Attachment B

AQS-Based, Hourly PM Characterization Analyses

General / Background:

This attachment describes the characterization analyses of hourly $PM_{2.5}$ data obtained from AQS. It also documents the analyses of hourly $PM_{10-2.5}$ estimates which were derived from the aforementioned $PM_{2.5}$ AQS dataset and a corresponding PM_{10} AQS dataset. Meteorological data from the nearest NWS site was used to convert the PM_{10} data to local temperature and pressure conditions.

Construction of $PM_{2.5}$ database

The database (db) utilized for all hourly $PM_{2.5}$ SP analyses was based on almost *all available hourly AQS $PM_{2.5}$ data*. The following statements document the creation of the db:
- Hourly duration (AQS duration code = '1') data for the time period 2001 to 2003 were polled from AS for parameter $88_{10}1$ ($PM_{2.5}$, local temperature and pressure conditions, LC) on August 24, 2004. [Deleted data with method codes of 740 or 741, per Tim Hanley of Ambient Air Monitoring (AAMG).]
- Data were processed on a monitor basis.
- To be used, a monitor had to meet the completeness goals of at least 75% of hours in a day (18+) at least 75% days in a quarter (68+). The most recent 4, 8, or 12 consecutive quarters that met those goals were utilized. 264 monitors met the completeness criteria: 128 monitors had 4 usable quarters, 72 had 8 usable quarters, and 64 had 12 'complete' quarters. Only data for those monitors, and for the corresponding days with 18+ hours, were kept; data for other monitors and/or days that did not have 18+ samples were discarded.
- SAS code ('raw from AQS.sas') was used to extract the raw data from AQS.

Construction of $PM_{10-2.5}$ database

The db utilized for all hourly $PM_{10-2.5}$ SP analyses was based on the $PM_{2.5}$ db specified above plus corresponding hourly PM_{10} data. A simplistic difference method (PM_{10} - $PM_{2.5}$) was used to generate the $PM_{10-2.5}$ estimates. PM_{10} data were retrieved for both 'local temperature and pressure conditions' (LC) and 'standard temperature and pressure' (STP) conditions. National Weather Service (NWS) data were used to convert the STP data to LC. Since $PM_{2.5}$ and PM_{10} data were then all in LC $\mu g/m^3$ units, resultant estimated $PM_{10-2.5}$ estimates were in the same units. $PM_{10-2.5}$ estimates were generated on site basis. The following statements provide additional detail:
- As noted above, hourly duration data for the time period 2001 to 2003 were polled from AQS for parameter $88_{10}1$ on August 24, 2004. Data for method codes of 740 or 741 were deleted.
- Hourly data for parameters $81_{10}2$ (PM_{10}, STP) and $85_{10}1$ (PM_{10} LC) were also retrieved from AQS on August 24, 2004.

- Raw NWS hourly data for 2001-2003 were obtained from Bill Cox of Air Quality Modeling Group (AQMG) on March 19, 2004. Utilized fields were relative humidity (RH), barometric pressure (BP), and temperature (T).
- PM_{10} STP data were converted to LC by using the corresponding (same date, same hour) met data from the nearest NWS site.
- Multiple site-date-hour measurements (from collocated monitors) of PM_{10} and/or $PM_{2.5}$ were averaged (independently)
- Hourly $PM_{10-2.5}$ was estimated by subtracting the $PM_{2.5}$ concentration from the PM_{10} (LC based) concentration.
- To be used, a site had to meet the completeness goals of at least 75% of hours in a day (18+) at least 75% days in a quarter (68+). The most recent 4, 8, or 12 consecutive quarters that met those goals were utilized. 31 sites met the completeness criteria: 14 sites had 12 usable quarters, 14 sites had 8 complete quarters, and 3 had all 12 quarters complete. Only data for those sites, and for the corresponding days with 18+ hours, were kept; data for other sites or days that did not have 18+ samples were discarded.
- SAS code was used to extract the raw data from AQS ('raw from AQS.sas') and to convert PM_{10} STP data to LC ('calc hourly coarse.sas').

Analysis 1 – Hourly versus 24-hour, $PM_{2.5}$ and $PM_{10-2.5}$

Goals:
- ? To determine how well correlated is the hourly daily maximum with the 24-hr average?
- ? How well do/would daily and annual standards control hourly peaks?
- ? How do the 1-hr distributions compare to the 24-hr distributions

Outputs:
- o Various tables and box-plots were generated for $PM_{2.5}$; Summary statistics were generated; see Output B.1a.
- o Various tables and box-plots were generated for $PM_{10-2.5}$; Summary statistics were generated; see Output B.1b.

Methods:
- 24-hour data were calculated from the hourly data and pseudo DV's (annual and 98th percentile) were constructed from the hourly-based daily averages. This technique was utilized instead of matching to collocated FRM data for 2 reasons: 1) To avoid sampler bias that would have resulted from comparing the continuous to the filter-based FRM measurements, and 2) To maximize the number of observations: FRM instruments may not have been collocated with the continuous monitors and even if they were, the FRM might have only sampled every 3rd day or every 6th day.
- SAS code was used for all of the analyses: 'correlations_maxvmean3.sas', 'correlations_maxvmean3_pmc.sas', 'hourly v 24 pmf.sas', 'hourly v 24 pmc.sas', 'hourly peak to mean pmf.sas', and 'hourly peak to mean pmc.sas'.

Analysis 2 – Diurnal distributions, seasonal plots, and episodic events of hourly measurements of $PM_{2.5}$ and $PM_{10-2.5}$ concentrations, 2001-2003.

Goals:

  ? To characterize and contrast short-term (diurnal) patterns of $PM_{2.5}$ and $PM_{10-2.5}$.
  ? To characterize differences in seasonal diurnal patterns.
  ? To investigate (and contrast) the effect of episodic events on hourly $PM_{2.5}$ and $PM_{10-2.5}$.

Outputs:

  o Diurnal boxplots, representing 4, 8, or 12 quarters of 2001 to 2003, were generated for every hourly $PM_{2.5}$ monitor and for every hourly $PM_{2.5}$ site. Regional aggregation plots were also included. Two 'example' sites were selected from the pools, one to basically represent 'eastern' sites and the other, to generally depict 'western' sites.
  o Seasonal line-plots were created for the two selected sites.
  o The effect of an episodic event on $PM_{2.5}$ and $PM_{10-2.5}$ concentrations over a 2-day period were plotted for the 'western' location.
  o All plot (except for the universe pools) are shown in output B.2.

Methods:

  • The data hourly data were adjusted for daylight savings time (both size cuts). SAS code ('hour_boxplot20012003_daylight_savings.sas', 'seasonal hour avg line plots.sas', and 'elpaso_gso_20012003.sas'), was used to make the adjustment and generate all of the plots.

Analysis 3 – Evaluation of hour-to-hour changes (increases) in $PM_{2.5}$

Goals:

  ? To characterize typical (median) monitor-level hour-to-hour increases in $PM_{2.5}$

Outputs:

  o See SAS output screen capture in Output B.3

Methods:

  • Only 'increases' from one hour to the next were evaluated. SAS code ('hour difference distribution.sas') was used for the evaluation.

Hourly vs. 24-hr – PM2.5

- Questions:
  1. How well correlated is the hourly daily max with the 24-hr average?
  2. How well do the daily and annual stds control hourly peaks?
  3. How do the 1-hr distributions compare to the 24-hr distributions

- Analyses details:
  1. Hourly data from AQS. TEOM, BAM, whatever. May or may not be 'adjusted' to be more FRM like.
  2. Only used sites that met completeness criteria of 75% hours in a day; 75% days in a quarter; most recent 4, 8, or 12 consecutive quarters. 264 sites met criteria (64 had 12 Q's, 72 had 8 Q's, and 128 had 4 Q's.) Only used data for those sites… days w/ 75%+
  3. 24-hr data calculated from hourly data.
  4. Pseudo DV's (annual and 98th percentile) constructed from hourly daily averages. … Instead of matching to collocated FRM DV. ….Rationale – Avoid sampler bias, continuous vs. filter

1. How well correlated is the hourly daily max with the 24-hr avg?

HEI region	# sites	Site Correlation			
		mean	median	minimum	maximum
National	264	0.82	0.84	0.53	0.95
Industrial Midwest	41	0.80	0.85	0.55	0.92
Northeast	51	0.89	0.90	0.72	0.95
Northwest	57	0.84	0.85	0.66	0.93
Southeast	65	0.78	0.80	0.53	0.91
Southern California	5	0.82	0.80	0.77	0.92
Southwest	26	0.81	0.83	0.71	0.94
Upper Midwest	17	0.74	0.76	0.56	0.87
Not in PMREG Reg	2	0.80	0.80	0.77	0.83

- Good correlation; consistent across geographic regions.

