

Analysis of Variance of Psychiatric Clinician Productivity  
At a Community Mental Health Center

Daniel E. Diamond  
8/3/19

## **Summary**

Data for this observational study was obtained from the author's employer, a Community Mental Health Center in the state of Indiana. Productivity data for clinical staff from the second half of the organization's most recent fiscal year (1/1/19 - 6/30/19) were analyzed to glean insight about the impact of both geographic location and staff level of experience on staff productivity. For the purposes of this study, some employees (such as those represented by multiple observations in the data set due to having transferred from one location to another during the date range) were excluded. Though observational, this study is akin to a 2-factor factorial design, with completely randomized selection. See Table 1 for a description of variables.

Variable	Meaning	Type	Scale/Levels
adj_prod (response variable)	Adjusted productivity of clinicians for the given time period, factoring in term of employment, vacation time, etc.	Continuous numeric	Floating point values representing the percentage of expected productivity the employee achieved: ( $\text{prod}_{\text{actual}} / \text{prod}_{\text{expected}}$ )
exp	Clinician level of experience.	Categorical, fixed	Assumes values: 0 when the employee has been employed less than 2 years, 1 when the employee has been employed between 2-4 years, 2 when the employee has been employed longer than 4 years.
central	Geographic location, indicating whether the clinician worked primarily in a 'central' (more populous) region, or whether they worked at a so-called 'outlying' location in a more rural setting.	Categorical, fixed	Assumes values: 0 when employee worked in rural setting, 1 when the employee worked in a central setting.

*Table 1 - Variable Summary*

## **Introduction**

The Community Mental Health Center employs approximately 250 clinicians, including case managers, therapists, psychologists, psychiatrists, and other clinical specialists who provide the communities they serve with psychiatric services, ranging from medication management to individual therapy to case management and employment services. To ensure the community is adequately served, the organization must make every effort to operate efficiently and make

effective use of limited resources. Clinicians, depending on their specific role, are given a ‘productivity expectation’ which is a target number of hours of service which they are to provide. It is the clinician’s responsibility, along with their supervisor, to then manage their caseload, ensuring the load is small enough that each client is properly served, but large enough to ensure the clinician meets their productivity expectations.

There are a number of factors which some clinicians suspect have significant bearing on whether they are able to consistently meet these expectations. Two of the more commonly suspected factors include the level of experience of the clinician and the geographic region of the clinician’s client base. It is suspected that clinicians with more experience have developed the tools and acquired the expertise to operate more efficiently and are thus at an advantage in meeting their expectation. Three such levels of experience are considered in this study. It is also suspected by some clinicians that working in more rural geographic regions (referred to as the ‘outlying branches’) where drifting snow can lead to increased appointment cancelations during the winter, and where presumably more travel is required for clinicians who do client home visits, puts clinicians at a disadvantage compared to those clinicians who work in the more populated ‘central’ service area, where more resources are available. These intuitions were tested utilizing the following hypotheses:

- 1) Mean productivity for central vs. outlying regions:

$$H_0: \mu_{cent} = \mu_{out}$$

$H_A$ : *mean productivity is not the same in populated regions and rural regions*

- 2) Mean productivity based on clinician experience level:

$$H_0: \mu_{exp1} = \mu_{exp2} = \mu_{exp3}$$

$H_A$ : *mean productivity is not the same for all levels of experience*

- 3) Interaction effects:

$$H_0: \text{there is no interaction between experience and regional factors}$$

$H_A$ : *an interaction does exist between experience and regional factors*

## **Methods**

The organization’s raw productivity data, after the removal of anomalous cases, contained 229 observations, distributed across the 6 possible treatment combinations as depicted in Table 2.

Treatment Combination	Number of Observations
Rural location, < 2 yrs experience	25
Rural location, 2-4 yrs experience	51
Rural location, > 4 yrs experience	32
Central location, < 2 yrs experience	50
Central location, 2-4 yrs experience	26
Central location, > 4 yrs experience	45

*Table 2 - Distribution of Raw Data across Treatments*

Minitab was employed to randomly select 25 observations from each (Region, Experience Level) treatment combination. For each treatment combination, each observation was assigned a sequential employee number (emp\_num). From this, a random index (r\_index) was derived. The observations for each treatment combination were then sorted by this random index and the first 25 observations were selected, as depicted in Figure 1.

