A | Research Topic

Investigation of the factors affecting the students mean performance in a Math 099 course.

B | Experiment

Following the performance of a class of students in a Math 099 course separated into 4 distinct sections, we decided to investigate any significant effect of two factors on the mean performance of the students. These two effects are the attendance rate of the students and the section to which the student is assigned.

1. What effects do attendance rate and section have on the performance of the students?
2. Is there an attendance rate that will give a uniformly better student's mean performance regardless of the section the student belongs to?

The experiment is conducted via student data collected throughout the Spring semester of 2019 in Western Illinois University. The data collected is presented below as a two-factor factorial design:

		Factor B: Sections			
		SEC1 – 45059	SEC2 – 45060	SEC3 – 45061	SEC4 – 45062
Factor B: Attendance Rate	< 80%	207.38 203.75 123.33	207.75 204.17 131.88	202.80 185.63 124.00	178.50 195.30 93.75
	>= 80%	207.55 231.90 176.25	230.11 262.13 195.45	194.82 228.18 147.78	216.50 229.71 184.17

The treatment is the test (1, 2, and 3) and the attendance rate is obtained from the beginning of classes till the test 1, test 1 – test 2, test 2 – test 3 respectively for each individual section. The data is then partitioned according to the attendance rate and a sample mean performance (response) is calculated for each partition (eliminating attendances of 0%, since this student is considered to have not participated in the test) of test scores.

C | Pre-diagnostics analysis

R-Code:

```
########################################################################################
#################
##### Two Factors Fixed Effet Design
##### Design: Factor A = Attendance Rate; Factor B = Section; Treatment = Test
numFacA=2;
numFacB=4;
numObvs=3;
facA=c(rep("< 80%",numFacB*numObvs),rep(">= 80%",numFacB*numObvs));
```

```
facB=c(rep(c(rep("SEC1",numObvs),rep("SEC2",numObvs),rep("SEC3",numObvs),rep("SEC4",numObvs)),numFacA));

yij=c(207.38,     203.75, 123.33, 207.75, 204.17, 131.88, 202.80, 185.63, 124.00, 178.50, 195.30,
         93.75, 207.55,   231.90, 176.25, 230.11, 262.13, 195.45, 194.82, 228.18, 147.78, 216.50,
         229.71, 184.17);

facA<-as.factor(facA); # makes row variable as a factor in r
facB<-as.factor(facB); # makes col variable as a factor in r

dat=data.frame(facA,facB,yij); dat;

##### Graphing multiple boxplots####
par(mfrow=c(1,2))# Partitioning your graphics window, this is optional, you can ignore it.
boxplot(yij~facA) #multiple box plots for facA effects
boxplot(yij~facB) #nultiple box plots for facB effects

##### Summary Statistics
tmp <- do.call(data.frame,
          list(mean = tapply(yij, facA, mean),
             sd = tapply(yij, facA, sd),
             median = tapply(yij, facA, median),
             min = tapply(yij, facA, min),
             max = tapply(yij, facA, max),
             n = tapply(yij, facA, length))); tmp
tmp <- do.call(data.frame,
          list(mean = tapply(yij, facB, mean),
             sd = tapply(yij, facB, sd),
             median = tapply(yij, facB, median),
             min = tapply(yij, facB, min),
             max = tapply(yij, facB, max),
             n = tapply(yij, facB, length))); tmp
```

R-Result:

```
> ############################################################################
##########################
> ##### Two Factors Fixed Effet Design
> ##### Design: Factor A = Attendance Rate; Factor B = Section; Treatment =
Test
> numFacA=2;
> numFacB=4;
> numObvs=3;
> facA=c(rep("< 80%",numFacB*numObvs),rep(">= 80%",numFacB*numObvs));
> facB=c(rep(c(rep("SEC1",numObvs),rep("SEC2",numObvs),rep("SEC3",numObvs),
rep("SEC4",numObvs)),numFacA));
>
> yij=c(207.38,203.75, 123.33, 207.75, 204.17, 131.88, 202.80, 185.63, 124.00
,       178.50, 195.30, 93.75, 207.55,  231.90, 176.25, 230.11, 262.13, 195.45
,       194.82, 228.18, 147.78, 216.50, 229.71, 184.17);
>
> facA<-as.factor(facA); # makes row variable as a factor in r
```

