STT 200 Spring 2009

Lecture Outline 1 - 16 - 09

Estimated Margin of Error for $\hat{p}_1 - \hat{p}_2$.

Consult page 567 and the supporting readings pp. 557-563.

Random page sampling. In the posted outline for lecture 1-12-09 we randomly sampled 36 page numbers from 001 to 767.

716	**32**	463
473	**200**	731
727	39	759
"(skip 944)"	43	"(skip 890 and 877)"
764	**364**	**132**
512	678	**98**
181	**27**	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**	71
359	412	**585**
428	**42**	

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 having a picture or graphic.

Here are the 36 sample pages sorted:

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

A Question. Some pages of our book have a picture or graphic. How do the frequencies of such pages compare in the first half of the book versus the last half?

Our solution. We'll estimate the difference $p_1 - p_2$ between the probabilities:

p_1 = probability that a page < 384 has a picture or graphic

p_2 = probability that a page > 383 has a picture or graphic

using the data above.

Examining the data we find

a. Sample pages < 384 vs > 383 total 17 early, 19 late.

b. Sample pages total 20 with picture/graphic, 16 without.

Try to be careful in the presence of such a *coincidence*. It will happen from time to time and can be a source of major error if we're not observant.

It helps to organize these counts in a ***contingency table***:

Contingency counts for 36 random pages:

	< 384	> 383	total
with p/g	**12**	**8**	**20**
without	**5**	**11**	**16**
total	**17**	**19**	**36**

* We may regard the 17 sample pages < 384 as a random sample of 17 pages from the first half of the book (383 pages).
* Likewise, we may regard the 19 sample pages > 383 as a random sample of 19 pages from the second half of the book (384 pages).
* The sample of 17 pages from the first half is ***statistically independent*** of the sample of 19 pages from the second half in the sense that *neither can tell us anything about the other that we would not know anyway*.

Statistical independence is a backbone concept in statistics, and is largely responsible for the fabled "Law of Averages" which we study later.

Recall the Question: Some pages of our book have a picture or graphic. How do the frequencies of such pages compare between the first half of the book and the last half?

Recall our solution. We'll estimate the difference $\boldsymbol{p_1 - p_2}$ between the probabilities:

p_1 = probability that a page < 384 has a picture or graphic (of 383 pages)

p_2 = probability that a page > 383 has a picture or graphic (of 384 pages)

using the data from the contingency table above.

1a. Needed estimates.

Point estimate of p_1 is $\frac{12}{17} \sim 0.70588$

(this is $\hat{p}_1$, we selected 17 pages < 384 of which 12 had a picture or graphic).

Point estimate of p_2 is $\frac{8}{19} \sim 0.42105$

(this is $\hat{p}_2$, we selected 19 pages > 383 of which 8 had a picture or graphic).

Our point estimate of $p_1 - p_2$ is $\hat{p}_1 - \hat{p}_2 \sim 0.70588 - 0.42105 \sim 0.28483$

(this seems to be a fairly large difference, but what about the estimated margin of error of $\hat{p}_1 - \hat{p}_2$)?

1b. Estimated margin of error of the estimator $\hat{p}_1 - \hat{p}_2 =$

$$1.96 \sqrt{\frac{\hat{p}_1(1-\hat{p}_1)}{n_1} \frac{N_1-n_1}{N_1-1} + \frac{\hat{p}_2(1-\hat{p}_2)}{n_2} \frac{N_2-n_2}{N_2-1}}$$

$$= 1.96 \sqrt{\frac{.70588\,(1-.70588)}{17} \frac{383-17}{383-1} + \frac{.42105\,(1-.42105)}{19} \frac{384-19}{384-1}}$$

$\sim$ **0.303185**

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**

$\hat{p}_1 - \hat{p}_2 \pm$ estimated margin of error

that will cover the actual value of $p_1 - p_2$.

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 has produced the interval

$$\hat{p} \pm \text{estimated margin of error}$$
$$= 0.28483 \pm 0.303185$$
$$= [-0.01835, 0.58801]$$

This 95% confidence interval for $p_1 - p_2$ just contains 0, suggesting that we cannot confidently exclude the poossibility that $p_1 = p_2$. It is within the estimated margin of error of our point estimate $\hat{\boldsymbol{p}}_1 - \hat{\boldsymbol{p}}_2 = 0.28483$ of $p_1 - p_2$.