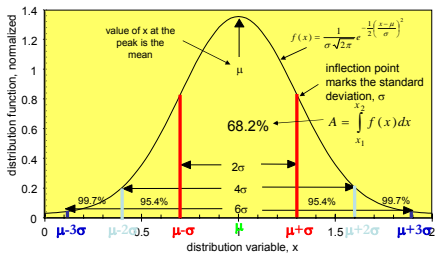


The Normal Distribution

Importance of normal or Gaussian distribution (ND)

- It is the most used distribution
- Most methods are based on the assumption of ND
- Sum of many independent, random contributions variables (grain size, height, environmental factors) are of ND
- Histogram - symmetrical (mean and STD).

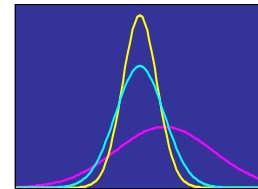
Normal Distribution Function



- Area from $-\sigma \leq x \leq +\sigma \rightarrow 68.2\%$ of the total area ($x_1 = -\sigma ; x_2 = \sigma$)
- Area from $-2\sigma \leq x \leq +2\sigma \rightarrow 95.4\%$ of the total area ($x_1 = -2\sigma ; x_2 = 2\sigma$)
- Area from $-3\sigma \leq x \leq +3\sigma \rightarrow 99.7\%$ of the total area ($x_1 = -3\sigma ; x_2 = 3\sigma$)

The Gaussian Curve

- The analytical computation of the area under the Gaussian curve is difficult.
- **Standardized tables** generated for use.
- By varying the parameters σ and μ , we obtain different normal distributions

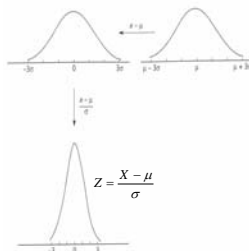


There are an infinite number of normal distributions!
Which table to use?

Standardized Z-score

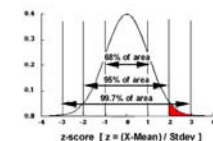
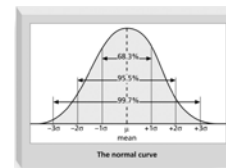
- The **standardization** assumes a mean of zero and variance of 1.
- Any random variable X with mean μ and std σ can be transformed to a standard ND using a **Z score (transformation)**

$$Z = (X - \mu) / \sigma$$



Advantages of using Z-score

- Permit numerical comparisons of measures from different scales by converting BOTH measures to a common scale.
- Permit numerical estimates of the expected frequency of an event
- Useful in determining the likelihood of an event occurring by chance.



Some commonly used standardized test scales...

Mean	Standard Deviation	Scale Name
500	100	SAT; GRE; LSAT; GMAT
100	15	Wechsler IQ
100	16	Stanford Binet IQ
20	5	ACT (Amer College Testing Co.)
50	10	T-scale (MMPI)

To convert raw scores to another scale, convert raw score to z score and then convert z score into the other raw score scale using the reverse transform.

Tables of Normal Distribution

- Predictions are made about ND populations by using Standard Table - **Appendix in many books**
($-3 < Z < 3$, $0.13\% < P < 99.87\%$).
- The easiest answer to the standard normal distribution table is to use **Excel function =NORMSDIST(Z)** (for $\mu = 0$ and $\sigma=1.0$)
- The inverse function **Z=NORMSINV(P)**

