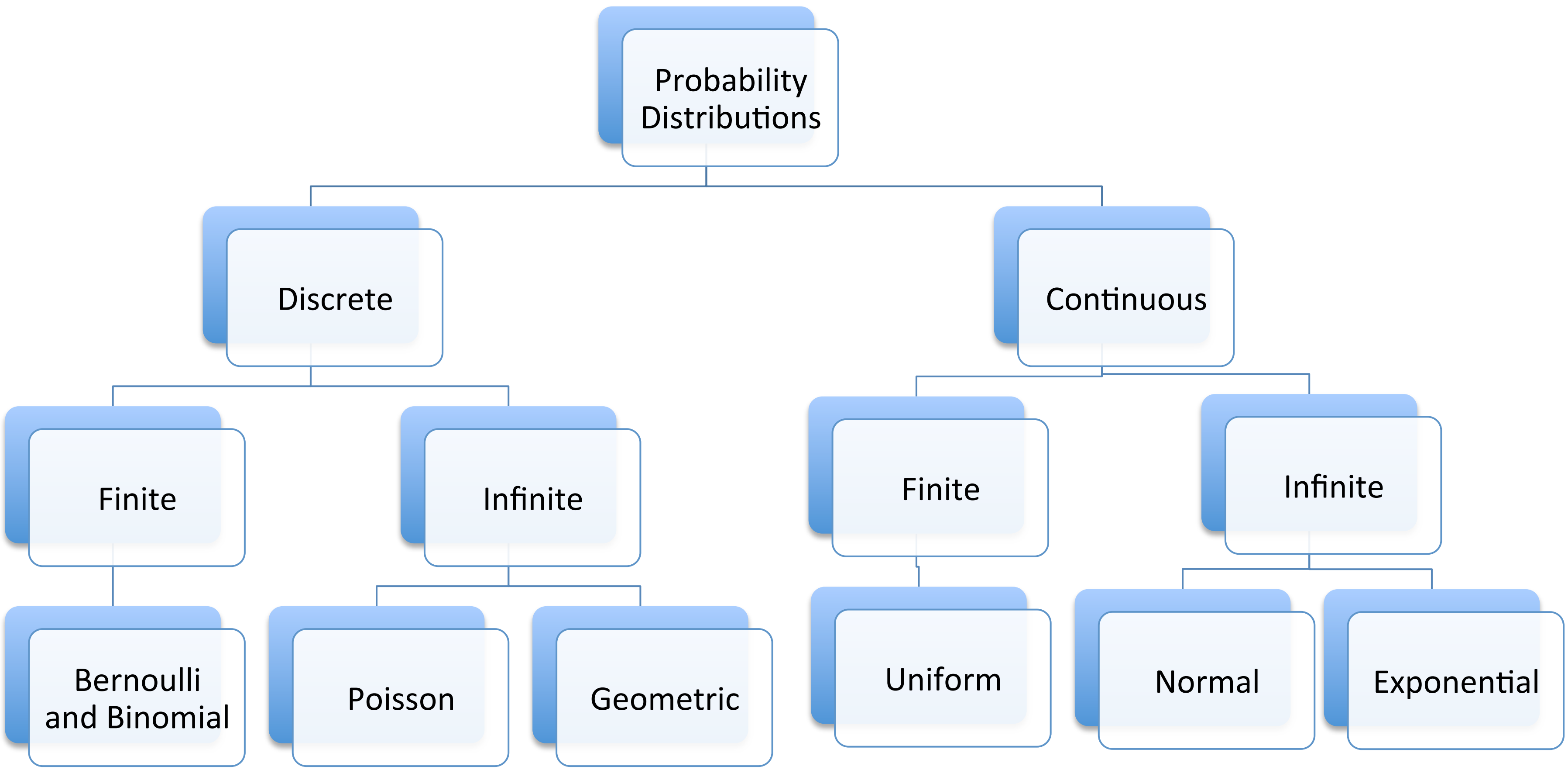
Continuous Distributions: Normal

Keegan Korthauer

Department of Statistics

UW Madison

Some Common Distributions



CONTINUOUS DISTRIBUTIONS

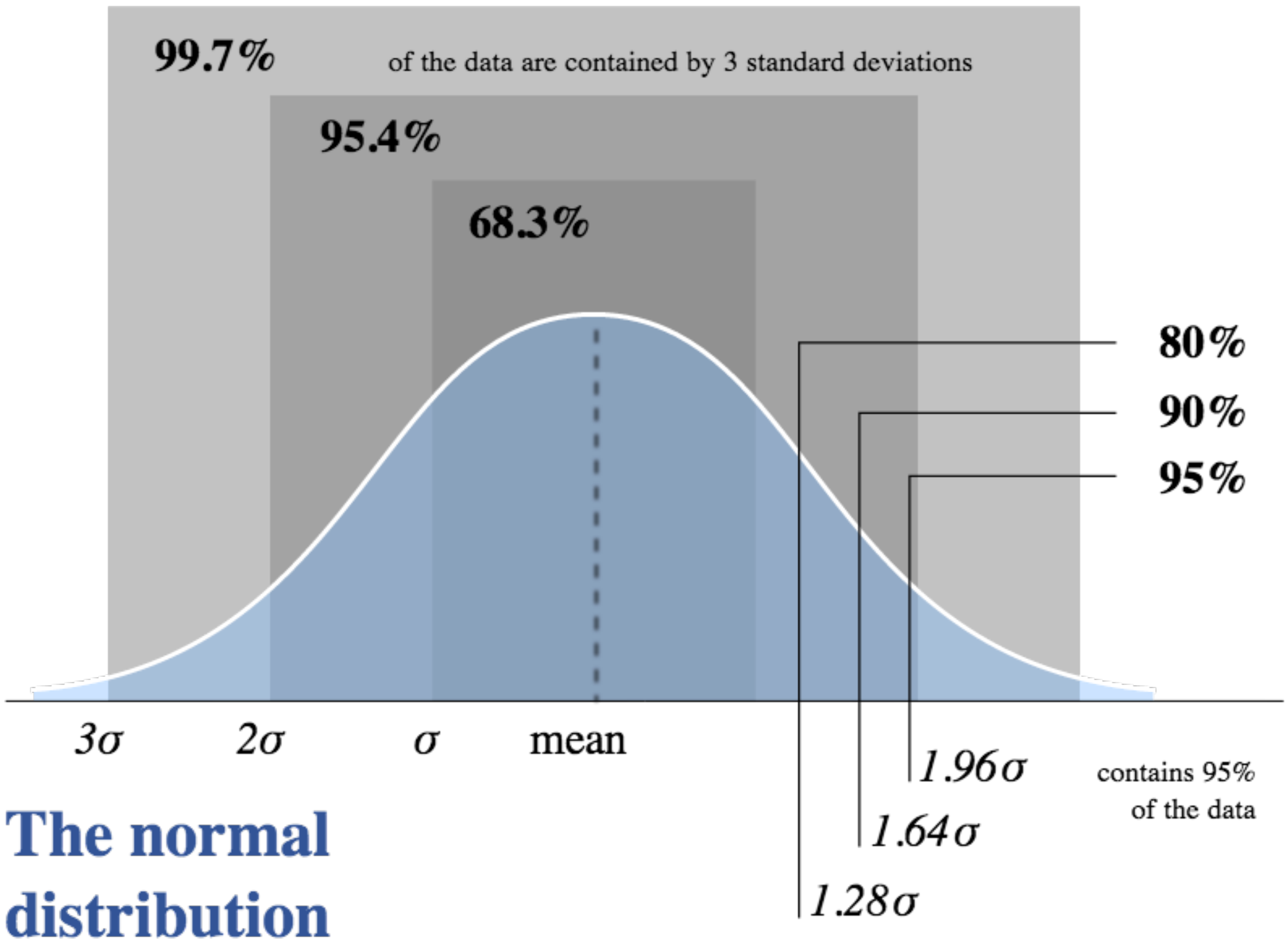
Normal distribution

