

Statistical and Mathematical Methods for Data Analysis

Dr. Syed Faisal Bukhari

Associate Professor

Department of Data Science

Faculty of Computing and Information Technology

University of the Punjab

Textbook

- **Probability & Statistics for Engineers & Scientists,**
Ninth Edition, Ronald E. Walpole, Raymond H.
Myer

Reference books

- ❑ **Probability Demystified**, Allan G. Bluman
- ❑ **Schaum's Outline of Probability and Statistics**
- ❑ **MATLAB Primer**, Seventh Edition
- ❑ **MATLAB Demystified** by McMahan, David

References

Readings for these lecture notes:

- ❑ **Schaum's Outline of Probability, Second Edition (Schaum's Outlines)** by by Seymour Lipschutz, Marc Lipson
- ❑ **Probability & Statistics for Engineers & Scientists**, Ninth Edition, Ronald E. Walpole, Raymond H. Myer
- ❑ <https://wordwatchtowers.wordpress.com/2009/12/21/underestimate-or-overestimate/>
- ❑ Elementary Statistics, Tenth Edition, Mario F. Triola
- ❑ <http://www.sjsu.edu/faculty/gerstman/>

These notes contain material from the above resources.

Sampling Distribution of Means and the Central Limit Theorem

$$\bar{X} = \frac{(x_1 + x_2 + x_3 + \dots + x_n)}{n}$$

has a normal distribution with **mean**

$$\begin{aligned}\mu_{\bar{X}} &= \frac{1}{n} (\mu + \mu + \mu + \dots + \mu) \text{ (n terms)} \\ &= \frac{n\mu}{n}\end{aligned}$$

$$\mu_{\bar{X}} = \mu$$

Sampling Distribution of Means and the Central Limit Theorem

and variance

$$\sigma_{\bar{X}}^2 = \frac{1}{n^2} (\sigma^2 + \sigma^2 + \sigma^2 + \dots + \sigma^2) \text{ (n terms)}$$

$$= \frac{n\sigma^2}{n^2}$$

$$\sigma_{\bar{X}}^2 = \frac{\sigma^2}{n}$$



Table A.3 Areas under the Normal Curve

z	.00	.01	.02	.03	.04	.05	.06	.07	.08	.09
-3.4	0.0003	0.0003	0.0003	0.0003	0.0003	0.0003	0.0003	0.0003	0.0003	0.0002
-3.3	0.0005	0.0005	0.0005	0.0004	0.0004	0.0004	0.0004	0.0004	0.0004	0.0003
-3.2	0.0007	0.0007	0.0006	0.0006	0.0006	0.0006	0.0006	0.0005	0.0005	0.0005
-3.1	0.0010	0.0009	0.0009	0.0009	0.0008	0.0008	0.0008	0.0008	0.0007	0.0007
-3.0	0.0013	0.0013	0.0013	0.0012	0.0012	0.0011	0.0011	0.0011	0.0010	0.0010
-2.9	0.0019	0.0018	0.0018	0.0017	0.0016	0.0016	0.0015	0.0015	0.0014	0.0014
-2.8	0.0026	0.0025	0.0024	0.0023	0.0023	0.0022	0.0021	0.0021	0.0020	0.0019
-2.7	0.0035	0.0034	0.0033	0.0032	0.0031	0.0030	0.0029	0.0028	0.0027	0.0026
-2.6	0.0047	0.0045	0.0044	0.0043	0.0041	0.0040	0.0039	0.0038	0.0037	0.0036
-2.5	0.0062	0.0060	0.0059	0.0057	0.0055	0.0054	0.0052	0.0051	0.0049	0.0048
-2.4	0.0082	0.0080	0.0078	0.0075	0.0073	0.0071	0.0069	0.0068	0.0066	0.0064
-2.3	0.0107	0.0104	0.0102	0.0099	0.0096	0.0094	0.0091	0.0089	0.0087	0.0084
-2.2	0.0139	0.0136	0.0132	0.0129	0.0125	0.0122	0.0119	0.0116	0.0113	0.0110
-2.1	0.0179	0.0174	0.0170	0.0166	0.0162	0.0158	0.0154	0.0150	0.0146	0.0143
-2.0	0.0228	0.0222	0.0217	0.0212	0.0207	0.0202	0.0197	0.0192	0.0188	0.0183
-1.9	0.0287	0.0281	0.0274	0.0268	0.0262	0.0256	0.0250	0.0244	0.0239	0.0233
-1.8	0.0359	0.0351	0.0344	0.0336	0.0329	0.0322	0.0314	0.0307	0.0301	0.0294
-1.7	0.0446	0.0436	0.0427	0.0418	0.0409	0.0401	0.0392	0.0384	0.0375	0.0367
-1.6	0.0548	0.0537	0.0526	0.0516	0.0505	0.0495	0.0485	0.0475	0.0465	0.0455
-1.5	0.0668	0.0655	0.0643	0.0630	0.0618	0.0606	0.0594	0.0582	0.0571	0.0559
-1.4	0.0808	0.0793	0.0778	0.0764	0.0749	0.0735	0.0721	0.0708	0.0694	0.0681
-1.3	0.0968	0.0951	0.0934	0.0918	0.0901	0.0885	0.0869	0.0853	0.0838	0.0823
-1.2	0.1151	0.1131	0.1112	0.1093	0.1075	0.1056	0.1038	0.1020	0.1003	0.0985
-1.1	0.1357	0.1335	0.1314	0.1292	0.1271	0.1251	0.1230	0.1210	0.1190	0.1170
-1.0	0.1587	0.1562	0.1539	0.1515	0.1492	0.1469	0.1446	0.1423	0.1401	0.1379
-0.9	0.1841	0.1814	0.1788	0.1762	0.1736	0.1711	0.1685	0.1660	0.1635	0.1611
-0.8	0.2119	0.2090	0.2061	0.2033	0.2005	0.1977	0.1949	0.1922	0.1894	0.1867
-0.7	0.2420	0.2389	0.2358	0.2327	0.2296	0.2266	0.2236	0.2206	0.2177	0.2148
-0.6	0.2743	0.2709	0.2676	0.2643	0.2611	0.2578	0.2546	0.2514	0.2483	0.2451
-0.5	0.3085	0.3050	0.3015	0.2981	0.2946	0.2912	0.2877	0.2843	0.2810	0.2776
-0.4	0.3446	0.3409	0.3372	0.3336	0.3300	0.3264	0.3228	0.3192	0.3156	0.3121
-0.3	0.3821	0.3783	0.3745	0.3707	0.3669	0.3632	0.3594	0.3557	0.3520	0.3483
-0.2	0.4207	0.4168	0.4129	0.4090	0.4052	0.4013	0.3974	0.3936	0.3897	0.3859
-0.1	0.4602	0.4562	0.4522	0.4483	0.4443	0.4404	0.4364	0.4325	0.4286	0.4247
-0.0	0.5000	0.4960	0.4920	0.4880	0.4840	0.4801	0.4761	0.4721	0.4681	0.4641

