Technical Appendix

Antonio Skilton

September 16, 2016

Data Import

The dataset contains 5 variables and 200 observations detailing the advertising spending and sales revenue of a product in each market. The "TV", "Radio", and "Newspaper" columns represent advertising spending in thousands of dollars, "Sales" represents sales revenue in millions of dollars, and "Market" specifies whether the market advertised and sold in was "Urban", "Rural", or "SubUrban".

I have also added "AdSpending", which is simply the sum of "TV", "Radio", and "Newspaper".

```
library(dplyr)
AdData <- read.csv("1_FDA_AdData.csv") %>%
  mutate(Sales = Sales * 1000,AdSpending = rowSums(.[1:3]))
```

Table 1: First 20 rows of Advertising data

TV	Radio	Newspaper	Sales	Market	AdSpending
265.6	20.0	0.3	17400	Urban	285.9
69.0	9.3	0.9	9300	Rural	79.2
8.6	2.1	1.0	4800	Rural	11.7
184.9	43.9	1.7	20700	SubUrban	230.5
293.6	27.7	1.8	20700	Urban	323.1
8.4	27.2	2.1	5700	Rural	37.7
31.5	24.6	2.2	9500	Rural	58.3
175.7	15.4	2.4	14100	Urban	193.5
220.3	49.0	3.2	24700	SubUrban	272.5
100.4	9.6	3.6	10700	Rural	113.6
166.8	42.0	3.6	19600	SubUrban	212.4
286.0	13.9	3.7	15900	Urban	303.6
214.7	24.0	4.0	17400	Urban	242.7
266.9	43.8	5.0	25400	SubUrban	315.7
238.2	34.3	5.3	20700	Urban	277.8
117.2	14.7	5.4	11900	Urban	137.3
255.4	26.9	5.5	19800	Urban	287.8
4.1	11.6	5.7	3200	Rural	21.4
39.5	41.1	5.8	10800	Rural	86.4
197.6	3.5	5.9	11700	Urban	207.0

Summary Statistics

```
library(psych)
sumStats <- describe(select(AdData,-Market),skew=FALSE)[,c(-1,-2,-8)]
kable(sumStats,digits=2,caption="Summary Statistics",align='c')
```

Table 2: Summary Statistics

	mean	sd	min	max	range
TV	147.04	85.85	0.7	296.4	295.7
Radio	23.26	14.85	0.0	49.6	49.6
Newspaper	30.55	21.78	0.3	114.0	113.7
Sales	14022.50	5217.46	1600.0	27000.0	25400.0
AdSpending	200.86	92.99	11.7	433.6	421.9

```
AdData %>%
  group_by(Market) %>%
  summarise(TV = sum(TV),
            Radio = sum(Radio),
            Newspaper = sum(Newspaper),
            Total = sum(AdSpending)) %>%
  kable(caption=("Sums of Advertising Spending"),align='c')
```