↓	C1	C2	C3	C4	C5
	emp_num	central	exp	adj_prod	r_index
1	21	1	0	1.03050	1
2	45	1	0	1.87223	2
3	48	1	0	1.44240	3
4	30	1	0	0.34380	4
5	29	1	0	1.18790	5
6	3	1	0	0.63000	6
7	23	1	0	1.26690	7
8	9	1	0	1.30725	8
9	2	1	0	0.99970	9
10	4	1	0	0.73440	10
11	5	1	0	0.76880	11
12	44	1	0	1.56967	12

*Figure 1 - Random Selection of Productivity Observations - Centrally Located, < 2yrs Experience*

The resultant data set included 150 observations, which were re-indexed for reference, 25 randomly selected observations for each of the 6 treatment combinations. See Appendix B for the resulting data set.

Initial exploratory data analysis was performed in Excel and SAS. See Appendix A for SAS code. To evaluate assumptions for employing ANOVA, a residual analysis was performed in SAS, examining whether residuals were normally distributed, had a mean of 0, were independent

and had equal variance among treatment levels. Mean comparisons were performed, also in SAS, employing the Tukey method to compare factor levels.

Both factors are of interest here, and it was not known whether there were interaction effects, so the initial model considered was:

$$Y_{ij} = \mu_{..} + \alpha_i + \beta_j + (\alpha\beta)_{ij} + \epsilon_{ijk}$$

$Y$  = clinician productivity tabulation

$\alpha$  = effect of geographic region

$\beta$  = effect of clinician level of experience

$(\alpha\beta)$  = interaction effect

$\epsilon$  = error

Analysis showed, however, there was not significant interaction between these factors and so the interaction term was dropped from the model, reducing to:

$$Y_{ij} = \mu_{..} + \alpha_i + \beta_j + \epsilon_{ijk}$$

See Results section for specifics about interaction effects.

## **Results**

To get an initial impression of the effects of these factors on mean clinician productivity, a basic exploratory data analysis was completed in SAS. PROC CONTENTS and PROC MEANS was performed. See Figures 2 and 3, respectively.

Variables in Creation Order			
#	Variable	Type	Len
1	emp_num	Num	8
2	central	Num	8
3	exp	Num	8
4	adj_prod	Num	8

Figure 2 - PROC CONTENTS

Analysis Variable : adj_prod							
central	exp	N Obs	N	Mean	Std Dev	Minimum	Maximum
0	0	25	25	1.16	0.35	0.48	1.97
	1	25	25	1.04	0.19	0.74	1.40
	2	25	25	0.85	0.30	0.11	1.49
1	0	25	25	1.10	0.39	0.34	1.87
	1	25	25	0.93	0.45	0.32	2.02
	2	25	25	0.78	0.31	0.30	1.59

Figure 3 - PROC MEANS

Mean values for each of the 6 treatment combinations were graphed in Excel, see Figure 4.