Area under the Normal Curve [2]

Table A.3 (continued) Areas under the Normal Curve

<i>z</i>	.00	.01	.02	.03	.04	.05	.06	.07	.08	.09
0.0	0.5000	0.5040	0.5080	0.5120	0.5160	0.5199	0.5239	0.5279	0.5319	0.5359
0.1	0.5398	0.5438	0.5478	0.5517	0.5557	0.5596	0.5636	0.5675	0.5714	0.5753
0.2	0.5793	0.5832	0.5871	0.5910	0.5948	0.5987	0.6026	0.6064	0.6103	0.6141
0.3	0.6179	0.6217	0.6255	0.6293	0.6331	0.6368	0.6406	0.6443	0.6480	0.6517
0.4	0.6554	0.6591	0.6628	0.6664	0.6700	0.6736	0.6772	0.6808	0.6844	0.6879
0.5	0.6915	0.6950	0.6985	0.7019	0.7054	0.7088	0.7123	0.7157	0.7190	0.7224
0.6	0.7257	0.7291	0.7324	0.7357	0.7389	0.7422	0.7454	0.7486	0.7517	0.7549
0.7	0.7580	0.7611	0.7642	0.7673	0.7704	0.7734	0.7764	0.7794	0.7823	0.7852
0.8	0.7881	0.7910	0.7939	0.7967	0.7995	0.8023	0.8051	0.8078	0.8106	0.8133
0.9	0.8159	0.8186	0.8212	0.8238	0.8264	0.8289	0.8315	0.8340	0.8365	0.8389
1.0	0.8413	0.8438	0.8461	0.8485	0.8508	0.8531	0.8554	0.8577	0.8599	0.8621
1.1	0.8643	0.8665	0.8686	0.8708	0.8729	0.8749	0.8770	0.8790	0.8810	0.8830
1.2	0.8849	0.8869	0.8888	0.8907	0.8925	0.8944	0.8962	0.8980	0.8997	0.9015
1.3	0.9032	0.9049	0.9066	0.9082	0.9099	0.9115	0.9131	0.9147	0.9162	0.9177
1.4	0.9192	0.9207	0.9222	0.9236	0.9251	0.9265	0.9279	0.9292	0.9306	0.9319
1.5	0.9332	0.9345	0.9357	0.9370	0.9382	0.9394	0.9406	0.9418	0.9429	0.9441
1.6	0.9452	0.9463	0.9474	0.9484	0.9495	0.9505	0.9515	0.9525	0.9535	0.9545
1.7	0.9554	0.9564	0.9573	0.9582	0.9591	0.9599	0.9608	0.9616	0.9625	0.9633
1.8	0.9641	0.9649	0.9656	0.9664	0.9671	0.9678	0.9686	0.9693	0.9699	0.9706
1.9	0.9713	0.9719	0.9726	0.9732	0.9738	0.9744	0.9750	0.9756	0.9761	0.9767
2.0	0.9772	0.9778	0.9783	0.9788	0.9793	0.9798	0.9803	0.9808	0.9812	0.9817
2.1	0.9821	0.9826	0.9830	0.9834	0.9838	0.9842	0.9846	0.9850	0.9854	0.9857
2.2	0.9861	0.9864	0.9868	0.9871	0.9875	0.9878	0.9881	0.9884	0.9887	0.9890
2.3	0.9893	0.9896	0.9898	0.9901	0.9904	0.9906	0.9909	0.9911	0.9913	0.9916
2.4	0.9918	0.9920	0.9922	0.9925	0.9927	0.9929	0.9931	0.9932	0.9934	0.9936
2.5	0.9938	0.9940	0.9941	0.9943	0.9945	0.9946	0.9948	0.9949	0.9951	0.9952
2.6	0.9953	0.9955	0.9956	0.9957	0.9959	0.9960	0.9961	0.9962	0.9963	0.9964
2.7	0.9965	0.9966	0.9967	0.9968	0.9969	0.9970	0.9971	0.9972	0.9973	0.9974
2.8	0.9974	0.9975	0.9976	0.9977	0.9977	0.9978	0.9979	0.9979	0.9980	0.9981
2.9	0.9981	0.9982	0.9982	0.9983	0.9984	0.9984	0.9985	0.9985	0.9986	0.9986
3.0	0.9987	0.9987	0.9987	0.9988	0.9988	0.9989	0.9989	0.9989	0.9990	0.9990
3.1	0.9990	0.9991	0.9991	0.9991	0.9992	0.9992	0.9992	0.9992	0.9993	0.9993
3.2	0.9993	0.9993	0.9994	0.9994	0.9994	0.9994	0.9994	0.9995	0.9995	0.9995
3.3	0.9995	0.9995	0.9995	0.9996	0.9996	0.9996	0.9996	0.9996	0.9996	0.9997
3.4	0.9997	0.9997	0.9997	0.9997	0.9997	0.9997	0.9997	0.9997	0.9997	0.9998

Example : An electrical firm manufactures light bulbs that have a length of life that is approximately normally distributed, with **mean** equal to **800 hours** and a **standard deviation** of **40 hours**. Find the probability that a random sample of **16 bulbs** will have an **average life** of less than **775 hours**.

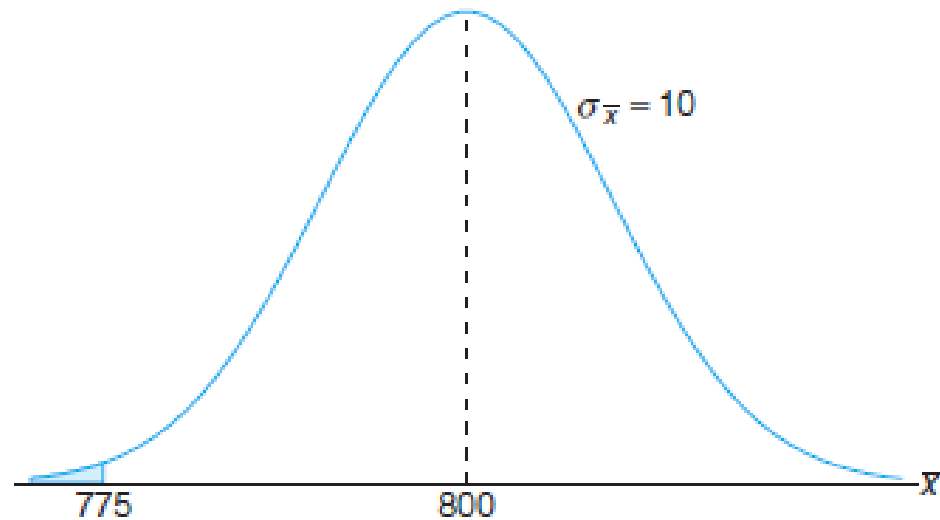
Solution : The sampling distribution of \bar{X} will be approximately normal, with $\mu_{\bar{X}} = 800$ and

