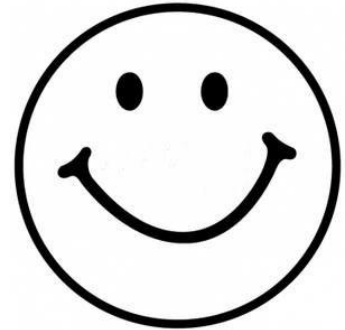


Happyville

Kevin S. Robinson, PhD
krobinson@millersville.edu

Department of Mathematics
Millersville University of Pennsylvania
Millersville, PA 17551

www.millersville.edu/~krobinson/happyville



Happyville

- a handout of a community map of households of various household sizes
- Source: **Beginnings of Interactive Statistics -- Aliaga and Gunderson**
- New? Innovative? In the spirit of ... *Activity-Based Statistics*
 - ✓ William Cochran's Box of Rocks
 - ✓ Random Rectangles, Rectangularity
 - ✓ Cents and the Central Limit Theorem
 - ✓ The JellyBlubber Colony

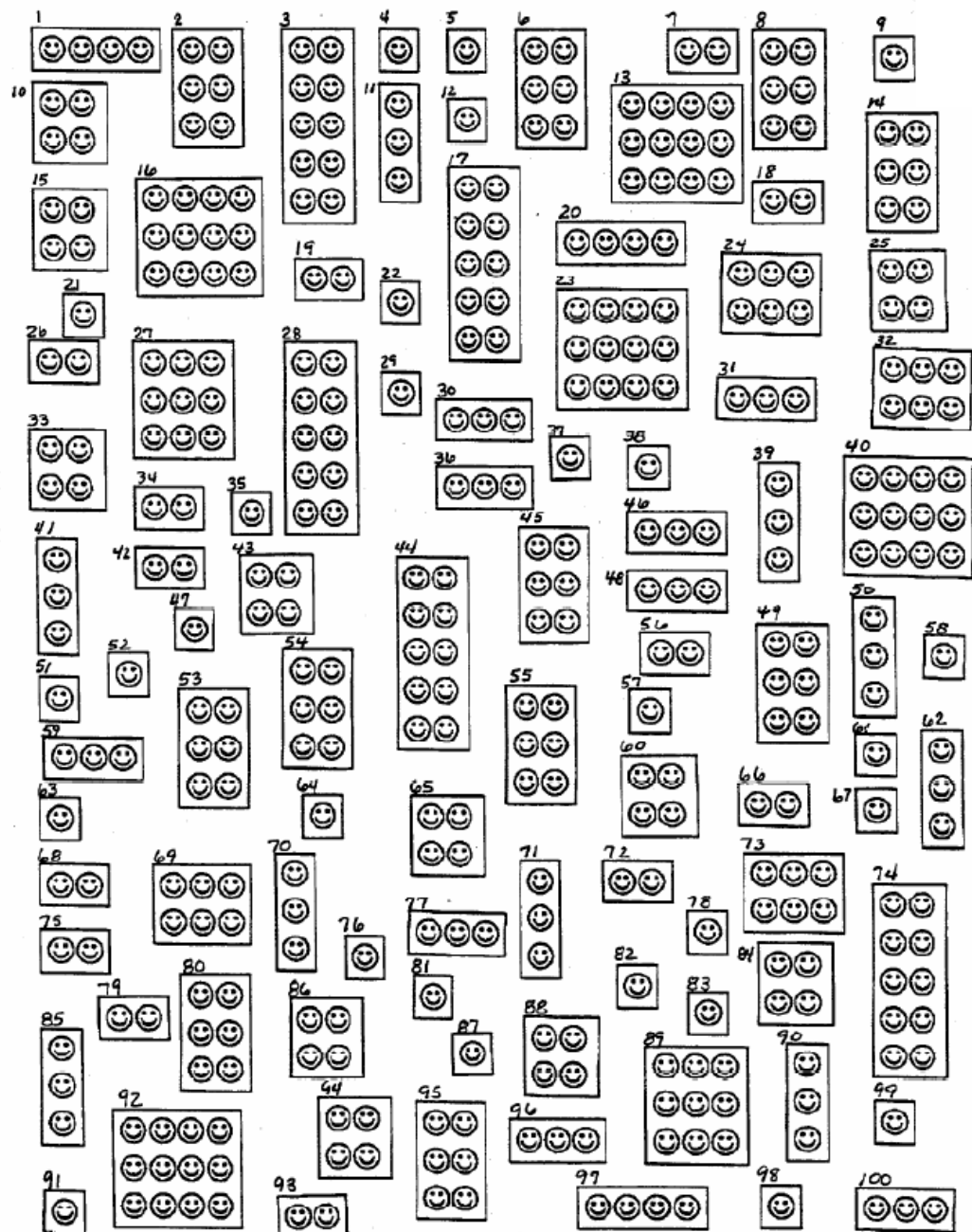
Motivation For Sharing ...