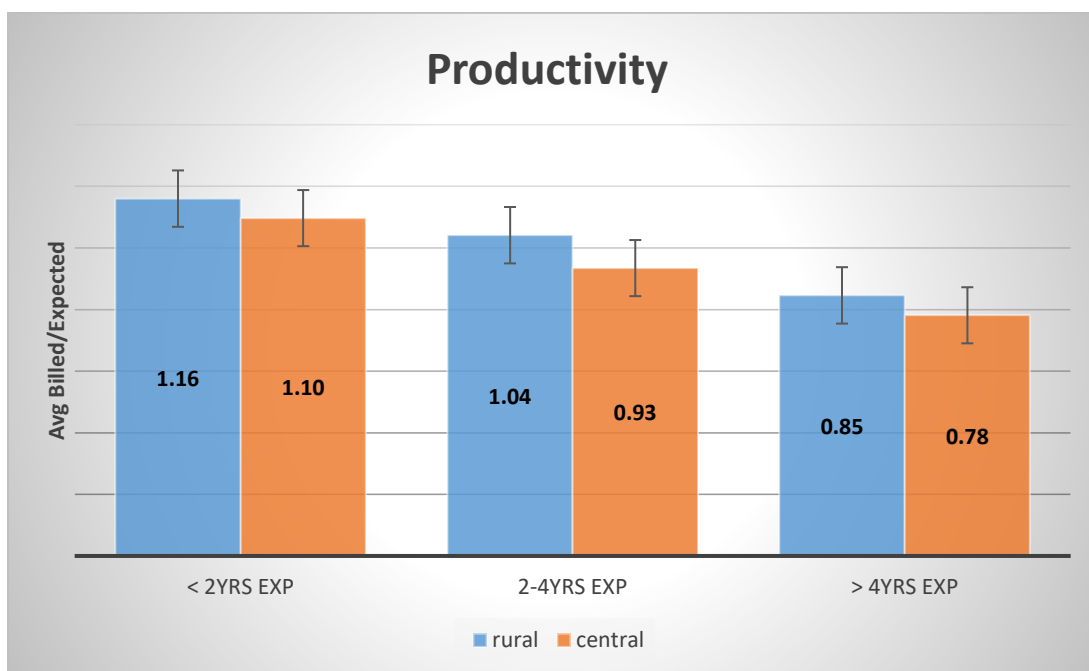


Figure 4 - Exploratory Mean Comparison with Standard Errors

From this preliminary visualization, we see that if any statistically significant relationships exist, they appear to run contrary to clinician intuition. It would appear that clinicians from centrally located, more populous regions actually appear to produce slightly less than their rurally stationed counterparts, and it appears that average productivity appears to deteriorate over time for both groups.

To examine the effect of the interaction between geographic region and level of experience of the clinician, an ANOVA was performed in SAS including the interaction term.

Type 3 Tests of Fixed Effects				
Effect	Num DF	Den DF	F Value	Pr > F
central	1	144	1.95	0.1648
exp	2	144	10.56	<.0001
central*exp	2	144	0.07	0.9370

Figure 5 - ANOVA with Interaction Term

We see from Figure 5 that the interaction effect has an F-value of 0.07 and a p-value of 0.9370, so the interaction between central location and clinician experience is not significant. This 2-Factor Factorial model is additive, and the interaction term will be dropped from the model.

The ANOVA was then run in SAS without the interaction term. Output pertaining to residual analysis appears in Figures 6-8.