$$\begin{aligned}\sigma_{\bar{X}} &= \sigma / \sqrt{n} \\ &= 40 / \sqrt{16}\end{aligned}$$

$$\sigma_{\bar{X}} = 10$$

$$\begin{aligned}Z &= \frac{\bar{X} - \mu}{\sigma / \sqrt{n}} \\ &= \frac{775 - 800}{10} \\ &= -2.5\end{aligned}$$

$$\begin{aligned}P(\bar{X} < 775) &= P(Z < -2.5) \\ &= 0.0062.\end{aligned}$$



```
from scipy.stats import norm
```

Given data

```
population_mean = 800  # Mean of the population
population_std = 40     # Standard deviation of the population
sample_size = 16       # Sample size
sample_mean = 775      # Sample mean
```

Calculate the standard error of the mean (SEM)

```
sem = population_std / (sample_size ** 0.5)
```

Calculate the z-score

```
z_score = (sample_mean - population_mean) / sem
```

Calculate the probability using the cumulative distribution function (CDF)

```
probability = norm.cdf(z_score)
```

```
print(f"The probability that a random sample of 16 bulbs will have  
an average life of less than 775 hours is: {probability:.4f}")
```

The probability that a random sample of 16 bulbs will have an average life of less than 775 hours is: 0.0062

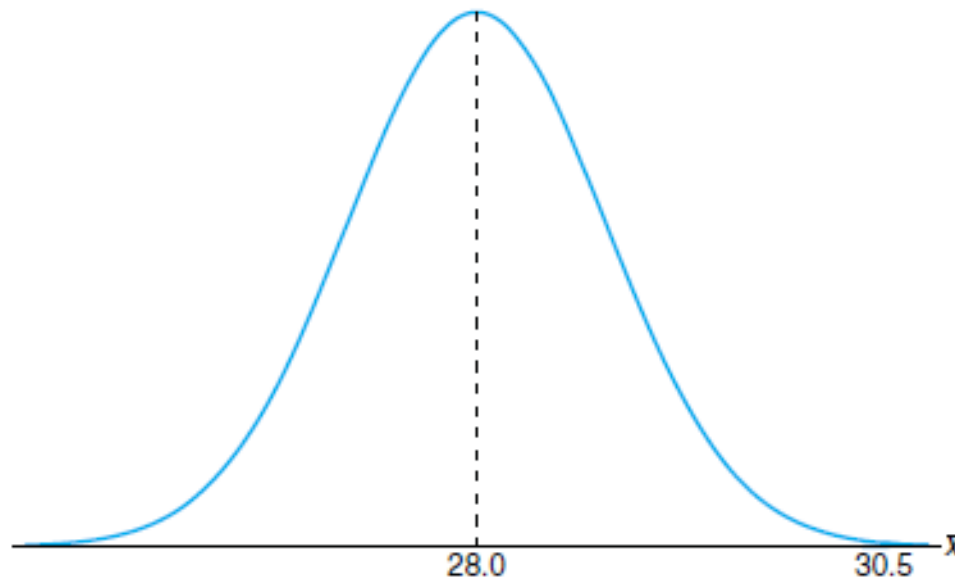
Example : Traveling between two campuses of a university in a city via shuttle bus takes, on average, **28 minutes** with a **standard deviation of 5** minutes. In a given week, a bus transported passengers **40 times**. What is the probability that the average transport time was more **than 30 minutes**? Assume the mean time is measured to the **nearest minute**

Solution: $\mu = 28$, $\sigma = 3$ and $n = 40$

$$P(\bar{X} > 30) = ?$$

Since the time is measured on a continuous scale to the **nearest minute**, an **\bar{X} greater than 30** is equivalent to **$\bar{x} \geq 30.5$** .

$$\begin{aligned}
 P(\bar{X} > 30) &= P\left(\frac{\bar{X}-28}{5/\sqrt{40}} \geq \frac{30.5-28}{5/\sqrt{40}}\right) \\
 &= P(Z \geq 3.16) \\
 &= 0.0008
 \end{aligned}$$



```
from scipy.stats import norm
```

Given data

```
population_mean = 28      # Mean of the population
population_std = 5        # Standard deviation of the population
sample_size = 40         # Sample size
sample_mean = 30.5       # Sample mean
```

Calculate the standard error of the mean (SEM)

```
sem = population_std / (sample_size ** 0.5)
```

Calculate the z-score

```
z_score = (sample_mean - population_mean) / sem
```

Calculate the probability using the complementary cumulative distribution function (1 - CDF)

```
probability = 1 - norm.cdf(z_score)
```

```
print(f"The probability that the average transport time was more  
than 30 minutes is: {probability:.4f}")
```

**# The probability that the average transport time was more than 30 minutes is:
0.0008**

Sampling Distribution of the Difference between Two Means

If independent samples of size n_1 and n_2 are drawn at random from two populations, discrete or continuous, with means μ_1 and μ_2 and variances σ_1^2 and σ_2^2 , respectively, then the sampling distribution of the differences of means,

$\bar{X}_1 - \bar{X}_2$, is approximately normally distributed with mean and variance given by

$$\mu_{\bar{X}_1 - \bar{X}_2} = \mu_1 - \mu_2 \text{ and } \sigma_{\bar{X}_1 - \bar{X}_2}^2 = \frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}$$

Hence,

$$Z = \frac{(\bar{X}_1 - \bar{X}_2) - (\mu_1 - \mu_2)}{\sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}}$$

is approximately a standard normal variable.

Sampling Distribution of the Difference between Two Means

- ❑ If both n_1 and n_2 are greater than or equal to **30**, the normal approximation for the distribution of $\bar{X}_1 - \bar{X}_2$ is very good when the underlying distributions are **not too far away** from normal.
- ❑ However, even when n_1 and n_2 are **less than 30**, the **normal approximation** is reasonably good **except** when the populations are decidedly **nonnormal**. Of course, if both populations are normal, then $\bar{X}_1 - \bar{X}_2$ has a normal distribution no matter what the sizes of n_1 and n_2 are.

