STAT 515 fa 2023 Lec 07 slides

The Normal distribution

"Gaussian"

Karl B. Gregory

University of South Carolina

$e^z = \exp(z)$

$e^{-\frac{(x-\mu)^2}{2\sigma^2}} = \exp\left[-\frac{(x-\mu)^2}{2\sigma^2}\right]$

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

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

μ = mu
σ = sigma

has the *Normal distribution* with mean μ and variance σ^2

Carl Friedrich Gauß

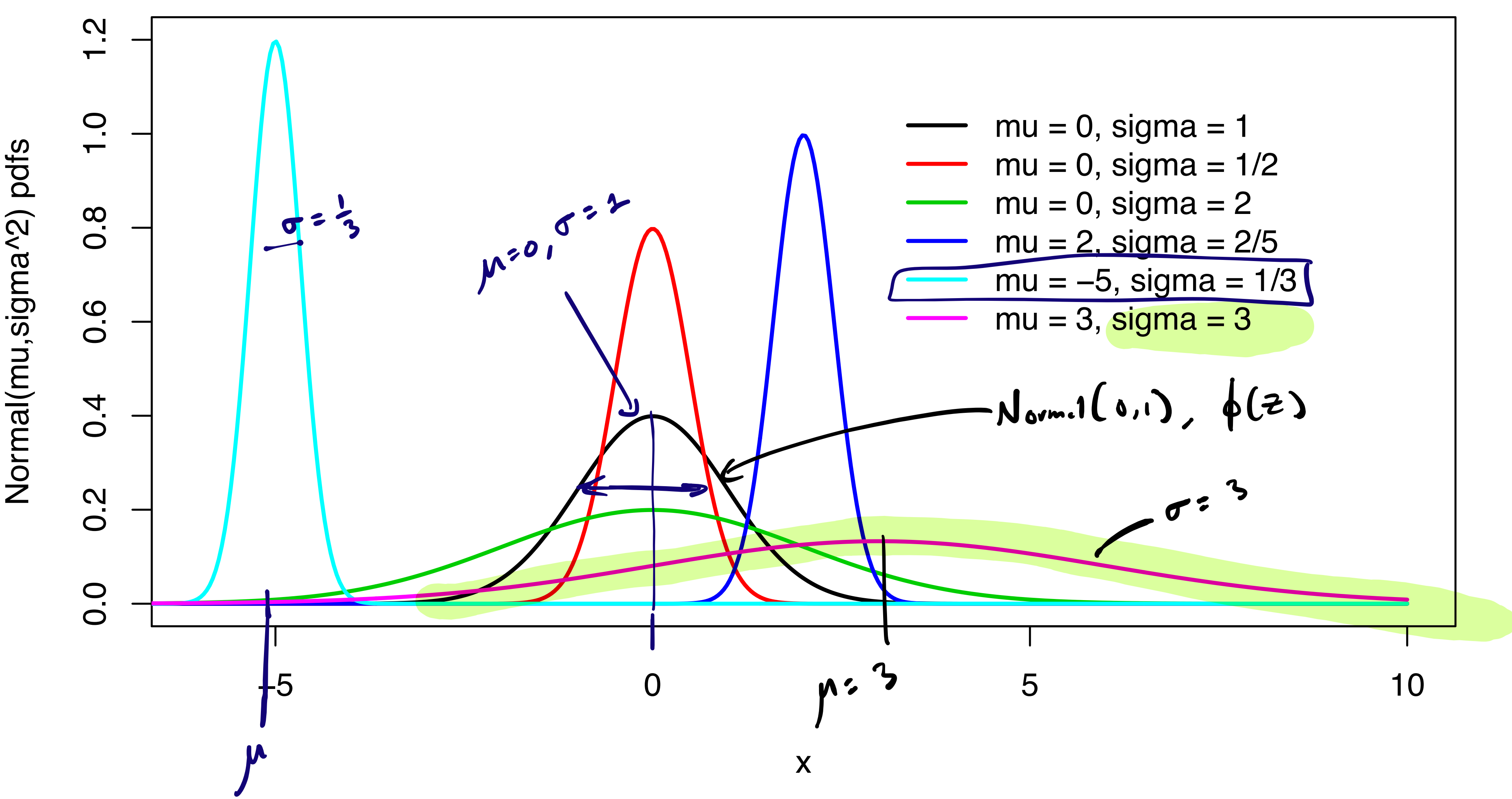
$\mathbb{E}X = \int_{-\infty}^{\infty} x f(x)\,dx = \mu$

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

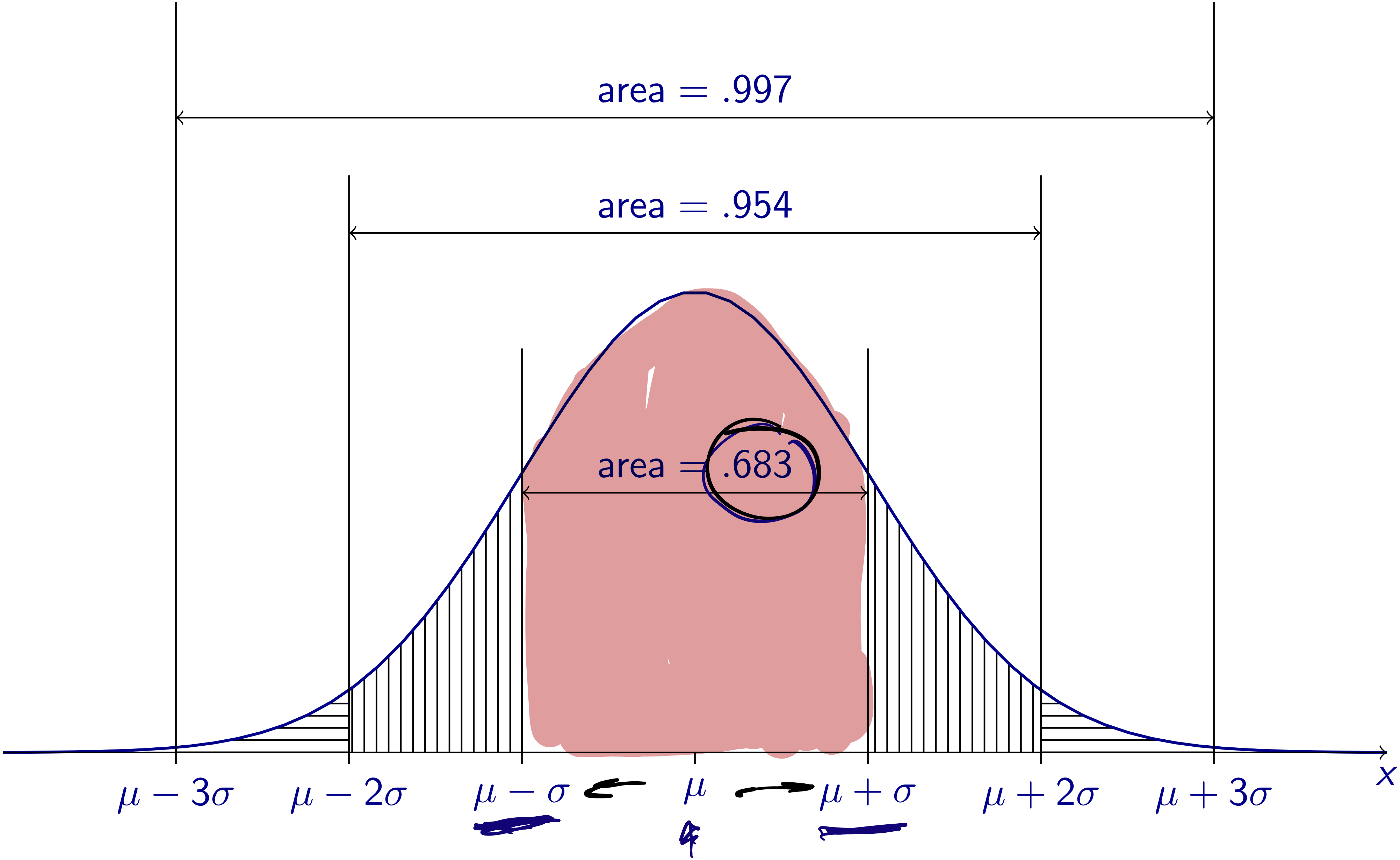
We write $X \sim \text{Normal}(\mu, \sigma^2)$.

pdfs of several Normal distributions

σ^2 = Variance , σ = standard deviation



The pdf of the Normal(μ, σ^2) distribution:



Mean and variance of Normal distribution

If $X \sim \text{Normal}(\mu, \sigma^2)$, then

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

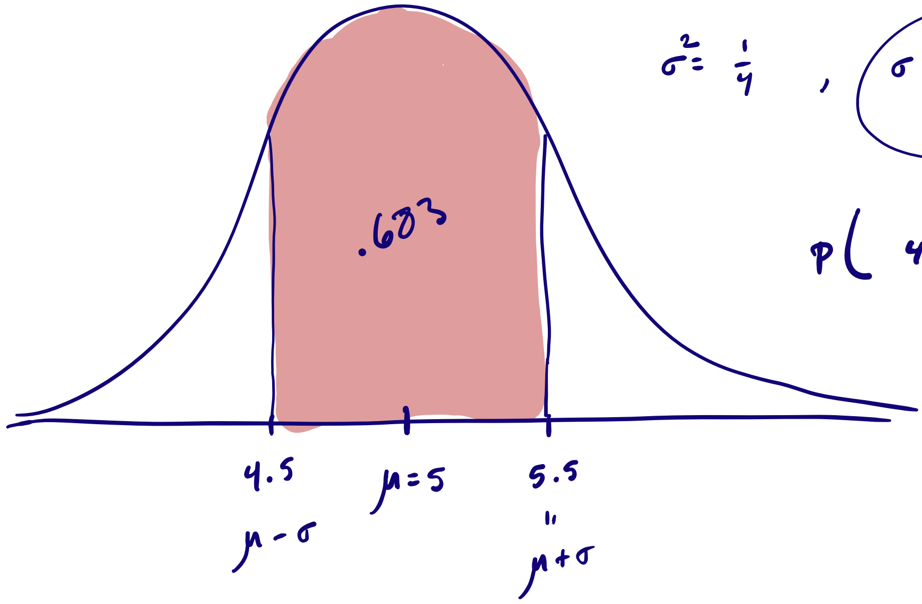
Exercise: Suppose growth in height (ft) of Loblolly pines from age three to five is $\text{Normal}(\mu = 5, \sigma^2 = 1/4)$. Give the probability that the growth of a randomly selected Loblolly pine is

1. between 4.5 and 5.5 feet. 0.683
2. more than 7 feet. very small.
3. less than 5.5 feet.
4. between 3.5 feet and 5.5 feet.

