Math 5330 Homework 5
Spring 2009 **Factoring--Results**

In cmpiling these results, I rounded things to an even 20 seed numbers and did 1000 consecutive integers for each number. This gave a total of 20,000 trials.

2. **How many of your 100 numbers were prime? How does this compare with the expected number of primes?**

From our 20,000 numbers, we generated 1355 primes. The expected number of primes was approximately $\sum_{i=1}^{20} \frac{1000}{\ln(\text{number } i)} = 1333$. So we were up about 22 from the expected value. This sounds bad, but put another way, we were within 2% of the expected value.

3. **How many of the 7-digit numbers were squares? How does this compare with the expected number of squares?**

There were a total of 7 squares. If the data had just been over 100 integers a piece, there would have been 2 squares. How might one estimate how many squares we should have gotten? If $f(x)$ is an approximation for the number of integers less than x which have some property, then $m f'(n)$ will be an estimate for the number of things having that property in a list of m numbers starting at n. In our case, $f(x) = \sqrt{x}$, so the expected number for any one person will be $\frac{1000}{2\sqrt{n}} = \frac{500}{\sqrt{n}}$. The total expected number for our class is: $.407 + .317 + \cdots + .229 \cong 5.7$. So we were slightly lucky. On lists of 100 numbers, the total would have been .57, so we were a bit lucy here as well.

4. **What is the distribution of the number of distinct primes dividing each of your numbers? What is the mean and standard deviation of this distribution?**

person	1	2	3	4	5	6	7	sq	mean	SD
MB	73	261	368	235	57	6	0	0	2.960	1.034
RD	69	246	385	226	67	7	0	0	2.997	1.039
ME	64	259	336	261	70	10	0	1	3.044	1.070
RF	65	233	397	228	69	8	0	0	3.027	1.034
BF	73	263	390	216	53	5	0	0	3.051	1.047
RG	73	263	390	216	53	5	0	1	2.928	1.011
JG	68	246	369	242	66	9	0	1	3.019	1.051
MH	78	230	379	242	62	9	0	0	3.007	1.055
ZH	66	267	359	239	63	6	0	1	2.984	1.037
CI	64	255	366	235	75	5	0	1	3.017	1.044
SJ	69	275	363	236	53	3	1	1	2.942	1.019
JK	72	232	364	241	87	4	0	0	3.051	1.069
ML	68	217	390	237	78	10	0	0	3.070	1.060
MN	70	237	369	254	63	7	0	0	3.024	1.042
EP	65	253	338	256	76	12	0	0	3.061	1.085
MR	69	258	385	219	64	5	0	0	2.966	1.026
DR	71	255	351	253	64	5	1	0	3.003	1.053
LS	59	287	357	232	60	5	0	1	2.962	1.019
SS	69	245	352	254	70	10	0	0	3.041	1.069
DW	65	253	353	253	69	7	0	0	3.029	1.051
Total	1355	5018	7347	4796	1340	141	3	7	3.009	1.047

What should we get for the mean and standard deviation? Here is one theorem: The average number of prime factors of n is $\ln(\ln(n)) + .261$

This means that our expected values should have been:
$2.916 + 2.951 + 3.001 + 2.994 + 3.003 + 2.889 + \cdots + 2.994$ for a class average of 2.971. I've found a reference that says the standard deviation should be $\sqrt{\ln(\ln(n))} \cong 1.6$ but this seems to be way too big for our data. I find this every year!

5. **How many of your numbers have their largest prime factor larger than $n^{3/4}$, $n^{1/2}$, $n^{1/3}$? What are the mean and standard deviation of the log of the largest prime factor divided by the log of n?**

person	$n^{3/4}$	$n^{1/2}$	$n^{1/3}$	mean	SD
MB	330	729	953	.6470	..2015
RD	313	729	955	.6434	.1999
ME	333	716	951	.6451	.2040
RF	322	725	960	.6481	.1991
BF	310	723	960	.6414	.1968
RG	322	731	958	.6483	.2008
JG	321	726	946	.6455	.2034
MH	314	731	960	.6454	.2003
ZH	330	734	955	.6493	.2022
CI	327	722	963	.6451	.2008
SJ	325	733	955	.6476	.2009
JK	318	722	960	.6449	.2021
ML	320	715	953	.6436	.2014
MN	310	725	959	.6434	.2003
EP	336	725	953	.6489	.2029
MR	329	730	959	.6468	.2002
DR	319	727	953	.6459	.2030
LS	330	724	949	.6456	.2016
SS	343	711	957	.6499	.2048
DW	327	740	959	.6478	.1999
AVE	324	726	956	.64615	.20131
Expected	288	693	951	.632	?

Here, mean and SD are the mean and standard deviation of $\frac{\ln p_k}{\ln n}$ where p_k is the largest prime divisor of the k'th number in our list of 1000 numbers. The answer is supposed to be .632, so our data was a little on the high side, but fairly close. I don't know what the expected standard deviation should be.

Note also that the expected numbers in the first three columns are consistently less than what the class found. This has been typical of the data past classes of mine have collected. I suspect that this is some systematic error in the expected values that gets smaller as n grows. Maybe if we had used 10 digit numbers instead of 7 digit numbers, we would have gotten better results here.

A good Master's project: Figure out what the right answers are supposed to be (theoretically) for the standard deviation in (5) and in (4), and try to find out how close, in general, these values should be to what we get in general for 7-digit numbers. In particular, our values in (5) all seemed a bit high. Why is this?