Fourth CBMS Forum
Teaching Teachers in the Era of the Common Core
B2: Teacher Preparation in Statistics
2:30 –4:00 Monday, October 3, 2011

Christine Franklin, University of Georgia
chris@stat.uga.edu
Jerry Moreno, John Carroll University
moreno@jcu.edu

AGENDA

- The CCSS in Statistics and the ASA GAISE Report
- A statistics example in the GAISE format
- Resources
- Recommendations for the pre-service course and professional development workshops
- Q&A

Getting to know you…

- Do you teach pre-service mathematics content courses? Methods courses?
  - Prek-2; 3-5; 6-8; HS
- How knowledgeable are you of CCSS in Statistics and Probability? Raise fingers.
  - 1: Haven't a clue.
  - 2:
  - 3:
  - 4:
  - 5: Completely familiar with.

Very brief history of Statistics in the School Curriculum

Before 1989: non-existent

1989: NCTM 14 content Standards

#10 Statistics; #11 Probability

2000: NCTM PSSM 5 content Standards

Data analysis and Probability – excellent

Many states adopted PSSM

Quality of assessment – sporadic

2010: CCSS

Very Brief Background Info on the Common Core State Standards (CCSS or CC)

www.corestandards.org

- CCSS released 6/2/10:
  - National Governors Association Center for Best Practices (NGA Center)
  - Council of Chief State School Officers (CCSSO)
- 44 states have adopted the CC Standards
  - Not AK, MN, MT, NE, TX, VA
- Two Assessment Consortia (2014-15)
  - PARCC (Partnership for Assessment of Readiness for College and Careers)
  - SBAC (SMARTER Balanced Assessment Consortium)

Very Brief Background Info on the Common Core State Standards (CCSS or CC)…cont.

8 Mathematical Practices Standards

- Describe "habits of mind"
- Foster reasoning and sense-making in mathematics
  1. Make sense of problems and persevere in solving them.
  2. Reason abstractly and quantitatively.
  3. Construct viable arguments and critique the reasoning of others.
  4. Model with mathematics.
  5. Use appropriate tools strategically.
  6. Attend to precision.
  7. Look for and make use of structure.
  8. Look for and express regularity in repeated reasoning.

Very Brief Background Info on the Common Core State Standards (CCSS or CC)…cont.

Re: Statistics and Probability

- K-5 Domain: Measurement and Data
- 6-8 Domain: Statistics and Probability
- HS Conceptual Category: Statistics and Probability

Let's look at the CC content standards that students are to master in order to get an idea of what needs to be in your university course(s) for pre-service teachers and/or in your professional development workshops for in-service teachers.

What will the CC Student have mastered in Statistics and Probability through CC k-12?

www.corestandards.org

- Grades K-5 Domain: Measurement and Data
  - Grade K: Classify objects into given categories; count the number of objects in each category; sort by count.
  - Grade 1: Organize, represent, and interpret data with up to three categories.
  - Grade 2: Make line plot for measurement data; picture and bar graphs for up to four categories.
  - Grade 3: Make bar graph in which each square represents k subjects; line plot for halves, quarters.
  - Grade 4: Make line plot for fractions; interpret largest minus smallest.
  - Grade 5: Redistribute total amount into k equal amounts.

ASIDE

- Let's look at that last Standard in Grade 5.
  I wrote it as: Redistribute total amount into k equal amounts.

- Here is the actual wording:
  - Domain: 5.MD
  - Cluster: Represent and interpret data.
  - Standard: Make a line plot to display a data set of measurements in fractions of a unit (1/2, 1/4, 1/8). Use operations on fractions for this grade to solve problems involving information presented in line plots. *For example, given different measurements of liquid in identical beakers, find the amount of liquid each beaker would contain if the total amount in all the beakers were redistributed equally."*

So, for all intents and purposes, our CC students know very little about statistics through grades k-5. Whatever is in the Standards is there more or less to motivate a mathematics concept.

What will the CC Student have mastered in Statistics and Probability through CC k-12? cont.

Grade 6 Domain: Statistics and Probability

Cluster: Develop understanding of statistical variability.

1. Recognize a statistical question as one that anticipates variability in the data related to the question and accounts for it in the answers.
2. Understand that a set of data collected to answer a statistical question has a distribution which can be described by its center, spread, and overall shape.
3. Recognize that a measure of center for a numerical data set summarizes all of its values with a single number, while a measure of variation describes how its values vary with a single number.

What will the CC Student have mastered in Statistics and Probability through CC k-12? cont.

Grade 6 Domain: Statistics and Probability

Cluster: Summarize and describe distributions.

4. Display numerical data in plots on a number including dot plots, histograms, and box plots.
5. Summarize numerical data sets in relation to their context.
   - Center (median, mean)
   - Variability (IQR, MAD)

Time out: GAISE k-12

www.amstat.org/education/gaise/index.cfm

The statistical process is a problem solving process consisting of four components:

1. Formulate a question that can answered by data.
2. Design and implement a plan to collect data.
3. Analyze the data by graphical and numerical methods.
4. Interpret the analysis in the context of the original question.

What will the CC Student have mastered in Statistics and Probability through CC k-12? cont.

Grade 7 Domain: Statistics and Probability

Cluster: Use random sampling to draw inferences about a population.

1. Understand that statistics can be used to gain information about a population by examining a representative sample from it.
2. Use data from a random sample to draw inferences about a population with an unknown characteristic of interest. Generate multiple samples (or simulated samples) of the same size to gauge the variation in estimates or predictions.

What will the CC Student have mastered in Statistics and Probability through CC k-12? cont.

Grade 7 Domain: Statistics and Probability

Cluster: Draw informal inferences about two populations.

3. Informally assess the degree of visual overlap of two numerical data distributions with similar variabilities, measuring the difference between the centers by expressing it as a measure of variability.
4. Use measures of center and measures of variability for numerical data from random samples to draw informal comparative inferences about two populations.

What will the CC Student have mastered in Statistics and Probability through CC k-12? cont.

Grade 7 Domain: Statistics and Probability cont.

