Running the Statistical Gauntlet in SPSS

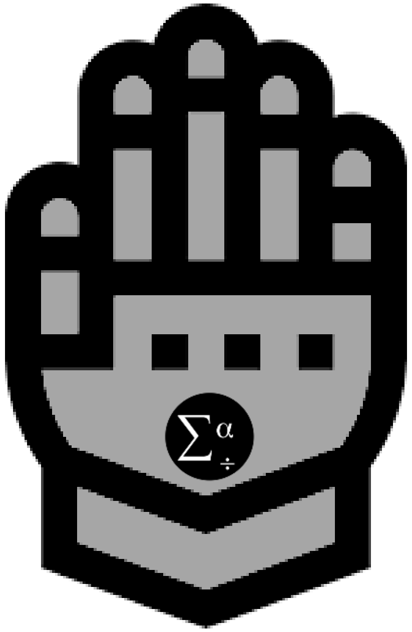
Dr. Mark Williamson

Biostatistics, Epidemiology, and Research Design Core

DaCCoTA, University of North Dakota

Introduction

- Often, in an introduction to statistics, a single example is used to display a model or technique
- This can lead to difficulty in adapting that example's particularities to your own work
- It also fails to train your eye in reading and understanding patterns across examples
- Here, we aim to remedy that by providing, exhaustive, back-to-back examples
- Aimed at intermediate learners
- Get ready for a gauntlet, I hope it will serve you well



DaCCoTA
DAKOTA CANCER COLLABORATIVE ON TRANSLATIONAL ACTIVITY

UNIVERSITY OF NORTH DAKOTA

Assessment

- Before continuing, please take the pre-test
  - **Pre-Test:** https://und.qualtrics.com/jfe/form/SV_2aUt01NaBj8FCqG

- After finishing, please take the post-test and survey
  - **Post-Test:** https://und.qualtrics.com/jfe/form/SV_0VBIEVJemS0YpN4
  - **Survey:** https://und.qualtrics.com/jfe/form/SV_6W42CskCOw9ZhEa

DaCCoTA
DAKOTA CANCER COLLABORATIVE ON TRANSLATIONAL ACTIVITY

UNIVERSITY OF NORTH DAKOTA

Overview

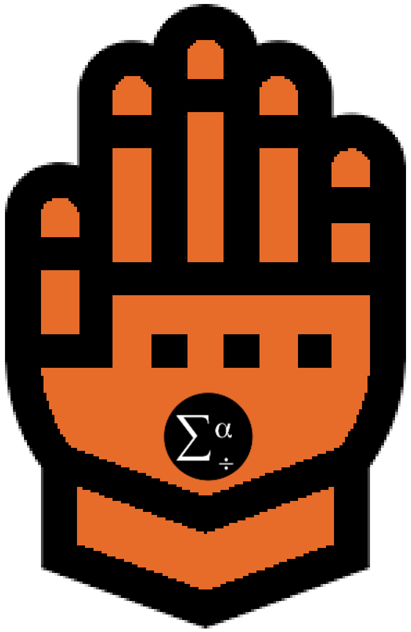
- Today, we'll be using SPSS
- Access SPSS at UND via the Citrix Workspace: https://und.teamdynamix.com/TDClient/2048/IT/KB/ArticleDet?ID=58677
- Sample files: https://www.ibm.com/docs/en/spss-statistics/28.0.0?topic=tutorial-sample-files#data_files
- Topics Covered
  - T-tests
    - One-sample t-test
    - Two-sample t-test
    - Paired t-test
  - ANOVA
    - One-way ANOVA
    - Two-way ANOVA
    - Advanced ANOVA
  - Regression
    - Simple Linear Regression
    - Multiple Linear Regression
    - Logistic Regression

DaCCoTA
DAKOTA CANCER COLLABORATIVE ON TRANSLATIONAL ACTIVITY

UNIVERSITY OF NORTH DAKOTA

Procedure

- Six examples per topic
- Ignoring most assumptions, condensing output for brevity
- The test statistic, p-value, and where appropriate, other variables will be outlined by color
- Sections will start with info on dropdown menus in SPSS
- Each example includes:
  - Research question in the form of a sentence
  - Relevant statistical results from SPSS
  - Written answer to research question
  - Figure or table when appropriate
    - Some graphs will be of null results for clarity (red)
    - Typically, only significant results are graphed
- Get ready to run the gauntlet!



