**YMCA “Hallowine” Silent Auction**

**Case Exercise: Data wrangling, modeling, analytics with descriptive statistics**

The Windward YMCA in Kailua, HI holds an annual fund raising event every Halloween called Hallow-wine. Business and individuals donate items to offer at a silent auction. Due to limited space and management overhead they can only offer a limited number of items at the auction. Items that do not get bid on at the auction are given away, given back to the donors, or thrown out (i.e. no salvage value). However generally a donor is willing to give a cash donation of approximately 40% of the value of the item in lieu of donating the item. The Executive Director Bill Hastings has a spreadsheet of items for the auction in the previous year but is unsure how to make use of this data to help improve the auction this year.

Tasks:

Let’s use the 6-steps in the BA problem solving process to address this case.

In this case you should make use of bar chart, percentiles, pareto, bubble chart, ANOVA

Lessons of this case are:

* Become familiar with the BA process
* Review some basic descriptive stats
* Review some inferential stats
* Experience using imperfect models to find “good enough” results (how to avoid going too far in pursuit of perfection or correctness)

Step 1. Recognizing the problem

1. What is the gap between what happened previously and what Bill thinks should be happening this year?

Many items went unsold or underbid. Missed opportunity to get cash donation for such items. Had to turn away some items due to lack of space. Would like to understand what kinds of items are of high interest and are likely to get bid up. Would like to get cash donation and turn away items that are not of high interest or get underbid.

Step 2. Defining the problem

1. What are the questions that if answered would address the gap?

Q1: What kinds of items raise the most money?

Q2: What kinds of items are of highest interest?

Q3: What do we use to determine if an item should be turned away?

Q4: What encourages sales?

1. What would satisfactory answers look like?

Q1: “80% of the funds raised came from X,Y,Z, …”, “On Average the highest amounts raised come from X,Y,Z,… in this order”

Q2: “On Average the highest amounts bids come from X,Y,Z,… in this order”

Q3: If the expected sale value is less than 40% of it’s assessed value ask for a donation.

Q4: Set the min bid for item X with value Y at Z

3. Structuring the problem

a. What are the key decisions to be made and how do the relate to the questions?

D1: What categories of items should we solicit or discourage? (prescriptive)

D2: Item X should be accepted or refused? (predictive, prescriptive)

D3: What should the min bid and increment should be set for item X? (descriptive)

b. What are the important factors to consider in making the decisions?

D1: Item expected sale amount per category, interest of category

D2: expected sale amount for item, likely alternative cash donation

D3: Impact of min bid and increment on interest and sale amount

c. What data is needed and is it available or easily obtained?

Available: Sales price, Value, Increment, Min bid, category

Add variables: Sales price/Value, # bids (Sales price/Increment), Value – Min bid, dummy “Sold”

Clean data (handle missing values in Increment, remove items not used)

4. Analyzing the problem

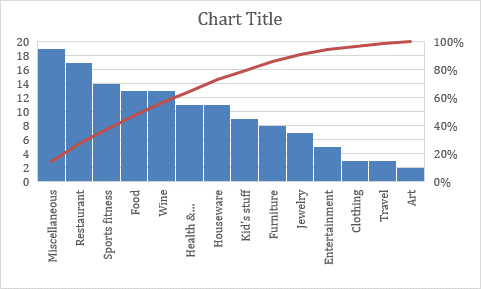
a. How will you represent the decision problems? What models can you use (descriptive, predictive, prescriptive)?



