

David Simpson

Dr. Donald Green

W 4768 Experimental Research

26 March 2019

### Do the Seeds Matter? – An Experiment with Guacamole and Jalapenos

**Introduction:** I enjoy making guacamole for dinner parties. I use a recipe that includes jalapeño peppers.<sup>1</sup> I make three recipe variants: low-spice, moderate, and spicy. Variants differ by the included jalapeño parts. Jalapeño parts are (1) the body, the outside flesh; (2) the seeds; and (3) the pith, the white to light green inside flesh where the seeds are attached. The low spice variant requires removing the jalapeño seeds and pith. To make the moderate variant, I remove the seeds and only add the pith and body to the other guacamole ingredients. The spicy variant uses all three parts of the jalapeño.

**Motivation:** From past observation, I am confident that the moderate and spicy variants are spicier than the low spice recipe. However, prior to testing I was unsure if a spice taste difference existed between the moderate and spicy variants. I find removing the seeds to be an annoying step in making the moderate recipe. Relatively speaking, it is easier to either remove the seeds and dice the body, or to just dice the entire pepper. Therefore, I designed an experiment to test whether the spicy and moderate variants differ in spice level. The null hypothesis in this experiment states, “Relative to the moderate recipe variant, inclusion of seeds in the spicy variant does not make a noticeable difference to the spice level.”

**Design:** The subjects are 30 guacamole taste observations. Control subjects are drawn from the moderate variant “control” guacamole. Treatment subjects are drawn from the spicy variant “treatment” guacamole. The two guacamoles differ only by the *treatment* status – the presence of seeds. The low-spice variant “baseline” guacamole was also created. Each batch originated from a single batch of all ingredients except for the jalapeños.<sup>2</sup>

Four individuals were recruited to serve as three guacamole taste-testers and one assistant in a double-blind experimental setting. For each observation, I supplied the assistant with a baseline guacamole serving followed by an experimental guacamole serving. The assistant then supplied each serving to a taste-tester. Experimental treatment or control servings were supplied using *complete randomization* within taste-tester block. Taste-testers rated each guacamole sample on a five-point scale for six ingredient categories and an overall taste category.<sup>3</sup> As such, taste-testers tasted 10 baseline-experimental pair observations and scored 20 guacamole samples. The main *outcome measure* of interest for each observation is the score used to rate the spice level of

---

<sup>1</sup> See the pre-analysis plan recipe in Appendix A1.

<sup>2</sup> See Appendix A2 for detailed pre-analysis plan food preparation information.

<sup>3</sup> See Appendix A3 for the score chart and further details on sample scoring.

the guacamole.<sup>4</sup> The spice score of the baseline serving is used as a “pretest” score for each observation to account for differences taste-tester spice sensitivity.

**Discussion:** The food preparation and serving procedures were designed to plausibly satisfy the core assumptions.<sup>5</sup> To satisfy *randomization*, I used complete randomization within three taste-tester blocks to allocate 5 control subjects and 5 treatment subjects to each taste-tester for outcome scoring. Taste-testers were told that they would rate 20 guacamole servings. Odd numbered servings were drawn from the baseline batch and even number serving were drawn according to treatment assignment.<sup>6</sup>

*Non-Interference* was the most difficult assumption to satisfy. Ideally, the experiment would use a unique taste-tester for each of the 30 subjects. With only three taste-testers, a non-interference violation could occur if a tester loses sensitivity to spice over subsequent tastings. To plausibly avoid non-interference, I implemented a palate reset procedure where testers must sip milk, sip water, and wait 1-minute before tasting each of the 20 guacamole samples. Taste testers were separated to prevent interference across blocks.

To satisfy *excludability*, I implemented the double-blind administration procedure discussed above. As such, there should not be a backdoor path from treatment assignment to score rating. No information was given to the assistant nor to the taste-testers regarding the batch identity of the 20 servings. No information was provided to suggest that spice level was the targeted outcome measure.

**Variable Distributions:** Table 1 presents mean and variance measures for four variables by treatment status and block.<sup>7</sup> *BaseSpice* is the baseline spice level score for each baseline-experimental pair. *Spicy* is the spice level outcome measure for treatment and control servings. *Overall* is the overall taste outcome measure for treatment and control servings. *SpicyDiff* is the spice level outcome measure scaled by *BaseSpice*. Figure 1 provides histograms of *BaseSpice*, *Spicy*, and *Overall* by treatment status and block.

**Results:** Table 2 presents relevant results from four models tested under randomization inference. Each model blocks on taste-tester. Model 1 estimates the ATE of seeds ( $z_i$ ) on spice score ( $Y_i$ ) using the difference-in-means estimator and includes the baseline spice score as a covariate ( $X_i$ ). Model 2 is a replication of Model 1; however, it does not include a covariate. Model 3 estimates the ATE using the difference-in-differences estimator. The pre-test spice score is subtracted from the outcome spice score to create a scaled outcome measure ( $Y_i - X_i$ ). Model 4 uses the overall taste score outcome ( $Y_i$ ) to estimate the ATE of seeds on overall taste.

The Model 1 ATE estimate is 0.5871. Thus, on a five-point spice scale, the addition of seeds causes treatment guacamole subjects to receive an average spice score rating that is 0.5871 points higher than that of control guacamole subjects. The standard error of the ATE estimate is

---

<sup>4</sup> Overall Taste scores was used as a secondary outcome measure.

<sup>5</sup> See Appendix A4 for pre-analysis plan information on serving samples.

<sup>6</sup> Appendix A5 gives the random assignment table. For clarity, Appendix A6 states the guacamole batch for each of the 20 samples supplied to each taste tester.