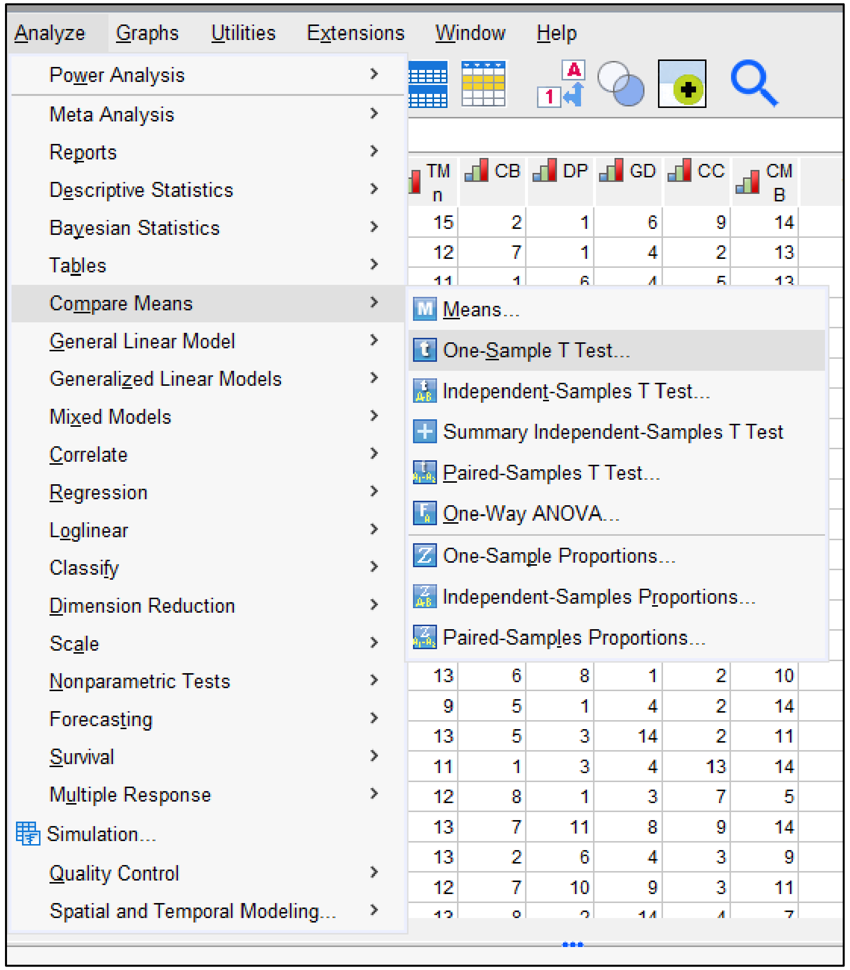
Datasets

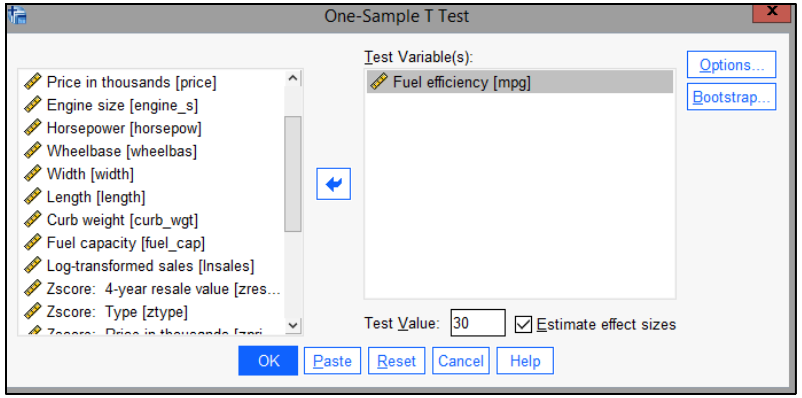
One-sample T-test	**One-way ANOVA**	**Simple Linear Regression**
breakfast_overall.sav	salesperformance.sav	advert.sav
car_sales.sav	aflatoxin.sav	catalog.sav
	carpet.sav	car_sales.sav
Two-sample T-test	**Two-way ANOVA**	**Multiple Linear Regression**
car_sales.sav	carpet.sav	car_sales.sav
breakfast_overall.sav	car_sales.sav	insurance_claims.csv
adl.sav	hourlywagedata.sav	catalog.sav
Paired t-test	**Advanced ANOVA**	**Logistic Regression**
dietstudy.sav	commercial_ratings.sav*	insurance_claims.csv
breakfast.sav	shampoo_ph.sav	hivassays.sav
	poll_cs_sample.sav	poll_cs_sample.sav
	worldsales.sav	

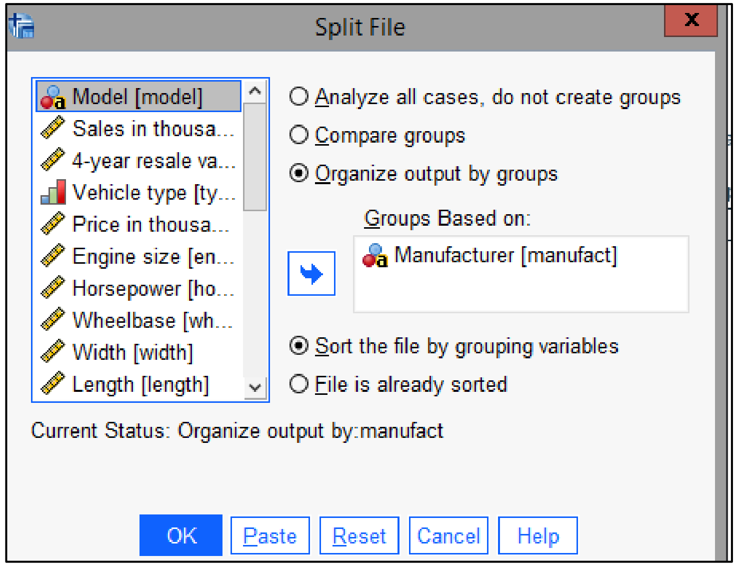
**https://www.spss-tutorials.com/spss-repeated-measures-anova/#run*