Example Paint Drying Time: Two independent experiments are run in which two different types of paint are compared. **Eighteen** specimens are painted using type A , and the drying time, in hours, is recorded for each. The same is done with type B . The **population standard deviations** are both known to be **1.0**.

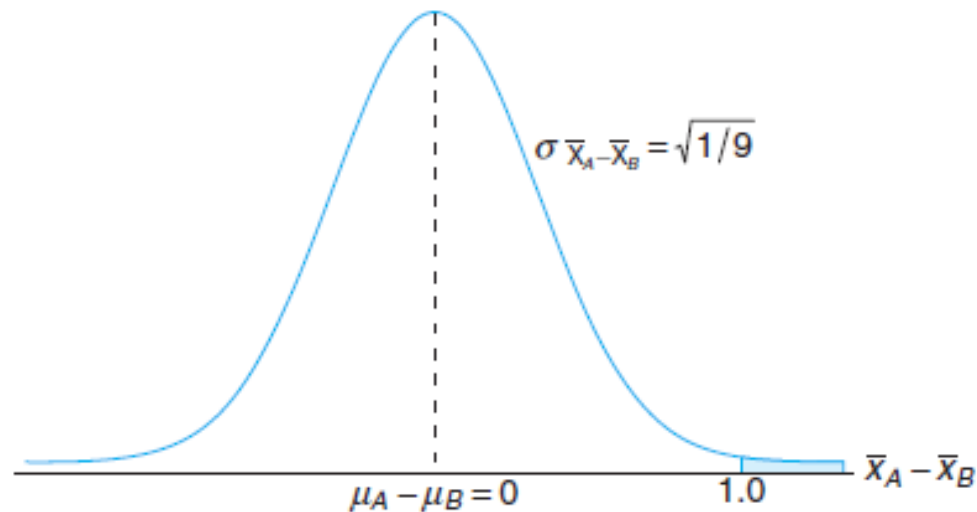
Assuming that the **mean drying time is equal** for the two types of paint, find **$P(\bar{X}_A - \bar{X}_B > 1.0)$** , where \bar{X}_A and \bar{X}_B are average drying times for samples of size **$n_A = n_B = 18$** .

Solution : $P(\bar{X}_A - \bar{X}_B > 1.0) = ?$

$$\mu_{\bar{X}_A - \bar{X}_B} = \mu_A - \mu_B = 0$$

$$\text{and } \sigma_{\bar{X}_A - \bar{X}_B}^2 = \frac{\sigma_A^2}{n_A} + \frac{\sigma_B^2}{n_B}$$

$$\sigma_{\bar{X}_1 - \bar{X}_2}^2 = \frac{1}{18} + \frac{1}{18} = \frac{1}{9}$$



$$Z = \frac{(\bar{X}_1 - \bar{X}_2) - (\mu_1 - \mu_2)}{\sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}}$$

$$\Rightarrow Z = \frac{1 - 0}{\sqrt{\frac{1}{9}}} = 3$$

$$\begin{aligned} P(Z > 3.0) &= 1 - P(Z < 3.0) \\ &= 1 - 0.9987 \\ &= 0.0013. \end{aligned}$$

```

from scipy.stats import norm
# Given data
sigma_A = 1.0          # Population standard deviation for type A
sigma_B = 1.0          # Population standard deviation for type B
n_A = 18               # Sample size for type A
n_B = 18               # Sample size for type B

# Calculate the standard error of the mean (SEM)
sem = ((sigma_A ** 2) / n_A + (sigma_B ** 2) / n_B) ** 0.5

# Calculate the z-score assuming equal means (mu_A - mu_B = 0)
z_score = 1.0 / sem

# Calculate the probability using the complementary cumulative
distribution function (1 - CDF)
probability = 1 - norm.cdf(z_score)
print(f"The probability that (mu_A - mu_B) > 1.0 is:
{probability:.4f}")
# The probability that (mu_A - mu_B) > 1.0 is: 0.0013

```

Interval Estimation

- There are many situations in which it is preferable to determine an interval within which we would expect to find the value of the parameter. Such an interval is called an **interval estimate**.
- An interval estimate of a population parameter θ is an interval of the form $\hat{\theta}_L < \theta < \hat{\theta}_U$, where $\hat{\theta}_L$ and $\hat{\theta}_U$ depend on the value of the **statistic $\hat{\theta}$** for a particular sample and also on the sampling distribution of $\hat{\theta}$.

Single Sample: Estimating the Mean

- The sampling distribution of \bar{X} is centered at μ , and in most applications the variance is smaller than that of any other estimators of μ .
- Thus, the sample mean \bar{x} will be used as a point estimate for the population mean μ . Recall that $\sigma^2_{\bar{X}} = \sigma^2/n$, so a large sample will yield a value of \bar{x} that comes from a sampling distribution with a small variance. Hence, \bar{x} is likely to be a very accurate estimate of μ when n is large.

Type I Error vs. Type II error

- ❑ Rejection of the **null hypothesis** when **it is true** is called a **type I error**.
- ❑ Nonrejection of the null hypothesis when it is false is called a **type II error**.
- ❑ The probability of committing a **type I error**, also called the **level of significance**, is denoted by the **Greek letter α**

Single Sample: Estimating the Mean

According to the **Central Limit Theorem**, we can expect the sampling distribution of \bar{X} to be approximately normally distributed with mean $\mu_{\bar{X}} = \mu$ and $\sigma_{\bar{X}} = \sigma / \sqrt{n}$

$$P(-z_{\alpha/2} < Z < z_{\alpha/2}) = 1 - \alpha, \text{ where } Z = \frac{\bar{X} - \mu}{\sigma / \sqrt{n}}$$

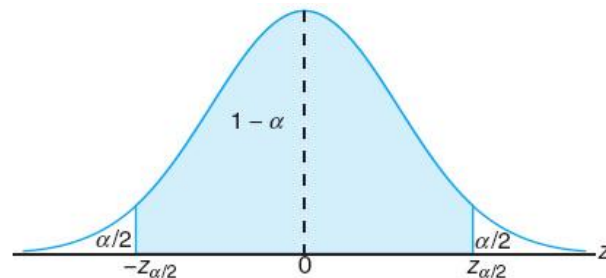
$$\Rightarrow P(-z_{\alpha/2} < \frac{\bar{X} - \mu}{\sigma / \sqrt{n}} < z_{\alpha/2}) = 1 - \alpha$$

$$\Rightarrow P(-z_{\alpha/2} \frac{\sigma}{\sqrt{n}} < \bar{X} - \mu < z_{\alpha/2} \frac{\sigma}{\sqrt{n}}) = 1 - \alpha$$

$$\Rightarrow P(-\bar{X} - z_{\alpha/2} \frac{\sigma}{\sqrt{n}} < -\mu < -\bar{X} + z_{\alpha/2} \frac{\sigma}{\sqrt{n}}) = 1 - \alpha$$

$$\Rightarrow P(\bar{X} + z_{\alpha/2} \frac{\sigma}{\sqrt{n}} > \mu > \bar{X} - z_{\alpha/2} \frac{\sigma}{\sqrt{n}}) = 1 - \alpha$$

$$\Rightarrow P(\bar{X} - z_{\alpha/2} \frac{\sigma}{\sqrt{n}} < \mu < \bar{X} + z_{\alpha/2} \frac{\sigma}{\sqrt{n}}) = 1 - \alpha$$



