Deliverable 7:

Testing the Centers of Weekday

Rebekah Sander

Research Question

Does day of the week predict number of casual bike rentals per day?

Research Variables

WHO	WHAT measurement is made on each		TYPE OF MEASURE
	Name of Variable	Question Asked	
A day in 2011 and 2012	Day of the Week	What day of the week is it?	Categorical Variable levels: Sunday, Monday, Tuesday, Wednesday, Thursday, Friday, Saturday
	Casual Counts	How many casual bike rentals per day?	Quantitative Variable Unit: Casual Bike Rentals
One categorical variable with K levels predicting one quantitative variable.			
1. One-way ANOVA 2. Welch's Test on Raw Data	3. Kruskal Wallis 4. Welch's Test on Ranked Data		

SAS Code: Examining the Data

```
/*Keeping what we want*/
data work.bike;
    set work.bike (keep = weekday casual);
    rename casual='Casual Counts'n;
run;


/* Check for and fix miscoding/missing values */
PROC FREQ DATA=WORK.bike;
TABLE weekday 'Casual Counts'n;
run;

Proc Contents data=work.bike varnum;
run;

Proc Means data = work.bike MAXDEC=2 n mean stddev median Qrange RANGE min Q1 Q3 max;
    var 'Casual Counts'n;
run;
```

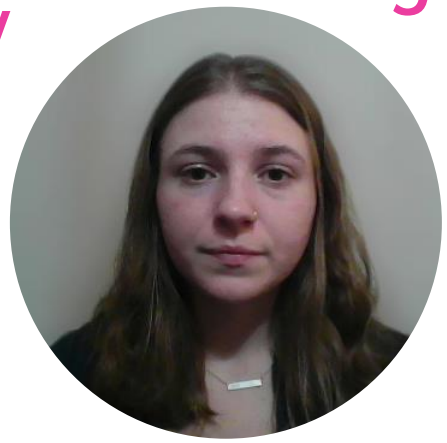
SAS Code: Examining the Data


```
/*fixing weekday*/
*0=sunday,1=monday,2=tuesday,3=wednesday,4=thursday,5=friday,6=saturday;
data work.bike;
set work.bike;
    if weekday = 0 then 'weekday category'n='A';
    else if weekday = 1 then 'weekday category'n='B';
    else if weekday = 2 then 'weekday category'n='C';
    else if weekday = 3 then 'weekday category'n='D';
    else if weekday = 4 then 'weekday category'n='E';
    else if weekday = 5 then 'weekday category'n='F';
    else if weekday = 6 then 'weekday category'n='G';
run;
```

```
Proc format;
Value $weekdayformat
'A'="Sunday"
'B'="Monday"
'C'="Tuesday"
'D'="Wednesday"
'E'="Thursday"
'F'="Friday"
'G'="Saturday";
run;
```

```
data work.bike;
set work.bike;
format 'weekday category'n weekdayformat.;
rename 'weekday category'n='Day of the Week'n;
drop weekday;
run;
```

SAS Code: Assessing Normality and Homogeneity



```
/*assessing normality and homogeneity*/
TITLE "Table 1: Table to Compare Standard Deviations for Homogeneity";
PROC MEANS DATA = work.bike mean stddev VAR maxdec=4;
      class 'Day of the Week'n;
      VAR 'Casual Counts'n;
RUN;
TITLE;
```

The MEANS Procedure

Analysis Variable : Casual Counts casual				
Day of the Week	N Obs	Mean	Std Dev	Variance
Sunday	105	1338.2952	809.3248	655006.5947
Monday	105	674.1333	493.7773	243816.0205
Tuesday	104	556.1827	342.7487	117476.6750
Wednesday	104	551.1442	401.8090	161450.4353
Thursday	104	590.9615	371.6803	138146.2509
Friday	104	752.2885	483.4224	233697.2170
Saturday	105	1465.2571	927.0829	859482.6736

