
TITLE

AUGUST 23, 2021

Group 2:

Andy Vu (1005244932)

David Pham (1005349053)

Gajan Gunasegaran (1005233481)

Steven Tran (1004927644)

Zun Li (1002555923)

1 Description of Case Study

Toilet paper is one of the top essential personal hygiene products of the century. It is widely used throughout western culture, namely in Canada. Consumer demand and market expansion for toilet paper have lead to the establishment of many brands with only a few claiming the majority of the market share. Good toilet paper is usually determined by its durability after absorbing liquids. We want to take a deeper look into the various popular toilet paper brands in Canada and see which brand offers the strongest toilet paper after exposure to various levels of moisture. Consequently, we want to devise an experiment to answer the question, “Does the brand of toilet paper and amount of liquid absorbed have an effect on its durability?”. The target population would be popular 2-ply toilet paper brands in Canada and the experimental unit would be a single square piece of 2-ply toilet paper. The results of this case study will allow us to understand the effects of brand and consumer usage on the credibility of the product.

2 Description of Experimental Design

2.1 Variables

The response variable of interest is the maximum number of coins that the wet toilet paper can support before it breaks in each trial. It is a numerical, interval variable.

We have two predictor variables:

- 1) Toilet paper brand (A, B, C, D, E) for (Cashmere, Charmin, Cottonelle, Kirkland, Royale) respectively. It is a factor (categorical predictor) with 5 levels.
- 2) The amount of water applied to the toilet paper (J, K, L) for (Low, Medium, High), respectively. It is also a factor (categorical variable) that describes the amount of water that was applied to the toilet paper, with 3 levels: Low = 0.5mL, Medium = 1.0mL, High = 1.5mL.

We treat Cashmere and the Low moisture level as the baseline level, and cross the two factors, resulting in $5 \times 3 = 15$ different treatments. We conduct each treatment 5 times (justification for sample size is in the Data Analysis section), which gives us a grand total of $15 \times 5 = 75$ samples. Since there is an even number of observations at each treatment combination, we have a complete (crossed) and balanced design. Furthermore, this experiment is a between-subjects design because the experimental unit (the single sheet of toilet paper) becomes ineffective after undergoing a single treatment.

As a side note, we did not consider any placebo treatment, control groups, or blindness as there are no psychological drawbacks in this experiment.

2.2 Procedure

We initialized a rigorous procedure to make sure there were constant, unchanging factors:

Materials needed: A glass cup exactly 5 cm in diameter, rubber band, quarters (made in 2000 and onward because composition of coins are different), a syringe, water, and toilet paper from a specific brand.

(**TODO**): take pictures of experiment to illustrate procedure

1. Carefully rip out all the sheets of toilet paper from the roll, making sure that none are damaged. Then, randomly select 15 of these sheets.
2. Take a single sheet of toilet paper and place it over the cup, fastening it in place with a rubber band.
3. Administer the previously determined amount of water using a syringe to the toilet paper.
4. After 5 seconds, place a coin on the center of the paper. If the coin does not tear and fall through the paper, wait another 5 seconds and place another coin. Continue waiting for 5 seconds and placing coins until the paper tears apart.
5. Record the number of coins needed to tear through the paper. Conduct the experiment 5 times for each factor (moisture) level of water (0.5ml, 1ml, 2ml), and repeat steps 2-5 until all trials have been completed.

2.3 Experiment Principles

To begin, there were many uncontrollable factors throughout the experiment, such as surrounding humidity, room temperature, and exact water temperature. These factors will be neutralized and taken into consideration by the randomization principle. However, there were plenty of other variables that we controlled for. These included: cup size (exactly 5 cm in diameter), the coins we used, the type of rubber band we used to fasten the toilet paper sheet in place, the method and initial force of placing a coin on the wet toilet paper, the syringe we used to administer water, the type of bathroom tissue (we only looked at 2-ply), and the time between applying water and dropping coins on the toilet paper.

We did not use blocking because we are primarily interested in the toilet paper brands themselves. Since we controlled most of the nuisance variables, we instead opt for a completely randomized (CR) design.

Next, because we did not utilize blocking, randomization becomes a key principle in our experiment. We rip off all the toilet paper sheets from a roll, take a random sheet from that pile, and randomly assign them to a moisture level in order to reduce confounding. Figure (**TODO** x) illustrates the random assignment of the experimental units. In addition to this, we will assume that a toilet paper sheet from a random toilet paper roll (from a random pack of bathroom tissue) is representative of that brand to allow for valid inferences made later on in the experiment.

