The **normal probability distribution** is the most important distribution for describing a continuous random variable.

- It is widely used in statistical inference.
Normal Probability Distribution

- $\mu - 3\sigma$ to $\mu + 3\sigma$: 99.72%
- $\mu - 2\sigma$ to $\mu + 2\sigma$: 95.44%
- $\mu - 1\sigma$ to $\mu + 1\sigma$: 68.26%
Standard Normal Probability Distribution

Converting to the Standard Normal Distribution

\[ z = \frac{x - \mu}{\sigma} \]
**Standard Normal Probability Distribution**

- **Standard Normal Density Function**

\[
f(x) = \frac{1}{\sqrt{2\pi}} e^{-z^2/2}
\]

where:

\[
z = \frac{x - \mu}{\sigma}
\]

\[
\pi = 3.14159
\]

\[
e = 2.71828
\]
Standard Normal Probability Distribution

Toko oliku menjual oli kendaraan bermotor. Ketika stok oli berada di bawah 20 gallon, dilakukan pemesanan ulang stok.

Manajer Toko tidak menginginkan kehilangan peluang penjualan akibat kehabisan stok selama menunggu kiriman barang. Dari data penjualan diketahui bahwa jumlah permintaan selama waktu pengisian kembali stok terdistribusi normal dengan mean 15 gallon dan standar deviasi 6 gallon.

Manajer tersebut ingin mengetahui probabilita terjadinya stockout, P(x > 20).
Standard Normal Probability Distribution

Tabel Normal Standar memperlihatkan luas daerah sebesar 0.2967 untuk wilayah antara \( z = 0 \) dan \( z = 0.83 \). Wilayah yang diarsir adalah \( 0.5 - 0.2967 = 0.2033 \). Probabilita kehabisan stok adalah 0.2033.

\[
z = \frac{x - \mu}{\sigma} = \frac{20 - 15}{6} = 0.83
\]
Menggunakan Standard Normal Probability Table

<table>
<thead>
<tr>
<th>z</th>
<th>.00</th>
<th>.01</th>
<th>.02</th>
<th>.03</th>
<th>.04</th>
<th>.05</th>
<th>.06</th>
<th>.07</th>
<th>.08</th>
<th>.09</th>
</tr>
</thead>
<tbody>
<tr>
<td>.0</td>
<td>.0000</td>
<td>.0040</td>
<td>.0080</td>
<td>.0120</td>
<td>.0160</td>
<td>.0199</td>
<td>.0239</td>
<td>.0279</td>
<td>.0319</td>
<td>.0359</td>
</tr>
<tr>
<td>.1</td>
<td>.0398</td>
<td>.0438</td>
<td>.0478</td>
<td>.0517</td>
<td>.0557</td>
<td>.0596</td>
<td>.0636</td>
<td>.0675</td>
<td>.0714</td>
<td>.0753</td>
</tr>
<tr>
<td>.2</td>
<td>.0793</td>
<td>.0832</td>
<td>.0871</td>
<td>.0910</td>
<td>.0948</td>
<td>.0987</td>
<td>.1026</td>
<td>.1064</td>
<td>.1103</td>
<td>.1141</td>
</tr>
<tr>
<td>.3</td>
<td>.1179</td>
<td>.1217</td>
<td>.1255</td>
<td>.1293</td>
<td>.1331</td>
<td>.1368</td>
<td>.1406</td>
<td>.1443</td>
<td>.1480</td>
<td>.1517</td>
</tr>
<tr>
<td>.4</td>
<td>.1554</td>
<td>.1591</td>
<td>.1628</td>
<td>.1664</td>
<td>.1700</td>
<td>.1736</td>
<td>.1772</td>
<td>.1808</td>
<td>.1844</td>
<td>.1879</td>
</tr>
<tr>
<td>.6</td>
<td>.2257</td>
<td>.2291</td>
<td>.2324</td>
<td>.2357</td>
<td>.2389</td>
<td>.2422</td>
<td>.2454</td>
<td>.2486</td>
<td>.2518</td>
<td>.2549</td>
</tr>
<tr>
<td>.7</td>
<td>.2580</td>
<td>.2612</td>
<td>.2642</td>
<td>.2673</td>
<td>.2704</td>
<td>.2734</td>
<td>.2764</td>
<td>.2794</td>
<td>.2823</td>
<td>.2852</td>
</tr>
<tr>
<td>.8</td>
<td>.2881</td>
<td>.2910</td>
<td>.2939</td>
<td>.2967</td>
<td>.2995</td>
<td>.3023</td>
<td>.3051</td>
<td>.3078</td>
<td>.3106</td>
<td>.3133</td>
</tr>
<tr>
<td>.9</td>
<td>.3159</td>
<td>.3186</td>
<td>.3212</td>
<td>.3238</td>
<td>.3264</td>
<td>.3289</td>
<td>.3315</td>
<td>.3340</td>
<td>.3365</td>
<td>.3389</td>
</tr>
</tbody>
</table>
Standard Normal Probability Distribution