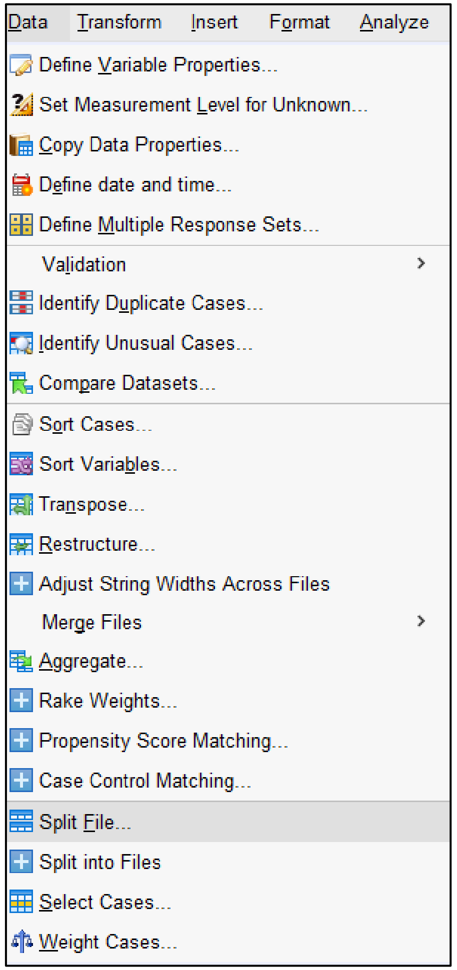
One-sample t-test

Tests if a variable's mean is different from a set value









DaCCoTA
DAKOTA CANCER COLLABORATIVE ON TRANSLATIONAL ACTIVITY

UNIVERSITY OF NORTH DAKOTA

1. One-sample t-test

Tests if a variable's mean is different from a set value

#1) Is coffee cake ranked lower than average (7.5)?

One-Sample Statistics

	N	Mean	Std. Deviation	Std. Error Mean
Coffee cake	42	5.19	3.833	.591

One-Sample Test

Test Value = 7.5

	t	df	Significance One-Sided p	Significance Two-Sided p	Mean Difference	95% Confidence Interval of the Difference Lower	95% Confidence Interval of the Difference Upper
Coffee cake	-3.905	41	<.001	<.001	-2.310	-3.50	-1.11

Yes, coffee cake preference was sign. lower than 7.5.

#2) Is buttered toast ranked higher than average (7.5)?

One-Sample Statistics

	N	Mean	Std. Deviation	Std. Error Mean
Buttered toast	42	7.93	4.480	.691

One-Sample Test

Test Value = 7.5

	t	df	Significance One-Sided p	Significance Two-Sided p	Mean Difference	95% Confidence Interval of the Difference Lower	95% Confidence Interval of the Difference Upper
Buttered toast	.620	41	.269	.539	.429	-.97	1.82

No, buttered toast preference was not higher than 7.5.

#3) Is the rank for jelly donuts different than glazed (7.55)?

One-Sample Statistics

	N	Mean	Std. Deviation	Std. Error Mean
Jelly donut	42	8.48	4.835	.746

One-Sample Test

Test Value = 5

	t	df	Significance One-Sided p	Significance Two-Sided p	Mean Difference	95% Confidence Interval of the Difference Lower	95% Confidence Interval of the Difference Upper
Jelly donut	4.659	41	<.001	<.001	3.476	1.97	4.98

Yes, jelly preference was sign. higher than glazed.

#4) Are average car sales higher than 50K?

One-Sample Statistics

	N	Mean	Std. Deviation	Std. Error Mean
Sales in thousands	157	52.99808	68.029422	5.429339

One-Sample Test

Test Value = 50

	t	df	Significance One-Sided p	Significance Two-Sided p	Mean Difference	95% Confidence Interval of the Difference Lower	95% Confidence Interval of the Difference Upper
Sales in thousands	.552	156	.291	.582	2.998076	-7.72643	13.72258

No, car sales were not higher than 50K.

#5) Are average Dodge car mpg's lower than average (23.84)?

One-Sample Statistics[a]

	N	Mean	Std. Deviation	Std. Error Mean
Fuel efficiency	10	20.10	4.771	1.509

a. Manufacturer = Dodge

One-Sample Test[a]

Test Value = 23.84

	t	df	Significance One-Sided p	Significance Two-Sided p	Mean Difference	95% Confidence Interval of the Difference Lower	95% Confidence Interval of the Difference Upper
Fuel efficiency	-2.479	9	.018	.035	-3.740	-7.15	-.33

a. Manufacturer = Dodge

Yes, Dodge mpg was sign. lower than 23.84.

#6) Are average Ford car mpg's different than Chevrolet (28.44)?

One-Sample Statistics[a]

	N	Mean	Std. Deviation	Std. Error Mean
Fuel efficiency	11	22.82	4.490	1.354

a. Manufacturer = Ford

One-Sample Test[a]

Test Value = 28.44

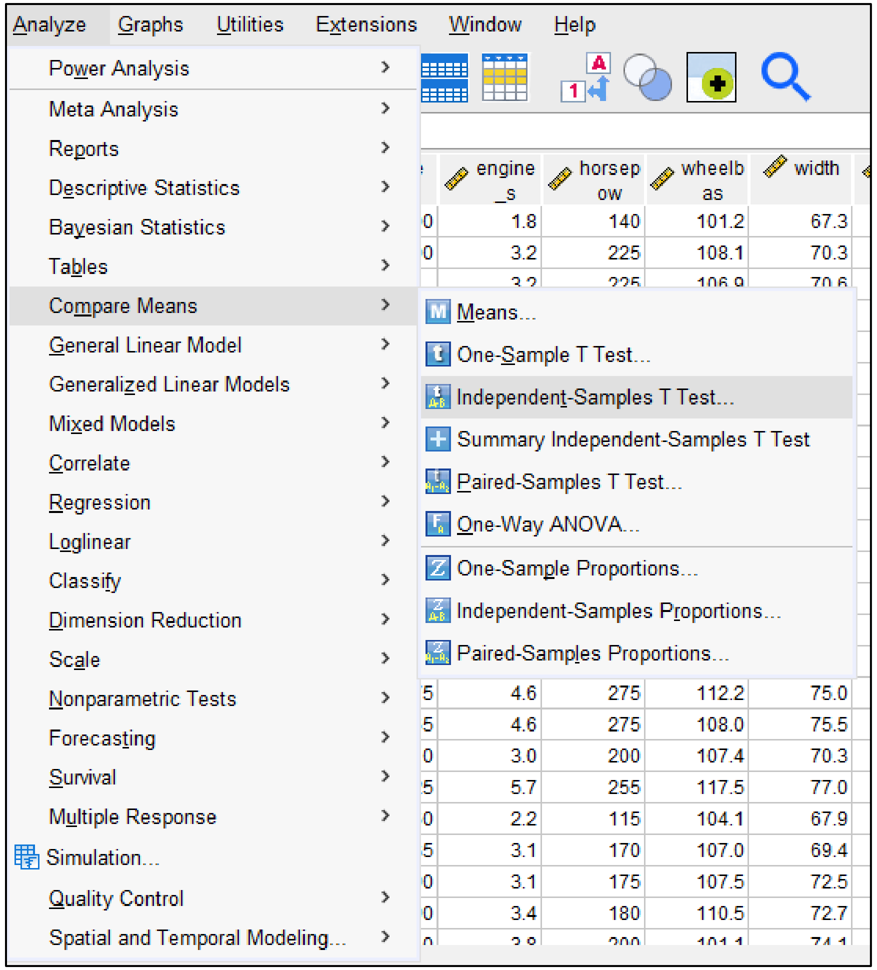
	t	df	Significance One-Sided p	Significance Two-Sided p	Mean Difference	95% Confidence Interval of the Difference Lower	95% Confidence Interval of the Difference Upper
Fuel efficiency	-4.152	10	<.001	.002	-5.622	-8.64	-2.61