Standard ND Tables

Areas Under the Standard Normal Curve from 0 to z											Areas Under the Standard Normal Curve from 0 to z										
z	.00	.01	.02	.03	.04	.05	.06	.07	.08	.09	z	.00	.01	.02	.03	.04	.05	.06	.07	.08	.09
0.0	.0000	.0040	.0080	.0120	.0160	.0199	.0239	.0279	.0319	.0358	0.0	.0000	.0040	.0080	.0120	.0160	.0199	.0239	.0279	.0319	.0358
0.1	.0398	.0438	.0478	.0519	.0559	.0599	.0639	.0679	.0719	.0758	0.1	.0398	.0438	.0478	.0519	.0559	.0599	.0639	.0679	.0719	.0758
0.2	.0798	.0838	.0878	.0918	.0958	.0997	.1036	.1075	.1114	.1153	0.2	.0798	.0838	.0878	.0918	.0958	.0997	.1036	.1075	.1114	.1153
0.3	.1193	.1233	.1273	.1312	.1351	.1390	.1429	.1468	.1506	.1545	0.3	.1193	.1233	.1273	.1312	.1351	.1390	.1429	.1468	.1506	.1545
0.4	.1584	.1623	.1661	.1699	.1736	.1773	.1811	.1848	.1885	.1922	0.4	.1584	.1623	.1661	.1699	.1736	.1773	.1811	.1848	.1885	.1922
0.5	.1959	.1995	.2031	.2068	.2104	.2139	.2174	.2209	.2244	.2279	0.5	.1959	.1995	.2031	.2068	.2104	.2139	.2174	.2209	.2244	.2279
0.6	.2315	.2350	.2384	.2419	.2454	.2488	.2522	.2556	.2590	.2625	0.6	.2315	.2350	.2384	.2419	.2454	.2488	.2522	.2556	.2590	.2625
0.7	.2659	.2693	.2726	.2759	.2791	.2824	.2856	.2888	.2919	.2950	0.7	.2659	.2693	.2726	.2759	.2791	.2824	.2856	.2888	.2919	.2950
0.8	.2981	.3015	.3048	.3080	.3112	.3143	.3174	.3205	.3236	.3266	0.8	.2981	.3015	.3048	.3080	.3112	.3143	.3174	.3205	.3236	.3266
0.9	.3296	.3324	.3353	.3381	.3409	.3438	.3465	.3493	.3520	.3547	0.9	.3296	.3324	.3353	.3381	.3409	.3438	.3465	.3493	.3520	.3547
1.0	.3564	.3591	.3618	.3645	.3671	.3697	.3723	.3749	.3774	.3799	1.0	.3564	.3591	.3618	.3645	.3671	.3697	.3723	.3749	.3774	.3799
1.1	.3815	.3841	.3867	.3892	.3917	.3942	.3967	.3992	.4017	.4041	1.1	.3815	.3841	.3867	.3892	.3917	.3942	.3967	.3992	.4017	.4041
1.2	.4059	.4085	.4110	.4135	.4159	.4183	.4207	.4231	.4255	.4279	1.2	.4059	.4085	.4110	.4135	.4159	.4183	.4207	.4231	.4255	.4279
1.3	.4292	.4315	.4338	.4361	.4384	.4406	.4429	.4451	.4473	.4495	1.3	.4292	.4315	.4338	.4361	.4384	.4406	.4429	.4451	.4473	.4495
1.4	.4515	.4535	.4554	.4573	.4592	.4611	.4630	.4648	.4667	.4685	1.4	.4515	.4535	.4554	.4573	.4592	.4611	.4630	.4648	.4667	.4685
1.5	.4693	.4711	.4729	.4746	.4764	.4781	.4798	.4815	.4832	.4849	1.5	.4693	.4711	.4729	.4746	.4764	.4781	.4798	.4815	.4832	.4849
1.6	.4856	.4871	.4886	.4900	.4915	.4930	.4944	.4958	.4972	.4985	1.6	.4856	.4871	.4886	.4900	.4915	.4930	.4944	.4958	.4972	.4985
1.7	.4990	.4999	.5008	.5017	.5026	.5035	.5044	.5053	.5061	.5069	1.7	.4990	.4999	.5008	.5017	.5026	.5035	.5044	.5053	.5061	.5069
1.8	.5076	.5084	.5092	.5100	.5108	.5116	.5124	.5132	.5140	.5147	1.8	.5076	.5084	.5092	.5100	.5108	.5116	.5124	.5132	.5140	.5147
1.9	.5154	.5161	.5168	.5176	.5183	.5190	.5197	.5204	.5211	.5218	1.9	.5154	.5161	.5168	.5176	.5183	.5190	.5197	.5204	.5211	.5218
2.0	.5224	.5231	.5238	.5245	.5252	.5259	.5266	.5272	.5279	.5285	2.0	.5224	.5231	.5238	.5245	.5252	.5259	.5266	.5272	.5279	.5285
2.1	.5291	.5298	.5304	.5310	.5317	.5323	.5329	.5335	.5341	.5347	2.1	.5291	.5298	.5304	.5310	.5317	.5323	.5329	.5335	.5341	.5347
2.2	.5353	.5359	.5364	.5370	.5375	.5381	.5386	.5392	.5398	.5403	2.2	.5353	.5359	.5364	.5370	.5375	.5381	.5386	.5392	.5398	.5403
2.3	.5408	.5413	.5418	.5424	.5429	.5434	.5439	.5444	.5449	.5454	2.3	.5408	.5413	.5418	.5424	.5429	.5434	.5439	.5444	.5449	.5454
2.4	.5458	.5463	.5468	.5473	.5478	.5483	.5488	.5493	.5498	.5503	2.4	.5458	.5463	.5468	.5473	.5478	.5483	.5488	.5493	.5498	.5503
2.5	.5507	.5512	.5517	.5521	.5526	.5531	.5536	.5541	.5546	.5551	2.5	.5507	.5512	.5517	.5521	.5526	.5531	.5536	.5541	.5546	.5551
2.6	.5556	.5561	.5566	.5570	.5575	.5580	.5585	.5590	.5595	.5600	2.6	.5556	.5561	.5566	.5570	.5575	.5580	.5585	.5590	.5595	.5600
2.7	.5605	.5609	.5613	.5618	.5623	.5627	.5632	.5637	.5641	.5646	2.7	.5605	.5609	.5613	.5618	.5623	.5627	.5632	.5637	.5641	.5646
2.8	.5650	.5654	.5658	.5663	.5667	.5671	.5676	.5680	.5685	.5689	2.8	.5650	.5654	.5658	.5663	.5667	.5671	.5676	.5680	.5685	.5689
2.9	.5693	.5697	.5701	.5705	.5709	.5713	.5717	.5721	.5725	.5729	2.9	.5693	.5697	.5701	.5705	.5709	.5713	.5717	.5721	.5725	.5729
3.0	.5733	.5737	.5740	.5744	.5748	.5752	.5756	.5760	.5764	.5768	3.0	.5733	.5737	.5740	.5744	.5748	.5752	.5756	.5760	.5764	.5768