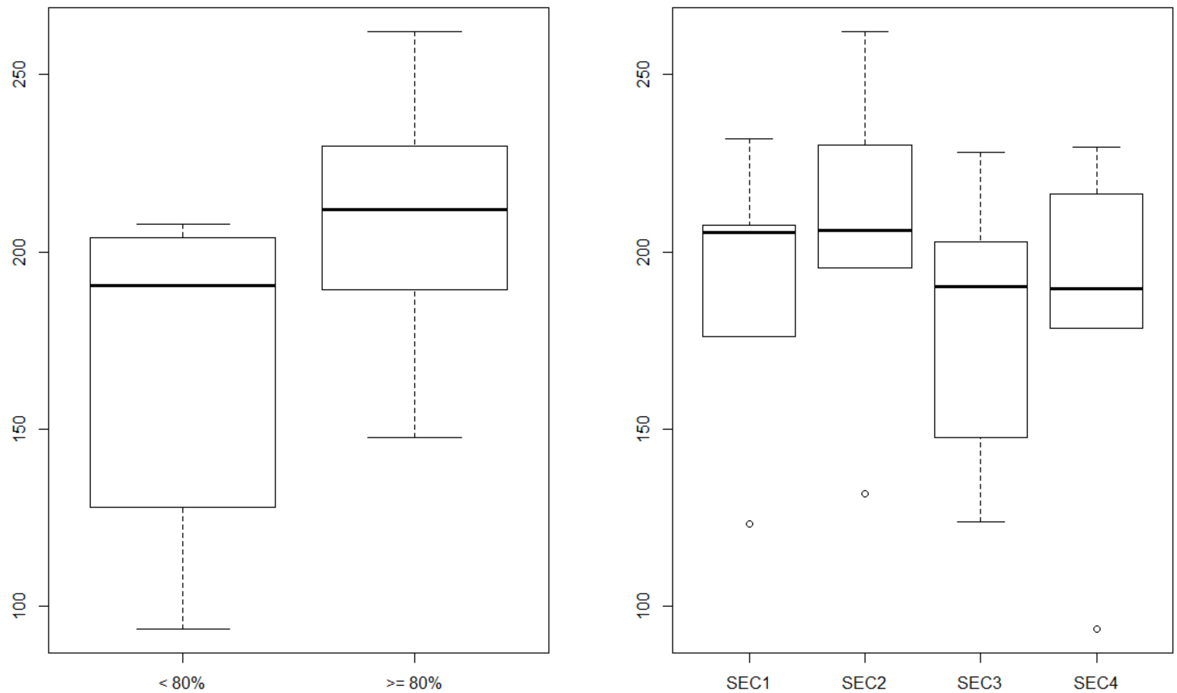
```
> facB<-as.factor(facB); # makes col variable as a factor in r
>
> dat=data.frame(facA,facB,yij); dat;
    facA facB    yij
1   < 80% SEC1 207.38
2   < 80% SEC1 203.75
3   < 80% SEC1 123.33
4   < 80% SEC2 207.75
5   < 80% SEC2 204.17
6   < 80% SEC2 131.88
7   < 80% SEC3 202.80
8   < 80% SEC3 185.63
9   < 80% SEC3 124.00
10  < 80% SEC4 178.50
11  < 80% SEC4 195.30
12  < 80% SEC4  93.75
13 >= 80% SEC1 207.55
14 >= 80% SEC1 231.90
15 >= 80% SEC1 176.25
16 >= 80% SEC2 230.11
17 >= 80% SEC2 262.13
18 >= 80% SEC2 195.45
19 >= 80% SEC3 194.82
20 >= 80% SEC3 228.18
21 >= 80% SEC3 147.78
22 >= 80% SEC4 216.50
23 >= 80% SEC4 229.71
24 >= 80% SEC4 184.17
>
> ##### Graphing multiple boxplots####
> par(mfrow=c(1,2))# Partitioning your graphics window, this is optional, y
ou can ignore it.
> boxplot(yij~facA) #multiple box plots for facA effects
> boxplot(yij~facB) #nultiple box plots for facB effects

>
```

