MBC 638 | EXERCISE IMPROVEMENT PROJECT

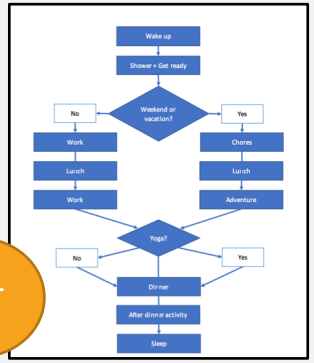
PROCESS OWNER: KENDRA OSBURN

KEY DATES → DEFINE: 7.19 | MEASURE: 7.22 | ANALYZE: 8.15 | IMPROVE: 8.20 | CONTROL: 9.15

DEFINE

PROBLEM TO SOLVE:

Not enough minutes spent being active during the day. According to the Center for Disease Control (CDC), I am not getting enough exercise for someone my age.



SQL 1

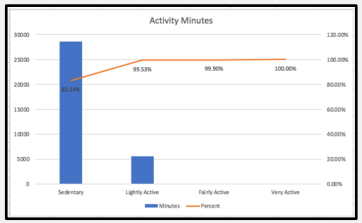
BUSINESS IMPACT:

According to the CDC, increased exercise leads to both physical health and mental health, both of which add to my quality of life and impact my ability to be an effective worker (and earn a wage). Additionally, lack of exercise can lead to health problems. Health problems are expensive.

MEASURE

COLLECT DATA ON 11 VARIABLES:

Calories (Y), Minutes sedentary (X5), Minutes Lightly Active (X6), Fairly Active (X7) Very active (X8) etc.



More time spent being sedentary than being active. 0% of my time is being "fairly active" or "very active."

BEFORE IMPROVEMENT

Mean	1973.93333
Standard Error	45.5475988
Median	2024.5
Mode	2013
Standard Deviation	249.474473
Sample Variance	62237.5126
Kurtosis	-0.1797323
Skewness	-0.3013388
Range	1000
Minimum	1443
Maximum	2443
Sum	59218
Count	30

ANALYZE

HYPOTHESIS TESTING:

P > a so null hypothesis is rejected. The average calories burned is greater than the old average calories burned.

PREDICTIVE EQUATION

Y=1438 +159distance - 5.2floors + 0.54minutesLightlyActive

3Xs have meaningful relationships with Y: Distance, Minutes Lightly Active, Floors Climbed

IMPROVE

CHANGES MADE TO THE PROCESS:

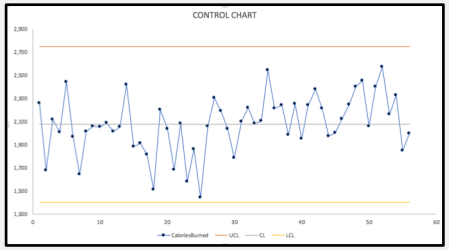
1.) Prioritize exercise by adding it directly into the daily schedule (instead of hoping it would just "happen")
2.) Acquire a stationary bike
3.) Use stationary bike 5x a week before morning shower

SQL 2.3!

