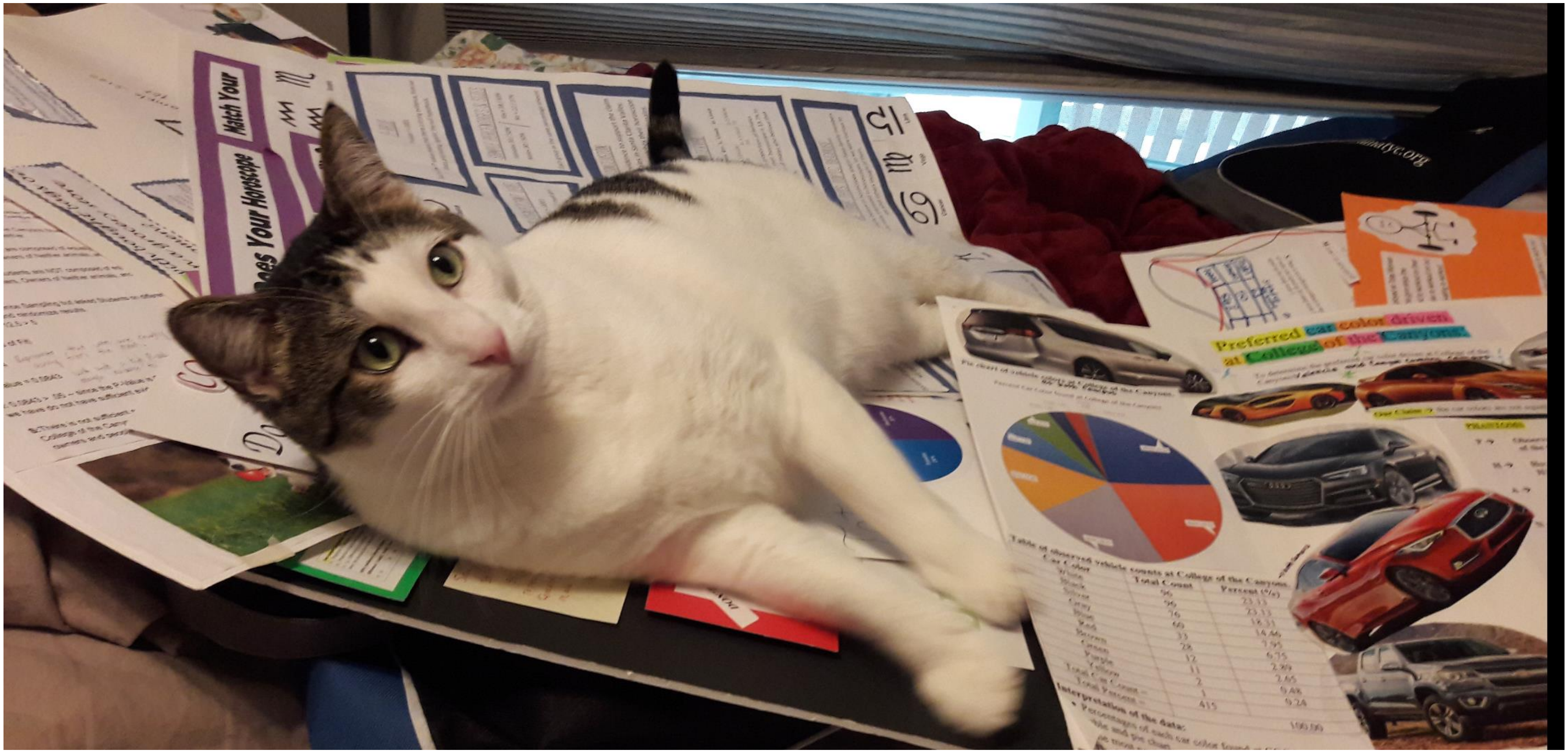
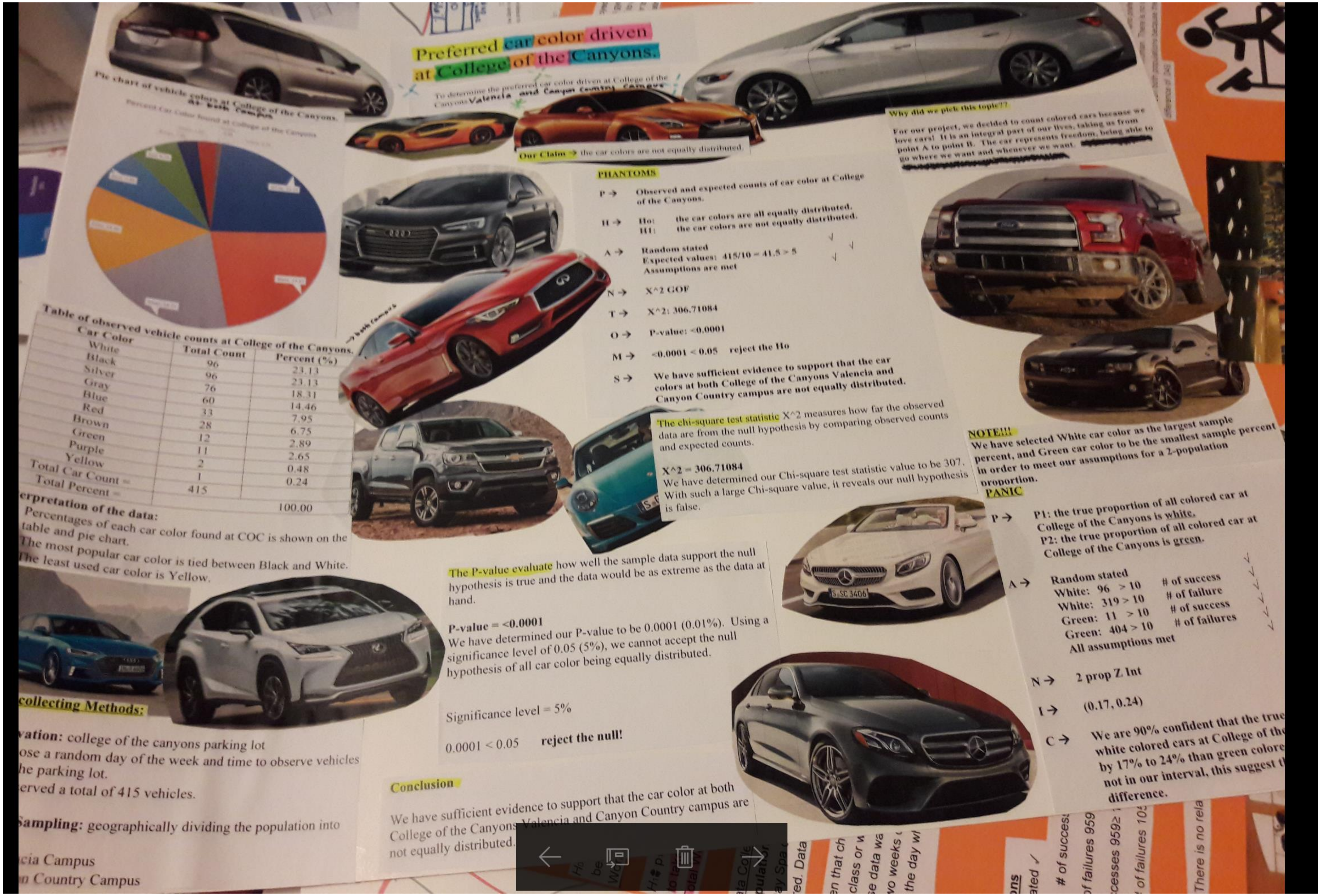
Professor Silva’s Cat loves to help grade:

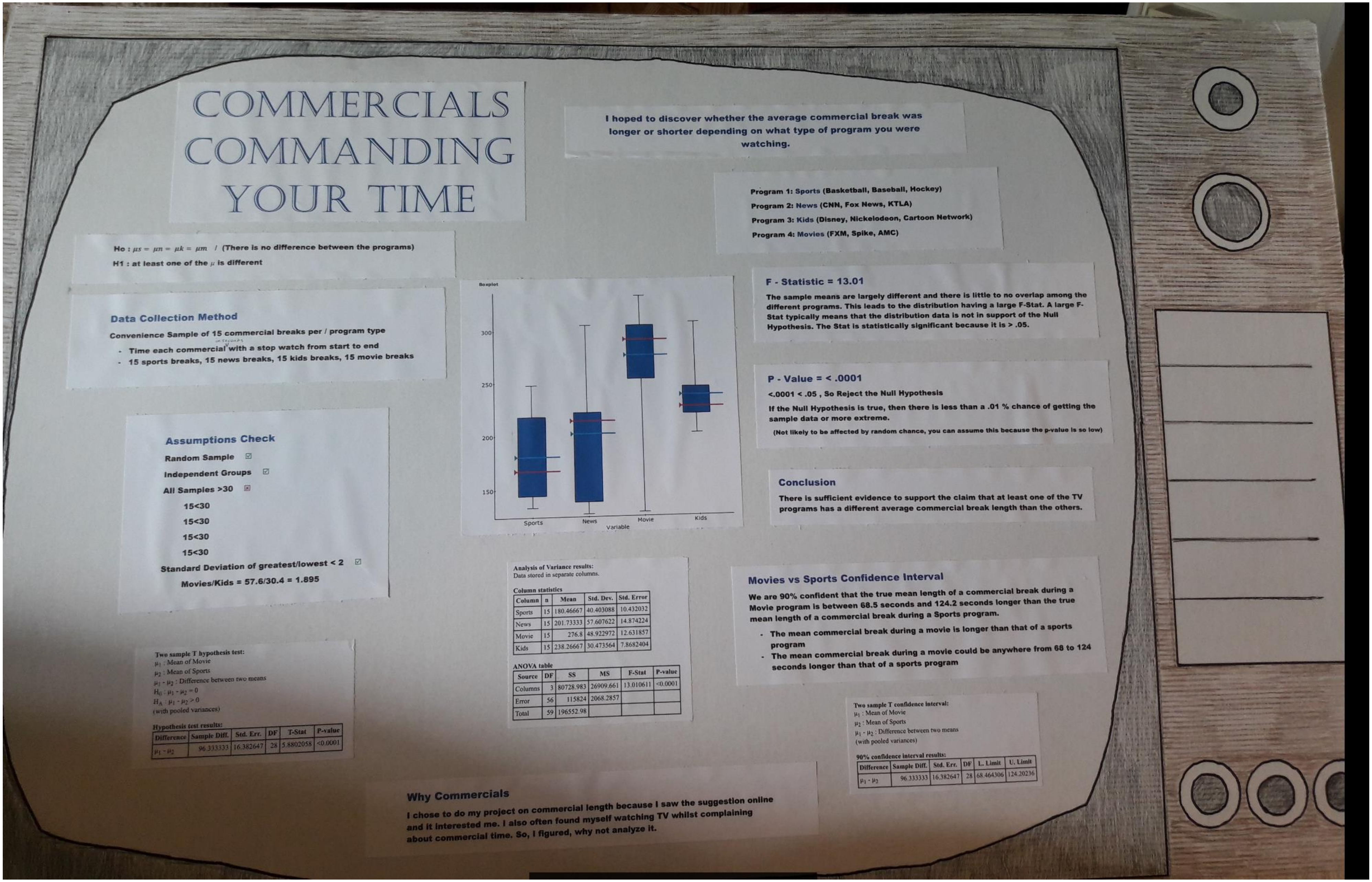


Here are some examples. These are meant to help guide you, but there are mistakes on all of these! And some of our requirements have changed over the years, so be sure to check the rubric for what’s needed on your poster!