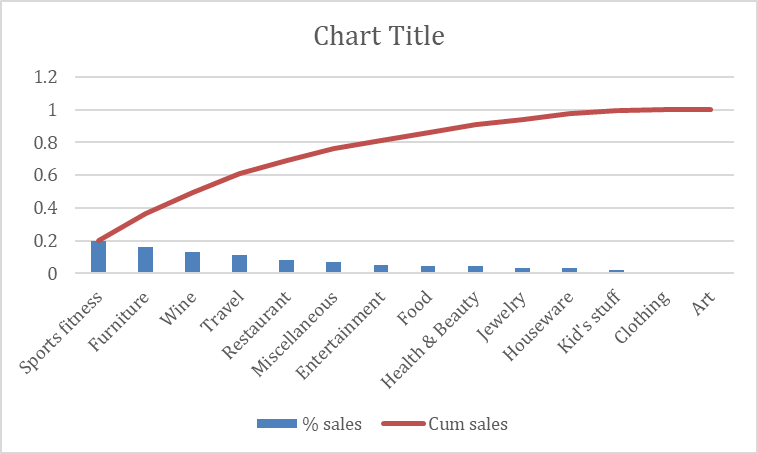
Create a pivot table

Sales price by category

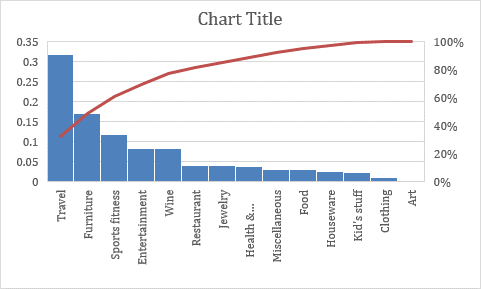
Pareto of items by category



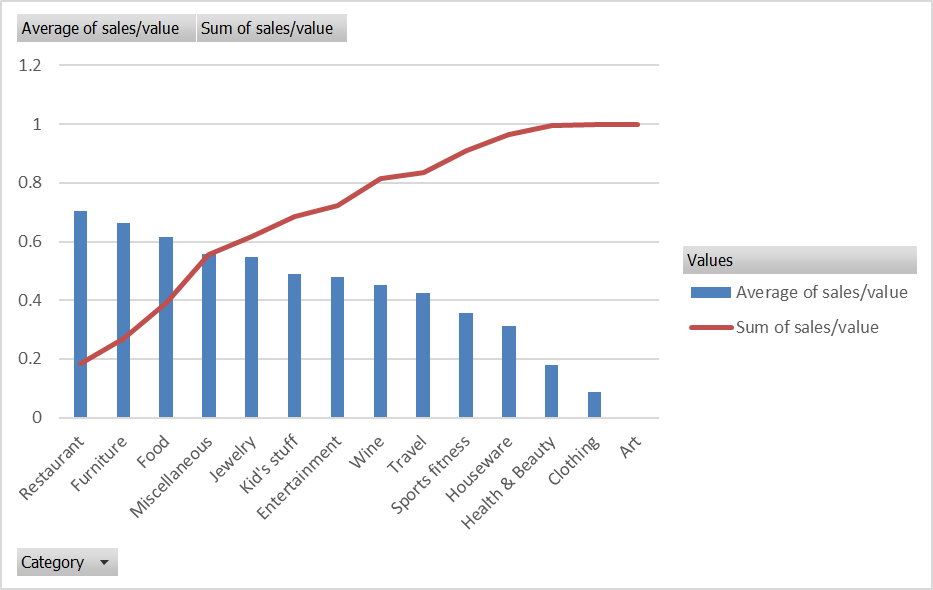
Pareto of % of sales by category



Pareto of % of average sales by category



Pareto of % of average sales/value by category



Sale frequency and expected sale amount per category



Average sales price / value per category



Average bids (sales price/increment) per category



Remove items with no min or increment (table filter, outliers, not sold

Scatter plot of min bid vs sales, increment vs sales





> BidValue <- YMCA$`Min. bid`/YMCA$Value; SalesValue <- YMCA$`sales price`/YMCA$Value;

> bid <- c(BidValue, SalesValue)

> model <- lm(SalesValue ~ BidValue)

> summary(model)

Call:

lm(formula = SalesValue ~ BidValue)

Residuals:

Min 1Q Median 3Q Max

-0.54416 -0.31168 -0.02792 0.18709 2.33334

Coefficients:

Estimate Std. Error t value Pr(>|t|)

(Intercept) 0.1567 0.1371 1.143 0.2554

BidValue 0.7749 0.3389 2.286 0.0239 \*

---

Signif. codes: 0 ‘\*\*\*’ 0.001 ‘\*\*’ 0.01 ‘\*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 0.3971 on 127 degrees of freedom

(6 observations deleted due to missingness)

Multiple R-squared: 0.03953, Adjusted R-squared: 0.03197

F-statistic: 5.227 on 1 and 127 DF, p-value: 0.02389



R version 3.4.3 (2017-11-30) -- "Kite-Eating Tree"

Copyright (C) 2017 The R Foundation for Statistical Computing

Platform: x86\_64-apple-darwin15.6.0 (64-bit)

R is free software and comes with ABSOLUTELY NO WARRANTY.

You are welcome to redistribute it under certain conditions.

Type 'license()' or 'licence()' for distribution details.

