

# Basic Methods of Comparing Data in Minitab Express

The following are some basic ways to compare data in Minitab Express based on whether the data is categorical or numeric. Specific direction on how to perform each comparison can be found on the Simple Data Comparisons videos found online in our class.

## Numeric and Categorical Data (Descriptive Statistics & Boxplot)

**Example Hypothesis:** People with pets exercise more than people without pets.

Variables used to analyze this data:

- Numeric: Number of hours a person spends exercising each week
- Categorical: Have a Pet?

### Directions

Minitab Express: Statistics → Describe → Descriptive Statistics

- Data Tab: Variable = Hours Exercise, Group variables = Pet
- Statistics Tab: Mean, SD, Min, Q1, Median, Q3, Max, N
- Display Tab: Boxplot

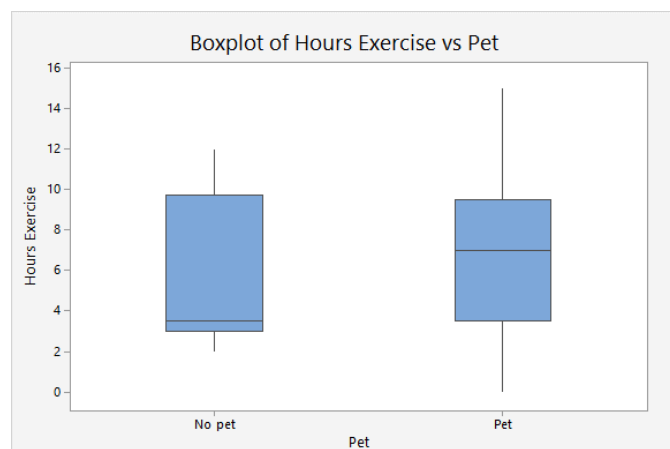
### Results

Example of what descriptive statistics from this type of comparison will look like are shown below.

#### Statistics

Variable	Pet	N	Mean	StDev	Minimum	Q1	Median	Q3	Maximum
Hours Exercise	No pet	9	5.556	3.803	2.000	3.000	3.500	9.750	12.000
	Pet	13	7.000	4.359	0.000	3.500	7.000	9.500	15.000

Example of the Boxplot that this comparison will create.



### Interpreting the Data & Boxplot

Start with the mean for both groups. Then look at the overall distribution of the data on the boxplot. What is the minimum value for each group? What is the median (remember, that means that half the data falls on either side of this point).

### Written Analysis of the Data

On average pet owners exercised 7 hours, while non-pet owners exercised 5.56 hours. The boxplot results show that even though pet owners had the lowest number of hours exercised, overall they exercised a lot more (median = 7) compared to the non-pet owners (median 3.5 hours). These results support the original hypothesis that people with pets exercise more than people without pets.

## Numeric and Categorical Data (Bar Chart)

**Example Hypothesis:** People with pets get less sleep than people without pets.

Variables used to analyze this data:

- Numeric: Hours sleep
- Categorical: Pet

### Directions

Minitab Express: Graphs → Bar Chart → Function of a variable → Simple (single Y variable)

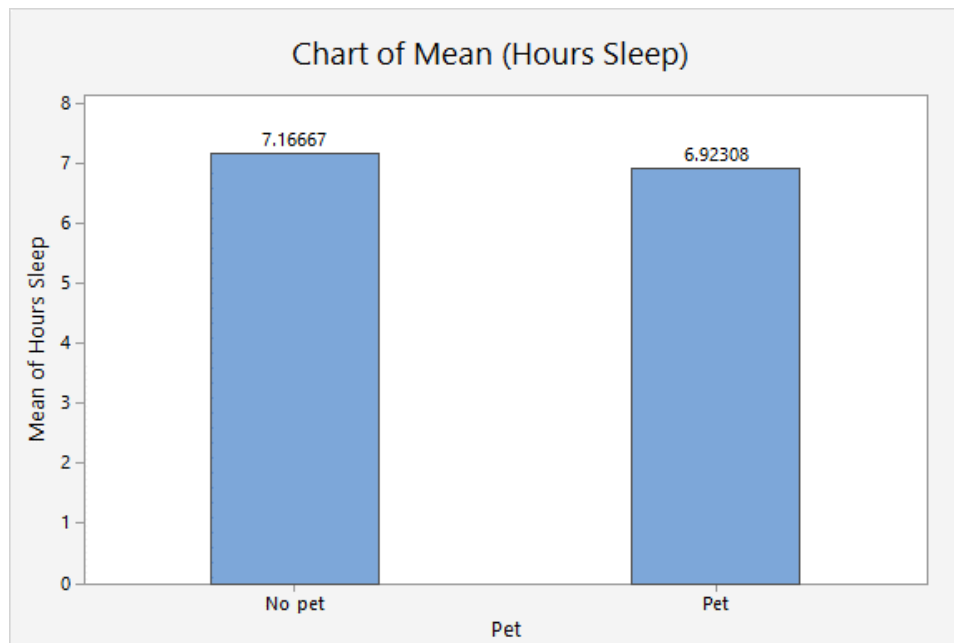
- Function: Mean
- Continuous variable: Hours Sleep
- Categorical Variable: Pet
- 

Add Mean to Chart

- Click on Chart, then click on the + icon with a circle around it (upper right corner)
- Check the “Data Labels” box (this defaults to a linear – straight – regression line)

### Results

Example of Bar Chart graph with means.