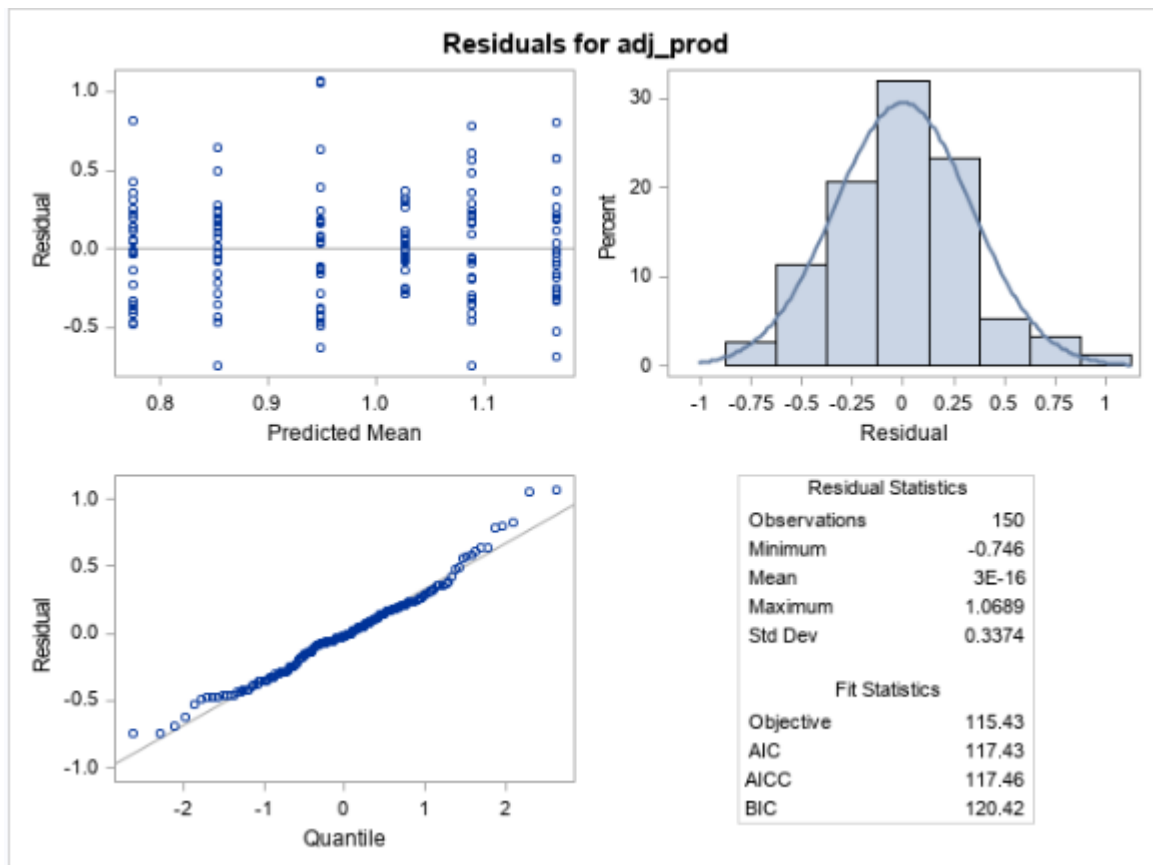


Figure 6 - Residual Analysis

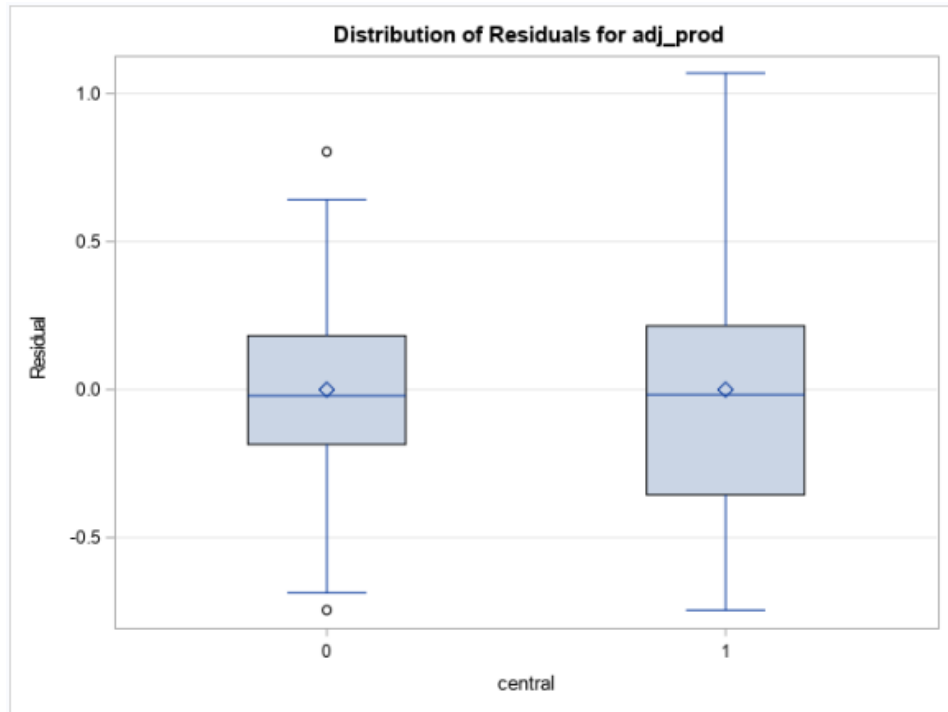


Figure 7 - Residual Distribution for Geo Region

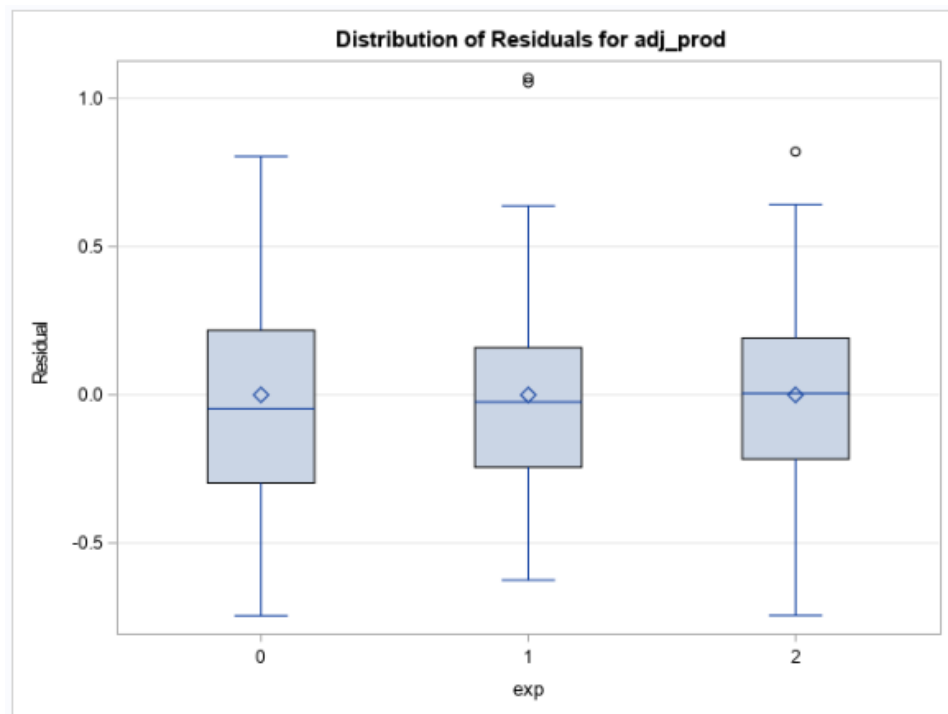


Figure 8 - Residual Distribution for Experience of Clinician



We see that these are well-behaved residuals, normally distributed about zero, with roughly equal variance, and no structure in the residual plot.

The results of the ANOVA are presented in Figure 9.

Type 3 Analysis of Variance								
Source	DF	Sum of Squares	Mean Square	Expected Mean Square	Error Term	Error DF	F Value	Pr > F
central	1	0.229511	0.229511	Var(Residual) + Q(central)	MS(Residual)	146	1.98	0.1620
exp	2	2.486564	1.243282	Var(Residual) + Q(exp)	MS(Residual)	146	10.70	<.0001
Residual	146	16.965069	0.116199	Var(Residual)	-	-	-	-

Figure 9 - ANOVA Result

The results of the Tukey means comparison and groupings appear in Figures 10-13.