Natural language support but running in an English locale

R is a collaborative project with many contributors.

Type 'contributors()' for more information and

'citation()' on how to cite R or R packages in publications.

Type 'demo()' for some demos, 'help()' for on-line help, or

'help.start()' for an HTML browser interface to help.

Type 'q()' to quit R.

[Workspace loaded from ~/.RData]

> View(Auction)

> View(Auction)

> View(data)

> View(data)

> p <- ggplot(data=YMCA, aes(x=Value, y=`sales price`, fill=Category, color=Category)) + geom\_point()

Error in ggplot(data = YMCA, aes(x = Value, y = `sales price`, fill = Category, :

could not find function "ggplot"

> View(YMCA)

> View(YMCA)

> library("ggplot2", lib.loc="/Library/Frameworks/R.framework/Versions/3.4/Resources/library")

> p <- ggplot(data=YMCA, aes(x=Value, y=`sales price`, fill=Category, color=Category)) + geom\_point()

> p <- ggplot(data=YMCA, aes(x=Value, y=`sales price`, fill=Category, color=Category)) + geom\_point()

> p + geom\_segment(data=YMCA, mapping=aes(x=0,xend=2000,y=0,yend=2000\*.4))

> q <- ggplot(data=YMCA, aes(x=YMCA$`Min. bid`, y=YMCA$`sales price`/YMCA$Value, fill=Category, color=Category)) + geom\_point()

> q <- ggplot(data=YMCA, aes(x=YMCA$`Min. bid`, y=YMCA$`sales price`/YMCA$Value, fill=Category, color=Category)) + geom\_point()

> p + geom\_segment(data=YMCA, mapping=aes(x=0,xend=2000,y=0,yend=2000\*.4))

> q <- ggplot(data=YMCA, aes(x=YMCA$`Min. bid`, y=YMCA$`sales price`/YMCA$Value, fill=Category, color=Category)) + geom\_point()

> q + geom\_segment(data=YMCA, mapping=aes(x=0,xend=2000,y=0,yend=2000\*.4))

Warning message:

Removed 6 rows containing missing values (geom\_point).

> plot(x=YMCA$`Min. bid`, y=YMCA$`sales price`/YMCA$Value)

> plot(x=YMCA$`Min. bid`/YMCA$Value, y=YMCA$`sales price`/YMCA$`sales price`)

> plot(x=YMCA$`Min. bid`/YMCA$Value, y=YMCA$`sales price`/YMCA$Value)

> bid <- c(BidValue=YMCA$`Min. bid`/YMCA$Value, SalesValue=YMCA$`sales price`/YMCA$Value )

> summary(bid)

Min. 1st Qu. Median Mean 3rd Qu. Max. NA's

0.0000 0.3000 0.4000 0.4349 0.5000 2.8000 6

> describe(bid)

Error in describe(bid) : could not find function "describe"