AFTER IMPROVEMENT

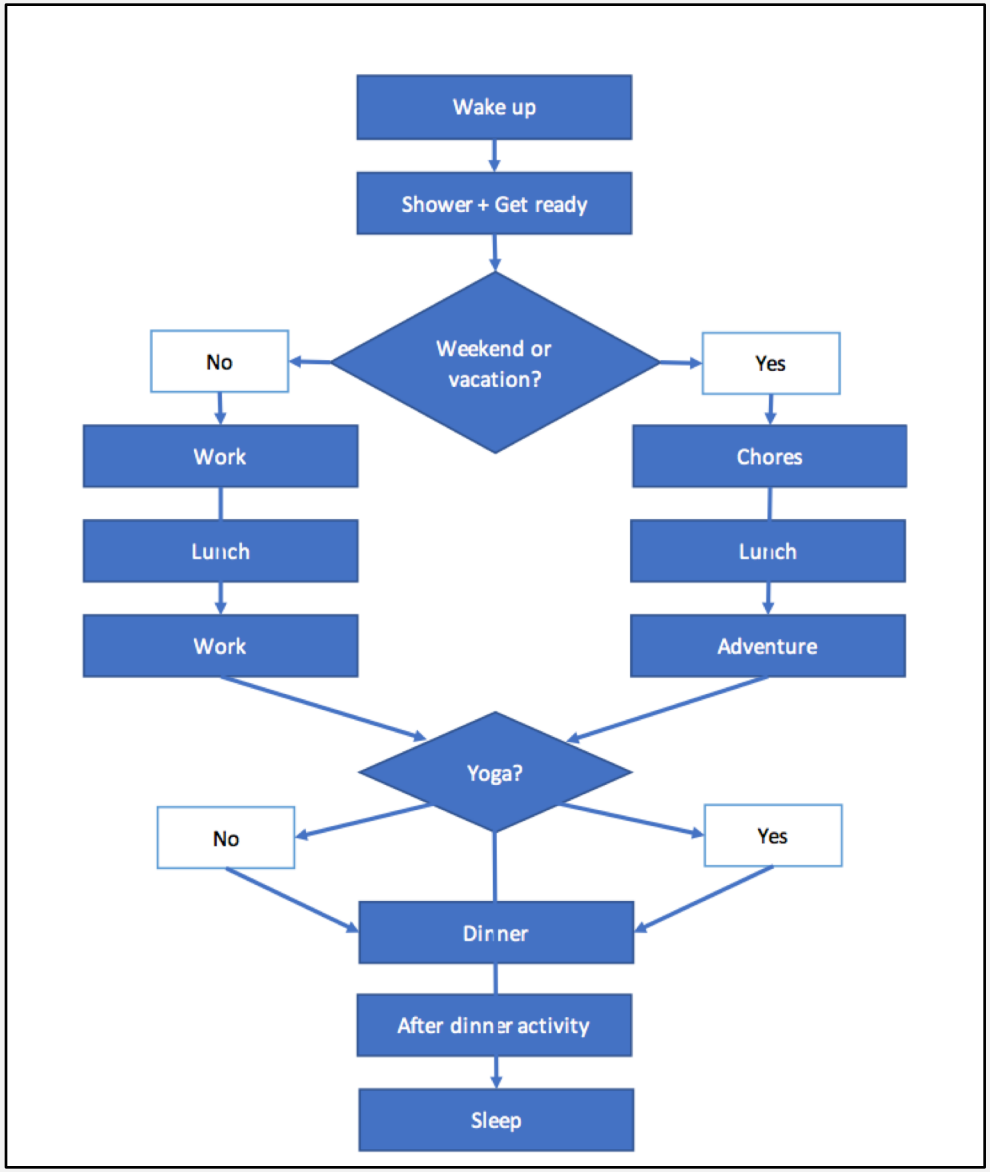
Mean	2196.76923
Standard Error	36.8175254
Median	2213
Mode	2213
Standard Deviation	187.733281
Sample Variance	35243.7846
Kurtosis	-0.5260345
Skewness	0.25343529
Range	724
Minimum	1847
Maximum	2571
Sum	57116
Count	26

CONTROL



The new process shows stable variation over time. It is within the control limits and if the process remains the same, should continue to be within the control limits.

CURRENT PROCESS: THE PROCESS MAP



- To better understand and visualize my current process, I analyzed my daily schedule and made a process map.
- This process map shows that, exercise has not been a priority when planning my day, despite strong scientific evidence* of the wellness benefits of exercise (including evidence that exercise is just as important a factor as sleep and diet).
- * see appendix for references

DATA MEASUREMENT PLAN

Performance Measure	Data Source & Location	How Data are Collected	Data Collector	Data Collection Timeframe	Sample Size
Calories Burned	Fitbit + Google Sheet	From Fitbit website into Google Sheet. Moved to Excel for final calculation.	Kendra Osburn	7/27 to 9/15	56
Day (Discrete)	Google Sheet	Autofill in Google Sheet. Moved to Excel for final calculation.	Kendra Osburn	7/27 to 9/15	56
Steps (Discrete)	Fitbit + Google Sheet	From Fitbit website into Google Sheet. Moved to Excel for final calculation.	Kendra Osburn	7/27 to 9/15	56
Distance	Fitbit + Google Sheet	From Fitbit website into Google Sheet. Moved to Excel for final calculation.	Kendra Osburn	7/27 to 9/15	56
Floors (Discrete)	Fitbit + Google Sheet	From Fitbit website into Google Sheet. Moved to Excel for final calculation.	Kendra Osburn	7/27 to 9/15	56
Minutes Sedentary	Fitbit + Google Sheet	From Fitbit website into Google Sheet. Moved to Excel for final calculation.	Kendra Osburn	7/27 to 9/15	56
Minutes Lightly Active	Fitbit + Google Sheet	From Fitbit website into Google Sheet. Moved to Excel for final calculation.	Kendra Osburn	7/27 to 9/15	56
Minutes Fairly Active	Fitbit + Google Sheet	From Fitbit website into Google Sheet. Moved to Excel for final calculation.	Kendra Osburn	7/27 to 9/15	56
Minutes Very Active	Fitbit + Google Sheet	From Fitbit website into Google Sheet. Moved to Excel for final calculation.	Kendra Osburn	7/27 to 9/15	56
Activity Calories	Fitbit + Google Sheet	From Fitbit website into Google Sheet. Moved to Excel for final calculation.	Kendra Osburn	7/27 to 9/15	56
Yoga (Discrete)	Google Sheet	Manual Entry into Google Sheet. Moved to Excel for final calculation	Kendra Osburn	7/27 to 9/15	56