Cluster: Investigate chance processes and develop, use, and evaluate probability models.

5. Understand that the probability of a chance event is a number between 0 and 1 that expresses the likelihood of the event occurring.
6. Approximate the probability of a chance event by collecting data on the chance process that produces it and observing its long-run relative frequency, and predict the approximate relative frequency given the probability.
7. Develop a probability model and use it to find probabilities of events.
   a. Develop a uniform probability model by assigning equal probability to all outcomes and use the model to determine probabilities of events.
   b. Develop a probability model by observing frequencies in data generated from a chance process.

What will the CC Student have mastered in Statistics and Probability through CC k-12? cont.

Grade 7 Domain: Statistics and Probability cont.

Cluster: Investigate chance processes and develop, use, and evaluate probability models. cont.

8. Find probabilities of compound events using organized lists, tables, tree diagrams, and simulation.

   a. Understand that, just as with simple events, the probability of a compound event is the fraction of outcomes in the sample space for which the compound event occurs.

   b. Represent sample spaces for compound events using methods such as organized lists, tables and tree diagrams.

   c. Design and use a simulation to generate frequencies for compound events.

What will the CC Student have mastered in Statistics and Probability through CC k-12? cont.

Grade 8 Domain: Statistics and Probability

Cluster: Investigate patterns of association in bivariate data.

1. Construct and interpret scatter plots for bivariate measurement data to investigate patterns of association between two quantities. Describe patterns such as clustering, outliers, positive or negative association, linear association, and nonlinear association.
2. Know that straight lines are widely used to model relationships between two quantitative variables. For scatter plots that suggest a linear association, informally fit a straight line, and informally assess the model fit by judging the closeness of the data points to the line.
3. Use the equation of a linear model to solve problems in the context of bivariate measurement data, interpreting the slope and intercept.
4. Understand that patterns of association can also be seen in bivariate categorical data by displaying frequencies and relative frequencies in a two-way table. Construct and interpret.

What will the CC Student have mastered in Statistics and Probability through CC k-12? cont.

Grade HS Conceptual Category : Statistics and Probability

Domain: Interpreting Categorical and Quantitative data

- Cluster: Summarize, represent, and interpret data on a single count or measurement variable.
- Cluster: Summarize, represent, and interpret data two categorical and quantitative variables.
- Cluster: Interpret linear models.

What will the CC Student have mastered in Statistics and Probability through CC k-12? cont.

Grade HS Conceptual Category : Statistics and Probability cont.

Domain: Making Inferences and Justifying Conclusions

- Cluster: Understand and evaluate random processes underlying statistical experiments.
- Cluster: Make inferences and justify conclusion from sample surveys, experiments, and observational studies.

What will the CC Student have mastered in Statistics and Probability through CC k-12? cont.

Grade HS Conceptual Category : Statistics and Probability

Cluster: Make inferences and justify conclusion from sample surveys, experiments, and observational studies.

- 3. Recognize the purposes of and differences among sample surveys, experiments, and observational studies; explain how randomization relates to each.
- 4. Use data from a sample survey to estimate a population mean or proportion; develop a margin of error through the use of simulation models for random sampling.
- 5. Use data from a randomized experiment to compare two treatments; use simulations to decide if differences between parameters are significant.
- 6. Evaluate reports based on data.

What will the CC Student have mastered in Statistics and Probability through CC k-12? cont.

Grade HS: Statistics and Probability Conceptual Category cont.

Domain: Conditional Probability and the Rules of Probability

- Cluster: Understand independence and conditional probability and use them to interpret data.
- Cluster: Use the rules of probability to compute probabilities of compound events in a uniform probability model.

CONNECTIONS TO FUNCTIONS and MODELING:

Functions may be used to describe data; if the data suggest a linear relationship, the relationship can be modeled with a regression line, and its strength and direction can be expressed through a correlation coefficient.

Summary of what the CC student will have mastered – both conceptually (interpretation) and by formula (with technology):

- Data analysis/Statistics:
  - The understanding of statistical variability.
  - The GAISE statistical process four-step model (but maybe not by name).
  - Graphs (pie, bar; dot, hist, box; scatter, time).
  - Characterizing numerical distributions:
    - Measures of center (mode, median, mean – as "fair share" and balance).
    - Measures of spread (range, IQR, MAD, standard deviation).
    - Shape (symmetric, skewed, outliers).
  - Correlation (not causal), coefficient r (with technology).
  - Regression – linear (median-median?, least squares) with residuals;

    quadratic, exponential fitting to data.
  - Inferences from sample surveys, observational studies, experiments.
  - Use of simulation for inferential or estimation purposes in one mean, one proportion, two means.

