STT 200 Spring 2009

Lecture Outline 1 - 21 - 09

Estimated Margin of Error for difference of sample means

$\bar{x}_1 - \bar{x}_2$. This follow-on to lecture 1-14-09 extends the margin of error concept given there for a single sample average $\bar{x}$. We require samples sizes n_1, N_1-n_1, n_2, N_2-n_2 be all "large enough" so that a bell curve approximation is justified. I will **no**t introduce Student's-t at this point, so the readings of pp. 616-619 are relevant except the bottom of 619.

A Question: What is the difference μ_1-μ_2 where

μ_1 = average number of graphics per page in pages 1 through 383

μ_2 = average number of graphics per page in pages 384 through 767

of our textboook?

A solution. Randomly sample 17 pages from 001 through 383. Independently of this sample 19 pages from 383 through 767 (384 pages).

Estimate μ_1-μ_2 by the difference of sample means $\bar{x}_1$- $\bar{x}_2$where

$\bar{x}_1$ = average (mean) of graphics & picture counts for 17 sample pages in 001-383

$\bar{x}_2$ = average (mean) of graphics & picture counts for 19 sample pages in 384-767

The estimated margin of error in this setup is

$$1.96 \sqrt{\frac{s_1^2}{n_1}\frac{N_1-n_1}{N_1-1} + \frac{s_2^2}{n_2}\frac{N_2-n_2}{N_2-1}}$$

The samples. 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). Those in **bold** have graphics and I've indicated the number of graphics in parentheses (graphic sub-components of a graphic display are not counted individually. It is a serious point, needing clarification for serious work, but let's just suppose that we'recounting graphics drawing attention to markedly different parts of the page.

716	**32 (1)**	463
473	**200 (1)**	731
727 (1)	39	759
"(skip 944)"	43	"(skip 890 and 877)"
764 (3)	**364 (1)**	**132 (1)**
512	678	**98 (3)**
181 (3)	**27 (1)**	133
622 (2)	"(skip 922)"	**666 (1)**
310 (1)	"(skip 844)"	**720 (1)**
"(skip 945)"	639	**112 (5)**
285 (1)	"(skip duplicate 112)"	429
471 (1)	"(skip duplicate 112)"	**647 (1)**
"(skip 770)"	**183 (1)**	71
359	412	**585 (1)**
428	**42 (2)**	

Here are the 36 sample pages sorted:

{**27(1)**, **32(1)**, 39, **42(2)**, 43, 71, **98(3)**, **112(5)**, **132(1)**, 133, **181(3)**, **183(1)**, **200(1)**, **285(1)**, **310(1)**, 359, **364(1)**, 412, 428, 429, 463, **471(1)**, 473, 512, **585(1)**, **622(2)**, 639, **647(1)**, **666(1)**, 678, 716, **720(1)**, **727(1)**, 731, 759, **764(3)**}

Those sample pages less than 384 number 17 and constitute a random sample of 17 from 383. The others number 19 and constitute a random sample of 19 from the 384 pages 384-767.

1a. The sample average number of graphics per page is (for pages < 384)

$$\bar{x}_1 = \frac{\sum_i x_{1i}}{n_1} = \frac{1+1+0+2+0+0+3+5+1+0+3+1+1+1+1+0+1}{17}$$

$$\sim 1.23529.$$

The sample standard deviation s_1 (see page 64) is

$$s_1 \sim 1.14564.$$

The sample average number of graphics per page is (for pages > 383)

$$\bar{x}_2 = \frac{\sum_i x_{2i}}{n_2} = \frac{0+0+0+0+1+0+0+1+2+0+1+1+0+0+1+1+0+0+3}{19}$$

$$\sim 0.656347$$

The sample standard deviation s (see page 64) is

$$s_2 \sim 0.781736.$$

The difference of sample means is

$$\bar{x}_1 - \bar{x}_2 \sim 1.23529 - 0.656347 \sim 0.656347.$$

1b.

Estimated margin of error of the estimator $\bar{x}_1 - \bar{x}_2$ is calculated as follows:

$$1.96\sqrt{\frac{s_1^2}{n_1}\frac{N_1-n_1}{N_1-1} + \frac{s_2^2}{n_2}\frac{N_2-n_2}{N_2-1}}$$

$$= 1.96\sqrt{\frac{1.3125}{17}\frac{383-17}{383-1} + \frac{0.61111}{19}\frac{384-19}{384-1}}$$

$$\sim 0.633975$$

1c. Claim made for estimated margin of error: Provided n_1, N_1-n_1, n_2, N_2-n_2 are all "large enough"

Around 95% **of random samples of 17 from 383 and 19 from 384 have**

$$\bar{x}_1 - \bar{x}_2 \pm \text{estimated margin of error}$$

that will cover the actual value of $\mu_1 - \mu_2$.

Our $n_1 = 17$ and $n_2 = 19$ are on the small side so we'll have to say the confidence level will not so closely approximate 95%. But let's press on snce the data are still informative.

Our random sample of 17 from 383 and 19 from 384 pages of the book has produced the 95% confidence interval

$$\bar{x}_1 - \bar{x}_2 \pm \text{estimated margin of error}$$

$$= 0.656347 \pm 0.633975$$

$$= [0.022372, 1.29032]$$

So either $\mu_1 - \mu_2$ is not zero or the confidence interval has failed to cover (~ 5% chance).