DATA COLLECTION METHOD

COLLECTION METHOD:

- Continuous and discrete data were collected every day between and including July 22nd and September 15th (56 days total). The first 30 days of data were part of the "**MEASURE**" phase, while the last 26 constituted the **"IMPROVEMENT"** phase.
- Nine of the 11 points of data were tracked and initially recorded by Fitbit. As "data collector," my first job was to wear my Fitbit device daily, while my second job was to export the data from Fitbit into Google Sheets (where I stored working numbers), then into Excel (where I performed my end analysis).
- The final two points of data were either provided by Google Sheets auto-fill (for "Day of the Week"), or manually entered by me, the data collector (for after work yoga, as "yes" or "no").

CONTINUOUS DATA

8 Variables | From Fitbit

Fitbit tracked calories burned, steps, distance traveled, minutes sedentary, minutes lightly active, minutes fairly active, minutes very active and active calories.

DISCRETE DATA

3 Variables | 1 From Fitbit

Fitbit tracked floors climbed. Google Sheets tracked day of the week. I tracked evening yoga.

SAMPLE SIZE

The total sample size is 56.

Before improvement: 30

After improvement: 26

SAMPLE SIZE FOR CONTINUOUS DATA

90% confidence: 1.645

Margin of Error: 45.55

Standard Deviation: 249.5

$$n = \left(\frac{z^* \hat{\sigma}}{E} \right)^2$$

= 81.18 (rounded up to nearest person is 82)

Unfortunately, given the time constraints of the project, I was unable to get the needed sample size for a 90% Confidence Interval. Fortunately, this will be an ongoing process!

LEVEL OF MEASUREMENT ERROR

The measurement error was a result of the limitations of my Fitbit. Like most wearables, Fitbit devices are susceptible to certain technical issues, and are therefore not wholly reliable sources of data. For example, on a handful of days (<4), my Fitbit did not record any minutes of any activity, even though my behavior was virtually identical to that of the previous day, when activity was recorded. Further research on Fitbit's website suggested that this may have been a result of multiple factors, including where my Fitbit sat on my wrist, the battery level of the device, or some other factor we don't yet understand ("the buildup of dead skin" was suggested, but not confirmed).

SUCCESS MEASURES

GOAL

The goal of this Process Improvement Project is to increase the amount of time spent exercising, here measured in calories burned.

