Continuous Random Variables and Probability Distributions

- **4.1 Continuous Random Variables and Probability Density Functions**
- **4.2 Cumulative Distribution Functions and Expected Values**
- **4.3 The Normal Distribution**
- **4.4 The Gamma Distribution and Its Relatives**
- **4.5 Other Continuous Distributions**
- **4.6 Probability Plots**

4.3 The Normal Distribution

- **Normal (Gaussian) Distribution**

A continuous rv 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

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

Note:
1.The normal distribution is the most important one in all of probability and statistics. Many numerical populations have distributions that can be fit very closely by an appropriate normal curve.
2.Even when the underlying distribution is discrete, the normal curve often gives an excellent approximation.
3.Central Limit Theorem (see next Chapter)

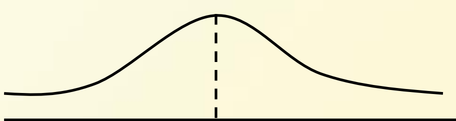
4.3 The Normal Distribution

- **Properties of f(x;μ,σ)**

$$f(x;\mu,\sigma)\geq 0,\ \int_{-\infty}^{+\infty}\frac{1}{\sqrt{2\pi}\sigma}e^{-(x-\mu)^2/(2\sigma^2)}dx=1$$ **Proof?**

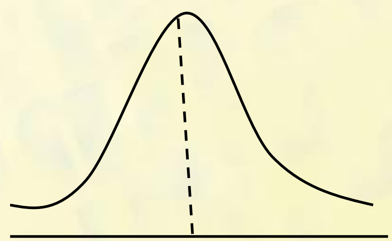
E(X) = μ & V(X) = σ^2 , X~ N(μ, σ^2)

σ is large



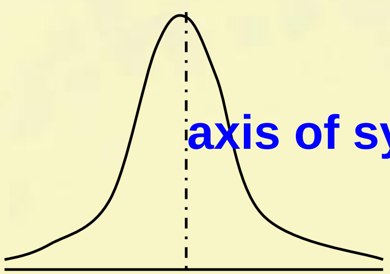
μ

σ is medium



μ

σ is small



μ

u is location of axis of symmetry
σcontrol the shape of the graph

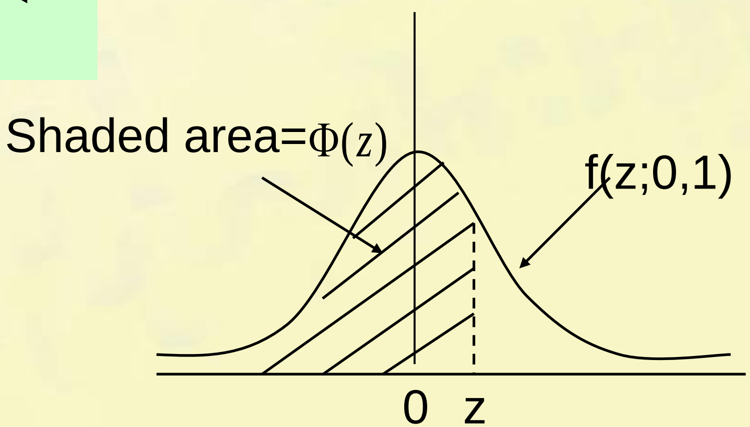
Symmetry Shape

- **Standard Normal Distribution**

The normal distribution with parameter values $\mu=0$ and $\sigma=1$ is called the standard normal distribution. A random variable that has a standard normal distribution is called a standard normal random variable and will be denoted by Z. **The pdf of Z is**

$$f(z;0,1) = \frac{1}{\sqrt{2\pi}} e^{-z^2/2}, \quad -\infty < z < \infty$$

The cdf of Z is



$$\Phi(z) = \int_{-\infty}^{z} f(t)dt = \int_{-\infty}^{z} \frac{1}{\sqrt{2\pi}} e^{-t^2/2} dt$$