2a. How well does an annual standard control hourly peaks?

Distribution of daily 1-hour maxes vs. annual mean PM2.5 concentrations, 2001-2003

Whiskers=5th,95th
Box=25th,75th
Line=Median

Daily 1-hour Max (ug/m3)

Annual Mean (ug/m3)	<=6	6-8	8-9	9-10	10-11	11-12	12-13	13-14	14-15	15-16	16-17	>17
N daily 1-hr maxes	9098	15190	19117	19031	17644	17294	15816	8445	15322	12231	4549	9760
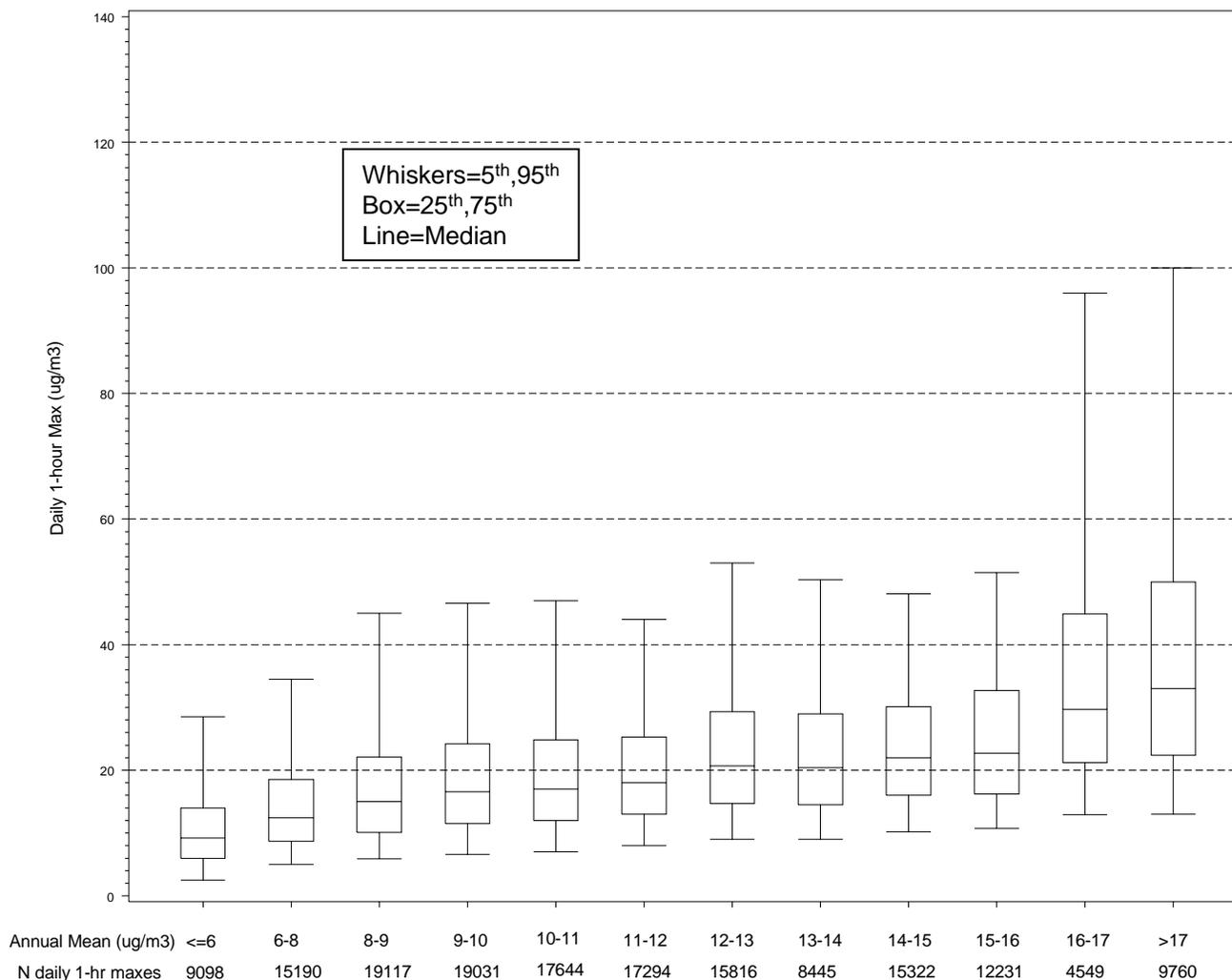
- More than 95% of daily max 1-hr's are $\leq$ 50 ug/m3 when annual DV $\leq$ 16
- [The 95th percentile (daily max 1-hr) is $\leq$ 50 for *most* of the annual mean intervals $\leq$ 16]

2b. How well does a daily standard control hourly peaks?

Distribution of daily 1-hour max's vs. 98th percentile 24-hour average PM2.5 concentrations, 2001-2003


Whiskers=5th,95th
Box=25th,75th
Line=Median

98th Percentile (ug/m3)	<=20	20-25	25-30	30-35	35-40	40-45	45-50	50-55	55-60	>65
N daily 1-hr maxes	19667	24520	37329	33182	26043	14417	3115	709	2442	2073
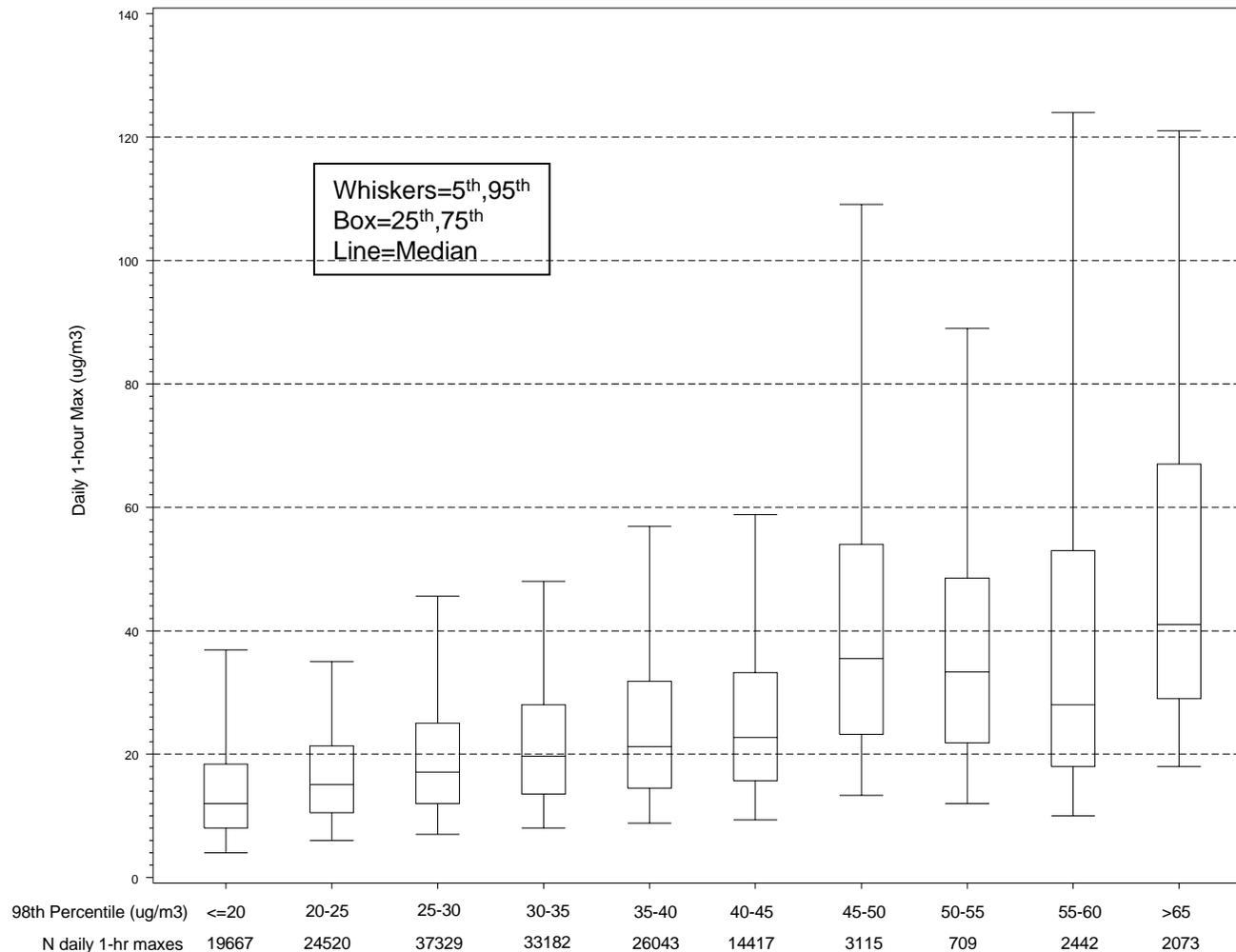
- More than 95% of daily max 1-hr's are $\leq$ 50 ug/m3 when daily DV (98th percentile) $\leq$ 45
- [The 95th percentile (daily max 1-hr) is $\leq$ 50 for *most* of the 98th percentile intervals $\leq$ 45]
- [The 95th percentile (daily max 1-hr) is $\leq$ 60 for *all* of the 98th percentile intervals $\leq$ 45]