Summary Statistics

	mean	sd	median	min	max	n
< 80%	171.52	41.22556	190.465	93.75	207.75	12
>= 80%	208.7125	30.84953	212.025	147.78	262.13	12

	mean	sd	median	min	max	n
SEC1	191.6933	37.87212	205.565	123.33	231.9	6
SEC2	205.2483	43.22029	205.96	131.88	262.13	6
SEC3	180.535	38.11796	190.225	124	228.18	6
SEC4	182.9883	47.83836	189.735	93.75	229.71	6

Box plots



- Comments: From this pre-diagnostic we can observe a significant difference in the mean performance of the attendance rates, with better mean performance recorded for the attendance rate of “>=80%”.
- Meanwhile, on the other box plot (rightmost) we observe no significant difference in the mean performance throughout sections.
- Additionally, there is the presence of some outliers on the data relative to sections 1,2, and 4.
- The effect (negative) of this outlier will be noticed in the process of `validating our analysis.
- Nevertheless, this is usually not a major concern since the effects of the estimates are reasonably robust to outliers.

D | ANOVA

- Hypothesis:
- Row treatment effects,
  - H_{0A}: $T_{<80\%} = T_{>=80\%} = 0$
  - H_{1A}: at least one $T_i \neq 0$
- Column treatment effects,
  - H_{0B}: $\beta_1 = \beta_2 = \beta_3 = \beta_4 = 0$
  - H_{1B}: at least one $\beta_i \neq 0$
- Row column treatment interaction
  - H_{0AB}: $(T\beta)_{ij} = 0$ for all i, j
  - H_{1AB}: at least one $(T\beta)_{ij} \neq 0$

R-Code:

```
##### ANOVA table construction with AOV command.
aov.out=aov(yij~facA+facB+facA*facB, data=dat); summary(aov.out);
```

R-Result:

```
> ##### ANOVA table construction with AOV command.
> aov.out=aov(yij~facA+facB+facA*facB, data=dat); summary(aov.out);
            Df Sum Sq Mean Sq F value Pr(>F)
facA         1   8300    8300   5.171 0.0371 *
facB         3   2244     748   0.466 0.7100
facA:facB    3   1237     412   0.257 0.8553
Residuals   16  25682    1605
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

>
```

- Observation: From the ANOVA table we observe a significant effect from FacA = Attendance rate at alpha = 0.10.
- With no significant effect from FacB = Sections nor from the interaction of FacA and FacB.
- That is, reject H_{0A}, fail to reject H_{1B} and H_{1AB}