Confidence Interval on μ , σ^2 Known [1]

If \bar{x} is the mean of a random sample of size n from a population with known variance σ^2 , a $100(1 - \alpha)\%$ confidence interval for μ is given by

$$\bar{x} - z_{\alpha/2} \frac{\sigma}{\sqrt{n}} < \mu < \bar{x} + z_{\alpha/2} \frac{\sigma}{\sqrt{n}}$$

Confidence Interval on μ , σ^2 Known [2]

Example 6.11: The **average** zinc concentration recovered from a sample of measurements taken in **36** different locations in a river is found to be **2.6** grams per milliliter. Find the **95% and 99%** confidence intervals for the mean zinc concentration in the river. Assume that the population standard deviation is **0.3** gram per milliliter.

Area under the Normal Curve [1]



Table A.3 Areas under the Normal Curve

z	.00	.01	.02	.03	.04	.05	.06	.07	.08	.09
-3.4	0.0003	0.0003	0.0003	0.0003	0.0003	0.0003	0.0003	0.0003	0.0003	0.0002
-3.3	0.0005	0.0005	0.0005	0.0004	0.0004	0.0004	0.0004	0.0004	0.0004	0.0003
-3.2	0.0007	0.0007	0.0006	0.0006	0.0006	0.0006	0.0006	0.0005	0.0005	0.0005
-3.1	0.0010	0.0009	0.0009	0.0009	0.0008	0.0008	0.0008	0.0008	0.0007	0.0007
-3.0	0.0013	0.0013	0.0013	0.0012	0.0012	0.0011	0.0011	0.0011	0.0010	0.0010
-2.9	0.0019	0.0018	0.0018	0.0017	0.0016	0.0016	0.0015	0.0015	0.0014	0.0014
-2.8	0.0026	0.0025	0.0024	0.0023	0.0023	0.0022	0.0021	0.0021	0.0020	0.0019
-2.7	0.0035	0.0034	0.0033	0.0032	0.0031	0.0030	0.0029	0.0028	0.0027	0.0026
-2.6	0.0047	0.0045	0.0044	0.0043	0.0041	0.0040	0.0039	0.0038	0.0037	0.0036
-2.5	0.0062	0.0060	0.0059	0.0057	0.0055	0.0054	0.0052	0.0051	0.0049	0.0048
-2.4	0.0082	0.0080	0.0078	0.0075	0.0073	0.0071	0.0069	0.0068	0.0066	0.0064
-2.3	0.0107	0.0104	0.0102	0.0099	0.0096	0.0094	0.0091	0.0089	0.0087	0.0084
-2.2	0.0139	0.0136	0.0132	0.0129	0.0125	0.0122	0.0119	0.0116	0.0113	0.0110
-2.1	0.0179	0.0174	0.0170	0.0166	0.0162	0.0158	0.0154	0.0150	0.0146	0.0143
-2.0	0.0228	0.0222	0.0217	0.0212	0.0207	0.0202	0.0197	0.0192	0.0188	0.0183
-1.9	0.0287	0.0281	0.0274	0.0268	0.0262	0.0256	0.0250	0.0244	0.0239	0.0233
-1.8	0.0359	0.0351	0.0344	0.0336	0.0329	0.0322	0.0314	0.0307	0.0301	0.0294
-1.7	0.0446	0.0436	0.0427	0.0418	0.0409	0.0401	0.0392	0.0384	0.0375	0.0367
-1.6	0.0548	0.0537	0.0526	0.0516	0.0505	0.0495	0.0485	0.0475	0.0465	0.0455
-1.5	0.0668	0.0655	0.0643	0.0630	0.0618	0.0606	0.0594	0.0582	0.0571	0.0559
-1.4	0.0808	0.0793	0.0778	0.0764	0.0749	0.0735	0.0721	0.0708	0.0694	0.0681
-1.3	0.0968	0.0951	0.0934	0.0918	0.0901	0.0885	0.0869	0.0853	0.0838	0.0823
-1.2	0.1151	0.1131	0.1112	0.1093	0.1075	0.1056	0.1038	0.1020	0.1003	0.0985
-1.1	0.1357	0.1335	0.1314	0.1292	0.1271	0.1251	0.1230	0.1210	0.1190	0.1170
-1.0	0.1587	0.1562	0.1539	0.1515	0.1492	0.1469	0.1446	0.1423	0.1401	0.1379
-0.9	0.1841	0.1814	0.1788	0.1762	0.1736	0.1711	0.1685	0.1660	0.1635	0.1611
-0.8	0.2119	0.2090	0.2061	0.2033	0.2005	0.1977	0.1949	0.1922	0.1894	0.1867
-0.7	0.2420	0.2389	0.2358	0.2327	0.2296	0.2266	0.2236	0.2206	0.2177	0.2148
-0.6	0.2743	0.2709	0.2676	0.2643	0.2611	0.2578	0.2546	0.2514	0.2483	0.2451
-0.5	0.3085	0.3050	0.3015	0.2981	0.2946	0.2912	0.2877	0.2843	0.2810	0.2776
-0.4	0.3446	0.3409	0.3372	0.3336	0.3300	0.3264	0.3228	0.3192	0.3156	0.3121
-0.3	0.3821	0.3783	0.3745	0.3707	0.3669	0.3632	0.3594	0.3557	0.3520	0.3483
-0.2	0.4207	0.4168	0.4129	0.4090	0.4052	0.4013	0.3974	0.3936	0.3897	0.3859
-0.1	0.4602	0.4562	0.4522	0.4483	0.4443	0.4404	0.4364	0.4325	0.4286	0.4247
-0.0	0.5000	0.4960	0.4920	0.4880	0.4840	0.4801	0.4761	0.4721	0.4681	0.4641

Area under the Normal Curve [2]