- **Colleague –** *I found myself jealous that I did not have such a resource when I teach similar material*

- **Instructor –** **Enjoyment/Effective, A Favorite, Reference Point/Common Thread Throughout The Course, Not A Lecture & Listen Format = GAISE, ...**

- **Student –** **Engagement, Student Interest, Improve Students' Attention, Motivation & Understanding, Reference Point/Common Thread Throughout The Course, Not A Lecture & Listen Format = GAISE, ...**

Community Map of Households



Formulate Questions That Can Be Addressed With Data:

What is the “typical” household size?

What is the max household size?

What is the “diversity” of household size?

What is the percentage above a household size of 2?

...

Population Information {Census} // Parameters

Versus

Sample Information {Sample} // Statistics

Uncertainty // Risk Of Using Sample Information

Data Production:

Happyville lends itself to:

Nonprobability Sampling:

**Judgment Sampling,
Convenience Sampling**

Probability Sampling:

**Simple Random Sampling,
Stratified Random Sampling (River),
Systematic Sampling,
Cluster Sampling (Wards)**

Examples:

Select 10 households that look to you to be "representative" of the town of Happyville. List the house numbers of the 10 you selected, along with the number of residents. Then calculate the mean number of residents per household for your sample.

Using the 10-sided die at the front of the class – generate a number at random from 1-10. Then sample every 10th household after that. For example, your sample could be H-holds # 3,13,23,33,43, 93.

...

Organize & Display Data

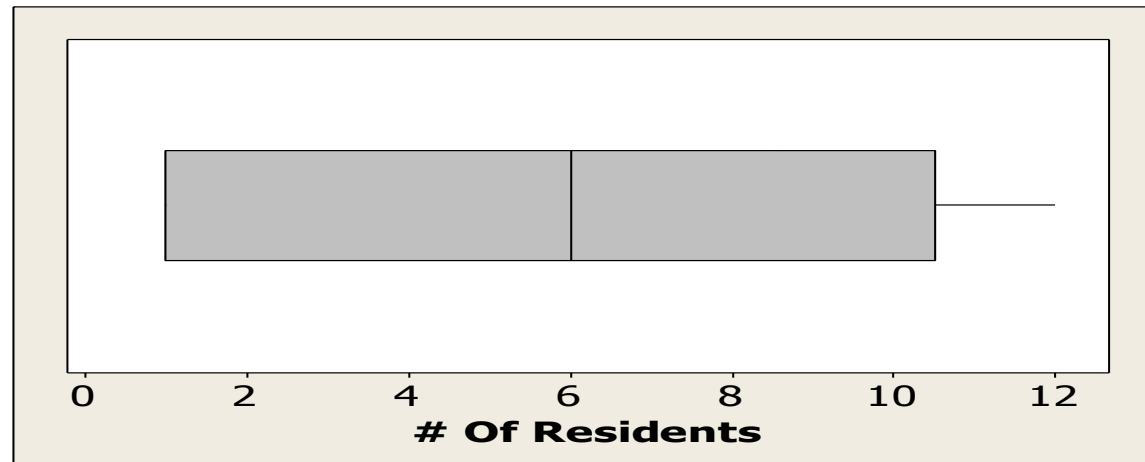
Make A Picture, Make A Picture, Make A Picture, ...

Quantitative Data:

**Dot-Plot, Stem & Leaf, Histogram,
Boxplot, Normal Probability Plot, ...**

Qualitative Data:

Pie Chart, Bar Graph, ...



Numerical Summary Statistics

Quantitative Data:

5-# Summary, Mean, Stdev, ...

Qualitative Data:

Frequency Tables, ...

Variable	N	Mean	SE Mean	StDev	Minimum	Q1	Median	Q3	Maximum
# Of Residents	10	6.10	1.42	4.48	1.00	1.00	6.00	10.50	12.00

Tally for Discrete Variables: Residents > 2

Residents

> 2	Count	Percent
N	4	40.00
Y	6	60.00
N=	10	

Sampling Variability

Using the last 4 digits of your phone number, take a sample of 2 households. Determine the sample maximum and the sample mean of your sample.