Use the picture on the previous slide.

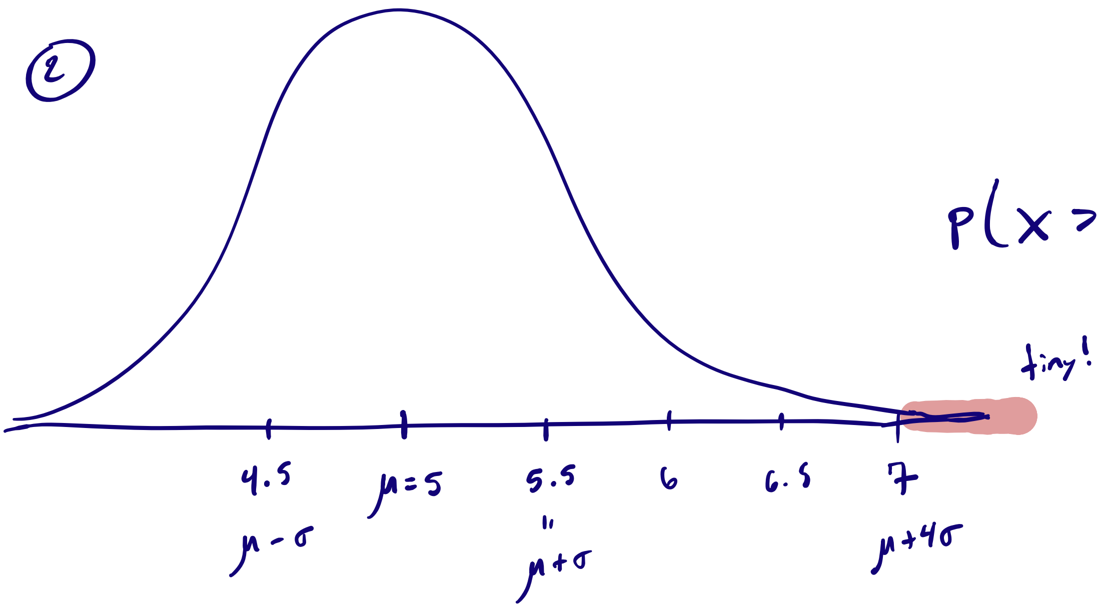
①

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



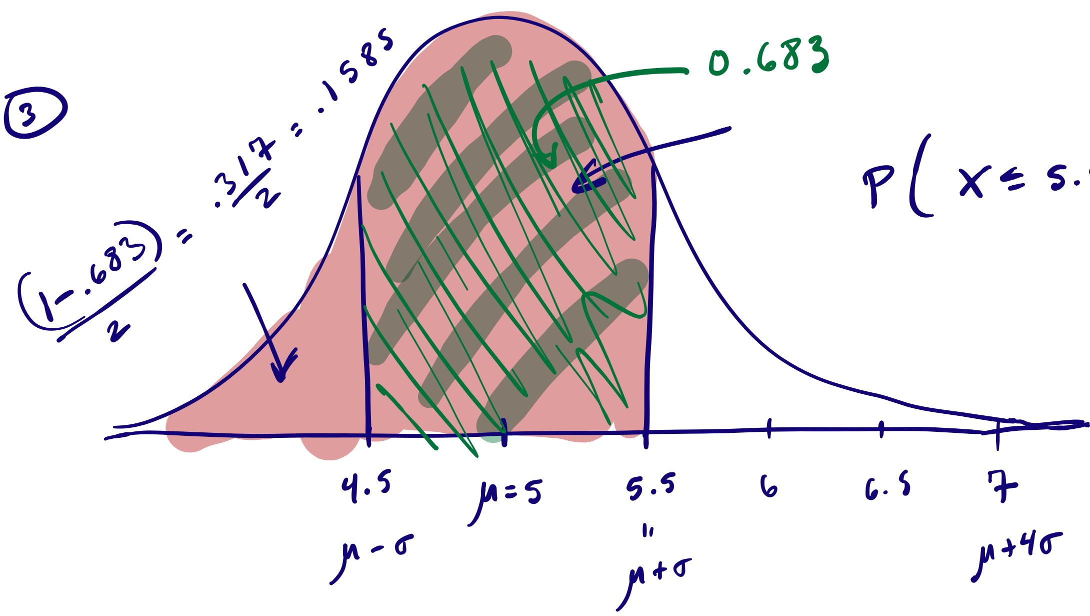
$P(4.5 < x < 5.5)$

②



$P(x > 7)$

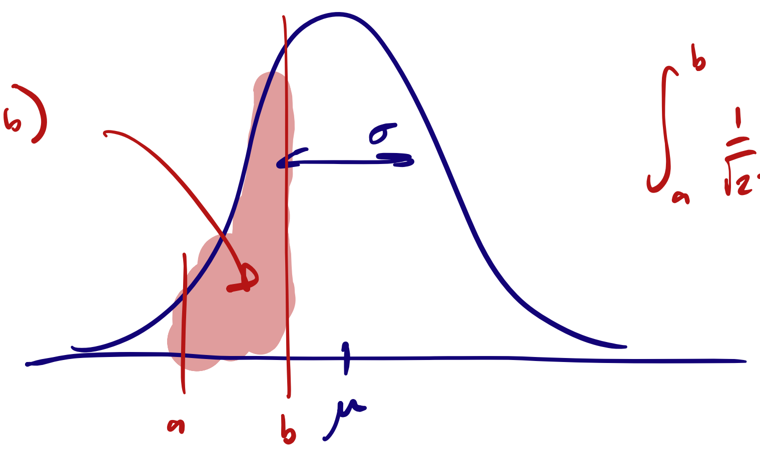
③



$P(x \le 5.5) = .683$

$$\begin{array}{r} .683 \\ .1585 \\ \hline 0.8415 \end{array}$$

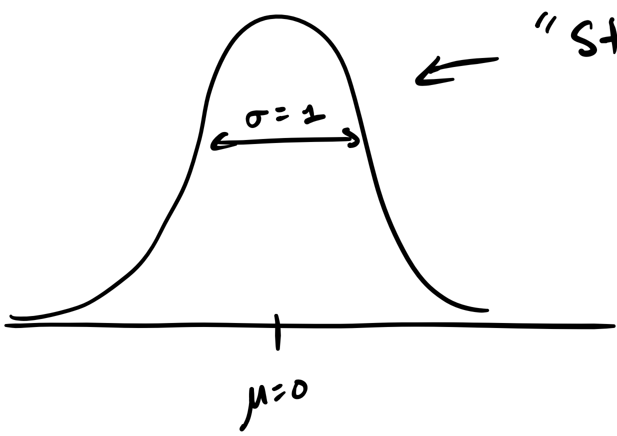
$P(a \le x \le b)$



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

↑ Cannot compute, so transform to

"Standard Normal distribution".



Get probabilities for $X \sim \text{Normal}(\mu, \sigma^2)$ 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 $X \sim \text{Normal}(\mu, \sigma^2)$, then

std. deviations from μ.