Jika manajer oliku menginginkan probabilita kehabisan stock tidak lebih dari 0.05, berapa seharusnya titik pemesanan kembali?

Misalkan \( z_{0.05} \) menggambarkan nilai \( z \) value batas area 0.05.
Using the Standard Normal Probability Table

Perhatikan luas daerah 0.4500 dalam Standard Normal Probability table untuk menentukan nilai $z_{0.05}$ yang sesuai.

<table>
<thead>
<tr>
<th>$z$</th>
<th>.00</th>
<th>.01</th>
<th>.02</th>
<th>.03</th>
<th>.04</th>
<th>.05</th>
<th>.06</th>
<th>.07</th>
<th>.08</th>
<th>.09</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1.5</td>
<td>.4332</td>
<td>.4345</td>
<td>.4357</td>
<td>.4370</td>
<td>.4382</td>
<td>.4394</td>
<td>.4406</td>
<td>.4418</td>
<td>.4429</td>
<td>.4441</td>
</tr>
<tr>
<td>1.6</td>
<td>.4452</td>
<td>.4463</td>
<td>.4474</td>
<td>.4484</td>
<td>.4495</td>
<td>.4505</td>
<td>.4515</td>
<td>.4525</td>
<td>.4535</td>
<td>.4545</td>
</tr>
<tr>
<td>1.7</td>
<td>.4554</td>
<td>.4564</td>
<td>.4573</td>
<td>.4582</td>
<td>.4591</td>
<td>.4599</td>
<td>.4608</td>
<td>.4616</td>
<td>.4625</td>
<td>.4633</td>
</tr>
<tr>
<td>1.8</td>
<td>.4641</td>
<td>.4649</td>
<td>.4656</td>
<td>.4664</td>
<td>.4671</td>
<td>.4678</td>
<td>.4686</td>
<td>.4693</td>
<td>.4699</td>
<td>.4706</td>
</tr>
<tr>
<td>1.9</td>
<td>.4713</td>
<td>.4719</td>
<td>.4726</td>
<td>.4732</td>
<td>.4738</td>
<td>.4744</td>
<td>.4750</td>
<td>.4756</td>
<td>.4761</td>
<td>.4767</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
Standard Normal Probability Distribution

Nilai \( x \) yang sesuai dapat dihitung sebagai berikut:

\[
\begin{align*}
x &= \mu + z_{0.05} \sigma \\
&= 15 + 1.645(6) \\
&= 24.87
\end{align*}
\]

Titik pemesanan kembali sebesar 24.87 gallon memiliki probabilita untuk terjadi kehabisan stok selama waktu pengisian kembali sebesar 0.05.

Dengan demikian oliku sebaiknya menentukan titik pengisian kembali pada 25 gallon untuk menjaga probabilita kehabisan stok sebesar 0.05.
Sampling and Sampling Distributions

- Simple Random Sampling
- Point Estimation
- Introduction to Sampling Distributions
- Sampling Distribution of $\bar{x}$
The purpose of statistical inference is to obtain information about a population from information contained in a sample.

A population is the set of all the elements of interest.

A sample is a subset of the population.

With proper sampling methods, the sample results can provide “good” estimates of the population characteristics.

A parameter is a numerical characteristic of a population.
Simple Random Sampling: Finite Population

- **Finite populations** are often defined by lists such as:
  - Organization membership roster
  - Credit card account numbers
  - Inventory product numbers

- A **simple random sample of size $n$ from a finite population of size $N$** is a sample selected such that each possible sample of size $n$ has the same probability of being selected.
Infinite populations are often defined by an ongoing process whereby the elements of the population consist of items generated as though the process would operate indefinitely.