Key Question: What could/would/can/... happen?

**What are the possible values for the max? ... the mean?
Will your results match your neighbor? ... the truth?**

Key Idea: Sample Statistics have possible values and distributions.

Sample Statistics: random variables w/ associated Sampling Distributions!

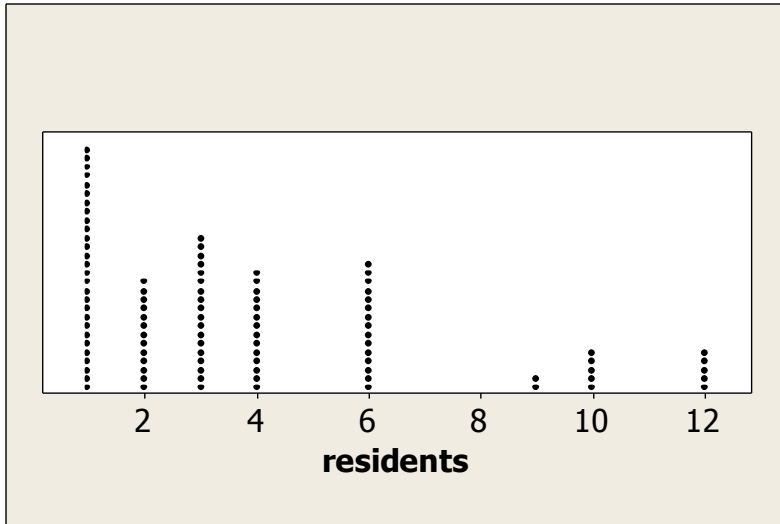
Example: 1347 – household 13 & household 47 -- # of residents = 12,1

