Distribution of High Emitters:
Perspective for Sample Selection for Inventory Model Development

ERG

Sandeep Kishan, Jim Lindner,

Cindy Palacios, Tim DeFries

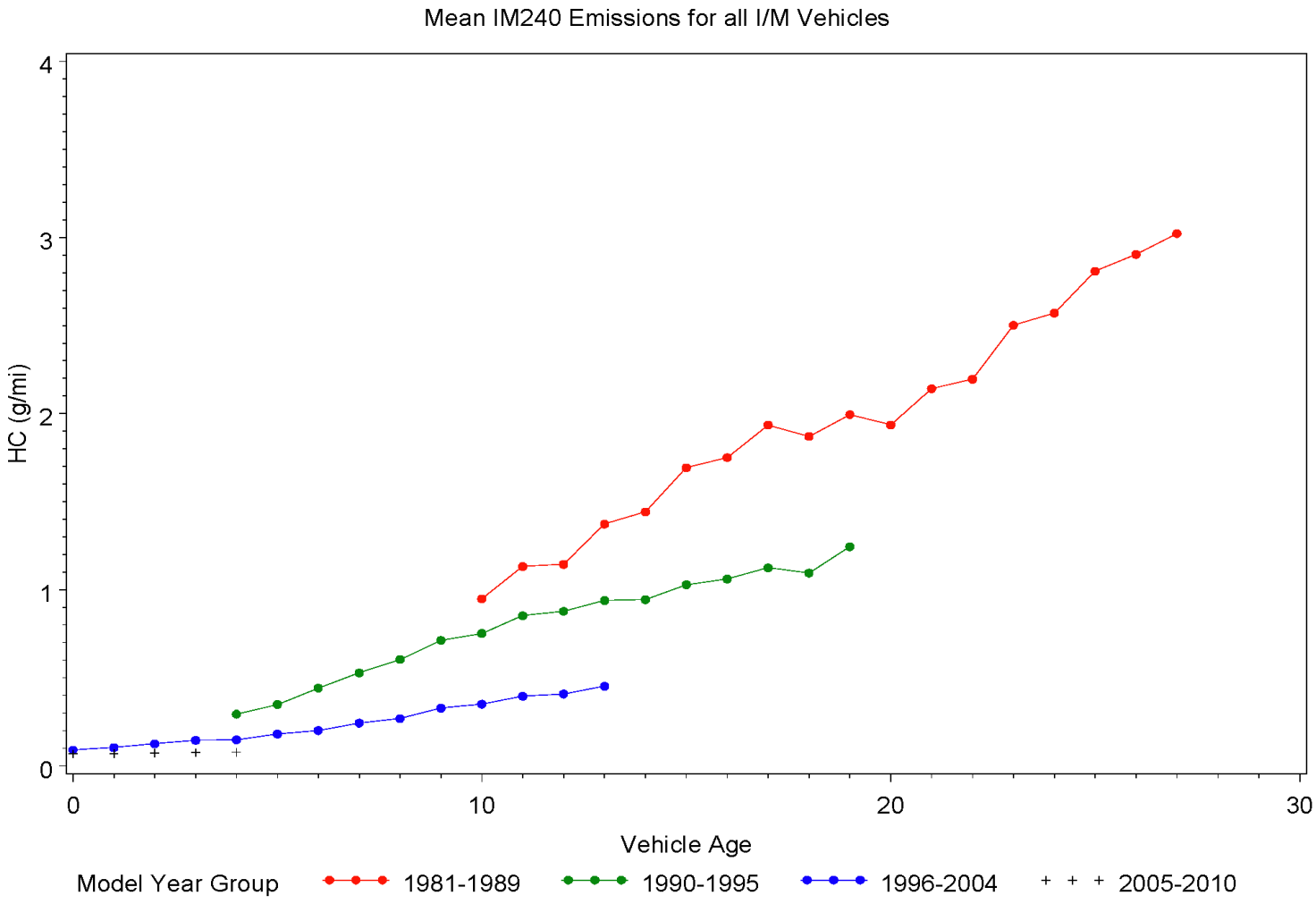
Ann Arbor

October 6th, 2010

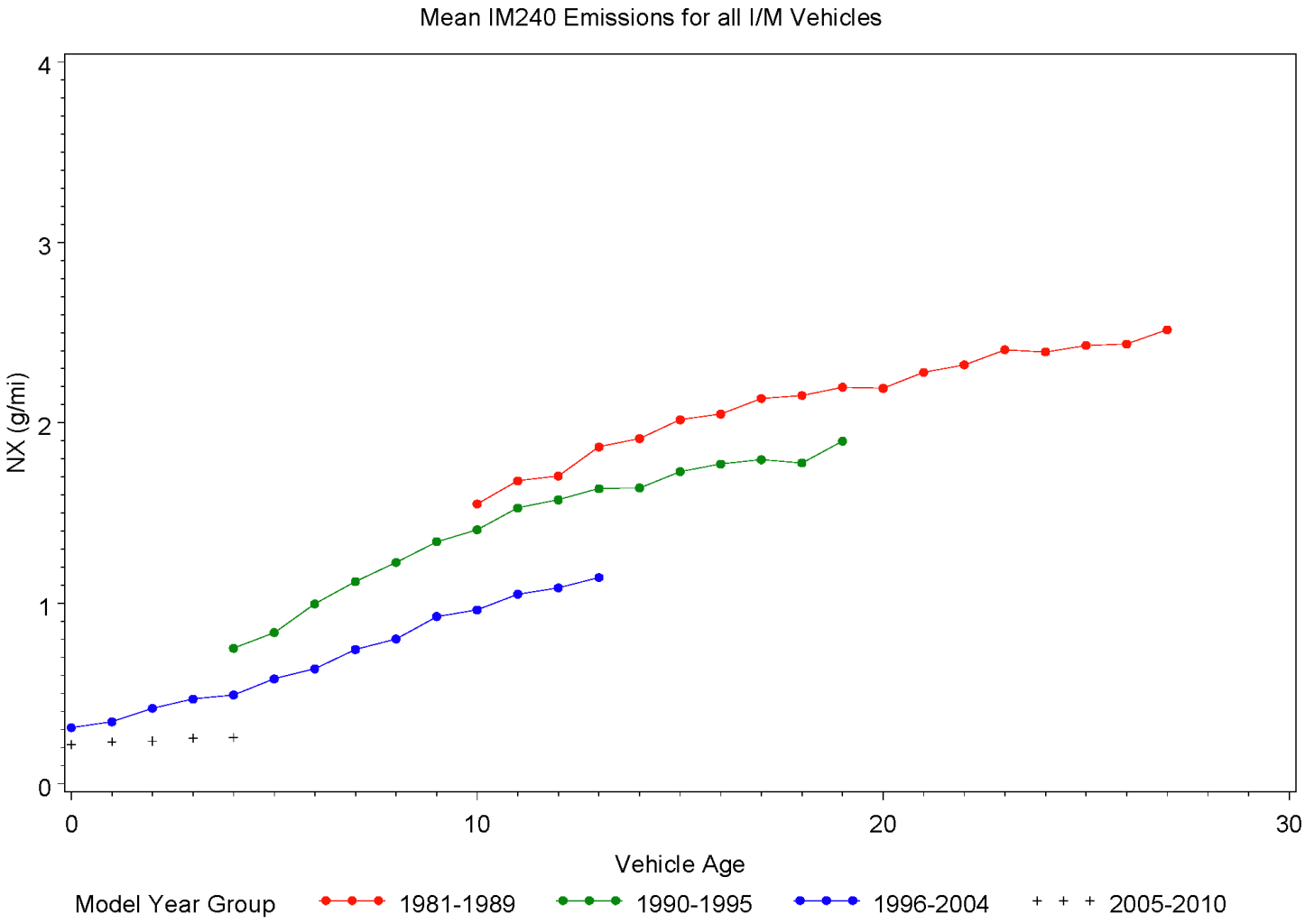
Objectives

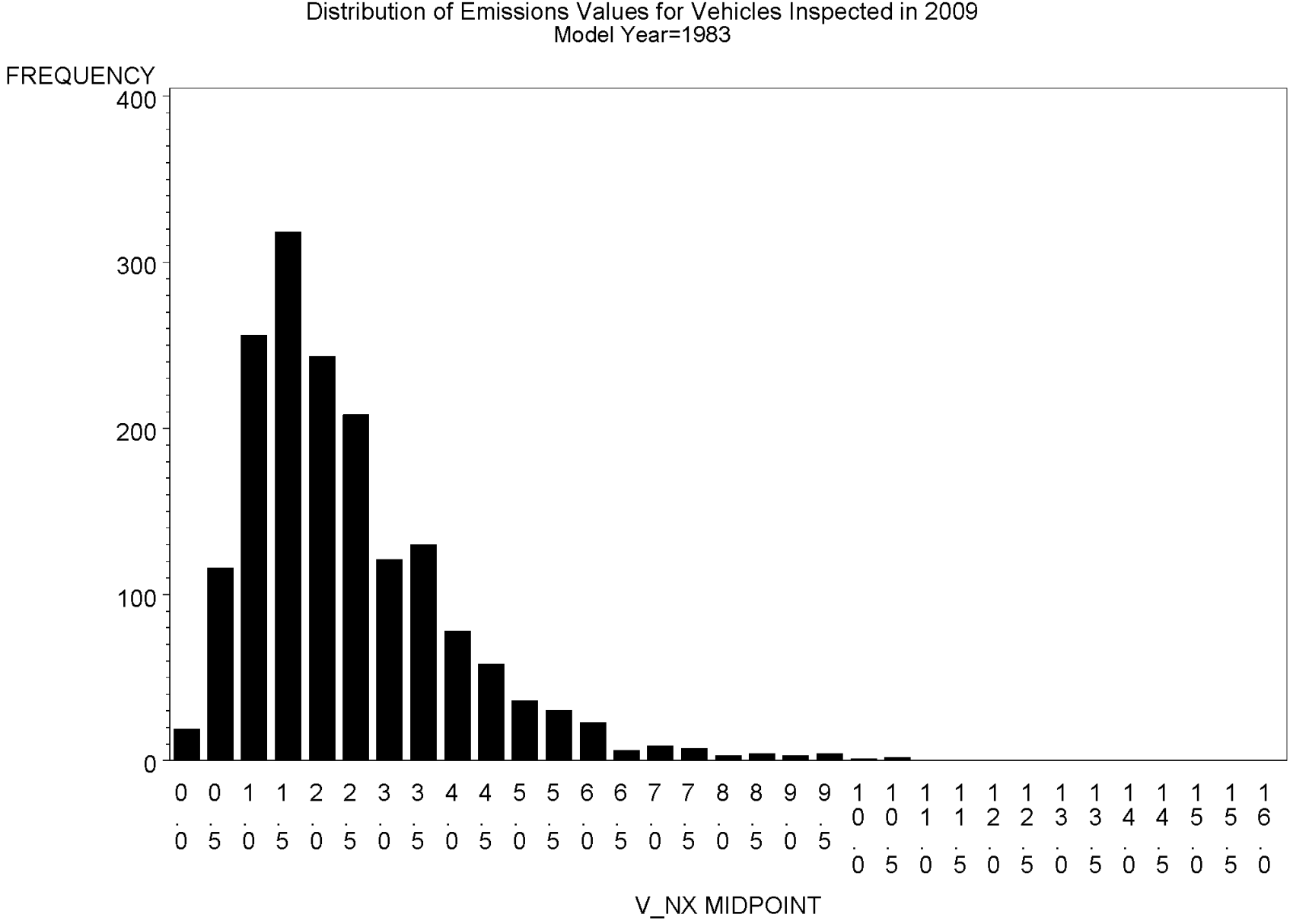
- What are contribution of High Emitters to Inventory?
- To estimate this, we must first identify these vehicles
- This presentation looks at recent Denver IM data, their distributions & trends to help further this discussion
- Tie in to EPA data collection efforts - longitudinal study example