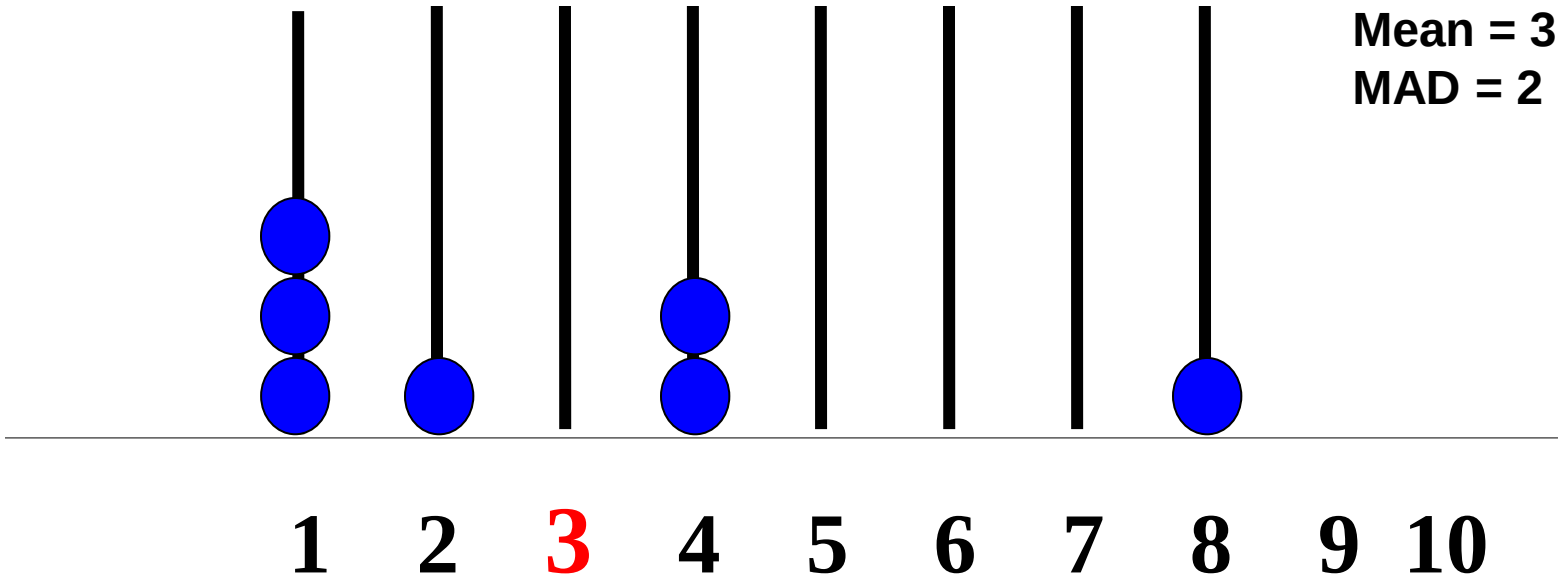
- Probability
  - Normal distribution calculation of probabilities.
  - Sample space; simple and compound events. Addition rule.
  - Independent events; conditional probability; extensive use of two-way tables.

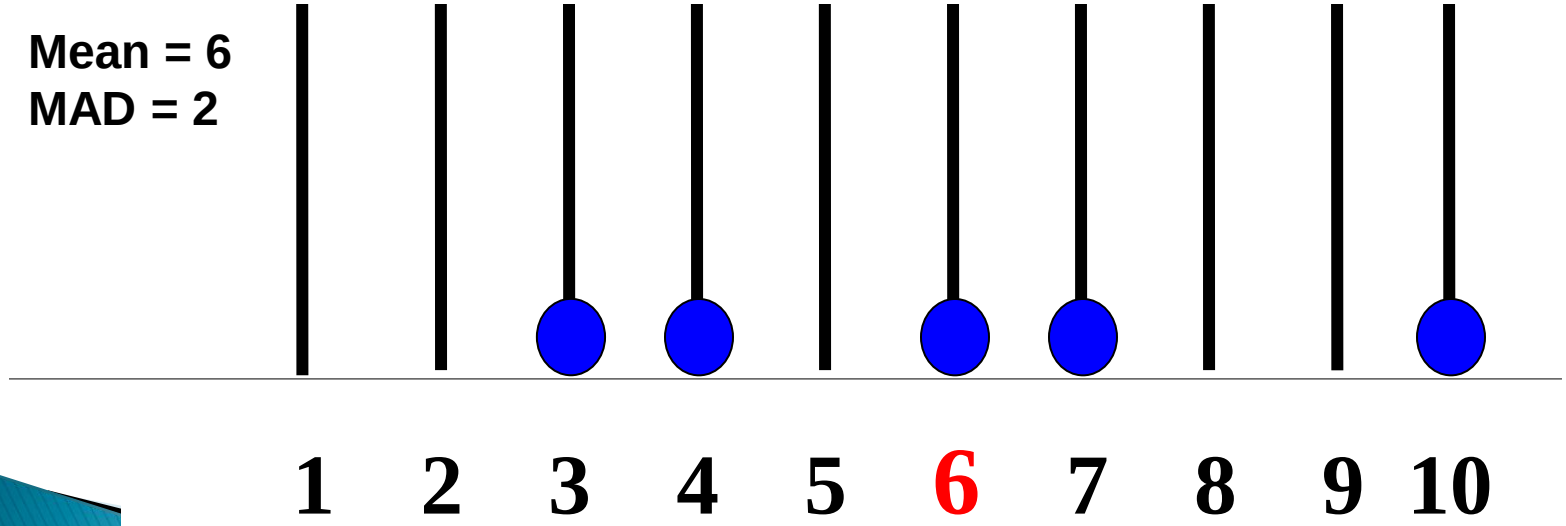
- *Aside: There is more probability but not for ALL students. The topics include the multiplication rule; permutations and combinations; random variable; expected value; theoretical probability distributions (e.g., two rolls of a fair die); probability distribution for empirical probabilities; probability distribution with weighted outcomes (e.g., payoffs); analysis of decisions and strategies using probability concepts (e.g., "pulling a hockey goalie at the end of a game.")*

	STATS 101 typical material	STATS 101 Not typical material
CC mastered	Graphs: (pie,bar; dot,hist,box; scatter,time). Measures: center(mmm); spread(range,IQR,s). Correlation: r. Regression (least squares); residual analysis. Surveys, observational studies, experiments. Probability: sample space; simple and compound events. Independent events. Two-way table; conditional probability.	Measures: spread (MAD). Regression: model fits for quadratic, exponential. Inference: randomization tests. Mathematical Practices.
Not in CC	Graphs: stem. Correlation: confounding. Central Limit Theorem. Normal theory-based inference.	

7.SP.3
Informally assess the degree of visual overlap of two numerical data distributions with similar variabilities, measuring the difference between the centers by expressing it as a multiple of a measure of variability.

- Consider the following two data sets.
- Choose a measure of variability, typically either IQR or MAD. Note that the measure needs to have the same value for the two sets.
- Then determine how many MAD's, say,
- separate the two means, say.





Number of Pets

7.SP.3
Informally assess the degree of visual overlap of two numerical data distributions with similar variabilities, measuring the difference between the centers by expressing it as a multiple of a measure of variability.

The two data sets have means that differ by 6 – 3 = 3 pets. The MAD for each is 2 pets. So, the number of MADs that separate the means is 1.5 pets.