3a. How does the max 1-hr distribution compare to the 24-hr distribution?
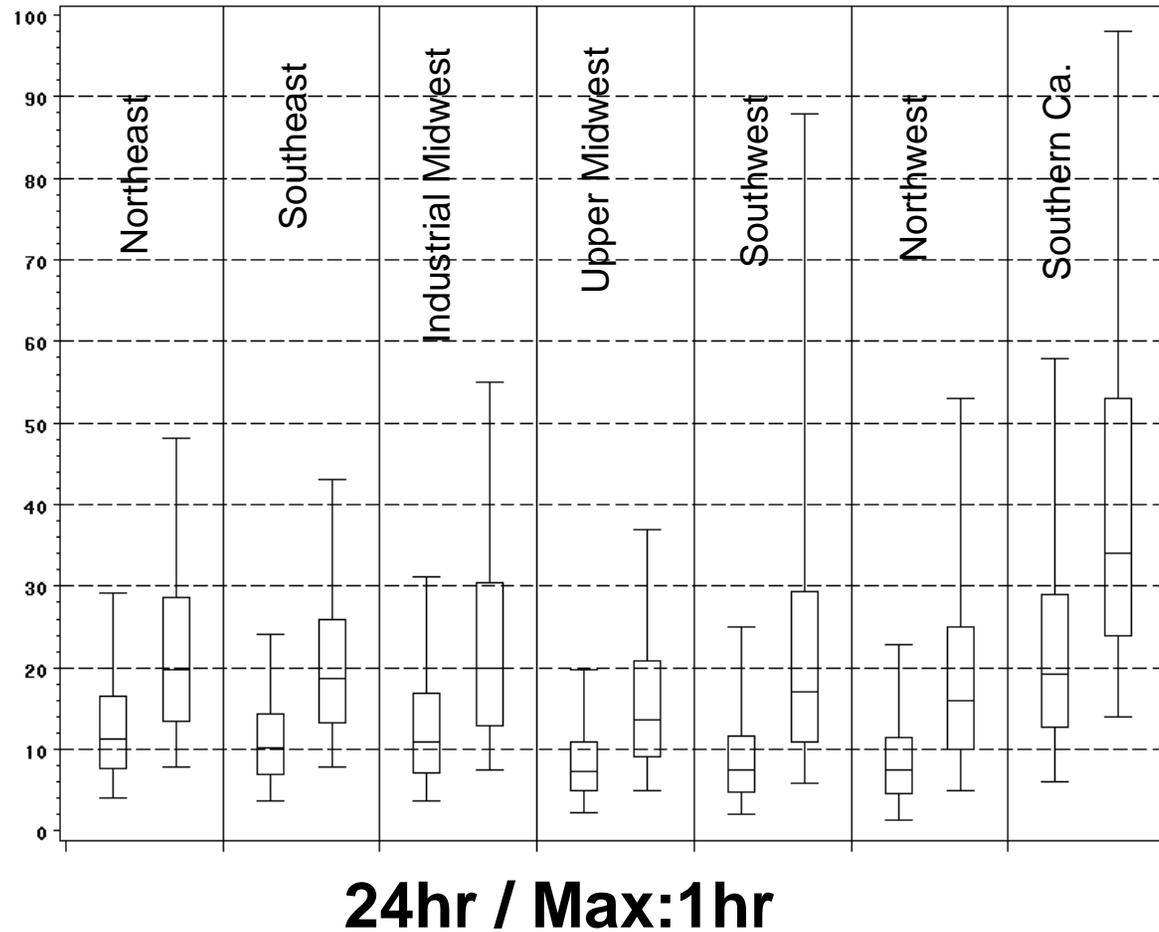
24hr / Max:1hr

3b. How does the 1-hr distribution (all hrs) compare to the 24-hr distribution?

24hr / 1hr
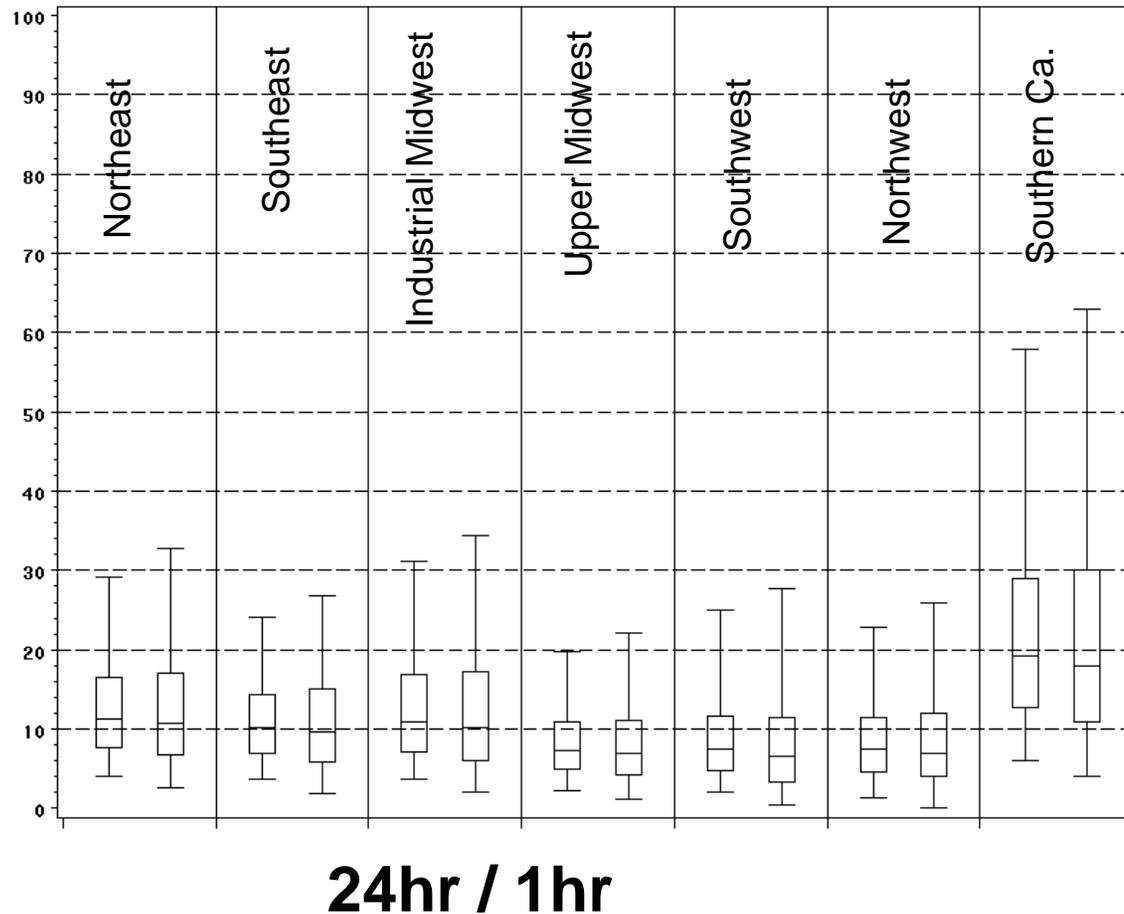
3c. How does the peak-to-mean ratio (max 1hr / 24hr avg) compare by region?
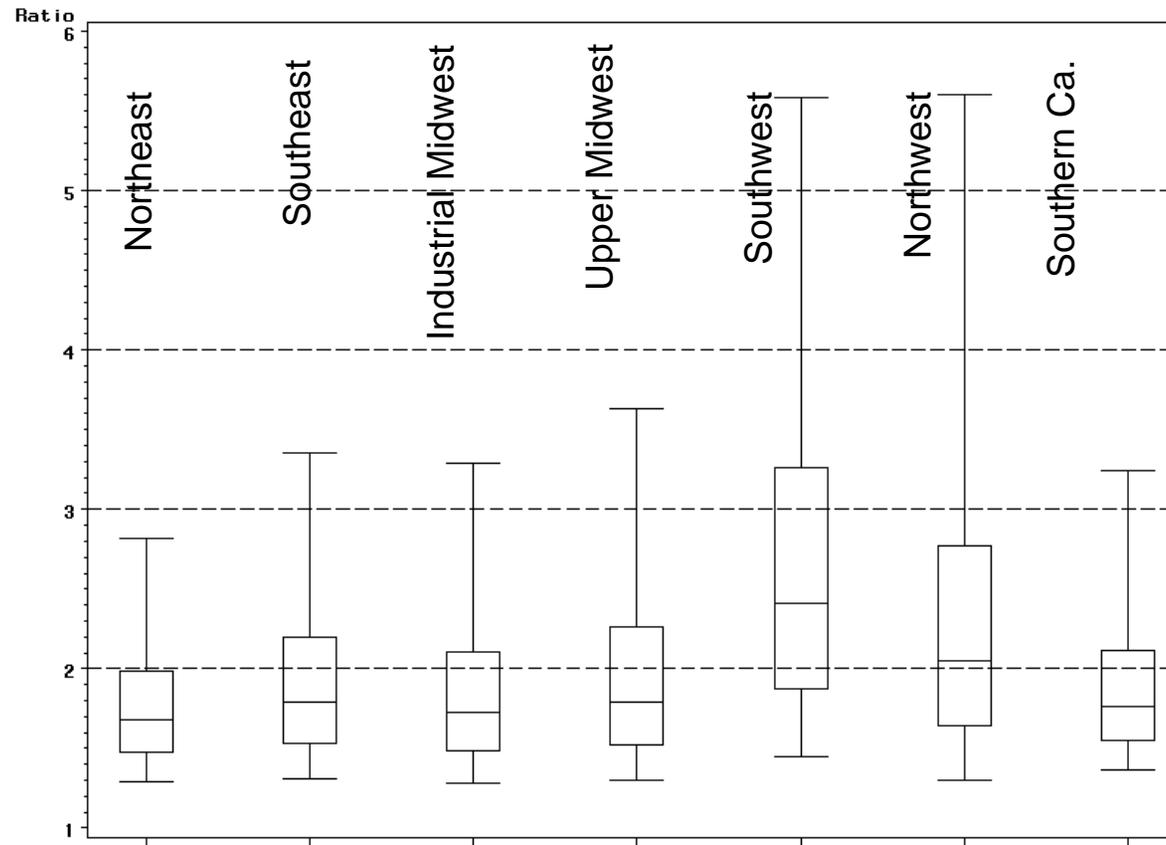
Ratio: max_1hr / 24hr_avg