Exponential distribution

Continuous Uniform Distribution

Average Diameter of 1000 Steel Rods

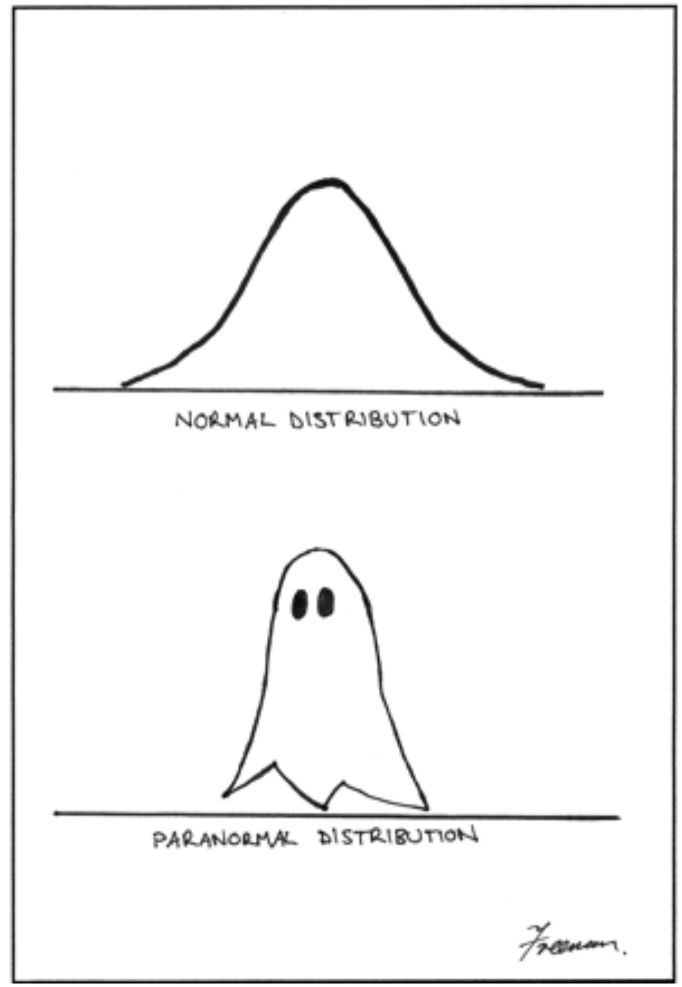
Bell-Shaped Curve



http://subsurfwiki.org/wiki/Normal_distribution

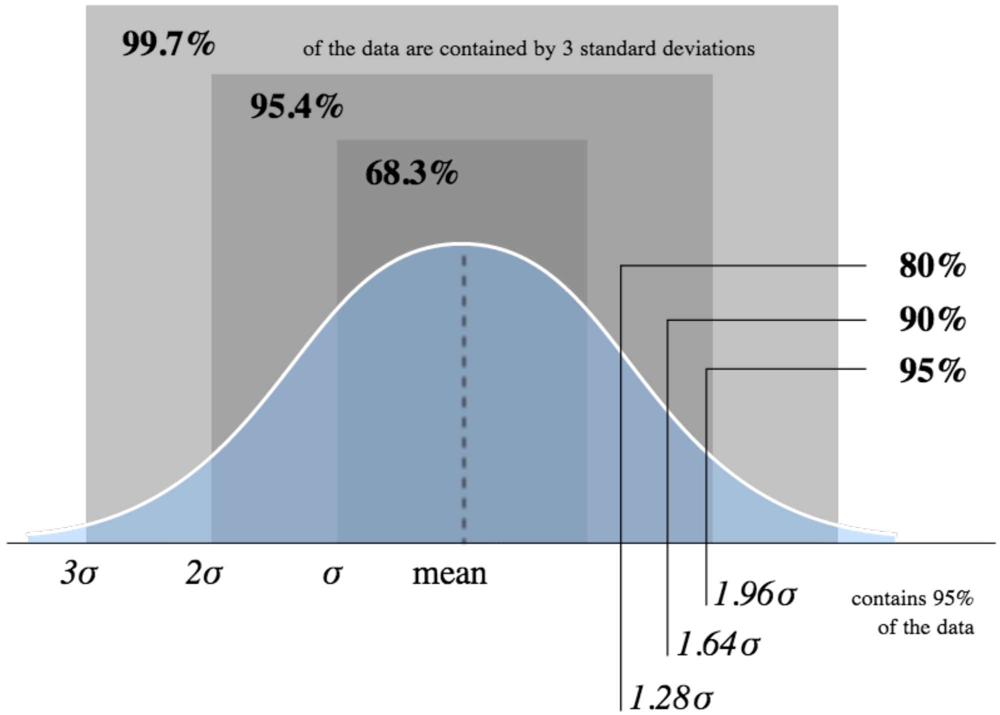
Properties of the Normal Distribution

- Continuous and infinite
- Symmetric about the median and mean
- Unimodal
- Observed in many naturally occurring phenomena
  - height and blood pressure
  - lengths of objects produced by machines
- Parameterized by its mean μ and variance σ^2