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

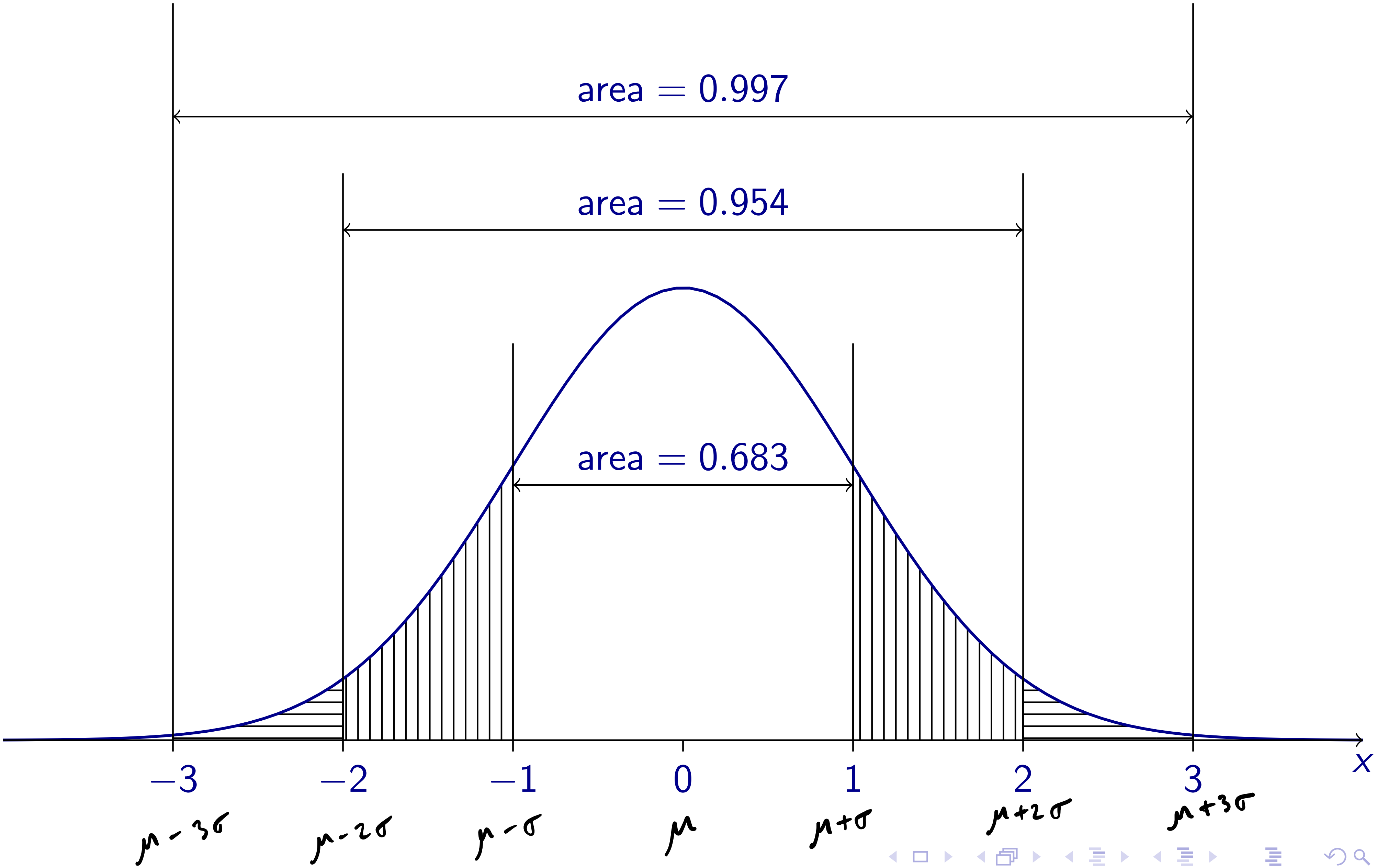
The $\text{Normal}(0,1)$ dist. is called the *Standard Normal distribution* and its pdf is

$\mu = 0$ $\quad \sigma^2 = 1$ $\quad$ "phi"

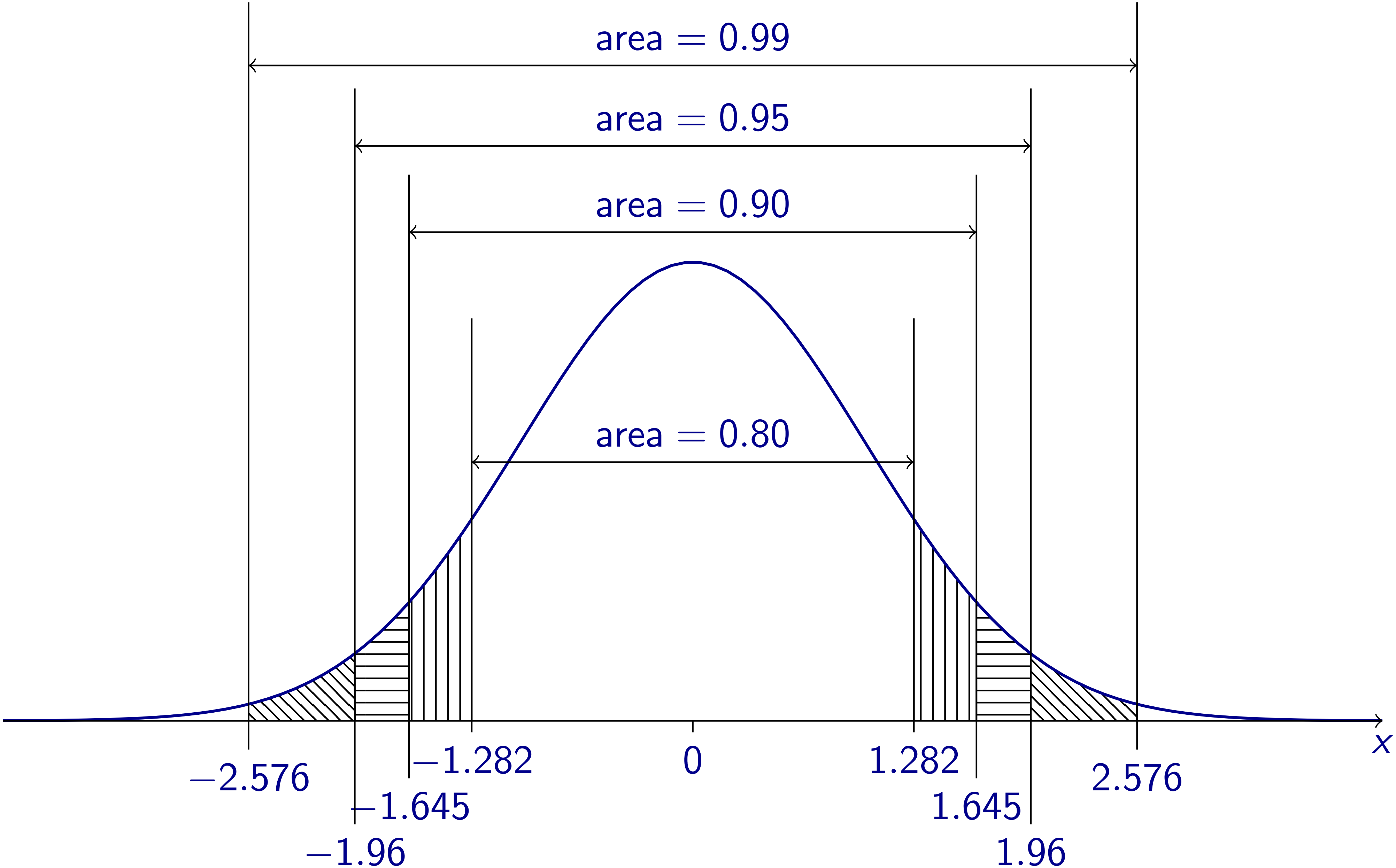
$$\phi(z) = \frac{1}{\sqrt{2\pi}} e^{-z^2/2}.$$

Can look up integrals over this pdf in a table.

The pdf of the Normal(0, 1) distribution:



The pdf of the Normal$(0, 1)$ distribution:



If $X \sim \text{Normal}(\mu, \sigma^2)$, 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

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

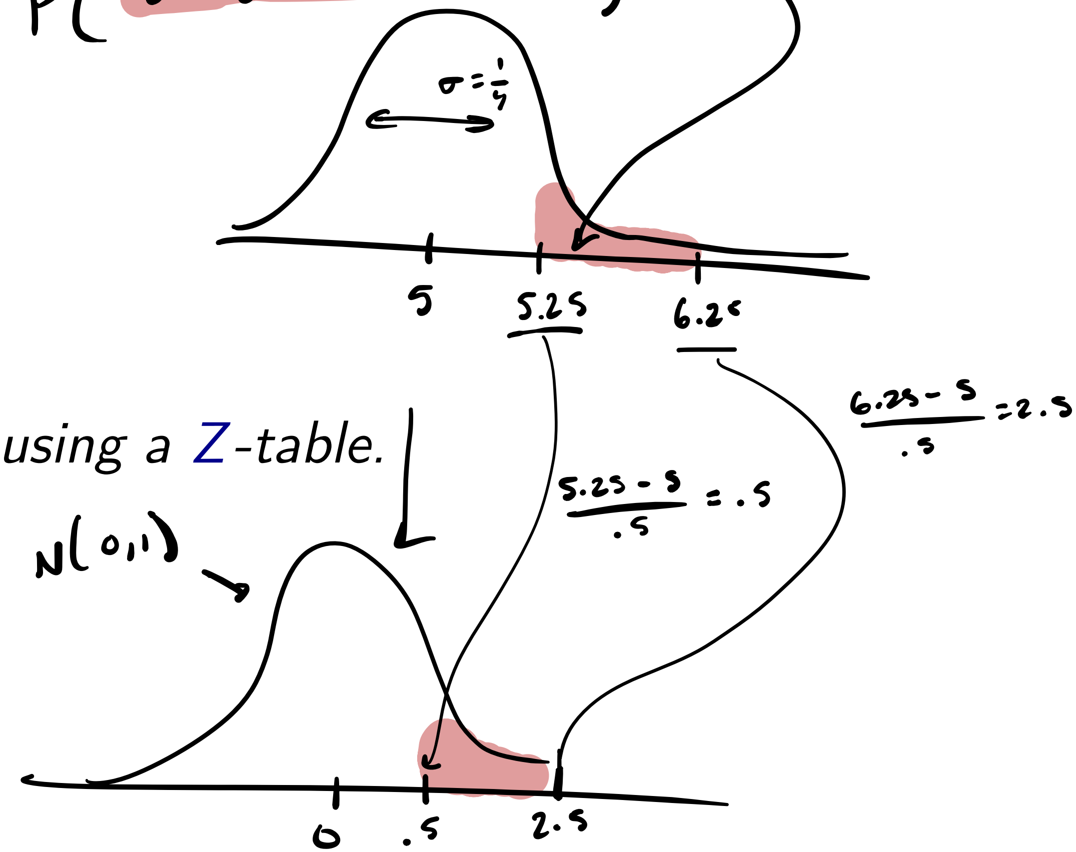
by using a "Z-table"—a table of Standard Normal probabilities.