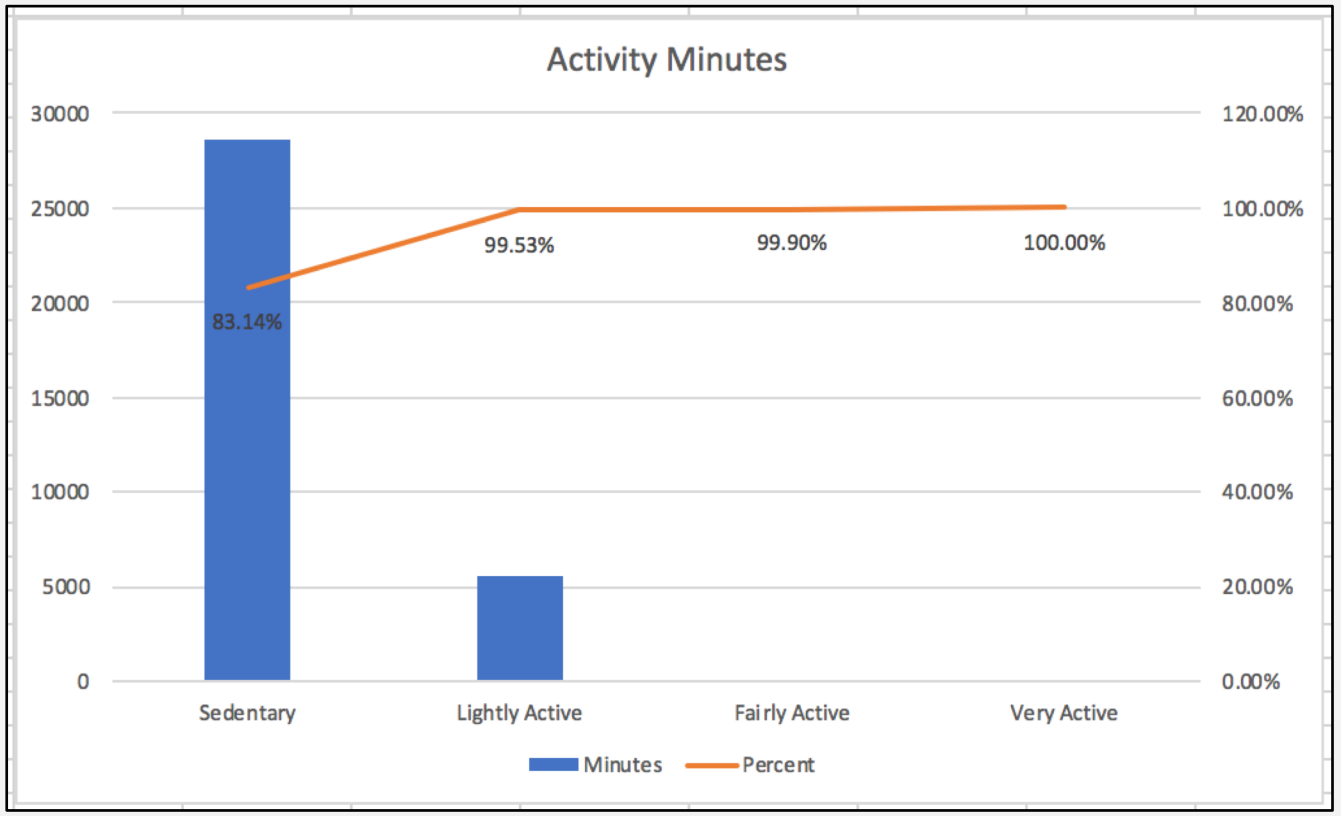
OPERATIONAL DEFINITIONS

- "Time spent exercising" can be calculated in many ways based on the data provided by Fitbit. Given the nature of the study and the main process improvement metric (i.e. the addition of a bicycle), "distance" didn't seem like the best measurement (in case Fitbit wasn't able to differentiate distance traveled while on a bike vs. while walking), while "steps" were considered discrete data which would cause larger problems for the analysis in the long-term. Ultimately, instead of creating my own metric ("Energy Units Exerted," as a potential unit of "step"), I decided to use calories as the continuous data value.
- In short, "Time spent exercising" is the total number of calories burned during the data collection period, as recorded by Fitbit.
- A defect is defined as a day without any "Fairly Active" or "Very Active" minutes.
- Day of the Week is defined as Saturday, Sunday, Monday, Tuesday, Wednesday, Thursday Friday, Saturday or Sunday, depending on the date when the data was collected.

BASELINE

Before the improvement, an average of 1974 calories were burned each day, with a standard deviation of 249.5 and an SQL value of 1.

CURRENT PROCESS: PARETO CHART



WHY THE PARETO CHART?

To visualize and understand the differences between minutes spent sedentary, lightly active, fairly active and very active minutes.

RESULTS

The Pareto Chart clearly indicates the dramatic difference in minutes spent being sedentary or active. For time to be categorized as "fairly active" or "very active," something dramatic must change.

DESCRIPTIVE STATISTICS

WHY DESCRIPTIVE STATISTICS?

To measure central tendency and dispersion in the dataset before and after the process improvement.

BEFORE IMPROVEMENT	
Mean	1973.93333
Standard Error	45.5475988
Median	2024.5
Mode	2013
Standard Deviation	249.474473
Sample Variance	62237.5126
Kurtosis	-0.1797323
Skewness	-0.3013388
Range	1000
Minimum	1443
Maximum	2443
Sum	59218
Count	30

AFTER IMPROVEMENT	
Mean	2196.76923
Standard Error	36.8175254
Median	2213
Mode	2213
Standard Deviation	187.733281
Sample Variance	35243.7846
Kurtosis	-0.5260345
Skewness	0.25343529
Range	724
Minimum	1847
Maximum	2571
Sum	57116
Count	26

RESULTS

Within the new process, the mean and median calories burned increased 10.14% and 8.52%, respectively. The range decreased due to the minimum increasing by 404 calories.

