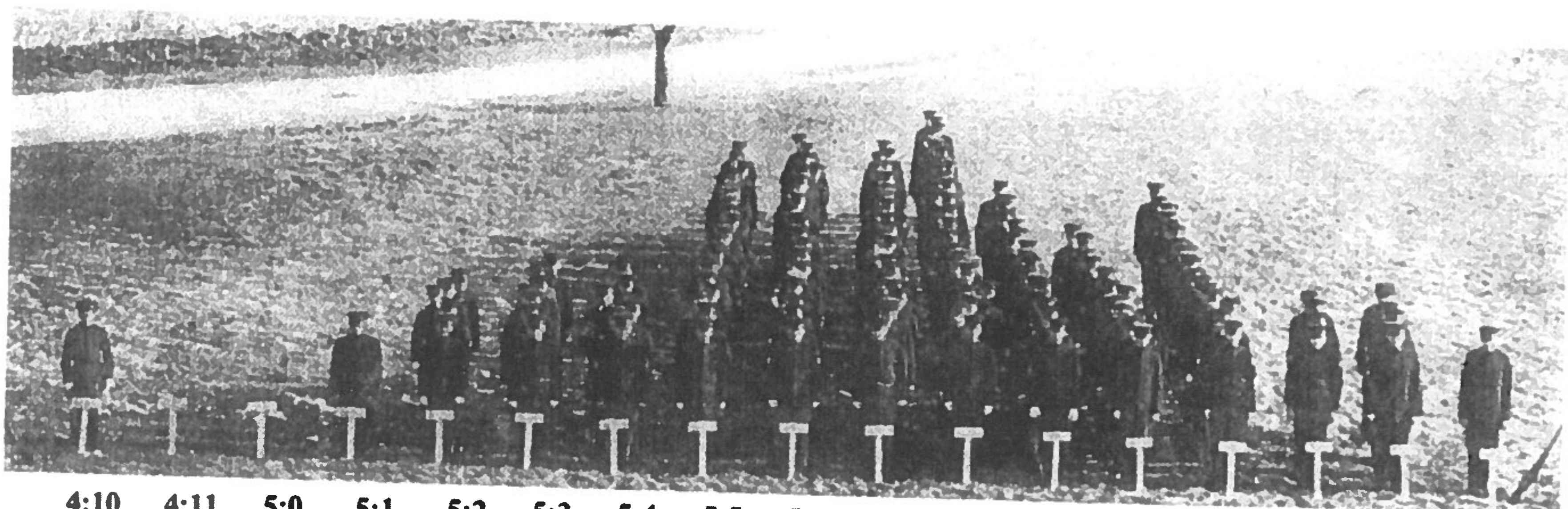
Chapter 5 Normal Distribution

Many phenomena give rise to data whose distribution is bell-shaped and roughly symmetric. For example, height, blood pressure and exam scores are often well described by a Normal model.



4:10 4:11 5:0 5:1 5:2 5:3 5:4 5:5 5:6 5:7 5:8 5:9 5:10 5:11 6:0 6:1 6:2

Figure 12. Living histogram of 175 male college students (Blakeslee 1914).

