$

1. $X \sim N(5, \frac{1}{4})$



$q = 5.3375$

$N(0,1)$



$q_z = 0.675$

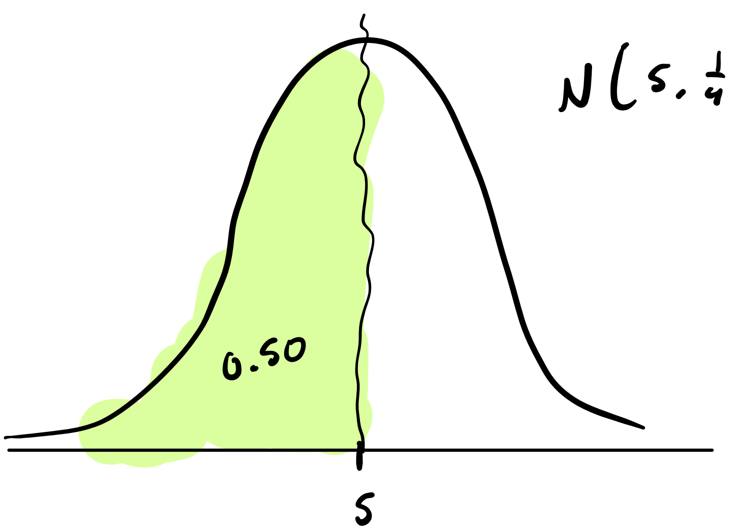
Find this using z-table!

$= 5 + 0.5(0.675)$

$= 5 + 0.3375$

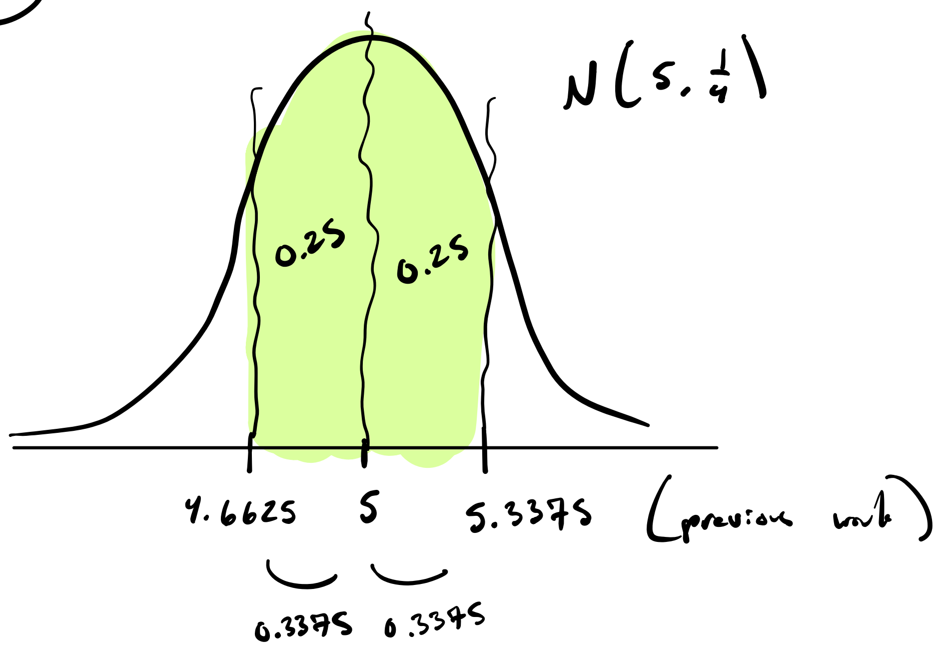
$= 5.3375$

2. Median $= 5$



$N(5, \frac{1}{4})$

3



Exercise: Redo some exercises with `pnorm` and `qnorm` functions in R.

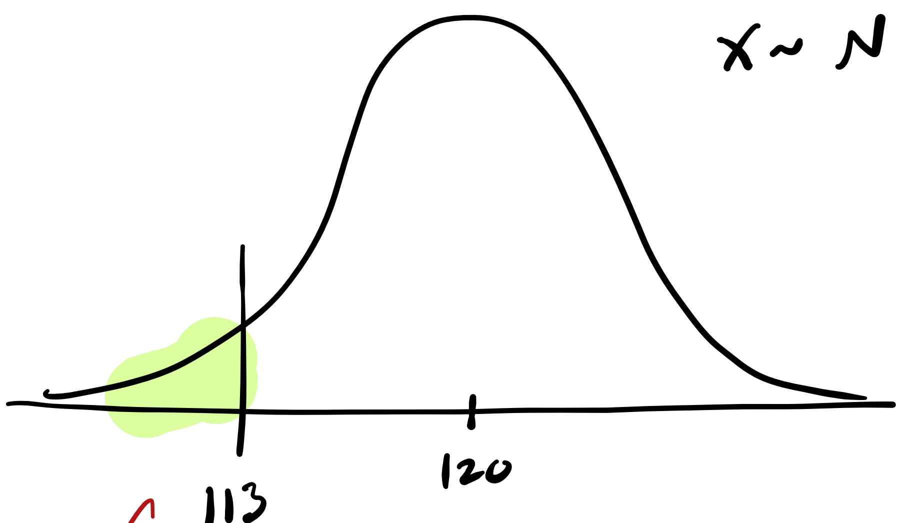
Exercise: You sell jars of baby food labelled as weighing $\boxed{4\text{oz} \approx 113\text{g.}}$ Suppose your process results in jar weights with the $\mathcal{N}(120, 4^2)$ distribution and that you will be fined if more than 2% of your jars weigh less than 113g.