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

Exercise: Suppose growth in height (ft) of Loblolly pines from age three to five is Normal($\mu = 5, \sigma^2 = 1/4$). 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 5.25 feet.
4. between 4.1 feet and 5.2 feet.

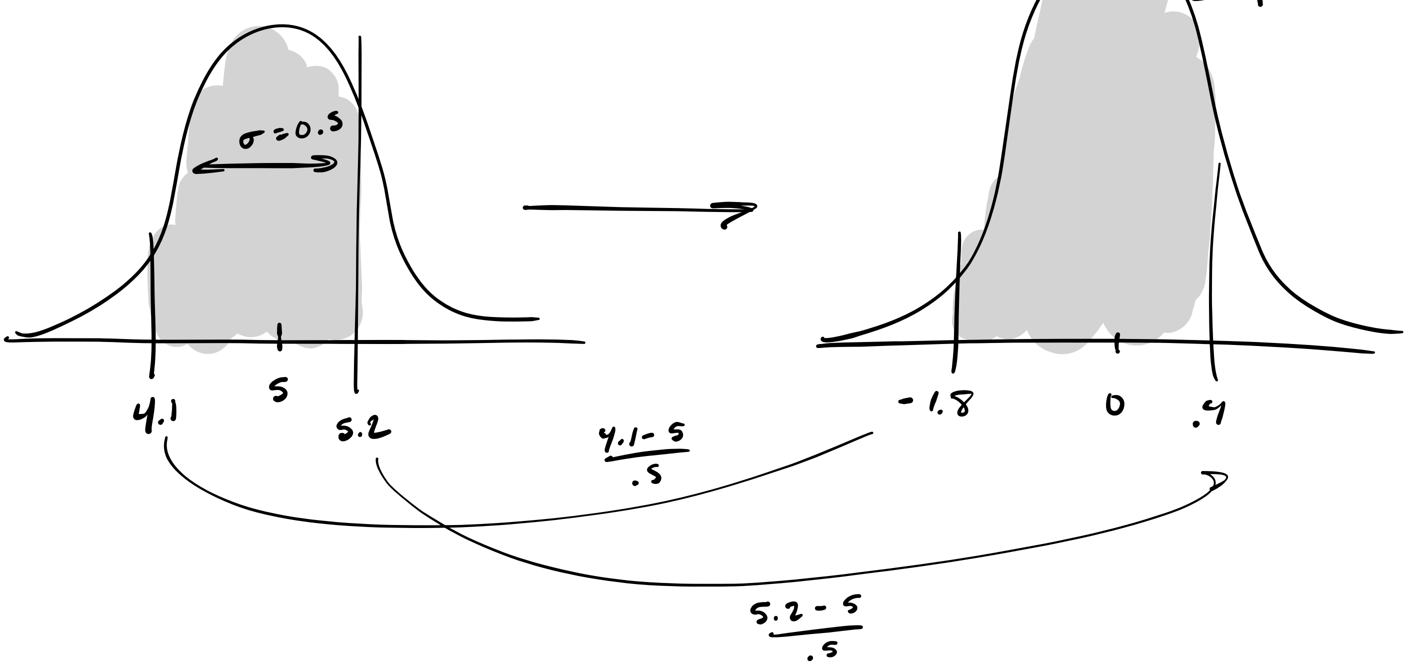
(1) $P(5.25 < X < 6.25)$

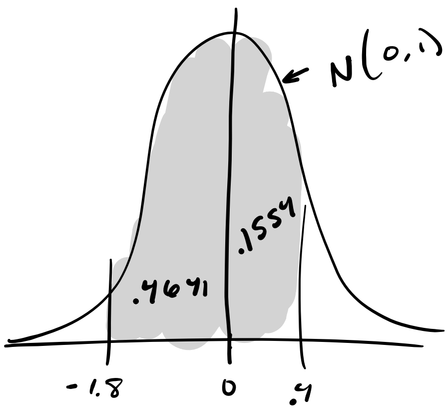


Find the probabilities in the "Z-world" using a Z-table.