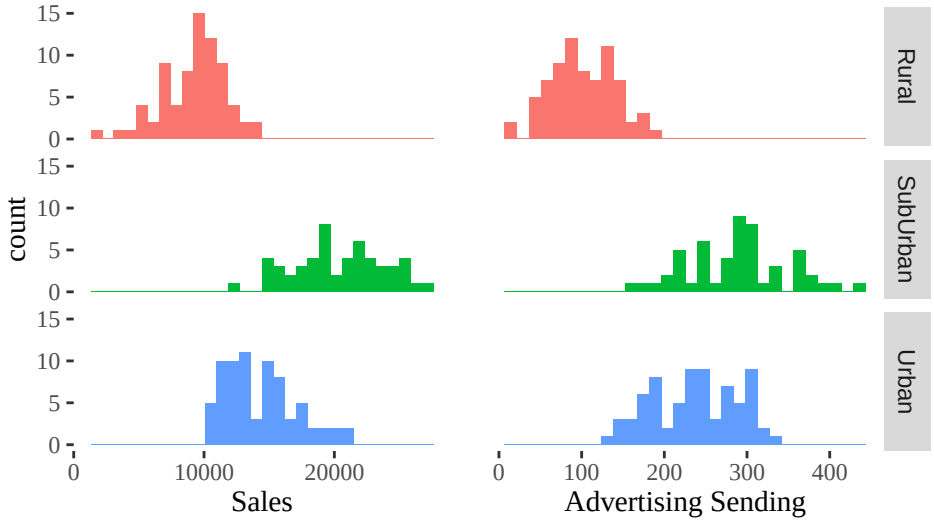
Table 3: Sums of Advertising Spending

Market	TV	Radio	Newspaper	Total
Rural	3979.6	1483.4	2065.5	7528.5
SubUrban	10684.8	2074.4	2527.6	15286.8
Urban	14744.1	1095.0	1517.7	17356.8

Histograms and total advertising spending per individual store by Market, in thousands of dollars below. It is clear that advertising spending is correlated with sales. Due to the categorical differences between markets however, we cannot assume any causative relationships

```
library(ggplot2)
plot_Sales <- ggplot(AdData,aes(x=Sales,fill=Market)) +
  geom_histogram(position="dodge") + facet_grid(Market~.) +
  theme(legend.position="none",strip.text.y=element_blank(),
        axis.title=element_text(family="Times"),axis.text=element_text(family="Times"),
        panel.background = element_rect(fill = "white"))
plot_Adspending <- ggplot(AdData,aes(x=AdSpending,fill=Market)) +
  geom_histogram(position="dodge") +
  facet_grid(Market~.) +
  theme(legend.position="none",axis.title.y=element_blank(),
        axis.text.y=element_blank(),axis.ticks.y=element_blank(),
        axis.title=element_text(family="Times"),axis.text=element_text(family="Times"),
        panel.background = element_rect(fill = "white")) +
  scale_y_continuous(limits=c(0,15)) +
  xlab("Advertising Sending")
library(gridExtra);library(grid)
grid.arrange(plot_Sales,
             plot_Adspending,
             ncol=2,
             top=textGrob("Figure 1: Histograms of Sales and Advertising Spending",
                          gp=gpar(fontsize=10,fontfamily="Times")))
```