Table A.3 (continued) Areas under the Normal Curve

<i>z</i>	.00	.01	.02	.03	.04	.05	.06	.07	.08	.09
0.0	0.5000	0.5040	0.5080	0.5120	0.5160	0.5199	0.5239	0.5279	0.5319	0.5359
0.1	0.5398	0.5438	0.5478	0.5517	0.5557	0.5596	0.5636	0.5675	0.5714	0.5753
0.2	0.5793	0.5832	0.5871	0.5910	0.5948	0.5987	0.6026	0.6064	0.6103	0.6141
0.3	0.6179	0.6217	0.6255	0.6293	0.6331	0.6368	0.6406	0.6443	0.6480	0.6517
0.4	0.6554	0.6591	0.6628	0.6664	0.6700	0.6736	0.6772	0.6808	0.6844	0.6879
0.5	0.6915	0.6950	0.6985	0.7019	0.7054	0.7088	0.7123	0.7157	0.7190	0.7224
0.6	0.7257	0.7291	0.7324	0.7357	0.7389	0.7422	0.7454	0.7486	0.7517	0.7549
0.7	0.7580	0.7611	0.7642	0.7673	0.7704	0.7734	0.7764	0.7794	0.7823	0.7852
0.8	0.7881	0.7910	0.7939	0.7967	0.7995	0.8023	0.8051	0.8078	0.8106	0.8133
0.9	0.8159	0.8186	0.8212	0.8238	0.8264	0.8289	0.8315	0.8340	0.8365	0.8389
1.0	0.8413	0.8438	0.8461	0.8485	0.8508	0.8531	0.8554	0.8577	0.8599	0.8621
1.1	0.8643	0.8665	0.8686	0.8708	0.8729	0.8749	0.8770	0.8790	0.8810	0.8830
1.2	0.8849	0.8869	0.8888	0.8907	0.8925	0.8944	0.8962	0.8980	0.8997	0.9015
1.3	0.9032	0.9049	0.9066	0.9082	0.9099	0.9115	0.9131	0.9147	0.9162	0.9177
1.4	0.9192	0.9207	0.9222	0.9236	0.9251	0.9265	0.9279	0.9292	0.9306	0.9319
1.5	0.9332	0.9345	0.9357	0.9370	0.9382	0.9394	0.9406	0.9418	0.9429	0.9441
1.6	0.9452	0.9463	0.9474	0.9484	0.9495	0.9505	0.9515	0.9525	0.9535	0.9545
1.7	0.9554	0.9564	0.9573	0.9582	0.9591	0.9599	0.9608	0.9616	0.9625	0.9633
1.8	0.9641	0.9649	0.9656	0.9664	0.9671	0.9678	0.9686	0.9693	0.9699	0.9706
1.9	0.9713	0.9719	0.9726	0.9732	0.9738	0.9744	0.9750	0.9756	0.9761	0.9767
2.0	0.9772	0.9778	0.9783	0.9788	0.9793	0.9798	0.9803	0.9808	0.9812	0.9817
2.1	0.9821	0.9826	0.9830	0.9834	0.9838	0.9842	0.9846	0.9850	0.9854	0.9857
2.2	0.9861	0.9864	0.9868	0.9871	0.9875	0.9878	0.9881	0.9884	0.9887	0.9890
2.3	0.9893	0.9896	0.9898	0.9901	0.9904	0.9906	0.9909	0.9911	0.9913	0.9916
2.4	0.9918	0.9920	0.9922	0.9925	0.9927	0.9929	0.9931	0.9932	0.9934	0.9936
2.5	0.9938	0.9940	0.9941	0.9943	0.9945	0.9946	0.9948	0.9949	0.9951	0.9952
2.6	0.9953	0.9955	0.9956	0.9957	0.9959	0.9960	0.9961	0.9962	0.9963	0.9964
2.7	0.9965	0.9966	0.9967	0.9968	0.9969	0.9970	0.9971	0.9972	0.9973	0.9974
2.8	0.9974	0.9975	0.9976	0.9977	0.9977	0.9978	0.9979	0.9979	0.9980	0.9981
2.9	0.9981	0.9982	0.9982	0.9983	0.9984	0.9984	0.9985	0.9985	0.9986	0.9986
3.0	0.9987	0.9987	0.9987	0.9988	0.9988	0.9989	0.9989	0.9989	0.9990	0.9990
3.1	0.9990	0.9991	0.9991	0.9991	0.9992	0.9992	0.9992	0.9992	0.9993	0.9993
3.2	0.9993	0.9993	0.9994	0.9994	0.9994	0.9994	0.9994	0.9995	0.9995	0.9995
3.3	0.9995	0.9995	0.9995	0.9996	0.9996	0.9996	0.9996	0.9996	0.9996	0.9997
3.4	0.9997	0.9997	0.9997	0.9997	0.9997	0.9997	0.9997	0.9997	0.9997	0.9998

Solution:

The point estimate of μ is $\bar{x} = 2.6$. The z-value leaving an area of 0.025 to the right, and therefore an area of 0.975 to the left, is $z_{0.025} = 1.96$ (Table A.3). Hence, the 95% confidence interval is

$$\bar{x} - z_{\alpha/2} \frac{\sigma}{\sqrt{n}} < \mu < \bar{x} + z_{\alpha/2} \frac{\sigma}{\sqrt{n}}$$

$$\alpha/2 = 0.05/2 = 0.0250$$

$$z_{\alpha/2} = z_{0.0250} = 1.96 \quad \because 1 - \alpha/2 = 1 - 0.250 = 0.9750$$

95% confidence interval for μ is

$$2.6 - (1.96)\left(\frac{0.3}{\sqrt{36}}\right) < \mu < 2.6 + (1.96)\left(\frac{0.3}{\sqrt{36}}\right)$$

$$2.47 < \mu < 2.73.$$

$$\alpha/2 = 0.01/2 = 0.005$$

$$\therefore 1 - \alpha/2 = 1 - 0.005 = 0.9950$$

$$Z_{\alpha/2} = Z_{0.005} = 2.575$$

99% confidence interval for μ is

$$2.6 - (2.575) \left(\frac{0.3}{\sqrt{36}} \right) < \mu < 2.6 + (2.575) \left(\frac{0.3}{\sqrt{36}} \right)$$

$$2.6 - 0.13 < \mu < 2.6 + 0.13$$

$$2.47 < \mu < 2.73.$$

```
import scipy.stats as stats
```

Given data

```
sample_mean = 2.6
```

Sample mean

```
population_std = 0.3  
deviation
```

Population standard

```
sample_size = 36
```

Sample size

```
confidence_levels = [0.95, 0.99] # Confidence levels
```

Calculate the standard error of the mean (SEM)

```
sem = population_std / (sample_size ** 0.5)
```

