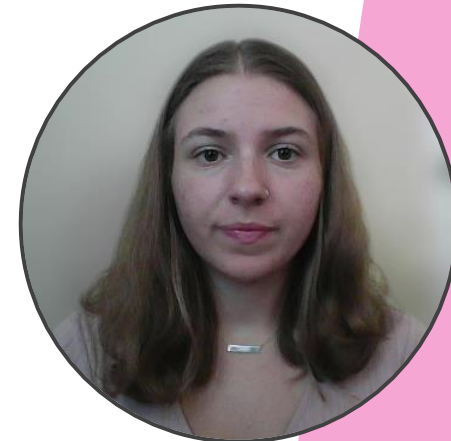


Deliverable 4:

Casual bike rentals on
holidays vs non-holidays

Rebekah Sander



Research Question

Is there a difference in the amount of daily count of casual bike rentals on holidays versus days that are not holidays?



Research Variable

WHO	WHAT measurement is made on each		TYPE OF MEASURE
	Name of Variable	Question Asked	
A day in 2011 and 2012	Holiday	Is it a Holiday?	Categorical Variable Unit: holiday/not a holiday
	Casual Counts	How many casual Bike rentals?	Quantitative Variable Unit: Bike Rentals
One quantitative variable being tested among one categorical variable to see difference amongst levels.			
1. Two Mean t Test (Pooled) 3. Wilcoxon Rank Sum 2. Two Mean t Test (Satterthwaite) 4. Welch's Test on Ranked Data			

SAS Code: Renaming



```
/* Make a new data set and RENAMING variables and cat. vars*/  
data work.bike;  
    set work.bike_full (keep = holiday casual);  
    length holidayC $50;  
    if holiday = 1 then holidayC = 'Holiday';  
    else if holiday = 0 then holidayC = 'Not a Holiday';  
    else holidayC = "Missing";  
    drop holiday;  
    rename holidayC=Holiday;  
    rename casual='Casual Counts'n;  
run;
```

Inspecting Data and Normality



```
/*QQ Plots and normality test*/
title "Figure 1: QQ Plots";
proc univariate data=work.bike normaltest plots;
class Holiday;
VAR 'Casual Counts';
run;
title;
```

Table 1: Summary Statistics for Holiday

Moments			
N	21	Sum Weights	21
Mean	1064.71429	Sum Observations	22359
Std Deviation	860.046809	Variance	739680.514
Skewness	1.0267872	Kurtosis	0.16546868
Uncorrected SS	38599557	Corrected SS	14793610.3
Coeff Variation	80.7772396	Std Error Mean	187.6776

Table 2: Summary Statistics for not a Holiday

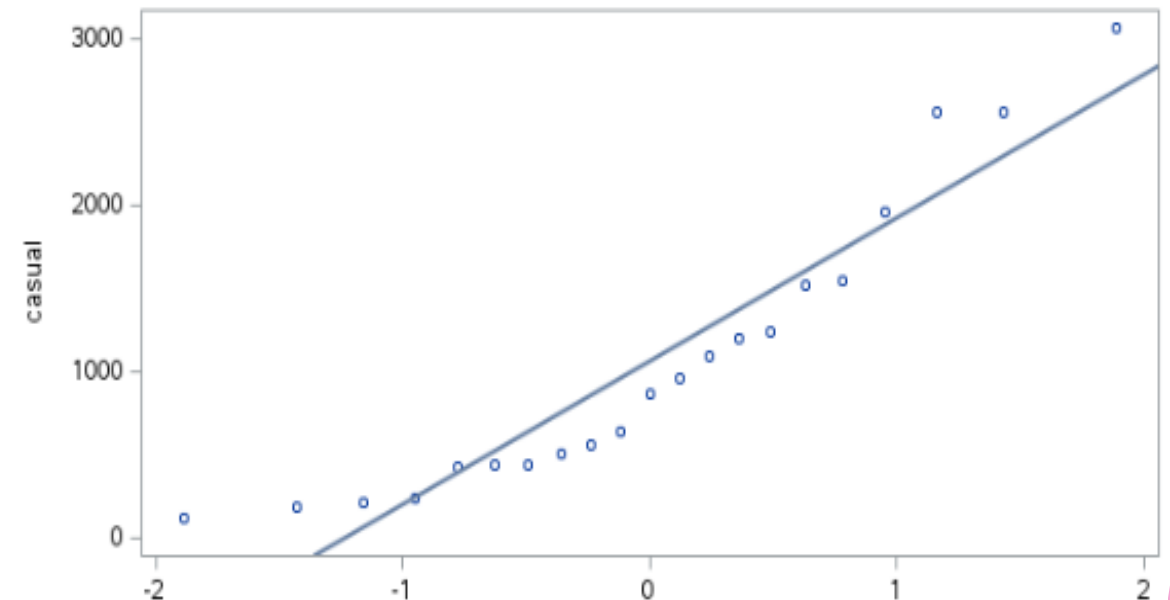
Moments			
N	710	Sum Weights	710
Mean	841.771831	Sum Observations	597658
Std Deviation	680.528529	Variance	463119.079
Skewness	1.27080148	Kurtosis	1.36760645
Uncorrected SS	831443096	Corrected SS	328351427
Coeff Variation	80.8447734	Std Error Mean	25.5397806