Assessing Normality

- H_0: The data came from a population where the number of casual rentals are normally distributed
- H_A: The data came from a population where the number of casual rentals are not normally distributed.
- $\alpha = 0.05$

Analysis Variable : Casual Counts casual				
Day of the Week	N Obs	Mean	Std Dev	Variance
Sunday	105	1338.2952	809.3248	655006.5947
Monday	105	674.1333	493.7773	243816.0205
Tuesday	104	556.1827	342.7487	117476.6750
Wednesday	104	551.1442	401.8090	161450.4353
Thursday	104	590.9615	371.6803	138146.2509
Friday	104	752.2885	483.4224	233697.2170
Saturday	105	1465.2571	927.0829	859482.6736

- The sample sizes for all seven categories are greater than 30 observations. Therefore, normality is supported.

Assessing Homogeneity

- H_0: $\sigma^2_{\text{Sunday}} = \sigma^2_{\text{Monday}} = \sigma^2_{\text{Tuesday}} = \sigma^2_{\text{Wednesday}}\ \sigma^2_{\text{Thursday}} = \sigma^2_{\text{Friday}} = \sigma^2_{\text{Saturday}}$

  H_A: At least one variance is different than the rest

- The ratio is greater than 2. Thus, the standard deviations are not close enough to use a test that requires homogeneity. The data is

Analysis Variable : Casual Counts casual				
Day of the Week	N Obs	Mean	Std Dev	Variance
Sunday	105	1338.2952	809.3248	655006.5947
Monday	105	674.1333	493.7773	243816.0205
Tuesday	104	556.1827	342.7487	117476.6750
Wednesday	104	551.1442	401.8090	161450.4353
Thursday	104	590.9615	371.6803	138146.2509
Friday	104	752.2885	483.4224	233697.2170
Saturday	105	1465.2571	927.0829	859482.6736

Ratio of standard deviations:

$$= \frac{SD_{Saturday}}{SD_{Tuesday}}$$

$$= \frac{927.0829}{342.7487}$$

$$= 2.7048 > 2$$

Choosing Hypothesis Test: Welch's Test on Raw Data

- Since the data is normal and heterogeneous, we will perform Welch's Test on Raw Data.
- The null hypothesis is that all weekdays have the same mean number casual bike rentals sent each month for the populations each weekday.

H_0: $\mu_{Sunday} = \mu_{Monday} = \mu_{Tuesday} = \mu_{Wednesday} = \mu_{Thursday} = \mu_{Friday} = \mu_{Saturday}$

- The alternative hypothesis is,

H_A: At least one day of the week has a different average number of casual bike rentals per day.

- The level of significance, $\alpha = 0.05$, tells us that 5% of the time the analysis will conclude that at least one mean is different when all means are equal for all days of the week.

SAS Code: Welch Test on Raw Data

```
/*Welch's Test on Data*/
TITLE "Welch Nonparametric Test for Mean Casual Counts across Days of the Week";
PROC GLM data=work.bike order=internal;
    class 'Day of the Week'n;
    model 'Casual Counts'n = 'Day of the Week'n;
    means 'Day of the Week'n / hovtest=levene(TYPE=square) welch;
run;
quit;

Proc Means data=work.bike n mean std;
var 'Casual Counts'n;
run;
title;
```

Welch's ANOVA for Casual Counts			
Source	DF	F Value	Pr > F
Day of the Week	6.0000	29.22	<.0001
Error	319.4		

The MEANS Procedure

Analysis Variable : Casual Counts casual		
N	Mean	Std Dev
731	848.1764706	686.6224883

Level of Day of the Week	N	Casual Counts	
		Mean	Std Dev
Sunday	105	1338.29524	809.324777
Monday	105	674.13333	493.777298
Tuesday	104	556.18269	342.748705
Wednesday	104	551.14423	401.808954
Thursday	104	590.96154	371.680307
Friday	104	752.28846	483.422400
Saturday	105	1465.25714	927.082884