Hourly vs. 24-hr – PM10-2.5

- Questions:
  1. How well correlated is the hourly daily max with the 24-hr average?
  2. How well would daily and annual standards control hourly peaks?
  3. How do the 1-hr distributions compare to the 24-hr distributions?

- Analyses details:
  1. Hourly data constructed by difference method from (AQS) collocated continuous PM10 and PM2.5.
  2. Only used sites that met completeness criteria of 75% hours in a day; 75% days in a quarter; most recent 4, 8, or 12 consecutive quarters. 31 sites met criteria (3 had 12 Q's, 14 had 8 Q's, and 14 had 4 Q's.) Only used data for those sites… days w/ 75%+
  3. 24-hr data calculated from hourly data.
  4. Constructed psuedo DV's (annual and 98th percentile) from hourly daily averages.

1. How well correlated is the hourly daily max with the 24-hr avg?

HEI Region	# sites	Site Correlation			
		mean	median	minimum	maximum
National	31	0.80	0.81	0.67	0.91
Industrial Midwest	9	0.81	0.83	0.75	0.86
Northeast	3	0.78	0.81	0.67	0.85
Northwest	5	0.77	0.77	0.69	0.88
Southeast	6	0.79	0.78	0.70	0.91
Southwest	5	0.84	0.85	0.78	0.91
Upper Midwest	3	0.83	0.84	0.80	0.86

No data (meeting completeness) for Southern California

•Good correlation; consistent across geographic regions.

2a. How well would a daily standard control hourly peaks?

Distribution of daily 1-hour maxes vs. 98th percentile 24-hour average PM10-2.5 concentrations, 2001-2003
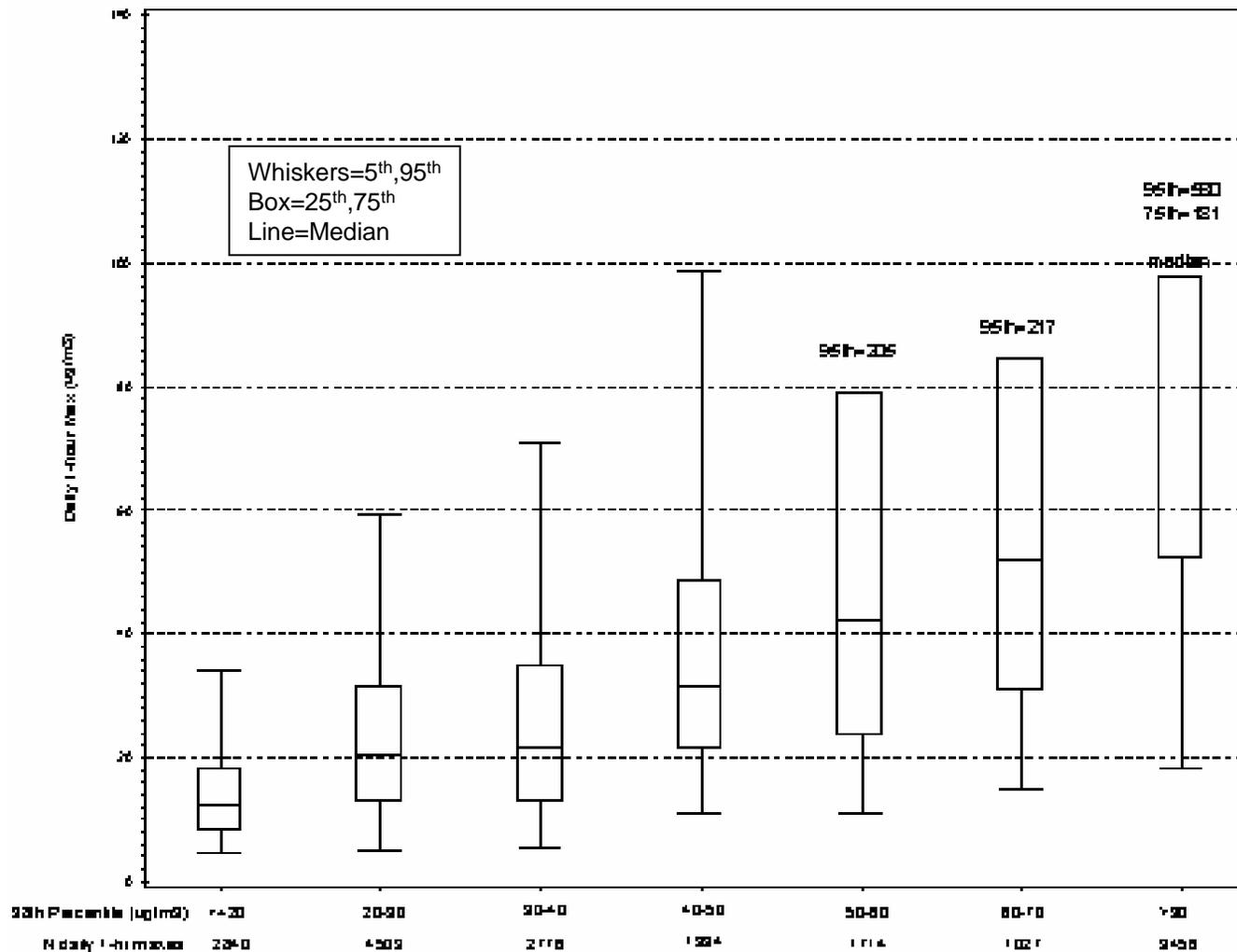
•A daily PM10-2.5 standard would appear not to control hourly peaks unless set on the low end (of the intervals shown here)

2b. How well would an annual standard control hourly peaks?