SQL: SIGMA QUALITY LEVEL

WHY SQL?

To determine a baseline for the process before any improvements began. This baseline provides a point of comparison for the improved data, which helps determine if the improvement plan was successful.

DEFINITION OF DEFECT

A "defective unit" is defined as a day where both "Fairly Active" and "Very Active" minutes were 0.

RESULT

According to the change in SQL values, the improvement project did in fact result in a decrease in "defective" units. Before the process improvement, the SQL was 1. Afterwards, the SQL increased to 2.3, while the defect rate decreased from 73% to 19%.

BEFORE IMPROVEMENT

Defect opportunities per unit	1
Units produced	30
Total possible defects	30
Total actual defects	22
Defect per opportunity rate	0.73
DPMO	733333.333
SQL value	1

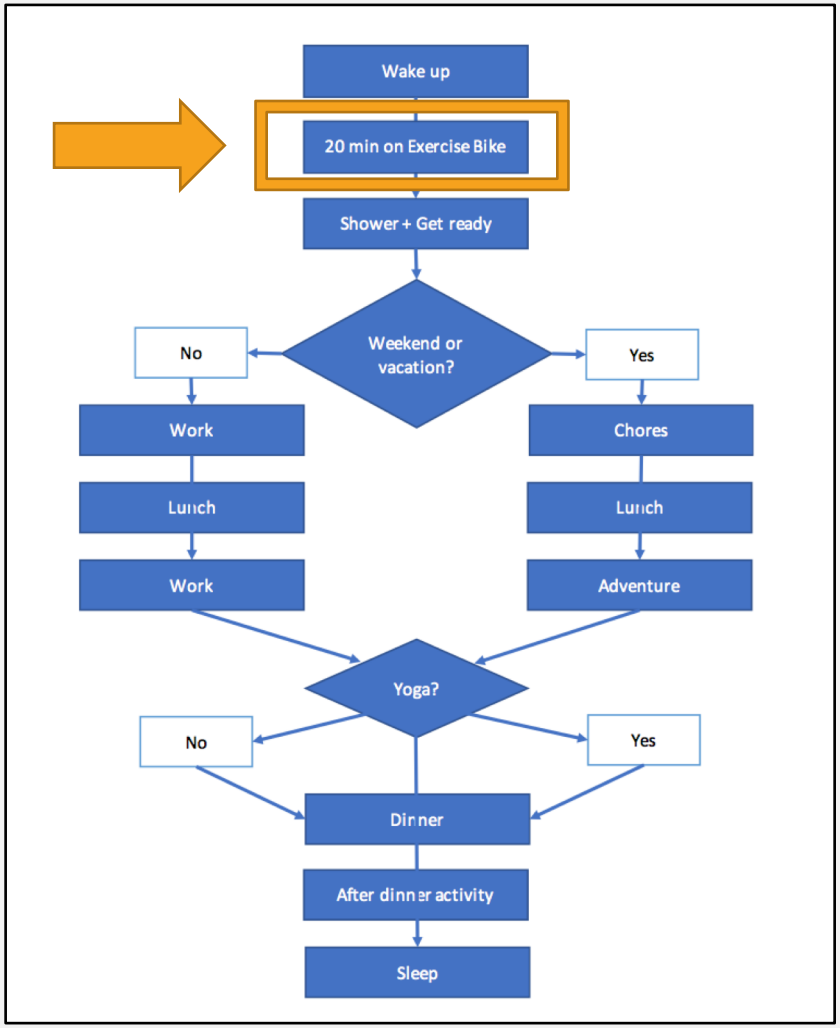
AFTER IMPROVEMENT

Defect opportunities per unit	1
Units produced	26
Total possible defects	26
Total actual defects	5
Defect per opportunity rate	0.19
DPMO	192307.692
SQL value	2.3

THE PROCESS CHANGE

After analyzing the 30 days of data, it became clear that the entire process needed to change in order to prioritize exercise and "moderate activity" ("fairly active" and "very active" combined).