Performing Welch's Test on Raw Data

- F Statistic: The variance between each day of the week's average amount of casual bike rentals from the overall average (848 Casual Rentals) for all 731 days recorded is 29.22 times the variance within the seven days of the week combined.
- P-value: There is a less than 0.01% chance of getting an F-value of 29.22 or more when the average amount of casual bike rentals is the same for all days of the week.
- Conclusion: Since less than 0.01% is less than 5%, we reject H_0. We are 95% confident that at least one day of the week has a different amount of casual bike rentals per day.

Welch's ANOVA for Casual Counts			
Source	DF	F Value	Pr > F
Day of the Week	6.0000	29.22	<.0001
Error	319.4		

The MEANS Procedure

Analysis Variable : Casual Counts casual		
N	Mean	Std Dev
731	848.1764706	686.6224883

Post-hoc Conclusions: PROC MEANS Table

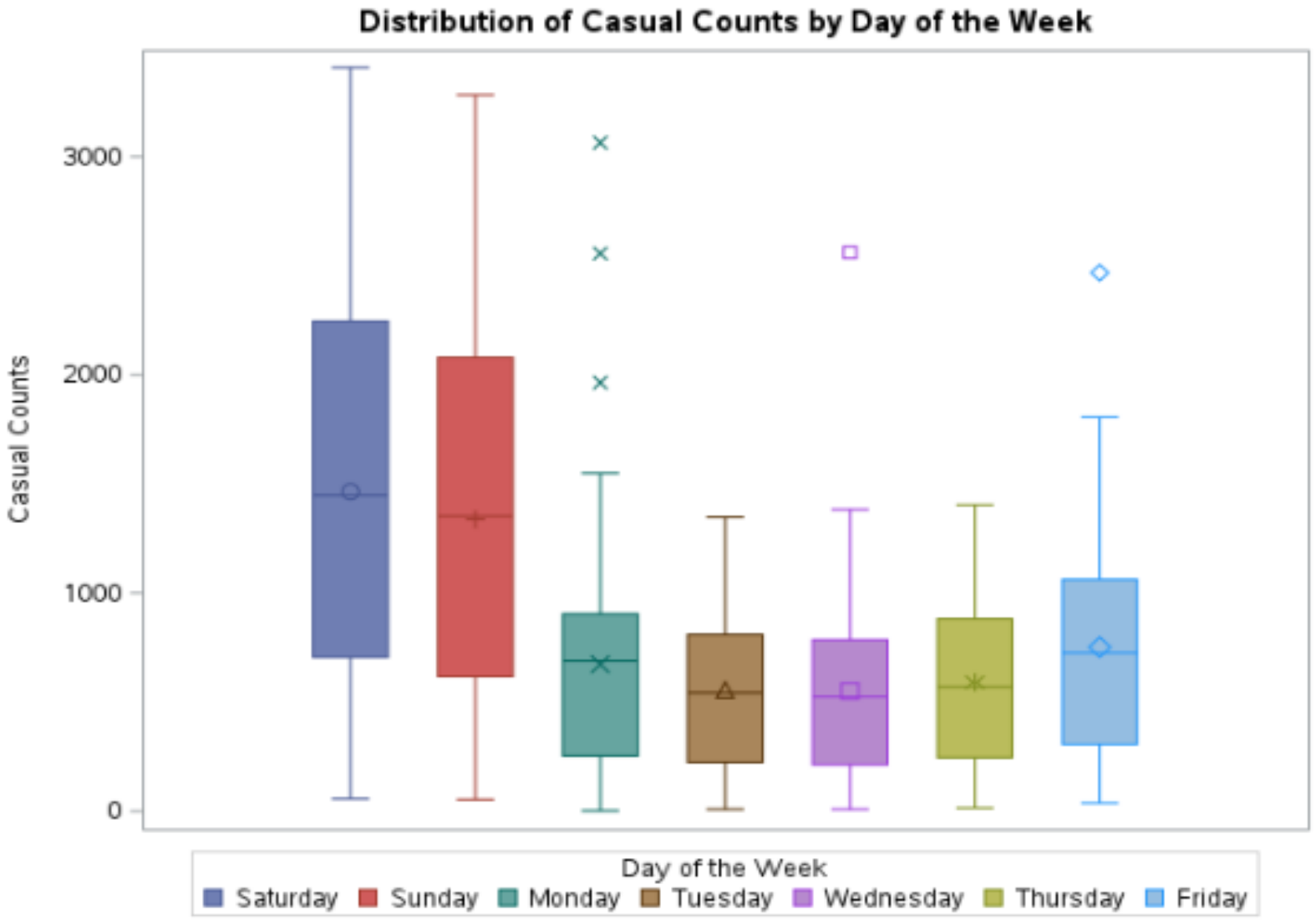
- Conclusion: We know at the minimum, that the largest mean (1465.3 casual bike rentals) is different from the smallest mean (551.1 casual bike rentals). Thus, Saturdays have a significantly higher amount of casual bike rentals than Wednesdays.