Assessing Normality--Holiday



- ▶ H_0 : The data came from a normal population
 H_A : The data did not come from a normal population
- ▶ **Normality Tests:** Out of the four tests for normality, three tests show p-values less than $\alpha = 0.05$. Thus, we have evidence to suggest the data is not normal.
- ▶ **QQ Plot:** The data shows large amounts of deviation from the agreement line, contradicting normality. Further, this agrees with the normality tests that say that the data is not normal.

Tests for Normality				
Test	Statistic		p Value	
Shapiro-Wilk	W	0.882208	Pr < W	0.0161
Kolmogorov-Smirnov	D	0.164655	Pr > D	0.1404
Cramer-von Mises	W-Sq	0.135192	Pr > W-Sq	0.0359
Anderson-Darling	A-Sq	0.872208	Pr > A-Sq	0.0218

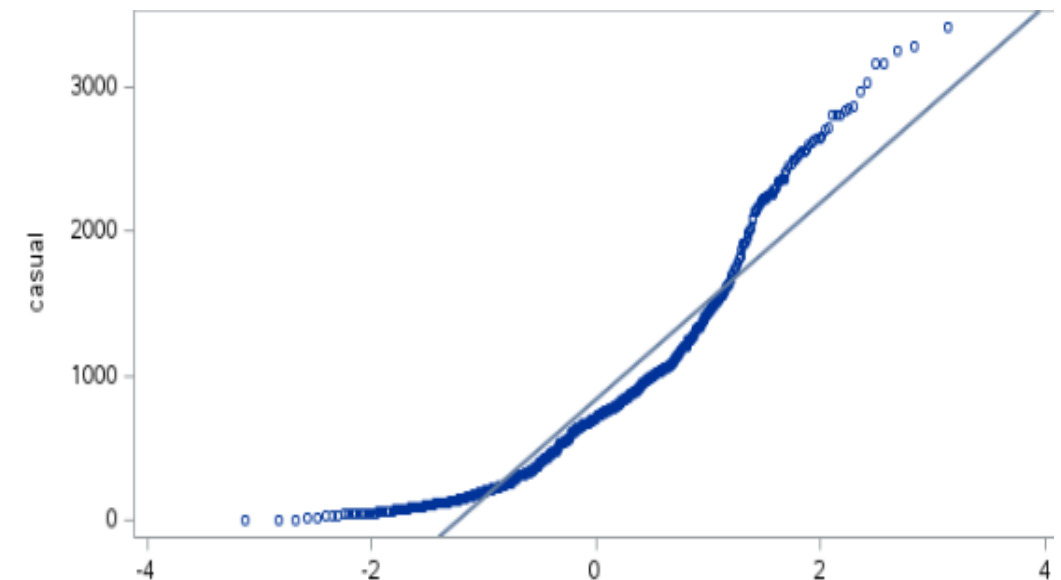




Assessing Normality—Not a Holiday

- ▶ H_0 : The data came from a normal population
 H_A : The data did not come from a normal population
- ▶ **Normality Tests**: All four of the tests for normality show p-values less than $\alpha = 0.05$. Thus, we have evidence to suggest the data is not normal.
- ▶ **QQ Plot**: The data shows large amounts of deviation from the agreement line, further contradicting that the data is normal
- ▶ **Sample size**: The sample size of non-holidays is 710 days. This follows the Central Limit Theorem as 710 is above 30. Thus, the data distribution is normal.