- **The Process Map** succinctly identified all the activities that happened within my standard day. Things like "lunch," "dinner" and even "getting ready" had dedicated spots in my schedule, but "exercise" did not. While time for "Yoga" was included the schedule, Fitbit frequently tracked this time as "Lightly Active" at best.
- **The Pareto Chart** showed just how sedentary I'd become, and clearly highlighted the scale of change needed to increase my "moderately active" minutes.
- **The SQL** showed a moderate level of defects. To increase the SQL for the new process**, I committed to at least 20 minutes of exercise before my morning shower and "getting ready" time, 5 days a week.** While I initially intended to assign this exercise to "weekdays," I decided to give myself the option to sleep in one weekday, given that I already do some form of exercise on Saturday or Sunday.



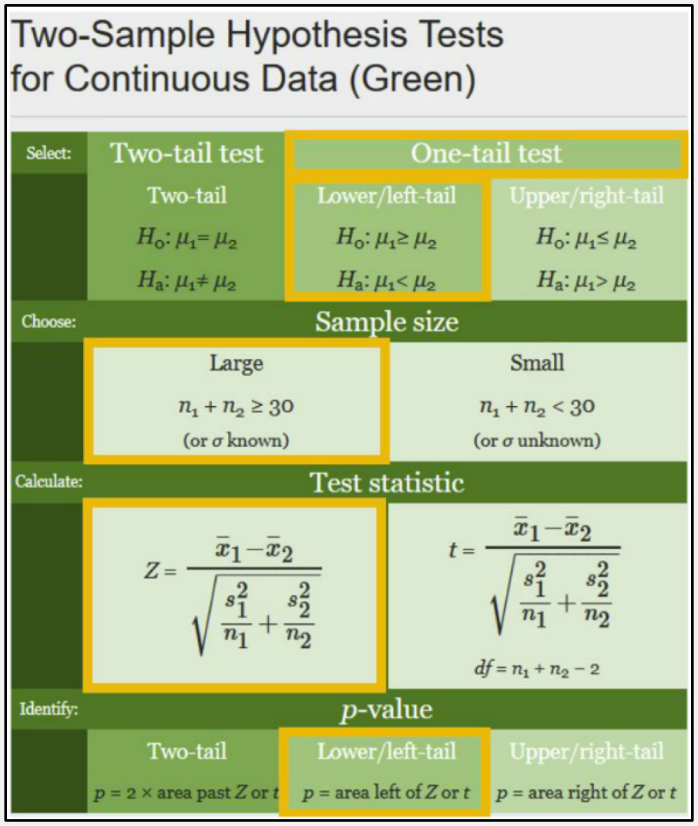
HYPOTHESIS TESTING

WHY HYPOTHESIS TESTING?

To determine whether the initial test data supports the hypothesis of this study: that after the process change, average exercise time will be greater than the average exercise time before the change.

HYPOTHESIS STATEMENTS

- Ho: Exercise time before is greater than or equal to exercise time after
  - Ho: μ1 >=μ2
- Ha: Exercise time before is less than exercise time after
  - Ha : μ1<μ2



RESULTS: *If p is low, Ho must go!*

Since the P-Value is well below the alpha of 0.05, the null hypothesis can be safely rejected.

	Variable 1	*Variable 2*
Mean	1973.933333	2196.769231
Known Variance	62237.5	35243.78
Observations	30	26
Hypothesized Mean Difference	0	
z	-3.804791996	
P(Z<=z) one-tail	7.09617E-05	
z Critical one-tail	1.644853627	
P(Z<=z) two-tail	0.000141923	
z Critical two-tail	1.959963985	

REGRESSION ANALYSIS (BEFORE THE IMPROVEMENT)

WHY REGRESSION ANALYSIS?

To determine which, if any, of the X variables have relationships with the Y variable, and the strength of those relationships. This will provide another baseline for comparison.

RESULTS

The results are excellent for predicting -- R .98 is a strong relationship. The adjusted R Square suggests that this accounts for 95% variability. Three, almost four X-values have a p-value < 0.05. Our significance is well below 0.05.