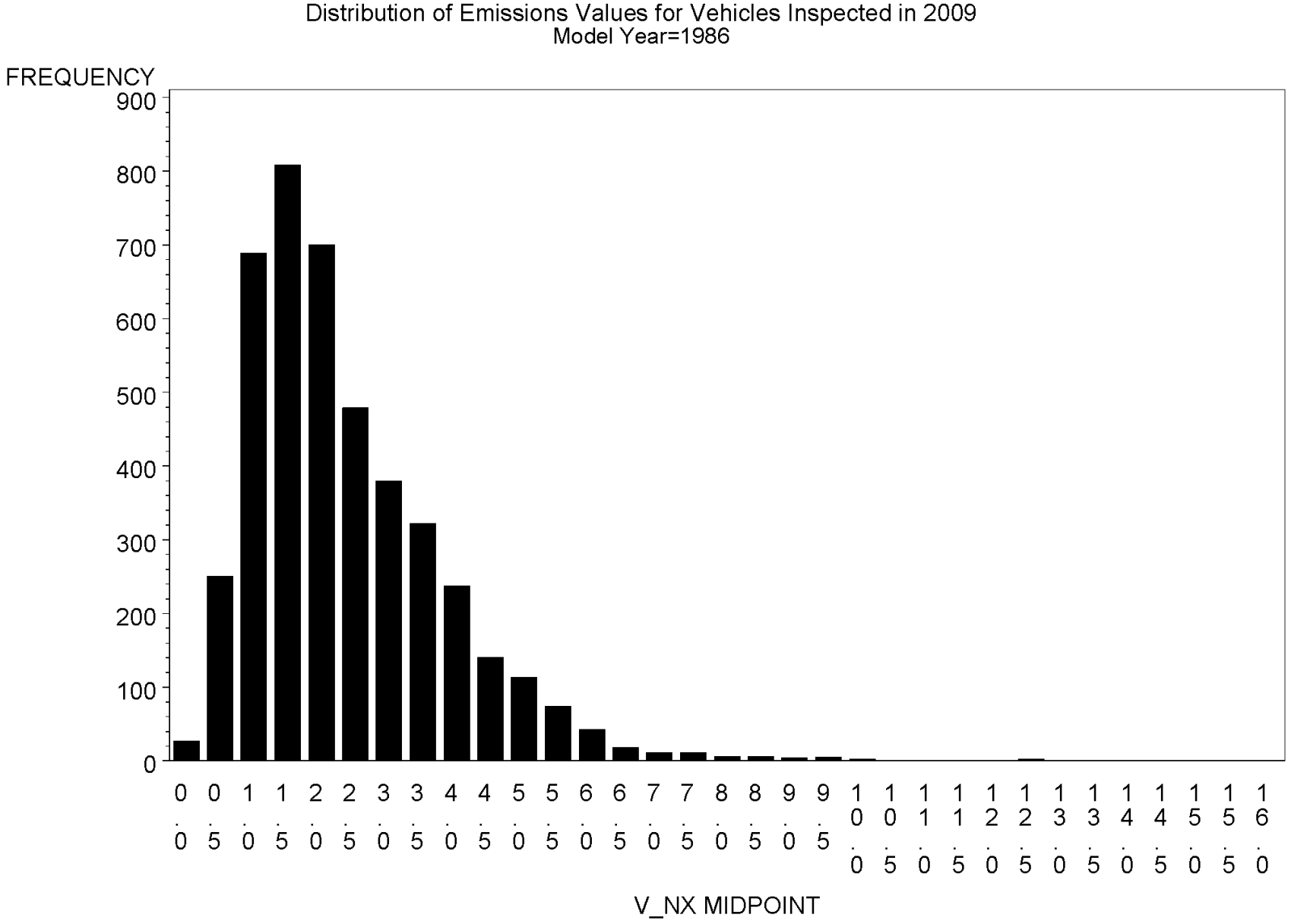
Mean HC Emissions

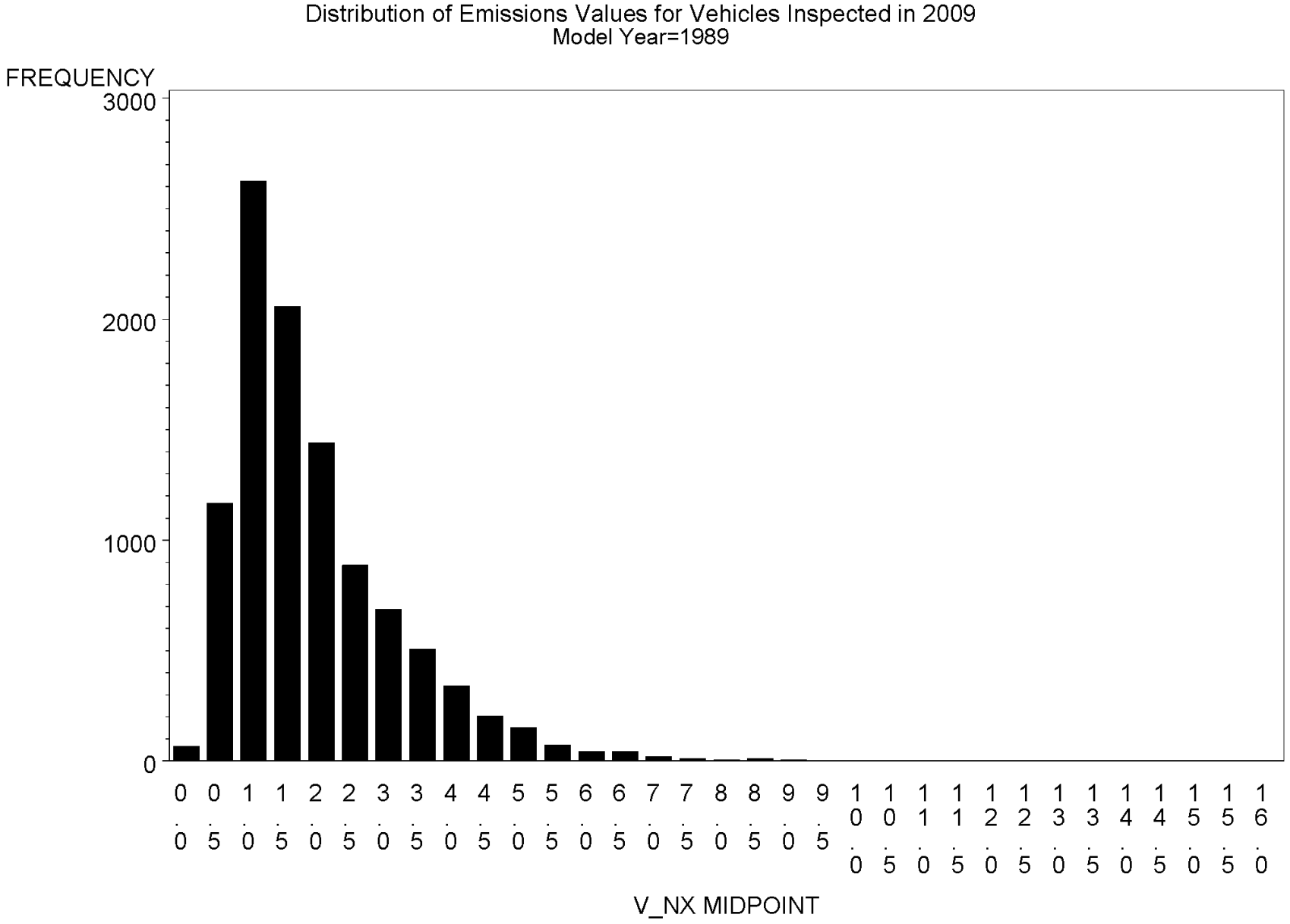


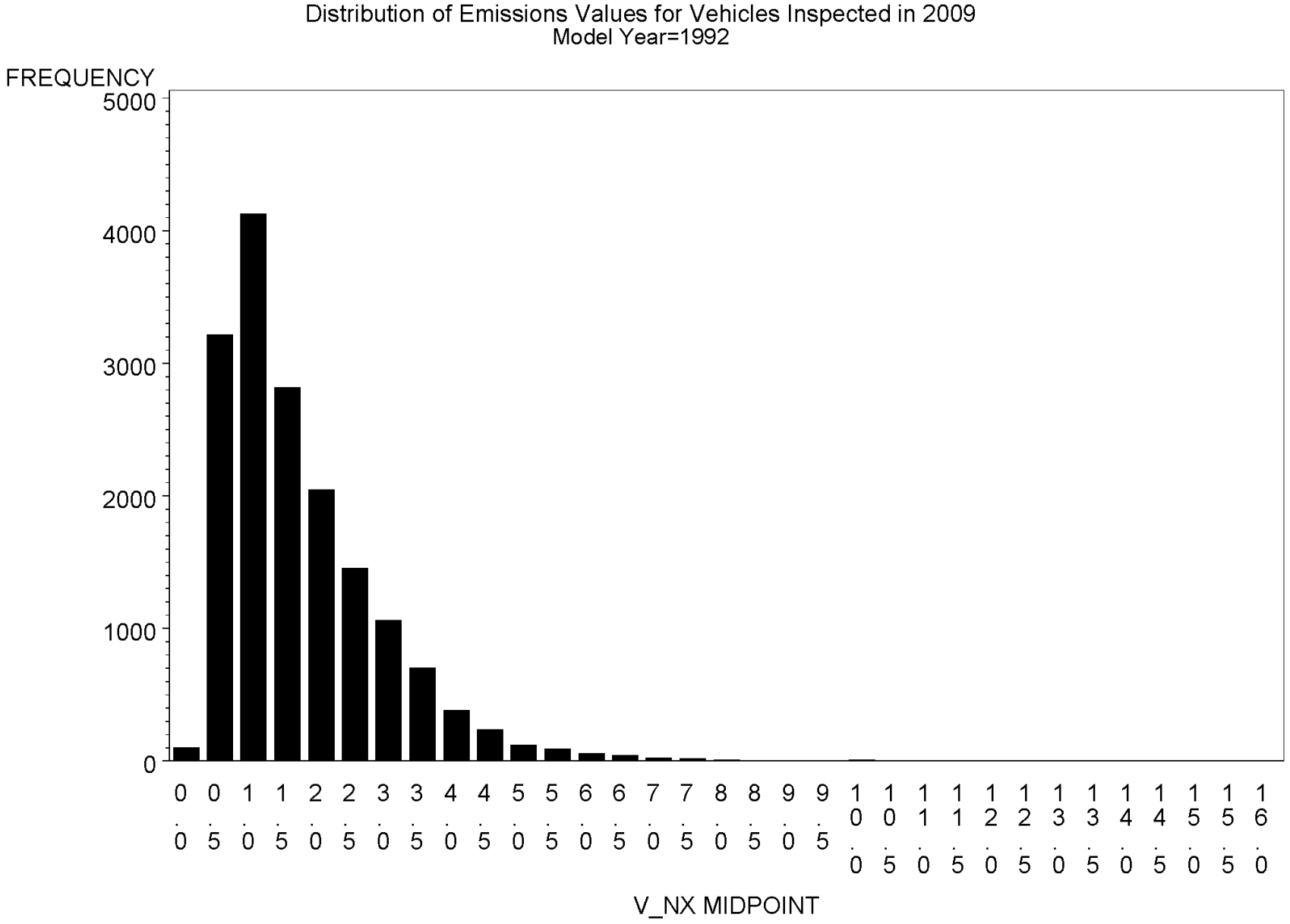
Mean NOx Emissions

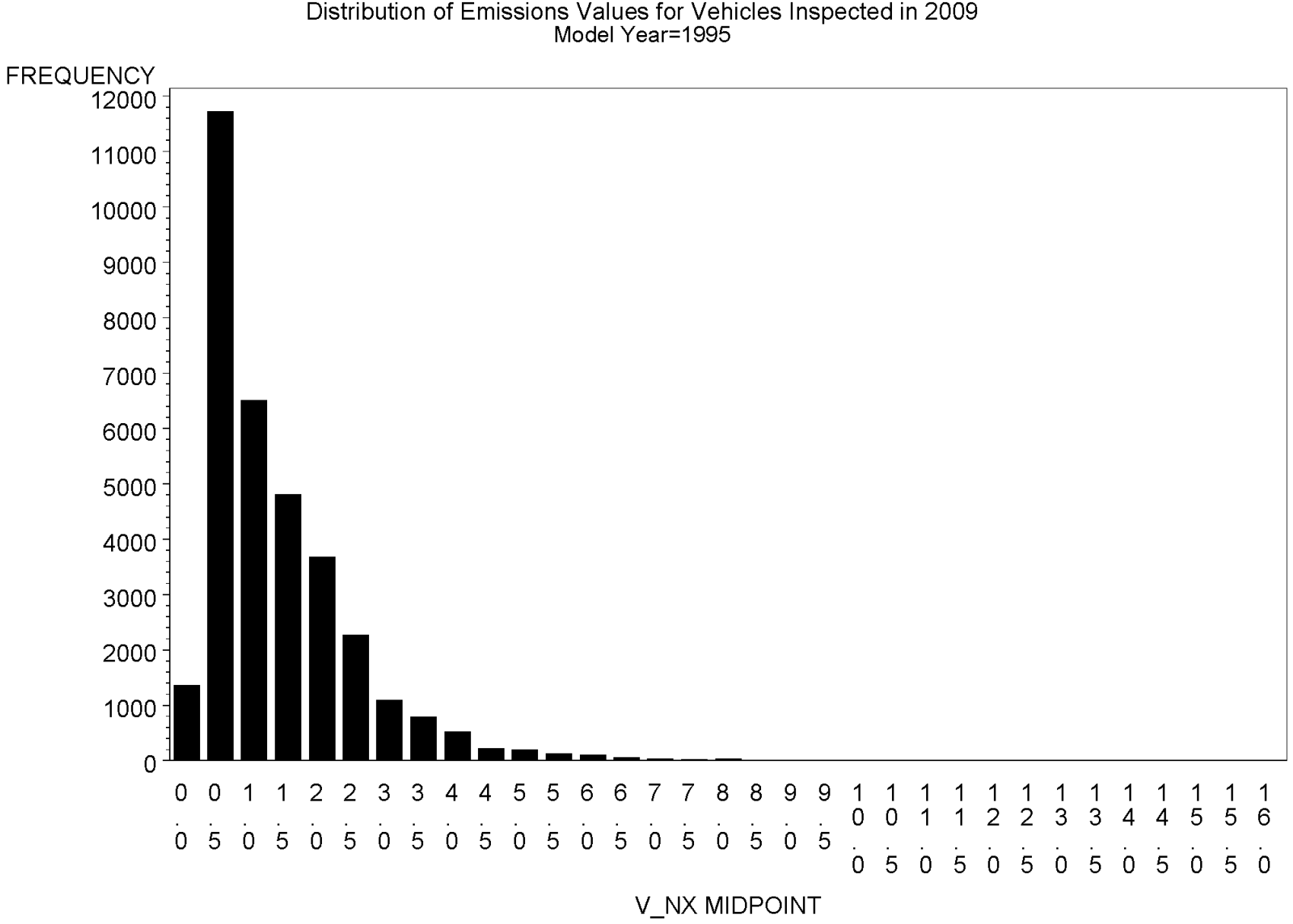


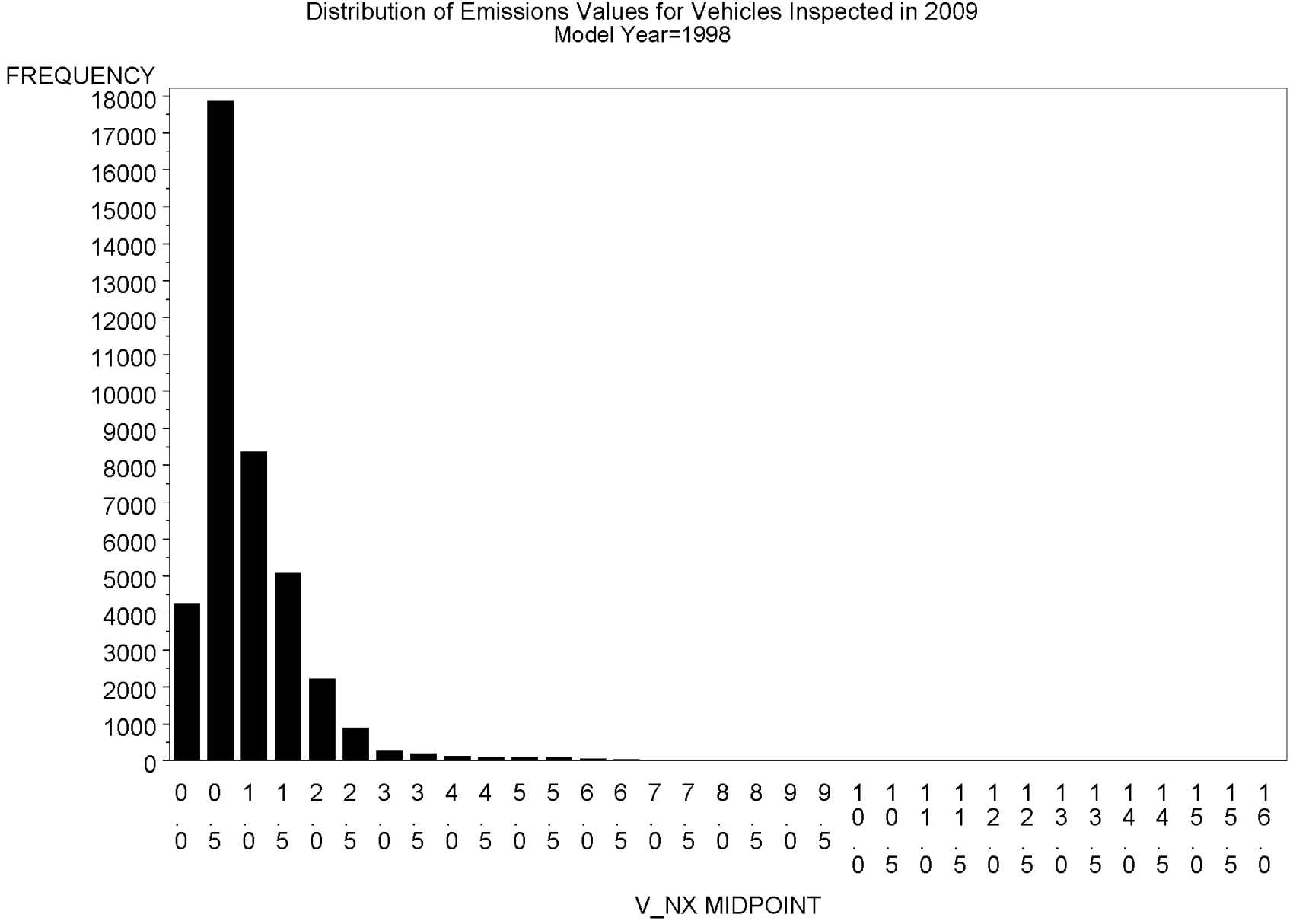

Distribution of Emissions Values for Vehicles Inspected in 2009
Model Year=1983
FREQUENCY
400
300
200
100
0
0.0
0.5
1.0
1.5
2.0
2.5
3.0
3.5
4.0
4.5
5.0
5.5
6.0
6.5
7.0
7.5
8.0
8.5
9.0
9.5
10.0
10.5
11.0
11.5
12.0
12.5
13.0
13.5
14.0
14.5
15.0
15.5
16.0