sample max = 12

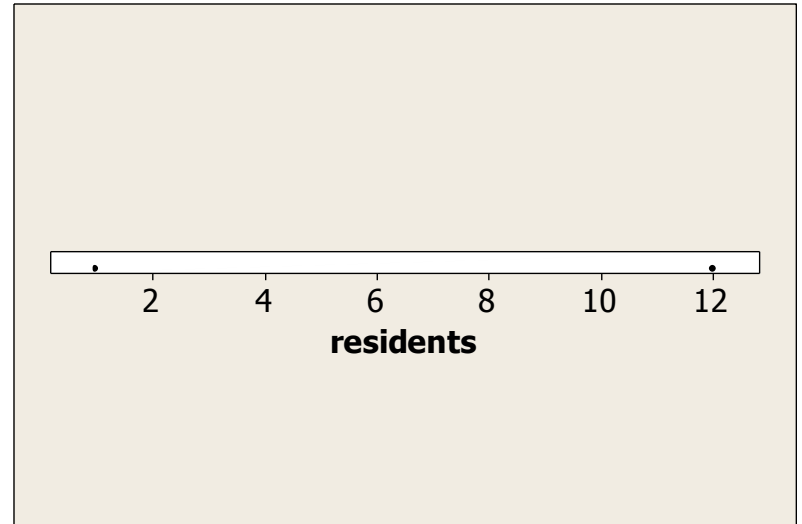
sample mean = 6.5

3 Different Distributions?!

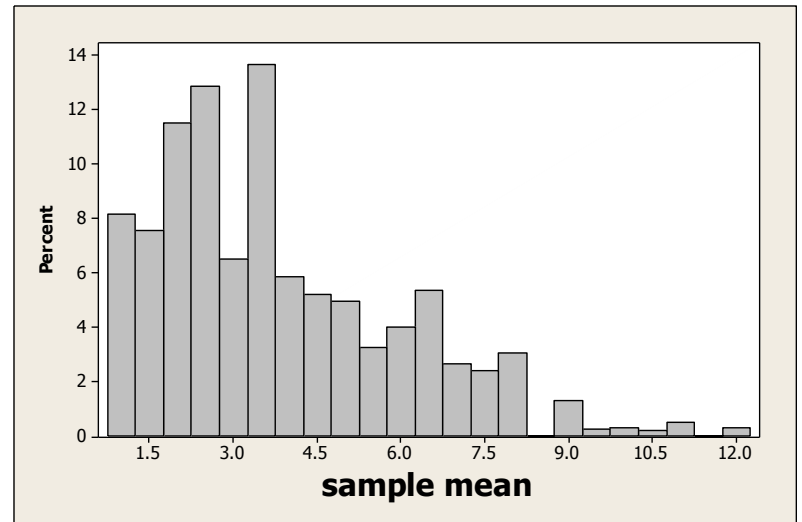
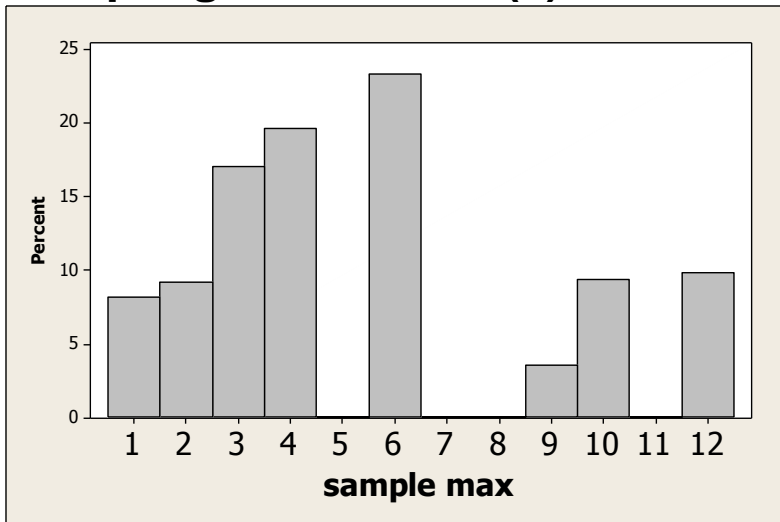
Population Distribution



Sample Distribution



Sampling Distribution(s)



Discovery Questions

What happens as n increases 5, 10, 20, ...?

... to the Population Distribution?

{Nothing! – “Happyville Is What It Is ...”}

... to the Sample Distribution?

{Better Reflection Of Population ...!}

... to the Sampling Distribution(s)?

{Confusion?, Requires Prompting, ...}

What happens to the possible values of sample max? {Nothing?! ...}

What has to happen in order to get a sample max of 1? ... of 12?

What happens to the distribution of sample max? {Moves To Right ...}

What happens to the possible values of sample mean?

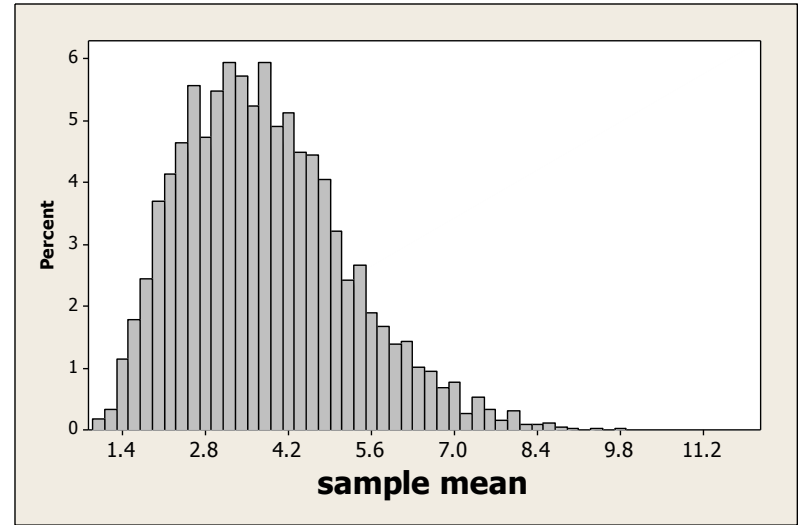
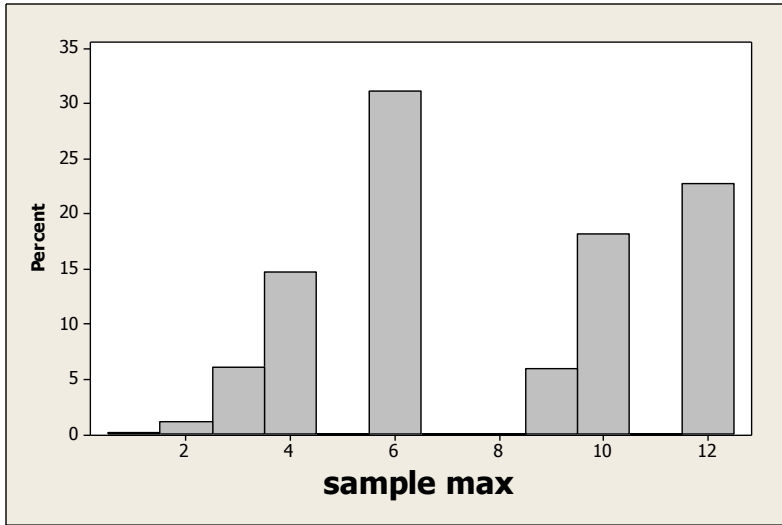
{Finer, Continuum, ...}

