Homework Assignment #1, Math 740/840, Fall 2019
Charlie Nitschelm

1. (10 pts) Civil engineers are studying four different water filtration systems for municipal water supplies. The response of interest is the level of impurities (in ppm) remaining in the water after filtration. The four systems were hooked up in parallel to the same input water source for the experiment. Over a period of one week, after the filters were properly conditioned, individual gallons of water were randomly obtained from each of the filtration system and the amount of remaining impurities in the water samples measured. Use the data set **Impurities.JMP** and the JMP software to answer questions about this experiment.

   a. What is the experimental unit (EU) for this experiment?

   Gallon of water

   b. Write out the one-way ANOVA model for an experiment with replication and no subsampling.

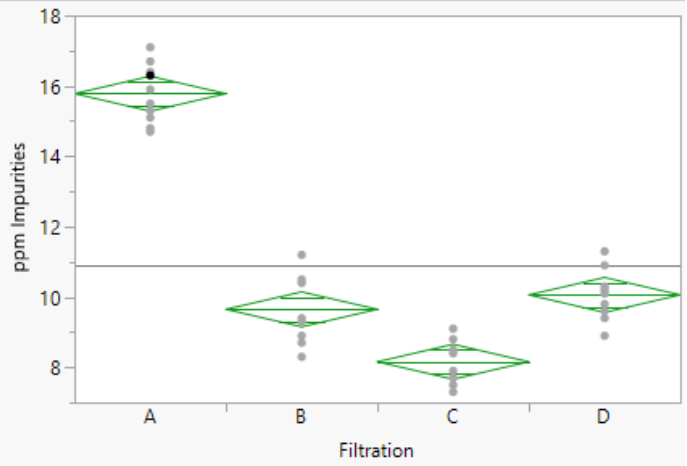
   $$Y_{ij} = \mu + \tau_i + \epsilon_{ij}$$

   c. State the null and alternative hypotheses for a one-way ANOVA analysis of the data.

   The null hypothesis is that the ppm collected does not differ by the type of filtration method. The alternative hypothesis is that the filtration method does affect the ppm of the data collected.

   d. Use the Fit Y by X platform in JMP to perform a one-way ANOVA on the experimental results. Be sure to include the relevant portions of the JMP output in your report.

Oneway Analysis of ppm Impurities By Filtration



Oneway Anova

Summary of Fit

Rsquare	0.939451
Adj Rsquare	0.934405
Root Mean Square Error	0.775475
Mean of Response	10.9125
Observations (or Sum Wgts)	40

Analysis of Variance

Source	DF	Sum of Squares	Mean Square	F Ratio	Prob > F
Filtration	3	335.89475	111.965	186.1858	<.0001*
Error	36	21.64900	0.601		
C. Total	39	357.54375			

Means for Oneway Anova

Level	Number	Mean	Std Error	Lower 95%	Upper 95%
A	10	15.7800	0.24523	15.283	16.277
B	10	9.6500	0.24523	9.153	10.147
C	10	8.1600	0.24523	7.663	8.657
D	10	10.0600	0.24523	9.563	10.557

Std Error uses a pooled estimate of error variance

e. Based upon the JMP output what is your conclusion concerning possible mean impurity level differences among the filtration methods.

Because the F-test is very small, we are forced to reject the null hypothesis in favor of the alternative. This means that there is a correlation between the ppm level and the filtration method.

f. **From the scatter plot in Fit Y by X which filtration may be best? Which may be worst?**

From the graph and the confidence intervals, it details that filtration method C supplies the best method for filtration. Its upper 95% confidence interval is still below the next best filtration method of B. The worst filtration method is A as its lower 95% confidence is still significantly higher then the other filtration methods.

2. (15 pts.) Use JMP to analyze the dataset **Chips.JMP**. Notice that this design involves subsampling, so you should use Fit Model to perform the correct analysis – see the Fit Model analysis for the Electroplate example in Chapter 2. The data consists of an experiment studying three different processes to manufacture semiconductors (chips). For each of the four processes 3 replicate chips were selected for electrical testing, and on each chip four measurements were taken at different locations. The response is resistance.

   a. What is the EU? What is the OU"

   The EU is the semi-conductor ship, and the OU is the location where the test occurs.

   b. Write out the One-way ANOVA model for an experiment with replication and subsampling and briefly explain the terms in the model.

   $$Y_{ijk} = \mu + \tau_i + \varepsilon_{ij} + \eta_{ijk}$$

   Where Y is the response, μ is the overall mean, τ_i is the shift in the mean response for the i-th treatment, ε_{ij} is random error or noise with respect to treatment and replicate, and η_{ijk} is the theoretical observation errors within a single EU, with respect to treatment, replicate and OU (k).

   c. State the null and alternative hypothesis for a treatment or process effect.

   The null hypothesis is that the process, wafer and location have no impact on the measured resistance. The alterative hypothesis is that the process, wafer and location have an impact on resistance.

   d. What is your decision concerning the null hypothesis based upon the F test for a treatment or process effect? What is the p-value for the test of the treatment effect?

   Do the analysis in the Fit Model platform of JMP using the electroplating example from the Chapter 2 notes as a guideline.

   The F-test provided a value of 0.2015, so we are forced to accept the null hypothesis.

Response Resistance

Effect Summary

Summary of Fit

RSquare	0.129602
RSquare Adj	0.070256
Root Mean Square Error	0.465768
Mean of Response	6.002917
Observations (or Sum Wgts)	48

Parameter Estimates

Term	Estimate	Std Error	DFDen	t Ratio	Prob>|t|
Intercept	6.0029167	0.058011	12	103.48	<.0001*
Process[A]	-0.377083	0.100477	12	-3.75	0.0028*
Process[B]	-0.037083	0.100477	12	-0.37	0.7185
Process[C]	0.0845833	0.100477	12	0.84	0.4164

Random Effect Predictions

REML Variance Component Estimates

Random Effect	Var Ratio	Var Component	Std Error	95% Lower	95% Upper	Wald p-Value	Pct of Total
Location[Process]	-0.085137	-0.01847	0.0284607	-0.074252	0.0373123	0.5164	0.000
Residual		0.2169396	0.0542349	0.1402992	0.3795394		100.000
Total		0.2169396	0.0542349	0.1402992	0.3795394		100.000

-2 LogLikelihood = 66.80171055

Note: Total is the sum of the positive variance components.

Total including negative estimates = 0.1984699

Covariance Matrix of Variance Component Estimates

Iterations

Fixed Effect Tests

Source	Nparm	DF	DFDen	F Ratio	Prob > F
Process	3	3	12	6.4222	0.0077*

Effect Details

Process

Least Squares Means Table

Level	Least Sq Mean	Std Error
A	5.6258333	0.11602103
B	5.9658333	0.11602103
C	6.0875000	0.11602103
D	6.3325000	0.11602103

Location[Process]

Least Squares Means Table

e. Which contributes more random variation in the observed resistance values, the replicate chips within each process or the measurements taken from within each chip? Explain.

Location within the chip adds more random variation then the other as the var component value for the location (process) is 0.1016, while the Residual, which is the location within the chip, is 0.128, slightly larger.

f. Suppose you could redo the experiment, however budgetary constraints prevented you from increasing the number of measurements taken. Would you prefer to include more replicates (chips) per process and fewer measurements per chip or perhaps fewer chips and more measurements on each chip? Explain? I am only asking for a brief qualitative answer based on your results from the analysis of the data.

There is a higher error for the wafer process category based on the analysis, so including more chips per process and fewer measurements per chip would decrease that error if it needed to be done based on budget requirements.