V_NX MIDPOINT
/proj1/Colorado_OBD_IM240/OTAQ/EPAPresentation/presentation.sas 24SEP10 00:30

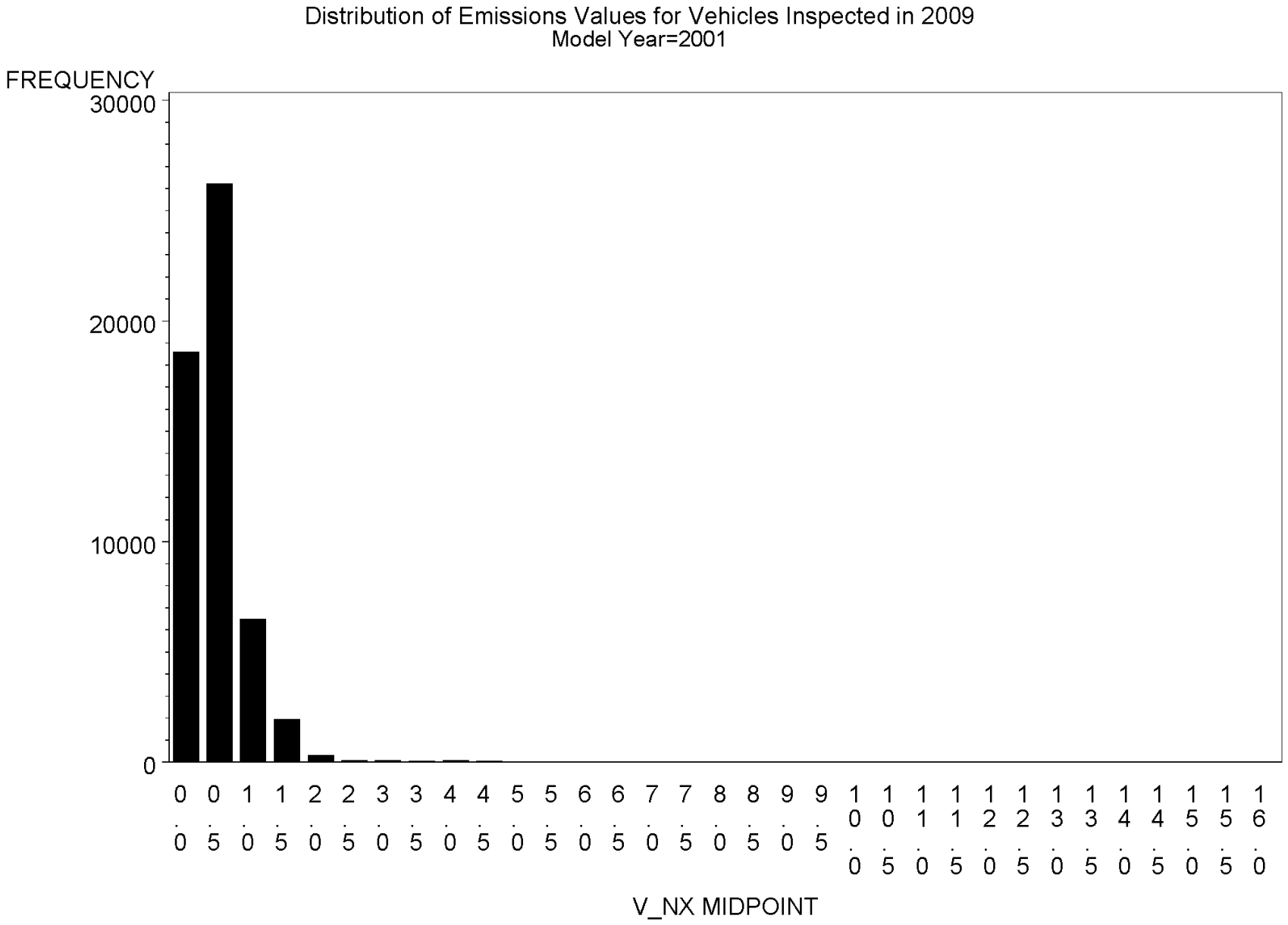

Distribution of Emissions Values for Vehicles Inspected in 2009
Model Year=1986
FREQUENCY
900
800
700
600
500
400
300
200
100
0
0.0
0.05
1.0
1.5
2.0
2.5
3.0
3.5
4.0
4.5
5.0
5.5
6.0
6.5
7.0
7.5
8.0
8.5
9.0
9.5
10.0
10.5
11.0
11.5
12.0
12.5
13.0
13.5
14.0
14.5
15.0
15.5
16.0
V_NX MIDPOINT
/proj1/Colorado_OBD_IM240/OTAQ/EPAPresentation/presentation.sas 24SEP10 00:30