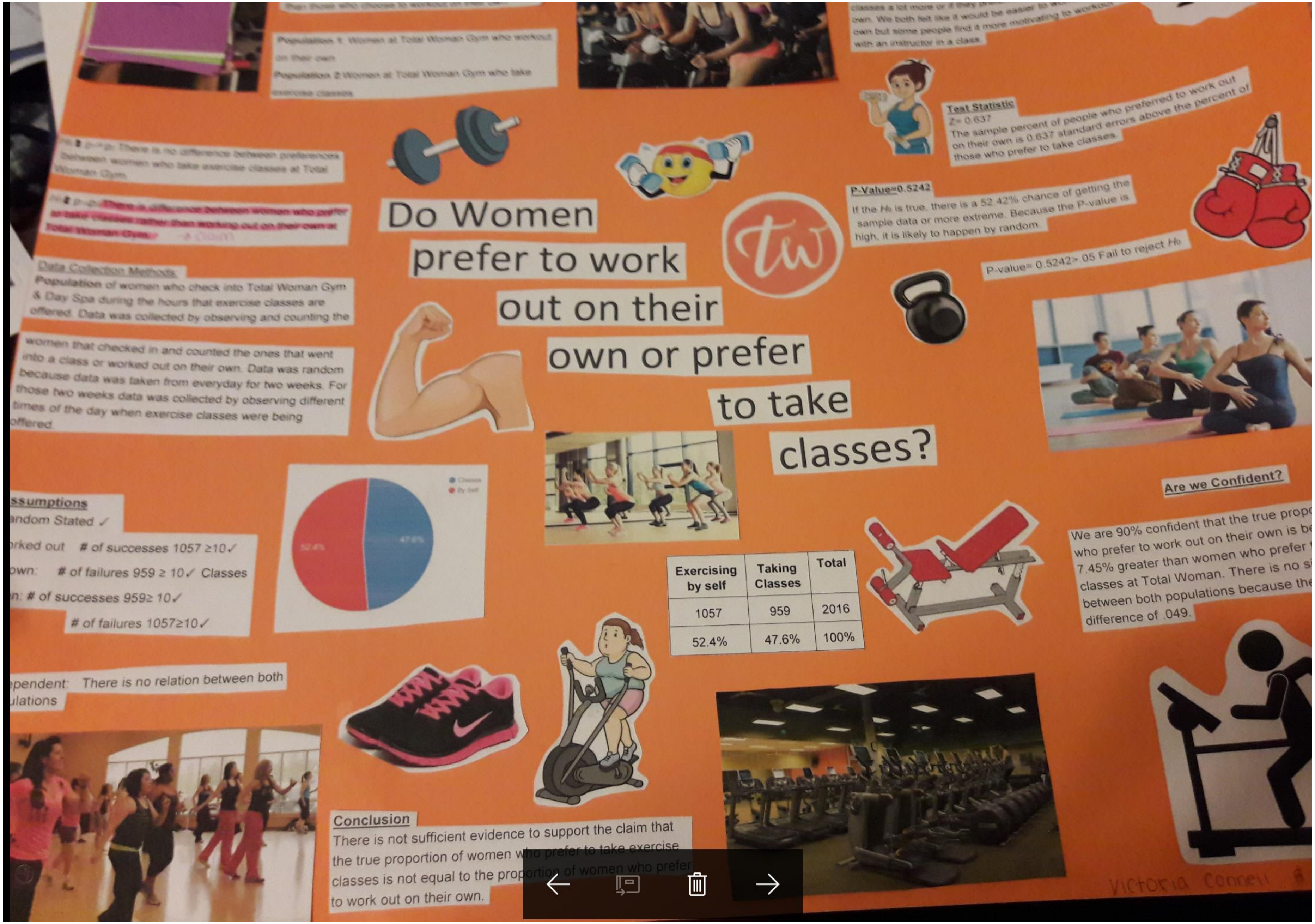
GOF:


Preferred car color driven at College of the Canyons.