### Interpreting this Table

Look at the bars for each category to see if there is a difference between them. The mean value will also help you when interpreting the results.

### Written Analysis of the Data

The bar chart of the data with mean shows very little difference between pet and non-pet owners for the mean number of hours they sleep. Non-pet owners indicated slightly higher sleep with an average of 7.17 hours, compared to pet owners with 6.92 average hours. These results do not support the original hypothesis that people with pets get less sleep than people without pets.

## Categorical and Categorical Data (Cross Tabulation)

**Example Hypothesis:** More women are on social media for 4-6 hours than men.

Variables used to analyze this data:

- Categorical: Gender?
- Categorical: How often do you use social media during a 24 hour period?

### Directions

Minitab Express: Statistics → Cross Tabulation and Chi-Square

- Data Tab: Row = Gender; Column = Social Media  
(Note, for this comparison it doesn't matter which variable gets used for the row and which is used for the column)
- Display Tab: Percent of row total, Percent of column total

### Results

Example of a cross tabulation and Chi-square from this type of comparison is shown below.

Rows: Gender		Columns: Social Media			
		0 hours	1-3 hours	4-6 hours	All
Female	Count	2	3	3	8
	% of Row	25.00	37.50	37.50	100.00
	% of Column	33.33	30.00	50.00	36.36
Male	Count	3	7	2	12
	% of Row	25.00	58.33	16.67	100.00
	% of Column	50.00	70.00	33.33	54.55
Other	Count	1	0	1	2
	% of Row	50.00	0.00	50.00	100.00
	% of Column	16.67	0.00	16.67	9.09
All	Count	6	10	6	22
	% of Row	27.27	45.45	27.27	100.00
	% of Column	100.00	100.00	100.00	100.00
Cell Contents: Count					
% of Row					
% of Column					

#### Interpreting this table

Which results should you look at when trying to determine if your hypothesis was supported or not? In this case, you would look at the % of row for each gender for the 4-6 hours category.

#### Written Analysis of the Data

The cross tabulation results showed that 37.5% of females and 16.67% of males were on social media 4-6 hours. This data supports our hypothesis that more women are on social media for 4-6 hours than men.

#### Why wouldn't you use the % of Column for this question?

For this question the column data (50% female, 33.33% males, & 16.67% other) is only looking at the people who were on social media 4-6 hours and then what gender they were. This would be like separating people by their social media use first and then looking at the percent gender breakdown. A hypothesis that would use this data would look like this – People who are on social media for 4-6 hours are more likely to be female than male. Yet, remember that for our initial hypothesis we started our claim with women (overall) being on social media more than men in the 4-6 hour range. This means we separated people by gender first and then looked at the number of hours they used social media.

## Numeric and Numeric Data (Simple Scatterplot)

**Example Hypothesis:** People who exercise more tend to drink more water.

Variables used to analyze this data:

- Numeric: Hours spent exercising each week
- Numeric: Number of 12oz glasses of water drank each day

### Directions

Minitab Express: Graphs → Scatterplot → Simple

- Y variable = Exercise
- X variable = Water

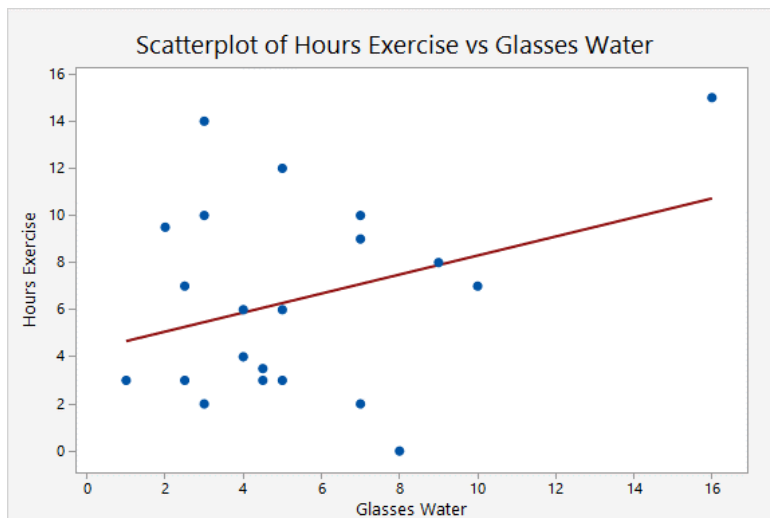
*Note, for this comparison it doesn't matter which variable gets used as the Y variable and which gets used for the X variable.*