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

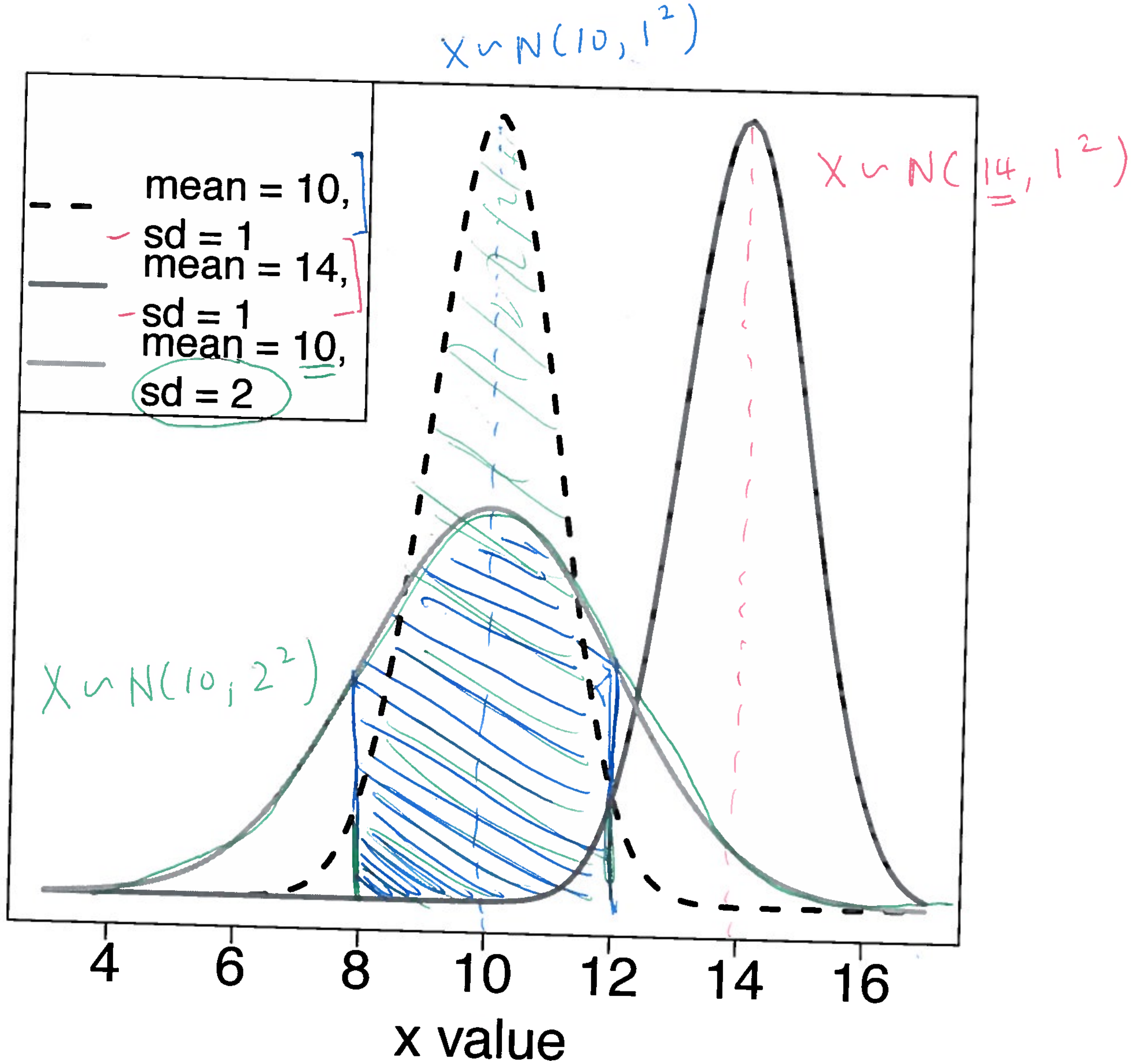
The Normal distribution is a continuous distribution with density

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

where μ (mean) and σ (standard deviation) are parameters, which control the central location and dispersion of the density.

(Note: we rarely use this in calculations for STAT241/251)

The Normal distribution is bell shaped, unimodal and symmetric about the mean (μ).



Z-scores:

- The Normal distribution with $\mu = 0$ and $\sigma = 1$ is called a standard Normal distribution. Notice that any random variable $X \sim N(\mu, \sigma^2)$ can be made into a standard Normal random variable. We often refer to a standard Normal as a **Z-score**, $Z \sim N(0, 1)$:

$$\boxed{Z = \frac{x-\mu}{\sigma}}$$

- A positive (negative) Z-score implies the observation has a value above (below) the mean.
- $Z = k$ (k positive) corresponds to an observation that is k many SDs above the mean. $Z = -k$ (k positive) corresponds to an observation that is k many SDs below the mean.