Distribution of daily 1-hour maxes vs. annual mean PM10-2.5 concentrations, 2001-2003
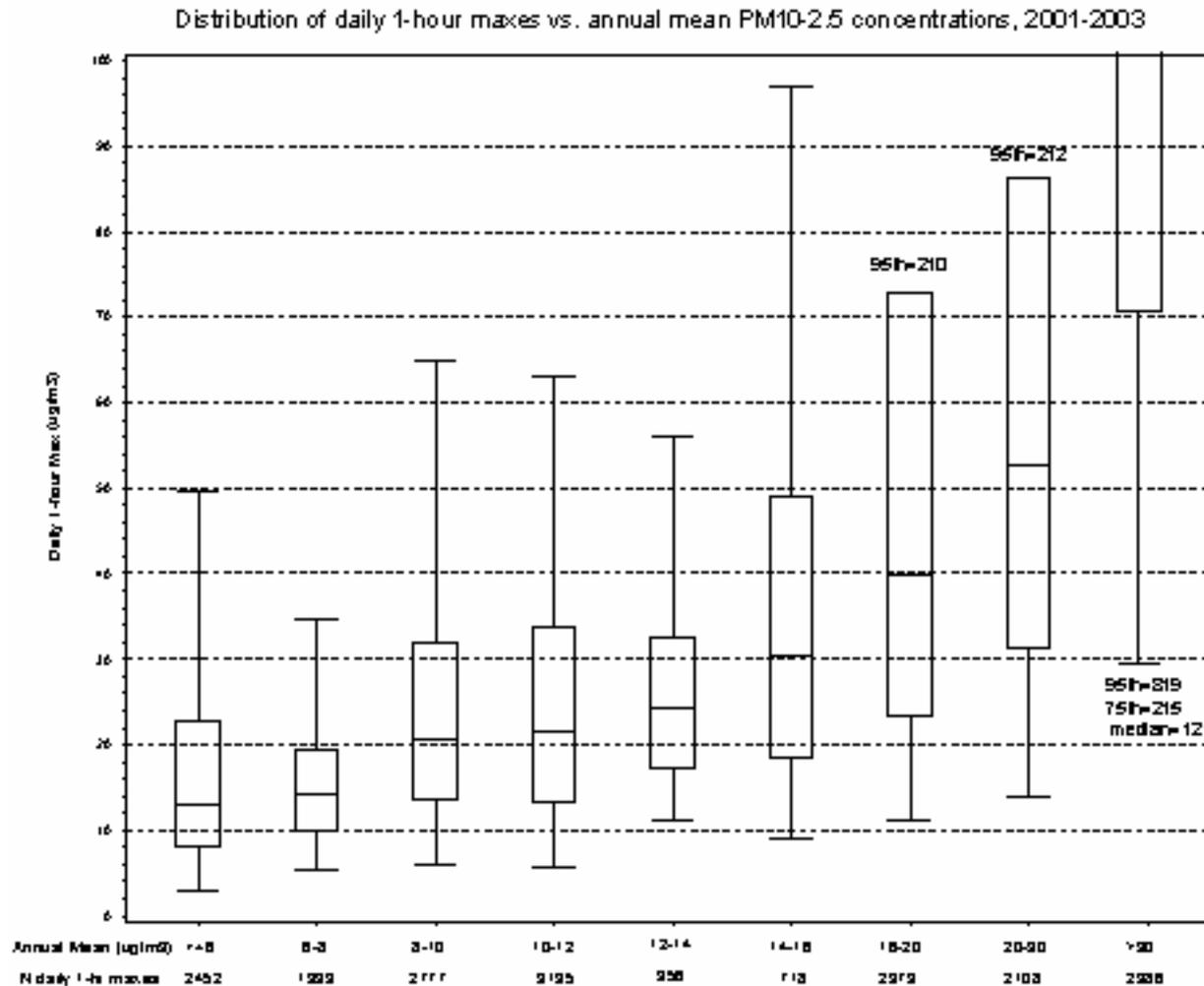
•An annual PM10-2.5 standard would appear not to control hourly peaks unless set on the low end (of the intervals shown here)

3a. How does the max 1-hr distribution compare to the 24-hr distribution?
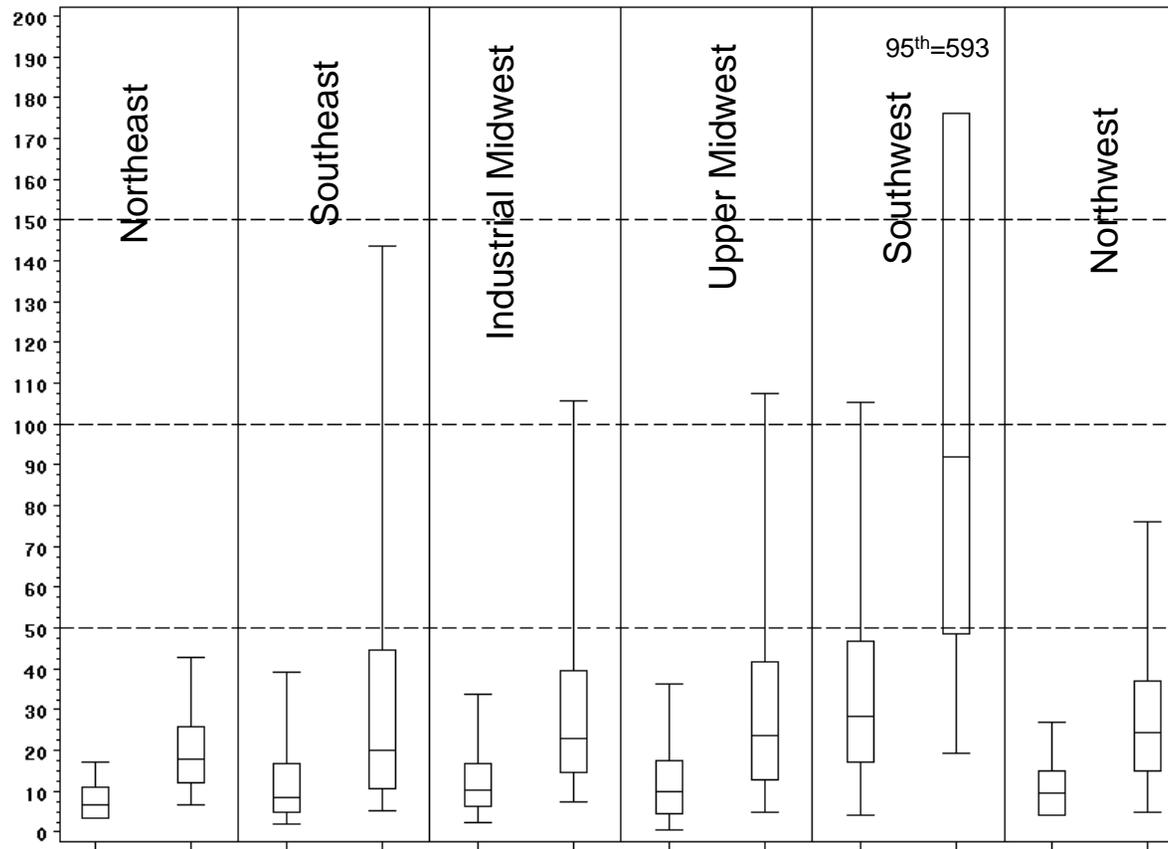
24hr / Max:1hr