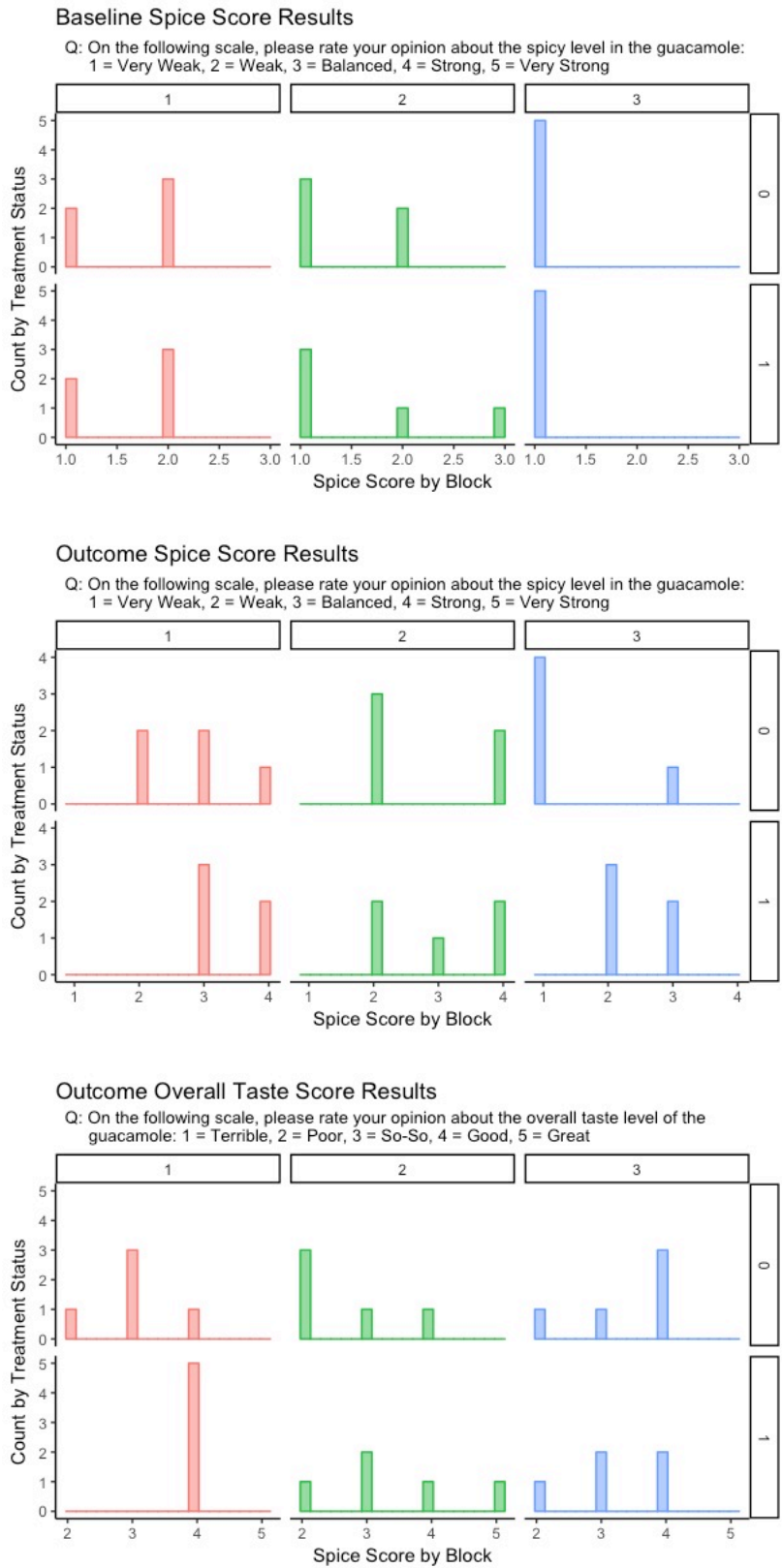
<sup>7</sup> Appendix A10 provides the complete data table.

the standard deviation of the sampling distribution. To calculate the standard error, I used randomization inference. The sharp null hypothesis states the average treatment effect for each subject is 0.5871. The resulting 95% confidence interval is (-0.0419, 1.2075). As such, this interval has a 95% chance of bracketing the true ATE in subsequent experiment replications.

**Table 1:** Summary Statistics

<b>Statistic</b>	<b>BaseSpice</b>	<b>Spicy</b>	<b>Overall</b>	<b>SpicyDiff</b>
<b>Block 1 Subjects</b>				
Mean (Control)	1.60	2.80	3.00	1.20
Var (Control)	0.24	0.56	0.40	0.56
Mean (Treatment)	1.60	3.40	4.00	1.80
Var (Treatment)	0.24	0.24	0.00	0.56
Mean (All)	1.60	3.10	3.50	1.50
Var (All)	0.24	0.49	0.45	0.65
<b>Block 2 Subjects</b>				
Mean (Control)	1.40	2.80	2.60	1.40
Var (Control)	0.24	0.96	0.64	1.84
Mean (Treatment)	1.60	3.00	3.40	1.40
Var (Treatment)	0.64	0.80	1.04	1.84
Mean (All)	1.50	2.90	3.00	1.40
Var (All)	0.45	0.89	1.00	1.84
<b>Block 3 Subjects</b>				
Mean (Control)	1	1.40	3.40	0.40
Var (Control)	0	0.64	0.64	0.64
Mean (Treatment)	1	2.40	3.20	1.40
Var (Treatment)	0	0.24	0.56	0.24
Mean (All)	1	1.90	3.30	0.90
Var (All)	0	0.69	0.61	0.69
<b>All Subjects</b>				
Mean (Control)	1.33	2.33	3.00	1.00
Var (Control)	0.22	1.16	0.67	1.20
Mean (Treatment)	1.40	2.93	3.53	1.53
Var (Treatment)	0.37	0.60	0.65	0.92
Mean (All)	1.37	2.63	3.27	1.27
Var (All)	0.30	0.97	0.73	1.13

**Figure 1 – Survey Score Measures by Block and Treatment Status**



I used randomization inference to test the sharp null hypothesis that the treatment effect of jalapeño seeds is zero for all observations. I used a one tailed test because there is no reason to believe jalapeño seeds reduce the observed spice level. The one-tailed p-value is 0.0439; it indicates that the ATE is statistically significant at the 95% confidence level. Therefore, the sharp null hypothesis is rejected.<sup>8</sup>

**Table 2:** Randomization Inference Results

Model	Outcome (Y)	Covariate (X)	ATE	SE	Sharp Null Hypotheses	
					$ATE = \overline{ATE}$	ATE = 0
					CI	P-Value
(1) Diff-in-Means	Spice Score	Baseline Spice Score	0.5871	0.3162	-0.0419, 1.2075	0.0439
(2) Diff-in-Means	Spice Score	NA	0.6000	0.2951	0.0133, 1.1600	0.0445
(3) Diff-in-Diff	Spice Score Diff	NA / Scales Y	0.5333	0.3809	-0.2311, 1.2889	0.1200
(4) Diff-in-Means	Overall Score	NA	0.5333	0.3030	-0.0356, 1.1111	0.0770