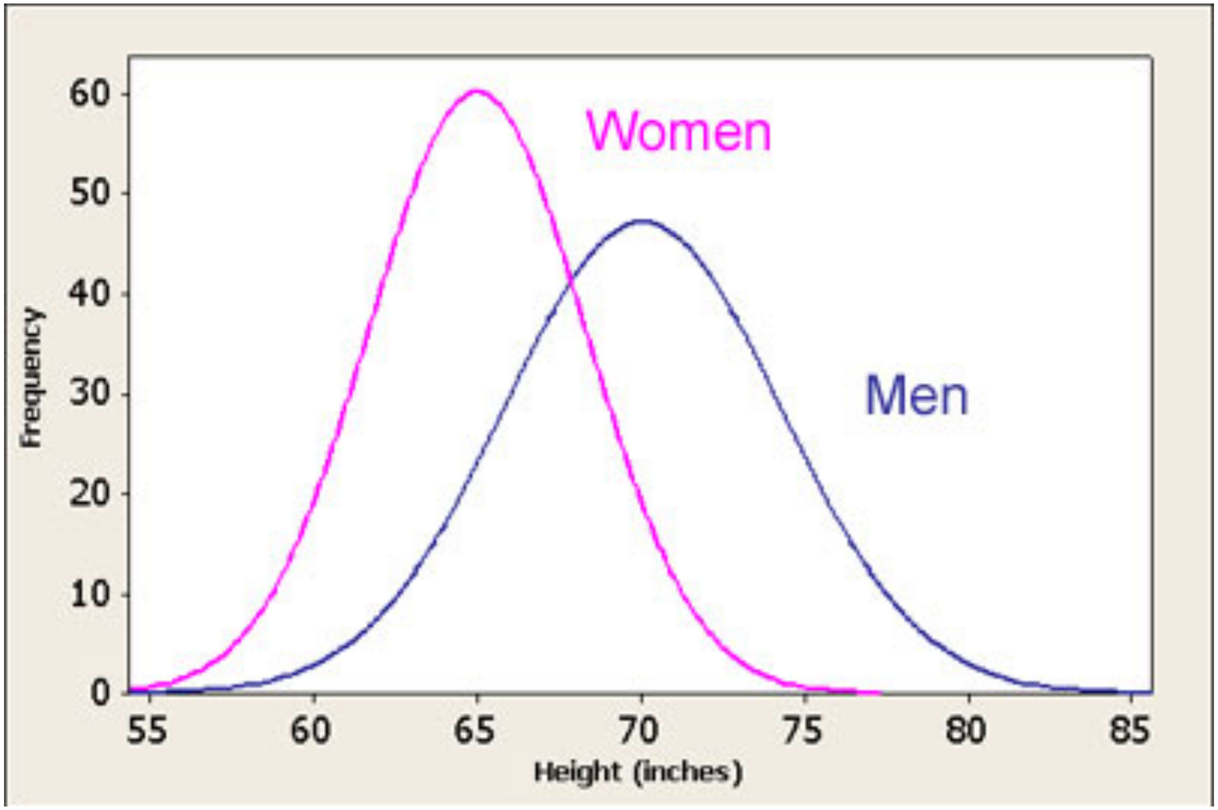
http://jech.bmj.com/content/60/1/6.full

Properties of the Normal Distribution



- About 68% of the population is in the interval $\mu \pm \sigma$
- About 95% of the population is in the interval $\mu \pm 2\sigma$
- About 99.7% of the population is in the interval $\mu \pm 3\sigma$

Example – Heights



- How likely are we to randomly sample a woman who is 73 inches tall?
- How about a man who is 61 inches tall?

X ~ Normal(μ,σ²)

- Probability density function