In higher level statistics (AP Stats), the MAD is typically replaced by the standard deviation "s" (or more precisely the standard error) and data sets are replaced by sampling distributions.

More resources

- Webinar Bill McCallum December 2010

http://educationnorthwest.org/event/1346

- Webinars on Teaching/Learning Statistics

http://www.causeweb.org/webinar/teaching/

Is Stats 101 Prepared for the CC Stats Prepared Student?

Jerry Moreno May 2011

- Website The Illustrative Mathematics Project

http://illustrativemathematics.org/

School curriculum series…

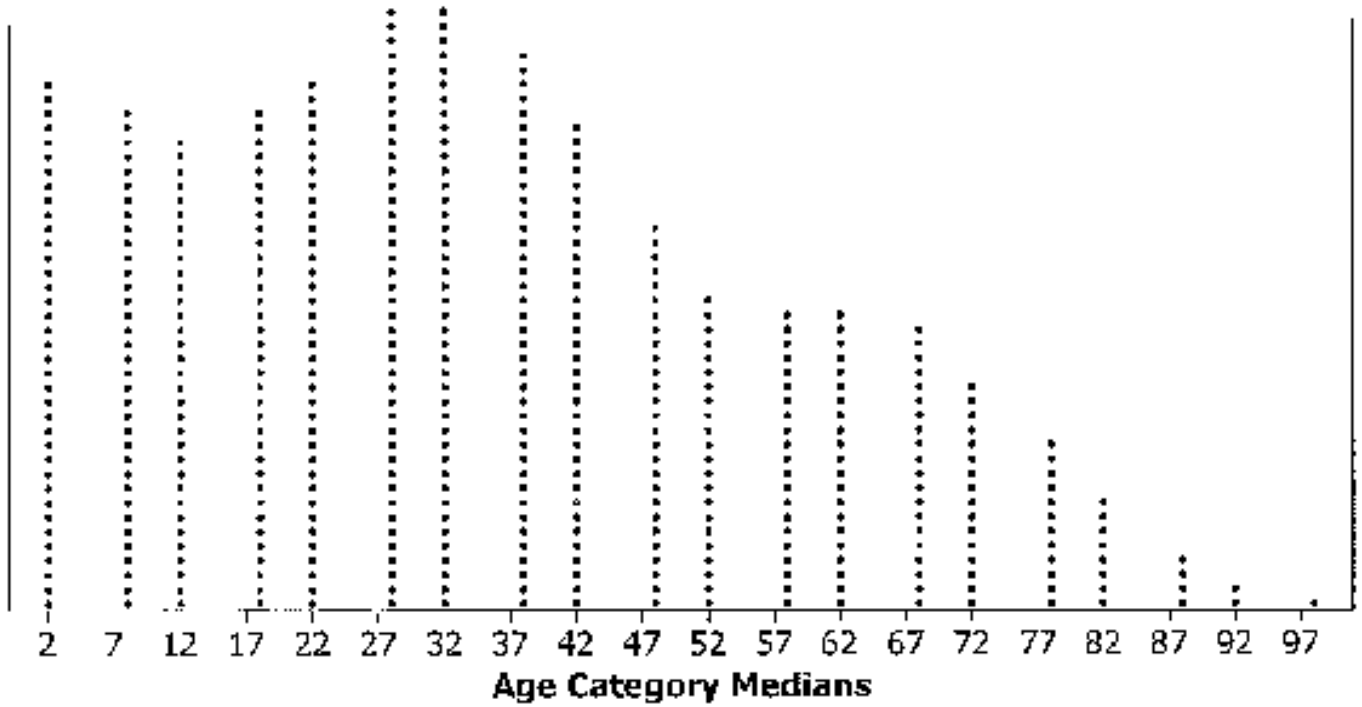
- Elementary…Investigations
  - Scott Foresman
- Middle…Connected Math
  - Pearson/Prentice Hall
- High School…Core Plus Math
  - http://www.wmich.edu/cpmp/

THE SHAPE of a POPULATION

Based on Henry's article in the NCTM publication Mathematics Teacher, *2004,* v97 n1, pp 58-66 entitled *People Count: Analyzing a Country's Future.*