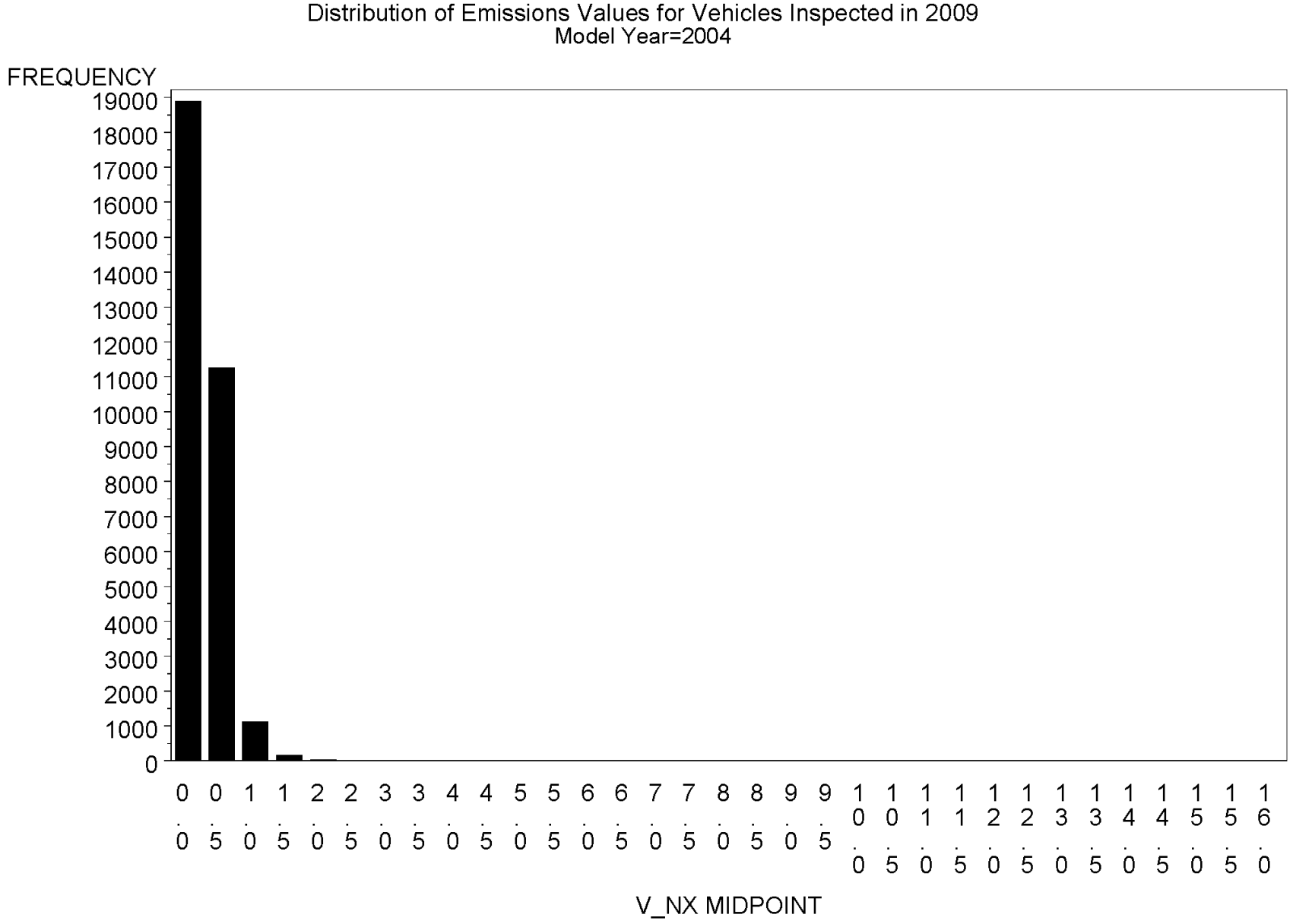

Distribution of Emissions Values for Vehicles Inspected in 2009
Model Year=1989
FREQUENCY
3000
2000
1000
0
0.0
0.5
1.0
1.5
2.0
2.5
3.0
3.5
4.0
4.5
5.0
5.5
6.0
6.5
7.0
7.5
8.0
8.5
9.0
9.5
10.0
10.5
11.0
11.5
12.0
12.5
13.0
13.5
14.0
14.5
15.0
15.5
16.0
V_NX MIDPOINT
/proj1/Colorado_OBD_IM240/OTAQ/EPAPresentation/presentation.sas 24SEP10 00:30

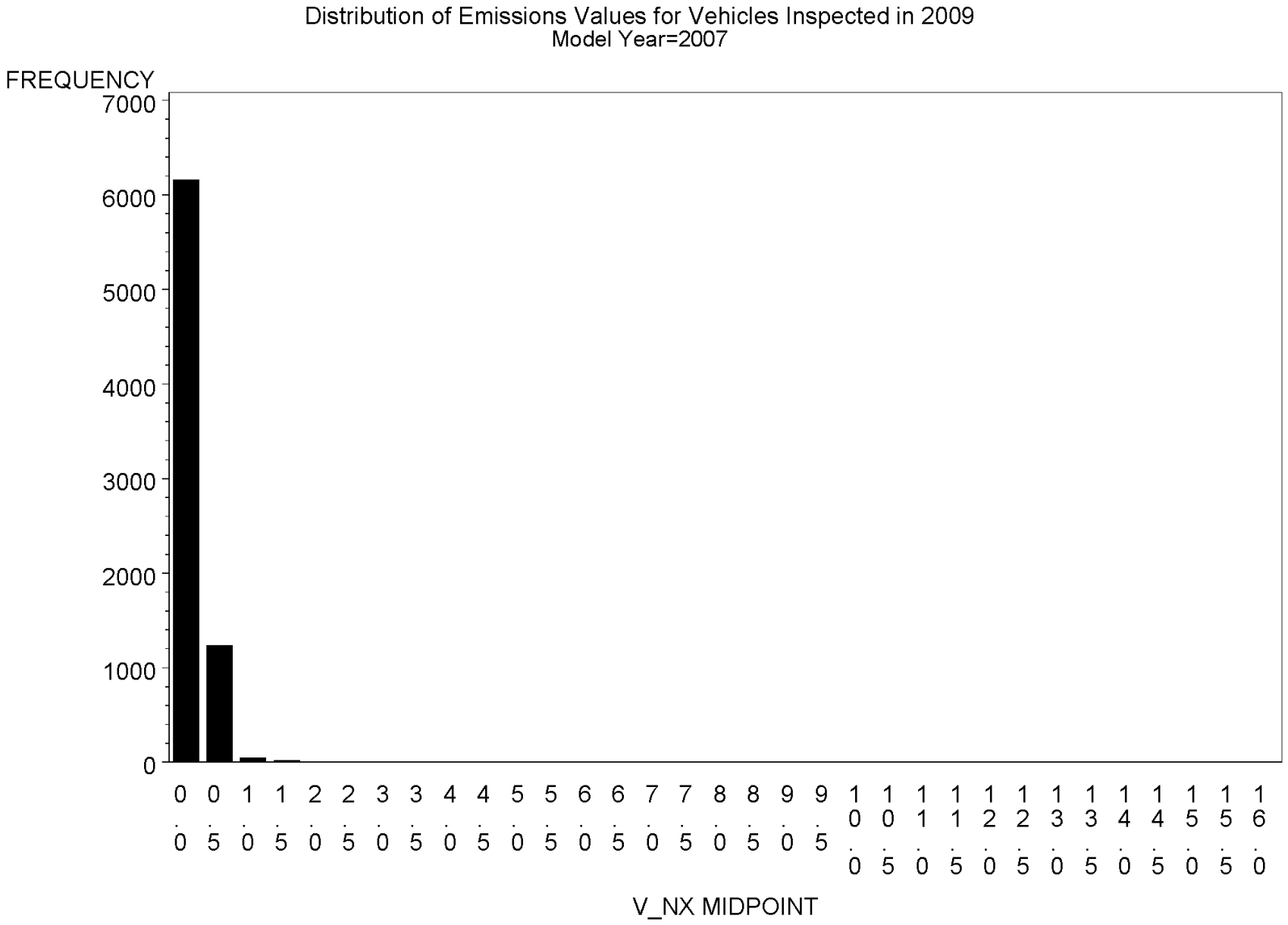

Distribution of Emissions Values for Vehicles Inspected in 2009
Model Year=1992
FREQUENCY
5000
4000
3000
2000
1000
0
0.0
0.5
1.0
1.5
2.0
2.5
3.0
3.5
4.0
4.5
5.0
5.5
6.0
6.5
7.0
7.5
8.0
8.5
9.0
9.5
10.0
10.5
11.0
11.5
12.0
12.5
13.0
13.5
14.0
14.5
15.0
15.5
16.0
V_NX MIDPOINT
/proj1/Colorado_OBD_IM240/OTAQ/EPAPresentation/presentation.sas 24SEP10 00:30

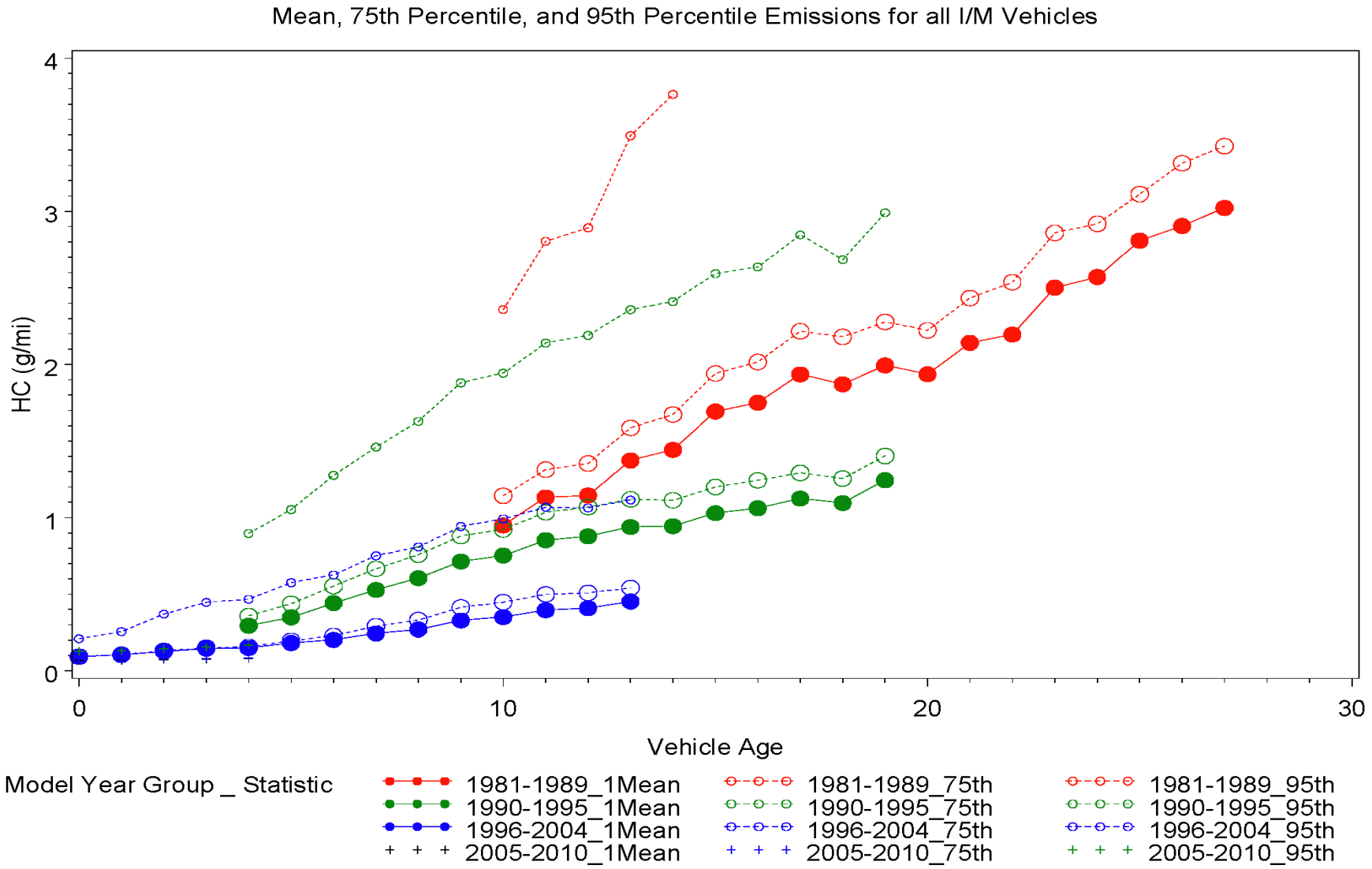

Distribution of Emissions Values for Vehicles Inspected in 2009
Model Year=1995
FREQUENCY
12000
11000
10000
9000
8000
7000
6000
5000
4000
3000
2000
1000
0
0.0
0.05
1.0
1.5
2.0
2.5
3.0
3.5
4.0
4.5
5.0
5.5
6.0
6.5
7.0
7.5
8.0
8.5
9.0
9.5
10.0
10.5
11.0
11.5
12.0
12.5
13.0
13.5
14.0
14.5
15.0
15.5
16.0
V_NX MIDPOINT
/proj1/Colorado_OBD_IM240/OTAQ/EPAPresentation/presentation.sas 24SEP10 00:30