To determine the preferred car color driven at College of the Canyons Valencia and Canyon Country Campus
Our Claim → the car colors are not equally distributed.
Why did we pick this topic??
For our project, we decided to count colored cars because we love cars! It is an integral part of our lives, taking us from point A to point B. The car represents freedom, being able to go where we want and whenever we want. [redacted]
Pie chart of vehicle colors at College of the Canyons. at both Campus
Percent Car Color found at college of the Canyons
Table of observed vehicle counts at College of the Canyons. → both Campus
Car Color | Total Count | Percent (%)
White | 96 | 23.13
Black | 96 | 23.13
Silver | 76 | 18.31
Gray | 60 | 14.46
Blue | 33 | 7.95
Red | 28 | 6.75
Brown | 12 | 2.89
Green | 11 | 2.65
Purple | 2 | 0.48
Yellow | 1 | 0.24
Total Car Count = 415
Total Percent = 100.00
erpretation of the data:
Percentages of each car color found at COC is shown on the table and pie chart.
The most popular car color is tied between Black and White.
The least used car color is Yellow.
PHANTOMS
P → Observed and expected counts of car color at College of the Canyons.
H → Ho: the car colors are all equally distributed.
H1: the car colors are not equally distributed.
A → Random stated
Expected values: 415/10 = 41.5 > 5
Assumptions are met
N → X^2 GOF
T → X^2: 306.71084
O → P-value: <0.0001
M → <0.0001 < 0.05 reject the Ho
S → We have sufficient evidence to support that the car colors at both College of the Canyons Valencia and Canyon Country campus are not equally distributed.
The chi-square test statistic X^2 measures how far the observed data are from the null hypothesis by comparing observed counts and expected counts.
X^2 = 306.71084
We have determined our Chi-square test statistic value to be 307. With such a large Chi-square value, it reveals our null hypothesis is false.
The P-value evaluate how well the sample data support the null hypothesis is true and the data would be as extreme as the data at hand.
P-value = <0.0001
We have determined our P-value to be 0.0001 (0.01%). Using a significance level of 0.05 (5%), we cannot accept the null hypothesis of all car color being equally distributed.
Significance level = 5%
0.0001 < 0.05 reject the null!
Conclusion
We have sufficient evidence to support that the car color at both College of the Canyons Valencia and Canyon Country campus are not equally distributed.
NOTE!!!
We have selected White car color as the largest sample percent, and Green car color to be the smallest sample percent in order to meet our assumptions for a 2-population proportion.
PANIC
P → P1: the true proportion of all colored car at College of the Canyons is white.
P2: the true proportion of all colored car at College of the Canyons is green.
A → Random stated
White: 96 > 10 # of success
White: 319 > 10 # of failure
Green: 11 > 10 # of success
Green: 404 > 10 # of failures
All assumptions met
N → 2 prop Z Int
I → (0.17, 0.24)
C → We are 90% confident that the true white colored cars at College of the by 17% to 24% than green colore not in our interval, this suggest t difference.
collecting Methods:
vation: college of the canyons parking lot
ose a random day of the week and time to observe vehicles he parking lot.
erved a total of 415 vehicles.