Refer to Appendix Table A.3

4.3 The Normal Distribution

- Properties of Φ(z)

$$\Phi(-z) = 1 - \Phi(z)$$

$$\Phi(0) = 0.5$$

$$P(|X| \le z) = 2\Phi(z) - 1$$

$$P(|X| \ge z) = 2[1 - \Phi(z)]$$

4.3 The Normal Distribution

- Example

Find Φ (1.65), Φ (-1.96) using appendix Table A.3

Solution:

Z	0.05	0.06
1.6	0.9505	
1.9		0.9750

$$\Phi(1.65) = 0.9505$$

$$\Phi(-1.96) = 1 - \Phi(1.96) = 1 - 0.9750 = 0.0250$$

Example

Given $X \sim N(0,1)$

Find $P(X \le 1.65)$

$$P(1.65 < X \le 2.09)$$

Solution:

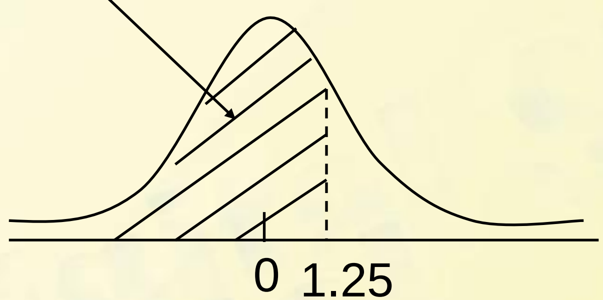
$$P(X<1.65)$$
$$=\Phi(1.65)=0.9505$$
$$P(1.65<X<2.09)$$
$$=\Phi(2.09)-\Phi(1.65)$$
$$=0.9817-0.9505=0.0312$$

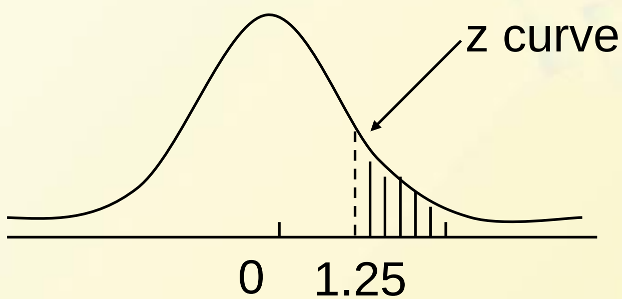
- Example 4.13

Compute the following standard normal probabilities:

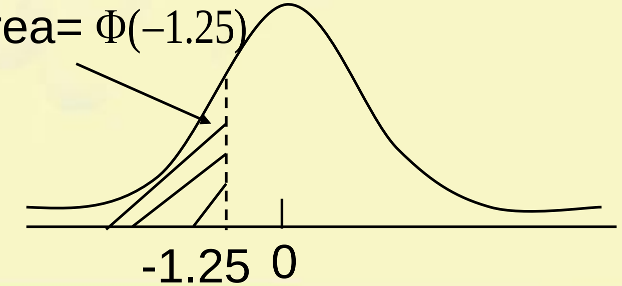
(a) $P(Z\leq 1.25)$ (b) $P(Z>1.25)$ (c) $P(Z\leq -1.25)$