A simple random sample from an infinite population is a sample selected such that the following conditions are satisfied.

- Each element selected comes from the same population.
- Each element is selected independently.
In point estimation we use the data from the sample to compute a value of a sample statistic that serves as an estimate of a population parameter.

We refer to $\bar{x}$ as the point estimator of the population mean $\mu$.

$s$ is the point estimator of the population standard deviation $\sigma$.

$p$ is the point estimator of the population proportion $p$. 
When the expected value of a point estimator is equal to the population parameter, the point estimator is said to be unbiased.

The absolute value of the difference between an unbiased point estimate and the corresponding population parameter is called the sampling error.

The sampling errors are:

- $|\bar{x} - \mu|$ for sample mean
- $|s - \sigma|$ for sample standard deviation
- $|\bar{p} - p|$ for sample proportion
IBS receives 900 applications annually from prospective students. The application form contains a variety of information including the individual’s scholastic aptitude test (SAT) score.
Conducting a Census

- Population Mean SAT Score
  \[ \mu = \frac{\sum x_i}{900} = 990 \]

- Population Standard Deviation for SAT Score
  \[ \sigma = \sqrt{\frac{\sum (x_i - \mu)^2}{900}} = 80 \]

- Population Proportion Wanting On-Campus Housing
  \[ p = \frac{648}{900} = .72 \]
Point Estimation

- \( \bar{x} \) as Point Estimator of \( \mu \)
  \[
  \bar{x} = \frac{\sum x_i}{30} = \frac{29,910}{30} = 997
  \]

- \( s \) as Point Estimator of \( \sigma \)
  \[
  s = \sqrt{\frac{\sum (x_i - \bar{x})^2}{29}} = \sqrt{\frac{163,996}{29}} = 75.2
  \]

- \( \bar{p} \) as Point Estimator of \( p \)
  \[
  \bar{p} = \frac{20}{30} = .68
  \]

Note: Different random numbers would have identified a different sample which would have resulted in different point estimates.
### Population Parameters and Point Estimates

<table>
<thead>
<tr>
<th>Population Parameter</th>
<th>Parameter Value</th>
<th>Point Estimator</th>
<th>Point Estimate</th>
</tr>
</thead>
<tbody>
<tr>
<td>( \mu ) = Population mean SAT score</td>
<td>990</td>
<td>( \overline{x} ) = Sample mean SAT score</td>
<td>997</td>
</tr>
<tr>
<td>( \sigma ) = Population std. deviation for SAT score</td>
<td>80</td>
<td>( s ) = Sample std. deviation for SAT score</td>
<td>75.2</td>
</tr>
<tr>
<td>( p ) = Population proportion wanting campus housing</td>
<td>.72</td>
<td>( \overline{p} ) = Sample proportion wanting campus housing</td>
<td>.68</td>
</tr>
</tbody>
</table>
Process of Statistical Inference

Population with mean $\mu = ?$

A simple random sample of $n$ elements is selected from the population.

The value of $\bar{x}$ is used to make inferences about the value of $\mu$.

The sample data provide a value for the sample mean $\bar{x}$.
The sampling distribution of $\bar{x}$ is the probability distribution of all possible values of the sample mean $\bar{x}$.

Expected Value of $\bar{x}$

$E(\bar{x}) = \mu$

where:

$\mu$ = the population mean
Standard Deviation of $\bar{x}$

**Finite Population**

$\sigma_{\bar{x}} = \left( \frac{\sigma}{\sqrt{n}} \right) \sqrt{\frac{N-n}{N-1}}$

**Infinite Population**

$\sigma_{\bar{x}} = \frac{\sigma}{\sqrt{n}}$

- A finite population is treated as being infinite if $n/N \leq .05$.
- $\sqrt{(N-n)/(N-1)}$ is the finite correction factor.
- $\sigma_{\bar{x}}$ is referred to as the standard error of the mean.
If we use a large \((n \geq 30)\) simple random sample, the \textbf{central limit theorem} enables us to conclude that the sampling distribution of \( \bar{x} \) can be approximated by a normal distribution.