Distribution of Emissions Values for Vehicles Inspected in 2009
Model Year=1998
FREQUENCY
18000
17000
16000
15000
14000
13000
12000
11000
10000
9000
8000
7000
6000
5000
4000
3000
2000
1000
0
0.0
0.5
1.0
1.5
2.0
2.5
3.0
3.5
4.0
4.5
5.0
5.5
6.0
6.5
7.0
7.5
8.0
8.5
9.0
9.5
10.0
10.5
11.0
11.5
12.0
12.5
13.0
13.5
14.0
14.5
15.0
15.5
16.0
V_NX MIDPOINT
/proj1/Colorado_OBD_IM240/OTAQ/EPAPresentation/presentation.sas 24SEP10 00:30

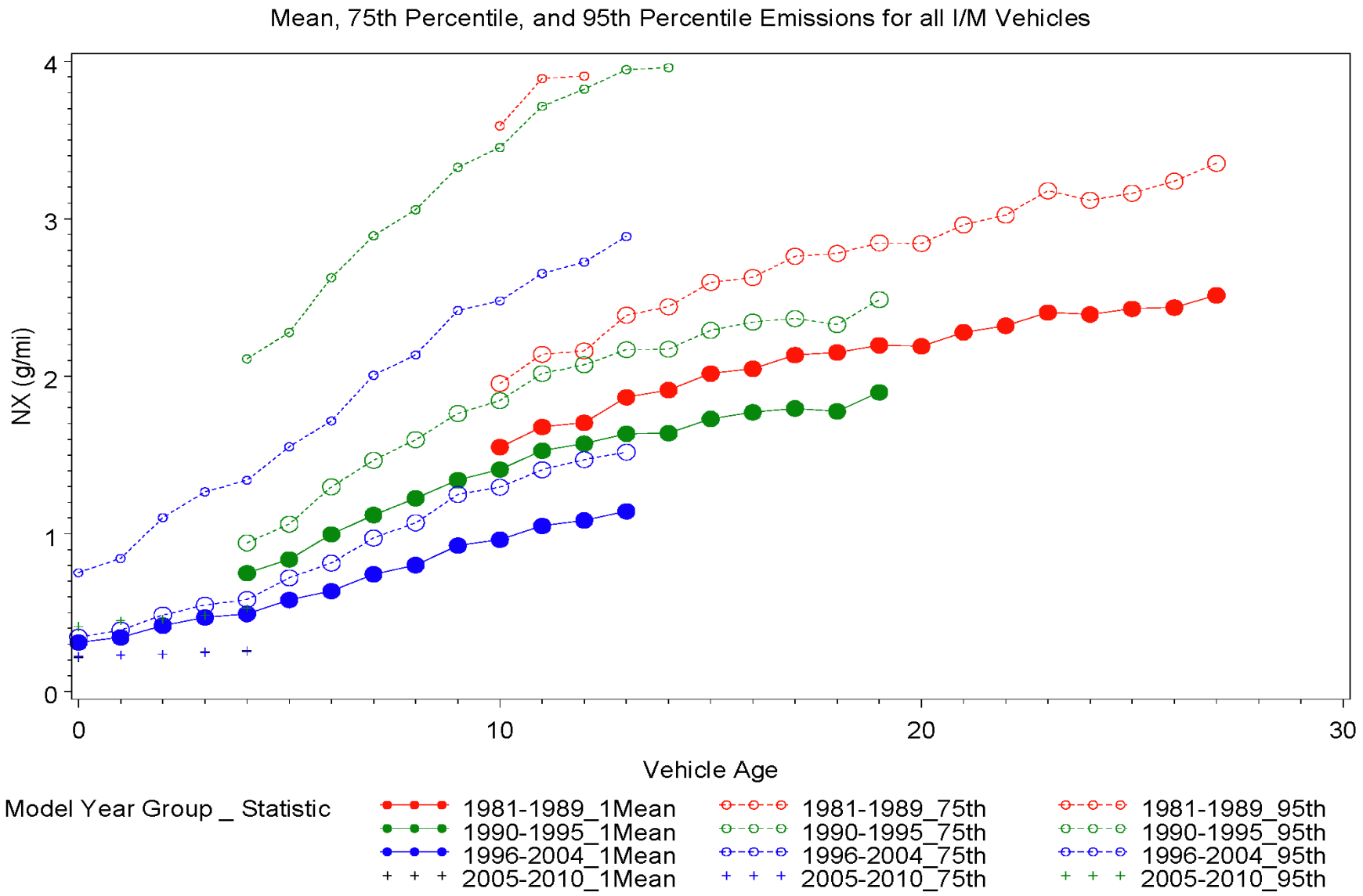

Distribution of Emissions Values for Vehicles Inspected in 2009
Model Year=2001
FREQUENCY
30000
20000
10000
0
0.0
0.05
1.0
1.5
2.0
2.5
3.0
3.5
4.0
4.5
5.0
5.5
6.0
6.5
7.0
7.5
8.0
8.5
9.0
9.5
10.0
10.5
11.0
11.5
12.0
12.5
13.0
13.5
14.0
14.5
15.0
15.5
16.0
V_NX MIDPOINT
/proj1/Colorado_OBD_IM240/OTAQ/EPAPresentation/presentation.sas 24SEP10 00:30


Distribution of Emissions Values for Vehicles Inspected in 2009
Model Year=2004
FREQUENCY
19000
18000
17000
16000
15000
14000
13000
12000
11000
10000
9000
8000
7000
6000
5000
4000
3000
2000
1000
0
0.0
0.05
0.10
0.15
0.20
0.25
0.30
0.35
0.40
0.45
0.50
0.55
0.60
0.65
0.70
0.75
0.80
0.85
0.90
0.95
1.00
1.05
1.10
1.15
1.20
1.25
1.30
1.35
1.40
1.45
1.50
1.55
1.60
V_NX MIDPOINT
/proj1/Colorado_OBD_IM240/OTAQ/EPAPresentation/presentation.sas 24SEP10 00:30


Distribution of Emissions Values for Vehicles Inspected in 2009
Model Year=2007
FREQUENCY
7000
6000
5000
4000
3000
2000
1000
0
0.0
0.05
0.10
0.15
0.20
0.25
0.30
0.35
0.40
0.45
0.50
0.55
0.60
0.65
0.70
0.75
0.80
0.85
0.90
0.95
1.00
1.05
1.10
1.15
1.20
1.25
1.30
1.35
1.40
1.45
1.50
1.55
1.60
V_NX MIDPOINT
/proj1/Colorado_OBD_IM240/OTAQ/EPAPresentation/presentation.sas 24SEP10 00:30