		μ	σ
→ you :	88%	78	5
→ Jerm	90%	85	5

$z = \frac{88 - 78}{5} = 2$

$z = \frac{90 - 85}{5} = 1$

$z = 2$ 2 SD above the mean.

$z = 0$

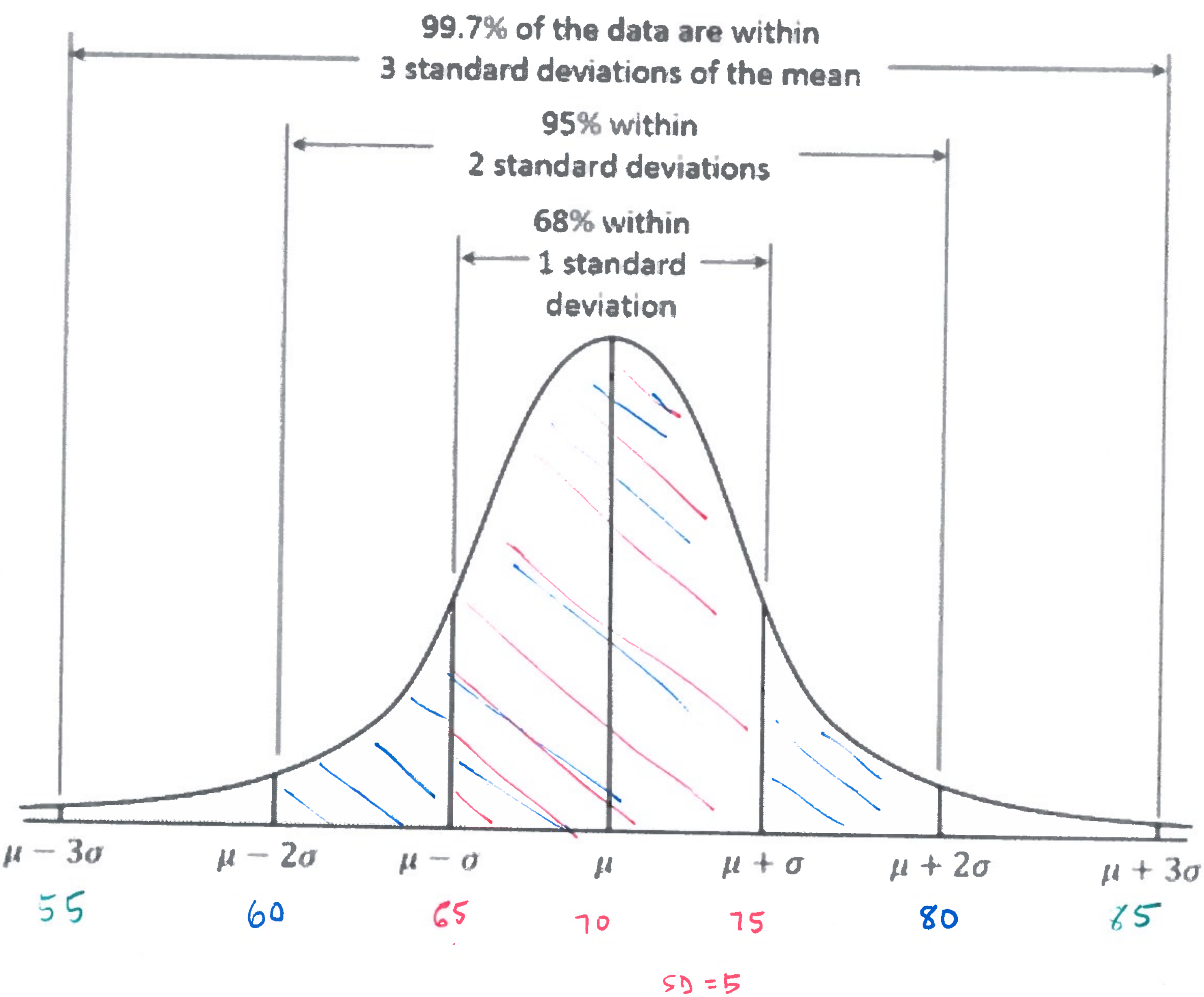
$z = -1$ 1 SD below the mean.