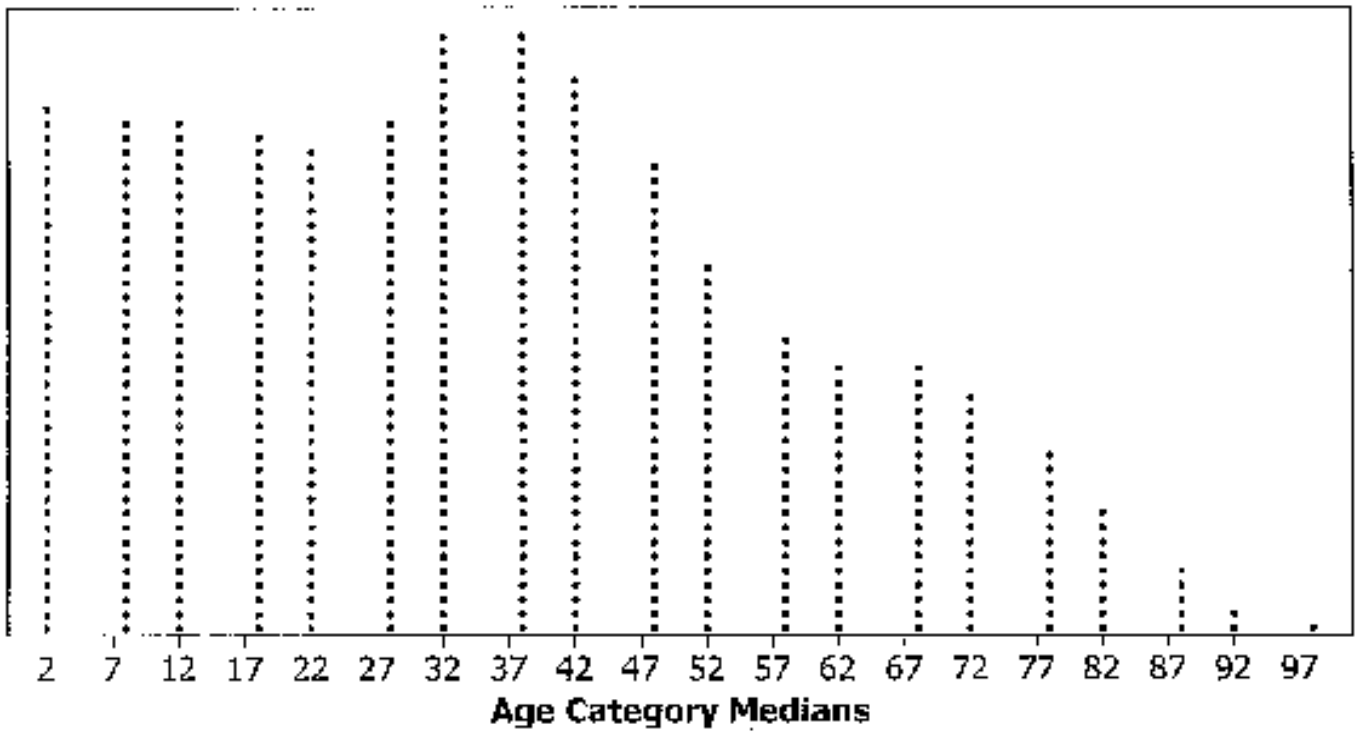
Ages	Median	1990	Ages	Median	1995
100+		36	100+		48
95-99	97	203	95-99	97	268
90-94	92	748	90-94	92	1017
85-89	87	2035	85-89	87	2352
80-84	82	3910	80-84	82	4478
75-79	77	6104	75-79	77	6700
70-74	72	7981	70-74	72	8831
65-69	67	10067	65-69	67	9926
60-64	62	10627	60-64	62	10046
55-59	57	10489	55-59	57	11086
50-54	52	11315	50-54	52	13642
45-49	47	13747	45-49	47	17458
40-44	42	17593	40-44	42	20259
35-39	37	19851	35-39	37	22296
30-34	32	21838	30-34	32	21825
25-29	27	21336	25-29	27	18905
20-24	22	19143	20-24	22	17982
15-19	17	17893	15-19	17	18203
10-14	12	17067	10-14	12	18853
5-9	7	18042	5-9	7	19096
0-4	2	18765	0-4	2	19532

The SHAPE of the 1990 Population



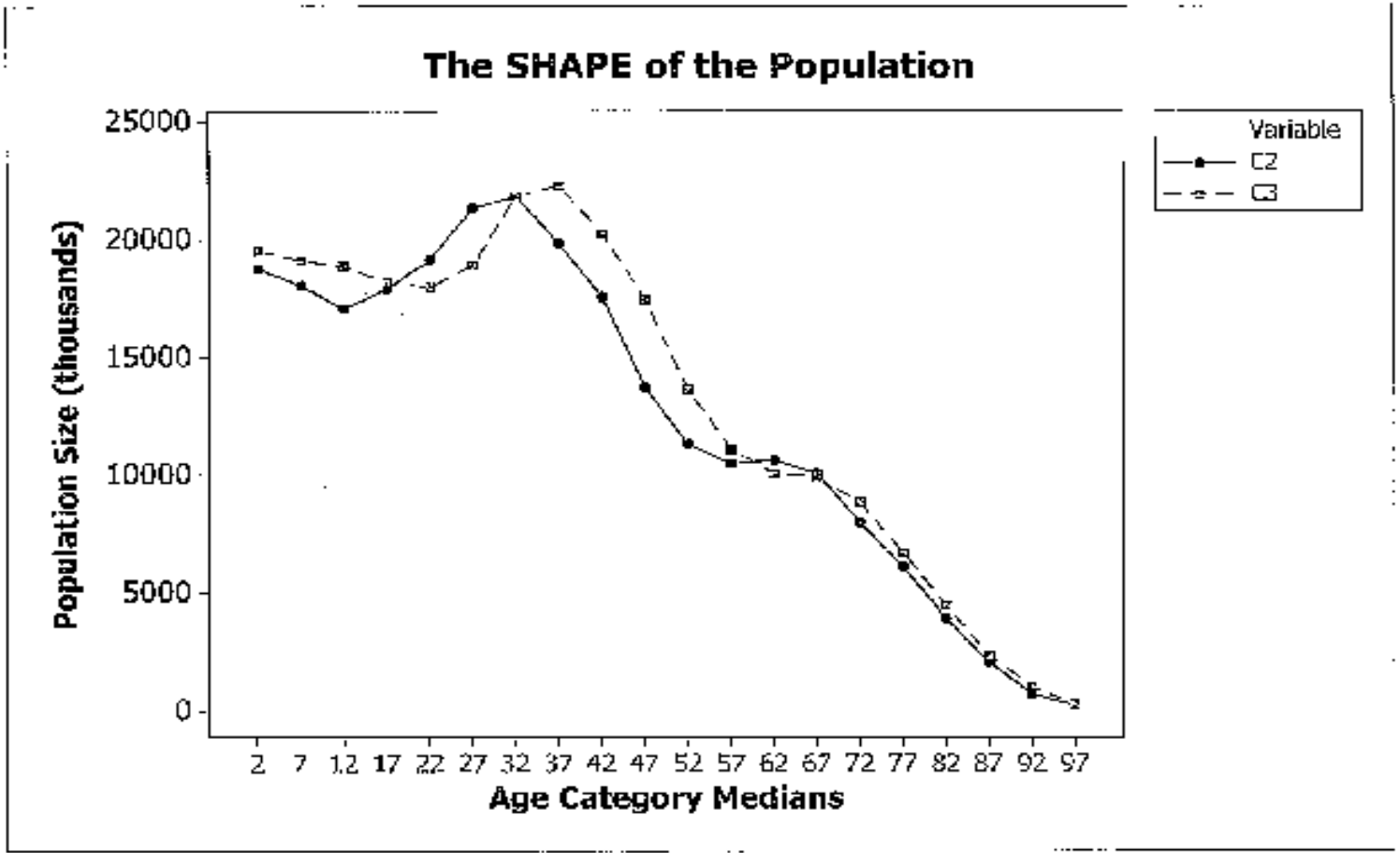
Each symbol represents up to 520 observations.