Area under the curve on each side of zero is 0.5. The ci



Normalization to use standard tables: $z = \frac{x - \mu}{\sigma}$

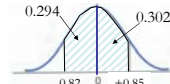
Example: if $z=0.82 \rightarrow$
Area under the curve for $[0, 0.82] : 0.294$
Total area for $[-\infty, 0.82]=0.5+0.294=0.794$
This value is the **probability** that $z < 0.82$

Example

Company Lentil[®] produces computer chip Pansium[®] XIX running at 66.666 THz. However, the rival company DAM[®] manufactures Craplon[®] 66+⁺, also running at 66.666 THz. However, DAM claims that Lentil's chip is flawed, and cannot run any faster than 63 THz. **Lentil decides to test its chips.** They take a sample of 1% (1000 chips). They find that the mean speed of these chips is $m = 65.980$ THz with a std. dev. of $s = 1.2$ THz. Assuming that the chip speed is normally distributed, **is Lentil's speed claim justifiable?** Assume that the claim is justifiable, if 95% of the chips lie in the speed limits of $65 < x < 67$ THz.

$$z = \frac{x - \mu}{\sigma} = \frac{65 - 65.980}{1.2} = -0.82$$

$$z = \frac{x - \mu}{\sigma} = \frac{67 - 65.980}{1.2} = 0.85$$



The probability that a Lentil chip has a speed in the $[65, 67]$ THz is $0.294+0.302=0.596$. Thus only 59.6% of the chips satisfy the criterion.

Now assume that the claim is justifiable, if 90% of the chips run faster than $x > 65.0$ THz.

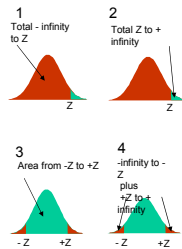
$$z = \frac{x - \mu}{\sigma} = \frac{65.0 - 65.980}{1.2} = -0.82$$

$$z = \frac{x - \mu}{\sigma} = \frac{63.0 - 65.980}{1.2} = -2.48$$

The probability that a Lentil chip has a speed larger than 65THz is $0.294+0.5=0.794$. That means, roughly 80% of the chips satisfy the criterion.
In any case, however, Lentil does better than DAM's claim of 63 THz. What % of Lentil chips run over 63THz? (Ans. 99.3%)
 $P(z > -2.48) = 0.493 + 0.5 = 0.993$

Tables of Normal Distribution - 1

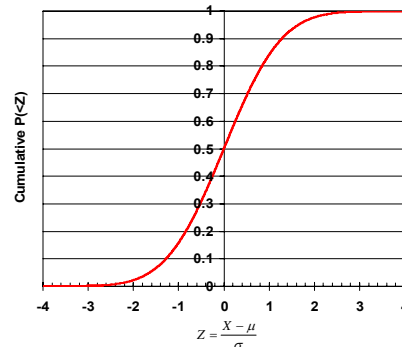
- $P(x < Z)$
- $P(x \geq Z) = 1 - P(x < Z)$
- $P(-Z < x < Z) = P(x < Z) - P(x \leq -Z)$
- $P(x \leq -Z, x \geq Z) = 1 - P(-Z < x < Z)$



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

Standard Normal Distribution

- Cumulative curve



Distribution of the Sample Mean

Distribution of the Sample Mean

- In real life, the MEAN and STD of a population are not known
- Need to use sample means to estimate or compare with those of populations.
- To test hypotheses about a product, a value or a statement.