Finding Probabilities

Areas under the Normal density can be found using:

- Computer software
- Statistical tables. Statistical tables only give areas under the $N(0,1)$ curve (Posted on course website)
- 68-95-99.7 Rule (Empirical rule)

68-95-99.7 Rule



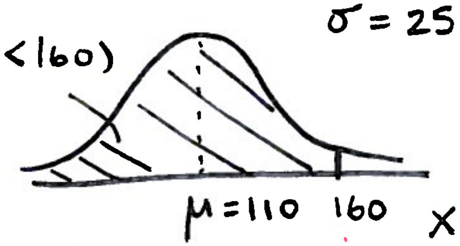
Example 1

Scores on a standard IQ test follow approximately the Normal model with mean $\mu = 110$ and standard deviation $\sigma = 25$.

(a) What percentage of people have IQ scores below 160? ✓

(b) What percentage of people have scores between 90 and 120?

a) $P(X<160)$ $\sigma = 25$

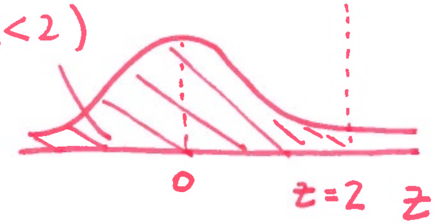


X rep. IQ score

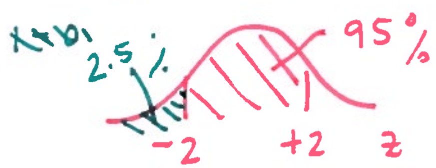
$X \sim N(110, 25^2)$

$$z = \frac{x-\mu}{\sigma} = \frac{160-110}{25} = 2$$

$P(z<2)$



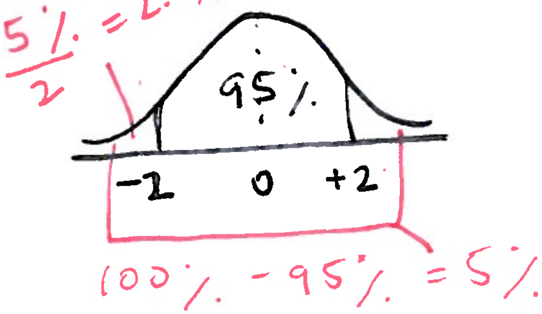
68 - 95 - 99.7

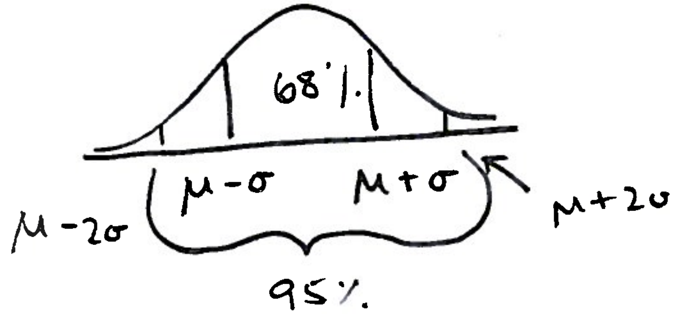


approx. $P(z<2) = 95\% + 2.5\%$

$= 97.5\%$

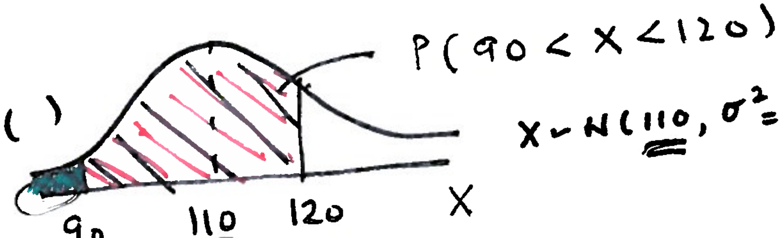
$Z \sim N(0,1)$





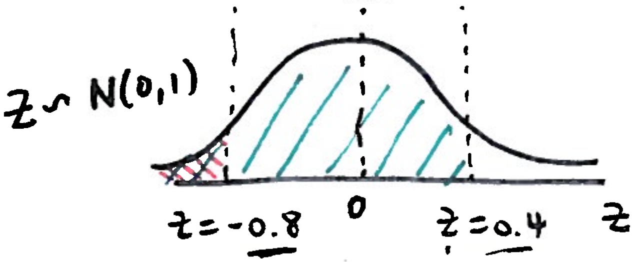
b)

$X \sim N(\ )$



$X \sim N(110, \sigma^2 = 25^2)$

$Z \sim N(0,1)$



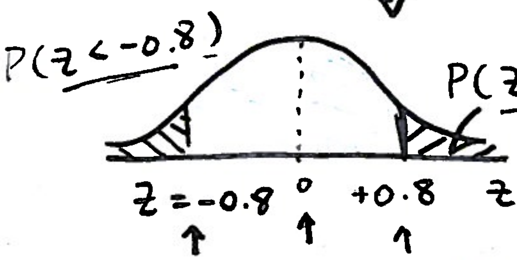
$P(90 < X < 120) = P(X < 120) - P(X < 90)$

$= P\left(Z < \frac{120-110}{25}\right) - P\left(Z < \frac{90-110}{25}\right)$

$= P(Z < 0.4) - P(Z < -0.8)$