1. What proportion of your jars weigh less than 113g?
2. To what must you increase μ to avoid being fined?
3. Keeping $\mu = 120$g, to what must you reduce σ to avoid being fined?

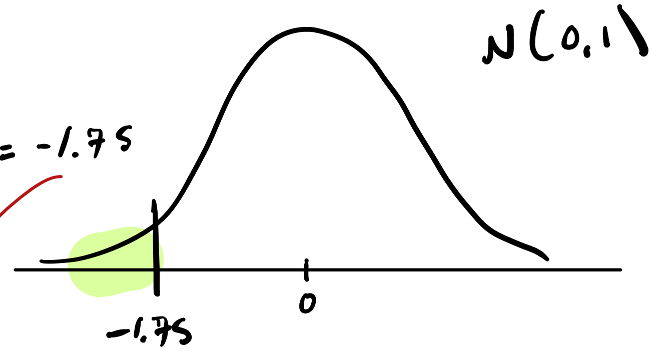
$X \sim N(120, 4^2)$

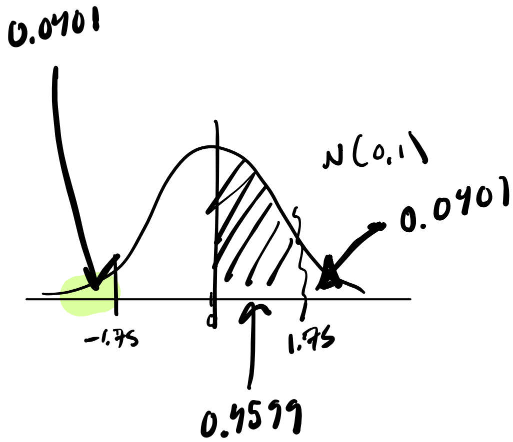
$\sigma = 4$

1



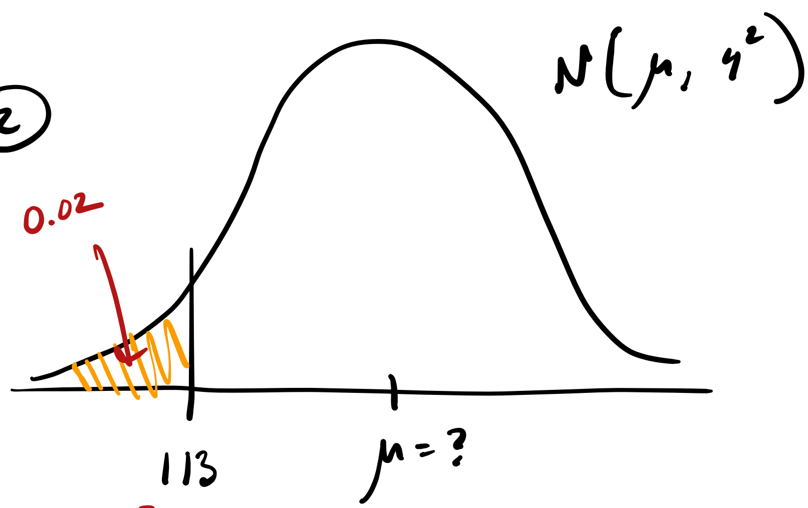
$$\frac{113-120}{4} = \frac{-7}{4} = -1.75$$