In Model 2, the ATE estimate is slightly larger (0.6000) and has a smaller standard error (0.2951). It is possible that the difference between Model 1 and Model 2 is due to the fact that the baseline score is not predictive of the outcome spice score. *BaseSpice* was used as a covariate because it could potentially account for difference in taste-tester spice sensitivity. However, since there are only three taste testers there is little variation in the *BaseSpice* and it is not prognostic of outcomes.<sup>9</sup>

Model 3, the difference-in-difference model provides a smaller ATE estimate (0.5333) and a larger standard error estimate than Model 1. Differencing is superior in ATE estimation when prognostic covariates are used. However, it does not yield an advantage if the covariate is not prognostic. Under randomization inference the ATE estimate is not significant at the 90% level (p-value = 0.1200). This is unsurprising. The above Table 1 shows that differencing the two spice scores increases rather than decreases the variance in treated and untreated outcomes.<sup>10</sup>

The Model 4 difference-in-means estimate of the ATE indicates that seeds cause treated subjects to receive an average overall score rating that is 0.5333 points higher than that of control subjects. Using randomization inference, I test the sharp null hypothesis that the treatment effect of seeds on overall score rating is zero for all subjects. The p-value of 0.0770 indicates that the null hypothesis cannot be rejected at the 95% level, but it can be rejected at the 90% level.

**Conclusion:** The above models provide strong evidence to reject the null hypothesis that seeds do not have an effect on the measured spice level of guacamole. Model 1 provides an estimate that is both substantively and statistically significant. A baseline spice score covariate was included to account for taste tester differences in perceived spice level; however, it is not prognostic. As such, its inclusion in Model 1 may dampen ATE estimate size as evidenced by

<sup>8</sup> Observing a spice score mean difference of 0.5871 is very unlikely if the null hypothesis were true.

<sup>9</sup> See Appendix A7 provides regressions to approximate whether *BaseSpice* is prognostic.

<sup>10</sup> Appendix A8 provides a balance check regression for *BaseSpice* that shows *BaseSpice* is not correlated with treatment assignment. As such, differencing the spice scores subtracts a non-prognostic, random variable from the spice score outcome measure thus increasing the variance of the outcome measure.

Model 2. Since the baseline score is not prognostic, the difference-in-differences Model 3 does not provide precision gains. Lastly, Model 4 provides substantively significant evidence that seeds do have a positive effect on the overall taste score rating. Together, the results suggest that seeds do increase the observed spice level of my guacamole recipe. Furthermore, the results provide evidence that including seeds improves the overall taste score rating. This information is useful, especially if I am making guacamole for people who enjoy spicy food.

## Appendix with Data and Accompanying R program.

**A1 - Guacamole Recipe:**<sup>11</sup> When I make guacamole, I follow a simple self-recipe:

- 3 Hass avocados (diced and mashed)
- 1/2 large Vidalia onion (diced)
- 1 medium tomato (diced) *Remove insides to prevent guacamole from becoming watery*
- 2 jalapeños (diced)
- 2 cloves of garlic (diced)
- 1 lime (squeezed for the juice)
- Mix the ingredients thoroughly in a bowl, and add salt to taste
- Variants: (1) Low Spice – remove the seeds and pith from the jalapeño, (2) Moderate Spice – remove the jalapeño seeds but keep the pith, (3) Spicy – entire jalapeño.
- Note: The recipe scales nicely for larger servings.

**A2 - Food Preparation:**<sup>12</sup> I will triple the guacamole recipe to ensure that there is enough for serving. There will be three guacamole batches, a baseline pre-treatment “Low-Spice” batch with only the jalapeño body, a control “Moderate” batch with the jalapeño body and pith but no seeds, and a treatment “Spicy” batch with the jalapeño body, pith, and seeds. I will make a large bowl of guacamole and then split it into pre-treatment, treatment, and control batches that will only differ by the jalapeño parts added later. Guacamole will be prepared as follows:

- Step 1 – Prepare all the guacamole ingredients, **except the 6 jalapeños**, and mix them in large mixing bowl.
- Step 2 – Pour one third of the guacamole into each of three smaller mixing bowls. For preparation, label one bowl “low-spice,” one bowl “moderate” and one bowl “spicy.”
- Step 3 – Cut open the jalapeños, remove the seeds, and mix together in a small bowl. Remove the pith from the jalapeños, dice, and mix together in a second small bowl. Dice the jalapeño bodies and mix together in a third small bowl. *(To prevent selection on jalapeños, the guacamole batches will receive equal shares from the jalapeños).*
- Step 4 – Add 1/3 of the jalapeño body mixture to each of the spice level bowls.
- Step 5 – Add 1/3 the jalapeño pith mixture to the spicy bowl and 1/3 to the moderate bowl and discard the rest.
- Step 6 – Add 1/3 of the jalapeño seed mixture to the spicy bowl and discard the rest.
- Step 7 – Mix each batch thoroughly.
- Step 8 – Prepare servings of each guacamole using clean plastic spoons. Prepare 30 pre-treatment “Low-Spice” servings, 15 “Moderate” control servings, and 15 “Spicy” treatment servings.

---

<sup>11</sup> Taken directly from the submitted pre-analysis plan.

<sup>12</sup> Taken directly from the submitted pre-analysis plan.

### A3 - Ingredient Score Chart

Each taste tester was asked to score their samples according to the following categories on a 1 to 5 scale as described below. The survey was administered through a Google Survey. The content rating questions were randomly ordered for the scoring of each guacamole serving. The content rating questions were followed by the overall taste rating question.

**Table A1:** Ingredient Scoring Chart

Contents	1	2	3	4	5
Garlic Level	Very Weak	Weak	Balanced	Strong	Very Strong
Salt Level	Very Weak	Weak	Balanced	Strong	Very Strong
Spicy Level	Very Weak	Weak	Balanced	Strong	Very Strong
Lime Level	Very Weak	Weak	Balanced	Strong	Very Strong
Onion Level	Way Too Little	Not Enough	Balanced	Too Much	Way Too Much
Tomato Level	Way Too Little	Not Enough	Balanced	Too Much	Way Too Much
Overall	1	2	3	4	5
Overall Taste	Terrible	Poor	So-So	Good	Great

**A4 - Serving:**<sup>13</sup> To ensure that the experiment is blind, I will give unlabeled samples to an assistant who will serve the guacamole to the taste testers. The samples will be given to the assistant in the order designated by the randomization procedure. Each individual will receive 10 baseline servings – one for each observation. Each baseline serving will be randomly followed by a treatment/control serving. Under the complete randomization process, each individual will receive 5 treatment servings and 5 control servings.