The SHAPE of the 1995 Population



Each symbol represents up to 531 observations.

IS THE POPULATION AGING?



A RECURSIVE PROCEDURE

0. Note that the people in age category x are the same people in age category x+1 in the succeeding time increment apart from those who died, emigrated, or immigrated.

1. To project population sizes for 2000, use the given data for 1990 and 1995. For each 5-year age interval (except 0-4 and over 100), calculate a "population factor" by taking population in 1995 in age category k+1 / population in 1990 in age category k.

Base Layer	Base Pop 1990	Next Layer	Next Pop 1995	Pop Factor Next/Base
100+	36			
95-99	203	100+	48	.2365
90-94	748	95-99	268	.3583
85-89	2035	90-94	1017	.4998
80-84	3910	85-89	2352	.6015
75-79	6104	80-84	4478	.7336
70-74	7981	75-79	6700	.8395
65-69	10067	70-74	8831	.8772
60-64	10627	65-69	9926	.9340
55-59	10489	60-64	10046	.9578
50-54	11315	55-59	11086	.9798
45-49	13747	50-54	13642	.9924
40-44	17593	45-49	17458	.9923
35-39	19851	40-44	20259	1.0206
30-34	21838	35-39	22296	1.0210
25-29	21336	30-34	21825	1.0229
20-24	19143	25-29	18905	.9876
15-19	17893	20-24	17982	1.0050
10-14	17067	15-19	18203	1.0666
5-9	18042	10-14	18853	18853/18042=1.0450
0-4	18765	5-9	19096	19096/18765=1.0176
		0-4	19532	

The POPULATION SIZES for 2000

For example, 2000 pop for 60-64 equals 1995 pop for 55-59 times the population factor, i.e., 11086(.9578) = 10618

1995 ages	1995	Pop Factor	2000 ages	Next Pop 2000 est.
100+			???	
95-99	268	.2365	100+	63
90-94	1017	.3583	95-99	364
85-89	2352	.4998	90-94	1176
80-84	4478	.6015	85-89	2694
75-79	6700	.7336	80-84	4915
70-74	8831	.8395	75-79	7414
65-69	9926	.8772	70-74	8707
60-64	10046	.9340	65-69	9383
55-59	11086	.9578	60-64	10618
50-54	13642	.9798	55-59	13366
45-59	17458	.9924	50-54	17325
40-44	20259	.9923	45-59	20103
35-39	22296	1.0206	40-44	22755
30-34	21825	1.0210	35-39	22283
25-29	18905	1.0229	30-34	19338
20-24	17982	.9876	25-29	17759
15-19	18203	1.0050	20-24	18294
10-14	18/853	1.0666	15-19	20109
5-9	19096	1.0450	10-14	19955
0-4	19532	1.0176	5-9	19876
			0-4	

What should we do with the 100+ age category?

For the moment, let's omit it from consideration since it is fairly small. In the future, we may have to create a new category or two if this age category becomes significantly large.

What should we do with the 0-4 age category for year 2000?

The 0-4 category in year 2000 obviously has no category in 1995 to project from. What to do?

Assuming a constant birth rate, the proportion of the 15-44 age category to the 0-4 age category should be about the same from 1995 to 2000.

Base Layer	Base Pop 1995	Pop Factor	Next Layer	Next Pop 2000 est.
100+	48			
95-99	268	.2365	100+	63
90-94	1017	.3583	95-99	364
85-89	2352	.4998	90-94	1176
80-84	4478	.6015	85-89	2694
75-79	6700	.7336	80-84	4915
70-74	8831	.8395	75-79	7414
65-69	9926	.8772	70-74	8707
60-64	10046	.9340	65-69	9383
55-59	11086	.9578	60-64	10618
50-54	13642	.9798	55-59	13366
45-49	17458	.9924	50-54	17325
40-44	20259	.9923	45-49	20103
35-39	22296	1.0206	40-44	22755
30-34	21825	1.0210	35-39	22283
25-29	18905	1.0229	30-34	19338
20-24	17982	.9876	25-29	17759
15-19	18203	1.0050	20-24	18294
10-14	18853	1.0666	15-19	20109
5-9	19096	1.0450	10-14	19955
0-4	19532	1.0176	5-9	19876
			0-4	