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

- Mean and variance

$$\mu_X = \mu$$

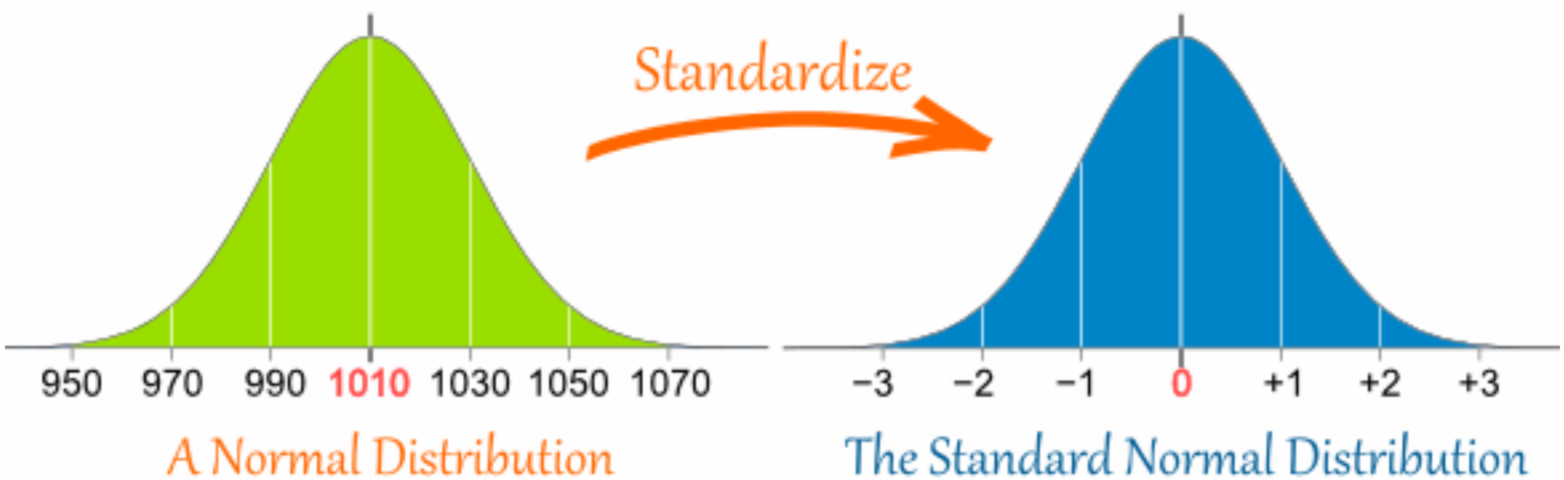
$$\sigma_X^2 = \sigma^2$$

Normal Distribution – Illustration

- http://www.intmath.com/counting-probability/normal-distribution-graph-interactive.php
- Shape of the distribution (PDF) is determined by the parameters μ and σ^2
  - Note that the range of x is all real numbers
  - Changing μ shifts the curve left and right
  - Increasing σ^2 flattens the curve
  - Decreasing σ^2 makes the curve taller and narrower