Sampling: geographically dividing the population into
cia Campus
n Country Campus

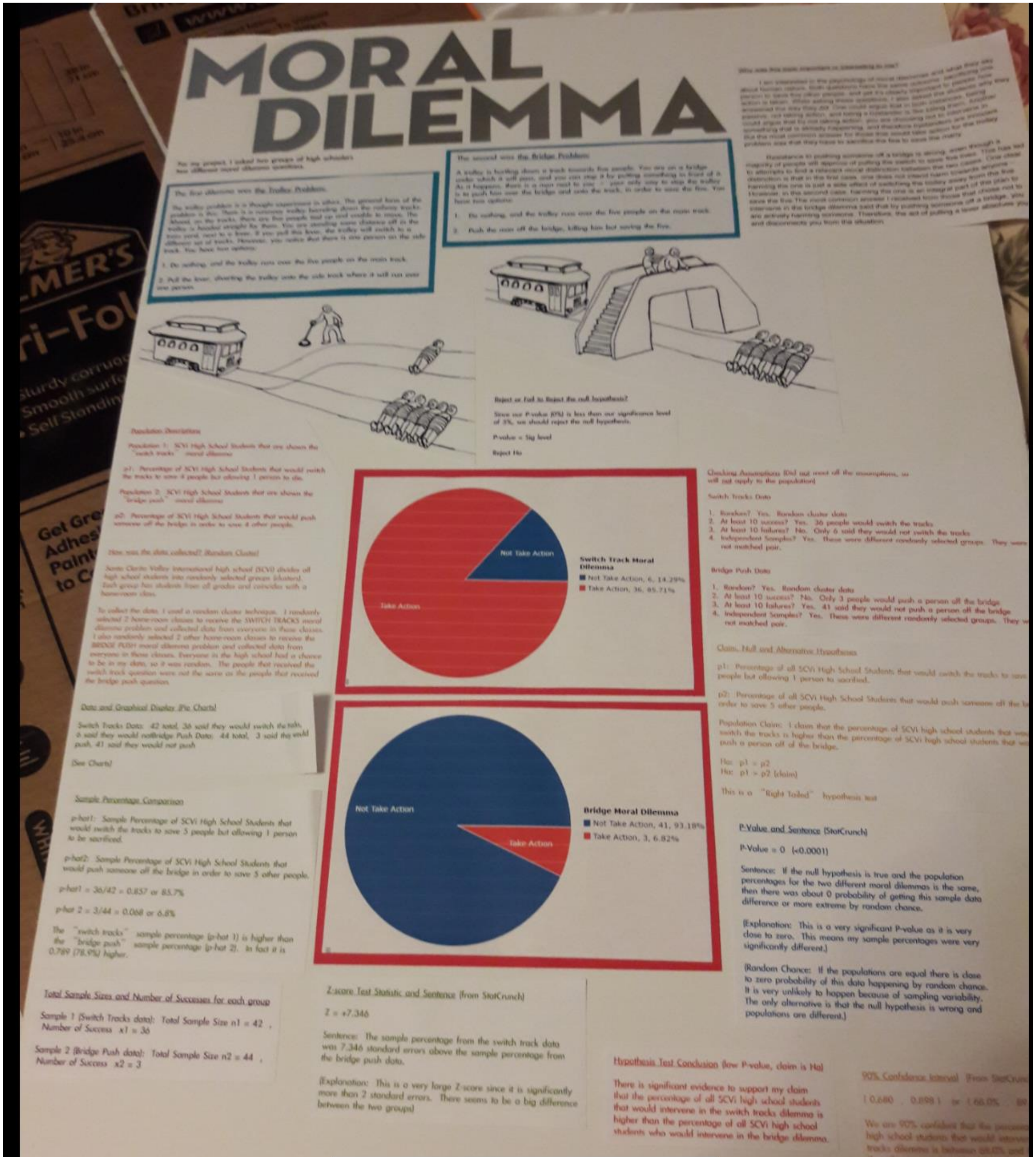
ANOVA:


COMMERCIALS COMMANDING YOUR TIME
I hoped to discover whether the average commercial break was longer or shorter depending on what type of program you were watching.
Program 1: Sports (Basketball, Baseball, Hockey)
Program 2: News (CNN, Fox News, KTLA)
Program 3: Kids (Disney, Nickelodeon, Cartoon Network)
Program 4: Movies (FXM, Spike, AMC)
Ho : μs = μn = μk = μm / (There is no difference between the programs)
H1 : at least one of the μ is different
Data Collection Method
Convenience Sample of 15 commercial breaks per / program type
Time each commercial with a stop watch from start to end
15 sports breaks, 15 news breaks, 15 kids breaks, 15 movie breaks
Assumptions Check
Random Sample
Independent Groups
All Samples >30
15<30
15<30
15<30
15<30
Standard Deviation of greatest/lowest < 2
Movies/Kids = 57.6/30.4 = 1.895
Boxplot
300
250
200
150
Sports
News
Movie
Kids
Variable
F - Statistic = 13.01
The sample means are largely different and there is little to no overlap among the different programs. This leads to the distribution having a large F-Stat. A large F-Stat typically means that the distribution data is not in support of the Null Hypothesis. The Stat is statistically significant because it is > .05.