Regression Statistics	
Multiple R	0.98257887
R Square	0.96546124
Adjusted R Square	0.95230362
Standard Error	54.4839825
Observations	30

ANOVA

	df	*SS*	*MS*	*F*	*Significance F*
Regression	8	1742549.28	217818.659	73.3765674	1.3191E-13
Residual	21	62338.5913	2968.50435		
Total	29	1804887.87			

	Coefficients	*Standard Erro*	*t Stat*	*P-value*	*Lower 95%*	*Upper 95%*	*Lower 95.0%*	*Upper 95.0%*
Intercept	1318.63592	68.4624928	19.2607056	7.9551E-15	1176.26037	1461.01146	1176.26037	1461.01146
Steps	0.28205331	0.07122225	3.96018557	0.00071454	0.13393854	0.43016808	0.13393854	0.43016808
Distance	-453.50722	161.6105	-2.8061743	0.0105821	-789.59466	-117.41979	-789.59466	-117.41979
Floors	-0.1194899	1.57843328	-0.0757016	0.94037307	-3.4020216	3.16304181	-3.4020216	3.16304181
MinutesSedentary	0.03643262	0.05473377	0.66563331	0.51289113	-0.0773925	0.15025772	-0.0773925	0.15025772
MinutesLightlyActive	2.20997697	1.10599813	1.99817424	0.05881184	-0.090072	4.51002598	-0.090072	4.51002598
MinutesFairlyActive	-3.7760658	2.07468149	-1.8200701	0.08303292	-8.0906022	0.53847053	-8.0906022	0.53847053
MinutesVeryActive	23.2193954	6.7827248	3.4233138	0.00255428	9.11394702	37.3248438	9.11394702	37.3248438
ActivityCalories	-0.6734568	0.36105055	-1.8652702	0.07617793	-1.4243026	0.0773889	-1.4243026	0.0773889

REGRESSION ANALYSIS (AFTER THE IMPROVEMENT)

RESULTS

Running the regression analysis after the Process Improvement proved even *more* useful for predictive modeling.

PREDICTIVE EQUATION

Y=1438 +159distance - 5.2floors + 0.54minutesLightlyActive

Regression Statistics	
Multiple R	0.94986601
R Square	0.90224544
Adjusted R Square	0.88362552
Standard Error	64.0427747
Observations	26

ANOVA

	df	*SS*	*MS*	*F*	*Significance F*
Regression	4	794963.599	198740.9	48.4559343	2.6094E-10
Residual	21	86131.0169	4101.47699		
Total	25	881094.615			

	Coefficients	*Standard Erro*	*t Stat*	*P-value*	*Lower 95%*	*Upper 95%*	*Lower 95.0%*	*Upper 95.0%*
Intercept	1438.14446	66.5435061	21.6120933	7.9157E-16	1299.75967	1576.52926	1299.75967	1576.52926
Distance	159.307147	15.5256943	10.260871	1.2318E-09	127.019698	191.594596	127.019698	191.594596
Floors	-5.2115202	2.12574747	-2.4516177	0.02305934	-9.632254	-0.7907863	-9.632254	-0.7907863
MinutesLightlyActive	0.54005055	0.23607635	2.28760975	0.0326374	0.04910291	1.03099819	0.04910291	1.03099819
ActivityCalories	-0.0842075	0.06269336	-1.3431638	0.19355594	-0.2145854	0.04617053	-0.2145854	0.04617053

