STAT 100 Chance and Data Analysis October 28, 2002(draft)

Today: More about Lotteries
Feedback from last Wednesday Oct 23, survey.
Response to confusing items – a start.

More about Lotteries:

Left over from last time, concerning the 6/49 lottery:

Recall: choose 6 numbers in range 1-49, match 6 numbers for the biggest prize (at least $2,000,000.)

Payout at Oct 25 draw:

$2,000,000	x 1	=2,000,000
$112,511	x 5	= 562,555
$2,113	x 213	= 450,069
$67	x 12932	= 866,444
$10	x 230152	= 2,301,520
Total Payout		$6,180,588

Quote from 6/49 literature:
"45% of sales are dedicated to the prize fund"

Tickets cost $1.
How many tickets were sold? 13,734,640 (except for rounding cents)
What would the payouts be if only 2,000,000 tickets sold?

This 6/49 lottery pays out amounts determined by the ticket sales. The BC/49 lottery is similar except that it pays out a fixed amount based on the chance of winning – no carry-overs. The revenue left over after prizes are paid is variable (determined by the chances and the particular numbers players choose) but the average yield of revenue is predictable.

When people can choose their own numbers, what numbers do they choose?

They tend to choose numbers 31 or less (birthdays). They tend to choose odd numbers (not even ones). They tend to avoid 10, 20, 30, 40. Numbers like 7, 11 are popular. They tend to choose combinations that are spread throughout the range 1-49. Would you choose the combination 1,2,3,4,5,6 ?

Why not?

Are there better numbers to choose?

Not much you can do to gain advantage, except to choose unpopular numbers, so if you win, you would not have to share your jackpot. But maybe a lot of people are doing this and the unpopular combinations suddenly become popular!

Does the square root law help to reduce variability of winnings from a lottery?

Yes and No.

No - If you invest in many public lotteries, or one lottery many times, you will probably still get no major prize but there is a small chance of a large prize. SD will not be small enough to provide a useful prediction.
SD of {one million $0s and 1 $1,000,000} = approx 1000

Yes - Average return will be about 50¢ per dollar invested (negative 50¢ net return) which will be realized if you buy a huge number of tickets. (e.g. all the possible ticket numbers!). Small SD in this case.
(Sampling without replacement, correction factor = sqrt(1-(n-1)(N-1)) = approx 0 when n is close to N, so SD of average is close to zero.)

Big Picture from lecture Oct 25:

1. Lotteries use random sampling. Lotteries are an example of a sampling process in which the "population" is known.
(In 6/49 it contains almost 14,000,000 equally likely selections).
2. Average payout in almost all lotteries is less than ticket cost (typically about 50%)
3. Carry-over jackpots may still not make a lottery a good investment (avg net return < 0 is likely)
4. Exact calculations for a real lottery can be complex, but can be simulated.
5. Square root law does not reduce variability enough to be useful in this context

Feedback from Survey:

I received 57 pieces of paper. Some had gaps, some had multiple answers. The following is an attempt to summarize the responses.

1. Most Important Idea:

Sampling Method, Random Sampling * * * * * * * * * * * * * * * *
- Sampling with and without replacement *
- Survey sampling * * *
- Quality of sampling information

Square Root Law (SD of Averages) * * * * * * *
- Distribution of Averages related to distribution of Proportions * * * *
- Law of Averages
- Averages tend to have Normal distributions

Randomized Response Technique * * * *
Estimation * * *
- Estimation of Animal populations

Zipf's Law * *
Normal Model – standard Normal *
Binomial Model
Correlation
Risk

2. Most Confusing Concept

Zipf's Law and logarithms * * * * * * * * * * * * * * * *
Averages and Proportions – connections, mean & SD * * * * * * **
Randomized Response Technique * *(neg ests) *(calc) * *
Square root law – which n? * * * *
Normal Model and Calculations * * *
Binomial Model * * *
Sampling * * *
- Sampling with and without replacement

Focus of sampling articles (Jury, Train Accounts) * * *
Correlation * *
Choice of study designs *
Solar System article *

3. Want More about …

Sampling * * * *
Binomial Model * * *
Randomized Response Technique * * *
- Calculations re Randomized Response Technique & privacy aspects