Standard Normal Distribution

- When X ~ $N(\mu,\sigma^2)$ with $\mu = 0$ and $\sigma^2 = 1$, we say that X comes from the **standard normal distribution**
- We can **standardize** any normally distributed random variable to be N(0,1) by subtracting its mean and dividing by its standard deviation:



http://www.mathsisfun.com/data/standard-normal-distribution.html

Z-Scores

- If X ~ Normal(μ,σ^2), then Z=(x-μ)/σ ~ N(0,1)
- The 'z-score' is used to standardize a sample to the standard normal distribution

$$z = \frac{x - \mu}{\sigma} = \text{"z-score"}$$

- It represents how many standard deviations away from the mean an observation is

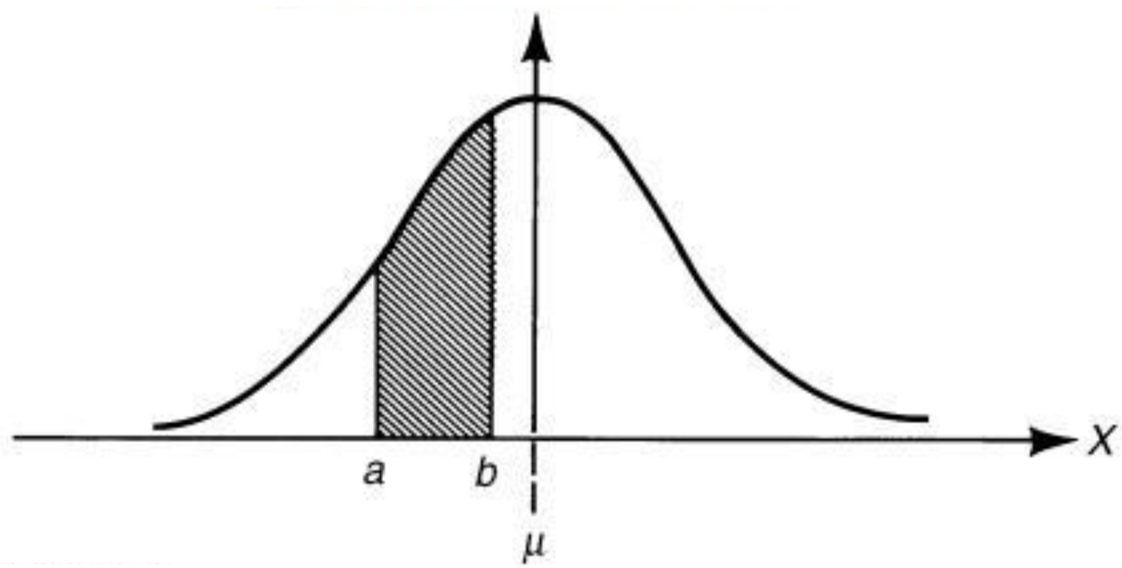
Examples 4.39-40

Aluminum sheets used to make beverage cans have a thickness (in thousandths of an inch) that are normally distributed with mean 10 and standard deviation 1.3

- A particular sheet is 10.8 thousandths of an inch thick. Find its z-score
- A particular sheet has a z-score of -1.7. Find the thickness of the original sheet (in thousandths of inches)