E | Assumptions check

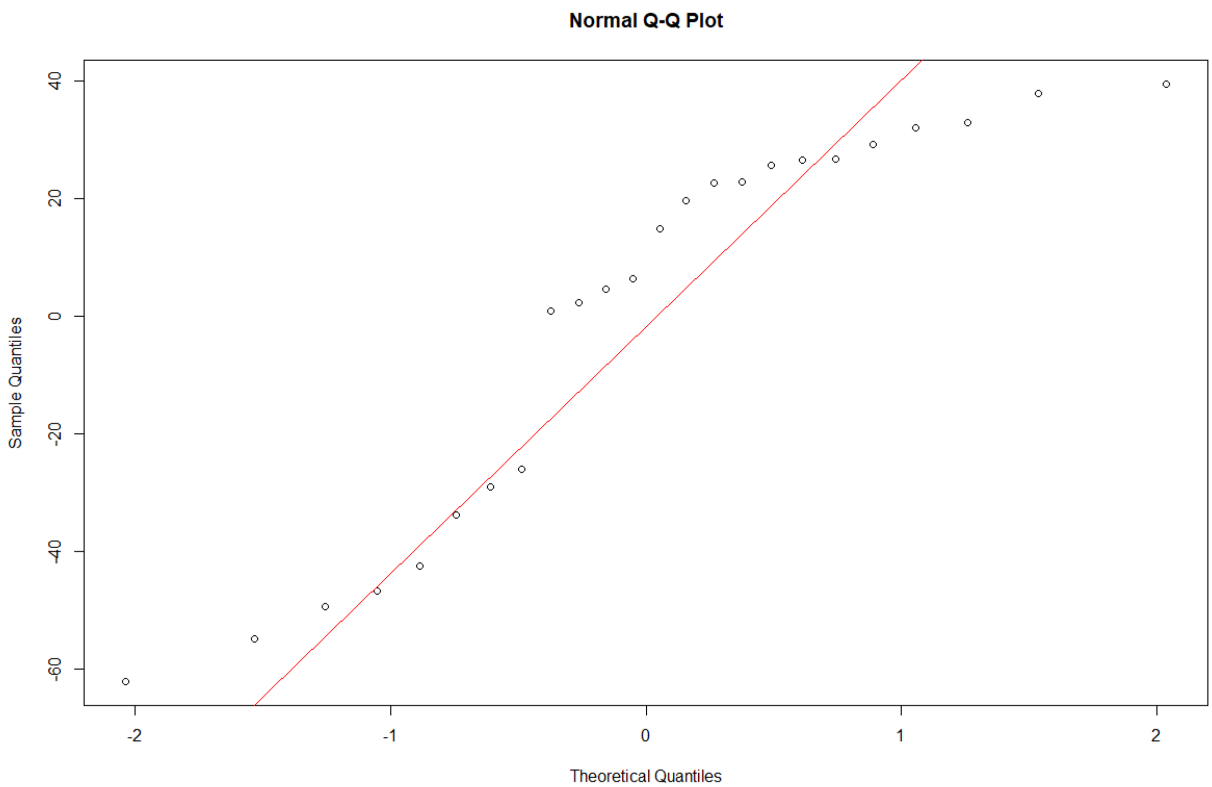
R-Code:

```
##### Normality assumption check of the residuals with Shapiro-Wilks test.
qqnorm(aov.out$residuals); qqline(aov.out$residuals,col="red")
shapiro.test(aov.out$residuals);

##### Homogeneity of variance assumption check with Levene's test.
library(car);
leveneTest(yij~facA, data=dat);
leveneTest(yij~facB, data=dat);
plot(aov.out$fitted.values,aov.out$residuals);abline(h=0,col="red",lwd=2);
```

R-Result:

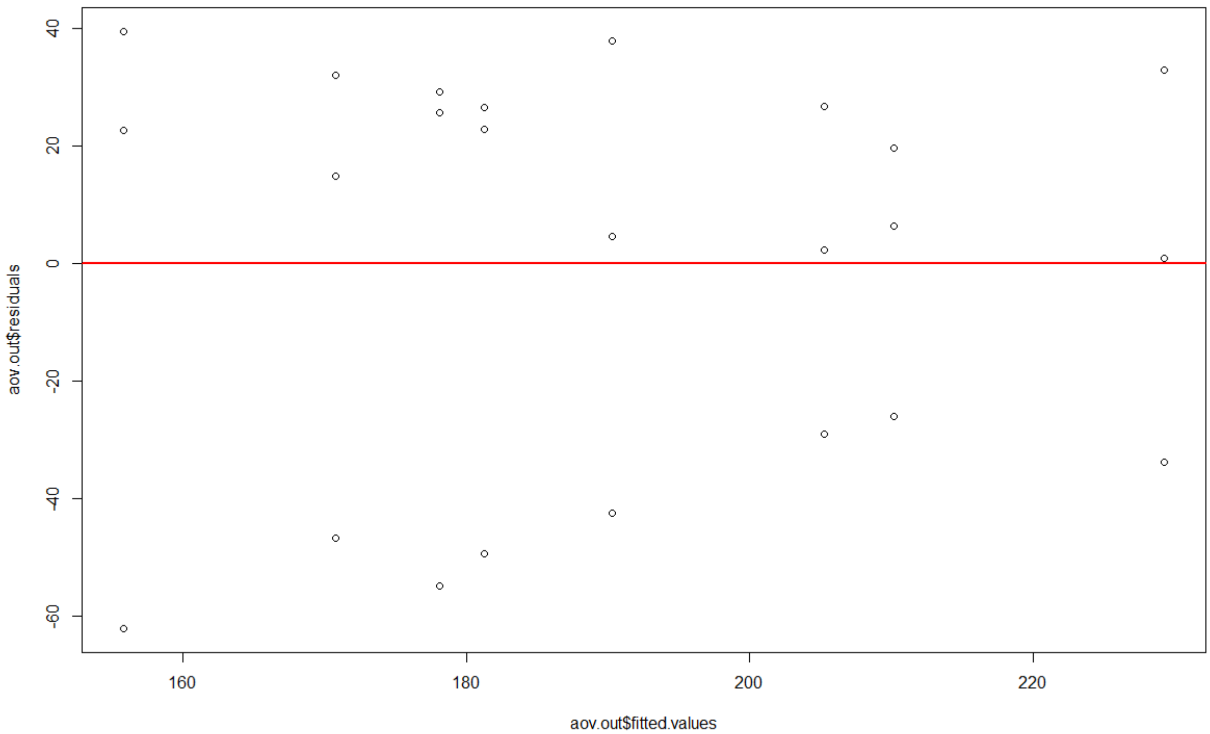
```
> ##### Normality assumption check of the residuals with Shapiro-Wilks test
.
> qqnorm(aov.out$residuals); qqline(aov.out$residuals,col="red")
> shapiro.test(aov.out$residuals);

        Shapiro-Wilk normality test

data:  aov.out$residuals
W = 0.87413, p-value = 0.006344

>
> ##### Homogeneity of variance assumption check with Levene's test.
> library(car);
> leveneTest(yij~facA, data=dat);
Levene's Test for Homogeneity of Variance (center = median)
      Df F value Pr(>F)
group  1  0.5418 0.4695
      22
> leveneTest(yij~facB, data=dat);
Levene's Test for Homogeneity of Variance (center = median)
      Df F value Pr(>F)
group  3  0.0527 0.9836
      20
> plot(aov.out$fitted.values,aov.out$residuals);abline(h=0,col="red",lwd=2)
;

>
```

- The normality plot:



- Observation: From the normal Q-Q plot we can conclude a roughly normal distribution due to the symmetric distribution about the straight line. But distribution has very flat tails as the dotted points tends to move too far away (due to outliers) from the straight line at the top and bottom.
- From the Shapiro's test at alpha = 0.10, we reached the conclusion of non-normality with p-value = 0.006344.
- For the purpose of this experiment (and the robustness to outliers) we will be more liberal and assume normality and nonetheless, perform the Turkey's pairwise comparison test as well.

❖ The residual vs fitted value:



❖ Observation: From the overall scattered nature of the fitted values vs residuals we can conclude a passed homogeneity of variance's assumption test.

❖ This conclusion is supported by both p-values obtained via the Levene's test at alpha = 0.10:
- ✓ 0.4695
- ✓ 0.9836

F | Pairwise comparisons

R-Code:

```
##### Pairwise comparisons with TukeyHSD test.
TukeyHSD(aov.out,"facA");
TukeyHSD(aov.out,"facB");
```