What has to happen in order to get a sample mean of 1? ... of 12?

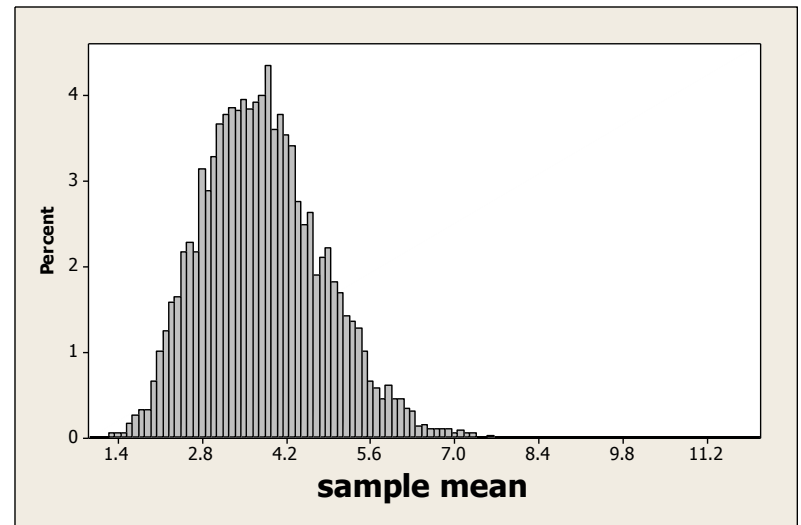
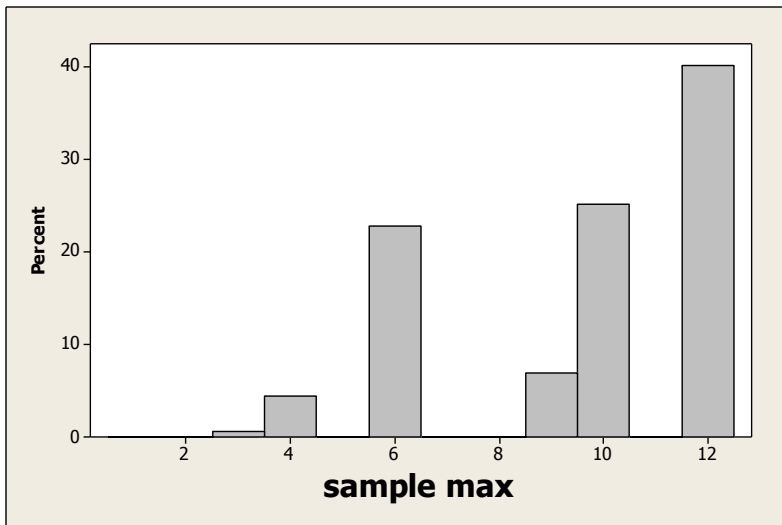
What happens to the distribution of sample mean?

{Moves To “Center”, Law of Large #'s, ...}

n = 5



n = 10



Discovered Concepts & Understanding ...

- distribution of sample max tends to a non-normal sampling distribution ...
- distribution of sample mean ... Location? Spread? Shape?
 - 1) centered at the population mean,
 - 2) less variability than the original population, **
 - 3) the variability decreases as n increases, **
 - 4) tends to a normal distribution- Central Limit Theorem.

** The Most Dangerous Equation

*Ignorance of how sample size affects statistical variation
has created havoc for nearly a millennium ...*

H Wainer - American Scientist, 2007

Assessment // Check of Understanding

Consider a balanced 6-sided die ... the die is to be studied by taking a random sample of size 10 and calculating the sample mean.