Calculate the confidence intervals for the given confidence levels

```
confidence_intervals = []
```

```
for confidence_level in confidence_levels:
```

Calculate the Z-score for the given confidence level

```
z_score = stats.norm.ppf((1 + confidence_level) / 2)
```

Calculate the margin of error

```
margin_of_error = z_score * sem
```


for loop continue from the previous slide

Calculate the confidence interval

```
lower_limit = sample_mean - margin_of_error
```

```
upper_limit = sample_mean + margin_of_error
```

```
confidence_intervals.append((lower_limit, upper_limit))
```

Print the results

```
for confidence_level in confidence_levels:
```

```
# Calculate the Z-score for the given confidence level
```

```
z_score = stats.norm.ppf((1 + confidence_level) / 2)
```

```
#95% Confidence Interval: (2.5020, 2.6980)
```

```
#99% Confidence Interval: (2.4712, 2.7288)
```

Error In Estimating μ By \bar{x} [1]

- The $100(1-\alpha)\%$ confidence interval provides an estimate of the accuracy of our point estimate. If μ is actually the center value of the interval, then \bar{x} estimates μ without error.
- Most of the time, however, **\bar{x} will not be exactly equal to μ** and the point estimate will **be in error**. The size of this error will be the absolute value of the difference between μ and \bar{x} , and we can be $100(1 - \alpha)\%$ confident that this **difference will not exceed**
 $z_{\alpha/2} \frac{\sigma}{\sqrt{n}}$

Error In Estimating μ By \bar{x} [2]

Theorem 9.1: If \bar{x} is used as an estimate of μ , we can be $100(1 - \alpha)\%$ confident that the error will not exceed $z_{\alpha/2} \frac{\sigma}{\sqrt{n}}$

Area under the Normal Curve [1]



Table A.3 Areas under the Normal Curve

z	.00	.01	.02	.03	.04	.05	.06	.07	.08	.09
-3.4	0.0003	0.0003	0.0003	0.0003	0.0003	0.0003	0.0003	0.0003	0.0003	0.0002
-3.3	0.0005	0.0005	0.0005	0.0004	0.0004	0.0004	0.0004	0.0004	0.0004	0.0003
-3.2	0.0007	0.0007	0.0006	0.0006	0.0006	0.0006	0.0006	0.0005	0.0005	0.0005
-3.1	0.0010	0.0009	0.0009	0.0009	0.0008	0.0008	0.0008	0.0008	0.0007	0.0007
-3.0	0.0013	0.0013	0.0013	0.0012	0.0012	0.0011	0.0011	0.0011	0.0010	0.0010
-2.9	0.0019	0.0018	0.0018	0.0017	0.0016	0.0016	0.0015	0.0015	0.0014	0.0014
-2.8	0.0026	0.0025	0.0024	0.0023	0.0023	0.0022	0.0021	0.0021	0.0020	0.0019
-2.7	0.0035	0.0034	0.0033	0.0032	0.0031	0.0030	0.0029	0.0028	0.0027	0.0026
-2.6	0.0047	0.0045	0.0044	0.0043	0.0041	0.0040	0.0039	0.0038	0.0037	0.0036
-2.5	0.0062	0.0060	0.0059	0.0057	0.0055	0.0054	0.0052	0.0051	0.0049	0.0048
-2.4	0.0082	0.0080	0.0078	0.0075	0.0073	0.0071	0.0069	0.0068	0.0066	0.0064
-2.3	0.0107	0.0104	0.0102	0.0099	0.0096	0.0094	0.0091	0.0089	0.0087	0.0084
-2.2	0.0139	0.0136	0.0132	0.0129	0.0125	0.0122	0.0119	0.0116	0.0113	0.0110
-2.1	0.0179	0.0174	0.0170	0.0166	0.0162	0.0158	0.0154	0.0150	0.0146	0.0143
-2.0	0.0228	0.0222	0.0217	0.0212	0.0207	0.0202	0.0197	0.0192	0.0188	0.0183
-1.9	0.0287	0.0281	0.0274	0.0268	0.0262	0.0256	0.0250	0.0244	0.0239	0.0233
-1.8	0.0359	0.0351	0.0344	0.0336	0.0329	0.0322	0.0314	0.0307	0.0301	0.0294
-1.7	0.0446	0.0436	0.0427	0.0418	0.0409	0.0401	0.0392	0.0384	0.0375	0.0367
-1.6	0.0548	0.0537	0.0526	0.0516	0.0505	0.0495	0.0485	0.0475	0.0465	0.0455
-1.5	0.0668	0.0655	0.0643	0.0630	0.0618	0.0606	0.0594	0.0582	0.0571	0.0559
-1.4	0.0808	0.0793	0.0778	0.0764	0.0749	0.0735	0.0721	0.0708	0.0694	0.0681
-1.3	0.0968	0.0951	0.0934	0.0918	0.0901	0.0885	0.0869	0.0853	0.0838	0.0823
-1.2	0.1151	0.1131	0.1112	0.1093	0.1075	0.1056	0.1038	0.1020	0.1003	0.0985
-1.1	0.1357	0.1335	0.1314	0.1292	0.1271	0.1251	0.1230	0.1210	0.1190	0.1170
-1.0	0.1587	0.1562	0.1539	0.1515	0.1492	0.1469	0.1446	0.1423	0.1401	0.1379
-0.9	0.1841	0.1814	0.1788	0.1762	0.1736	0.1711	0.1685	0.1660	0.1635	0.1611
-0.8	0.2119	0.2090	0.2061	0.2033	0.2005	0.1977	0.1949	0.1922	0.1894	0.1867
-0.7	0.2420	0.2389	0.2358	0.2327	0.2296	0.2266	0.2236	0.2206	0.2177	0.2148
-0.6	0.2743	0.2709	0.2676	0.2643	0.2611	0.2578	0.2546	0.2514	0.2483	0.2451
-0.5	0.3085	0.3050	0.3015	0.2981	0.2946	0.2912	0.2877	0.2843	0.2810	0.2776
-0.4	0.3446	0.3409	0.3372	0.3336	0.3300	0.3264	0.3228	0.3192	0.3156	0.3121
-0.3	0.3821	0.3783	0.3745	0.3707	0.3669	0.3632	0.3594	0.3557	0.3520	0.3483
-0.2	0.4207	0.4168	0.4129	0.4090	0.4052	0.4013	0.3974	0.3936	0.3897	0.3859
-0.1	0.4602	0.4562	0.4522	0.4483	0.4443	0.4404	0.4364	0.4325	0.4286	0.4247
-0.0	0.5000	0.4960	0.4920	0.4880	0.4840	0.4801	0.4761	0.4721	0.4681	0.4641