Tests for Normality				
Test	Statistic		p Value	
Shapiro-Wilk	W	0.885581	Pr < W	<0.0001
Kolmogorov-Smirnov	D	0.114097	Pr > D	<0.0100
Cramer-von Mises	W-Sq	3.314881	Pr > W-Sq	<0.0050
Anderson-Darling	A-Sq	21.89839	Pr > A-Sq	<0.0050





Assessing Homogeneity

- ▶ H_0 : Data is Homogeneous
 H_A : Data is Non-Homogeneous
- ▶ To assess homogeneity, we will look at the ratio of the standard deviations.
- ▶ The ratio is less than 2. Thus, the standard deviations are close enough to use a test that requires homogeneity.
- ▶ **Conclusion:** Use Wilcoxon Rank Sum Test.

Ratio of standard deviations:

$$\begin{aligned} &= \frac{SD_{Holiday}}{SD_{Not a Holiday}} \\ &= \frac{860.05}{680.53} \\ &= 1.2638 \end{aligned}$$

Choosing Hypothesis Test: Wilcoxon Rank Sum Test

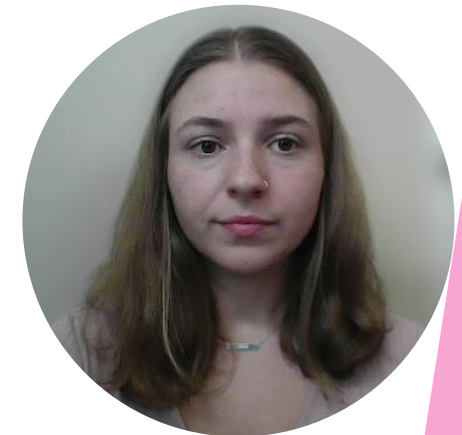
- ▶ Since the data is not normal and homogeneous, we will perform the Wilcoxon Rank Sum Test.

$$H_0: \eta_{non-holiday} = \eta_{holiday}$$

$$H_A: \eta_{non-holiday} \neq \eta_{holiday}$$

$$\alpha = 0.05$$

- ▶ The null hypothesis is that the median of casual rentals on non-holidays is the same as the median of casual rentals on holidays.
- ▶ The alternative hypothesis is that the median of casual rentals on non-holidays is different than the median of casual rentals on holidays.
- ▶ The level of significance, $\alpha = 0.05$, tells us that 5% of the time we will conclude $\eta_{non-holiday} \neq \eta_{holiday}$ when $\eta_{non-holiday} = \eta_{holiday}$ is actually true.



Wilcoxon Rank Sum Test--Calculations

9



The Expected
Sum of Rank

$$= \left(\frac{n_{\text{smallest}}}{n} \right) \left(\frac{(n)(n+1)}{2} \right)$$

$$= \left(\frac{21}{731} \right) \left(\frac{(731)(732)}{2} \right)$$

$$= 7,686$$

Z-value

=

$$\frac{(\text{Sum of Scores} + 0.5) - \text{Expected Sum of Ranks}}{\text{Standard Deviation}}$$

$$= \frac{(8740.5 + 0.5) - 7,686}{953.681216}$$

$$= 1.1062$$

Performing Wilcoxon Rank Sum Test



```
/*Running the Wilcoxon Rank Sum with correction*/
PROC NPAR1WAY DATA = work.bike WILCOXON CORRECT=YES;
  CLASS Holiday;
  VAR 'Casual Counts';
RUN;
```

Wilcoxon Scores (Rank Sums) for Variable Casual Counts Classified by Variable Holiday					
Holiday	N	Sum of Scores	Expected Under H0	Std Dev Under H0	Mean Score
Not a Holiday	710	258805.50	259880.0	953.681216	364.514789
Holiday	21	8740.50	7688.0	953.681216	416.214286
Average scores were used for ties.					