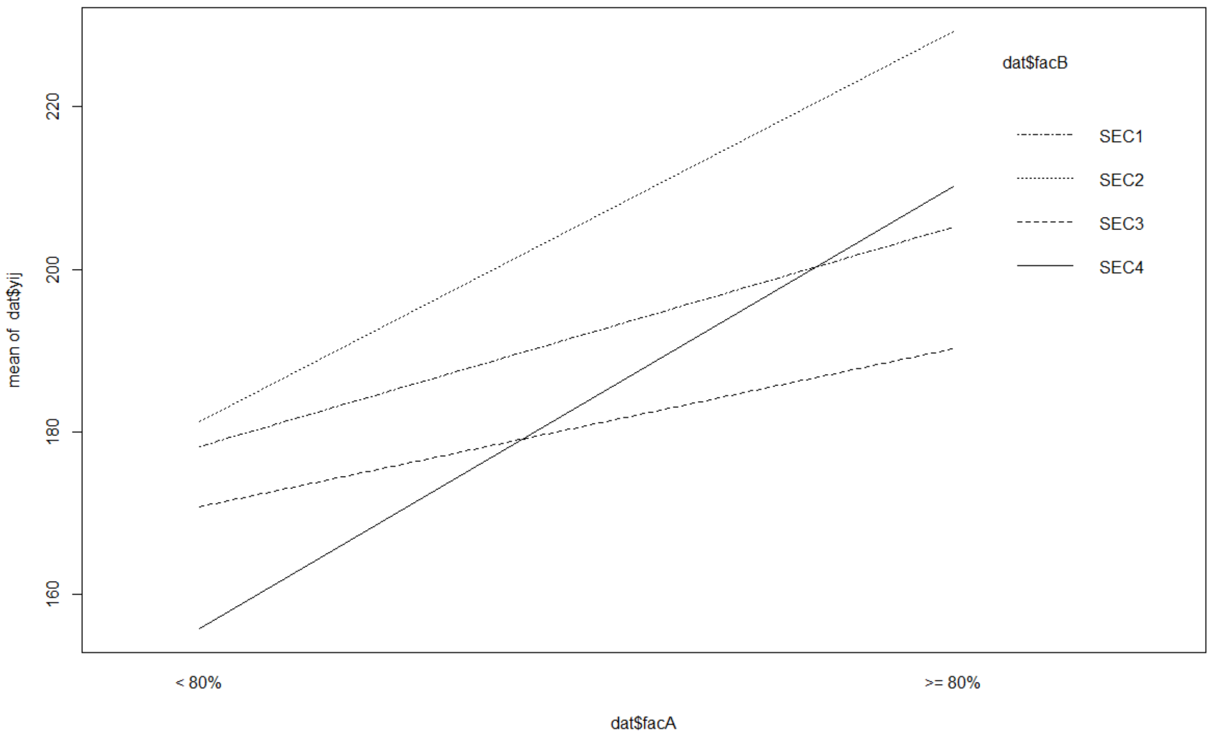
R-Result:

```
> ##### Pairwise comparisons with TukeyHSD test.
> TukeyHSD(aov.out,"facA");
  Tukey multiple comparisons of means
    95% family-wise confidence level

Fit: aov(formula = yij ~ facA + facB + facA * facB, data = dat)

$`facA`
                diff      lwr      upr     p adj
>= 80%-< 80% 37.1925 2.519335 71.86567 0.0370953
```

```
> TukeyHSD(aov.out,"facB");
  Tukey multiple comparisons of means
    95% family-wise confidence level

Fit: aov(formula = yij ~ facA + facB + facA * facB, data = dat)

$`facB`
                diff       lwr      upr     p adj
SEC2-SEC1  13.555000 -52.62289 79.73289 0.9348585
SEC3-SEC1 -11.158333 -77.33622 55.01955 0.9619061
SEC4-SEC1  -8.705000 -74.88289 57.47289 0.9811858
SEC3-SEC2 -24.713333 -90.89122 41.46455 0.7128849
SEC4-SEC2 -22.260000 -88.43789 43.91789 0.7721766
SEC4-SEC3   2.453333 -63.72455 68.63122 0.9995540
```

- Comments: At a significance level of 0.10 we can deduce the following:
- There is a significant difference between the attendance rates “>=80%” and “<80%”.
- There isn’t any significant difference across sections.

| Interaction Graph



- The interaction graph shows more interaction between section 4 and sections 2, 3.

Conclusion

Following the analyses at a significance level of 0.10 we reached the following conclusions.

- There is a significant effect of attendance rate on the mean performance of the students.
- There is no significant effect of the sections on the mean performance of the students.