Finally, in terms of replication, we replicate at the *treatment* level 5 times per treatment to ensure that the resulting effects are not due to the experimental unit. Furthermore, this better allows us to estimate experimental error.

3 Data Analysis

3.1 Sample Size Determination

The sample size for this experiment was calculated using the `ss.2way` function from the `pwr2` R package; the corresponding output is in Table 2. The function takes as input the number of groups in each factor,

significance level chosen for the hypothesis test, desired power, and the effect sizes for the factors. The number of groups was set to 5 for the brand factor (Cashmere, Charmin, Cottonelle, Kirkland, Royale), and 3 in the moisture factor (Low, Medium, high). The significance level chosen for the hypothesis test is 0.05. Since the purpose of this study was to understand the properties of various toilet paper products and inform future consumer choice, the ability to replicate the findings of this study is a key objective. As a result, a high power is necessary as it corresponds to a low probability of Type II error. Thus, the power was set to 0.8, or 80%. And finally, in order to “fine tune” the experiment such that even relatively small effect gets detected, the effect size for each factor was set to 0.25. Our end result for the sample size calculations tell us that each factor level must have 5 units assigned, which means our total sample size is 75.

3.2 Two-Way ANOVA

The first step of our two-way ANOVA analysis is to determine if there is a significant interaction effect between the two factors (brand and moisture). From the interaction plot seen in Figure 2 we observed mostly parallel lines, except for cashmere having a significant higher mean (# of coins) under low moisture relative to when under medium and high moisture level. This implies minimal interaction effect between the two factors, so we fitted the interaction model and formally tested the interaction effect with ANOVA F-test (see Tables 3 and 4). From the ANOVA table we obtained a p-value of 0.38, which implies that we do not have significant evidence for the existence of an interaction effect between the two factors. And together with the interaction plot (Figure 2), we concluded that there is no significant interaction effect between the toilet paper brands and the moisture levels on the number of coins that the toilet paper can support, and we proceed to fitting a reduced additive model with just the main effects as our final model.

Observing the additive model summary (Table 5), we see that on average, Cashmere can support up to 15 coins before tearing, regardless of moisture level. There is not enough evidence to say anything about Charmin, Royale and Cottonelle compared to the baseline, but Kirkland was deemed statistically significant to conclude that it can support an additional 6-7 coins more than Cashmere (while keeping moisture constant). Finally, for our second factor, it is clear that moisture levels affect the number of coins supported; in fact, we have strong evidence to deduce that the average number of coins held by the sheet of toilet paper decreases by roughly 3 coins and 7 coins for the `medium` and `high` moisture levels respectively.

From the ANOVA table of the additive model (Table 6), we obtained two extremely small p-values ($< .000001$) that are well below the 5% significance level, which provides strong evidence that both brand and moisture level have a statistically significant effect on the # of coins supported when the other respective factor is controlled for.

3.3 Tukey’s HSD Post-Hoc Analysis

Table 1: Tukey’s HSD Confidence Interval Plot

term	contrast	null value	estimate	conf low	conf high	adj p value
brand	B-A	0	0.8666667	-3.0968779	4.8302112	0.9726002
brand	C-A	0	-2.4000000	-6.3635446	1.5635446	0.4427332

term	contrast	null value	estimate	conf low	conf high	adj p value
brand	D-A	0	6.4666667	2.5031221	10.4302112	0.0001992
brand	E-A	0	0.8000000	-3.1635446	4.7635446	0.9795715
brand	C-B	0	-3.2666667	-7.2302112	0.6968779	0.1544432
brand	D-B	0	5.6000000	1.6364554	9.5635446	0.0016610
brand	E-B	0	-0.0666667	-4.0302112	3.8968779	0.9999989
brand	D-C	0	8.8666667	4.9031221	12.8302112	0.0000003
brand	E-C	0	3.2000000	-0.7635446	7.1635446	0.1699278
brand	E-D	0	-5.6666667	-9.6302112	-1.7031221	0.0014202
moisture	K-J	0	-3.1600000	-5.7850884	-0.5349116	0.0143439
moisture	L-J	0	-7.5600000	-10.1850884	-4.9349116	0.0000000
moisture	L-K	0	-4.4000000	-7.0250884	-1.7749116	0.0004367