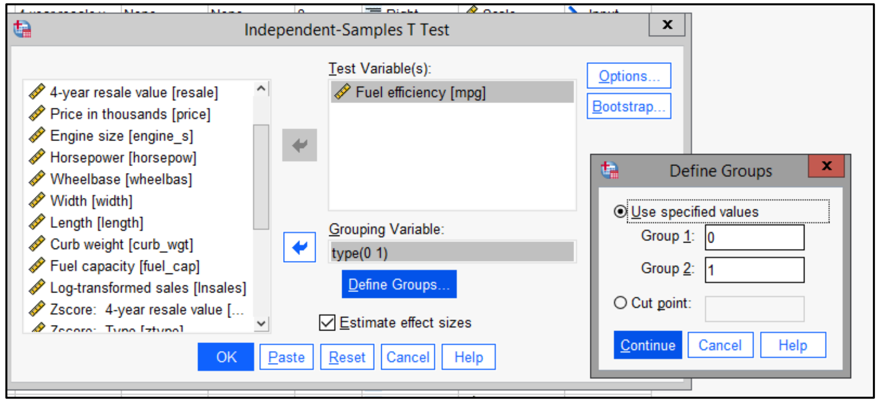
a. Manufacturer = Ford

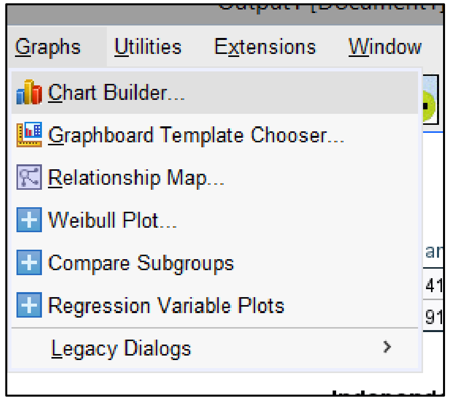
Yes, Ford mpg was sign. lower than 28.44.

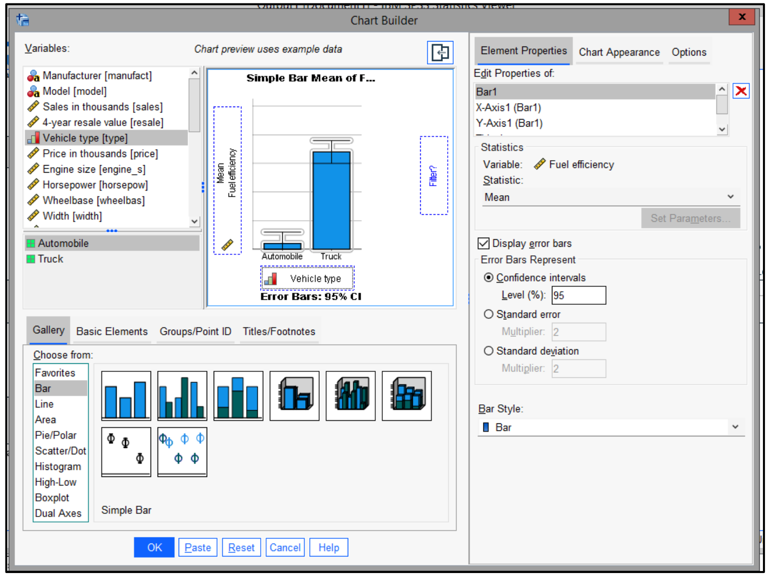
Two-sample t-test

Tests if the mean of two different groups is different









2. Two-sample t-test

Tests if the mean of two different groups is different

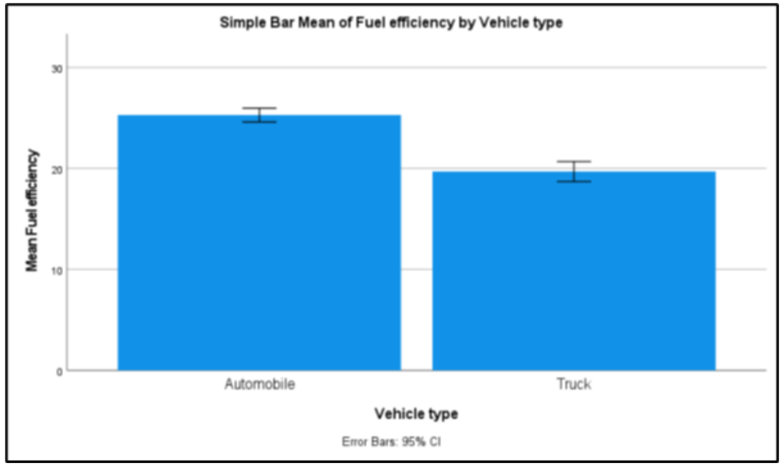
#1) Is mpg greater in automobiles compared to trucks?