15-44
119470

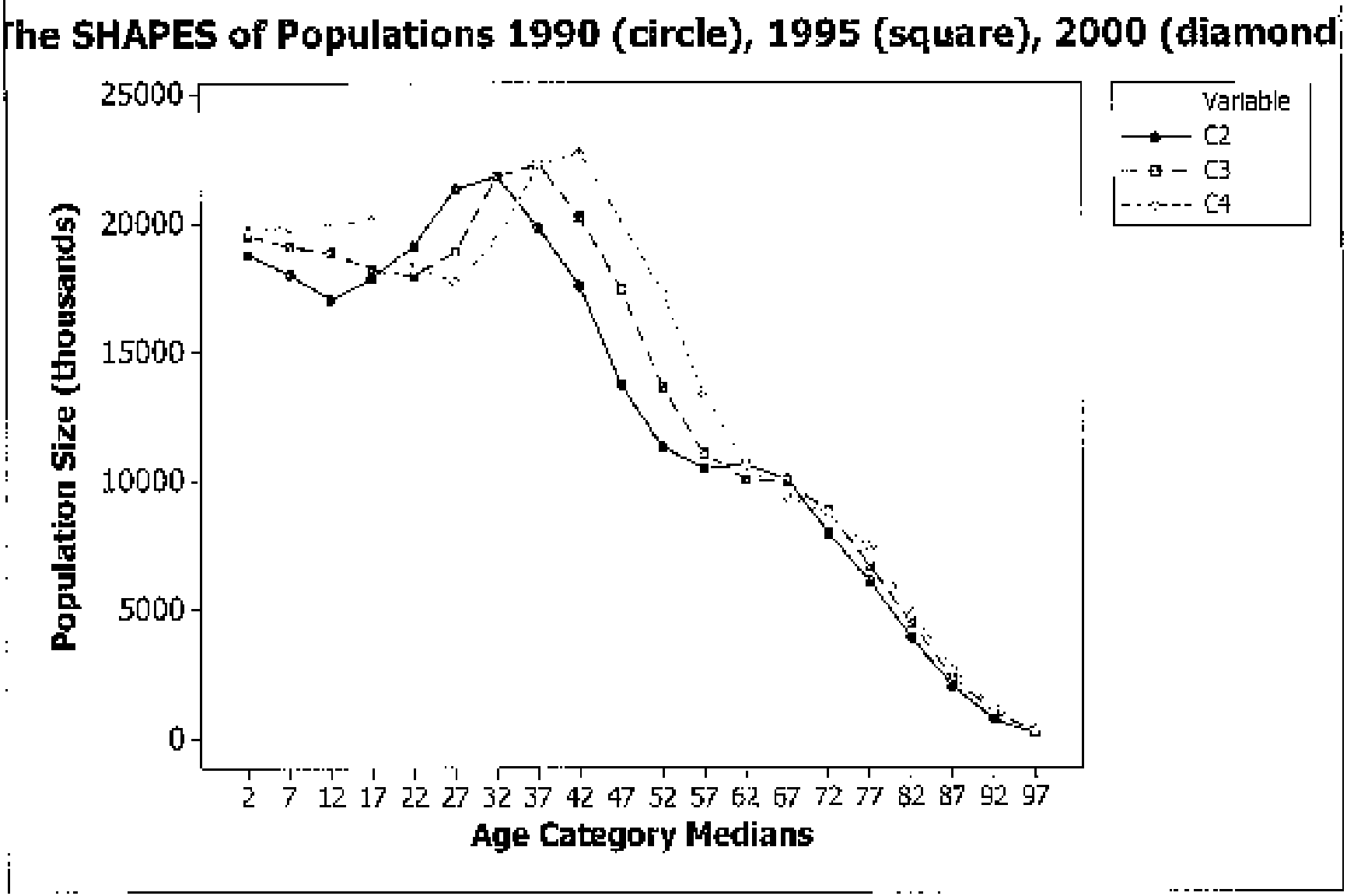
15-44
120538

19532 →

← x

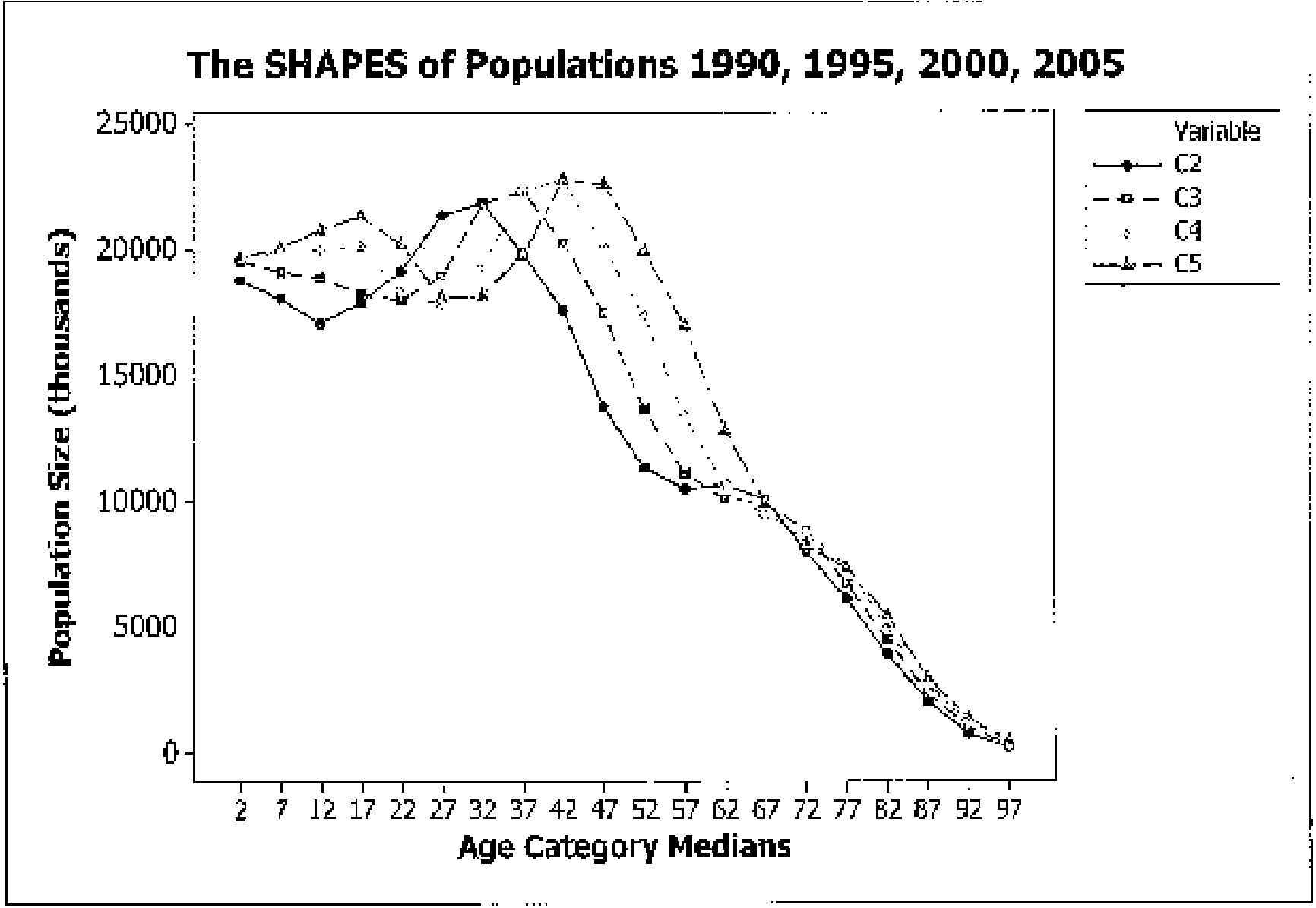
$$\frac{119470}{19532} = \frac{120538}{x}$$