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

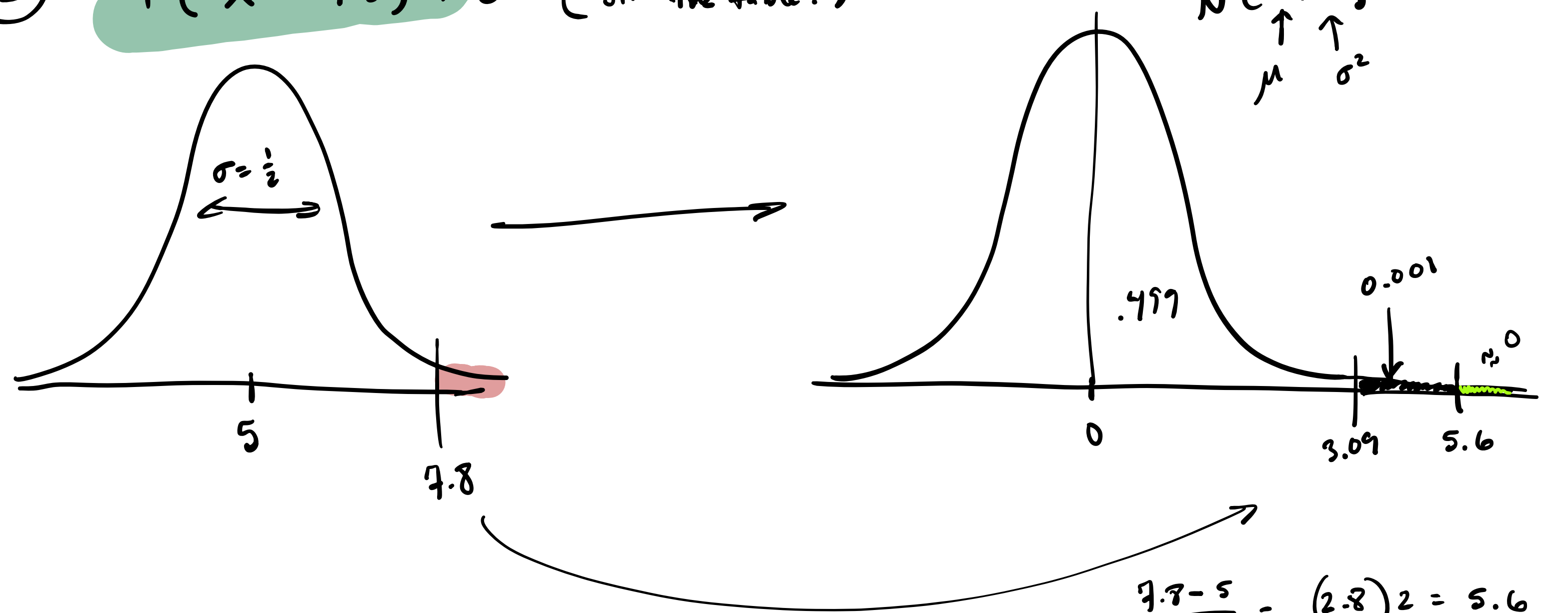
④ $P(4.1 < x < 5.2)$





$P(4.1 < x < 5.2) = .4641 + .1554 = .6195$

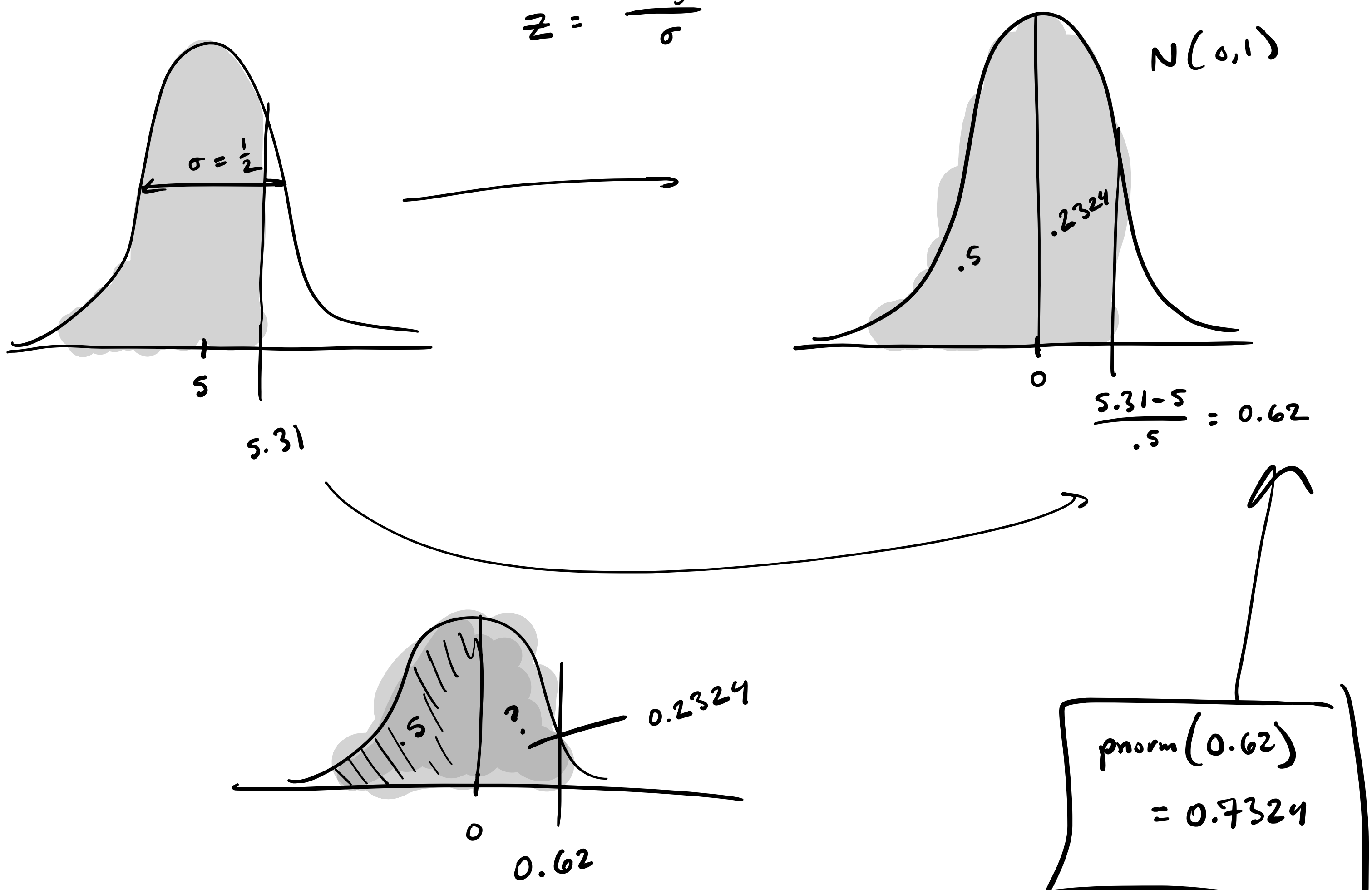
② $P(X > 7.8) \approx 0$ (off the table!)



$$\frac{7.8-5}{.5} = (2.8)2 = 5.6$$

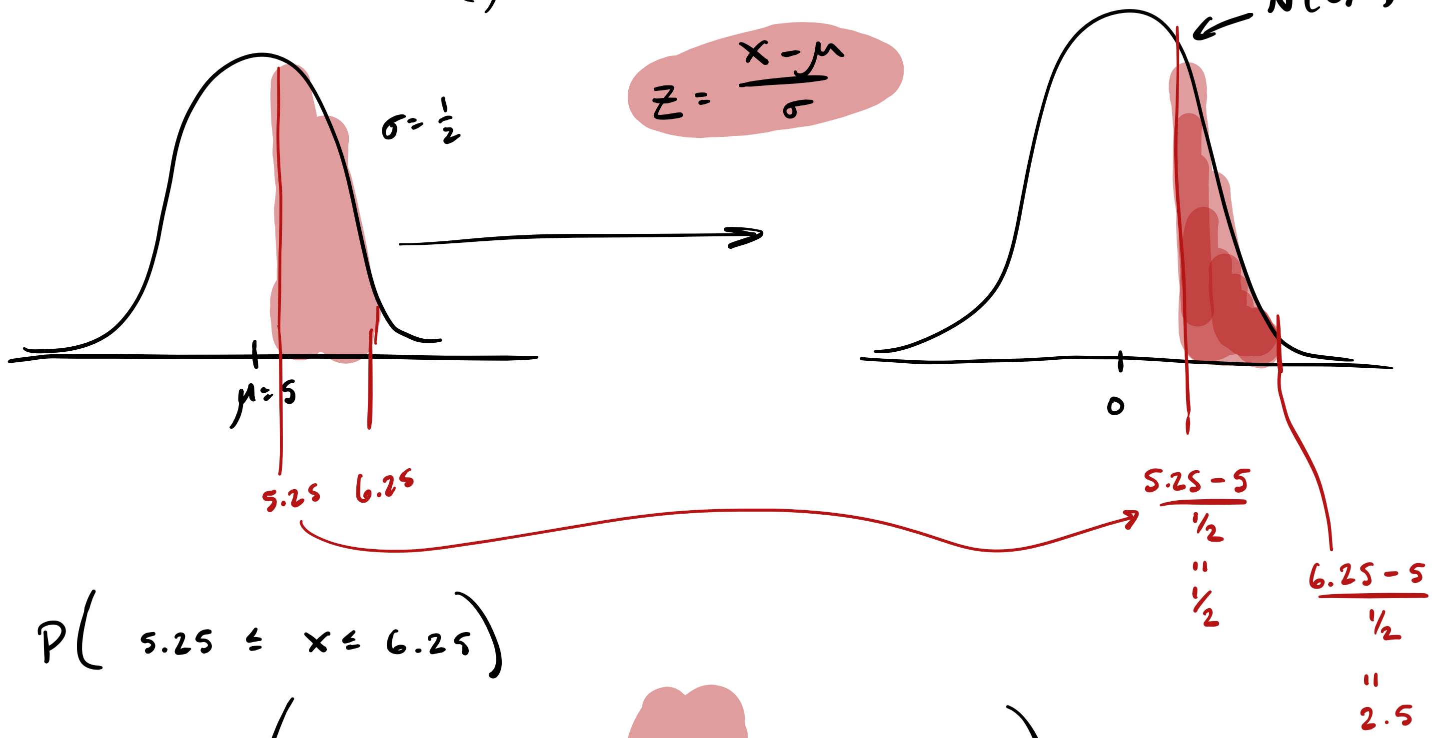
① $P(X < 5.31)$

$$z = \frac{x-\mu}{\sigma}$$



$P(X < 5.25) = .7324$

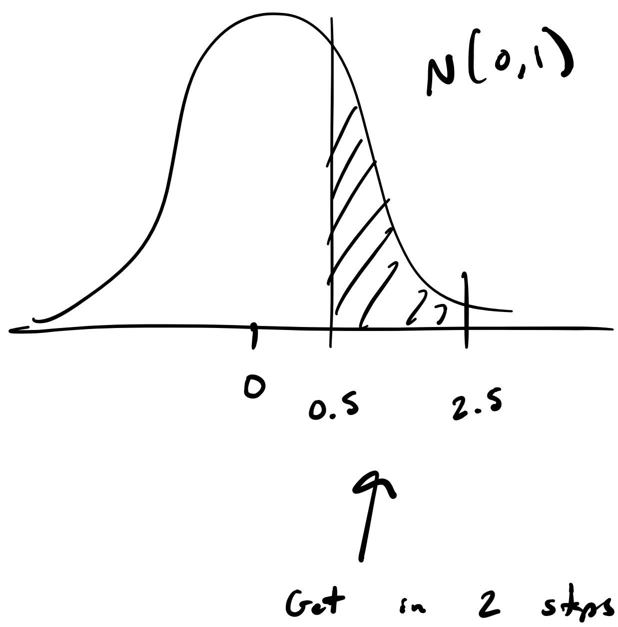
① $X \sim \text{Normal}\left(\mu = 5,\ \sigma^2 = \frac{1}{4}\right)$

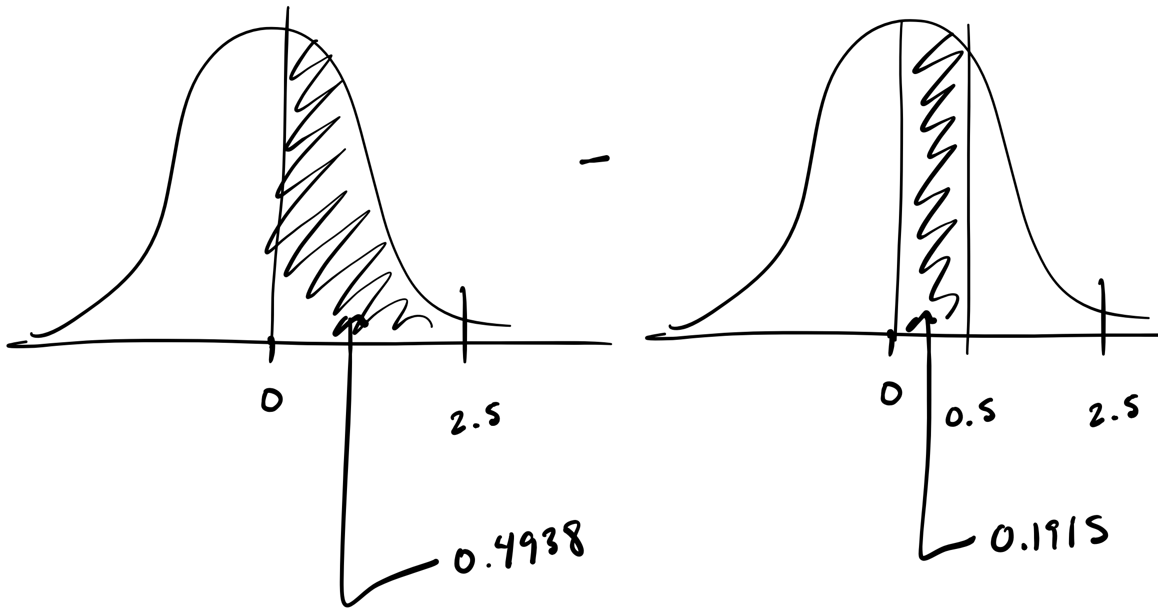


$$P(5.25 \le X \le 6.25)$$

$$= P\left(\frac{5.25 - 5}{1/2} \le \frac{X - 5}{1/2} \le \frac{6.25 - 5}{1/2}\right)$$

$$= P\left(\tfrac{1}{2} \le Z \le 2.5\right)$$

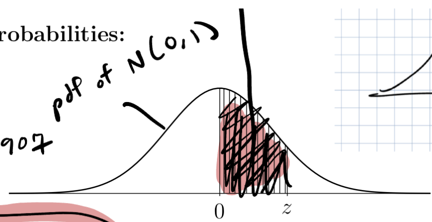


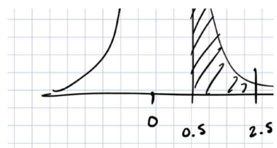


$$.4938 - .1915 = .3023$$

$$P(5.25 < x < 6.25) = 0.3023.$$

Standard Normal probabilities:





$P(0 \le Z \le 1.23) = 0.3907$

Illustration of how to use Z table

	0.00	0.01	0.02	0.03	0.04	0.05	0.06	0.07	0.08	
0.0	0.0000	0.0040	0.0080	0.0120	0.0160	0.0199	0.0239	0.0279	0.0319	0
0.1	0.0398	0.0438	0.0478	0.0517	0.0557	0.0596	0.0636	0.0675	0.0714	0
0.2	0.0793	0.0832	0.0871	0.0910	0.0948	0.0987	0.1026	0.1064	0.1103	0
0.3	0.1179	0.1217	0.1255	0.1293	0.1331	0.1368	0.1406	0.1443	0.1480	0
0.4	0.1554	0.1591	0.1628	0.1664	0.1700	0.1736	0.1772	0.1808	0.1844	0
0.5	0.1915	0.1950	0.1985	0.2019	0.2054	0.2088	0.2123	0.2157	0.2190	0
0.6	0.2257	0.2291	0.2324	0.2357	0.2389	0.2422	0.2454	0.2486	0.2517	0
0.7	0.2580	0.2611	0.2642	0.2673	0.2704	0.2734	0.2764	0.2794	0.2823	0
0.8	0.2881	0.2910	0.2939	0.2967	0.2995	0.3023	0.3051	0.3078	0.3106	0
0.9	0.3159	0.3186	0.3212	0.3238	0.3264	0.3289	0.3315	0.3340	0.3365	0
1.0	0.3413	0.3438	0.3461	0.3485	0.3508	0.3531	0.3554	0.3577	0.3599	0
1.1	0.3643	0.3665	0.3686	0.3708	0.3729	0.3749	0.3770	0.3790	0.3810	0
1.2	0.3849	0.3869	0.3888	0.3907	0.3925	0.3944	0.3962	0.3980	0.3997	0
1.3	0.4032	0.4049	0.4066	0.4082	0.4099	0.4115	0.4131	0.4147	0.4162	0
1.4	0.4192	0.4207	0.4222	0.4236	0.4251	0.4265	0.4279	0.4292	0.4306	0
1.5	0.4332	0.4345	0.4357	0.4370	0.4382	0.4394	0.4406	0.4418	0.4429	0
1.6	0.4452	0.4463	0.4474	0.4484	0.4495	0.4505	0.4515	0.4525	0.4535	0
1.7	0.4554	0.4564	0.4573	0.4582	0.4591	0.4599	0.4608	0.4616	0.4625	0

End of exam I material

Quantiles of a continuous random variable

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

$$P(X \leq q_\theta) = \theta,$$

where θ is a value in $[0, 1]$.

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

Example: If X is the weight of a fresh chicken egg:

- With probability 0.90, a randomly selected egg has weight $\leq q_{0.90}$.
- With probability 0.25, a randomly selected egg has weight exceeding $q_{0.75}$.
- The median weight is $q_{0.50}$.

to "Z world": $Z = \frac{X - \mu}{\sigma}$.   From "Z world" to "X world": $X = \mu + \sigma Z$

If $X \sim \text{Normal}(\mu, \sigma^2)$, we can find q_θ such that $P(X \leq q_\theta) = \theta$ in two steps:

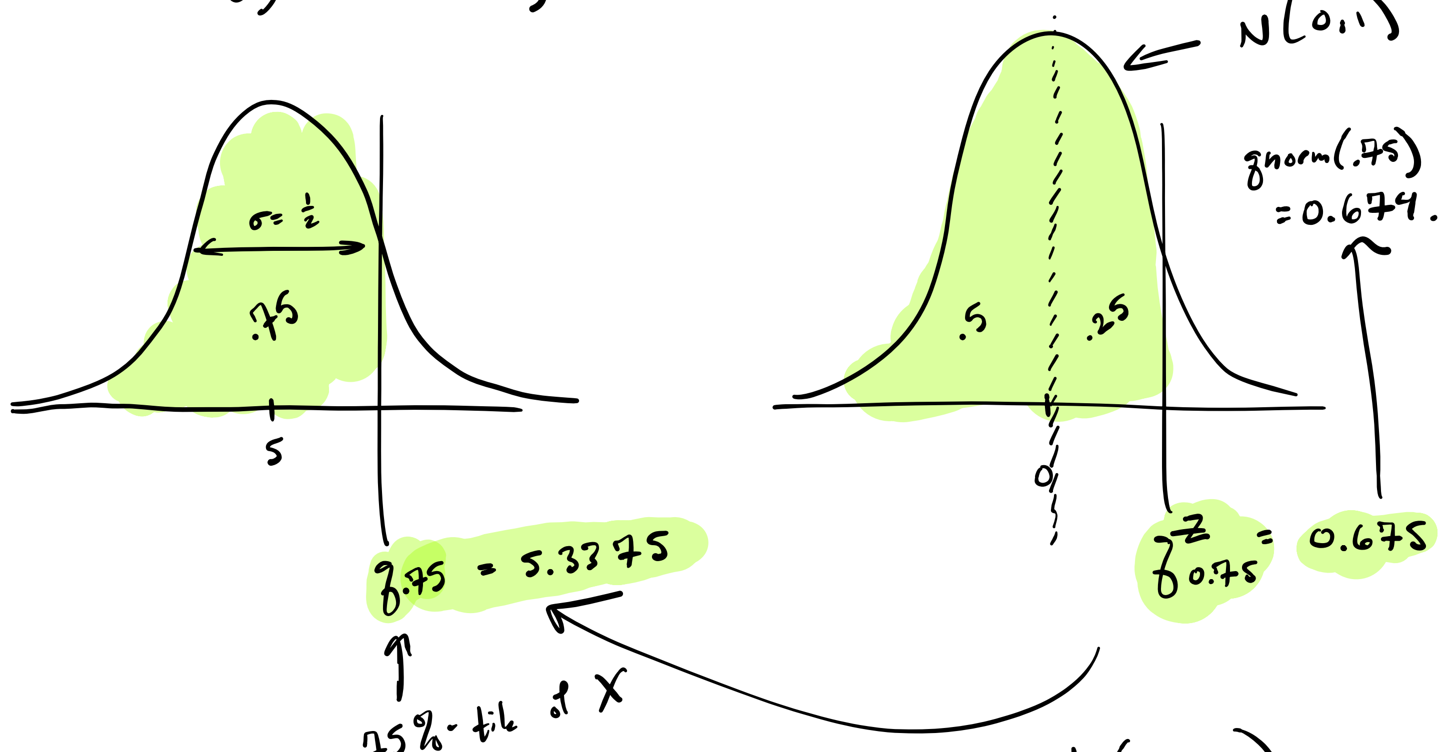
1. Find q_θ^Z such that $P(Z < q_\theta^Z) = \theta$ using a "Z-table".
2. Get the corresponding quantile in the X-world as

$$q_\theta = \mu + \sigma q_\theta^Z$$

Exercise: Suppose growth in height (ft) of Loblolly pines from age three to five is $\text{Normal}(\mu = 5, \sigma^2 = 1/4)$. Let X denote the ~~height~~ growth of a randomly selected Loblolly pine and find

1. the 75%-tile of growth.
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.

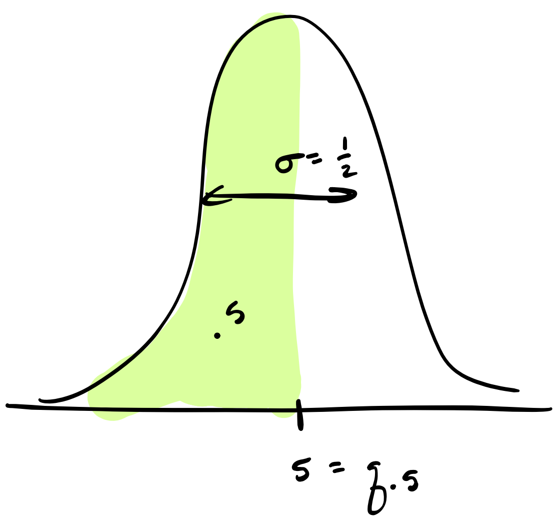
① $X \sim N(\mu = 5, \sigma^2 = \frac{1}{4})$



$$q_{.75} = 5 + \frac{1}{2}(0.675) = 5 + .3375$$
$$= 5.3375$$

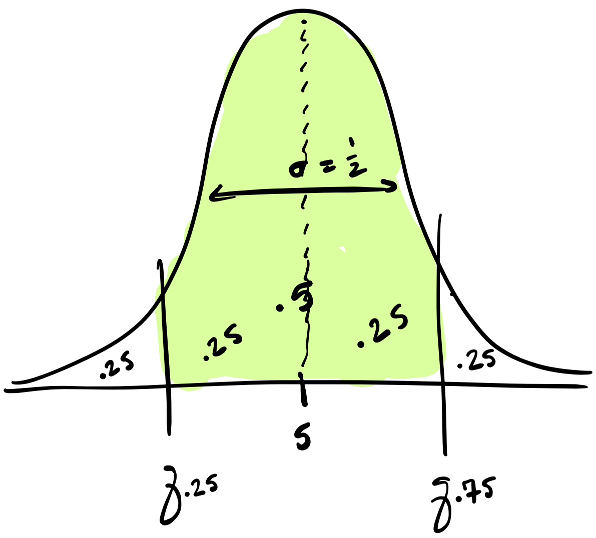
$$z = \frac{x - \mu}{\sigma} \iff x = \mu + \sigma z$$

② 50%-tile of X



For a symmetric distribution, the mean and median (50%-tile) are equal!!