Group Statistics

	Vehicle type	N	Mean	Std. Deviation	Std. Error Mean
Fuel efficiency	Automobile	114	25.30	3.646	.341
	Truck	40	19.70	3.107	.491

Independent Samples Test

		Levene's Test for Equality of Variances		t-test for Equality of Means							
						Significance				95% Confidence Interval of the Difference	
		F	Sig.	t	df	One-Sided p	Two-Sided p	Mean Difference	Std. Error Difference	Lower	Upper
Fuel efficiency	Equal variances assumed	.004	.948	8.664	152	<.001	<.001	5.597	.646	4.321	6.874
	Equal variances not assumed			9.356	79.405	<.001	<.001	5.597	.598	4.407	6.788



Yes, mpg was significantly higher in automobiles.

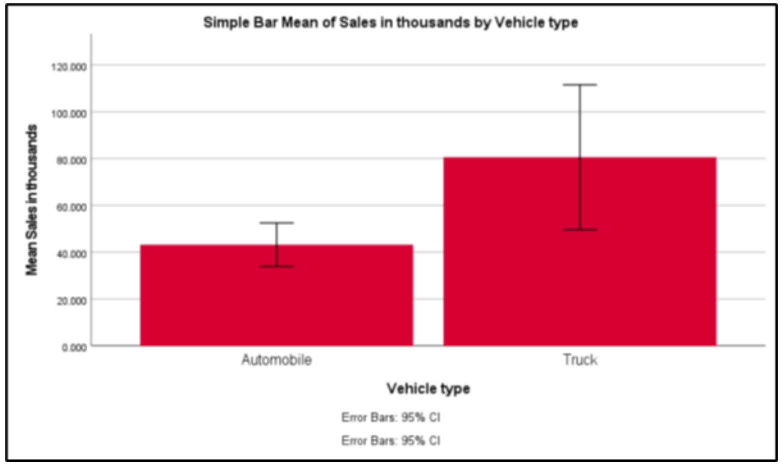
#2) Is price different between automobiles and trucks?

Group Statistics

	Vehicle type	N	Mean	Std. Deviation	Std. Error Mean
Price in thousands	Automobile	115	27.76320	15.566574	1.451591
	Truck	40	26.31998	10.169436	1.607929

Independent Samples Test

		Levene's Test for Equality of Variances		t-test for Equality of Means							
						Significance				95% Confidence Interval of the Difference	
		F	Sig.	t	df	One-Sided p	Two-Sided p	Mean Difference	Std. Error Difference	Lower	Upper
Price in thousands	Equal variances assumed	5.909	.016	.547	153	.293	.585	1.443225	2.640462	-3.773246	6.659696
	Equal variances not assumed			.666	104.686	.253	.507	1.443225	2.166230	-2.852158	5.738608



No, price was not different between automobiles and trucks.

2. Two-sample t-test

Tests if the mean of two different groups is different

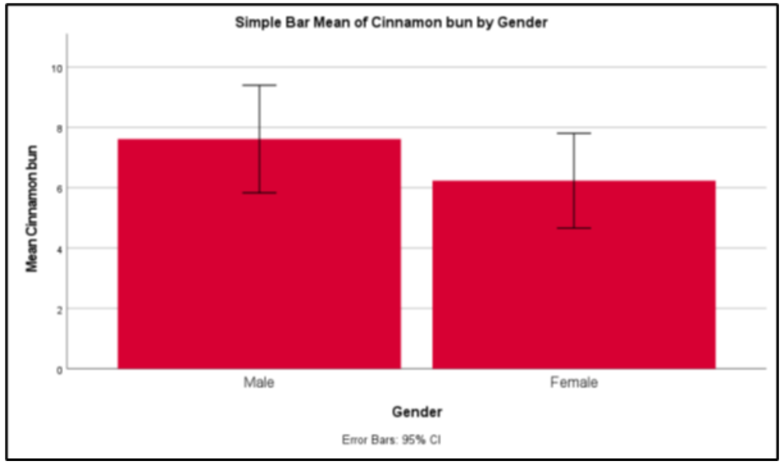
#3) Is cinnamon bun ranking different between males and females?

Group Statistics

	Gender	N	Mean	Std. Deviation	Std. Error Mean
Cinnamon bun	Male	21	7.62	3.918	.855
	Female	21	6.24	3.448	.752

Independent Samples Test

		Levene's Test for Equality of Variances		t-test for Equality of Means							
						Significance				95% Confidence Interval of the Difference	
		F	Sig.	t	df	One-Sided p	Two-Sided p	Mean Difference	Std. Error Difference	Lower	Upper
Cinnamon bun	Equal variances assumed	.865	.358	1.213	40	.116	.232	1.381	1.139	-.921	3.683
	Equal variances not assumed			1.213	39.366	.116	.233	1.381	1.139	-.922	3.684



No, cinnamon bun ranking was not different in males.