Correlation * *

Square Root Law Applications * *
Recent articles from Tanur (Zipf's Law, Jury, CPI, Census, Accounts) *
    Zipf's Law *
Randomization *
Lotteries *
Standard Normal Distribution
Business Applications
    Stock Market Indicators
Statistics in Accounting
Graphical summary and interpretation
Intentional bias in statistical reports
Randomness
Statistics Canada
Risk
Profitable strategies in a casino

4. Miscellaneous Comments

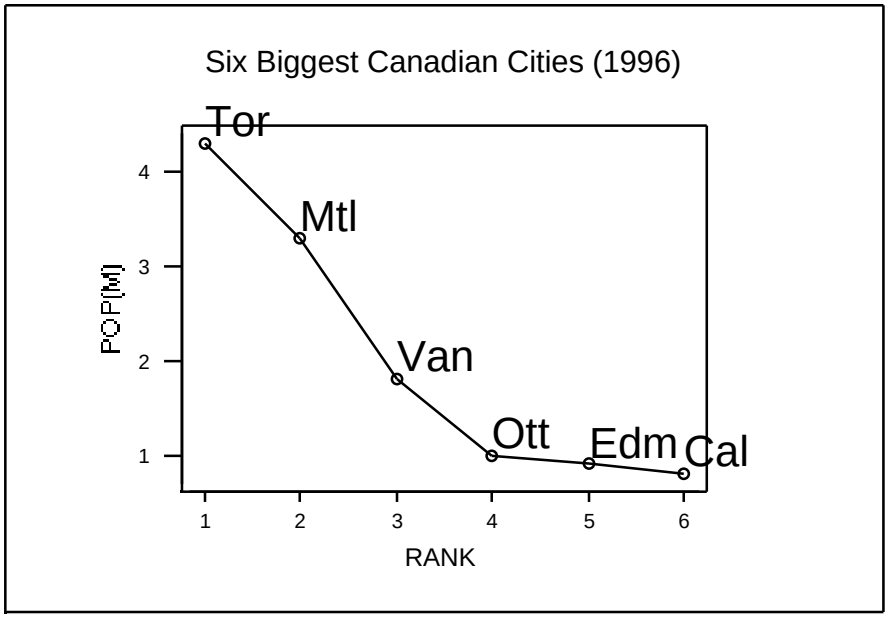
Want more time for Midterm or shorter test *
    More clear questions on midterm (#3) *
    More about Midterm 2
Liked class on randomized response technique and interactive lectures *
    Slower pace, more interaction, more overhead use * * *
Nice examples *
Assignment 4 worth too much for one question
Summary that compares and contrasts all methods in course
Pics of daughter's graduation!
Class average so far ?
More study materials
Bring Calculator and do not estimate

Clarification of Confusing Items:

1. Zipf's Law and logarithms * * * * * * * * * * * * * * *

Zipf's Law is a curiosity, not a key concept like the square root law.
But interesting that sizes of things, when ranked from biggest to smallest, decline in jumps that are approx predictable no matter what the context!

E.G.