Add Regression Fit line to Scatterplot

- Click on Scatterplot, then click on the + icon with a circle around it (upper right corner)
- Check the "Regression Fit" box (this defaults to a linear – straight – regression line)

### Results

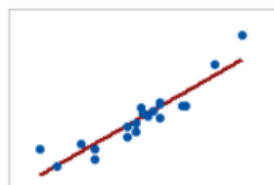
Example of Scatterplot with Regression Fit Line



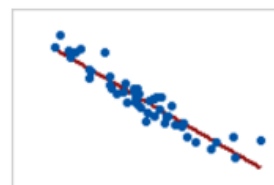
#### Interpreting this Graph

Look at the Regression Fit line to see if there is a positive, negative, or no relationship between the variables. *Positive relationship* means when one variable goes up, the other variable goes up. *Negative relationship* means when one variable goes up, the other variable goes down. *No relationship* means the variables don't seem to be related. The closer the dots (data points) are to the line, the better the fit!

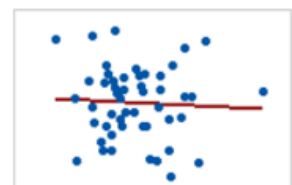
#### Example of each type of Relationship



Linear: positive



Linear: negative



No relationship

### Written Analysis of the Data

The scatterplot of the data with the regression fit line shows a positive relationship between number of hours a person exercises and the number of glasses of water they drink each day. The dots aren't that close to the line, so it's not a strong relationship. Yet, these results do seem to mostly support the original hypothesis that people who exercise more tend to drink more water.

## Numeric compared by Categorical & Categorical

**Example Hypothesis:** Men who have a pet exercise the same amount as women who have a pet.

Variables used to analyze this data:

- Numeric: Hours spent exercising
- Categorical: Gender
- Categorical: Pet Owner

### Directions

Minitab Express: Graphs → Bar Chart → Function of a variable → Clustered

- Function: Mean
- Continuous variable: Hours Exercise
- Categorical variables: Pet, Gender

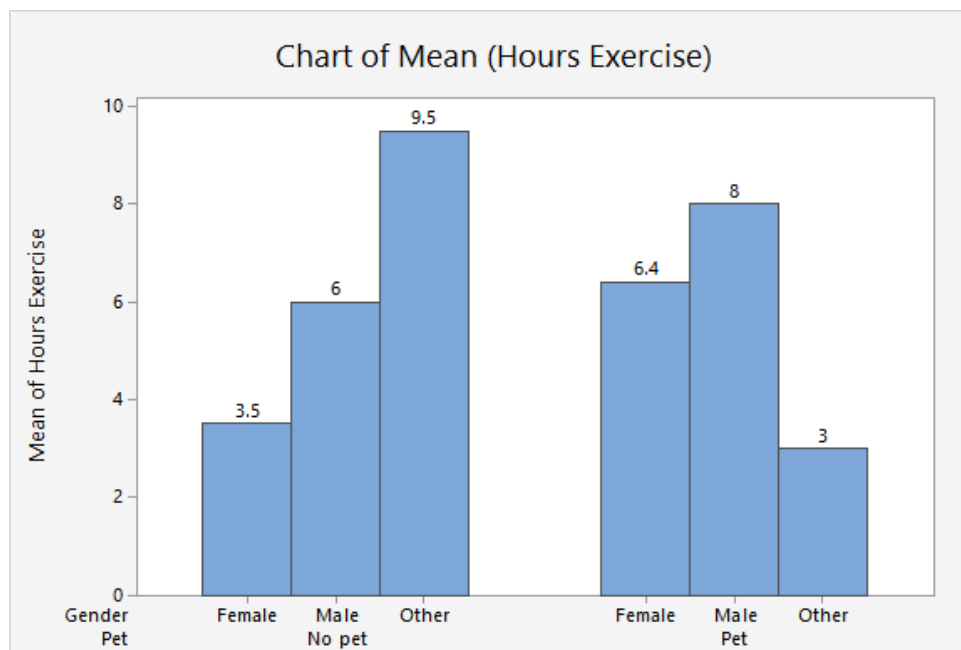
*Note: Since we are only comparing people who have a pet, add Pet first and then gender in the Categorical box.*

Add Mean to Chart

- Click on Chart, then click on the + icon with a circle around it (upper right corner)
- Check the “Data Labels” box.

### Results

Example of Clustered Bar Chart graph with means.