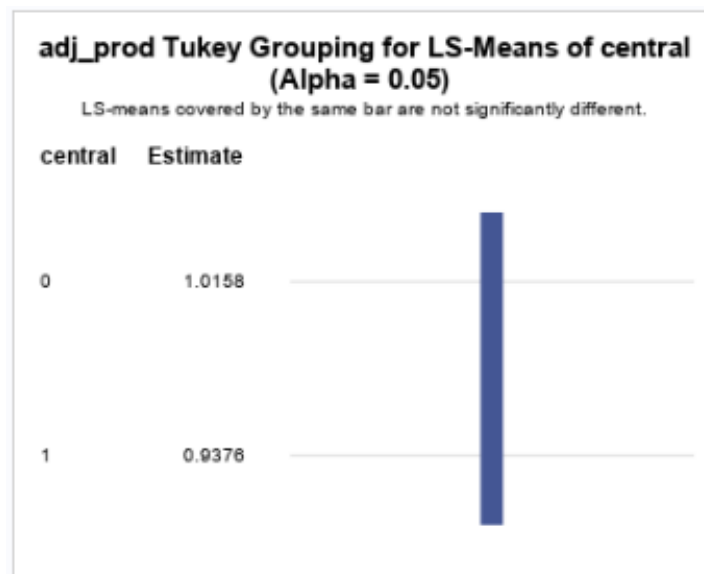


Figure 10 - Tukey Grouping - Geo Region

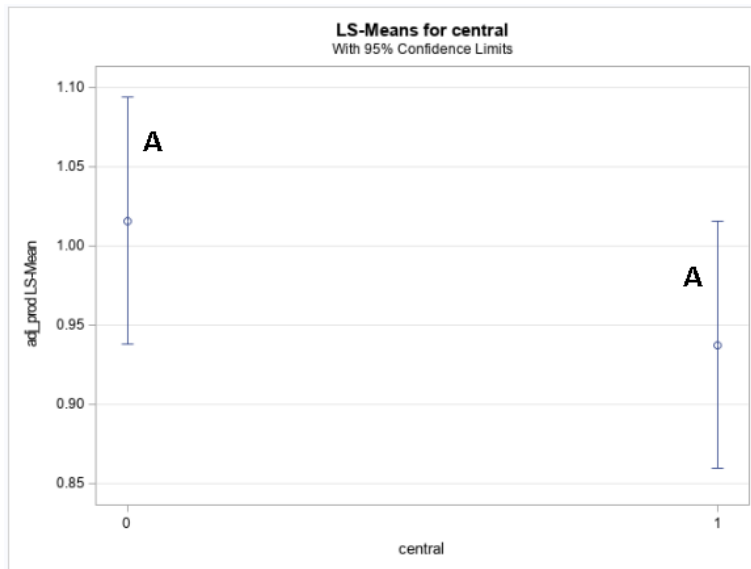


Figure 11 - Means Plot with Tukey Grouping - Geo Region

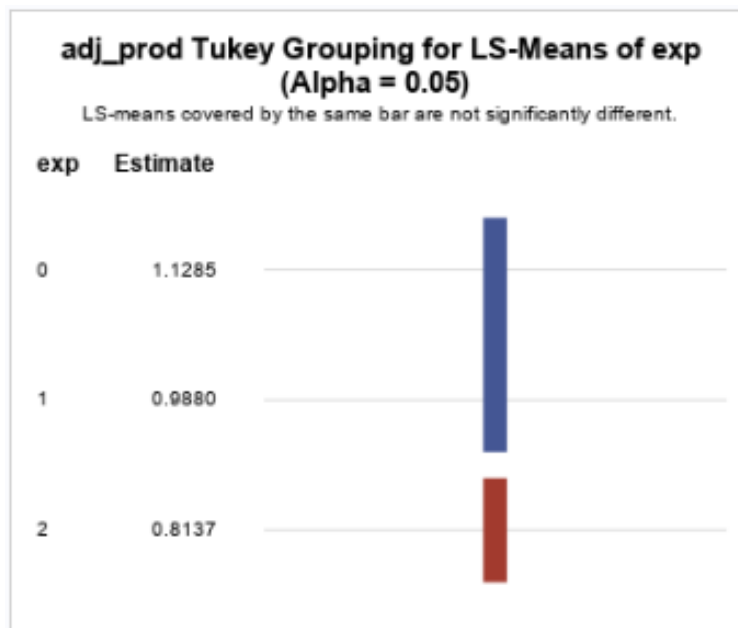


Figure 12 - Tukey Grouping -Clinician Experience

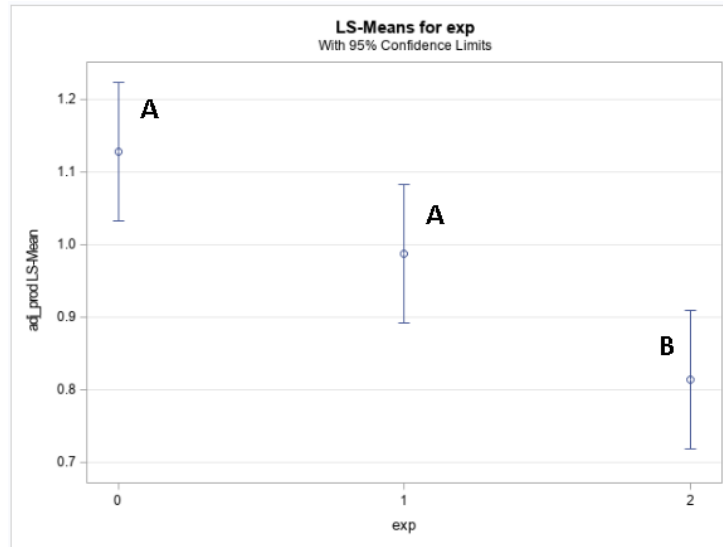


Figure 13 - Means Plot with Tukey Grouping - Clinician Experience

Results for Hypotheses 1-3 at significance level 0.05, and Tukey Means Comparison interpretation:

- 1) With an F-value of 1.98 and a p-value of 0.1620, we fail to reject the null hypothesis and conclude that mean clinician productivity is statistically the same for both centrally located clinicians and those located in outlying, rural areas.
- 2) With an F-value of 10.70 and a p-value of <.0001, we reject the null hypothesis and conclude that mean clinician productivity is not statistically the same for all levels of clinician experience. The Tukey means comparison shows no statistical difference between clinicians of less than 2 years of experience and those with 2-4 years of experience, but that clinicians with more than 4 years of experience show statistically significantly lower mean productivity than the other two groups.
- 3) As noted above, with an F-value of 0.07 and a p-value of 0.9370, we fail to reject the null hypothesis and conclude there is no interaction effect between geographic region and clinician level of experience which affects clinician productivity.

## **Conclusion**

Contrary to clinician intuition, the effect of geographic location on clinician productivity is not significant. While clinician experience does appear to have a statistically significant effect on productivity, it is *in the opposite direction* of what the clinicians suspected, with more

experienced staff actually showing lower mean productivity than those clinicians of lesser experience.