For each observation ( $i$ ), I will follow the following steps:

- Step 1 – A tester will be asked to sip a glass of milk and a glass of water to “reset” and “clear” their palette from any residual spice flavor.
- Step 2 – An assistant will blindly give the tester a serving from the pre-treatment batch.
- Step 3 – **Baseline Score:** The tester will then score the serving using the below score chart. This score will act as a baseline covariate variable.
- Step 4 – The tester will then sip a glass of milk and a glass of water to “reset” and clear their palette.
- Step 5 – **Outcome Score:** An assistant will blindly and randomly give the tester either a treatment serving or control serving. This score will act as the potential outcome variable.
- Step 6 – The tester will then score the serving using the below chart.

<sup>13</sup> Taken directly from the submitted pre-analysis plan.



## A5 – Random Assignment Table

Table A2: Treatment Assignment Table

Observation	Block	Treatment	Type
1	1	1	Spicy
2	1	0	Moderate
3	1	0	Moderate
4	1	0	Moderate
5	1	1	Spicy
6	1	1	Spicy
7	1	1	Spicy
8	1	1	Spicy
9	1	0	Moderate
10	1	0	Moderate
11	2	1	Spicy
12	2	0	Moderate
13	2	0	Moderate
14	2	1	Spicy
15	2	1	Spicy
16	2	1	Spicy
17	2	1	Spicy
18	2	0	Moderate
19	2	0	Moderate
20	2	0	Moderate
21	3	1	Spicy
22	3	0	Moderate
23	3	1	Spicy
24	3	1	Spicy
25	3	1	Spicy
26	3	0	Moderate
27	3	1	Spicy
28	3	0	Moderate
29	3	0	Moderate
30	3	0	Moderate

**A6 – Serving Administration Table**

The following reference table was used to supply guacamole samples to the recruited assistant. The assistant was not given any information about which guacamole batch the samples were drawn from. Note: Baseline = Low Spice, Control = Moderate, Treatment = Spicy

**Table A3:** Serving Administration

<b>Serving N</b>	<b>Block 1</b>	<b>Block 2</b>	<b>Block 3</b>
1	Baseline	Baseline	Baseline
2	Treatment	Treatment	Treatment
3	Baseline	Baseline	Baseline
4	Control	Control	Control
5	Baseline	Baseline	Baseline
6	Control	Control	Treatment
7	Baseline	Baseline	Baseline
8	Control	Treatment	Treatment
9	Baseline	Baseline	Baseline
10	Treatment	Treatment	Treatment
11	Baseline	Baseline	Baseline
12	Treatment	Treatment	Control
13	Baseline	Baseline	Baseline
14	Treatment	Treatment	Treatment
15	Baseline	Baseline	Baseline
16	Treatment	Control	Control
17	Baseline	Baseline	Baseline
18	Control	Control	Control
19	Baseline	Baseline	Baseline
20	Control	Control	Control

## A7 – Prognostic Check

Without a full schedule of potential outcomes, we cannot compare the covariance of spice level potential outcomes with baseline covariate spice score to the variance of the baseline covariance score. However, the below regressions provide strong evidence that *BaseSpice* is not prognostic. Model 1 is a regression of the spice level outcome measure on *BaseSpice*. Model 2 and Model 3 provide this regression respectively using only control subjects and treatment subjects. The negative coefficient on *BaseSpice* should cause us to doubt that *BaseSpice* is prognostic.<sup>14</sup>

**Table A4:** Covariate Prognostic Evaluation

	Spicy		
	Model 1	Model 2	Model 3
BaseSpice	-0.304	-0.417	-0.273
	-0.334	-0.627	-0.354
Constant	3.587***	3.467***	3.836***
	-0.603	-1.093	-0.657
Blocked	Yes	Yes	Yes
Subjects	All	Control	Treatment
N	30	15	15
R-squared	0.307	0.401	0.32
Adj. R-squared	0.228	0.238	0.135

\*\*\*p < .01; \*\*p < .05; \*p < .1

## A8 – Balance Check

**Table A5:** Balance Check for  
Baseline Spice Score

	BaseSpice
Treatment	0.067
	-0.188
Constant	1.567***
	-0.188
Blocked	Yes
N	30
R-squared	0.234
Adj. R-squared	0.146

\*\*\*p < .01; \*\*p < .05; \*p < .1

<sup>14</sup> With a full schedule of potential outcomes, a prognostic covariate implies  $\frac{Cov(Y_i(0), X_i)}{Var(X_i)} + \frac{Cov(Y_i(1), X_i)}{Var(X_i)} > 1$ .

## A9 – Regression Results for Other Outcome Measures

Table A6: Other Outcome Measures

	<b>Garlic</b>	<b>Salt</b>	<b>Lime</b>	<b>Onion</b>	<b>Tomato</b>
	<b>Model 1</b>	<b>Model 2</b>	<b>Model 3</b>	<b>Model 4</b>	<b>Model 5</b>
Treatment	0.267	-0.067	0.2	-0.2	-0.4
	-0.196	-0.218	-0.147	-0.272	-0.275
Constant	2.667***	2.833***	1.100***	3.200***	3.000***
	-0.196	-0.218	-0.147	-0.272	-0.275
Blocked	Yes	Yes	Yes	Yes	Yes
N	30	30	30	30	30
R-squared	0.152	0.397	0.532	0.326	0.24
Adj. R-squared	0.054	0.327	0.478	0.248	0.152