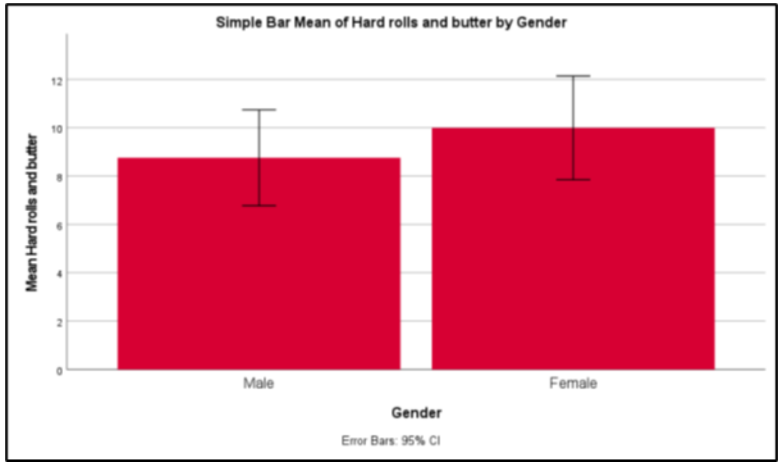
#4) Is hard roll ranking lower in males compared to females?

Group Statistics

	Gender	N	Mean	Std. Deviation	Std. Error Mean
Hard rolls and butter	Male	21	8.76	4.358	.951
	Female	21	10.00	4.712	1.028

Independent Samples Test

		Levene's Test for Equality of Variances		t-test for Equality of Means							
						Significance				95% Confidence Interval of the Difference	
		F	Sig.	t	df	One-Sided p	Two-Sided p	Mean Difference	Std. Error Difference	Lower	Upper
Hard rolls and butter	Equal variances assumed	.335	.566	-.884	40	.191	.382	-1.238	1.401	-4.069	1.592
	Equal variances not assumed			-.884	39.759	.191	.382	-1.238	1.401	-4.069	1.593



No, hard roll ranking was not lower in males.

2. Two-sample t-test

Tests if the mean of two different groups is different

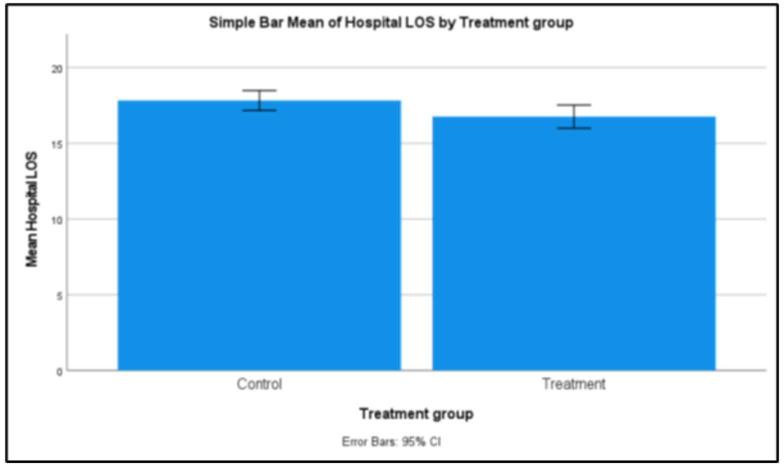
#5) Is length of stay greater for patients in the control compared to the treatment?

Group Statistics

	Treatment group	N	Mean	Std. Deviation	Std. Error Mean
Hospital LOS	Control	46	17.83	2.224	.328
	Treatment	54	16.76	2.801	.381

Independent Samples Test

		Levene's Test for Equality of Variances		t-test for Equality of Means							
						Significance				95% Confidence Interval of the Difference	
		F	Sig.	t	df	One-Sided p	Two-Sided p	Mean Difference	Std. Error Difference	Lower	Upper
Hospital LOS	Equal variances assumed	1.749	.189	2.083	98	.020	.040	1.067	.512	.051	2.083
	Equal variances not assumed			2.122	97.549	.018	.036	1.067	.503	.069	2.065



Yes, length of stay was significantly greater for control patients.

#6) Is age lower for patients in the control compared to the treatment?

Group Statistics

	Treatment group	N	Mean	Std. Deviation	Std. Error Mean
Pt. age	Control	46	72.20	3.215	.474
	Treatment	54	71.39	4.470	.608