$P(X < 113) = 0.0401.$

②

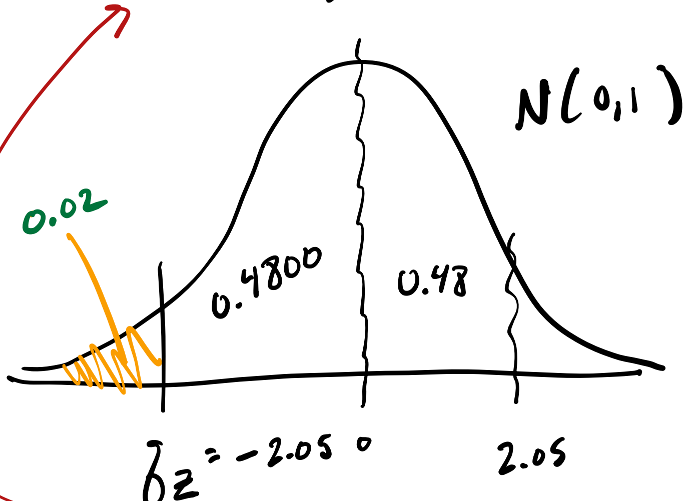


$\sigma = 4$

$$Z = \frac{X-\mu}{\sigma}$$

$\iff$

$$X = \mu + \sigma Z$$



$113 = \mu + 4(-2.05)$ ← solve for μ

$\Leftrightarrow$

$113 + 4(2.05) = 113 + 8.2 = 121.2$

set $\mu = 121.2$

unknown

③ $113 = 120 + \sigma(-2.05)$

$\Leftrightarrow$ $\sigma = \dfrac{113 - 120}{-2.05} = \dfrac{120 - 113}{2.05} = 3.41$

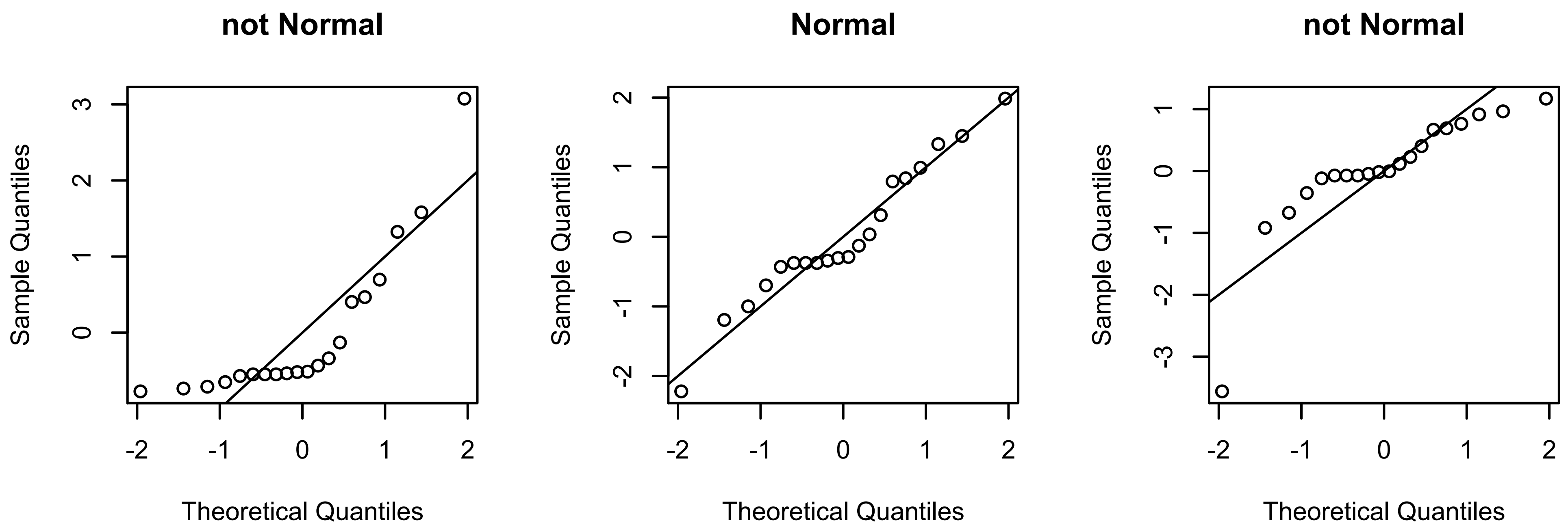
Do my data come from a Normal distribution?

Example: These are the commute times (sec) to class of a sample of students.

1832	1440	1620	1362	577	934	928	998	1062	900
1380	913	654	878	172	773	1171	1574	900	900

Check with a Q-Q plot whether the data quantiles match those of a Normal distribution.

Some Q-Q plots (middle is for the commute time data):



Sum of independent Normal random variables

If $X_1 \sim \mathcal{N}(\mu_1, \sigma_1^2), \dots, X_n \sim \mathcal{N}(\mu_1, \sigma_n^2)$ are independent random variables, then

$$\sum_{i=1}^n X_i \sim \mathcal{N}\left(\sum_{i=1}^n \mu_i, \sum_{i=1}^n \sigma_i^2\right).$$

In the above, *independent* means that the values of the rvs don't affect one other.

Exercise: Consider boxes containing 25 jars of baby food (from previous).

1. What is the expected weight of the boxes?
2. What is the standard deviation of the box weights?
3. Give the probability that the box weighs less than 2,975 grams.