standard norm table $\rightarrow$ $= 0.6554$

$P(Z < -0.8)$



$P(Z > 0.8) = 1 - P(Z < 0.8)$

$P(Z < -0.8) = 1 - P(Z < 0.8)$

$= 0.6554 - (1 - 0.7881)$

$= 0.4435$

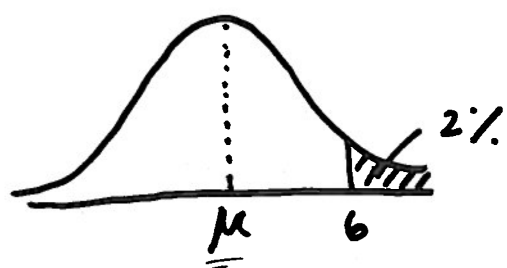
Example 2

A machine used to control the amount of dye dispensed for mixing shades of paint can be set so that it dispenses an average of μ milliliters (mL) of dye per paint can. The amount of dye dispensed follows a Normal model with a standard deviation of 0.4 mL. If more than 6 mL of dye are dispensed when making a certain shade of blue paint, the shade is deemed unacceptable. Determine the setting for the mean μ such that only 2% of the cans of paint will be unacceptable.

$\mu = ?$ $\quad X$ = amount dye dispensed

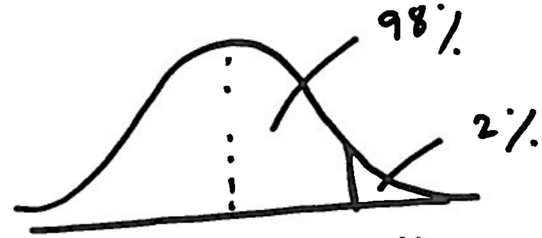
$X \sim N(\mu, \sigma^2 = 0.4^2)$

$P(X>6) = 0.02$



$$P(X>6) = P\left(Z > \frac{6-\mu}{0.4}\right) = 0.02$$

$$= 1 - P\left(Z < \frac{6-\mu}{0.4}\right) = 0.02$$



$$z = \frac{6-\mu}{0.4} =$$

$$P\left(Z < \frac{6-\mu}{0.4}\right) = 0.98$$

$$\frac{6-\mu}{0.4} = 2.055$$ ← from table.

$$\mu = 5.178$$

Independent Normal Samples

- The sums (and differences) of independent Normal variables are also Normal
  - If $X \sim N(10, 25)$ and $Y \sim N(5, 16)$ are independent, then $X + Y$ is also Normal.