\*\*\* p < .01; \*\* p < .05; \* p < .1

## A10 – Data Table

Table A7: Data Table

Observation	Block	Treatment	Type	BaseSpice	Garlic	Salt	Spicy	Lime	Onion	Tomato	Overall	SpicyDiff
1	1	1	Spicy	2	3	3	3	2	3	3	4	1
2	1	0	Moderate	1	3	1	3	1	2	4	3	2
3	1	0	Moderate	2	1	3	3	1	4	2	2	1
4	1	0	Moderate	1	3	3	2	1	4	3	3	1
5	1	1	Spicy	1	3	3	4	1	3	2	4	3
6	1	1	Spicy	2	3	3	4	2	3	2	4	2
7	1	1	Spicy	2	3	3	3	1	3	3	4	1
8	1	1	Spicy	1	3	3	3	1	3	3	4	2
9	1	0	Moderate	2	3	3	2	1	3	3	3	0
10	1	0	Moderate	2	3	3	4	1	3	3	4	2
11	2	1	Spicy	1	3	2	2	2	3	2	4	1
12	2	0	Moderate	2	3	3	2	2	3	2	2	0
13	2	0	Moderate	1	2	3	4	1	1	3	4	3
14	2	1	Spicy	3	3	3	3	1	3	2	3	0
15	2	1	Spicy	1	3	3	4	2	3	4	3	3
16	2	1	Spicy	2	3	2	2	2	3	2	2	0
17	2	1	Spicy	1	3	3	4	3	4	3	5	3
18	2	0	Moderate	1	4	3	4	2	4	3	3	3
19	2	0	Moderate	2	3	2	2	2	4	2	2	0
20	2	0	Moderate	1	3	3	2	2	3	3	2	1
21	3	1	Spicy	1	2	1	2	1	4	1	2	1
22	3	0	Moderate	1	2	2	1	1	5	1	3	0
23	3	1	Spicy	1	3	1	2	1	3	1	3	1
24	3	1	Spicy	1	3	2	3	1	5	2	4	2
25	3	1	Spicy	1	3	2	3	1	3	1	4	2
26	3	0	Moderate	1	3	1	3	1	4	3	4	2
27	3	1	Spicy	1	3	2	2	1	4	3	3	1
28	3	0	Moderate	1	2	2	1	1	5	2	2	0
29	3	0	Moderate	1	2	3	1	1	4	3	4	0
30	3	0	Moderate	1	3	2	1	1	4	3	4	0

**A11 – R Code**

```

# Author: David Simpson
# Title: Guacamole Experiment
# Class: W 4768 Experimental Research
# Created: 10 March 2019
# Edited: 26 March 2019
# Adapted From: Professor Donald Green - "Drying wood practicum 2018 - with 5 day outcomes.R"
# References: (1) Gerber, Alan S. and Donald P. Green. 2012. Field Experiments: Design, Analysis,
#             and Interpretation. New York: W.W. Norton.
#             (2) {ri} R Documentation
#####
# Outline:
# Guacamole Recipe given in Pre-Analysis Plan
# Research Question: Does removing jalapeno seeds change the observed spice level
#             in guacamole that contains jalapenos.
# Treatment == Guacamole with jalapeno seeds;
# Control == Guacamole without jalapeno seeds
# Measure == 5 point "spice" scale measured by taste tester
# (1) Testers are "blindly" given a sample of the baseline guacamole
# Pre-treatment measure is a tester spice scoring of the baseline guacamole
# (2) Testers are then "blindly" and randomly given either a sample of the treatment or control
# Outcome measure is a tester spice scoring of the random guacamole
# Estimation model: Outcome Score = a + b(treatment) + c(Baseline-Score) + u
# Test is one-tailed:
# Blocking: I expect to have only three taste testers, so I will block on tester
#####
###
setwd("/Users/dsimp/Dropbox/ColumbiaPoliSci/02.Spring2019/03_Experiments/02_ExperimentProject/Analysis")
rm(list=ls())
library(ri)
library(stargazer)
library("kableExtra")
library("tidyr")
library("foreign")
library("ggplot2")
set.seed(1234567)

#####
## SECTION 1 - PRE-ANALYSIS
#####

N = 30 # Total number of subjects i
n = 10 # Number of subjects per block
m = 5 # Number of treated subjects per block

guac <- data.frame("Observation"=1:N) # Create data frame with 30 observations
guac$Block <- c(rep(1,10),rep(2,10),rep(3,10)) # Create block variable - for 3 blocks

#####
# Randomization Procedure
#####
# Complete Random Assignment within each block
# Per block, 0 < m=5 < 10
guac$Treatment[1:10] <- ifelse(1:n %in% sample(1:10, m), 1, 0) # Random Assignment of treatment for Block 1

```

```

guac$Treatment[11:20] <- ifelse(11:20 %in% sample(11:20, m), 1, 0) # Random Assignment of treatment for
Block 2
guac$Treatment[21:30] <- ifelse(21:30 %in% sample(21:30, m), 1, 0) # Random Assignment of treatment for
Block 3

table(guac$Treatment) # Show total treatment vs control. Confirms 15 in each status

#####
## Random Assignment Table (To report in pre-analysis plan)
#####
guac$Type<- "Spicy" # Variable to describe the treatment type (Spicy as default)
guac$Type[guac$Treatment==0] <- "Moderate" # Change Type to "Moderate" for observations with Treatment == 0
kable(guac, "html", booktabs = T, caption="Random Assignment Table") %>% # Generate Random Assignment
Table
  kable_styling(latex_options = "striped")

#####
## Administrative Table (Make a Table to easily administer the experiment)
#####
# Note: There will be 60 total tastings (30 baseline and 30 treatment or control)
# Note: As such each individual will taste 10 baseline servings, 5 treatment serving, and 5 control servings
# Note: Each taste tester will taste and rate a baseline guacamole serving before tasting and rating each treatment or
control guac

admin <- data.frame("Order"=1:60,"Type"=1:60) # Destination data frame for the administration data

# Loop to insert a baseline observation between each treatment or control serving
counter <- c(1:60)
count <- 0
for (i in 1:60){
  if((counter[i] %% 2)==0){
    print(paste(i,"even"))
    count = i/2
    print(count)
    admin$Type[i] = guac$Treatment[count]+1
  } else{
    print(paste(i,"odd"))
    admin$Type[i]=0
  }
}

admin$Name <- "Baseline" # Describe the batch identity of Baseline servings
admin$Name[admin$Type==1] <- "Control" # Describe the batch identity of Control servings
admin$Name[admin$Type==2] <- "Treatment" # Describe the batch identity of Treatment servings
admin <- cbind(admin[1:20,c(1,3)],admin[21:40,3],admin[41:60,3]) # Organize data by administration rounds.
Assistant will administer 20 rounds
colnames(admin) = c("Order", "Block 1","Block 2","Block 3") # Rename Columns

kable(admin, "html", booktabs = T, caption = "Experiment Administration Table") %>% # Generate Administration
Table
  kable_styling(latex_options = "striped")

#####
## SECTION 2 - EXPERIMENTAL RESULTS ANALYSIS
#####

