**STT 200** **Spring 2009**

# Lecture Outline 1 - 12 - 09

You will learn a lot of useful statistics before Spring Break. My style is to quickly introduce backbone ideas and skills and I will not put too fine a point on it. When something new is placed before you expect that examples of its usefulness will be not far behind.

**Keep up.** Don't think you can cram later, it won't work. Wednesday's lecture will move on to another topic and so it goes. The first exam is in Tuesday recitation just before Spring Break. Get your questions answered in class, in recitation, in office hours, in help-room C100 Wells Hall, or through cooperative study with your class-mates. Don't be shy, **keep up**.

## Estimated Margin of Error for $\hat{p}$.

Consult page 501 and the supporting readings pp. 486-500.

**A Question:** What is the probability p that a page selected at random from pp. 1 to 767 in our textbook has a (non-table) picture or a graphic? I've randomly sampled 36 pages by perusing the table of random digits A-94, skipping over those outside the range 001 to 767 (consult the table).

716, **032**, 463  473, **200**, 731  **727**, 039, 759  (skip 944)
043 (skip 890 and 877)  **764**, **364**, **132**  512, 678, **098**
**181**, **027**, 133  **622** (skip 922)  **666**, **310** (skip 844)
**720** (skip 945)  639, **112**, **285** (skip duplicate 112)
429, **471** (skip duplicate 112)  **647** (skip 770)
**183**, 071, 359  412, **585**, 428  **042**

**1a.** Every page of the book had the same chance of making it into this sample of n = 36 pages. I've highlighted in **bold** those sample pages with a picture or graphic.

**Point estimate of p** = $\frac{20}{36}$ ~ 0.555556.

**Estimated margin of error of the estimator $\hat{p}$** is calculated as follows:

$$1.96\ \frac{\sqrt{\hat{p}\,(1-\hat{p})}}{\sqrt{n}}\ \sqrt{\frac{N-n}{N-1}} = 1.96\ \frac{\sqrt{0.555556\,(1-0.555556)}}{\sqrt{36}}\ \sqrt{\frac{767-36}{767-1}}$$

$$\sim 0.15857$$

**1b.** Claim made for estimated margin of error:

Around 95% **of random samples of n = 36 pages** from the book of N = 767 pages will produce an interval

$\hat{p}$ ± estimated margin of error

that will cover the actual value of p = fraction of pages with a picture or graphic.

Our random sample of n = 36 pages from the N = 767 pages of the book has produced the interval

$\hat{p}$ ± estimated margin of error

= 0.555556 ± 0.15857

= [0.396986, 0.714126]

This interval is called a "95% confidence interval for p."

**1c.** Around 95% of samples of 36 are "good samples" whose 95% confidence interval **$\hat{p}$ ± estimated margin of error** covers the true value of p. Is our sample of 36 a "good one?" We don't know. Nonetheless, the method of reporting the sampling results does give a pretty good idea of the reliability of the findings. ***For this 95% chance to apply we require that n is "large enough" and so too is N-n. At this point we're not putting too fine a point on it. Consult the chapter readings. Our example 3b below (n = 10 and N = 26) is pushing the boundaries of applicability.***

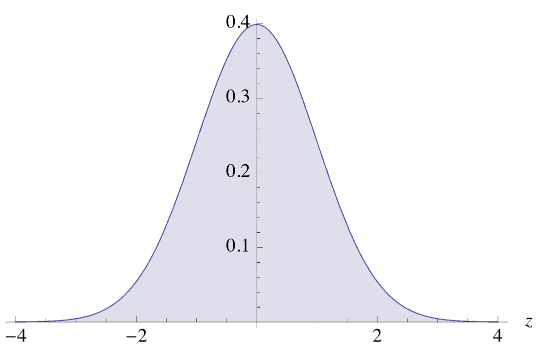
**2. BELL CURVE.** Consult Table Z, pp. A-95-A96. This is a table of areas captured under the Standard Normal Curve (sometimes called the "bell curve" or "z-curve"). At this early point of the course you are only tasked with learning the skill of using this table to find the area under the curve between two specified z values.