$E(X+Y) = E(X) + E(Y)$
$= 10 + 5 = 15$

$$\implies X + Y \sim N(15, 41)$$

$Var(X+Y) = V(X) + V(Y)$
$= 25 + 16$
$= 41$

$E(X-Y) = 10 - 5$
$= 5$

What about $X - Y$?

$\implies X - Y \sim N(5, 41)$

$Var(X-Y)$
$= V(X) + V(Y)$
$= 41$

  - If $X_1, X_2, ..., X_n$ are n independent observations of Normal variable $X_i \sim N(\mu, \sigma^2)$, then

$E(X_1 + ... + X_n) = E(X_1) + ... + E(X_n)$
$= n\mu$

$$X_1 + X_2 + ... + X_n \sim N(n\mu, n\sigma^2)$$

$Var(X_1) + ... + Var(X_n)$

  - Let $X_1, X_2, ..., X_n$ be a random sample of n independent observations. If $X_i \sim N(\mu, \sigma^2)$, then the sample mean $\bar{X}$ also follows a Normal distribution and

$$\bar{X} \sim N(\mu, \frac{\sigma^2}{n})$$

$$E(\bar{X}) = E\left(\frac{X_1 + \ldots + X_n}{n}\right) = \frac{1}{n}\left[E(X_1) + \ldots + E(X_n)\right] = \frac{\cancel{n}\mu}{\cancel{n}} = \mu$$

$$Var(\bar{X}) = Var\left(\frac{X_1 + \ldots + X_n}{n}\right) = \frac{1}{n^2}\left[Var(X_1) + \ldots + Var(X_n)\right]$$

$$= \frac{1}{n^{\cancel{2}}}\left[\cancel{n}\sigma^2\right] = \frac{\sigma^2}{n}$$

a) Packing 2 sys. > 20

$P_i \sim N(9, 1.5^2) \quad i = 1, 2$

$T = P_1 + P_2 \qquad P(T > 20)$

$T \sim\ ? \qquad T \sim N(\mu_T = ?, \sigma_T^2 = ?)$

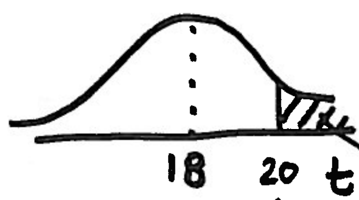
$E(T) = E(P_1 + P_2) = E(P_1) + E(P_2) = 9 + 9 = 18$

$\sigma_T^2 = Var(T) = Var(P_1 + P_2) = Var(P_1) + V(P_2) \quad P_i$ indep.

$= 1.5^2 + 1.5^2$

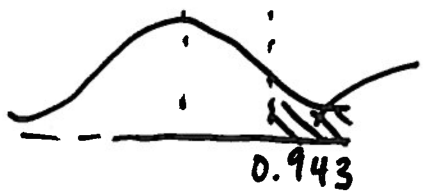
$= 4.5$

$T \sim N(18, 4.5)$



$\sigma^2 = 4.5$

$P(T > 20) = P\left(Z > \frac{20 - 18}{\sqrt{4.5}}\right)$



$P(Z > 0.943)$

$= P(Z > 0.943)$

$= 1 - P(Z < 0.943)$

$= 0.173$

Example 3

A certain company manufactures stereo systems. The times required to pack the stereos can be described by a Normal model with mean of 9 minutes and standard deviation of 1.5 minutes. The times for the boxing stage can be modeled as Normal with a mean of 6 minutes and standard deviation of 1 minute. Assume the two packing times are independent.

(a) What is the probability that packing two consecutive systems takes over 20 minutes?