```

```

results <- read.csv("Results.csv") # Read in Google Survey Results
glimpse(results) #60 Observations

#####
## Prepare Data
#####
# Note: Every 20 observations is a different block (eg. Block 1: 1-20, Block 2: 21-40, Block 3: 41-60)
# Note: Within each block, observations alternate between Baseline servings (odd numbered) and randomly assigned
servings (even-numbered)
# Note: Want to create a data set where each observation has the block assignment, treatment assignment,
#   prior baseline spice score, and outcome ratings
# Note: Will add results data to the "guac" data set

baseline <- results[(results$Number %% 2)>0,1:9] # Baseline pre-test scores are odd numbered
guac$BaseSpice <- baseline$Spicy # Add baseline spice score "BaseSpice" to the guac dataset

outcomes <- results[(results$Number %% 2)==0,1:9] # Treatment and Control outcome measures are even
numbered
guac<-cbind(guac,outcomes[,3:9]) # Add outcomes measures to the guac dataset
guac$SpicyDiff <- guac$Spicy - guac$BaseSpice # Create Differenced Outcome Variable

#####
## Data Summary
#####
kable(guac, "html", booktabs = T, caption = "Data Table") %>% # Generate Data Table
  kable_styling(latex_options = "striped")

# Summary Data Destination Data Frame
summ<-data.frame("Statistic"=1:6,"BaseSpice"=1:6,"Garlic"=1:6,
  "Salt"=1:6,"Spicy"=1:6,"Lime"=1:6,
  "Onion"=1:6,"Tomato"=1:6,"Overall"=1:6,"SpicyDiff"=1:6)
summ$Statistic <-
c("Mean(Control)","Var(Control)","Mean(Treatment)","Var(Treatment)","Mean(All)","Var(All)")

# Loop to create summary statistics
for (j in 1:4){ # For loop identify data group to summarize
  print(paste("Block ",j," Round"))
  if (j<4){ # If statement to create data group and variance adjustments
    guac_summ<-guac[guac$Block==j,]
    adj1<-4/5 # Within block | treatment or control adjustment (obs = 5)
    adj2<-9/10 # Within block total adjustment (obs =10)
  } else{
    guac_summ<-guac
    adj1<-14/15 # Total sample | treatment or control adjustment (obs = 15)
    adj2<-29/30 # Total sample (obs =30)
  }
}
for (i in 5:13){ # For loop to create summary stats
  print(names(guac_summ[i]))
  summ[1,i-3]<-mean(guac_summ[guac_summ$Treatment==0,i])
  summ[2,i-3] <-adj1*var(guac_summ[guac_summ$Treatment==0,i])
  summ[3,i-3]<-mean(guac_summ[guac_summ$Treatment==1,i])
  summ[4,i-3] <-adj1*var(guac_summ[guac_summ$Treatment==1,i])
  summ[5,i-3]<-mean(guac_summ[,i])
  summ[6,i-3] <-adj2*var(guac_summ[,i])
} # End for loop

```



```

summ[,2:10]<-round(summ[,2:10],2)
print(summ)
if (j==1){      # If statement to store data summary
  summ1<-summ
} else if (j==2){
  summ2<-summ
} else if (j==3){
  summ3<-summ
} else {
  summall<-summ
}              # End If Statment

} # End Summary for loop

# Tables With All Data
kable(summ1, "html", booktabs = T, caption = "Block 1 Data Summary Table") %>% # Generate Data Summary
  kable_styling(latex_options = "striped")
kable(summ2, "html", booktabs = T, caption = "Block 2 Data Summary Table") %>% # Generate Data Summary
  kable_styling(latex_options = "striped")
kable(summ3, "html", booktabs = T, caption = "Block 3 Data Summary Table") %>% # Generate Data Summary
  kable_styling(latex_options = "striped")
kable(summall, "html", booktabs = T, caption = "All Blocks Data Summary Table") %>% # Generate Data
Summary
  kable_styling(latex_options = "striped")

# Tables With Key Measures
kable(summ1[,c(1,2,5,9,10)], "html", booktabs = T, caption = "Block 1 Data Summary Table") %>% # Generate
Data Summary
  kable_styling(latex_options = "striped")
kable(summ2[,c(1,2,5,9,10)], "html", booktabs = T, caption = "Block 2 Data Summary Table") %>% # Generate
Data Summary
  kable_styling(latex_options = "striped")
kable(summ3[,c(1,2,5,9,10)], "html", booktabs = T, caption = "Block 3 Data Summary Table") %>% # Generate
Data Summary
  kable_styling(latex_options = "striped")
kable(summall[,c(1,2,5,9,10)], "html", booktabs = T, caption = "All Blocks Data Summary Table") %>% #
Generate Data Summary
  kable_styling(latex_options = "striped")

ggplot(guac,aes(x=BaseSpice,color=factor(Block),fill=factor(Block)))+
  geom_histogram(alpha=0.5,bins=20)+
  #geom_vline(data=mu, aes(xintercept=grp.mean), linetype="dashed",show.legend = FALSE)+
  facet_grid(factor(Treatment)~factor(Block))+
  labs(title="Baseline Spice Score Results",x="Spice Score by Block",y="Count by Treatment Status",fill="
",color=" ",
  subtitle=" Q: On the following scale, please rate your opinion about the spicy level in the guacamole:\
  1 = Very Weak, 2 = Weak, 3 = Balanced, 4 = Strong, 5 = Very Strong")+
  theme_classic()+
  theme(legend.position="none")

#mu<-ddply(data3,"year",summarise,grp.mean=mean(turnout)) #Df of Mean Stats
ggplot(guac,aes(x=Spicy,color=factor(Block),fill=factor(Block)))+
  geom_histogram(alpha=0.5,bins=20)+
  facet_grid(factor(Treatment)~factor(Block))+