Probabilities from the Normal Distribution

- Just like before we can get the probability that $X \sim N(\mu, \sigma^2)$ is between a and b by integrating the density function between a and b.
- But… the integral involves an infinite series and needs to be approximated numerically



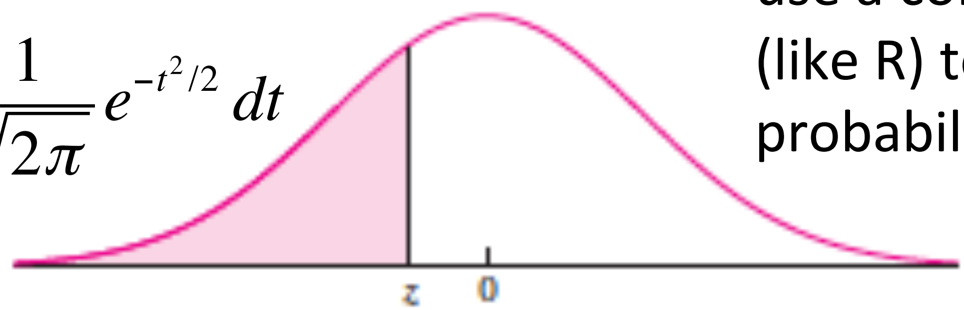
Standard Normal Table

Values of the integral P(Z < z) for Z~N(0,1) are pre-computed for many values of z and displayed in a table in the front cover (and appendix) of your text

In Practice, statisticians use a computer program (like R) to evaluate the probability

TABLE A.2 Cumulative normal distribution (z table)

$$P(Z \le z) = \int_{-\infty}^{z} \frac{1}{\sqrt{2\pi}} e^{-t^2/2}\, dt$$

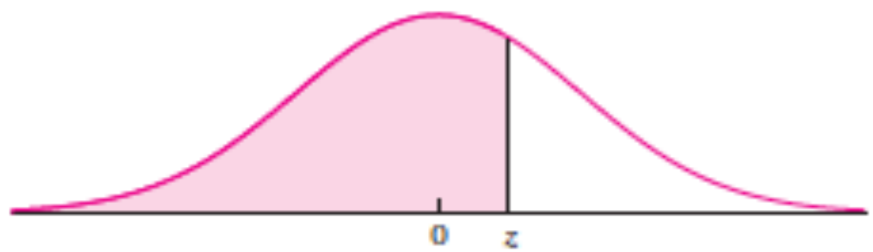


z	0.00	0.01	0.02	0.03	0.04	0.05	0.06	0.07	0.08	0.09
−3.6	.0002	.0002	.0001	.0001	.0001	.0001	.0001	.0001	.0001	.0001
−3.5	.0002	.0002	.0002	.0002	.0002	.0002	.0002	.0002	.0002	.0002
−3.4	.0003	.0003	.0003	.0003	.0003	.0003	.0003	.0003	.0003	.0002
−3.3	.0005	.0005	.0005	.0004	.0004	.0004	.0004	.0004	.0004	.0003
−3.2	.0007	.0007	.0006	.0006	.0006	.0006	.0006	.0005	.0005	.0005
−3.1	.0010	.0009	.0009	.0009	.0008	.0008	.0008	.0008	.0007	.0007
−3.0	.0013	.0013	.0013	.0012	.0012	.0011	.0011	.0011	.0010	.0010

Example 4.41 – Using the Normal Table

Find the area under the normal curve to the left of z=0.47

TABLE A.2 Cumulative normal distribution (continued)