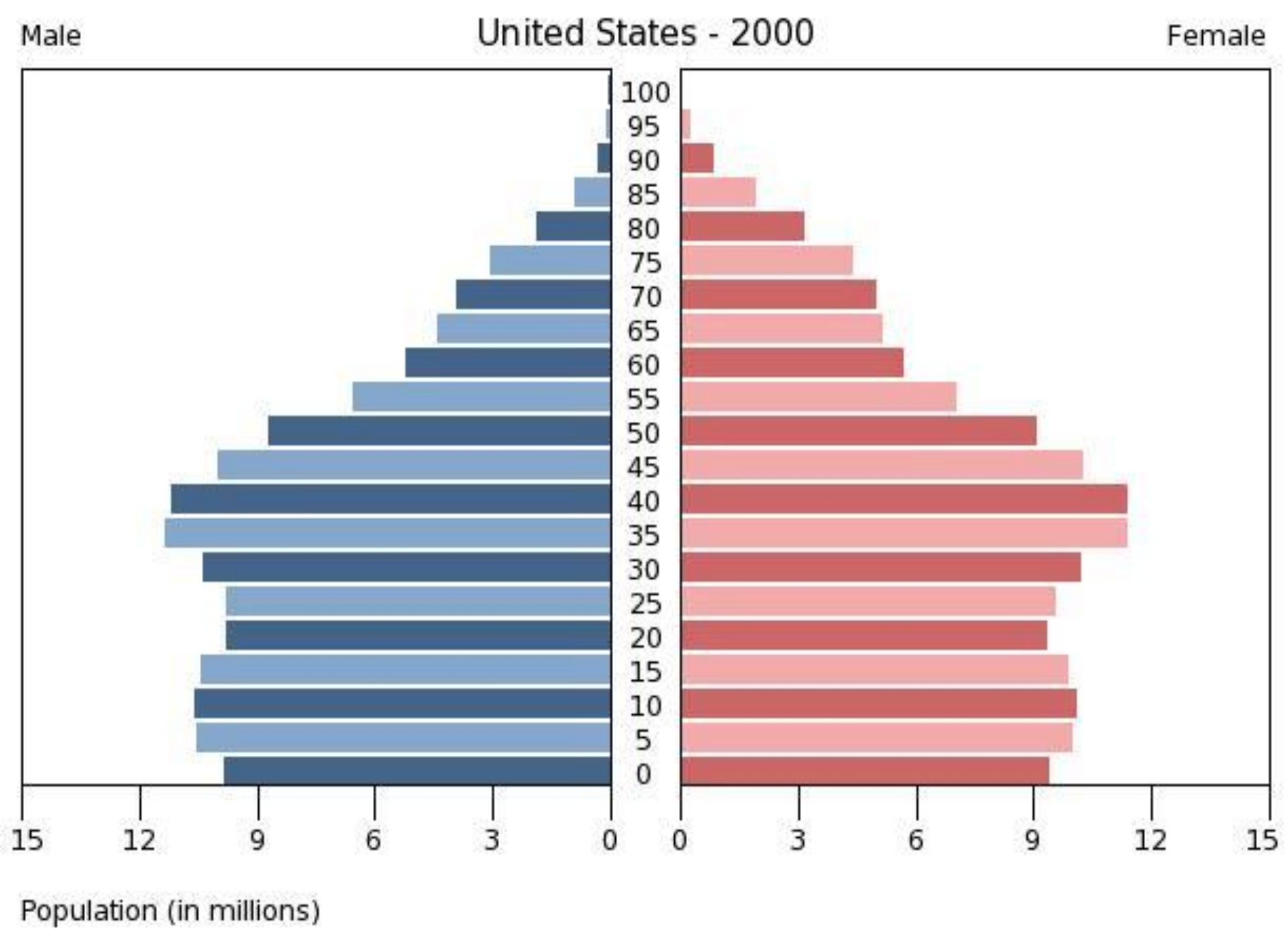
$$x = 19707$$


The SHAPES of Populations 1990 (circle), 1995 (square), 2000 (diamond
Variable
C2
C3
C4
Population Size (thousands)
25000
20000
15000
10000
5000
0
2 7 12 17 22 27 32 37 42 47 52 57 62 67 72 77 82 87 92 97
Age Category Medians

PREDICT THE AGE CATEGORY POPULATION SIZES FOR 2005

Base Layer	Base Pop 2000	Pop Factor	Next Layer	Next Pop 2005 est.
95-99	364	.2365	100+	
90-94	1176	.3583	95-99	
85-89	2694	.4998	90-94	
80-84	4915	.6015	85-89	
75-79	7414	.7336	80-84	
70-74	8707	.8395	75-79	
65-69	9383	.8772	70-74	
60-64	10618	.9340	65-69	
55-59	13366	.9578	60-64	
50-54	17325	.9798	55-59	
45-49	20103	.9924	50-54	
40-44	22755	.9923	45-49	
35-39	22283	1.0206	40-44	
30-34	19338	1.0210	35-39	
25-29	17759	1.0229	30-34	
20-24	18294	.9876	25-29	
15-19	20109	1.0050	20-24	
10-14	19955	1.0666	15-19	
5-9	19876	1.0450	10-14	
0-4	19707	1.0176	5-9	
			0-4	


The SHAPES of Populations 1990, 1995, 2000, 2005
Population Size (thousands)
25000
20000
15000
10000
5000
0
2 7 12 17 22 27 32 37 42 47 52 57 62 67 72 77 82 87 92 97
Age Category Medians
Variable
C2
C3
C4
C5


United States - 2000
Male
Female
100
95
90
85
80
75
70
65
60
55
50
45
40
35
30
25
20
15
10
5
0
15
12
9
6
3
0
0
3
6
9
12
15
Population (in millions)