Figure 1: Histograms of Sales and Advertising Spending

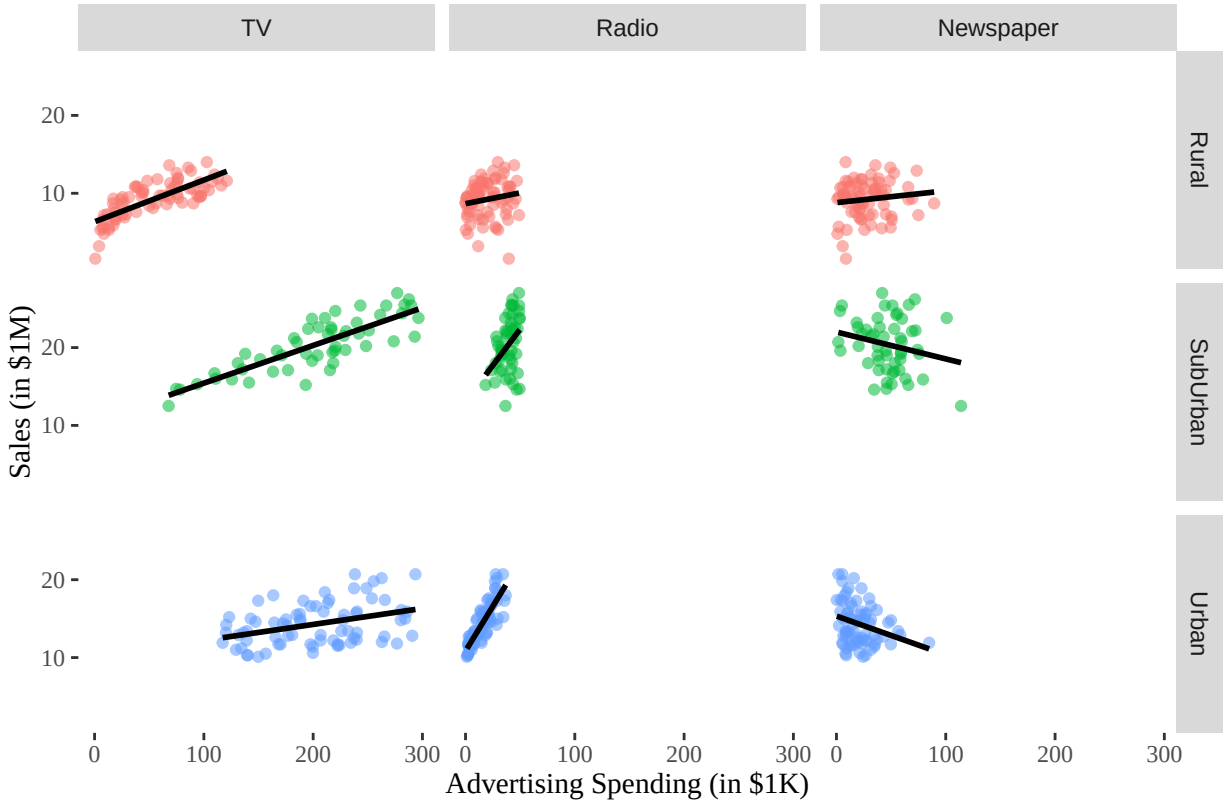


Correlations

We use the reshape2 package in order to reshape the data; each observation in the new dataset represents a unique combination of store and type of advertising. The resulting data frame is plotted by the ggplot2 package, which facets the data by both type of market and type of advertising spending. Each dot represents an individual observation as described, while the lines are trendlines that include 95% confidence interval boundaries.

```
AdData <- mutate(AdData,row_number=1:n())
library(reshape2)
AdData.melt <- melt(AdData,id=c("row_number","Market","Sales"))
AdData <- select(AdData,-row_number)
library(ggplot2)
ggplot(filter(AdData.melt,variable!="AdSpending"),aes(x=value,y=Sales/1000,alpha=0.75)) +
  geom_point(aes(color=Market)) +
  theme(legend.position="none",
        axis.title=element_text(family="Times"),
        axis.text=element_text(family="Times"),
        plot.title=element_text(family="Times"),
        panel.background = element_rect(fill = "white")) +
  stat_smooth(color="black",se=FALSE,method='lm') +
  facet_grid(Market~variable) +
  ggtitle("Figure 2: Sales by Advertising Spending") +
  ylab("Sales (in $1M)") +
  xlab("Advertising Spending (in $1K)")
```

Figure 2: Sales by Advertising Spending



Linear Models

In order to determine how to most efficiently allocate advertising spending, we use linear regression to find the model that most accurately predicts sales. Note that television ad spending has an exponential relationship with sales in Rural markets.

```
Rural.mod <- lm(Sales~TV+Radio+Newspaper,data=AdData[AdData$Market=="Rural",-5])
Urban.mod <- lm(Sales~TV+Radio+Newspaper,data=AdData[AdData$Market=="Urban",-5])
SubUrban.mod <- lm(Sales~TV+Radio+Newspaper,data=AdData[AdData$Market=="SubUrban",-5])
```

Table 4: Linear Models by Market

	Dependent variable:		
	Sales		
TV	67.625***	53.279***	26.134***
	(3.984)	(1.321)	(1.458)
Radio	86.670***	252.633***	241.059***
	(10.251)	(11.123)	(7.502)
Newspaper	5.858	−3.806	−6.211
	(6.927)	(3.595)	(4.845)
Constant	3,682.606***	−98.403	5,532.044***
	(378.655)	(623.726)	(356.228)
Observations	74	53	73
Adjusted R^2	0.803	0.974	0.950
Note:		*p<0.1; **p<0.05; ***p<0.01	