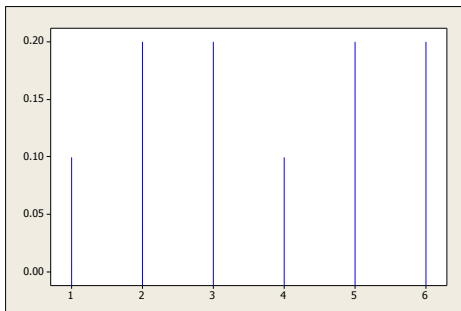
Match the graph (from below) with the appropriate distribution.

Population Distribution For Balanced 6-Sided Die _____

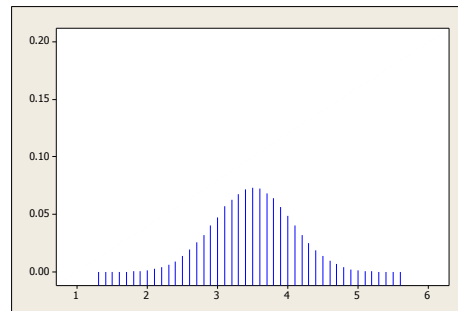
Sample Distribution For 10 Tosses Of The Die _____

Sampling Distribution Of Sample Mean For Random Samples Of Size 10 _____

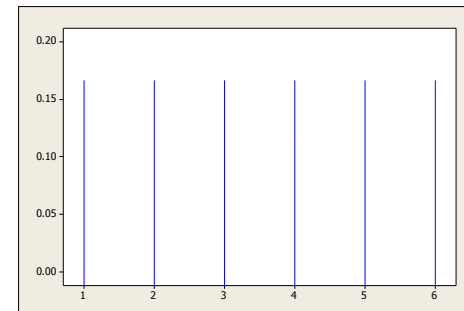
A



B



C



Assessment // Check of Understanding

Use of http://onlinestatbook.com/stat_sim/sampling_dist/index.html

And sampling distribution quiz –

see <http://www.millersville.edu/~krobinson/happyville>

A waiter believes the distribution of his tips has a model that is right tailed, with a mean = \$9.60 and a stdev = \$20.

Can you determine the probability that a given table will tip him at least \$10.

Consider instead the 25 tables he will be assigned during his next shift,

Describe the sampling distribution of the mean tip resulting from these 25 tables.

Would it be reasonable for the 25 tables to result in a mean tip of at least \$10. Explain.

Estimation

Happyville – Parameter of Interest = Population Mean Household Size

Problem of Point Estimation

Can NOT guarantee statistic = parameter!

Probability [statistic = parameter] is very very low!!

Interval Estimation

Range of plausible values for the parameter of interest ...

Interpretation of Confidence Level? Interpretation of Interval?

Classroom Activity

Provide each student with sample of 10 households from happyville.

**Each student constructs a 90% confidence interval
for the mean household size of happyville.**

Confirm each student's interval is correct.

Questions for Student to ponder ...

Does your interval match your neighbor's?

Do you think your interval contains the mean household size?

Does your interval contain the mean household size?

Interpret your interval – {Misconceptions?!, ...}

What does 90% confidence mean? –

{Reveal Truth – Students Who Missed Stand}

{Confidence In Method ...}

Testing, Hypotheses, Type I & II Error

US Census – mean household size = 2.59 – Does happyville differ?

$$H_1: \mu \neq 2.59$$

US Census – second most common type of household (26 percent)
consisted of people living alone – higher in happyville?

$$H_1: p > 0.26$$

Using sample info introduces the reality of Type I & II Error ...

Type I & II Error – discussion in context of happyville ...

Using sample info to reach conclusion – nature of evidence, p-value, ...

Future Endeavors:

Mathematical Statistics Course

Recognition Of The Finite Population

With/Without – Binomial/Hypergeometric

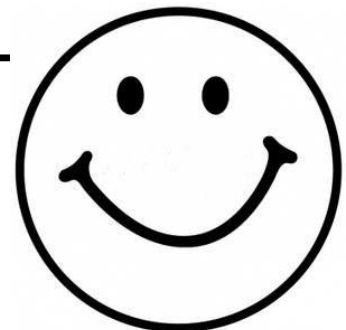
**Properties Of The Mean: Finite Population
{Finite Population Correction Factor}**

Conclusion:

Statistics is fun!

**Learning to think clearly with data will open
your eyes to seeing the world more clearly ...**

**Happyville is a useful and fun resource –
promoting conceptual understanding,
encouraging active participation,
& generating student interest.**



Enjoy!