Sampling and Estimation

- Random Variables Used to Estimate a Population Parameter
- Sample Mean Is an Estimator of Population Mean μ

Sampling Error

- The difference between a value (a statistic) computed from a sample and the corresponding value (a parameter) computed from a population.

$$\text{Sampling Error} = \bar{x} - \mu$$

- Where: \bar{x} = Sample mean
 μ = Population mean

Sampling Distribution

A *sampling distribution* is a distribution of the possible values of a statistic for a given size sample selected from a population.

Z-VALUE FOR SAMPLING DISTRIBUTION OF \bar{x}

$$z = \frac{(\bar{x} - \mu)}{\frac{\sigma}{\sqrt{n}}}$$

where:

- \bar{x} = Sample mean
- μ = Population mean
- σ = Population standard deviation
- n = Sample size

Standard Error of Mean

- Less Than Pop. Standard Deviation
- Formula (Sampling With Replacement)

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

Properties of Mean

1. Unbiasedness

- Mean of Sampling Distribution Equals Population Mean

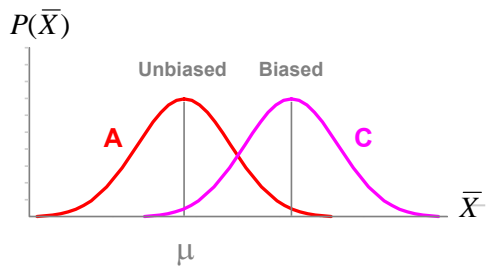
2. Efficiency

- Sample Mean Comes Closer to Population Mean Than Any Other Unbiased Estimator

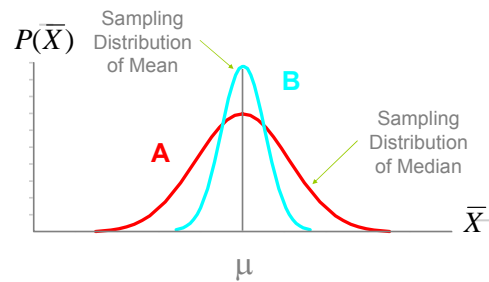
3. Consistency

- As Sample Size Increases, Variation of Sample Mean from Population Mean Decreases

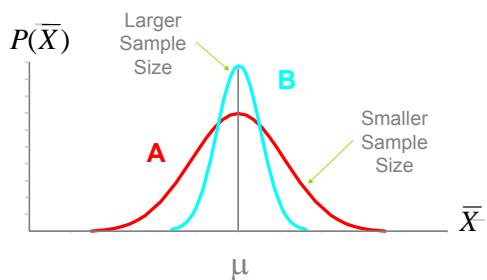
Unbiasedness



Efficiency



Consistency



Sampling from Non-Normal Populations

Central Tendency

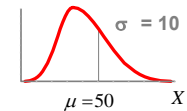
$$\mu_{\bar{X}} = \mu$$

Dispersion

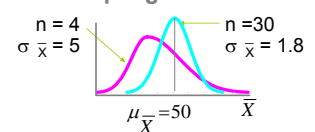
$$\sigma_{\bar{X}} = \frac{\sigma}{\sqrt{n}}$$

Sampling **With** Replacement

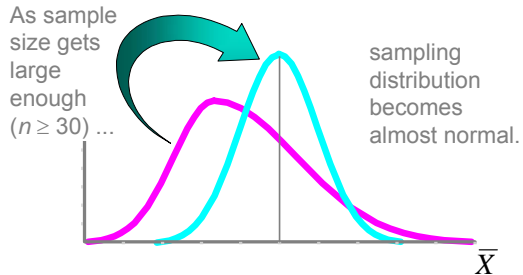
Population Distribution



Sampling Distribution



Central Limit Theorem



Central Limit Theorem

- For samples of n observations taken from a population with mean μ and standard deviation σ , **regardless of the population's distribution**, provided the sample size is sufficiently large, the distribution of the sample mean, will be normal with a mean equal to the population mean. Further, the standard deviation will equal the population standard deviation divided by the square-root of the sample size.
- The larger the sample size, the better the approximation to the normal distribution.