```

```

labs(title="Outcome Spice Score Results",x="Spice Score by Block",y="Count by Treatment Status",fill="
",color=" ",
  subtitle=" Q: On the following scale, please rate your opinion about the spicy level in the guacamole:\
  1 = Very Weak, 2 = Weak, 3 = Balanced, 4 = Strong, 5 = Very Strong")+
theme_classic()+
theme(legend.position="none")

#mu<-ddply(data3,"year",summarise,grp.mean=mean(turnout)) #Df of Mean Stats
ggplot(guac,aes(x=Overall,color=factor(Block),fill=factor(Block)))+
  geom_histogram(alpha=0.5,bins=20)+
  #geom_vline(data=mu, aes(xintercept=grp.mean), linetype="dashed",show.legend = FALSE)+
  facet_grid(factor(Treatment)~factor(Block))+
  labs(title="Outcome Overall Taste Score Results",x="Spice Score by Block",y="Count by Treatment Status",fill="
",color=" ",
  subtitle=" Q: On the following scale, please rate your opinion about the overall taste level of the
  guacamole: 1 = Terrible, 2 = Poor, 3 = So-So, 4 = Good, 5 = Great")+
  theme_classic()+
  theme(legend.position="none")

#####
## Balance Check
#####

# Balance Check with Main Covariate the Baseline Spice Variable
summary(b1<-lm(BaseSpice~Treatment+factor(Block),data=guac)) # Balance Check - Spice

stargazer(b1,
  header = FALSE,
  title = "Balance Check for BaseSpice",
  type = "text",
  omit.stat=c("LL","ser","f"),
  #table.placement = "!htbp",
  #column.labels =c("1950","1951","1952"),
  omit = "Block",
  add.lines = list(c("Blocked","Yes")),
  style = "ajps"
)

# Balance Check with Other Baseline Covariates
summary(lm(baseline$Garlic~Treatment+factor(Block),data=guac)) # Balance Check - Garlic
summary(lm(baseline$Salt~Treatment+factor(Block),data=guac)) # Balance Check - Salt
summary(lm(baseline$Lime~Treatment+factor(Block),data=guac)) # Balance Check - Lime
summary(lm(baseline$Onion~Treatment+factor(Block),data=guac)) # Balance Check - Onion
summary(lm(baseline$Tomato~Treatment+factor(Block),data=guac)) # Balance Check - Tomato
summary(lm(baseline$Overall~Treatment+factor(Block),data=guac)) # Balance Check - Overall

#####
## Analyze Main Outcome Measures with Regression
#####

mean(guac$BaseSpice) # Average Baseline Spice = 1.366667
mean(guac$Spicy) # Average Spice Level of Moderate and Spicy Guac = 2.633333
mean(guac$Spicy[guac$Treatment==0]) # Average Control Guac Spice = 2.333333
mean(guac$Spicy[guac$Treatment==1]) # Average Treated Guac Spice = 2.933333

summary(lm(Spicy~Treatment+factor(Block),data=guac))

```

```

summary(lm(Spicy~Treatment+factor(Block)+BaseSpice,data=guac))
summary(lm(SpicyDiff~Treatment+factor(Block),data=guac))

#####
## Analyze Other Outcome Measures with Regression
#####
# Overall - Slightly Better Tasting overall
summary(lm(Overall~Treatment+factor(Block),data=guac)) # Overall

# Other Ingredients - None test better
summary(c1<-lm(Garlic~Treatment+factor(Block),data=guac)) # Garlic
summary(c2<-lm(Salt~Treatment+factor(Block),data=guac)) # Salt
summary(c3<-lm(Lime~Treatment+factor(Block),data=guac)) # Lime
summary(c4<-lm(Onion~Treatment+factor(Block),data=guac)) # Onion
summary(c5<-lm(Tomato~Treatment+factor(Block),data=guac)) # Tomato

stargazer(c1,c2,c3,c4,c5,
  header = FALSE,
  title = "Other Outcome Measures",
  type = "html",
  omit.stat=c("LL","ser","f"),
  #table.placement = "!htbp",
  #column.labels =c("1950","1951","1952"),
  omit = "Block",
  add.lines = list(c("Blocked","Yes", "Yes", "Yes", "Yes", "Yes")),
  style = "ajps"
)

#####
## RI Reporting Table
#####
# Note: Will Create location to store results from randomization inference
# Note: Collect: ATE estimate, SE, 95% CI, and P-value from one tailed test
RiReport <-
data.frame("Y"=1:4,"Covariate"=1:4,"ATE"=1:4,"SE"=1:4,"CI_Lower"=1:4,"CI_Upper"=1:4,"PValue"=1:4)

RiReport$Y<-c("Spice Score","Spice Score","Spice Score Difference","Overall Score")
RiReport$Covariate <-c("Baseline Spice","NA","NA Used to Scale Y","NA")

#####
# Randomization Inference 1 - Difference-in-Means Estimator with Baseline: Y = Spicy Level
#####
Z <- guac$Treatment
Y <- guac$Spicy
X <- guac$BaseSpice
block <- guac$Block

perms <- genperms(Z=Z, blockvar=block,maxiter = 10000) # Permutations
probs <- genprobexact(Z = Z, blockvar = block) # Probability of treatment
table(probs,block) # Confirm treatment probability is identical across blocks

ate <- estate(Y,Z,X,prob=probs) # Estimate the ATE

# Store Estimates (1)
RiReport$ATE[1] <- round(ate,4) # Store ATE estimate in the RiReport Table

```

```

# Get Confidence Intervals
Ys <- genouts(Y,Z,ate=ate)          # Generate potential outcomes under tau = ATE
distout <- gendist(Ys, perms, X=X, prob=probs) # Generate sampling dist. under sharp null
dispdist(distout,ate)              # Display characteristics of sampling dist. for inference
mean(distout)                      # Mean of Sampling Distribution
ate

# Store Estimates (2)
RiReport$SE[1] <- round(dispdist(distout,ate,display.plot = FALSE)$sd,4)      # Store SE from test
ATE=ATE in the RiReport Table
RiReport$CI_Lower[1] <- round(dispdist(distout,ate,display.plot = FALSE)$quantile[1],4) # Store CI_L from test
ATE=ATE in the RiReport Table
RiReport$CI_Upper[1] <- round(dispdist(distout,ate,display.plot = FALSE)$quantile[2],4) # Store CI_U from test
ATE=ATE in the RiReport Table

# Test Sharp Null (No Effect for Every Unit)
Ys <- genouts(Y,Z,ate=0) # generate potential outcomes under sharp null
distout <- gendist(Ys, perms, X=X, prob=probs) # generate sampling dist. under tau = 0
dispdist(distout,ate) # display characteristics of sampling dist. for inference
sum(distout >= ate)    # one-tailed comparison used to calculate p-value (greater than)

# Store Estimates (3)
RiReport$PValue[1] <- round(dispdist(distout,ate,display.plot = FALSE)$greater.p.value,4) # Store one tailed p-
value from Sharp Null Test in the RiReport Table

# plot results with and without covariate adjustment
resresplot(Y,Z,probs,scale=1)
resresplot(Y,Z,X,prob=probs,scale=1)

#####
# Randomization Inference 2 - Difference-in-Means Estimator without Baseline: Y = Spicy Level
#####
ate <- estate(Y,Z,prob=probs) # Estimate the ATE

# Store Estimates (1)
RiReport$ATE[2] <- round(ate,4) # Store ATE estimate in the RiReport Table

# Get Confidence Intervals
Ys <- genouts(Y,Z,ate=ate)          # Generate potential outcomes under tau = ATE
distout <- gendist(Ys, perms, prob=probs) # Generate sampling dist. under sharp null
dispdist(distout,ate)              # Display characteristics of sampling dist. for inference
mean(distout)                      # Mean of Sampling Distribution
ate

# Store Estimates (2)
RiReport$SE[2] <- round(dispdist(distout,ate,display.plot = FALSE)$sd,4)      # Store SE from test
ATE=ATE in the RiReport Table
RiReport$CI_Lower[2] <- round(dispdist(distout,ate,display.plot = FALSE)$quantile[1],4) # Store CI_L from test
ATE=ATE in the RiReport Table
RiReport$CI_Upper[2] <- round(dispdist(distout,ate,display.plot = FALSE)$quantile[2],4) # Store CI_U from test
ATE=ATE in the RiReport Table

# Test Sharp Null (No Effect for Every Unit)
Ys <- genouts(Y,Z,ate=0) # generate potential outcomes under sharp null
distout <- gendist(Ys, perms, prob=probs) # generate sampling dist. under tau = 0