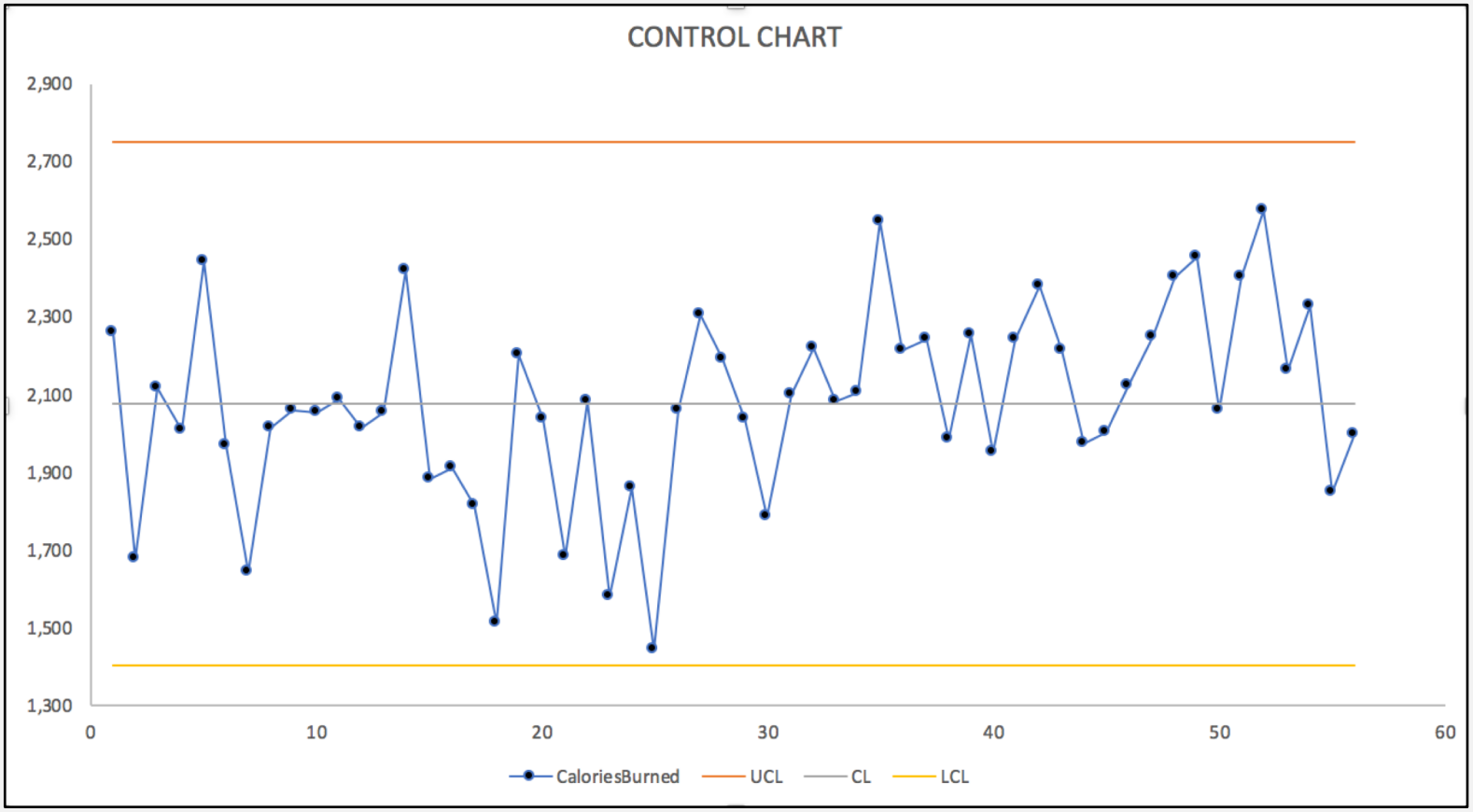
IMR (XMR) CONTROL CHART

WHY CONTROL CHART?

To plot exercise time over the entire data collection period and observe any unnatural outliers.

RESULTS

The old process, day 30 and before, was mostly below the CL. The new process, day 31+ was mostly above the CL, meaning the improvement is working!



CONCLUSION: SUMMARY

PURPOSE

The purpose of this project is to increase time spent exercising. The original goal was to increase time spent exercising, as measured in step count. After it was determined (via the Process Definition Worksheet) that "step count" is a discrete measurement, the performance metric was changed to "distance traveled." However, when the time came to make an improvement to the process, the metric shifted again, this time to calories burned per day, to avoid any potential Fitbit measurement error between biking and walking "distance."

TOOLS

To reach this goal, I employed several statistical methods to analyze the process and determine if the intervention actually improved calories burned per day.

- *Tools used:* The original Problem Definition Worksheet, the Process Map (before and after), the Pareto Chart, Descriptive Statistics (before and after), SQL and Regression Analysis

CONCLUSION: RESULTS

RESULTS

- The **Descriptive Statistics** are the first indication that the Improvement Process is a success. The **mean** and **median** increased 10.14% and 8.52% respectively once the new process was introduced.
- The change of **SQL** values before and after the Process Improvement also indicates that the improvement attempt was a success. Before the improvement was introduced, the SQL was 1. After, the SQL increased to 2.3 and the defect rate decreased from 73% to 19%.
- The strength of the **Regression Analysis** gives credence to both data sets with the strong R value and low p-value. Additionally, the post-improvement regression analysis shows the *very* strong predictive capabilities of certain X variables (Distance, Minutes Lightly Active and Floors Climbed).
- The **IMR (XmR) Control Chart** has been set in motion to continue to monitor the new and improved process
- **The data collected and the measurement metrics used determined that the improvement process succeeded in increasing exercise activity.** However, a larger study with additional data points is definitely needed and, as always, further study is encouraged.

APPENDIX | SOURCES

- https://www.cdc.gov/physicalactivity/data/facts.htm
- https://www.cdc.gov/physicalactivity/basics/adults/index.htm
- https://www.cdc.gov/physicalactivity/basics/pa-health/index.htm
- https://www.cdc.gov/diabetes/prevention/pdf/postcurriculum_session8.pdf