The **numerator of the variance equation** is called the **"SUM OF SQUARES"**

Short for: "the sum of the squared deviations about the arithmetic mean"

$$SS = \sum_{i=1}^n (x_i - \bar{X})^2$$

The sample Standard Deviation

- Taking the square root of the variance **scales the variance back to the number scale of the original measurements...**

$$STDEV = \sqrt{\frac{\sum_{i=1}^n (x_i - \bar{X})^2}{n - 1}}$$

The Bienayme - Chebyshev Inequality

Suppose $Z = k$. In any population or sample of measures no more than $1/k^2$ of the raw scores can differ from the mean by k or more STDEVs... at least $(1 - 1/k^2)$ of the values will fall within $k = 1$ STDEV of the mean.

Example: ($|Z| = 4$) no more than 1/16th (6.25%) of the total scores in a sample can differ from the mean by 4 or more STDEVs .. At least $(1 - 1/4^2)$ or 93.75% of the values will fall within 2 STDEV of the mean.

Generally: the larger the Z score, the rarer the event...

Sampling Distribution of mean

- Conclusions

- **Sample means are less scattered** about the overall mean than are the individual observations.
- We can **rely more on the sample means** of independent, randomly chosen specimens than on a single specimen.

Sampling from Normal Populations

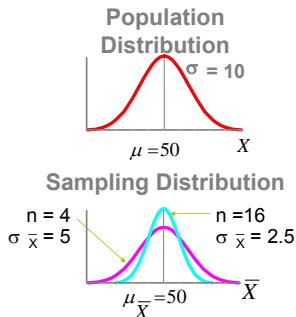
Central Tendency

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

Dispersion

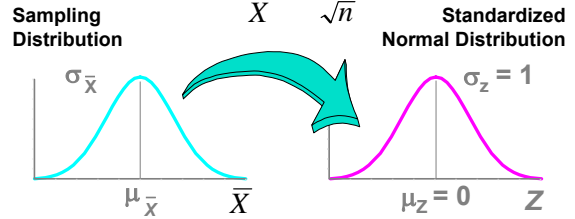
$$\sigma_{\bar{X}} = \frac{\sigma}{\sqrt{n}}$$

Sampling **With**
Replacement

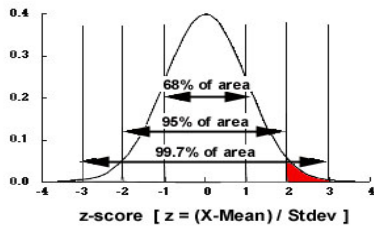


Standardizing Sampling Distribution of Mean – Z-score

$$Z = \frac{\bar{X} - \mu}{\frac{\sigma}{\sqrt{n}}} = \frac{\bar{X} - \mu}{\sigma_{\bar{X}}}$$



Standard Normal Distribution (μ=0, σ=1)



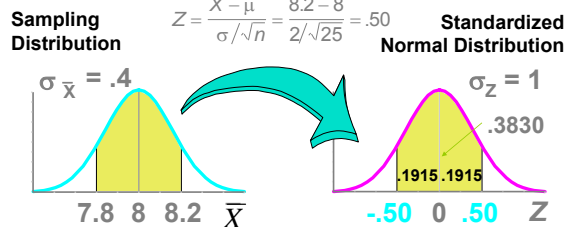
Exercise

You're an operations analyst for AT&T. Long-distance telephone calls are normally distributed with $\mu = 8$ min. & $\sigma = 2$ min. If you select random samples of **25** calls, what percentage (probability) of the **sample means** would be between **7.8 & 8.2** minutes?

Example (continued)

$$Z = \frac{\bar{X} - \mu}{\frac{\sigma}{\sqrt{n}}} = \frac{7.8 - 8}{2/\sqrt{25}} = -.50$$

$$Z = \frac{\bar{X} - \mu}{\frac{\sigma}{\sqrt{n}}} = \frac{8.2 - 8}{2/\sqrt{25}} = .50$$



Normal Distribution

- Computer exercise

- Select a parameter in the data set.
- Calculate for mean and Stdev
- Find probabilities of a value that is
 - 1.25*mean or less
 - 1.2*mean or more
 - 0.75*mean or less, and
 - between (0.75-1.25)*mean