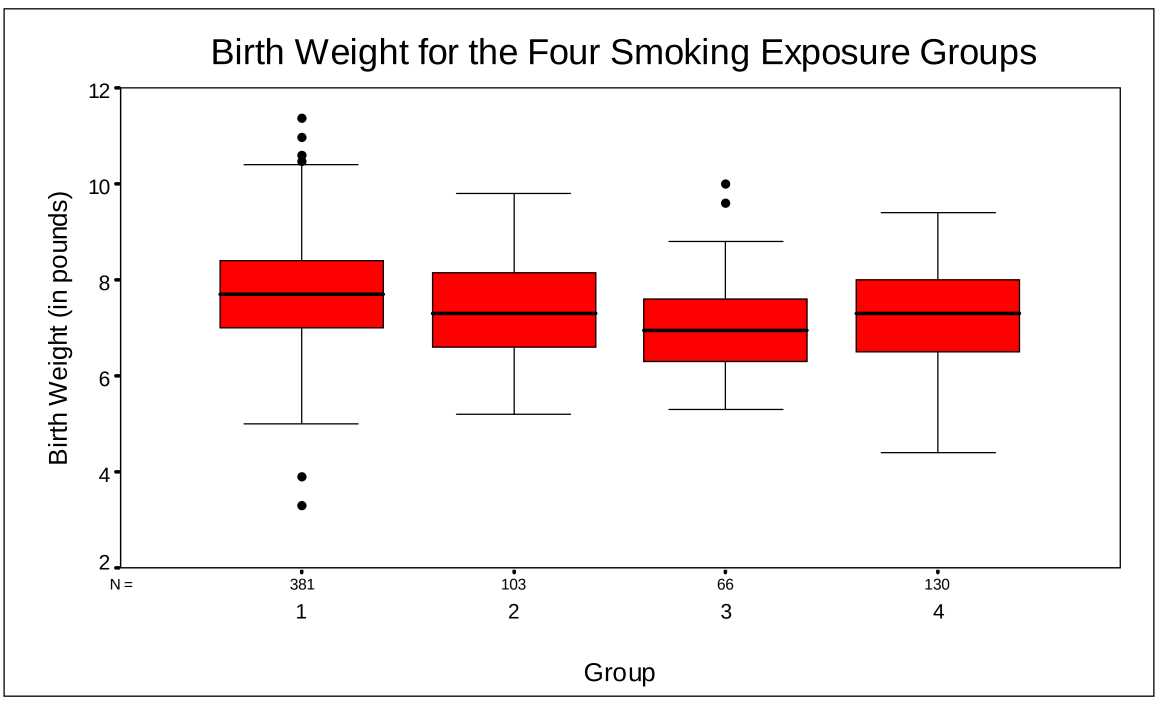
11. Assessing the Effect of Smoking with ANOVA

In this section we will use the child health and development data to examine the effect of maternal smoking with ANOVA. The effects from maternal smoking on birth weight can be investigated by comparing the mean birth weight of infants classified by maternal smoking exposure.

The maternal smoking is described by the variable MNOCIG that expresses the information about the number of cigarettes smoked per day by a mother. The SPSS output reveals that the possible values of the variable are 0, 2, 7, 12, 17, 25, 35, and 50 with the corresponding frequencies: 381, 43, 60, 42, 24, 105, 16, and 9. Therefore, it is reasonable to define the following smoking exposure classes: 0 (non-smoker), 2-7 (occasional smoker), 12-17 (less than pack/day), 25-50 (more than pack/day).

GROUP	MNOCIG	SIZE	MEAN BIRTH WEIGHT	STANDARD DEVIATION
1	0 (Non-smoker)	381	7.7328	1.0523
2	2-7 (Occasional smoker)	103	7.3806	1.1329
3	12-17 (Less than pack/day)	66	6.9727	0.9408
4	25-50 (More than pack/day)	130	7.2662	1.0909

The side-by-side boxplots of birth weight for the four smoking exposure groups are displayed below:



As you can see, the birth weights follow approximately a symmetric distribution with similar spread in each of the four groups. The assumption of normality and equal variances for the four groups is not likely to be seriously violated.

We would like to know whether there are significant differences in birth weight for the four smoking exposure groups. An appropriate statistical technique to examine the differences is one-way ANOVA. The purpose of ANOVA is to assess whether the observed differences among the five groups are statistically significant.

The following is an SPSS output of a one-way analysis of variance for the four smoking groups:

Analysis of Variance

Source	**D.F.**	**Sum of Squares**	**Mean Squares**	**F Ratio**	**F Prob.**
Between Groups	3	47.3925	15.7975	13.9998	.0000
Within Groups	676	762.8030	1.1284		
Total	679	810.1955			

The analysis of variance F-statistic is F=13.9998, with 3 and 676 degrees of freedom, giving a p-value of less than 0.0001. That small p-value indicates strong evidence against the null hypothesis of no difference among the average birth weights for the five groups. In other words, there is strong evidence of differences among the group means. The within-group mean square is 1.1284, so the pooled estimate of a common standard deviation is the square root of the value, which is equal to 1.0623 pounds

According to the ANOVA assumptions, the groups should come from treatments with equal variances. The above side-by-side boxplots indicate that this assumption is not likely to be violated in this case. Moreover, the Levene homogeneity-of-variance test produces the p-value of 0.173.

Levene Test for Homogeneity of Variances

Statistic	df1	df2	2-tail Sig.
1.6645	3	676	.173

That high p-value shows very weak evidence that the assumption of equal variances might be violated.

It is also worthy to determine which group means differ. In order to find some significant differences between the group means, we use the Tukey multiple comparisons test. The SPSS procedure provides the following output:

```
Multiple Range Tests:  Tukey-HSD test with significance level .050

The difference between two means is significant if
MEAN(J)-MEAN(I)  >= .7511 * RANGE * SQRT(1/N(I) + 1/N(J))
with the following value(s) for RANGE: 3.65

 (*) Indicates significant differences which are shown in the lower triangle

                    Group
                    3 4 2 1
  Mean    SMOK

  6.9727  Group 3
  7.2662  Group 4
  7.3806  Group 2
  7.7328  Group 1   * * *
```

As you can see, the test shows that there are significant differences in birth weights between the first group (non-smokers) and each of the three remaining groups (smokers).

What we have found above can be stated as follows: there is strong evidence of differences between birth weights of children born to smokers and non-smokers. However, we cannot make any causal statements about the relationship between smoking and birth weight.

Remark.

According to the analysis above, birth weight remains significantly associated with mother's smoking exposure status. However, observe that birth weight is also strongly influenced by length of gestation (correlation coefficient of 0.426). Our conclusions above are not adjusted for the differences in lengths of gestation. Let us compare the mean gestation time across the smoking exposure classes.

GROUP	MNOCIG	SIZE	MEAN GESTATION	STANDARD DEVIATION
1	0 (Non-smoker)	381	39.9344	1.8662
2	2-7 (Occasional smoker)	103	39.5340	1.9088
3	12-17 (Less than pack/day)	66	39.3939	1.5872
4	25-50 (More than pack/day)	130	39.6692	1.9742

The table indicates that increases in smoking exposure tend to be associated with shorter gestation. That is, average length of gestation differs among the exposure categories. Although these differences are not large, it is of interest to adjust the mean levels of birth weight so the influence from differing lengths of gestation is removed from the comparisons of birth weights among the four categories of smoking exposure. This analysis can be carried out with analysis of covariance. You can learn about analysis of variance in higher level courses in statistics.