Shaded area= $\Phi(1.25)$ =0.8944



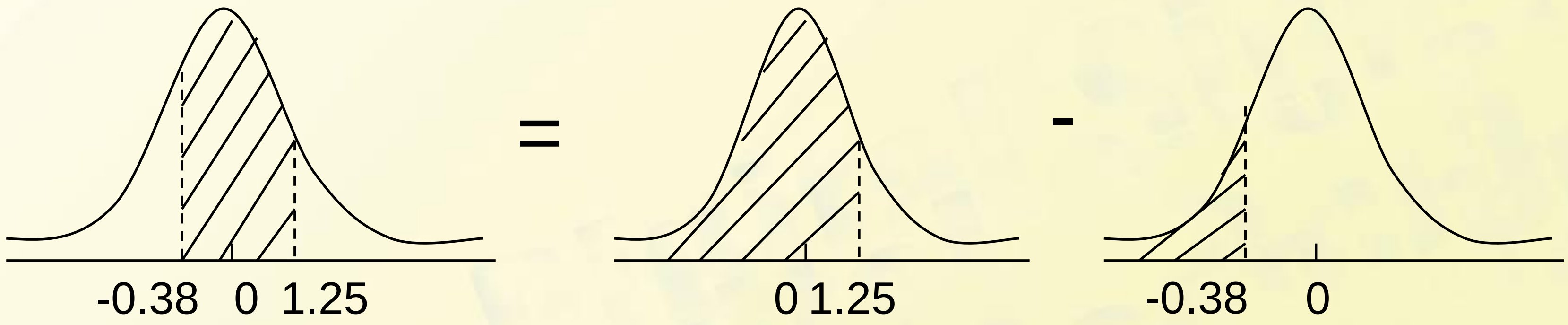


Shaded area= $\Phi(-1.25)$



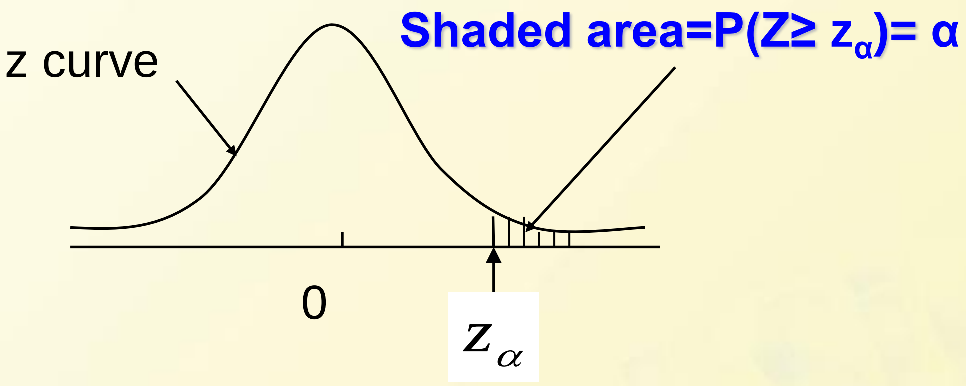
- Example 4.13(Cont')

(d) $P(-0.38 \leq Z \leq 1.25)$



z_α notation

z_α will denote the values on the measurement axis for which α of the area under the z curve lies to the right of z_α



Note: Z_α is the 100(1- α)th percentile of the standard normal distribution

Table 4.1 standard normal percentiles and critical values

Percentile	90	95	97.5	99	99.5	99.9	99.95
α (tail area)	0.1	0.05	0.025	0.01	0.005	0.001	0.0005
$z_\alpha = 100(1-\alpha)th$ percentile	1.28	1.645	1.96	2.33	2.58	3.08	3.27

4.3 The Normal Distribution

- **Nonstandard Normal Distribution**

If X has the normal distribution with mean μ and standard deviation σ, then

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

has a standard normal distribution (why?).

Relationship between Nonstandard Distribution and Normal Distribution

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

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

If we let $\frac{x-\mu}{\sigma} = t,\ x = \mu + \sigma t,\ dx = \sigma dt$

Then F(x) $= \frac{1}{\sqrt{2\pi}}\int_{-\infty}^{\frac{x-\mu}{\sigma}} e^{-\frac{t^2}{2}} dt$

$$= \Phi\left(\frac{x-\mu}{\sigma}\right)$$

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

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

Proposition

If X has the normal distribution with mean μ and standard deviation σ, then

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

has a standard normal distribution. Thus

$$P(a \le X \le b) = P\left(\frac{a-\mu}{\sigma} \le Z \le \frac{b-\mu}{\sigma}\right) = \Phi\left(\frac{b-\mu}{\sigma}\right) - \Phi\left(\frac{a-\mu}{\sigma}\right)$$

$$P(X \le a) = \Phi\left(\frac{a-\mu}{\sigma}\right) \qquad P(X \ge b) = 1 - \Phi\left(\frac{b-\mu}{\sigma}\right)$$

Example

Given $X \sim N(3,2^2)$ **, find** $P(X>2)$

Solution:

If X>2, then $Y=\frac{X-\mu}{\sigma}=\frac{X-3}{2}>-\frac{1}{2}$

$$P(X>2)=P(Y>-\frac{1}{2})$$

$$Y \sim N(0,1)$$

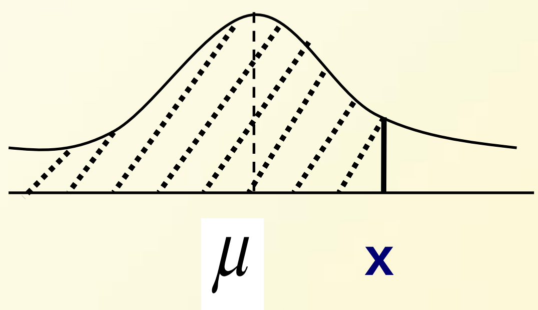
$$=1-P(Y\le -\frac{1}{2})=1-\Phi(-\frac{1}{2})$$

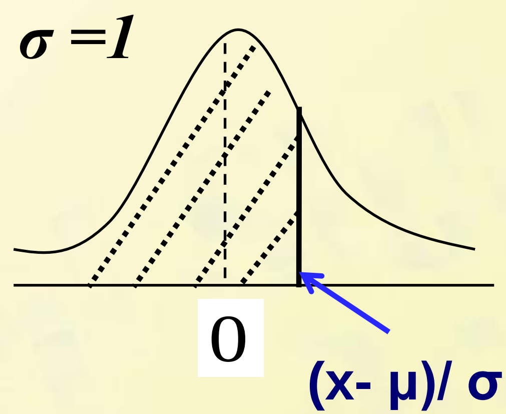
$$=\Phi(\frac{1}{2})=0.6915$$

4.3 The Normal Distribution

- Equality of nonstandard and standard normal curve area

$$P(Z \le z) = P\left(X \le \sigma z + \mu\right) = \int_{-\infty}^{\sigma z+\mu} f(x;\mu,\sigma)dx$$





Percentiles of an Arbitrary Normal Distribution

$$(100p)\text{th percentile for normal } (\mu,\sigma)$$
$$= \mu + \left[(100p)\text{th for standard normal}\right]\cdot\sigma$$

Refer to Ex. 4.17

- **Example 4.16**

The time that it takes a driver to react to the brake lights on a decelerating vehicle is critical in helping to avoid rear-end collisions . Reaction time for an in-traffic response to a brake signal from standard brake lights can be modeled with a **normal distribution** having mean value 1.25 sec and standard deviation of .46 sec . What is the probability that reaction time is between 1.00 sec and 1.75 sec?

4.3 The Normal Distribution

Solution:

$$P(1.00 \le X \le 1.75) = P\left(\frac{1.00-1.25}{0.46} \le Z \le \frac{1.75-1.25}{0.46}\right)$$

$$= \Phi(1.09) - \Phi(-0.54)$$

$$= 0.8621 - 2.946 = 0.5675$$

- **Example 4.17**

  The breakdown voltage of a randomly chosen diode of a particular type is known to be normally distributed. What is the probability that a diode's breakdown voltage is within 1 standard deviation of its mean value?

4.3 The Normal Distribution

Solution:

$$P\left(X \text{ is within 1 standard deviation of its mean }\right)$$

$$= P\left(\mu - \sigma \le X \le \mu + \sigma\right) = P\left(\frac{\mu - \sigma - \mu}{\sigma} \le Z \le \frac{\mu + \sigma - \mu}{\sigma}\right)$$