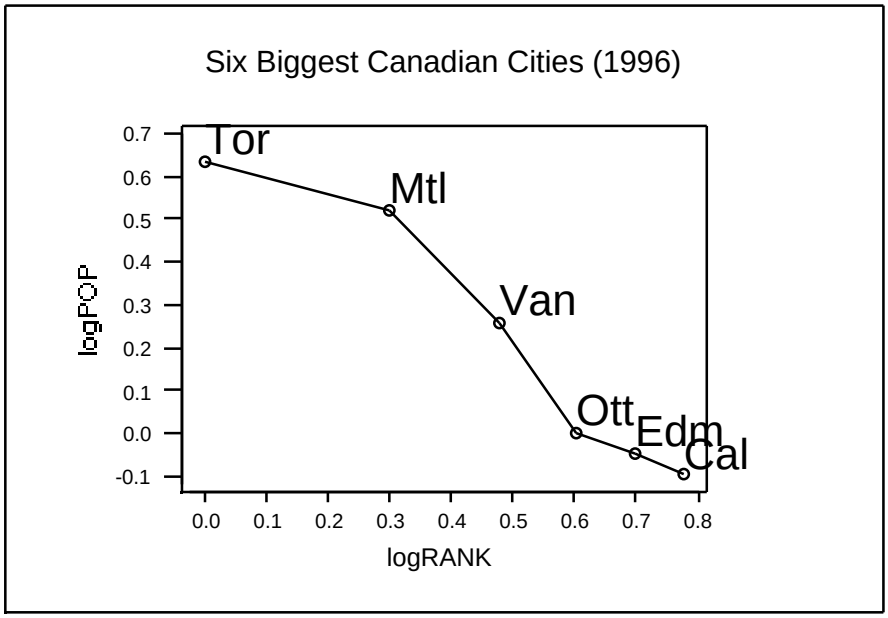


Base 10 logs … Population in millions

RANK		Population (Millions, 1996)
1	Toronto	4.3
2	Montreal	3.3
3	Vancouver	1.8
4	Ottawa	1.0
5	Edmonton	0.9
6	Calgary	0.8

Log 1 = 0.00	Toronto	Log Pop = 0.63
Log 2 = 0.30	Montreal	Log Pop = 0.51
Log 3 = 0.48	Vancouver	Log Pop = 0.26
Log 4 = 0.60	Ottawa	Log Pop = 0.00
Log 5 = 0.70	Edmonton	Log Pop = -0.045

Log 6 = 0.78 Calgary Log Pop = -0.097



Definition: the base 10 logarithm of a number is the power to which 10 must be raised to equal the number. For example $10^3=1000$ so $\log_{10}(1000)=3$

2. Averages and Proportions – connections, mean & SD * * * * * * **

Theory about variability of sample averages (i.e. square root law) applies to sample proportions.
Why? Because a sample proportion IS a sample average. Just code the things you want the proportion of as 1 and everything else as 0. Proportion of 1s = avg of 0s and 1s.
See notes Oct 23 for details.

3. Randomized Response Technique * *(neg ests) *(calc) * *

81 students toss fair coin,
estimate that 40.5 get heads, and 40.5 get tails. (81/2=40.5)
So approx 40.5 answer Yes to question about coin outcome, and 40.5 answer question about marijuana use. We counted 46 Yes answers. So approx 46-40.5 = 5.5 were answering Yes to the sensitive question (about marijuana). But that was 5.5 out of approx 40.5, or 100 x 5.5/40.5 = approx 14%. So we ESTIMATE that the proportion of the class using marijuana weekly is about 14%.

Q: What if we had less than 40.5 Yes answers? Instead of reporting a negative estimate for the proportion, just use 0.

4. Square root law – which n? * * * *

SD of avg = SD of things averaged / $\sqrt{number-of-things-averaged}$

5. Normal Model and Calculations * * *

For a normal distribution,
Mean ± 1SD is typical range containing 68%,
Mean ± 2SD is 95%,
Mean ± 3SD is 100% (almost)
Remember, averages tend to have Normal Distributions.

Demo: Use deck of cards: shuffle and deal 5 to self. Record and pass on to next student. Use this coding to compute average: 2-10=2-10, J Q K A = 15 so the distribution of codes in the deck is

Code	Freq
2	4
3	4
4	4
5	4
6	4
7	4
8	4
9	4
10	4
15	16
---	----
all	52

Certainly not normal!

However the averages of 5 cards will be approx normal.

What about Mean and SD?

Can compute mean = 8.0 SD=4.7. So SD of average of 5 is 4.7/$\sqrt{5}$ or about 2.1. Compare with class result. (For a normal distribtion, 68% within 2.1 of 8, that is in the range 5.9 to 10.1?)

(Data from class is next page)

Here is what we actually got in class (only a few students had time to record 5-card hands so the number of hands is smaller than ideal for this demo.)

Rounded Mean Score	Frequency
4	2
5	2
6	4
7	6
8	5
9	4
10	6
11	4
12	3
13	3

Mean = 8.7
SD = 2.5
Percent within 1 SD of Mean (based on sample, range is 6.2 to 11.2) is 25/39 or 64%
Percent within 1 SD of Mean (based on population, range is 5.9 to 10.1) is 25/39 or 64%
Normal distribution would give 68%

Binomial Model * * *
Sampling * * *
- Focus of sampling articles (Jury, Train Accounts) * * *
- Sampling with and without replacement

Correlation * *
Choice of study designs *
Solar System article *