II.5 The Normal Approximation for Data

- The Normal Curve
  - the general case, $N(\mu, \sigma^2)$ - sheet 1a
  - the case standard normal, $N(0,1)$ - sheet 1b

- Computing $P(a \leq N(0,1) \leq b)$ - sheet 1c

- Computing $P(c \leq N(\mu, \sigma^2) \leq d)$ changes "c" to standard unit
  Use
  \[
P(c \leq N(\mu, \sigma^2) \leq d) = P\left(\frac{c-\mu}{\sigma} \leq N(0,1) \leq \frac{d-\mu}{\sigma}\right)
  \]
  compute using $N(0,1)$ table

Example: Trees in a forest have a mean diameter of 10" with a standard deviation $\sigma = 2". What is the probability that a randomly selected tree has a diameter between 6" and 15"? Assume a Normal Distribution (approx).

Then
\[
P(6 \leq N(10,2^2) \leq 15) = P\left(\frac{6-10}{2} \leq N(0,1) \leq \frac{15-10}{2}\right)
\]
\[
= P(-2 \leq N(0,1) \leq 2.5)
\]
\[
= 0.971.
\]

Normal Approximation to Histograms - Sheet 1d
The Normal Distribution

- Density function of form:
  \[ f(x) = \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{1}{2\sigma^2}(x-\mu)^2}, \quad -\infty < x < \infty. \]

Each choice of \((\mu, \sigma)\) defines a particular normal distribution. The case \(\mu=0, \sigma=1\) is called "standard normal."

- Mean and variance: \(\mu\) and \(\sigma^2\)

- Picture and properties:
  - Curve is symmetric & bell-shaped
  - Points of inflection:
    - \(\mu-3\sigma\)
    - \(\mu-2\sigma\)
    - \(\mu+2\sigma\)
    - \(\mu+3\sigma\)
  - \(N(\mu, \sigma^2)\) density
  - \(68.27\%\) between \(\mu-1\sigma\) and \(\mu+1\sigma\)
  - \(95.45\%\) between \(\mu-2\sigma\) and \(\mu+2\sigma\)
  - \(99.73\%\) between \(\mu-3\sigma\) and \(\mu+3\sigma\)
  - \(\mu \pm 1.96\sigma\) : 95\% exactly
  - \(\mu \pm 2.58\sigma\) : 99\% exactly
The Standard Normal Distn

- \( f(x) = \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}x^2} \), \(-\infty < x < \infty\)
  (or \( \phi(x) \))

- mean = 0
- variance = 1

- Table gives a tabulation for \( N(0,1) \)

- CDF:
  \[
  \Phi(x) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{x} e^{-\frac{1}{2}u^2} \, du
  \]

- \( \Phi(-\infty) = 1 - \Phi(0) \), \#70
  (Symmetry) \( \Phi(-x) = 1 - \Phi(x) \)

- Computing \( P(a \leq N(0,1) \leq b) \)

  \[
  P(a \leq N(0,1) \leq b) = \Phi(b) - \Phi(a)
  \]

- \( a \quad 0 \quad b \)
  \( \Phi(a) \) \( \Phi(b) \)
Example: Find $P(-1.5 \leq N(0,1) \leq +0.67)$

\[
\frac{1}{2} \left[ \begin{array}{c}
\frac{1}{2} [0.866 - 0.504] \\
\frac{1}{2} (0.866) + \frac{1}{2} (0.504)
\end{array} \right]
\]

\[
= \frac{1}{2} (0.866) + \frac{1}{2} (0.504)
\]

\[
= 0.685
\]

86.64% from Table

0.8664

0.504 (approx.) from table
- **Getting Percentiles of \( N(0,1) \)**

Find value \( z_a \) such that \( \frac{3}{4} = \text{area } A \)

**Example:** Find \( z_{25} \) (1st quartile)

\[ 25\% \quad 50\% \quad 25\% \]

\[ 0.25 \quad 0.50 \quad 0.25 \]

\( z_{25} = -0.675 \) from table

\[ z_{25} = -0.68 \]

**Example:** Find \( z_{75} \) (\( = +0.68 \))

**Example:** Find \( z_{50} \) (\( = \text{median} = 0 \))

- **Getting Percentiles of \( N(\mu, \sigma^2) \)**

The \( P \)th percentile of \( N(\mu, \sigma^2) \)

\[ = \mu + \left[ \frac{P \text{th percentile of } N(0,1)}{\sigma} \right] \sigma \]

**Example (text, p. 90)**

Math SAT scores of all applicants to a university average 535 with SD=100. Use \( N(535, 100^2) \) dist'n.

Find 95th %tile.

**Solution:** \( z_{95} \approx 1.65 \)

Then 95th percentile of the SAT scores is

\[ 535 + 1.65 \times 100 = 535 + 165 = 700. \]