3b. How does the 1-hr distribution (all hrs) compare to the 24-hr distribution?
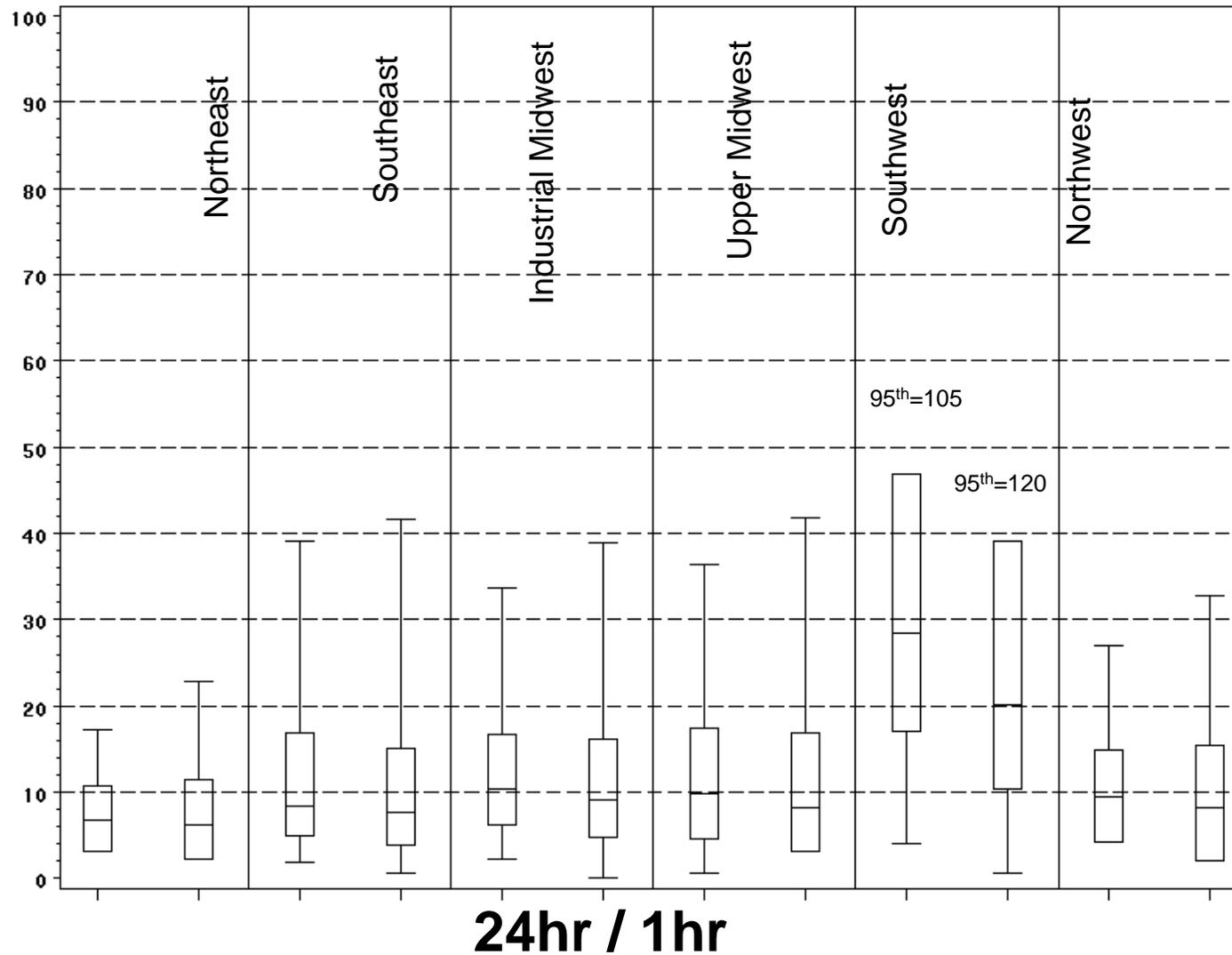
3c. How does the peak-to-mean ratio (max 1hr / 24hr avg) compare by region?
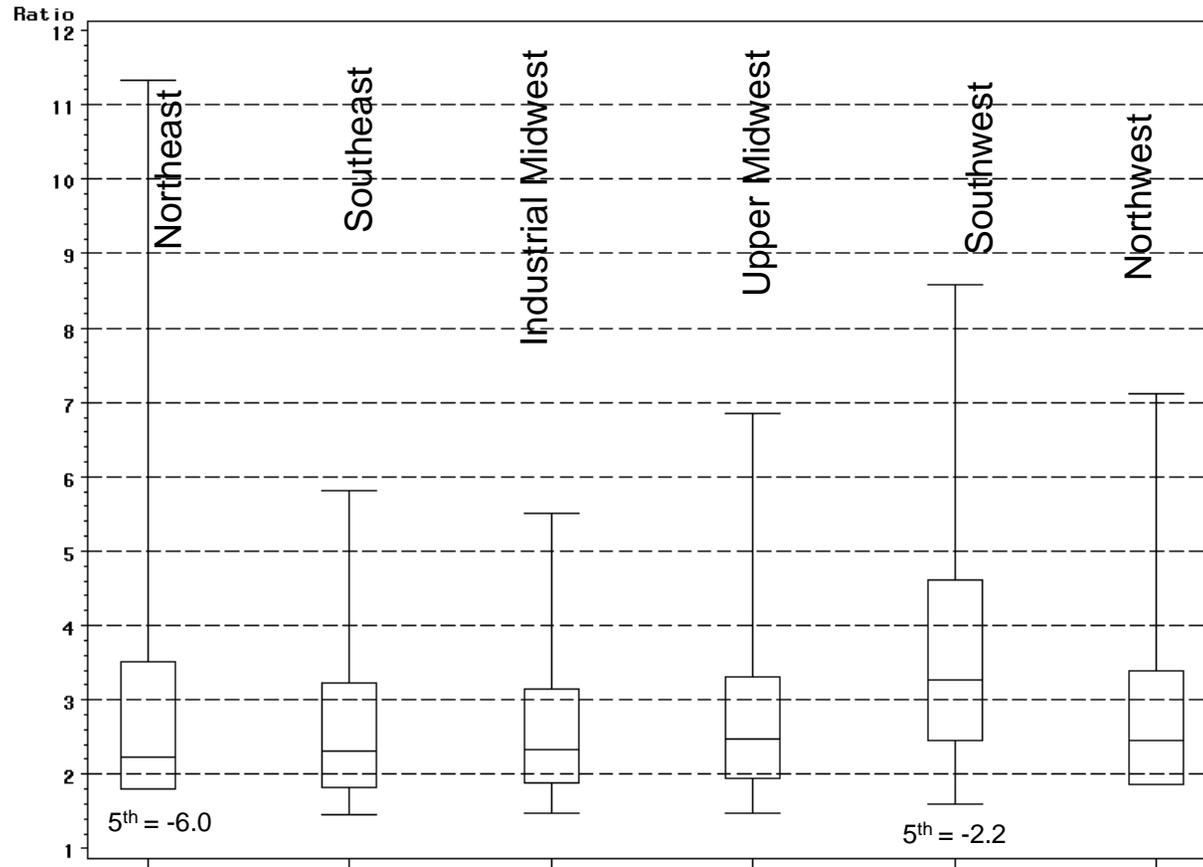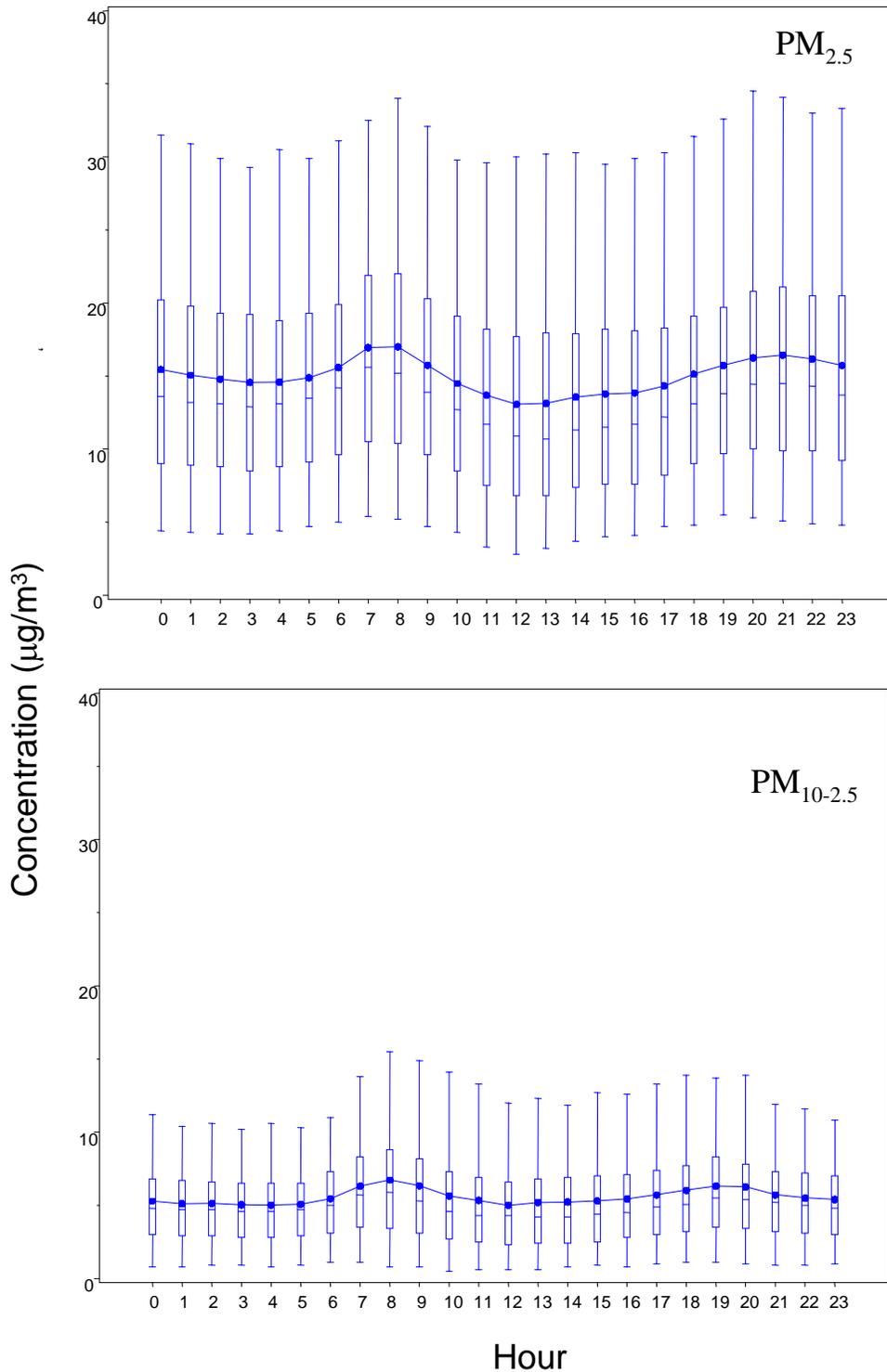
Ratio: max_1hr / 24hr_avg