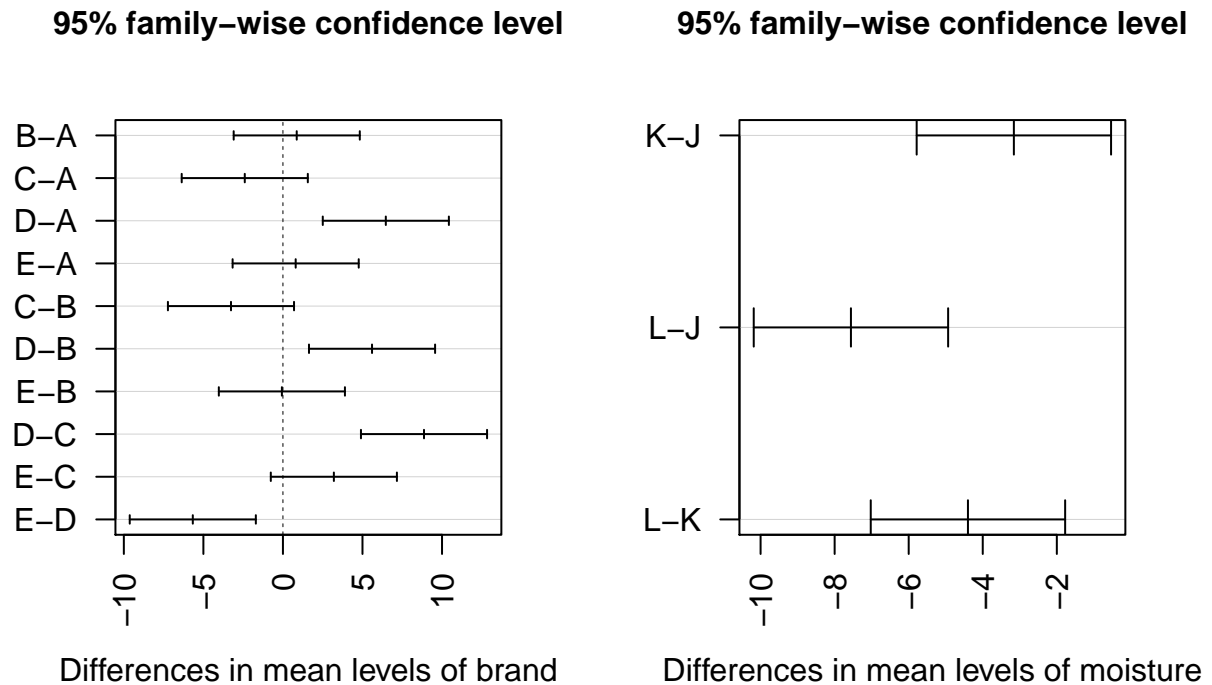


Figure 1: Tukey's HSD Confidence Interval Plot

To determine if the mean difference between specific pairs of groups are statistically significant, we shift our attention to Tukey's HSD. Table 1 shows us the mean difference estimates of all pairs of each factor level, along with their confidence intervals and p-values. The coding for the factor levels in these contrasts were described in Section 2.1.

Looking only at the confidence intervals in 1 without 0 and the p-values less than our benchmark significance level, we see that $D - A$, $D - B$, $D - C$, $E - D$, and all the moisture contrasts are significant. It is notable that D (Kirkland) appears in all these significant contrasts; each pair has positive estimates (except $E - D$,

since Kirkland's estimate is being subtracted from Royale's), making it a strong contender for durable toilet paper brands since it can hold more coins than any other brand on average. Hence, we reject H_0 , affirming that the mean difference of coins supported between Kirkland and any other brand is non-zero, and its difference lays in each respective interval with 95% confidence.

Finally, each pair of moisture levels is also statistically significant, so we can conclude that the mean difference of coins supported for each pairwise comparison is also non-zero (holding brand constant).

4 Other Relevant Information

It is no surprise that a human executed experiment will have some degree of human error. During data collection, we found that there was difficulty keeping the coins balanced on top of each other. This in theory varied the amount of pressure induced by the coins on the toilet paper between each trial, as well as which points of the toilet paper experienced more force. In addition, we also found that during some of the trials for the medium and especially the high moisture levels, some of the water would not get absorbed by the toilet paper and leak through into the cup. This directly varies the amount of water absorbed by the toilet paper that was supposed to be held constant.