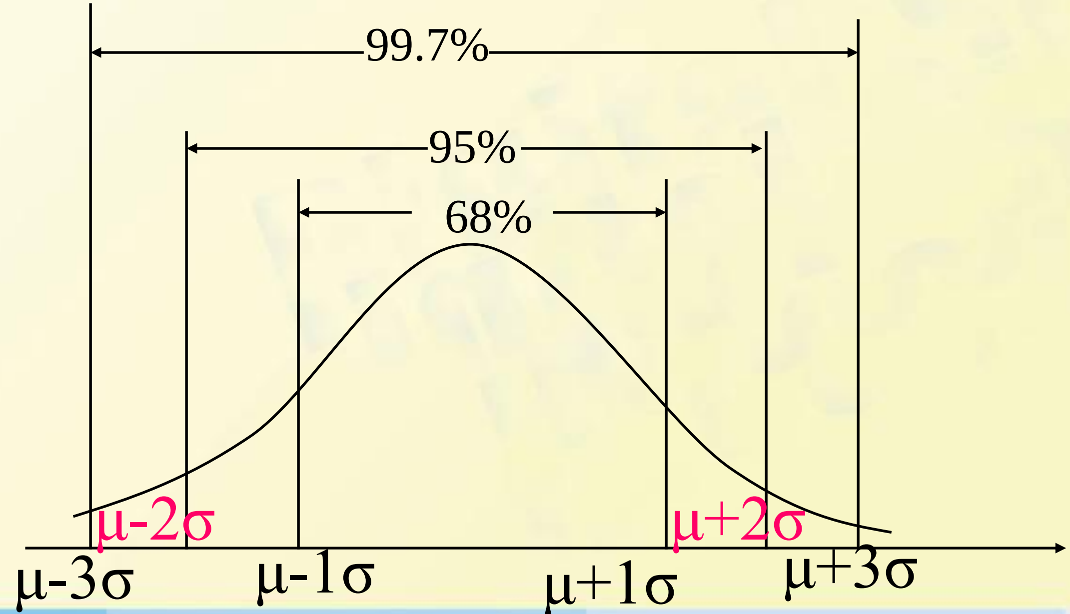
$$= P\left(-1.00 \le Z \le 1.00\right) = \Phi\left(1.00\right) - \Phi\left(-1.00\right) = 0.6826$$

Note: This question can be answered without knowing either μ or σ, as long as the distribution is known to be normal; in other words , the answer is the same for any normal distribution:

4.3 The Normal Distribution

If the population distribution of a variable is (approximately) normal, then

1. Roughly **68%** of the values are within **1 SD** of the mean.
2. Roughly **95%** of the values are within **2 SDs** of the mean
3. Roughly **99.7%** of the values are within **3 SDs** of the mean



4.3 The Normal Distribution

- **The Normal Distribution and Discrete Populations**

The normal distribution is often used as an approximation to the distribution of values in a discrete population.

Ex. 4.19:

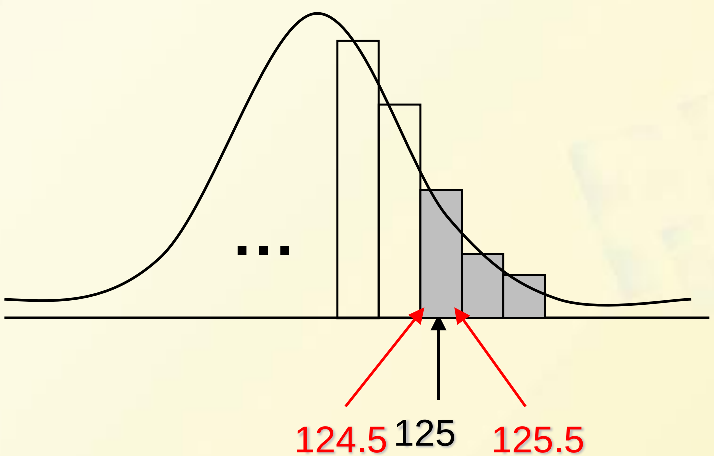
IQ in a particular population is known to be approximately normally distributed with $\mu = 100$ and $\sigma = 15$. What is the probability that a randomly selected individual has an IQ of at least 125?

The Normal Distribution and Discrete Populations

Solution:

Letting X = the IQ of a randomly chosen person, we wish $P(X \geq 125)$.