When the simple random sample is small \((n < 30)\), the sampling distribution of \( \bar{x} \) can be considered normal only if we assume the population has a normal distribution.
Sampling Distribution of $\bar{x}$ for SAT Scores

$E(\bar{x}) = 990$

$\sigma_{\bar{x}} = \frac{\sigma}{\sqrt{n}} = \frac{80}{\sqrt{30}} = 14.6$
What is the probability that a simple random sample of 30 applicants will provide an estimate of the population mean SAT score that is within $+/-10$ of the actual population mean $\mu$?

In other words, what is the probability that $\bar{x}$ will be between 980 and 1000?
Step 1: Calculate the z-value at the upper endpoint of the interval.

\[ z = \frac{(1000 - 990)}{14.6} = 0.68 \]

Step 2: Find the area under the curve to the left of the upper endpoint.

\[ P(z \leq 0.68) = 0.7517 \]
### Sampling Distribution of $\bar{x}$ for SAT Scores

#### Cumulative Probabilities for the Standard Normal Distribution

<table>
<thead>
<tr>
<th>$z$</th>
<th>.00</th>
<th>.01</th>
<th>.02</th>
<th>.03</th>
<th>.04</th>
<th>.05</th>
<th>.06</th>
<th>.07</th>
<th>.08</th>
<th>.09</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>.5</td>
<td>.6915</td>
<td>.6950</td>
<td>.6985</td>
<td>.7019</td>
<td>.7054</td>
<td>.7088</td>
<td>.7123</td>
<td>.7157</td>
<td>.7190</td>
<td>.7224</td>
</tr>
<tr>
<td>.6</td>
<td>.7257</td>
<td>.7291</td>
<td>.7324</td>
<td>.7357</td>
<td>.7389</td>
<td>.7422</td>
<td>.7454</td>
<td>.7486</td>
<td>.7517</td>
<td>.7549</td>
</tr>
<tr>
<td>.7</td>
<td>.7580</td>
<td>.7611</td>
<td>.7642</td>
<td>.7673</td>
<td>.7704</td>
<td>.7734</td>
<td>.7764</td>
<td>.7794</td>
<td>.7823</td>
<td>.7852</td>
</tr>
<tr>
<td>.8</td>
<td>.7881</td>
<td>.7910</td>
<td>.7939</td>
<td>.7967</td>
<td>.7995</td>
<td>.8023</td>
<td>.8051</td>
<td>.8078</td>
<td>.8106</td>
<td>.8133</td>
</tr>
<tr>
<td>.9</td>
<td>.8159</td>
<td>.8186</td>
<td>.8212</td>
<td>.8238</td>
<td>.8264</td>
<td>.8289</td>
<td>.8315</td>
<td>.8340</td>
<td>.8365</td>
<td>.8389</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
Sampling Distribution of $\bar{x}$ for SAT Scores

$\sigma_{\bar{x}} = 14.6$

Area = 0.7517

990 1000

$\bar{x}$
Step 3: Calculate the z-value at the lower endpoint of the interval.

\[ z = \frac{(980 - 990)}{14.6} = -0.68 \]

Step 4: Find the area under the curve to the left of the lower endpoint.

\[ P(z \leq -0.68) = P(z \geq 0.68) \]

\[ = 1 - P(z \leq 0.68) \]

\[ = 1 - 0.7517 \]

\[ = 0.2483 \]
Sampling Distribution of $\bar{x}$ for SAT Scores

$\sigma_{\bar{x}} = 14.6$

Area = 0.2483

980 990
Step 5: Calculate the area under the curve between the lower and upper endpoints of the interval.

\[ P(-.68 < z < .68) = P(z \leq .68) - P(z \leq -.68) \]
\[ = .7517 - .2483 \]
\[ = .5034 \]

The probability that the sample mean SAT score will be between 980 and 1000 is:

\[ P(980 \leq \bar{x} \leq 1000) = .5034 \]
Sampling Distribution of $\bar{x}$ for SAT Scores

$\sigma_{\bar{x}} = 14.6$

Area = 0.5034

$P(980 \leq \bar{x} \leq 1000) = 0.5034$
Other Sampling Methods

- Stratified Random Sampling
- Cluster Sampling
- Systematic Sampling
- Convenience Sampling
- Judgment Sampling
Thank You