5 Limitations

A major limitation of our design is the sample. Since we are only experimenting on the top toilet paper brands in Canada, our results cannot be extrapolated to all toilet paper. As such, our target population is limited to only toilet paper brands in Canada. A simple improvement to this would be conducting the experiment again with toilet paper brands from various countries in the world. For reducing human error in our data collection, we can design a rig to place the coins in the exact spot with the same amount of force each time.

6 Conclusion

In the final analysis, both moisture level and brand affect the durability of toilet paper. With this in mind, we found that all the moisture levels are statistically significant holding brand constant, but not all brands are statistically significant holding moisture level constant. More specifically, while holding moisture level constant, we found that the Kirkland brand was supporting on average an additional 6-7 coins more than the brand default, Cashmere. Furthermore, by comparing individual pairs of brands, we also found that the pairs with the Kirkland brand have a greater average difference than other pairs. Hence, we conclude that the Kirkland brand of toilet paper is more durable on average compared to the other brands.

To answer our question of interest, it is clear that regardless of moisture level, the durability of toilet paper is dependent on its brand. This tells us that even though the same product is being created by different brands, they are not created equal. Also, it is important to notice that even though the brands we have selected for our study were popular Canadian brands, there was a variation in quality. That is to say, we can deduce that

quality is independent of brand popularity. Ultimately, consumers should consider the quality of a product over the brand that is producing it.

7 Appendix

7.1 Tables and Plots

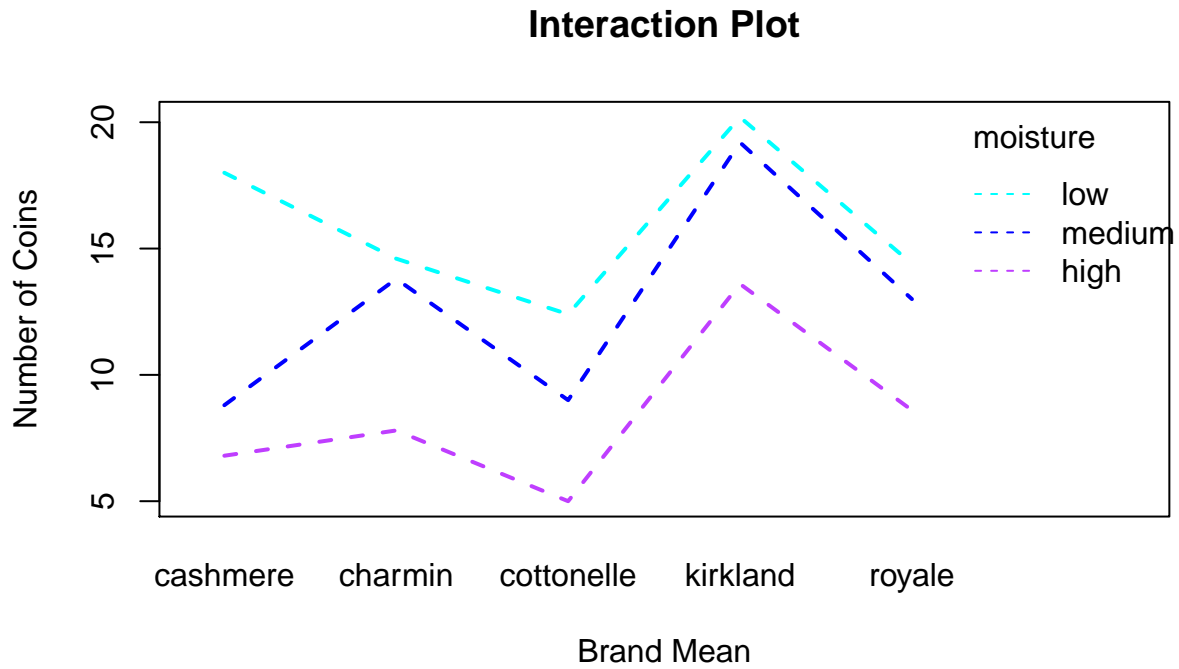


Figure 2: Interaction Plot for Brand and Moisture

Table 2: Sample size computation

Parameter	Value
a	5
b	3
Significance Level	0.05
Power	0.8
N	5
Note	n is number in each group, total sample = 75
Method	Balanced two-way analysis of variance sample size adjustment