> bid

BidValue1 BidValue2 BidValue3 BidValue4 BidValue5 BidValue6 BidValue7

0.20000000 0.26666667 0.28571429 0.35000000 0.35714286 0.35714286 0.37500000

BidValue8 BidValue9 BidValue10 BidValue11 BidValue12 BidValue13 BidValue14

0.33333333 0.40000000 0.41666667 0.33333333 0.33333333 0.41666667 0.30000000

BidValue15 BidValue16 BidValue17 BidValue18 BidValue19 BidValue20 BidValue21

0.40000000 0.50000000 0.30000000 0.44444444 0.44444444 0.50000000 0.50000000

BidValue22 BidValue23 BidValue24 BidValue25 BidValue26 BidValue27 BidValue28

0.33333333 0.33333333 0.50000000 0.50000000 0.40000000 0.20000000 0.20000000

BidValue29 BidValue30 BidValue31 BidValue32 BidValue33 BidValue34 BidValue35

0.40000000 0.40000000 0.40000000 0.40000000 0.40000000 0.40000000 0.45454545

BidValue36 BidValue37 BidValue38 BidValue39 BidValue40 BidValue41 BidValue42

0.44642857 0.33333333 0.33333333 NA 0.37500000 0.37500000 0.37500000

BidValue43 BidValue44 BidValue45 BidValue46 BidValue47 BidValue48 BidValue49

0.50000000 0.33333333 0.33333333 0.38461538 0.37500000 0.41666667 0.33333333

BidValue50 BidValue51 BidValue52 BidValue53 BidValue54 BidValue55 BidValue56

0.46666667 0.38095238 NA 0.49333333 0.69811321 0.69811321 0.49333333

BidValue57 BidValue58 BidValue59 BidValue60 BidValue61 BidValue62 BidValue63

0.35714286 0.41666667 0.35000000 0.36000000 0.42857143 0.30000000 0.33333333

BidValue64 BidValue65 BidValue66 BidValue67 BidValue68 BidValue69 BidValue70

0.33333333 0.50000000 0.42857143 0.33333333 0.35000000 0.33333333 0.50000000

BidValue71 BidValue72 BidValue73 BidValue74 BidValue75 BidValue76 BidValue77

0.41666667 0.40000000 0.40000000 0.50000000 0.50000000 0.50000000 0.50000000

BidValue78 BidValue79 BidValue80 BidValue81 BidValue82 BidValue83 BidValue84

NA NA NA 0.50000000 0.29411765 0.20000000 0.28571429

BidValue85 BidValue86 BidValue87 BidValue88 BidValue89 BidValue90 BidValue91

0.28571429 0.28571429 0.20000000 0.20000000 0.50000000 0.20000000 0.50000000

BidValue92 BidValue93 BidValue94 BidValue95 BidValue96 BidValue97 BidValue98

NA 0.60000000 0.60000000 0.29411765 0.01176471 0.40000000 0.50000000

BidValue99 BidValue100 BidValue101 BidValue102 BidValue103 BidValue104 BidValue105

0.60000000 0.60000000 0.40000000 0.35000000 0.50000000 0.50000000 0.50000000

BidValue106 BidValue107 BidValue108 BidValue109 BidValue110 BidValue111 BidValue112

0.50000000 0.46666667 0.50000000 0.40000000 0.25000000 0.33333333 0.46666667

BidValue113 BidValue114 BidValue115 BidValue116 BidValue117 BidValue118 BidValue119

0.30000000 0.32258065 0.40000000 0.32258065 0.33333333 0.37878788 0.37878788

BidValue120 BidValue121 BidValue122 BidValue123 BidValue124 BidValue125 BidValue126

0.37878788 0.40000000 0.42857143 0.36363636 0.40000000 0.41666667 0.36363636

BidValue127 BidValue128 BidValue129 BidValue130 BidValue131 BidValue132 BidValue133

0.41666667 0.33333333 0.36363636 0.40000000 0.40000000 0.40000000 0.40000000

BidValue134 BidValue135 SalesValue1 SalesValue2 SalesValue3 SalesValue4 SalesValue5

0.42857143 0.10000000 0.00000000 0.00000000 0.00000000 0.00000000 0.00000000

SalesValue6 SalesValue7 SalesValue8 SalesValue9 SalesValue10 SalesValue11 SalesValue12

0.00000000 0.00000000 0.00000000 0.00000000 0.00000000 0.00000000 0.00000000

SalesValue13 SalesValue14 SalesValue15 SalesValue16 SalesValue17 SalesValue18 SalesValue19

0.00000000 0.00000000 0.00000000 0.00000000 0.00000000 0.00000000 0.00000000

SalesValue20 SalesValue21 SalesValue22 SalesValue23 SalesValue24 SalesValue25 SalesValue26

0.00000000 0.00000000 0.00000000 0.00000000 0.00000000 0.00000000 0.00000000

SalesValue27 SalesValue28 SalesValue29 SalesValue30 SalesValue31 SalesValue32 SalesValue33

0.00000000 0.00000000 0.00000000 0.00000000 0.00000000 0.00000000 0.26666667

SalesValue34 SalesValue35 SalesValue36 SalesValue37 SalesValue38 SalesValue39 SalesValue40

0.26666667 0.90909091 0.89285714 0.33333333 1.00000000 0.80000000 0.50000000

SalesValue41 SalesValue42 SalesValue43 SalesValue44 SalesValue45 SalesValue46 SalesValue47

0.50000000 0.50000000 0.60000000 1.00000000 1.00000000 0.53846154 0.87500000

SalesValue48 SalesValue49 SalesValue50 SalesValue51 SalesValue52 SalesValue53 SalesValue54

0.66666667 0.66666667 0.60000000 0.76190476 0.84000000 0.49333333 0.69811321

SalesValue55 SalesValue56 SalesValue57 SalesValue58 SalesValue59 SalesValue60 SalesValue61