(b) What percentage of the stereo systems take longer to pack than to box? Assume that packing and boxing times are independent.

b)

P = packing time

B = boxing " "

$P(P > B) =$

$P > B \rightarrow P - B > 0$

$P - B \sim N(\mu, \sigma^2)$

b) $E(P-B) = E(P) - E(B) = 9 - 6 = 3 \text{ min}$

$Var(P-B) = Var(P) + Var(B) = 1.5^2 + 1^2 = 3.25$

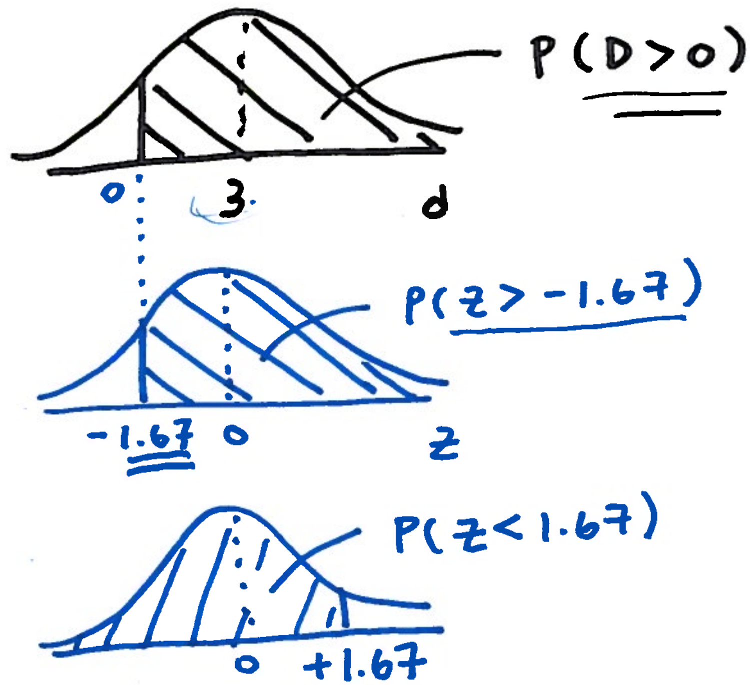
$D = P - B \sim N(3,\ 3.25)$

$P(D>0) = P\left(Z > \dfrac{0-3}{\sqrt{3.25}}\right)$

$= P(Z > -1.67)$

$= P(Z < 1.67)$

Z-table $\rightarrow$ $= 0.9525$



Example 4

Assume that the true mean weight of summertime airline passengers is 85 kg, with a standard deviation of 15 kg. The population of weights is reasonably Normally distributed. A commuter plane carries 25 passengers. What is the probability that the total weight of the passengers exceeds 2350 kg?

$T = \underline{X_1 + \dots + X_{25}}$ X rep. weight passengers $\sim N(85, 15^2)$

$P(T > 2350)$

$E(T) = E(X_1 + \dots + X_{25}) = 25\,E(X) = \underline{2125}$

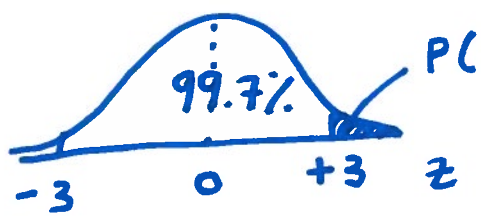
$$
\begin{aligned}
Var(T) &= Var(X_1 + \dots + X_{25}) \\
&= Var(X_1) + \dots + Var(X_{25}) \\
&= 25 \times 15^2 \\
&= \underline{5625}
\end{aligned}
$$

assume indep. X_i

$T \sim N(\underline{2125}, 5625)$

$$P(T > 2350) = P\left(Z > \frac{2350 - 2125}{\sqrt{5625}}\right) = \underline{P(Z > 3)} \approx 0.0015$$

68-95-99.7



$P(Z > 3) = \frac{0.3\%}{2} = 0.15\%$