The temptation here is to standardize $X \geq 125$ immediately as in previous example. However, the IQ population is actually discrete, since IQs are integer-valued, **so the normal curve is an approximation to a discrete** probability histogram,



continuity correction

$$P(X \geq 125) = P(Z \geq [(125-0.5)-100)]/15)$$
$$= P(Z \geq 1.63) = 0.0516$$

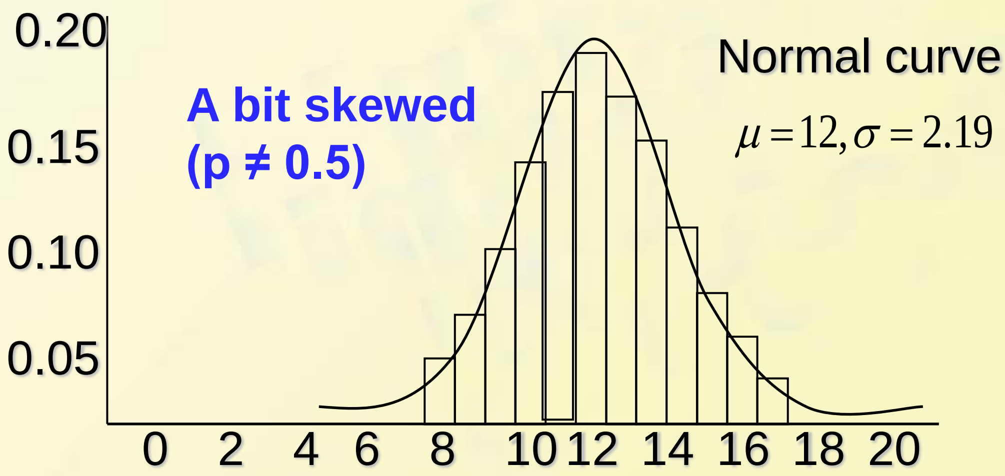
$$P(X = 125)$$
$$= P([(125-0.5)-100]/15 \leq Z \leq [(125+0.5)-100]/15)$$
$$= P(1.63 \leq Z \leq 1.7) \quad \neq 0$$

4.3 The Normal Distribution

- **The Normal Approximation to the Binomial Distribution**

Recall that the mean value and standard deviation of a binomial random variable X are $\mu_X = np$ and $\sigma_X = (npq)^{1/2}$.

Consider the binomial probability histogram with $n = 20$, $p = 0.6$. It can be approximated by the normal curve with $\mu = 12$ and $\sigma = 2.19$ as follows.



- **Proposition**

Let X be a binominal rv based on n trials with success probability p. Then if the binomial probability histogram is not too skewed, X has approximately a normal distribution with $\boldsymbol{\mu = np}$ and $\boldsymbol{\sigma_X = (npq)^{1/2}}$.

In particular, for x = a possible value of X ,

$$
\begin{aligned}
&p(X \le x) = B(x; n, p) \\
&\approx (\text{area under the normal curve to the left of } x + 0.5) \\
&= \Phi\left(\frac{x + 0.5 - np}{\sqrt{npq}}\right)
\end{aligned}
$$

Rule: In practice, the approximation is adequate provided that both *np≥10* and *nq ≥10*. (where *q=1-p*)

- **Example 4.20**

Suppose that 25% of all licensed drivers in a particular state do not have insurance. Let X be the number of uninsured drivers in a random sample of size 50, so that p=0.25.

Find: (A) $P(X \leq 10)$

(B) $P(5 \leq X \leq 15)$

Solution:

(A) Since np=50(0.25)=12.5≥10 and nq=37.5 ≥ 10, **the approximation can safely be applied**. Then $\mu = 12.5$ and $\sigma = 3.06$.

$$P(X \le 10) = B(10; 50, 0.25) \approx \Phi\left(\frac{10 + 0.5 - 12.5}{3.06}\right)$$

$$= \Phi(-0.65) = 0.2578$$

(B) Similarly , the probability that between 5 and 15 (inclusive) of the selected drivers are uninsured is

$$P(5 \le X \le 15) = B(15; 50, 0.25) - B(4; 50, 0.25)$$

$$\approx \Phi\left(\frac{15.5 - 12.5}{3.06}\right) - \Phi\left(\frac{4.5 - 12.5}{3.06}\right) = 0.8320$$