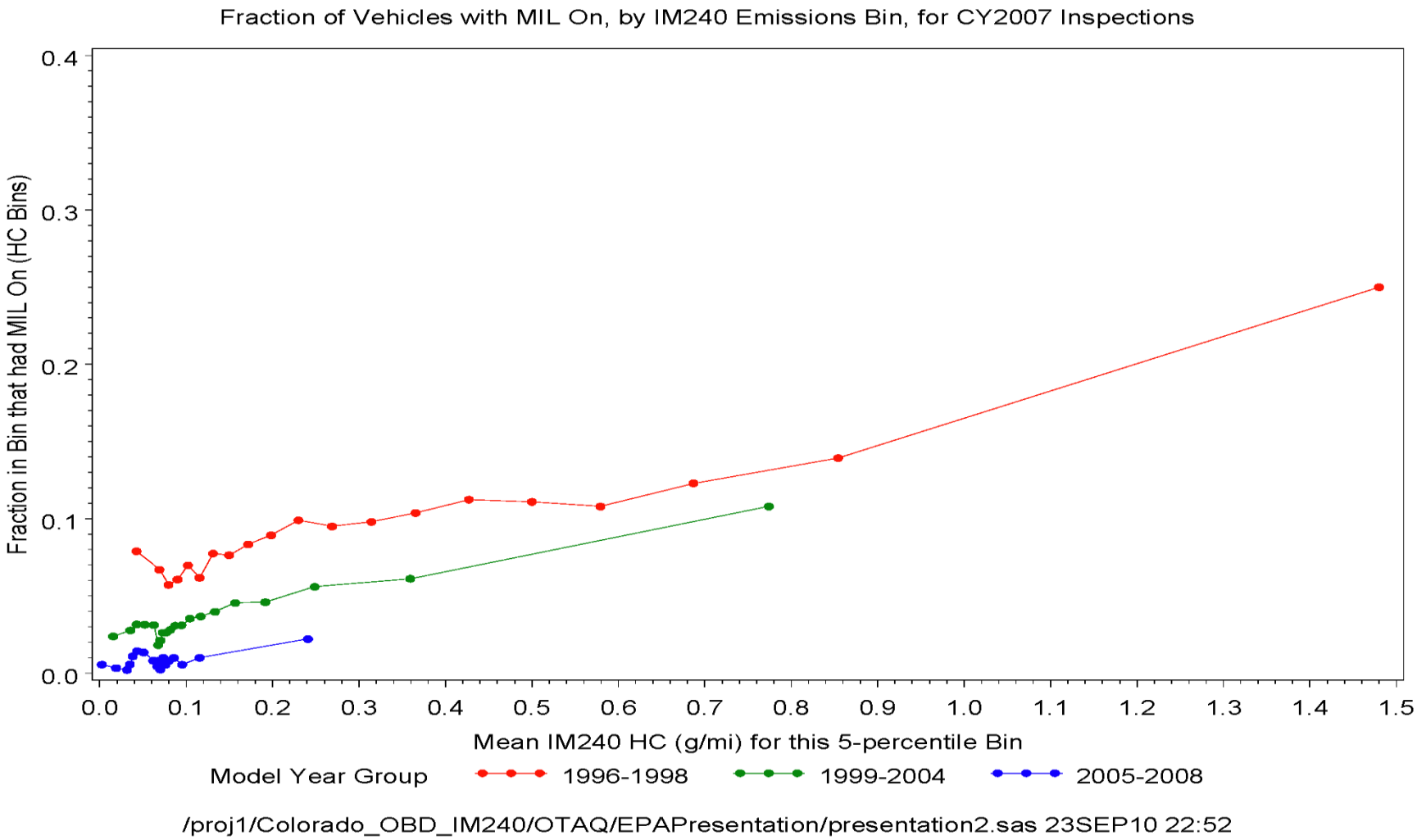


/proj1/Colorado_OBD_IM240/OTAQ/EPAPresentation/presentation.sas 24SEP10 00:30

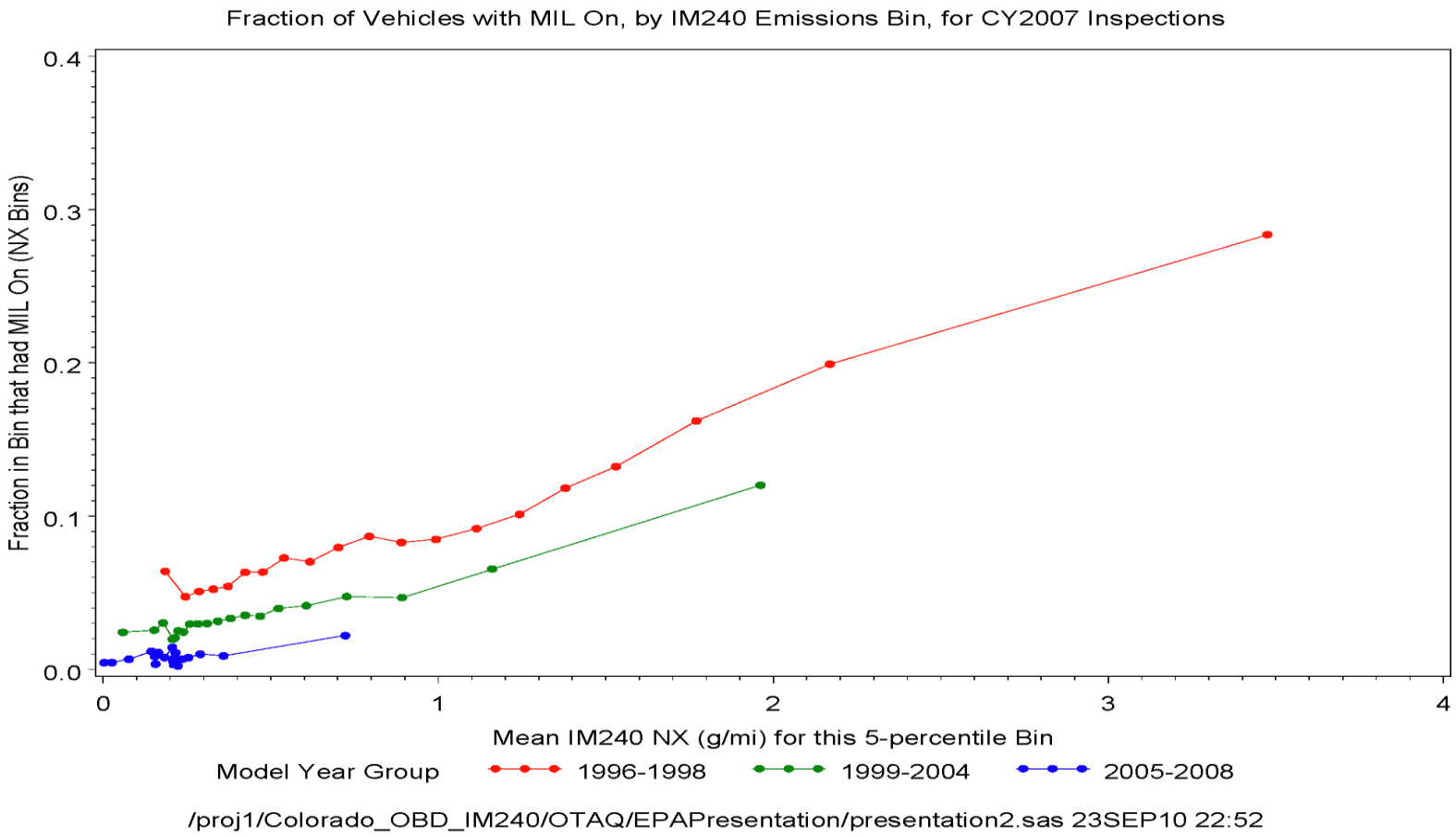


/proj1/Colorado_OBD_IM240/OTAQ/EPAPresentation/presentation.sas 24SEP10 00:30

Fraction of vehicles with MIL On – HC Emissions Bins



Fraction of vehicles with MIL On – NOx Emissions Bins

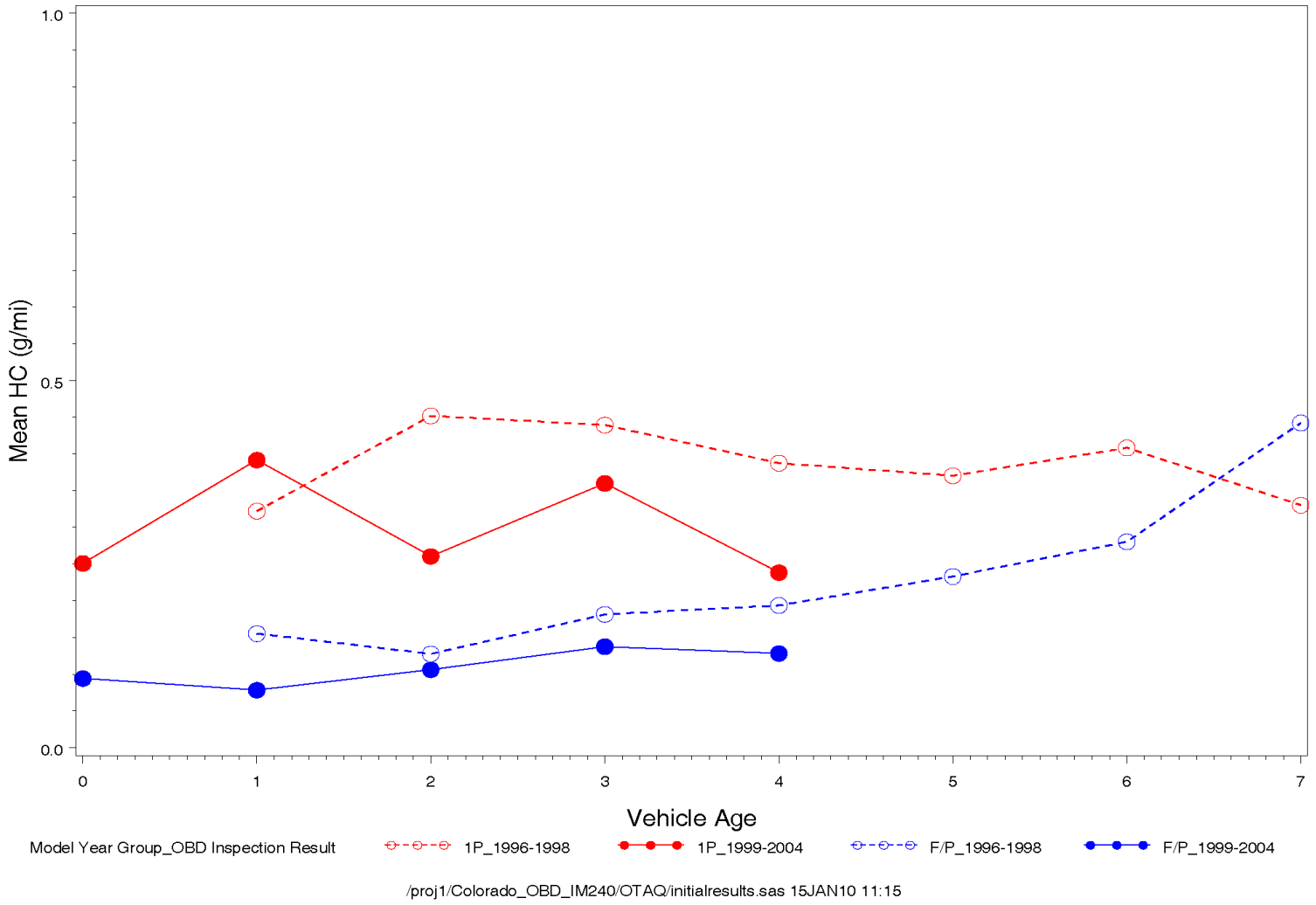


OBD Repair Effectiveness

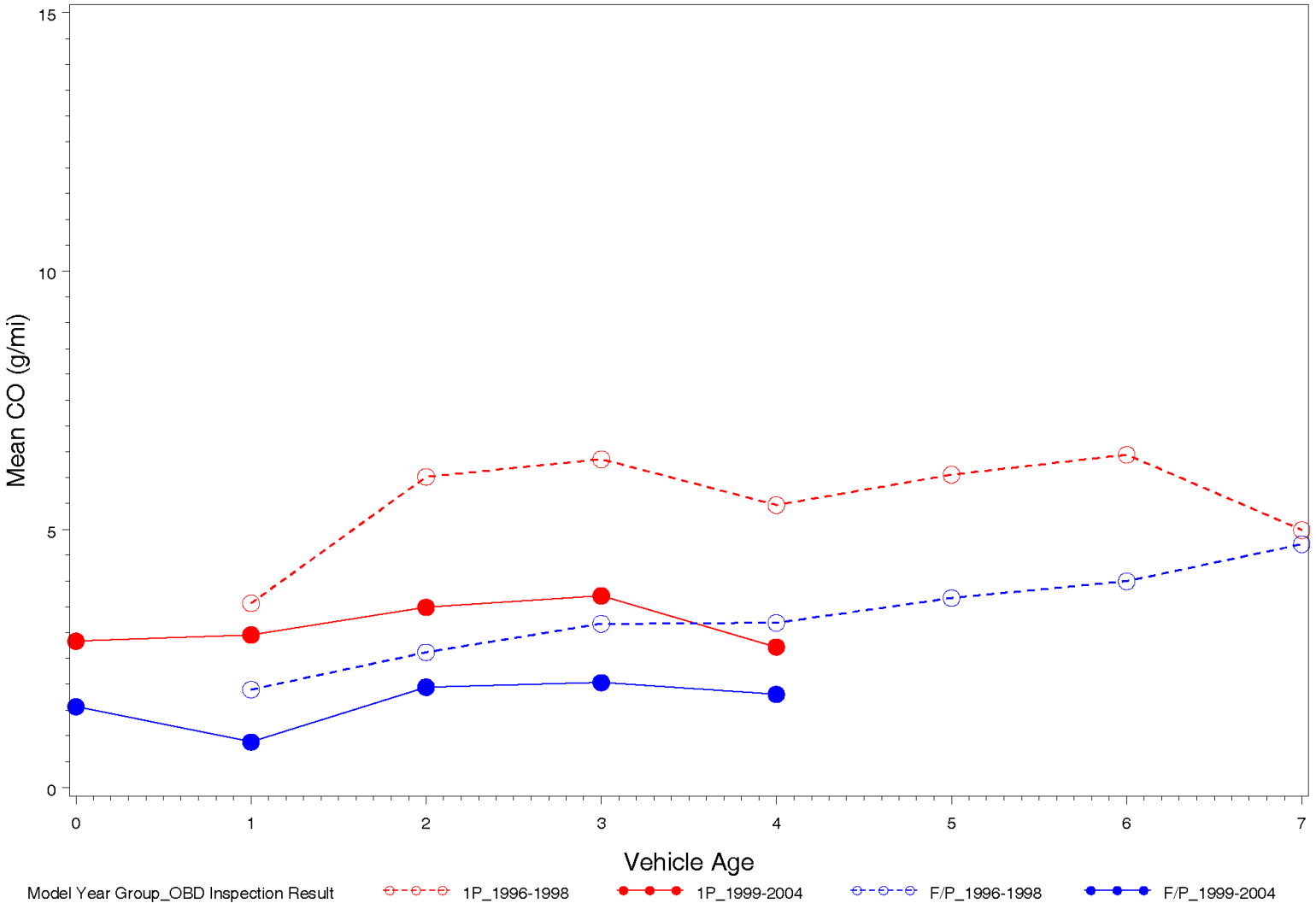
- Compare the fail/repair/pass emissions levels to initial pass emissions levels using only OBD result for pass/fail identification & IM240 measurement for emissions delta
- It was seen that the “pass’ retest emission levels for vehicles that the initially failed their OBD test were in fact lower than vehicles with an initial passing OBD result.
- This is a significant finding.
- Our observations over the years indicate that emission rates of vehicles repaired as a result of emission tests in I/M programs never get as low, never mind below, that of initially passing vehicles.
- Validation that OBD enhances the repairs of vehicles in ways that an emission test alone are unable to accomplish.