Wilcoxon Two-Sample Test					
Statistic	Z	Pr > Z	Pr > Z	t Approximation	
				Pr > Z	Pr > Z
8740.500	1.1052	0.1345	0.2691	0.1347	0.2694
Z includes a continuity correction of 0.5.					

Performing Wilcoxon Rank Sum Test



- **Z-value:** The Wilcoxon statistic (sum of ranks for holiday sample), which is 8740.50 is 1.1052 standard deviations to the right of the hypothesized sum of the ranks, 7686. It is the proportion of the sum of the ranks 1 to 21 that we should see for the holiday sample.
- **p-value:** There is a 26.91% chance of observing this sum of ranks for the smallest sample (holiday) if there was no difference in the medians for casual bike rentals on non-holidays and holidays.
- **Conclusion:** Since the sample results are so likely to happen by chance when $\eta_{non-holiday} = \eta_{holiday}$, p-value = 0.2691 > $\alpha = 0.05$, we cannot say that there is a difference in the casual bike rental counts for holidays and non-holidays.

Wilcoxon Scores (Rank Sums) for Variable Casual Counts Classified by Variable Holiday					
Holiday	N	Sum of Scores	Expected Under H0	Std Dev Under H0	Mean Score
Not a Holiday	710	258805.50	258860.0	953.681216	364.514789
Holiday	21	8740.50	7686.0	953.681216	416.214286
Average scores were used for ties.					

Wilcoxon Two-Sample Test					
Statistic	Z	Pr > Z	Pr > Z	t Approximation	
				Pr > Z	Pr > Z
8740.500	1.1052	0.1345	0.2691	0.1347	0.2694
Z includes a continuity correction of 0.5.					

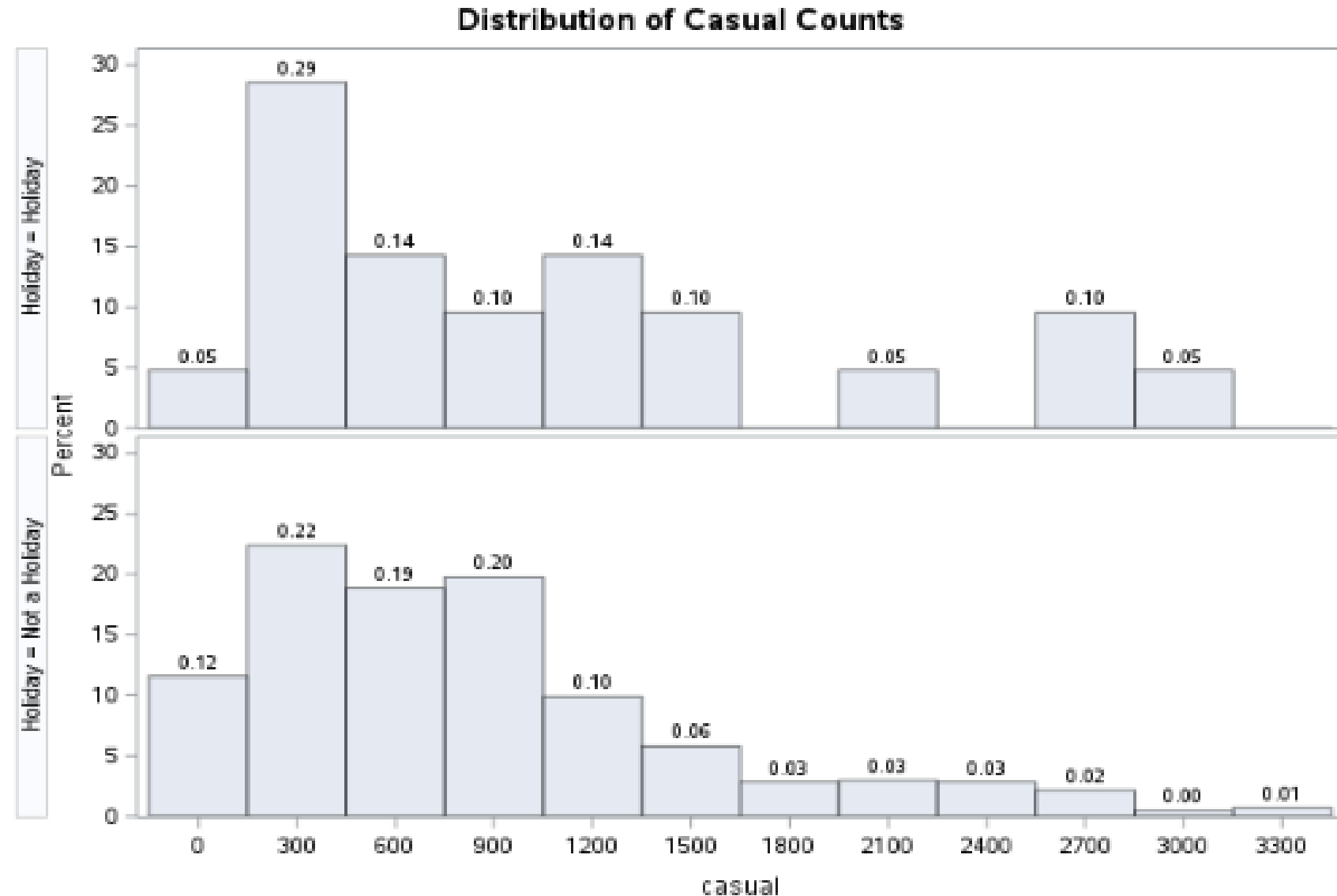
Figure 3: Histogram of Casual Bike Rental Counts by Whether it is a Holiday