```

```

dispdist(distout,ate) # display characteristics of sampling dist. for inference
sum(distout >= ate)      # one-tailed comparison used to calculate p-value (greater than)

# Store Estimates (3)
RiReport$PValue[2] <- round(dispdist(distout,ate,display.plot = FALSE)$greater.p.value,4) # Store one tailed p-
value from Sharp Null Test in the RiReport Table

#####
# Randomization Inference 3 - Difference-in-Differences Estimator: Y = Spicy Level - Baseline Spice Level
#####
Y <- guac$SpicyDiff      # Set new Outcome Variable
ate <- estate(Y,Z,prob=probs) # estimate the ATE

# Store Estimates (1)
RiReport$ATE[3] <- round(ate,4) # Store ATE estimate in the RiReport Table

# Get Confidence Intervals
Ys <- genouts(Y,Z,ate=ate) # generate potential outcomes under tau = ATE
distout <- gendist(Ys, perms, prob=probs) # generate sampling dist. under sharp null
dispdist(distout,ate) # display characteristics of sampling dist. for inference
mean(distout)
ate

# Store Estimates (2)
RiReport$SE[3] <- round(dispdist(distout,ate,display.plot = FALSE)$sd,4)      # Store SE from test
ATE=ATE in the RiReport Table
RiReport$CI_Lower[3] <- round(dispdist(distout,ate,display.plot = FALSE)$quantile[1],4) # Store CI_L from test
ATE=ATE in the RiReport Table
RiReport$CI_Upper[3] <- round(dispdist(distout,ate,display.plot = FALSE)$quantile[2],4) # Store CI_U from test
ATE=ATE in the RiReport Table

# Test Sharp Null (No Effect for Every Unit)
Ys <- genouts(Y,Z,ate=0) # generate potential outcomes under sharp null
distout <- gendist(Ys, perms, prob=probs) # generate sampling dist. under tau = 0
dispdist(distout,ate) # display characteristics of sampling dist. for inference
sum(distout >= ate)      # one-tailed comparison used to calculate p-value (greater than)

# Store Estimates (3)
RiReport$PValue[3] <- round(dispdist(distout,ate,display.plot = FALSE)$greater.p.value,4) # Store one tailed p-
value from Sharp Null Test in the RiReport Table

#####
# Randomization Inference 4 - Difference-in-Means Estimator: Y = Overall Taste (no covariate)
#####
Y <- guac$Overall      # Set new Outcome Variable
ate <- estate(Y,Z,prob=probs) # estimate the ATE

# Store Estimates (1)
RiReport$ATE[4] <- round(ate,4) # Store ATE estimate in the RiReport Table

# Get Confidence Intervals
Ys <- genouts(Y,Z,ate=ate) # generate potential outcomes under tau = ATE
distout <- gendist(Ys, perms, prob=probs) # generate sampling dist. under sharp null
dispdist(distout,ate) # display characteristics of sampling dist. for inference

```

```

mean(distout)
ate

# Store Estimates (2)
RiReport$SE[4] <- round(dispdist(distout,ate,display.plot = FALSE)$sd,4) # Store SE from test
ATE=ATE in the RiReport Table
RiReport$CI_Lower[4] <- round(dispdist(distout,ate,display.plot = FALSE)$quantile[1],4) # Store CI_L from test
ATE=ATE in the RiReport Table
RiReport$CI_Upper[4] <- round(dispdist(distout,ate,display.plot = FALSE)$quantile[2],4) # Store CI_U from test
ATE=ATE in the RiReport Table

# Test Sharp Null (No Effect for Every Unit)
Ys <- genouts(Y,Z,ate=0) # generate potential outcomes under sharp null
distout <- gendist(Ys, perms, prob=probs) # generate sampling dist. under tau = 0
dispdist(distout,ate) # display characteristics of sampling dist. for inference
sum(distout >= ate) # one-tailed comparison used to calculate p-value (greater than)

# Store Estimates (3)
RiReport$PValue[4] <- round(dispdist(distout,ate,display.plot = FALSE)$greater.p.value,4) # Store one tailed p-
value from Sharp Null Test in the RiReport Table

#####
## Create RI Reporting Table
#####
kable(RiReport, "html", booktabs = T, caption = "Randomization Inference Results") %>% # Generate
Administration Table
  kable_styling(latex_options = "striped")

#####
## Covariate Prognostic Evaluation
#####
summary(p1<-lm(Spicy~BaseSpice+factor(Block),data=guac))
summary(p2<-lm(Spicy~BaseSpice+factor(Block),data=subset(guac,guac$Treatment==0)))
summary(p3<-lm(Spicy~BaseSpice+factor(Block),data=subset(guac,guac$Treatment==1)))

stargazer(p1,p2,p3,
  header = FALSE,
  title = "Covariate Prognostic Evaluation",
  type = "html",
  omit.stat=c("LL","ser","f"),
  #table.placement = "!htbp",
  #column.labels =c("1950","1951","1952"),
  omit = "Block",
  add.lines = list(c("Blocked","Yes", "Yes", "Yes"),
    c("Subjects","All","Control","Treatment")),
  style = "ajps"
)

```