IM240 HC Initial Pass & Final Pass for Fail/Repair/Pass

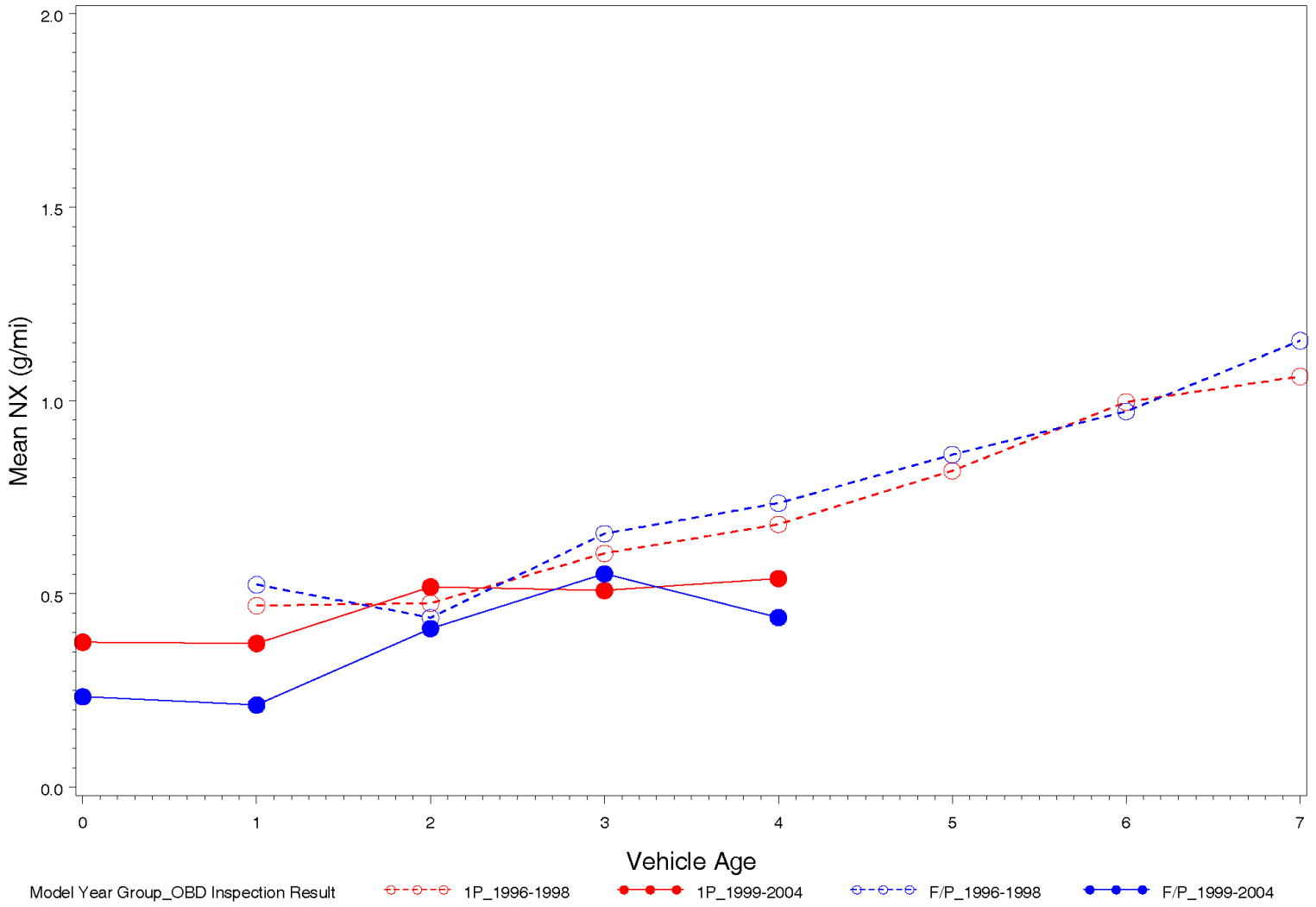
,



IM240 CO Initial Pass & Final Pass for Fail/Repair/Pass



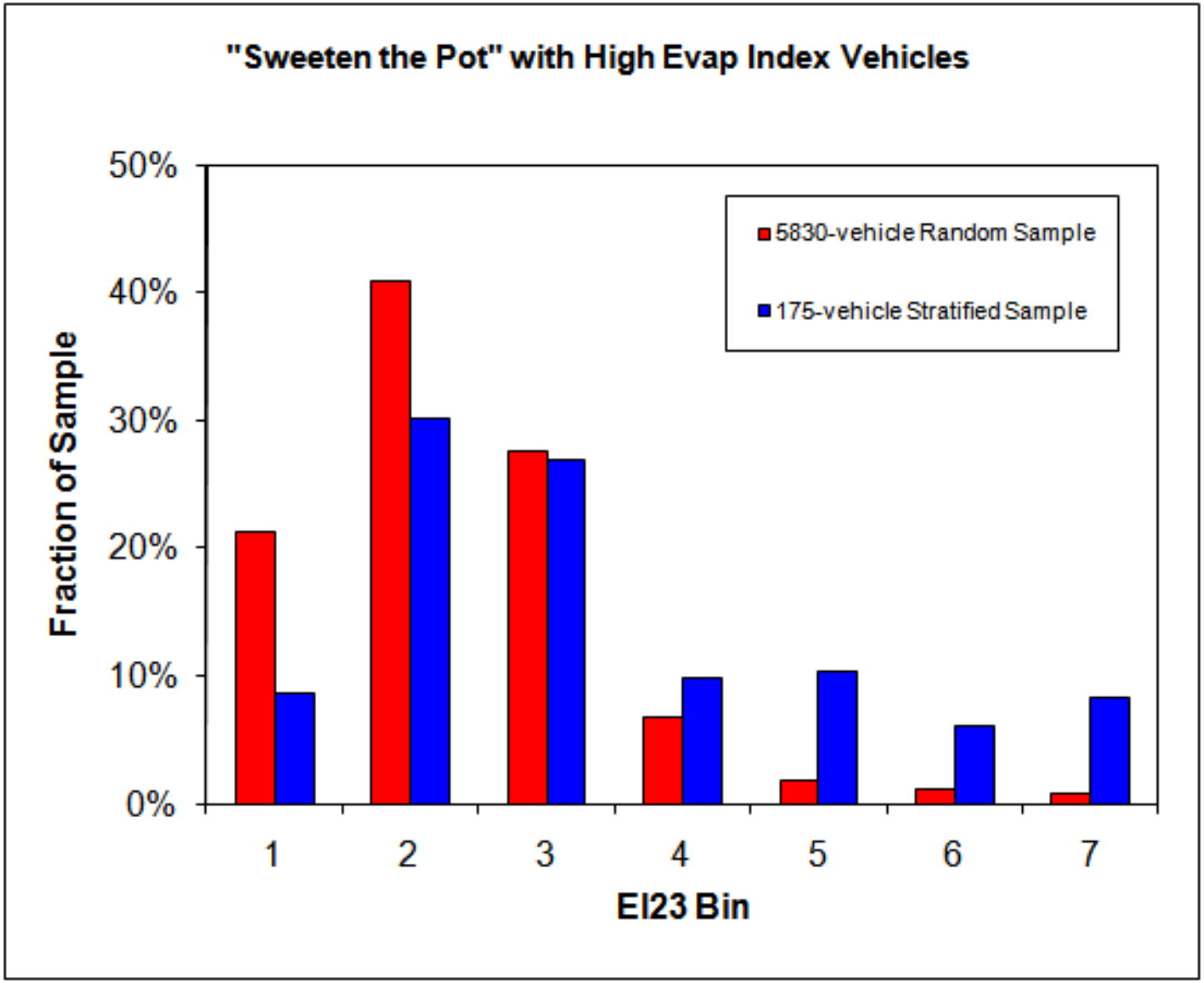
IM240 NOx Initial Pass & Final Pass for Fail/Repair/Pass



How should we sample a fleet for model development

- In light of these skewed distributions, random sampling is not sufficient
- EPA/ERG have been using a surrogate measure to develop a stratified random sample in the following studies:
  - Kansas City Study
  - Midwest Non Road Construction Vehicles
  - Houston Port Drayage Vehicles
  - Colorado High Evaporative Emissions
  - Planned Tier 2 Longitudinal Deterioration Study
- Use of RSD as a preliminary surrogate measure

Example 1 : Sampling for the Colorado High Evaporative Emitter Project



Example 2 : Sampling for Tier 2 Longitudinal Deterioration Study

- Relate RSD to IM240
  - Based on data collected at Lipan Station (Denver)
    - Remote sensing outside station
    - IM240 tests in lanes (within minutes of the RSD)
    - Obtained 2,096 paired results
- Develop a Stratified sampling design (Jim Warila/EPA assisted in this effort)
- Seek high-emitting vehicles
  - For any of HC, CO, NOx
  - Rather than targeting one pollutant

Sampling for Tier 2 - Model Development

- Modelled exceedance probabilities
  - Using Logistic Regression
- Applied several thresholds to full IM240 values
  - to generate successive exceedance probabilities
- Multiple models by pollutant considered:
  - model year,
  - measured IM240 emission rates for HC, CO, and NX,
  - artificial IM240 cutpoints for HC, CO, and NX,
  - the IM240 pass/fail result for HC, CO, and NX,
  - RSD Method A emissions concs for HC, CO, and NX, and
  - the fractile assignment of the RSD HC, CO, and NX.

"Fractiles ???"

a.k.a. "reverse normalized rank"

... For example, out of 1,000 NOx measurements (FTP Bag 2), we rank in Reverse order and Calculate the fractiles ...

Why use "fractiles"?

1) *Incorporates negative RSDs*
2) *Puts RSD for all emissions in same scale*
3) *Incorporates non-linear, non-normal behavior*
4) *Convenience*

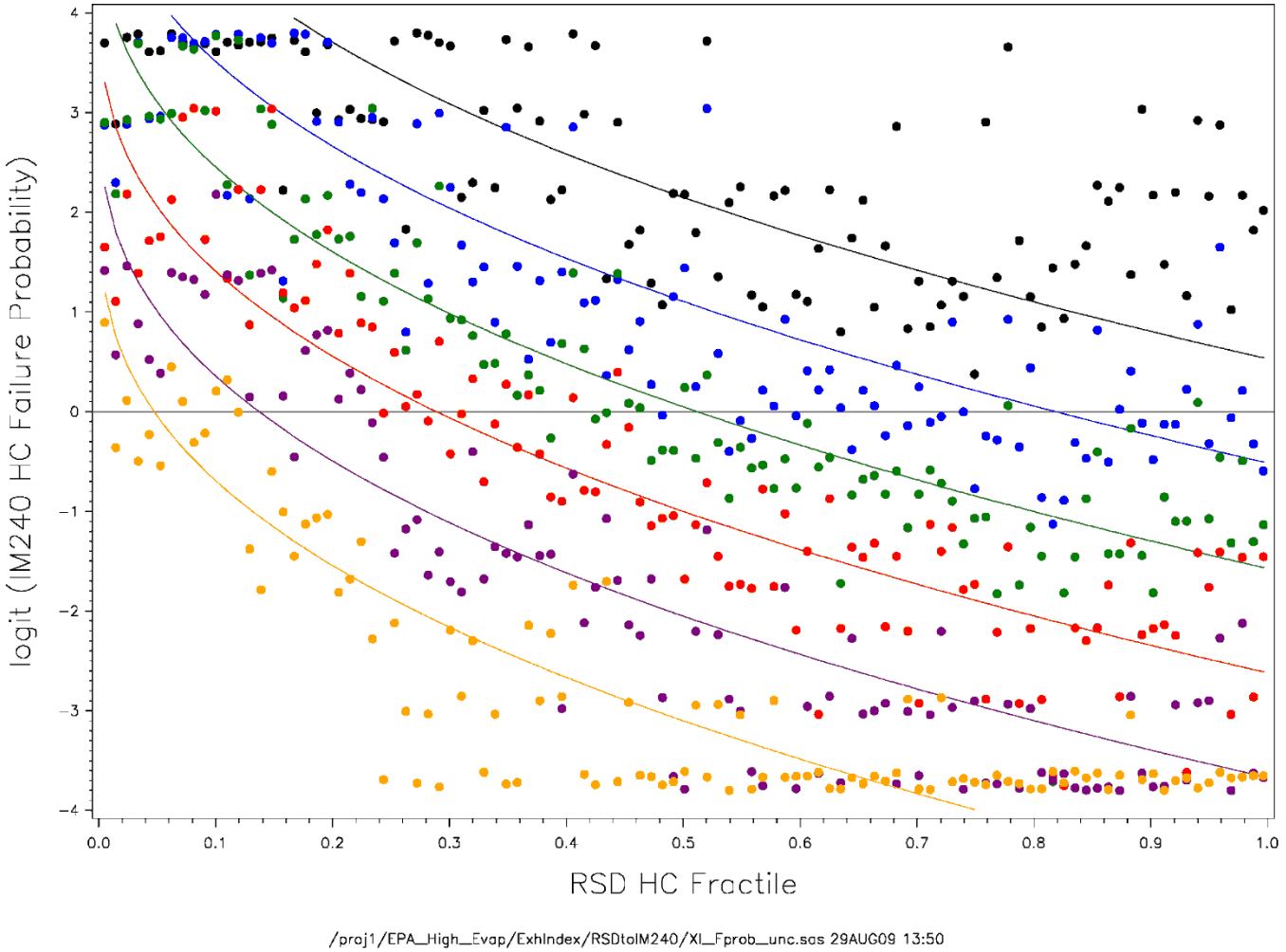
Question: *does it absolutely have to be done this way?*