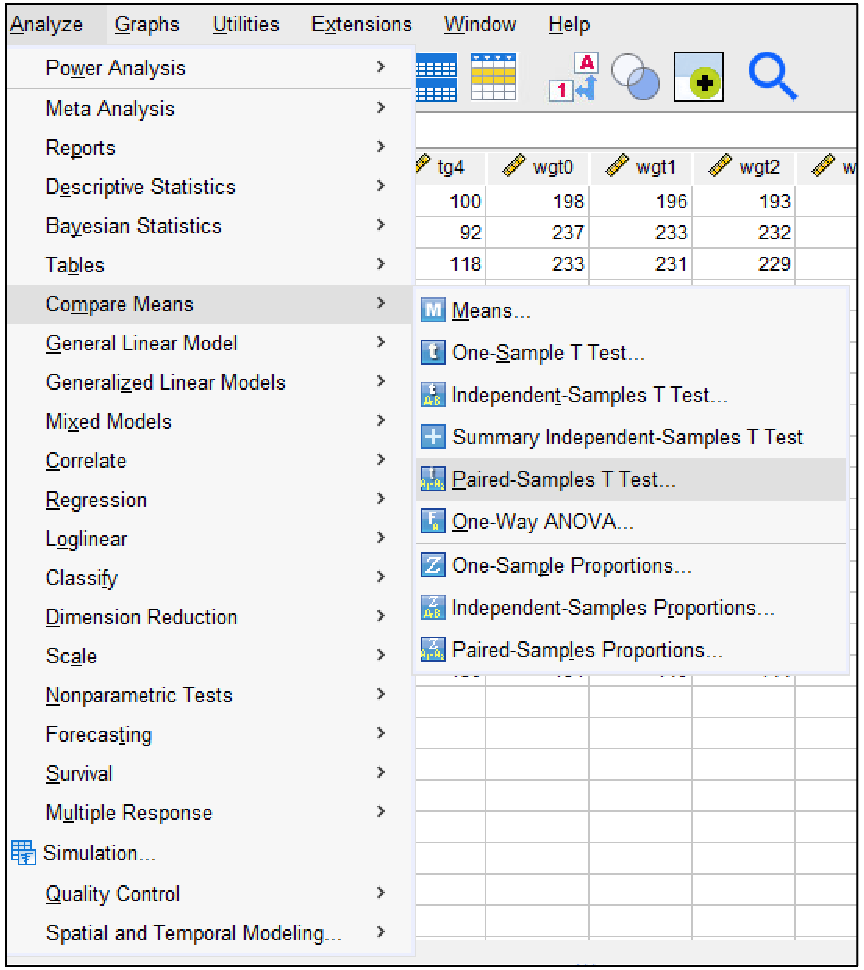
Independent Samples Test

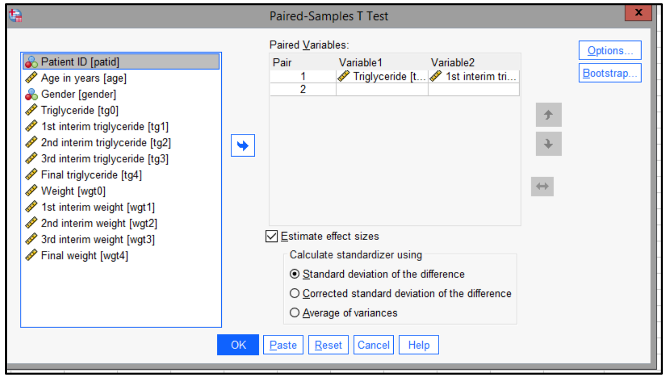
		Levene's Test for Equality of Variances		t-test for Equality of Means							
						Significance				95% Confidence Interval of the Difference	
		F	Sig.	t	df	One-Sided p	Two-Sided p	Mean Difference	Std. Error Difference	Lower	Upper
Pt. age	Equal variances assumed	1.892	.172	1.020	98	.155	.310	.807	.791	-.763	2.377
	Equal variances not assumed			1.046	95.461	.149	.298	.807	.771	-.724	2.338

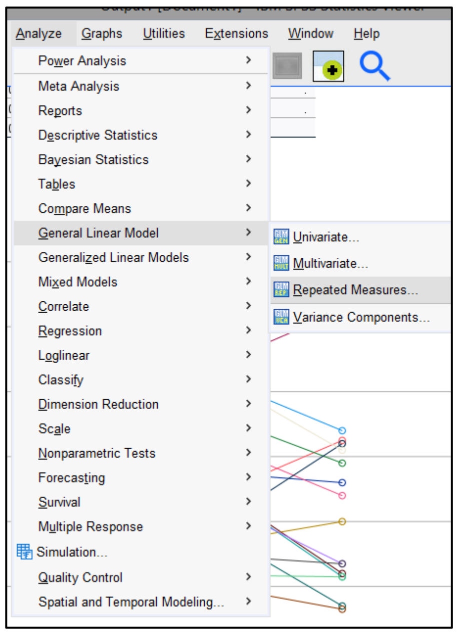
No, age was not lower in control patients.

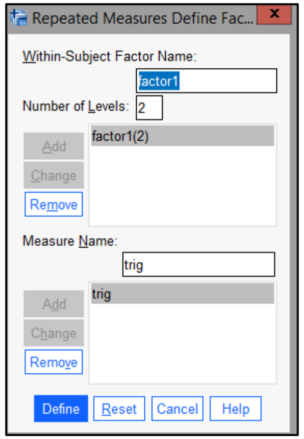
3. Paired t-test

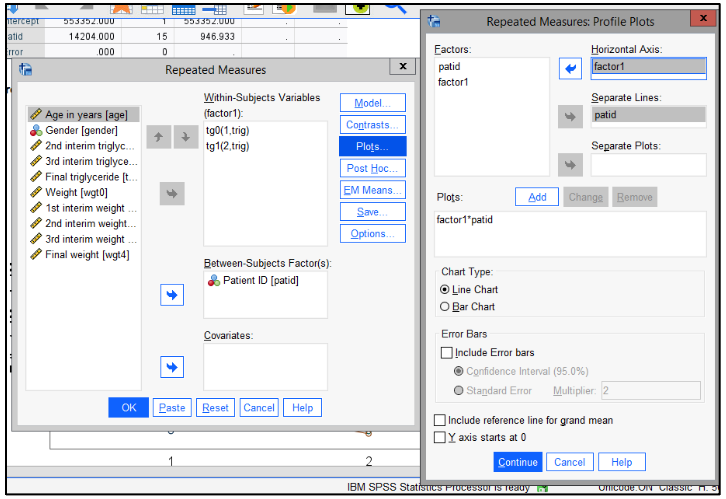
Tests if the means of two different paired groups are different











DaCCoTA
DAKOTA CANCER COLLABORATIVE ON TRANSLATIONAL ACTIVITY

3. Paired t-test

UNIVERSITY OF NORTH DAKOTA

Tests if the means of two different paired groups are different

#1) Are triglyceride measurements in patients different between the initial and 1st interim visit?