③ $X \sim N\left(\mu = 5,\ \sigma^2 = \frac{1}{4}\right)$

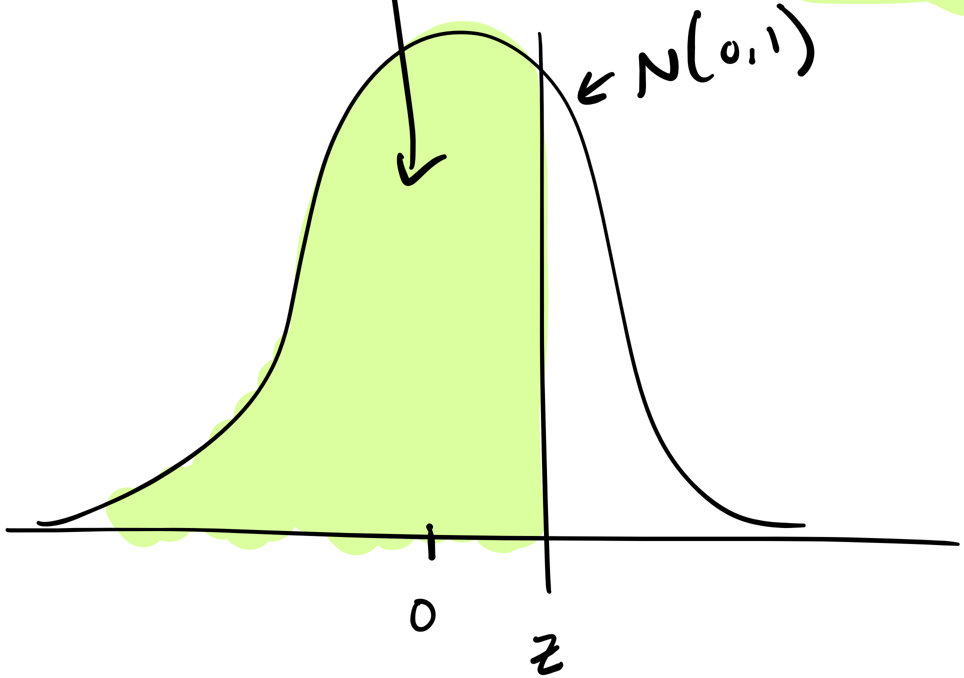


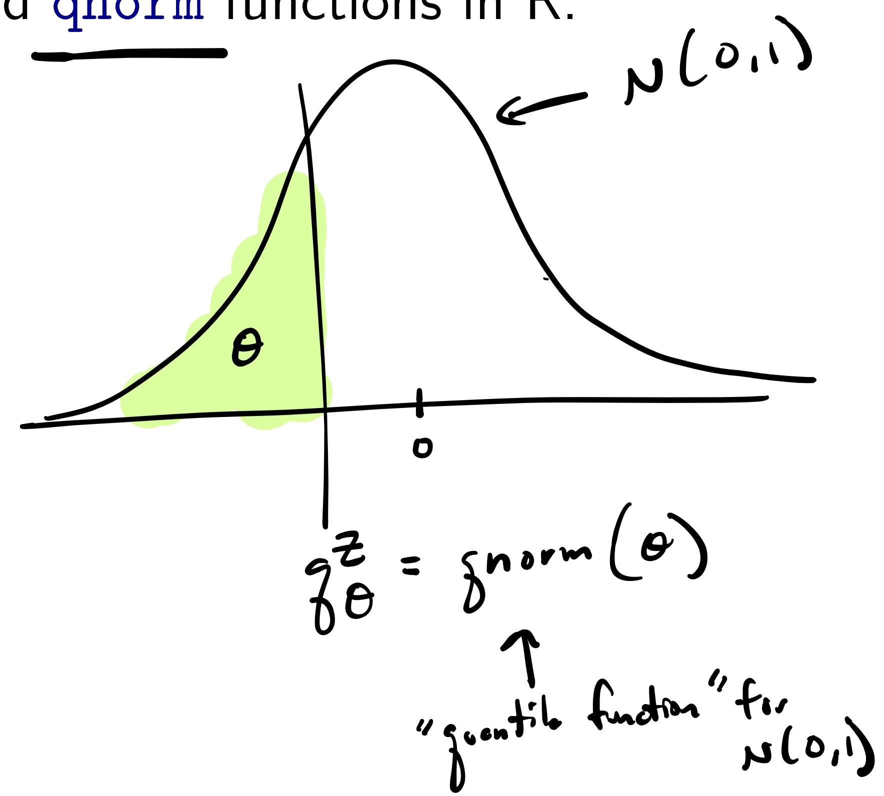
$z_{.75} = 5.3375$

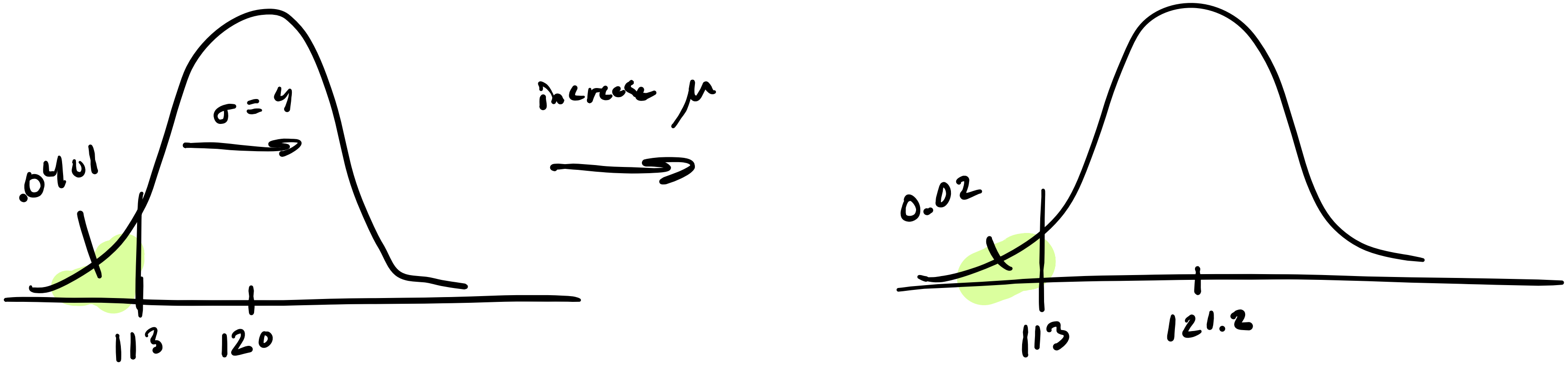
$z_{.25} = 4.6625 = (5 - 0.3375)$

pnorm(z) (this is the cdf – cumulative dist. function – of N(0,1)).

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





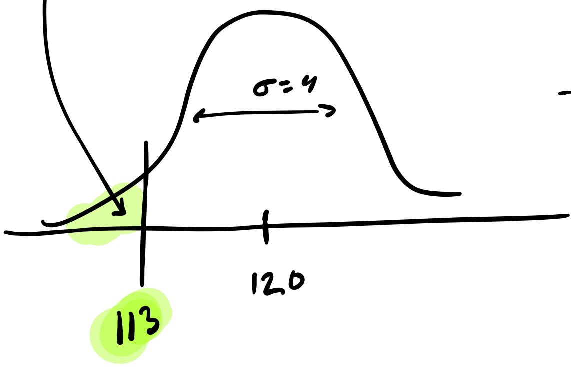


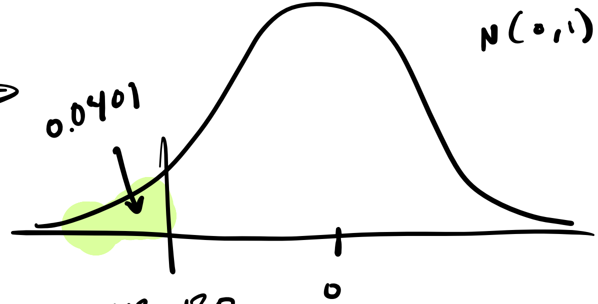
Exercise: You sell jars of baby food labelled as weighing 4oz ≈ 113g. Suppose your process results in jar weights with the Normal($\mu = 120, \sigma^2 = 4^2$) distribution and that you will be fined if more than 2% of your jars weigh less than 113g.

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?