Furthermore, in order to guarantee a uniformly better mean performance of the student in Math 099 course, the students should observe a min attendance rate of 80%.

| Problems Encountered

- The presence of outliers:
  - ✓ Possible fix: The retrieval of the outlier point will lead to a more normal plot, easier to interpret, and exactly the same significant effects will be identified when the full set of experimental data is used.

Raw Data:

Section	Att 1	Test 1	Att 2	Test 2	Att 3	Test 3
Crosslisted Section 45059	0%	0	0%	0	0%	0
Crosslisted Section 45059	50%	168	0%	0	0%	0
Crosslisted Section 45059	50%	237	29%	153	13%	0
Crosslisted Section 45059	67%	258	43%	230	38%	80
Crosslisted Section 45059	67%	264	71%	275	50%	200
Crosslisted Section 45059	83%	195	57%	211	63%	120
Crosslisted Section 45059	67%	126	71%	99	63%	110
Crosslisted Section 45059	83%	255	71%	270	63%	200
Crosslisted Section 45059	100%	195	86%	189	63%	120
Crosslisted Section 45059	100%	72	100%	45	63%	60
Crosslisted Section 45059	83%	183	71%	171	75%	80
Crosslisted Section 45059	100%	198	100%	279	75%	140
Crosslisted Section 45059	83%	213	57%	221	88%	100
Crosslisted Section 45059	67%	213	86%	171	88%	130
Crosslisted Section 45059	100%	261	86%	288	88%	250
Crosslisted Section 45059	100%	249	100%	266	88%	160
Crosslisted Section 45059	83%	213	100%	275	88%	220
Crosslisted Section 45059	17%	219	86%	252	100%	200
Crosslisted Section 45059	83%	249	86%	270	100%	220
Crosslisted Section 45059	67%	174	100%	284	100%	130
Crosslisted Section 45060	83%	147	86%	238	13%	0
Crosslisted Section 45060	100%	267	100%	288	13%	0
Crosslisted Section 45060	67%	192	71%	279	25%	0
Crosslisted Section 45060	33%	273	57%	225	38%	220
Crosslisted Section 45060	67%	150	57%	198	50%	30
Crosslisted Section 45060	67%	216	86%	225	50%	115
Crosslisted Section 45060	83%	225	86%	300	50%	250
Crosslisted Section 45060	83%	189	71%	194	63%	95
Crosslisted Section 45060	83%	273	71%	248	63%	155
Crosslisted Section 45060	83%	180	71%	81	63%	110
Crosslisted Section 45060	83%	210	86%	252	75%	80
Crosslisted Section 45060	83%	209	86%	288	88%	245
Crosslisted Section 45060	100%	267	86%	284	88%	190
Crosslisted Section 45060	100%	198	86%	257	88%	90
Crosslisted Section 45060	100%	300	100%	288	88%	260
Crosslisted Section 45060	100%	216	100%	234	88%	140
Crosslisted Section 45060	83%	192	86%	216	100%	100