Interaction + main Effect Model:

$$\begin{aligned}
E(Y_i) = & \beta_0 + \underbrace{\beta_1 I_{B,i} + \beta_2 I_{C,i} + \beta_3 I_{D,i} + \beta_4 I_{E,i}}_{\text{Terms for brands}} \\
& + \underbrace{\beta_5 I_{K,i} + \beta_6 I_{L,i}}_{\text{Terms for water applied}} \\
& + \underbrace{\beta_7 (I_{B,i} \times I_{K,i}) + \beta_8 (I_{B,i} \times I_{L,i})}_{\text{Charmin-(Med/High) ITs}} + \underbrace{\beta_9 (I_{C,i} \times I_{K,i}) + \beta_{10} (I_{C,i} \times I_{L,i})}_{\text{Cottonelle ITs}} \\
& + \underbrace{\beta_{11} (I_{D,i} \times I_{K,i}) + \beta_{12} (I_{D,i} \times I_{L,i})}_{\text{Kirkland-(Med/High) ITs}} + \underbrace{\beta_{13} (I_{E,i} \times I_{K,i}) + \beta_{14} (I_{E,i} \times I_{L,i})}_{\text{Royale-(Med/High) ITs}} + \epsilon_i
\end{aligned}$$

$$H_0 : \beta_7 = \beta_8 = \beta_9 = \beta_{10} = \beta_{11} = \beta_{12} = \beta_{13} = \beta_{14} = 0$$

$$H_A : \text{at least one } \beta_i \neq 0 \text{ for } i \in \{7, \dots, 14\}$$

Table 3: Interaction Model Summary

	Estimate	Std. Error	t value	Pr(> t)
(Intercept)	18.0	1.723175	10.4458318	0.0000000
brandcharmin	-3.4	2.436938	-1.3951935	0.1681005
brandcottonelle	-5.6	2.436938	-2.2979658	0.0250654
brandkirkland	2.2	2.436938	0.9027723	0.3702569
brandroyale	-3.6	2.436938	-1.4772637	0.1448350
moisturemedium	-9.2	2.436938	-3.7752295	0.0003689
moisturehigh	-11.2	2.436938	-4.5959315	0.0000227
brandcharmin:moisturemedium	8.4	3.446351	2.4373608	0.0177763
brandcottonelle:moisturemedium	5.8	3.446351	1.6829396	0.0975840
brandkirkland:moisturemedium	8.2	3.446351	2.3793284	0.0205416
brandroyale:moisturemedium	7.8	3.446351	2.2632636	0.0272501
brandcharmin:moisturehigh	4.4	3.446351	1.2767128	0.2066229
brandcottonelle:moisturehigh	3.8	3.446351	1.1026156	0.2745972
brandkirkland:moisturehigh	4.6	3.446351	1.3347452	0.1870027
brandroyale:moisturehigh	5.4	3.446351	1.5668748	0.1224033

Table 4: Interaction Model ANOVA Table

	Df	Sum Sq	Mean Sq	F value	Pr(>F)
brand	4	635.9200	158.98000	10.708127	0.0000013
moisture	2	720.8267	360.41333	24.275707	0.0000000

	Df	Sum Sq	Mean Sq	F value	Pr(>F)
brand:moisture	8	129.4400	16.18000	1.089807	0.3828059
Residuals	60	890.8000	14.84667	NA	NA

Additive Model with only main effect:

$$E(Y_i) = \beta_0 + \underbrace{\beta_1 I_{B,i} + \beta_2 I_{C,i} + \beta_3 I_{D,i} + \beta_4 I_{E,i}}_{\text{Terms for brands}} + \underbrace{\beta_5 I_{K,i} + \beta_6 I_{L,i}}_{\text{Terms for water applied}} + \epsilon_i$$