The bell curve has mathematical form

$$f(z) = \frac{e^{-\frac{z^2}{2}}}{\sqrt{2\pi}}, -\infty < z < \infty.$$

**2a.** The area under f is exactly one. All but a tiny fraction of this area is found between the limits of z = -∞ and z = 4. Table Z approximates this area

| **z** | **0.00** | |
|---|---|---|
| **3.9** | **1.0000** | **(= 1 to four decimals)** |



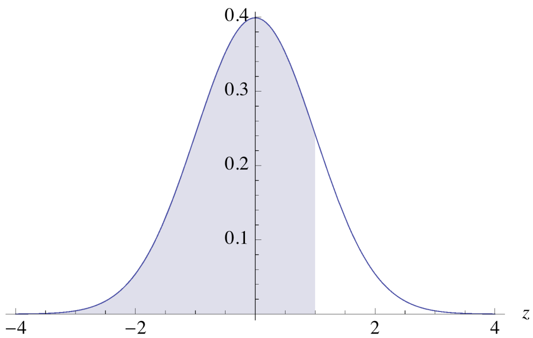

**~ 1.0000 (Table Z)**

A typical laptop computer can quickly give us much more accuracy

0.999968328758166880078746229243277848701556166624519723491450668..

**2b.** The area between z = -∞ and z = 1.00 is found

| **z** | **0.00** |
|---|---|
| **1.0** | **0.8413** |

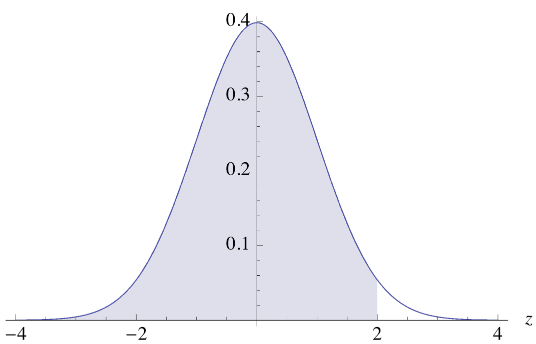



**~ 0.8413 (Table Z)**

**2c.** The area between z = -1.00 and z = 1.00 is found using symmetry as
~ 2 (.8413 - 0.5) = 0.6826 (twice the area from z = 0 to z = 1).

**2d.** The area between z = -2 and z = 2 is found as twice the area between z = 0 and z = 2 using

| **z** | **0.00** |
|---|---|
| **2.0** | **0.97725** |




**~ 0.97725 (Table Z)**

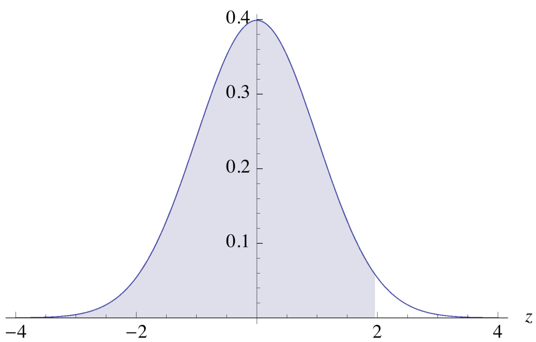
The area between z = -2 and z = 2 is
~ 2 (.97725 - 0.5) = 0.9545 (Table Z)

**2e.** RULE OF THUMB

z-curve (bell curve) puts ~ 68% of area between limits z = -1 and z = 1

~ 95% of area between limits z = -1.96 and z = 1.96

| **z** | **0.06** |
|---|---|
| **1.9** | **0.9750** |




**~ 0.9750 (Table Z)**

The area between -1.96 and 1.96 is ~ 2 (0.9750 - 0.5) = 0.95 (Table Z).

# 3. Examples for you.

**3a.** A random sample of 300 registered voters is selected and each is asked whether they favor a particular ballot initiative. Of the 300 there are 167 who favor the proposal.

$\hat{p}$ =

estimated margin of error of $\hat{p}$ =

Let p denote the fraction of all registered voters who are in favor of the proposal.

You may assume any really large number N for the number of registered voters.

95% confidence interval for p (the fraction of all registered voters
=

Approximate chance that the 95% confidence interval covers p =

**3b.** (unrelated to 3a) Use the table of random digits A-94 to select a random sample of 10 different letters of the alphabet. Hint: number the letters
a = 00, b = 01, ....., z = 26 (vowels a, e, i, o, u)

**3c.** Use your sample to estimate p = fraction of vowels in the alphabet.

$\hat{p}$ =

Estimated margin of error of $\hat{p}$ =

95% confidence interval for p =

Approximate chance a 95% confidence interval using n = 10 will cover p =

Has your particular sample of 10 produced a 95% confidence interval covering p?

**4.** We require an estimate of the fraction of telephones (517) 353xxxx that will be answered by a person between 6 and 7 p.m. Wednesday.

**4a.** How would you use the table of random digits to dial a random sample of 50 telephone numbers (517) 353-xxxx?

**4b.** If you find that 18 are answered by a person what is $\hat{p}$?

**4c.** Per 4b, give your 95% confidence interval for p. Use any large N and notice that your answer is not sensitive to changes in (a large) N.