0.69811321 0.54666667 0.35714286 0.41666667 0.40000000 0.36000000 0.45714286

SalesValue62 SalesValue63 SalesValue64 SalesValue65 SalesValue66 SalesValue67 SalesValue68

0.30000000 0.33333333 0.33333333 0.50000000 0.85714286 0.66666667 0.45000000

SalesValue69 SalesValue70 SalesValue71 SalesValue72 SalesValue73 SalesValue74 SalesValue75

0.33333333 0.85000000 0.66666667 0.60000000 0.60000000 0.78571429 0.17500000

SalesValue76 SalesValue77 SalesValue78 SalesValue79 SalesValue80 SalesValue81 SalesValue82

0.25000000 0.27500000 0.86666667 1.00000000 0.60000000 1.25000000 0.29411765

SalesValue83 SalesValue84 SalesValue85 SalesValue86 SalesValue87 SalesValue88 SalesValue89

0.60000000 0.28571429 0.28571429 0.28571429 1.00000000 1.10000000 0.60000000

SalesValue90 SalesValue91 SalesValue92 SalesValue93 SalesValue94 SalesValue95 SalesValue96

1.30000000 0.70000000 0.90000000 0.76000000 0.80000000 1.17647059 0.12941176

SalesValue97 SalesValue98 SalesValue99 SalesValue100 SalesValue101 SalesValue102 SalesValue103

0.40000000 0.70000000 0.84000000 0.88000000 0.96000000 0.40000000 1.33333333

SalesValue104 SalesValue105 SalesValue106 SalesValue107 SalesValue108 SalesValue109 SalesValue110

0.90000000 0.90000000 0.50000000 0.80000000 0.92857143 2.80000000 0.62500000

SalesValue111 SalesValue112 SalesValue113 SalesValue114 SalesValue115 SalesValue116 SalesValue117

0.40000000 0.73333333 0.35000000 0.37634409 0.30000000 0.43010753 0.50000000

SalesValue118 SalesValue119 SalesValue120 SalesValue121 SalesValue122 SalesValue123 SalesValue124

0.42613636 0.42613636 0.42613636 0.60000000 0.42857143 0.22727273 1.04000000

SalesValue125 SalesValue126 SalesValue127 SalesValue128 SalesValue129 SalesValue130 SalesValue131

0.50000000 0.27272727 0.66666667 0.33333333 0.45454545 0.40000000 0.45000000

SalesValue132 SalesValue133 SalesValue134 SalesValue135

0.45000000 0.50000000 0.57142857 0.25000000

> BidValue

Error: object 'BidValue' not found

> BidValue <- YMCA$`Min. bid`/YMCA$Value; SalesValue <- YMCA$`sales price`/YMCA$Value;

> bid <- c(BidValue, SalesValue)

> model <- lm(SalesValue ~ BidValue)

> summary(model)

Call:

lm(formula = SalesValue ~ BidValue)

Residuals:

Min 1Q Median 3Q Max

-0.54416 -0.31168 -0.02792 0.18709 2.33334

Coefficients:

Estimate Std. Error t value Pr(>|t|)

(Intercept) 0.1567 0.1371 1.143 0.2554

BidValue 0.7749 0.3389 2.286 0.0239 \*

---

Signif. codes: 0 ‘\*\*\*’ 0.001 ‘\*\*’ 0.01 ‘\*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 0.3971 on 127 degrees of freedom

(6 observations deleted due to missingness)

Multiple R-squared: 0.03953, Adjusted R-squared: 0.03197

F-statistic: 5.227 on 1 and 127 DF, p-value: 0.02389

> BidValue <- YMCA$`Min. bid`;

> Bid <- YMCA$`Min. bid`;

> bid <- c(Bid, SalesValue)

> model <- lm(SalesValue ~ Bid)

> summary(model)

Call:

lm(formula = SalesValue ~ Bid)

Residuals:

Min 1Q Median 3Q Max

-0.5045 -0.3550 0.0086 0.2086 2.3035

Coefficients:

Estimate Std. Error t value Pr(>|t|)

(Intercept) 0.5065510 0.0460658 10.996 <2e-16 \*\*\*

Bid -0.0010102 0.0006403 -1.578 0.117

---

Signif. codes: 0 ‘\*\*\*’ 0.001 ‘\*\*’ 0.01 ‘\*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 0.4012 on 127 degrees of freedom

(6 observations deleted due to missingness)