(1) $X \sim N(\mu = 120, \sigma^2 = 4^2)$. $\sigma = 4$

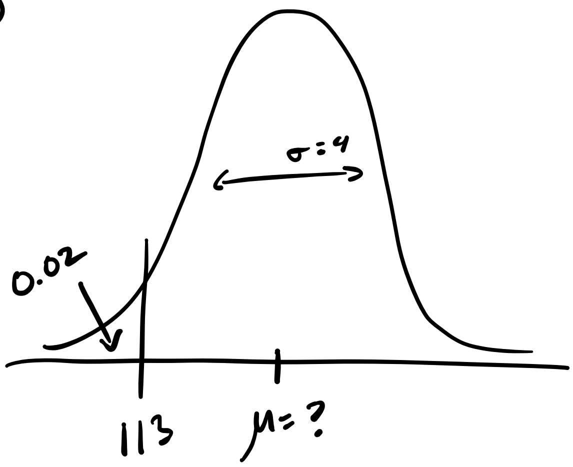
$P(X < 113) = 0.0401$ $\quad z = \frac{x - \mu}{\sigma}$

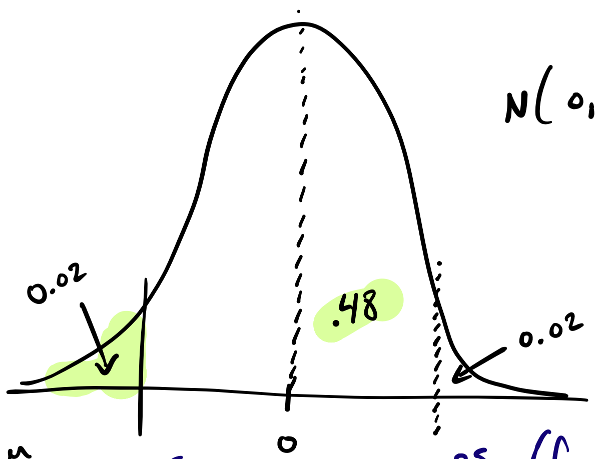




$$\frac{113-120}{4} = \frac{-7}{4} = -1.75$$

②





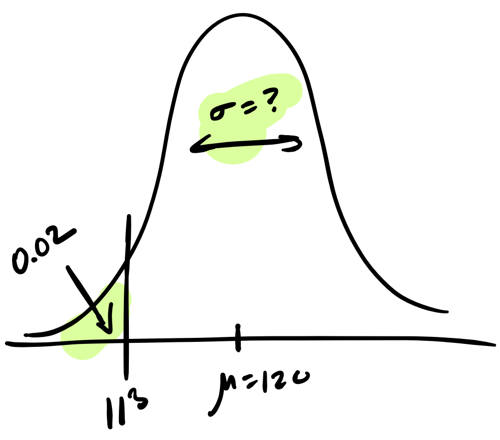
$$\frac{113-\mu}{4} = -2.05 \qquad 2.05 \text{ (from Z table)}$$

$$\Leftrightarrow \quad 113 - \mu = (-2.05)4$$

$$\Leftrightarrow \quad 113 + \underbrace{(2.05)\cdot 4}_{8.2} = \mu$$

$$\Leftrightarrow \quad \mu = 121.2$$

③



Do my data come from a Normal distribution?

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

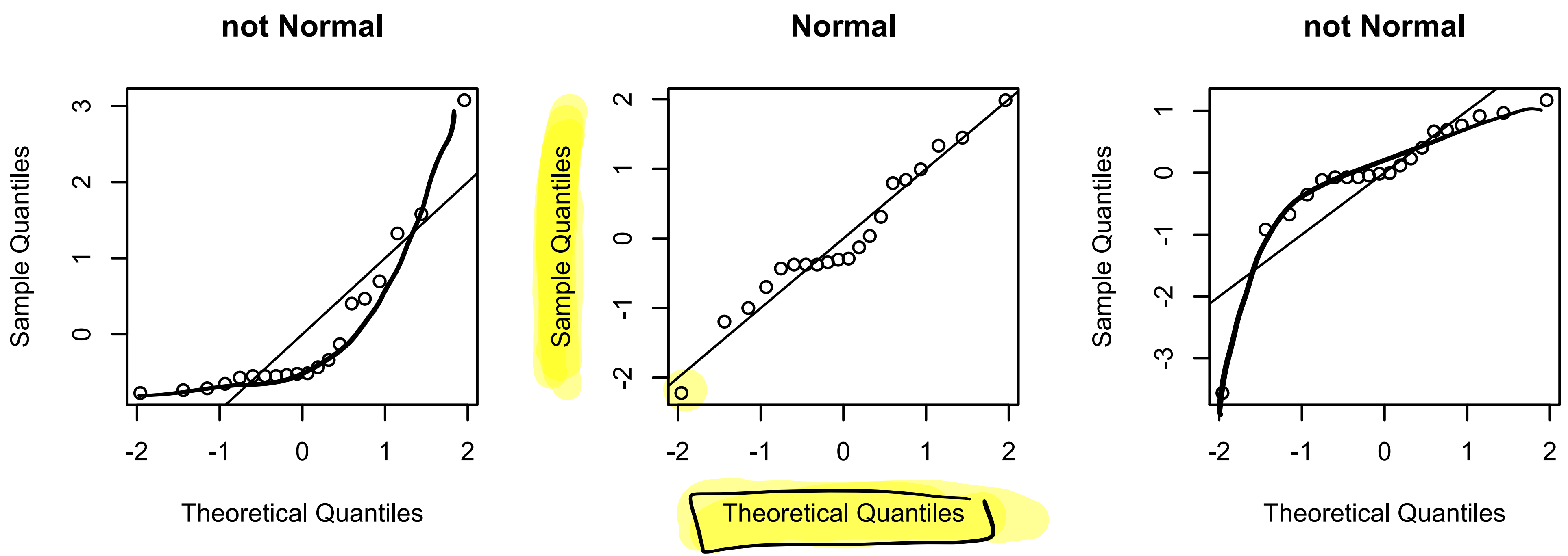
1832	1440	1620	1362	577	934	928	998	1062	900
1380	913	654	878	172	773	1171	1574	900	900

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

$\bar{X}_n$ = average, estimate for μ

S_n^2 = sample variance, estimate for σ^2.

Some more Q-Q plots:



Sum of independent Normal random variables

If $X_1 \sim \text{Normal}(\mu_1, \sigma_1^2), \ldots, X_n \sim \text{Normal}(\mu_1, \sigma_n^2)$ are independent random variables, then

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

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.

$X \sim N(\mu = 120, \sigma^2 = 4^2)$ ← each jar.

$Y = X_1 + \dots + X_{25}$ (Sum of 25 jar weights)

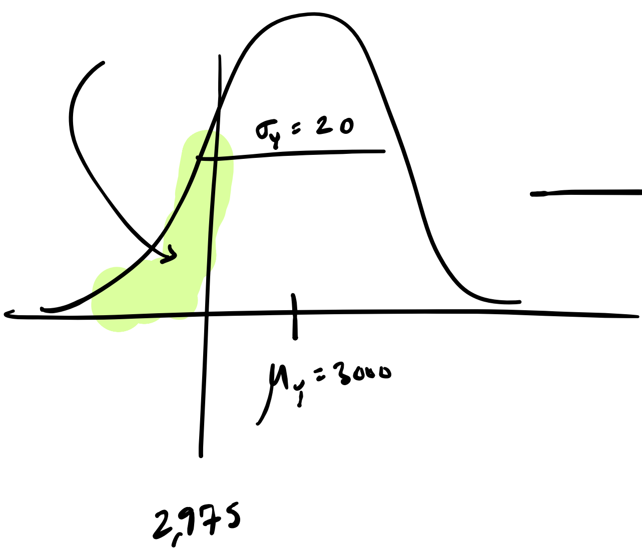
Then $Y \sim \text{Normal}\left(25 * 120,\ 25 * 4^2\right)$

① $\mathbb{E}Y = 25 * 120 = 2500 + 500 = 3{,}000$

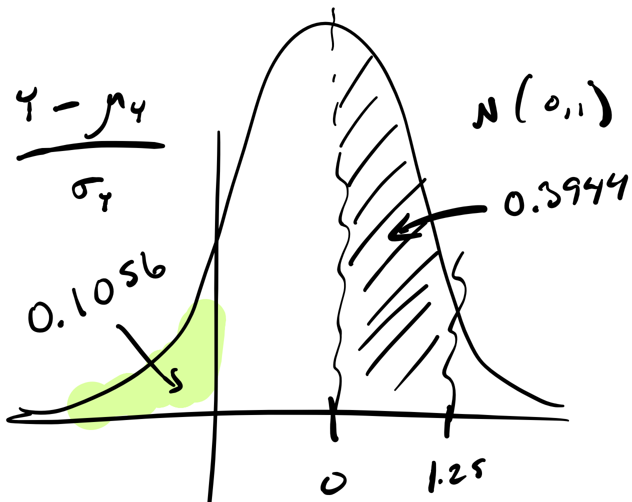
② $\sqrt{\text{Var}\, Y} = \sqrt{25 \cdot 4^2} = 5 \cdot 4 = 20$

③ $Y \sim \text{Normal}\left(\mu_Y = 3000,\ \sigma_Y^2 = 20^2\right)$

$P(Y \le 2{,}975) = 0.1056$



$Z = \dfrac{Y - \mu_Y}{\sigma_Y}$



$$\frac{2975 - 3000}{20} = \frac{-25}{20} = -1.25$$