Answer: *No. options can be considered.*

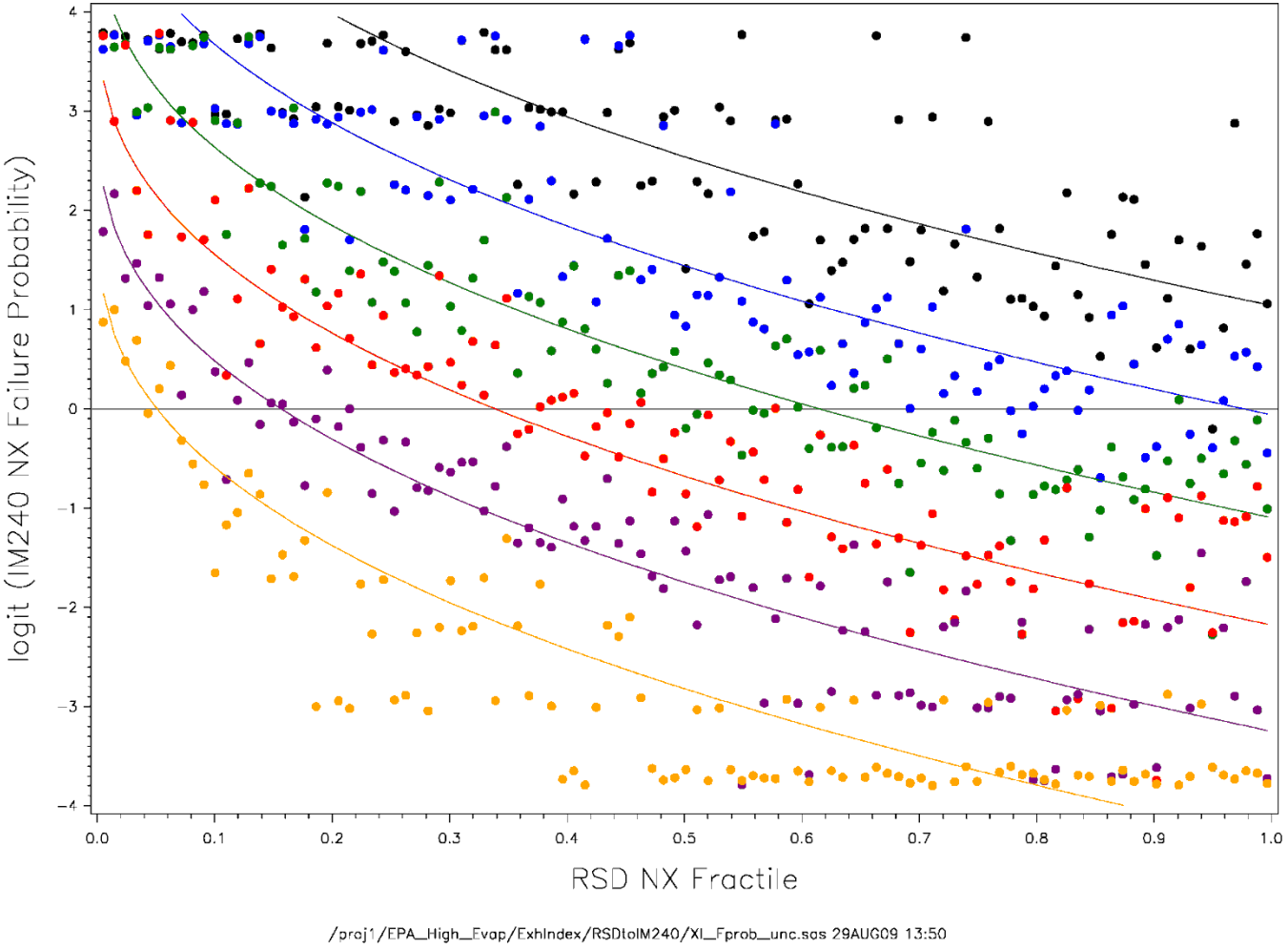
$$f_{\mathrm{NOx}} = \frac{r_{\mathrm{NOx}}}{n}$$

NOx (g/mi)	rank (*r*)	*fractile (f)*
0.1000	1	0.001
0.0800	2	0.002
0.0750	3	0.003
0.5600	4	
...	...	
...	...	
...	...	
0.0056	998	0.998
0.0045	999	0.999
0.0030	1,000	1.000

Model Fit to RSD HC and IM240 HC Data



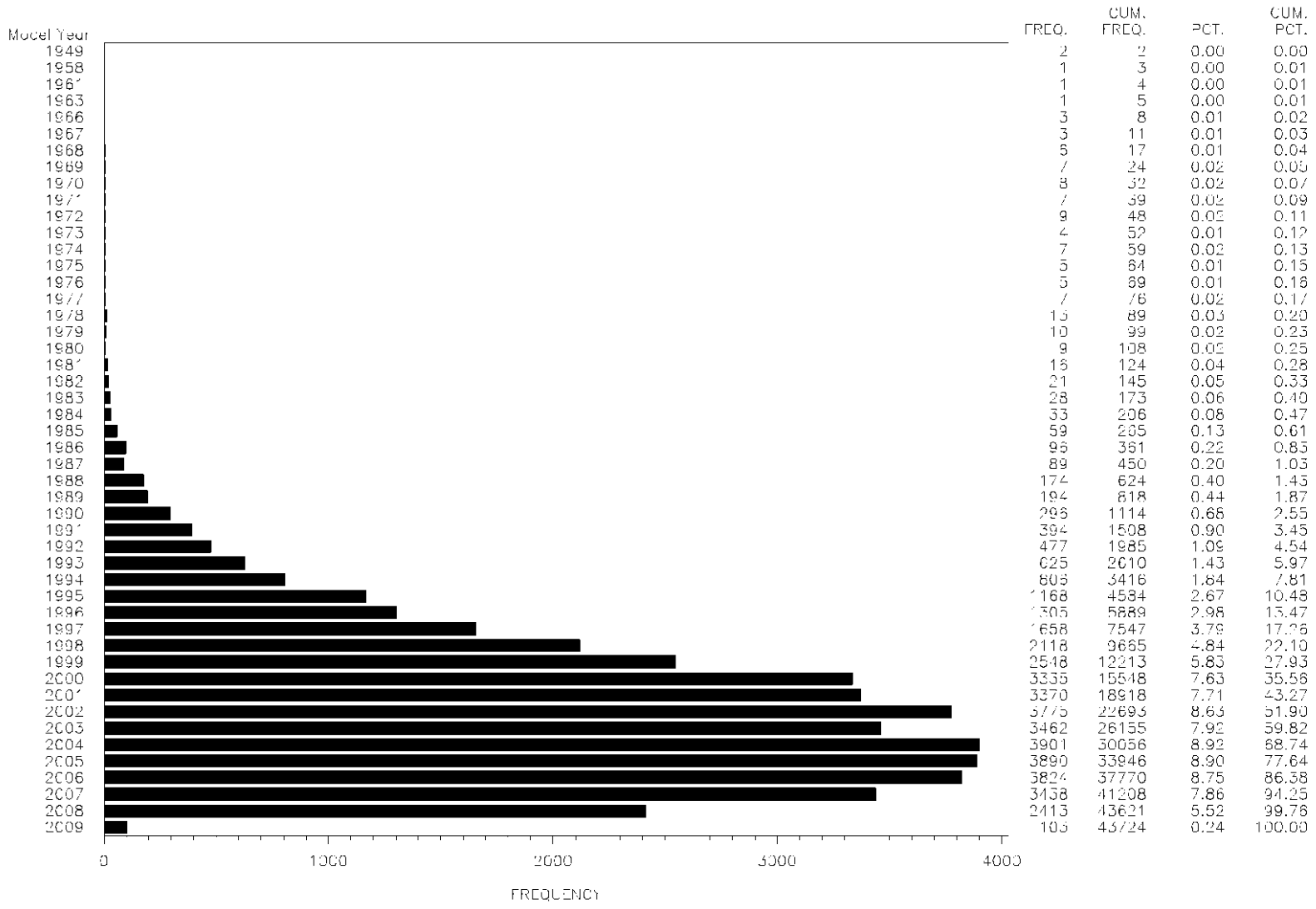
Model Fit to RSD NX and IM240 NX Data



Application Data Set

- Obtained 2 days of On-Road RSD
  - in Denver Area,
  - from about 8 Sites (July 31 and August 4, 2008)
  - contained 43,724 observations
- The models described above were applied to this dataset
- Example thresholds were IM240 “stringent” cutpoints
  - HC: 0.8 g/mile,
  - CO: 15 g/mile
  - NOx: 2.0 g/mile HC, CO, and NOx.

Model Year Distribution of the Application Dataset



Observations in Exceedance Probability Bins

1996 and Newer Model Years

Stringent IM240 HC, CO, NX Cutpoints (0.8, 15, 2 g/mile)

HC, CO, NX PrEx Bin Label	Exceedance Probabilities	Number of Observations in PrEx Bin		
		HC	**CO**	**NX**
2	>82%	83	0	111
1	62 to 82%	145	5	515
0	38 to 62%	574	250	2,428
-1	18 to 38%	4,177	1,339	5,530
-2	8 to 18%	12,888	6,583	9,981
-3	3 to 8%	21,273	13,349	14,600
-4	<3%	0	17,614	5,975
	Total	**39,140**	**39,140**	**39,140**

How do we select?

- Get about equal numbers
  - in each probability bin
- End up with a target sample
  - n ~ 250 vehicles
- Assuming a participation rate of 30%
- Target 830 vehicles for selection

IM240 Exceedance Probabilities of Tested Vehicles
Expected Distribution

HC, CO or NX PrEx Bin Label	Exceedence Probabilities	Number of Observations		
		HC	CO	NX
2	>82%	14	0	11
1	62 to 82%	20	0	28
0	38 to 62%	51	31	59
-1	18 to 38%	75	71	71
-2	8 to 18%	48	65	39
-3	3 to 8%	42	44	28
-4	<3%	0	39	14
	Total	250	250	250

Conclusions

- It is important to understand the distribution of emissions for inventory purposes
- The emissions distributions are highly skewed with the high emitters constituting the long tail
- EPA is attempting to understand the full range of emissions across the vehicle fleet by considering the importance of high emitters