Hourly average PM$_{2.5}$ and PM$_{10\text{-}2.5}$ concentrations at a Greensboro, NC monitoring site, 2001-2003. Upper panel shows the distribution of PM$_{2.5}$ concentrations and the lower panel shows the distribution of PM$_{10\text{-}2.5}$ concentrations (box plot of interquartile range, mean, median, 5th and 95th percentiles)
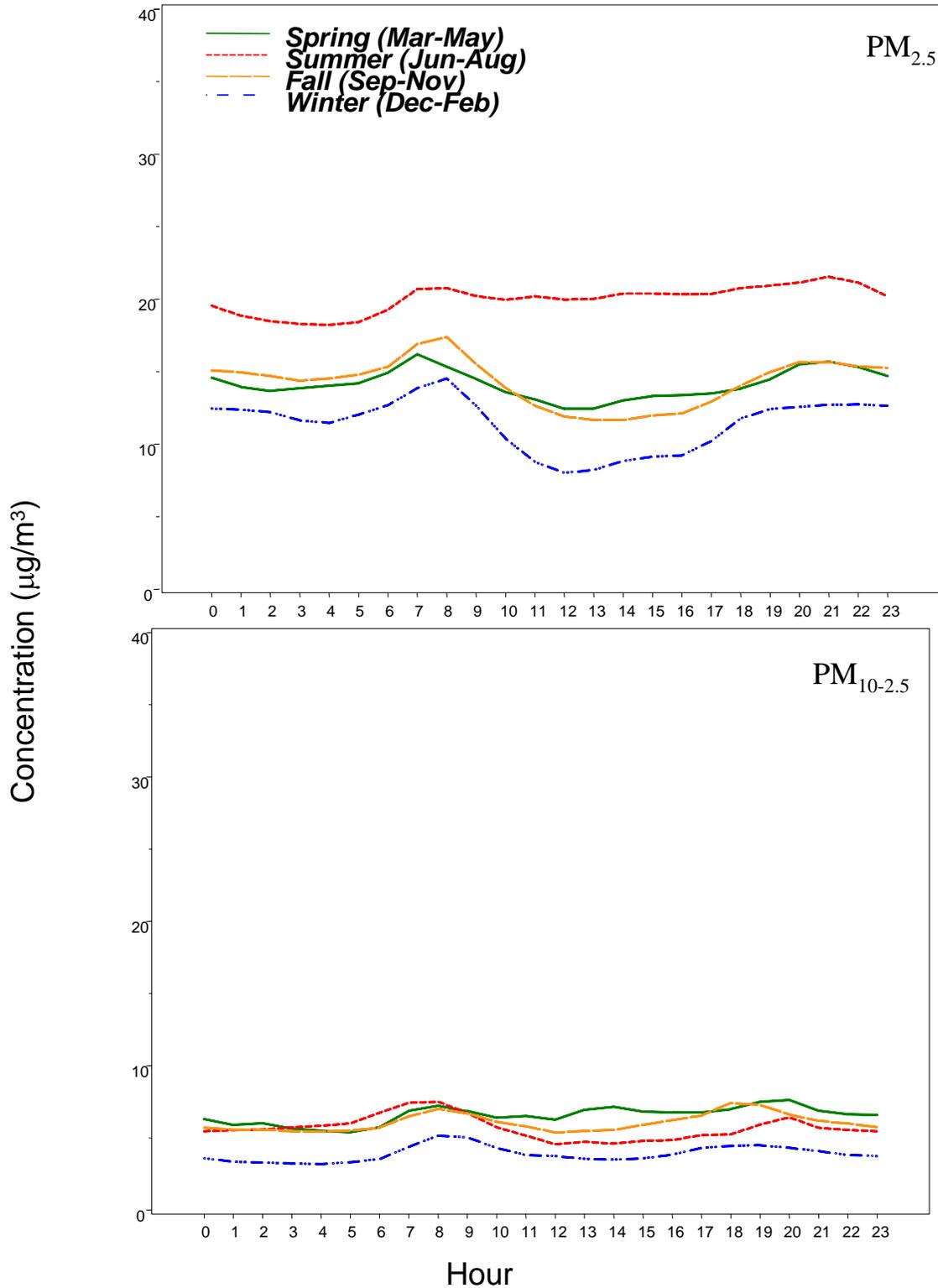
Seasonal hourly average PM$_{2.5}$ and PM$_{10-2.5}$ concentrations at a Greensboro, NC monitoring site, 2001-2003. Upper panel shows the PM$_{2.5}$ concentrations and the lower panel shows the PM$_{10-2.5}$ concentrations.
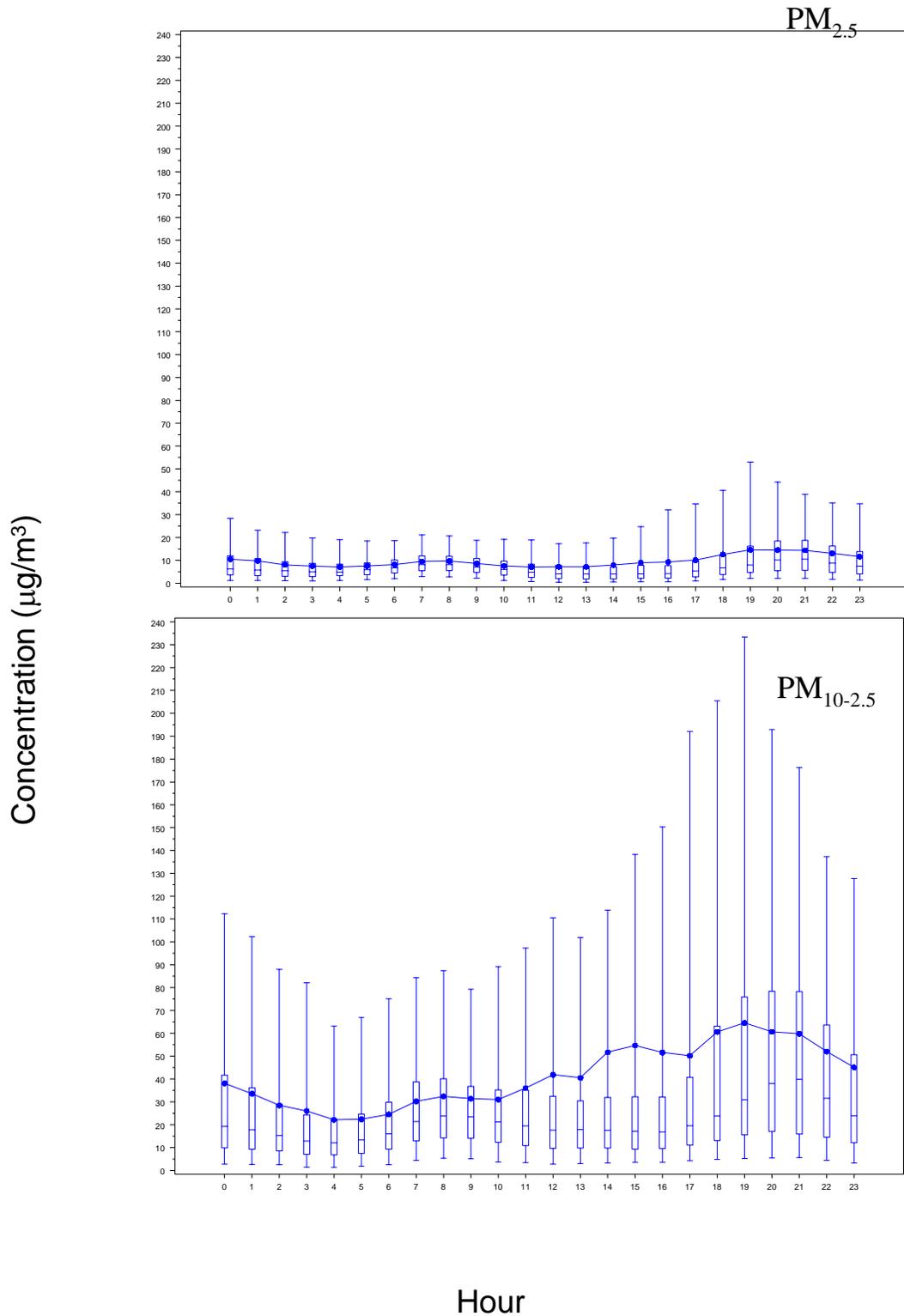
Hourly average PM$_{2.5}$ and PM$_{10-2.5}$ concentrations at an El Paso. TX monitoring site, 2001-2003. Upper panel shows the distribution of PM$_{2.5}$ concentrations and the lower panel shows the distribution of PM$_{10-2.5}$ concentrations (box plot of interquartile range, mean, median, 5th and 95th percentiles)
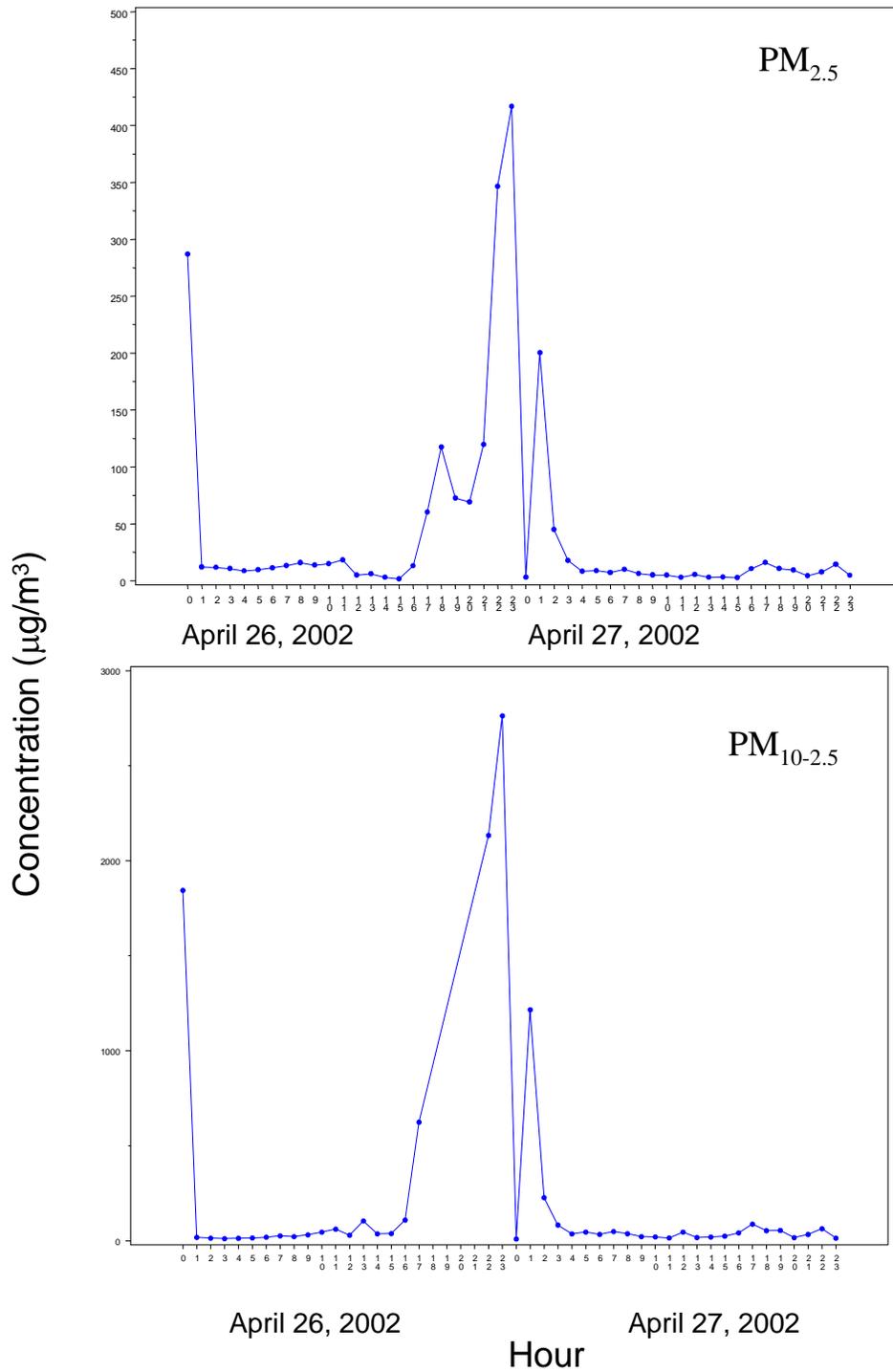
Hourly PM$_{2.5}$ and PM$_{10-2.5}$ concentrations at a El Paso, TX monitoring site for April 26, 2002-April 27, 2002. Upper panel shows the hourly PM$_{2.5}$ concentrations and the lower panel shows the hourly PM$_{10-2.5}$ concentrations.

```
                          The UNIVARIATE Procedure
                       Variable:  median  (the median, diff)


                                  Moments

N                             264    Sum Weights                   264
Mean                   1.81780303    Sum Observations            479.9
Std Deviation          0.99451218    Variance               0.98905447
Skewness                2.0698422    Kurtosis               4.43380155
Uncorrected SS           1132.485    Corrected SS           260.121326
Coeff Variation        54.7095676    Std Error Mean         0.06120799



                        Basic Statistical Measures

              Location                       Variability

          Mean     1.817803     Std Deviation          0.99451
          Median   1.500000     Variance               0.98905
          Mode     2.000000     Range                  5.50000
                                 Interquartile Range    0.80000



                        Tests for Location: Mu0=0

           Test             -Statistic-      -----p Value------

           Student's t    t  29.69878     Pr > |t|     <.0001
           Sign           M       132     Pr >= |M|    <.0001
           Signed Rank    S     17490     Pr >= |S|    <.0001



                         Quantiles (Definition 5)

                         Quantile        Estimate

                         100% Max           6.0
                         99%                5.5
                         95%                4.0
                         90%                3.0
                         75% Q3             2.0
                         50% Median         1.5
                         25% Q1             1.2
                         10%                1.0
                         5%                 1.0