- Brand Effect

$$H_0 : \beta_1 = \beta_2 = \beta_3 = \beta_4 = 0$$

$$H_A : \text{at least one } \beta_i \neq 0 \text{ for } i \in \{1, \dots, 4\}$$

- Moisture Effect

$$H_0 : \beta_5 = \beta_6 = 0$$

$$H_A : \text{at least one } \beta_i \neq 0 \text{ for } i \in \{5, 6\}$$

Table 5: Additive Model Summary

	Estimate	Std. Error	t value	Pr(> t)
(Intercept)	14.7733333	1.183355	12.4842769	0.0000000
brandcharmin	0.8666667	1.414380	0.6127538	0.5420831
brandcottonelle	-2.4000000	1.414380	-1.6968567	0.0942970
brandkirkland	6.4666667	1.414380	4.5720860	0.0000210
brandroyale	0.8000000	1.414380	0.5656189	0.5735146
moisturemedium	-3.1600000	1.095574	-2.8843328	0.0052491
moisturehigh	-7.5600000	1.095574	-6.9004924	0.0000000

Table 6: Additive Model ANOVA Table

	Df	Sum Sq	Mean Sq	F value	Pr(>F)
brand	4	635.9200	158.98000	10.59617	9.868e-07
moisture	2	720.8267	360.41333	24.02190	1.280e-08

	Df	Sum Sq	Mean Sq	F value	Pr(>F)
Residuals	68	1020.2400	15.00353	NA	NA

7.2 R Code

7.2.1 Sample size calculation

```
sample_size_calc <- ss.2way(a=5, b=3, alpha=0.05, beta=1-0.8, f.A=0.25, f.B=0.25, B=4)
```

7.2.2 Interaction plot

```
interaction.plot(brand, moisture, coins,
                 col=c("cyan", "darkorchid1", "blue"),
                 lty=2,
                 lwd=2,
                 xlab="Brand Mean",
                 ylab="Number of Coins")
```

7.2.3 Model generation

```
# Function to convert fitted model to df with convenient col/rownames
model_to_df <- function(model) {
  df <- tidy(model) %>% as.data.frame

  # Use cleaned predictors names as dataframe row names
  varnames <- df$term %>%
    lapply(function(x) gsub("[^A-Za-z0-9]", "", x)) # keep only alphanumeric chars in row names
  rownames(df) <- varnames
  df$term <- NULL
  return(df)
}

# Function to add useful terms to a named list with its model.
gen_model_summaries <- function(vec) {
  # Pull out useful values from model summaries to data frame and to vec scope
  vec$dataframe <- model_to_df(vec$model)
```

```

vec$modelsummary <- summary(vec$model)
vec$anovatable <- anova(vec$model) %>% as.data.frame

return(vec)
}

# Loading in the data & attaching col vectors to global scope, rename "some" -> "medium"
toiletPaperData <- read.csv("toilet_paper_data.csv") %>%
  mutate(moisture = str_replace(moisture, "some", "medium"))
  replace("some", "medium")
attach(toiletPaperData)

# Force an order for the data variables
toiletPaperData$moisture <- factor(toiletPaperData$moisture, levels=c("low", "medium", "high"))

# Model with interaction
int <- list(model=lm(coins ~ brand * moisture))

# Pull out useful values from model summaries to data frame and to object scope
int <- gen_model_summaries(int)

# Additive Model: model with main effects only
additive <- list(model=lm(coins ~ brand + moisture))

# Pull out useful values from model summaries to data frame and to object scope
additive <- gen_model_summaries(additive)

```

7.2.4 Assessing Model Assumptions

```

# Assessing independence of the error terms using residual vs fitted plots
par(mfrow=c(1, 2))
plot(int$model, which = 1, main="Interaction Model")
plot(additive$model, which = 1, main="Additive Model")

# Assessing identical distribution of the error terms using scale-location plots
par(mfrow=c(1, 2))
plot(int$model, which=3, main="Interaction Model")
plot(additive$model, which=3, main="Additive Model")

# Assessing normality of the error terms using QQ plots
par(mfrow=c(1, 2))
plot(int$model, which=2)

```

```

plot(additive$model, which=2)

# Assessing normality of the error terms using Shapiro-Wilk tests
int$resid <- residuals(object = int$model)
additive$resid <- residuals(object = additive$model)
data.frame(
  Interaction=shapiro.test(int$resid) %>% tidy %>% select(-method),
  Additive=shapiro.test(additive$resid) %>% tidy %>% select(-method)
) %>%
  setNames(colnames(.) %>% lapply(function(x) gsub("\\.", " ", x))) %>%
  kable(caption="Interaction and Additive Model Shapiro-Wilk Tests")

```