Note that this study does not take into consideration confounding factors which are likely significant, such as the characteristics of employees who prefer to live in rural areas and of those who have a tendency to stay at jobs longer than others. A more thorough examination of these factors is required before conclusions can be drawn about the effects of geographic region and level of clinician experience on clinician productivity, but analysis of this data set's variance does not support clinician intuition.

### **References/Sources**

- Statistical procedures were taken from:  
<https://newonlinecourses.science.psu.edu/stat502>
- Data was provided by Valley Oaks Health, Inc. Community Mental Health Center

### **Appendix A - Code**

```
/*  
    proj.sas  
  
    Dan Diamond  
    STAT 502 - ANOVA/DOE  
    Summer 2019  
  
    Analysis of Variance of clinician productivity data.  
*/  
  
/* read in 25 randomly selected observations for each of 6  
   combinations of treatments. */  
data prod;  
    input emp_num central exp adj_prod;  
  
datalines;  
1      0      0      1.009  
2      0      0      1.1381  
3      0      0      0.9192  
.  
.  
.  
[abbreviated]  
.  
.  
.  
148    1      2      0.3584
```

```

149    1    2    0.7577
150    1    2    1.5949
;
run;

* display data;
proc print data=prod;
title 'Raw Productivity Data';
run;

* exploratory examination of data set;
proc contents data=prod position;
title 'Variable Summary';
run;

proc means data=prod fw=8 maxdec=2;
class central exp;
var adj_prod;
title 'Exploratory Data Analysis';
run;

ods graphics on;

* create model to evaluate interaction term;
proc mixed data=prod method=type3 plots=all;
class central exp;
model adj_prod = central exp central*exp;
store prod_mu;
title 'ANOVA of Productivity Data';
run;

ods html style=statistical sge=on;

* revised model without interaction term;
proc mixed data=prod method=type3 plots=all;
class central exp;
model adj_prod = central exp;
store prod_mu;
title 'ANOVA of Productivity Data';
run;

* mean comparison without interaction;
proc plm restore=prod_mu;
lsmeans central exp /_adjust=tukey plot=meanplot cl lines;
ods exclude diffplot;
run; title; run;