Level of Day of the Week	N	Casual Counts	
		Mean	Std Dev
Sunday	105	1338.29524	809.324777
Monday	105	874.13333	493.777298
Tuesday	104	556.18269	342.748705
Wednesday	104	551.14423	401.808954
Thursday	104	590.96154	371.680307
Friday	104	752.28846	483.422400
Saturday	105	1465.25714	927.082884

SAS Code: Stratified Box Plot

```
/*Stratified Box Plot*/
TITLE 'Box Plot for Casual Counts by Day of the Week';
PROC sgplot data= work.bike;
    vbox 'Casual Counts'n / group= 'Day of the Week'n;
    title 'Distribution of Casual Counts by Day of the Week';
    yaxis label= 'Casual Counts';
    xaxis label= 'Day of the Week';
    ODS graphics
        /   attrpriority=none;
RUN;
```

Supporting Graphic: Stratified Box Plot



Mean:
1465.26

Mean:
551.14

Taking Action

- Benefit to Companies: Bike Sharing companies may be interested in which days get more casual bike rentals on average as it helps them gain insight on how to price, when and how to promote, and upkeeping resources.

SAS Code: Screen Recording

SAS® Studio — SAS Programme

SANDER Deliverable 7 SAS CODE.sas

CODE | LOG | RESULTS

```
1  /*Research Question: Does day of the week predict number of casual bike rentals per day?
2   Use alpha = 0.05 */
3  /*Unit of observation: A day in 2011 or 2012*/
4  /*Categorical: Weekday; What day of the week is it?;
5                levels= Sunday, Monday, Tuesday, Wednesday, Thursday, Friday, Saturday*/
6  /*Quantitative: Casual Counts; How may casual bike rentals per day?; units: bike rentals*/
7
8
9  /* Determine hypothesis testing options for answering the question.*/
10 /*
11 1. One-way ANOVA<-normal, homogeneous
12 2. Welch's Test on Raw Data<-normal, heterogeneous
13 3. Kruskal Wallis<-not normal, homogeneous
14 4. Welch's Test on Ranked Data<-not normal, heterogeneous
15 */
16
17
18 /* Import the data set.*/
19 %web_drop_table(WORK.bike);
20
21 FILENAME REFFILE '/home/u62685438/sasuser.v94/stat3130/data/bike.xlsx';
22
23 PROC IMPORT DATAFILE=REFFILE
24     DBMS=XLSX
25     OUT=WORK.bike;
26     GETNAMES=YES;
27 RUN;
28
```

/home/u62685438/sasuser.v94/stat3130/deliverables/SANDER Deliverable 7 SAS CODE.sas