Crosslisted Section 45060	100%	264	100%	306	100%	270
Crosslisted Section 45060	100%	213	100%	194	100%	130
Crosslisted Section 45060	100%	219	100%	279	100%	290
Crosslisted Section 45060	100%	300	100%	266	100%	265
Crosslisted Section 45060	100%	273	100%	279	100%	170
Crosslisted Section 45061	0%	0	0%	0	0%	0
Crosslisted Section 45061	33%	162	0%	0	0%	0
Crosslisted Section 45061	83%	234	57%	103.5	0%	0
Crosslisted Section 45061	67%	291	43%	306	25%	270
Crosslisted Section 45061	50%	90	43%	144	38%	80
Crosslisted Section 45061	33%	234	57%	189	38%	70
Crosslisted Section 45061	100%	186	100%	198	63%	0
Crosslisted Section 45061	67%	237	86%	270	75%	170
Crosslisted Section 45061	83%	210	86%	225	75%	130
Crosslisted Section 45061	83%	45	86%	108	75%	10
Crosslisted Section 45061	83%	147	100%	234	75%	95
Crosslisted Section 45061	100%	261	100%	252	75%	185
Crosslisted Section 45061	100%	168	100%	153	75%	80
Crosslisted Section 45061	100%	234	100%	216	75%	150
Crosslisted Section 45061	83%	243	100%	252	88%	180
Crosslisted Section 45061	100%	225	100%	270	88%	100
Crosslisted Section 45061	100%	204	100%	288	100%	190
Crosslisted Section 45061	100%	90	100%	162	100%	80
Crosslisted Section 45061	100%	243	100%	261	100%	170
Crosslisted Section 45061	100%	192	100%	198	100%	180
Crosslisted Section 45061	100%	159	100%	207	100%	100
Crosslisted Section 45061	100%	270	100%	306	100%	240
Crosslisted Section 45061	100%	201	100%	279	100%	90
Crosslisted Section 45062	0%	0	0%	0	0%	0
Crosslisted Section 45062	0%	243	0%	297	13%	65
Crosslisted Section 45062	17%	0	0%	0	0%	0
Crosslisted Section 45062	17%	186	0%	126	50%	80
Crosslisted Section 45062	33%	48	43%	0	0%	0
Crosslisted Section 45062	33%	156	43%	27	13%	0
Crosslisted Section 45062	33%	129	57%	252	63%	90
Crosslisted Section 45062	33%	243	71%	270	75%	110
Crosslisted Section 45062	33%	270	86%	297	75%	240
Crosslisted Section 45062	50%	255	29%	198	13%	0
Crosslisted Section 45062	50%	117	86%	162	75%	50
Crosslisted Section 45062	50%	162	71%	198	88%	80
Crosslisted Section 45062	67%	141	43%	90	38%	0

Crosslisted Section 45062	67%	192	86%	221	63%	55
Crosslisted Section 45062	83%	153	86%	117	75%	60
Crosslisted Section 45062	83%	231	57%	252	88%	170
Crosslisted Section 45062	83%	210	71%	243	100%	190
Crosslisted Section 45062	83%	231	86%	270	100%	200
Crosslisted Section 45062	83%	240	86%	248	100%	220
Crosslisted Section 45062	83%	234	86%	293	100%	245

		Section											
		SEC1 - 45059			SEC2 - 45060			SEC2 - 45061			SEC3 - 45062		
		T1	T2	T3	T1	T2	T3	T1	T2	T3	T1	T2	T3
Att	Sample from < 80%	219	153	80	273	225	220	162	144	270	243	297	65
		168	230	200	150	198	30	234	306	80	186	126	80
		237	211	120	192	279	115	90	189	70	48	198	90
		258	221	110	216	194	250	237	103.5	170	156	27	55
		264	99	200		248	95	291		130	129	90	110
		126	171	120		81	155			10	270	252	240
		213	270	60			110			95	243	252	50
		174	275	80			80			185	255	270	60
				140						80	117	198	
										150	162	243	
											141		
											192		
	Sample Mean	**207.38**	**203.75**	**123.33**	**207.75**	**204.17**	**131.88**	**202.80**	**185.63**	**124.00**	**178.50**	**195.30**	**93.75**
	Sample from >= 80%	195	171	100	189	225	245	243	270	180	153	297	170
		255	189	130	209	288	190	234	225	100	231	162	80
		183	252	250	147	238	90	147	108	190	210	221	190
		213	270	160	273	216	260	210	252	80	231	117	200
		213	288	220	192	252	140	45	234	170	240	270	220
		249	45	200	210	300	100	204	288	180	234	248	245
		195	266	220	180	284	270	261	252	100		293	
		72	275	130	225	257	130	90	162	240			

		198	279		300	288	290	243	261	90			
		261	284		216	234	265	192	198				
		249			267	288	170	159	207				
					267	306		225	270				
					198	194		270	306				
					264	279		168	153				
					213	266		186	198				
					219	279		234	216				
					300			201	279				
					273								
	Sample Mean	**207.55**	**231.90**	**176.25**	**230.11**	**262.13**	**195.45**	**194.82**	**228.18**	**147.78**	**216.50**	**229.71**	**184.17**