```

## **Appendix B - Data**

emp_num	central	exp	adj_prod
1	0	0	1.009

2	0	0	1.1381
3	0	0	0.9192
4	0	0	1.0658
5	0	0	0.8685
6	0	0	1.132
7	0	0	1.3494
8	0	0	1.1574
9	0	0	1.745625
10	0	0	1.53165
11	0	0	0.8965
12	0	0	1.3843
13	0	0	0.6369
14	0	0	1.97145
15	0	0	1.0859
16	0	0	0.8424
17	0	0	0.4809
18	0	0	0.8469
19	0	0	0.9192
20	0	0	1.2018
21	0	0	0.9826
22	0	0	1.748475
23	0	0	1.367775
24	0	0	1.4358
25	0	0	1.2847
26	0	1	0.9699
27	0	1	1.4011
28	0	1	0.9831
29	0	1	1.1393
30	0	1	1.0251
31	0	1	0.7603
32	0	1	1.0778
33	0	1	0.969
34	0	1	0.7466
35	0	1	1.3254
36	0	1	1.34745
37	0	1	1.0817
38	0	1	1.0226
39	0	1	0.7828
40	0	1	1.2917
41	0	1	0.9703
42	0	1	1.0486

43	0	1	0.9753
44	0	1	1.1198
45	0	1	0.9626
46	0	1	0.7403
47	0	1	1.1035
48	0	1	0.893
49	0	1	0.9459
50	0	1	1.3524
51	0	2	1.0236
52	0	2	0.8695
53	0	2	0.4142
54	0	2	1.0925
55	0	2	0.7768
56	0	2	0.7827
57	0	2	0.5668
58	0	2	0.5016
59	0	2	1.0894
60	0	2	0.6365
61	0	2	0.7864
62	0	2	0.1082
63	0	2	0.8194
64	0	2	0.8323
65	0	2	1.131
66	0	2	1.349
67	0	2	0.3882
68	0	2	1.0447
69	0	2	0.9999
70	0	2	0.9561
71	0	2	0.6909
72	0	2	0.8712
73	0	2	0.9282
74	0	2	0.9965
75	0	2	1.4949
76	1	0	1.0305
77	1	0	1.872225
78	1	0	1.4424
79	1	0	0.3438
80	1	0	1.1879
81	1	0	0.63
82	1	0	1.2669
83	1	0	1.30725

84	1	0	0.9997
85	1	0	0.7344
86	1	0	0.7688
87	1	0	1.569675
88	1	0	1.6954
89	1	0	1.6518
90	1	0	0.6337
91	1	0	0.7336
92	1	0	1.2547
93	1	0	1.3329
94	1	0	0.6746
95	1	0	1.0192
96	1	0	1.3028
97	1	0	0.8967
98	1	0	0.9045
99	1	0	1.3758
100	1	0	0.792
101	1	1	2.0025
102	1	1	0.4816
103	1	1	0.3238
104	1	1	0.4557
105	1	1	0.5676
106	1	1	0.4789
107	1	1	0.8412
108	1	1	0.8206
109	1	1	0.791
110	1	1	1.1176
111	1	1	1.0247
112	1	1	0.5272
113	1	1	1.1076
114	1	1	0.5128
115	1	1	0.9856
116	1	1	1.1876
117	1	1	1.0323
118	1	1	1.1326
119	1	1	0.6604
120	1	1	1.0003
121	1	1	1.3391
122	1	1	0.5578
123	1	1	1.586475
124	1	1	0.8124



125	1	1	2.017775
126	1	2	1.1318
127	1	2	0.7517
128	1	2	0.8196
129	1	2	0.7452
130	1	2	0.8378
131	1	2	0.8907
132	1	2	0.3
133	1	2	0.634
134	1	2	0.9896
135	1	2	1.0878
136	1	2	1.0249
137	1	2	0.542
138	1	2	0.9734
139	1	2	0.7415
140	1	2	0.7668
141	1	2	0.9158
142	1	2	0.3074
143	1	2	0.9159
144	1	2	0.4158
145	1	2	0.4387
146	1	2	0.3893
147	1	2	1.2038
148	1	2	0.3584
149	1	2	0.7577
150	1	2	1.5949