P - Value = < .0001
<.0001 < .05 , So Reject the Null Hypothesis
If the Null Hypothesis is true, then there is less than a .01 % chance of getting the sample data or more extreme.
(Not likely to be affected by random chance, you can assume this because the p-value is so low)
Conclusion
There is sufficient evidence to support the claim that at least one of the TV programs has a different average commercial break length than the others.
Analysis of Variance results:
Data stored in separate columns.
Column statistics
Column n Mean Std. Dev. Std. Error
Sports 15 180.46667 40.403088 10.432032
News 15 201.73333 57.607622 14.874224
Movie 15 276.8 48.922972 12.631857
Kids 15 238.26667 30.473564 7.8682404
ANOVA table
Source DF SS MS F-Stat P-value
Columns 3 80728.983 26909.661 13.010611 <0.0001
Error 56 115824 2068.2857
Total 59 196552.98
Two sample T hypothesis test:
μ1 : Mean of Movie
μ2 : Mean of Sports
μ1 - μ2 : Difference between two means
H0 : μ1 - μ2 = 0
HA : μ1 - μ2 > 0
(with pooled variances)
Hypothesis test results:
Difference Sample Diff. Std. Err. DF T-Stat P-value
μ1 - μ2 96.333333 16.382647 28 5.8802058 <0.0001
Movies vs Sports Confidence Interval
We are 90% confident that the true mean length of a commercial break during a Movie program is between 68.5 seconds and 124.2 seconds longer than the true mean length of a commercial break during a Sports program.
The mean commercial break during a movie is longer than that of a sports program
The mean commercial break during a movie could be anywhere from 68 to 124 seconds longer than that of a sports program
Two sample T confidence interval:
μ1 : Mean of Movie
μ2 : Mean of Sports
μ1 - μ2 : Difference between two means
(with pooled variances)
90% confidence interval results:
Difference Sample Diff. Std. Err. DF L. Limit U. Limit
μ1 - μ2 96.333333 16.382647 28 68.464306 124.20236
Why Commercials
I chose to do my project on commercial length because I saw the suggestion online and it interested me. I also often found myself watching TV whilst complaining about commercial time. So, I figured, why not analyze it.

2 proportion:



Exercising by self	Taking Classes	Total
1057	959	2016
52.4%	47.6%	100%


MORAL DILEMMA
The first dilemma was the Trolley Problem:
1. Do nothing, and the trolley runs over the five people on the main track.
2. Pull the lever, diverting the trolley onto the side track where it will run over one person.
The second was the Bridge Problem:
1. Do nothing, and the trolley runs over the five people on the main track.
2. Push the man off the bridge, killing him but saving the five.
Reject or Fail to Reject the null hypothesis?
Since our P-value (0%) is less than our significance level of 5%, we should reject the null hypothesis.
P-value < Sig level
Reject Ho
Population Descriptions
Population 1: SCVi High School Students that are shown the "switch tracks" moral dilemma
p1: Percentage of SCVi High School Students that would switch the tracks to save 4 people but allowing 1 person to die.
Population 2: SCVi High School Students that are shown the "bridge push" moral dilemma
p2: Percentage of SCVi High School Students that would push someone off the bridge in order to save 4 other people.
How was the data collected? (Random Cluster)
Santa Clarita Valley International high school (SCVi) divides all high school students into randomly selected groups (clusters). Each group has students from all grades and coincides with a home-room class.