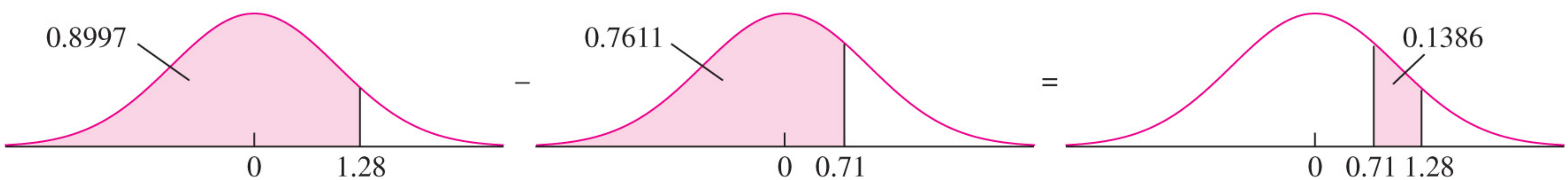
z	0.00	0.01	0.02	0.03	0.04	0.05	0.06	0.07	0.08	0.09
0.0	.5000	.5040	.5080	.5120	.5160	.5199	.5239	.5279	.5319	.5359
0.1	.5398	.5438	.5478	.5517	.5557	.5596	.5636	.5675	.5714	.5753
0.2	.5793	.5832	.5871	.5910	.5948	.5987	.6026	.6064	.6103	.6141
0.3	.6179	.6217	.6255	.6293	.6331	.6368	.6406	.6443	.6480	.6517
0.4	.6554	.6591	.6628	.6664	.6700	.6736	.6772	.6808	.6844	.6879

Finding Center Areas/Right Tails

- The table only gives us 'left tails': the probability that the normal RV Z is less than z
- We can use these to find other areas since for a continuous RV X,

$$P(a < X < b) = P(X < b) - P(X < a)$$ and

$$P(X > b) = 1 - P(X < b)$$



Examples 4.47-49

A process manufactures ball bearings with normally distributed diameters with mean 2.505cm and sd 0.008cm.

- We need the diameter to be in the interval 2.5 ± 0.01cm. What proportion will meet this criteria?
- If the process is recalibrated so that the mean is exactly 2.5cm, what proportion will meet the criteria?
- To what value must the sd be lowered so that 95% will meet the criteria?

What if μ and σ are Unknown?

- Estimate the mean with the sample mean

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

  – Bias: $E(\bar{X}) - \mu = \mu - \mu = 0$

  – Uncertainty: $\sigma_{\bar{X}} = \sigma / n \approx s / n$

- Estimate the variance with the sample variance

$$\hat{\sigma}^2 = s^2$$

Linear Functions of Independent Normal Random Variables

- Let $X \sim N(\mu,\sigma^2)$ and let *a* and *b* be constants (where *a* is nonzero). Then

$$\boldsymbol{aX + b \sim N(a\mu + b,\ a^2\sigma^2)}$$

- Generalize to n independent normal RVs $X_1,\dots,X_n$ with means $\mu_1,\dots,\mu_n$ and variances $\sigma^2{}_1,\dots,\ \sigma^2{}_n$. Let $a_1,\dots,a_n$ be constants. Then

$$\boldsymbol{a_1X_1 + \dots + a_nX_n \sim N(a_1\mu_1 + \dots + a_n\mu_n,\ \ a^2{}_1\sigma^2{}_1 + \dots + a^2{}_n\sigma^2{}_n\ )}$$

Important Linear Combinations

- Sample Mean:

$$\bar{X} \sim N\left(\mu, \frac{\sigma^2}{n}\right)$$

- Sum and Difference of Two Normal RVs

$$X + Y \sim N(\mu_X + \mu_Y,\ \sigma_X^2 + \sigma_Y^2)$$

$$X - Y \sim N(\mu_X - \mu_Y,\ \sigma_X^2 + \sigma_Y^2)$$

Example – Exercise 4.5.21

- A light fixture holds two bulbs:
  - Type A with normally distributed lifetime, mean 800 hrs and sd 100 hrs
  - Type B with normally distributed lifetime, mean 900 hrs and sd 150 hrs

- What is the probability that bulb B lasts longer than bulb A?

Next

Two more continuous distributions:

- Uniform
- Exponential