Multiple R-squared: 0.01922, Adjusted R-squared: 0.0115

F-statistic: 2.489 on 1 and 127 DF, p-value: 0.1171

> BidValue <- YMCA$`Min. bid`/YMCA$Value; SalesValue <- YMCA$`sales price`/YMCA$Value;

> IncValue <- YMCA$Increment/YMCA$Value;

> inc <- c(IncValue, SalesValue)

> plot(SalesValue ~ IncValue)

> model <- lm(SalesValue ~ IncValue)

> summary(model)

Call:

lm(formula = SalesValue ~ IncValue)

Residuals:

Min 1Q Median 3Q Max

-0.4645 -0.3360 -0.0367 0.2110 2.3391

Coefficients:

Estimate Std. Error t value Pr(>|t|)

(Intercept) 0.46568 0.08202 5.678 8.81e-08 \*\*\*

IncValue -0.06009 0.76831 -0.078 0.938

---

Signif. codes: 0 ‘\*\*\*’ 0.001 ‘\*\*’ 0.01 ‘\*’ 0.05 ‘.’ 0.1 ‘ ’ 1

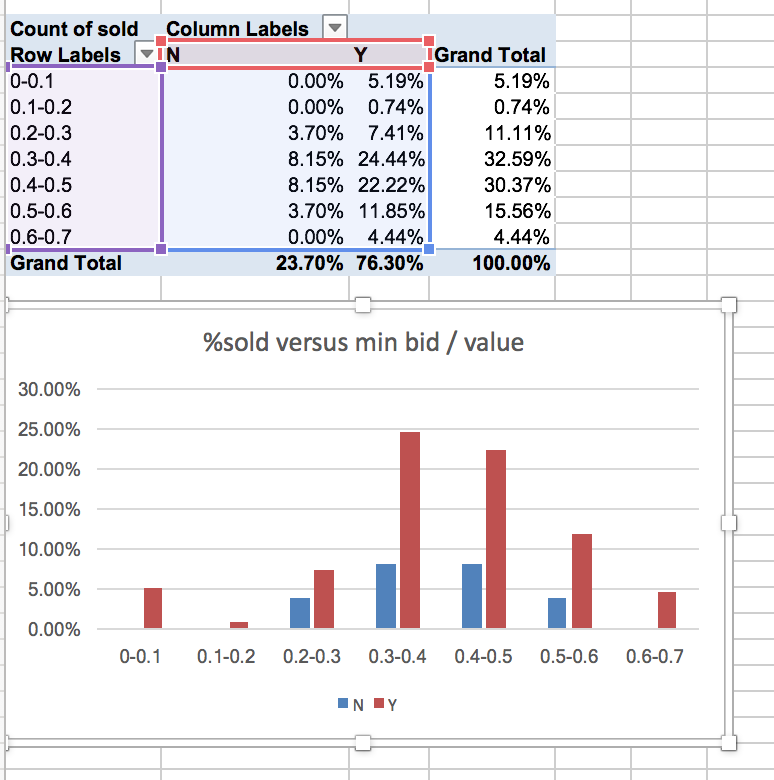
Residual standard error: 0.4051 on 127 degrees of freedom

(6 observations deleted due to missingness)

Multiple R-squared: 4.817e-05, Adjusted R-squared: -0.007825

F-statistic: 0.006118 on 1 and 127 DF, p-value: 0.9378





b. What confidence can you have in the results?

Consider variability when looking at averages. More variability implies lower confidence. Use confidence intervals, hypothesis tests (ANOVA, Tukey-Kramer), or variability diagrams to indicate confidence.

c. Are the assumptions reasonable?

Consider assumptions about representativeness of previous auction data, sample sizes, quality of the data (e.g. do sales always start with the minimum bid?), any issues that would affect the answers to the decision questions!

5. Interpreting results and making a decision

a. Analyze the models and discuss the implications for the decisions

Use the models to answer the decision questions (details and specifics not generalities!)

6. Implementing the solution

a. How should the results be used for this year’s auction?

Detail how the answers to the questions should be used when making decisions e.g. “Given the value of a donation for a given item, use the sale frequency and average sales/value for the item category to estimate the expected sale amount for that item. It is less than 40% of the value, ask for a cash donation”

b. What are the limitations, cautions, and possible exceptions?

There are obvious high-value or high-interest items that should be dealt with on a case by case basis. Also think about “boundary” cases or when there are not many donations.