#### Interpreting this Table

Look at the bars for female and men pet owners. Is there a difference? Make sure to look at the mean for each group.

### Written analysis of the data

The clustered bar chart shows a slight difference between the hours spent exercising for men and women pet owners. Men exercised more with an average of 8 hours per week compared to women who exercised on average 6.4 hours per week. These results do not support the original hypothesis that there isn't a difference between the amount exercised between men and women pet owners.

## Numeric & Numeric Data compared by a Categorical Variable (Scatterplot with Groups)

**Example Hypothesis:** There is a difference in the relationship between time spent exercising each week and glasses of water drank each day for pet owners versus non-pet owners.

Variables used to analyze this data:

- Numeric: Hours spent exercising each week
- Numeric: Number of 12oz glasses of water drank each day
- Categorical: Pet Owner

### Directions

Minitab Express: Graphs → Scatterplot → With Groups

- Y variable = Exercise
- X variable = Water
- Group Variable = Pet

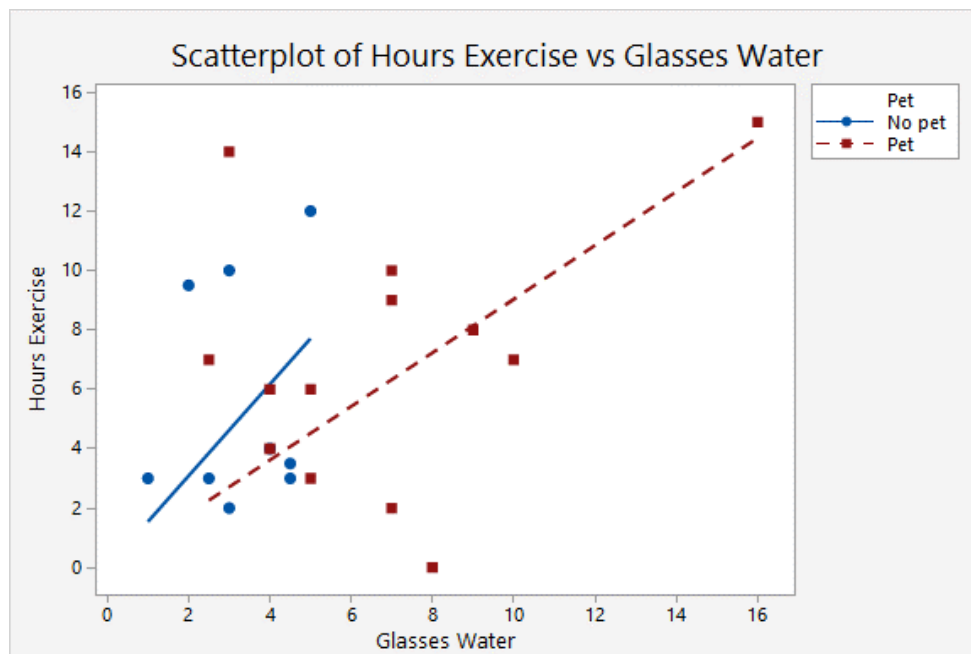
*Note, for this comparison it doesn't matter which variable gets used as the Y variable and which gets used for the X variable.*

Add Regression Fit line to Scatterplot

- Click on Scatterplot, then click on the + icon with a circle around it (upper right corner)
- Check the "Regression Fit" box (this defaults to a linear – straight – regression line)
- Click the arrow next to "Regression Fit" and in the box that comes up, uncheck the "Fit Intercept" box

### Results

Example of Scatterplot with Groups and Regression Line Fit



### Interpreting this Graph

Look at the Regression Fit line for each group to see if there is a positive, negative, or no relationship between the variables (see previous page for more on this). Then compare the lines to see if there is any difference between them.

Example Summary Statistics that are also generated with the Scatterplot with Groups

### Summary Statistics

Variable	Pet	N	Mean	StDev	Minimum	Maximum
Hours Exercise	No pet	9	5.556	3.803	2.000	12.000
	Pet	13	7.000	4.359	0.000	15.000
Glasses Water	No pet	9	3.2778	1.3255	1.0000	5.0000
	Pet	13	6.731	3.609	2.500	16.000

### Written Analysis of the Data

The scatterplot of the data with the regression fit line shows a positive relationship for pet & non-pet owners between the variables (exercise & water). The dots for each group aren't close to their line, so it's not a strong relationship. The scatterplot data indicates a similar trend for both pet and non-pet owners, but their lines are at different angles, indicating different relationships between their data. The Summary statistics back up that there are differences in the data, with non-pet owners having a lower mean for exercise (5.56 hours) and water (3.28 glasses) compared to pet owners (exercise = 7.0 hours; water = 6.73 glasses). These results help show that we can support our original hypothesis that there is a difference in the relationship between time spent exercising each week and glasses of water drank each day for pet owners versus non-pet owners.