BZAN 535 **Homework 4** **Due September 21**

1. We wish to measure the impact of having an item appear in the Kroger weekly advertisement. You will study five 26 oz. Ragu pasta sauce with upc in (3620000250, 3620000300, 3620000350, 3620000441, 3620000446). For three Kentucky stores (333, 352, 377), these five upcs were all displayed in the weekly flyer (Interior Page Feature) for 18 weeks, with 14 of those weeks in year 2, i.e., with weeks > 52. This information is contained in the table carbo_causal_lookup.

    a. Show a query that will capture all the transaction data for these upcs and stores, aggregating the data by week, and computing the total unit sales, summed across stores and upcs. (You may find it useful to include other columns later, but these are not required for answering 1a.)

    b. Using the carbo_causal_lookup table, identify the 18 weeks where these upcs appear in the interior of the weekly flyer.

    c. Create a plot of total unit sales by week, highlighting the 18 weeks identified in b. Annotate any other weeks as seems relevant, based on other information available from the transactions or causal_lookup tables.

    d. Discuss what you can learn from the plot. Using just the plot, does there appear to be any indication that the weekly advertisements help sales of these products?

2. Propose a paired sample method to evaluate the effect of the flyer.
    a. First, identify the criteria you would propose for choosing pairs of weeks. What should they have in common?

    b. Second, select weeks to be paired with at least 10 of the 18 weeks from 1b. Explain the rationale for your choices, arguing why these matches were provided and why any weeks not used were omitted.

    c. Conduct a paired sample hypothesis test. Write out the null and alternative hypotheses very clearly in the language of the problem. (Make sure that both hypotheses include the word ‘average’ or ‘expected’.) Show all the details of the test using computer output. Interpret the p-value clearly and state your conclusion.

    d. Comment about the correlation between the pairs of your data, and what this indicates about the effectiveness of your pairing.

3. One might also try to use more of the data and give no thought to pairing, just treating the data as two independent samples.
    a. Out of the 104 -18 weeks you did not have the products displayed in the inner pages, which should be included for this comparison. Justify your choices of exclusions and inclusions.

    b. Conduct an independent sample test using the Interior Page Feature weeks and the other weeks you identified in 3a. Show the test results using output from JMP, interpret the p-value, and state your conclusion.