Area under the Normal Curve [2]

Table A.3 (continued) Areas under the Normal Curve

<i>z</i>	.00	.01	.02	.03	.04	.05	.06	.07	.08	.09
0.0	0.5000	0.5040	0.5080	0.5120	0.5160	0.5199	0.5239	0.5279	0.5319	0.5359
0.1	0.5398	0.5438	0.5478	0.5517	0.5557	0.5596	0.5636	0.5675	0.5714	0.5753
0.2	0.5793	0.5832	0.5871	0.5910	0.5948	0.5987	0.6026	0.6064	0.6103	0.6141
0.3	0.6179	0.6217	0.6255	0.6293	0.6331	0.6368	0.6406	0.6443	0.6480	0.6517
0.4	0.6554	0.6591	0.6628	0.6664	0.6700	0.6736	0.6772	0.6808	0.6844	0.6879
0.5	0.6915	0.6950	0.6985	0.7019	0.7054	0.7088	0.7123	0.7157	0.7190	0.7224
0.6	0.7257	0.7291	0.7324	0.7357	0.7389	0.7422	0.7454	0.7486	0.7517	0.7549
0.7	0.7580	0.7611	0.7642	0.7673	0.7704	0.7734	0.7764	0.7794	0.7823	0.7852
0.8	0.7881	0.7910	0.7939	0.7967	0.7995	0.8023	0.8051	0.8078	0.8106	0.8133
0.9	0.8159	0.8186	0.8212	0.8238	0.8264	0.8289	0.8315	0.8340	0.8365	0.8389
1.0	0.8413	0.8438	0.8461	0.8485	0.8508	0.8531	0.8554	0.8577	0.8599	0.8621
1.1	0.8643	0.8665	0.8686	0.8708	0.8729	0.8749	0.8770	0.8790	0.8810	0.8830
1.2	0.8849	0.8869	0.8888	0.8907	0.8925	0.8944	0.8962	0.8980	0.8997	0.9015
1.3	0.9032	0.9049	0.9066	0.9082	0.9099	0.9115	0.9131	0.9147	0.9162	0.9177
1.4	0.9192	0.9207	0.9222	0.9236	0.9251	0.9265	0.9279	0.9292	0.9306	0.9319
1.5	0.9332	0.9345	0.9357	0.9370	0.9382	0.9394	0.9406	0.9418	0.9429	0.9441
1.6	0.9452	0.9463	0.9474	0.9484	0.9495	0.9505	0.9515	0.9525	0.9535	0.9545
1.7	0.9554	0.9564	0.9573	0.9582	0.9591	0.9599	0.9608	0.9616	0.9625	0.9633
1.8	0.9641	0.9649	0.9656	0.9664	0.9671	0.9678	0.9686	0.9693	0.9699	0.9706
1.9	0.9713	0.9719	0.9726	0.9732	0.9738	0.9744	0.9750	0.9756	0.9761	0.9767
2.0	0.9772	0.9778	0.9783	0.9788	0.9793	0.9798	0.9803	0.9808	0.9812	0.9817
2.1	0.9821	0.9826	0.9830	0.9834	0.9838	0.9842	0.9846	0.9850	0.9854	0.9857
2.2	0.9861	0.9864	0.9868	0.9871	0.9875	0.9878	0.9881	0.9884	0.9887	0.9890
2.3	0.9893	0.9896	0.9898	0.9901	0.9904	0.9906	0.9909	0.9911	0.9913	0.9916
2.4	0.9918	0.9920	0.9922	0.9925	0.9927	0.9929	0.9931	0.9932	0.9934	0.9936
2.5	0.9938	0.9940	0.9941	0.9943	0.9945	0.9946	0.9948	0.9949	0.9951	0.9952
2.6	0.9953	0.9955	0.9956	0.9957	0.9959	0.9960	0.9961	0.9962	0.9963	0.9964
2.7	0.9965	0.9966	0.9967	0.9968	0.9969	0.9970	0.9971	0.9972	0.9973	0.9974
2.8	0.9974	0.9975	0.9976	0.9977	0.9977	0.9978	0.9979	0.9979	0.9980	0.9981
2.9	0.9981	0.9982	0.9982	0.9983	0.9984	0.9984	0.9985	0.9985	0.9986	0.9986
3.0	0.9987	0.9987	0.9987	0.9988	0.9988	0.9989	0.9989	0.9989	0.9990	0.9990
3.1	0.9990	0.9991	0.9991	0.9991	0.9992	0.9992	0.9992	0.9992	0.9993	0.9993
3.2	0.9993	0.9993	0.9994	0.9994	0.9994	0.9994	0.9994	0.9995	0.9995	0.9995
3.3	0.9995	0.9995	0.9995	0.9996	0.9996	0.9996	0.9996	0.9996	0.9996	0.9997
3.4	0.9997	0.9997	0.9997	0.9997	0.9997	0.9997	0.9997	0.9997	0.9997	0.9998

$$z_{\alpha/2} = z_{0.0250} = 1.96 \quad \because 1 - \alpha/2 = 1 - 0.250 = 0.9750$$

$$z_{\alpha/2} = z_{0.005} = 2.575 \quad \because 1 - \alpha/2 = 1 - 0.005 = 0.9950$$

□ We are **95% confident** that the sample mean $\bar{x} = 2.6$ differs from the true mean μ by an amount less than **$(1.96)(0.3)/\sqrt{36} = 0.1$** and **99% confident** that the difference is less than **$(2.575)(0.3)/\sqrt{36} = 0.13$** .

□ Frequently, we wish to know how **large a sample** is necessary to ensure that the error in estimating μ will be less than a specified amount **e** .

□ By Theorem 9.1, we must choose n such that

$$z_{\alpha/2} \frac{\sigma}{\sqrt{n}} = e$$

Theorem 9.2: If \bar{x} is used as an estimate of μ , we can be $100(1 - \alpha)\%$ confident that the error will not exceed a specified amount e when the sample size is

$$n = \left(\frac{\sigma Z_{\alpha/2}}{e} \right)^2$$

When solving for the sample size, n , we round all fractional values up to the next whole number. By adhering to this principle, we can be sure that our degree of confidence never falls below $100(1 - \alpha)\%$.

Example: How **large a sample** is required if we want to be 95% confident that our estimate of μ in the previous Example is off by **less than 0.05**?

Solution:

The population standard deviation is $\sigma = 0.3$. Then, by Theorem 9.2

$$n = \{(1.96)(0.3) / 0.05\}^2 = 138.3.$$

Therefore, we can be 95% confident that a random sample of size **139** will provide an estimate \bar{x} differing from μ by an amount less than 0.05.