                         1%                 0.7
                         0% Min             0.5
```

```
                        The UNIVARIATE Procedure
                  Variable:  median  (the median, diff)


                        Extreme Observations


                  ----Lowest----         ----Highest---

                  Value      Obs         Value       Obs

                    0.5      228           5.3         36
                    0.6      252           5.3         40
                    0.7      224           5.5         48
                    0.8      116           5.7         47
                    0.8      144           6.0         14
```

```
                          The UNIVARIATE Procedure
                    Variable:  pct95  (the 95th percentile, diff)

                                  Moments

     N                         264      Sum Weights                  264
     Mean                8.46041667     Sum Observations         2233.55
     Std Deviation       4.47284534     Variance              20.0063455
     Skewness            2.91291781     Kurtosis              13.3914067
     Uncorrected SS      24158.4325     Corrected SS          5261.66885
     Coeff Variation     52.8679085     Std Error Mean         0.2752846


                          Basic Statistical Measures

           Location                        Variability

      Mean      8.460417      Std Deviation             4.47285
      Median    7.000000      Variance                 20.00635
      Mode      5.000000      Range                    38.00000
                              Interquartile Range       4.40000


                        Tests for Location: Mu0=0

         Test              -Statistic-      -----p Value------

         Student's t    t  30.73335      Pr > |t|     <.0001
         Sign           M       132      Pr >= |M|    <.0001
         Signed Rank    S     17490      Pr >= |S|    <.0001


                         Quantiles (Definition 5)

                         Quantile      Estimate

                         100% Max         41.0
                         99%              27.0
                         95%              16.0
                         90%              13.5
                         75% Q3           10.0
                         50% Median        7.0
                         25% Q1            5.6
                         10%               5.0
                         5%                4.8
                         1%                4.0
                         0% Min            3.0
```

```
                    The UNIVARIATE Procedure
           Variable:  pct95  (the 95th percentile, diff)


                      Extreme Observations

            ----Lowest----          ----Highest---

            Value      Obs          Value      Obs

             3.0       226          23.0       161
             3.9       116          26.6       106
             4.0       257          27.0        13
             4.0       256          28.4       189
             4.0       232          41.0        14
```