To collect the data, I used a random cluster technique. I randomly selected 2 home-room classes to receive the SWITCH TRACKS moral dilemma problem and collected data from everyone in those classes. I also randomly selected 2 other home-room classes to receive the BRIDGE PUSH moral dilemma problem and collected data from everyone in those classes. Everyone in the high school had a chance to be in my data, so it was random. The people that received the switch track question were not the same as the people that received the bridge push question.
Data and Graphical Display (Pie Charts)
Switch Tracks Data: 42 total, 36 said they would switch the tracks, 6 said they would notBridge Push Data: 44 total, 3 said they would push, 41 said they would not push
(See Charts)
Sample Percentage Comparison
p-hat1: Sample Percentage of SCVi High School Students that would switch the tracks to save 5 people but allowing 1 person to be sacrificed.
p-hat2: Sample Percentage of SCVi High School Students that would push someone off the bridge in order to save 5 other people.
p-hat1 = 36/42 = 0.857 or 85.7%
p-hat 2 = 3/44 = 0.068 or 6.8%
The "switch tracks" sample percentage (p-hat 1) is higher than the "bridge push" sample percentage (p-hat 2). In fact it is 0.789 (78.9%) higher.
Total Sample Sizes and Number of Successes for each group
Sample 1 (Switch Tracks data): Total Sample Size n1 = 42 , Number of Success x1 = 36
Sample 2 (Bridge Push data): Total Sample Size n2 = 44 , Number of Success x2 = 3
Not Take Action
Take Action
Switch Track Moral Dilemma
Not Take Action, 6, 14.29%
Take Action, 36, 85.71%
Not Take Action
Take Action
Bridge Moral Dilemma
Not Take Action, 41, 93.18%
Take Action, 3, 6.82%
Z-score Test Statistic and Sentence (from StatCrunch)
Z = +7.346
Sentence: The sample percentage from the switch track data was 7.346 standard errors above the sample percentage from the bridge push data.
(Explanation: This is a very large Z-score since it is significantly more than 2 standard errors. There seems to be a big difference between the two groups)
Checking Assumptions (Did not meet all the assumptions, so will not apply to the population)
Switch Tracks Data
1. Random? Yes. Random cluster data
2. At least 10 success? Yes. 36 people would switch the tracks
3. At least 10 failures? No. Only 6 said they would not switch the tracks
4. Independent Samples? Yes. These were different randomly selected groups. They were not matched pair.
Bridge Push Data
1. Random? Yes. Random cluster data
2. At least 10 success? No. Only 3 people would push a person off the bridge
3. At least 10 failures? Yes. 41 said they would not push a person off the bridge
4. Independent Samples? Yes. These were different randomly selected groups. They w
not matched pair.
Claim, Null and Alternative Hypotheses
p1: Percentage of all SCVi High School Students that would switch the tracks to save
people but allowing 1 person to sacrified.
p2: Percentage of all SCVi High School Students that would push someone off the b
order to save 5 other people.
Population Claim: I claim that the percentage of SCVi high school students that wo
switch the tracks is higher than the percentage of SCVi high school students that w
push a person off of the bridge.
Ho: p1 = p2
Ha: p1 > p2 (claim)
This is a "Right Tailed" hypothesis test
P-Value and Sentence (StatCrunch)
P-Value = 0 (<0.0001)
Sentence: If the null hypothesis is true and the population percentages for the two different moral dilemmas is the same, then there was about 0 probability of getting this sample data difference or more extreme by random chance.
(Explanation: This is a very significant P-value as it is very close to zero. This means my sample percentages were very significantly different.)
(Random Chance: If the populations are equal there is close to zero probability of this data happening by random chance. It is very unlikely to happen because of sampling variability. The only alternative is that the null hypothesis is wrong and populations are different.)
Hypothesis Test Conclusion (low P-value, claim is Ha)
There is significant evidence to support my claim that the percentage of all SCVi high school students that would intervene in the switch tracks dilemma is higher than the percentage of all SCVi high school students who would intervene in the bridge dilemma.
90% Confidence Interval (From StatCrun