Bin width is 300 rental counts

The UNIVARIATE Procedure



Supporting
Graphic:
Histogram



SAS Code—Stratified Histogram



```
/*Stratified Histogram*/  
PROC UNIVARIATE DATA = work.bike noprint;  
  VAR 'Casual Counts'n;  
  CLASS Holiday;  
  HISTOGRAM/barlabel=proportion midpoints=0 to 3500 by 300 ;  
  TITLE1 height=16pt 'Figure 3: Histogram of Casual Bike Rental Counts by Whether it is a Holiday';  
  Title2 height=12pt 'Bin width is 300 rental counts';  
RUN;  
title;
```

Taking Action



This analysis may be helpful not only for bike sharing companies, but also to the cities that use these companies.

- ▶ **Companies:** Bike Sharing companies may be interested in casual bike rentals on holidays vs non-holidays as it helps them gain insight on how to price, when and how to promote, and upkeeping resources.
- ▶ **Cities:** Cities may be interested in casual bike rentals on holidays vs non-holidays when assessing the mobility of the city, going-ons, and other patterns about their population.

SAS Code Screen Recording

SAS® Studio

SAS Programmer

SANDER Deliverable 4 SAS CODE.sas

CODE LOG RESULTS OUTPUT DATA

Line #

```
1 /*Research Question: Is there a difference in the amount of daily count of casual bike rentals
2  on holidays versus days that are not holidays?*/
3 /*Unit of observation: A day in 2011 and 2012*/
4 /*Categorical: holiday; whether or not day is a holiday; 0-not a holiday , 1-holiday*/
5 /*Quantitative: casual; count of casual rentals; casual rentals per day*/
6
7
8 /* Determine hypothesis testing options for answering the question.*/
9 /*
10 1. Two mean t-test (pooled)
11 2. Two mean t-test (satterthwaite)
12 3. wilcox rank sum
13 4. welch's test on ranked data
14 */
15
16
17 /* Import the data set.*/
18 %web_drop_table(WORK.bike_full);
19
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_full;
```

/home/u62685438/sasuser.v94/stat3130/deliverables/SANDER Deliverable 4 SAS CODE.sas

Line 1, Column 1

Message