Paired Samples Statistics

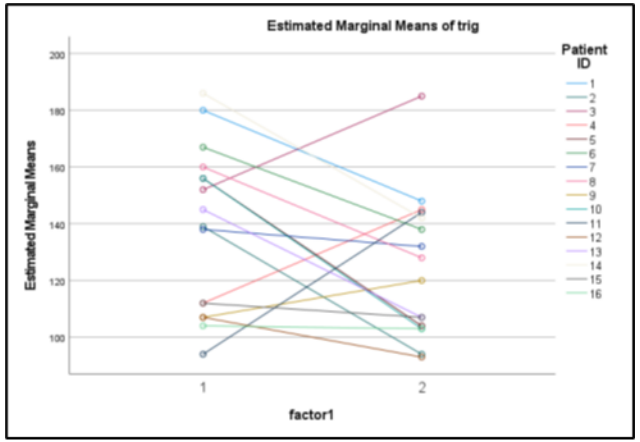
		Mean	N	Std. Deviation	Std. Error Mean
Pair 1	Triglyceride	138.44	16	29.040	7.260
	1st interim triglyceride	124.56	16	25.126	6.282

Paired Samples Correlations

		N	Correlation	Significance One-Sided p	Significance Two-Sided p
Pair 1	Triglyceride & 1st interim triglyceride	16	.287	.140	.281

Paired Samples Test

		Paired Differences Mean	Paired Differences Std. Deviation	Paired Differences Std. Error Mean	95% Confidence Interval of the Difference Lower	95% Confidence Interval of the Difference Upper	t	df	Significance One-Sided p	Significance Two-Sided p
Pair 1	Triglyceride - 1st interim triglyceride	13.875	32.488	8.122	-3.436	31.186	1.708	15	.054	.108



No, triglyceride measurements were not different.

#2) Are triglyceride measurements in patients different between the initial and final visit?

Paired Samples Statistics

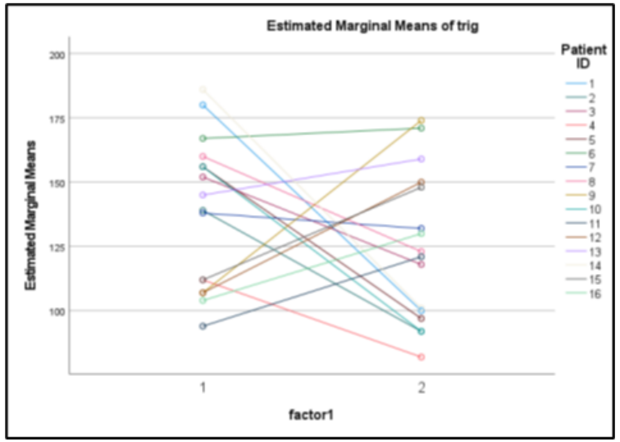
		Mean	N	Std. Deviation	Std. Error Mean
Pair 1	Triglyceride	138.44	16	29.040	7.260
	Final triglyceride	124.38	16	29.412	7.353

Paired Samples Correlations

		N	Correlation	Significance One-Sided p	Significance Two-Sided p
Pair 1	Triglyceride & Final triglyceride	16	-.286	.141	.283

Paired Samples Test

		Paired Differences Mean	Paired Differences Std. Deviation	Paired Differences Std. Error Mean	95% Confidence Interval of the Difference Lower	95% Confidence Interval of the Difference Upper	t	df	Significance One-Sided p	Significance Two-Sided p
Pair 1	Triglyceride - Final triglyceride	14.063	46.875	11.719	-10.915	39.040	1.200	15	.124	.249



No, triglyceride measurements were not different.

DaCCoTA
DAKOTA CANCER COLLABORATIVE ON TRANSLATIONAL ACTIVITY

UND UNIVERSITY OF NORTH DAKOTA

3. Paired t-test

Tests if the means of two different paired groups are different

#3) Are weight measurements lower in patients between the initial and final visit?

Paired Samples Statistics

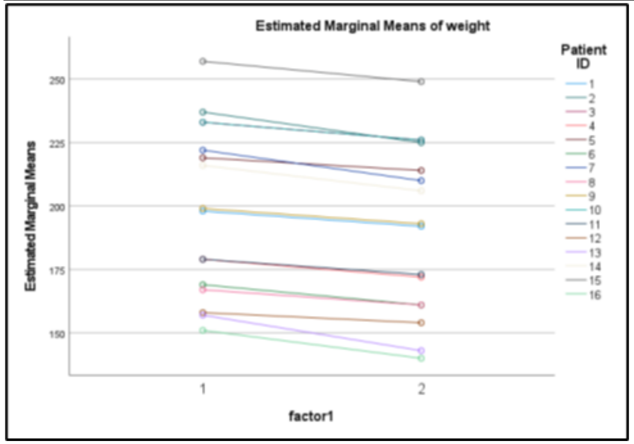
		Mean	N	Std. Deviation	Std. Error Mean
Pair 1	Weight	198.38	16	33.472	8.368
	Final weight	190.31	16	33.508	8.377

Paired Samples Correlations

		N	Correlation	Significance One-Sided p	Significance Two-Sided p
Pair 1	Weight & Final weight	16	.996	<.001	<.001

Paired Samples Test

		Paired Differences Mean	Paired Differences Std. Deviation	Paired Differences Std. Error Mean	95% Confidence Interval of the Difference Lower	95% Confidence Interval of the Difference Upper	t	df	Significance One-Sided p	Significance Two-Sided p
Pair 1	Weight - Final weight	8.063	2.886	.722	6.525	9.600	11.175	15	<.001	<.001



Yes, weight was lower in the final visit.

#4) Is glazed donut preference different between overall and snack preference?

Paired Samples Statistics

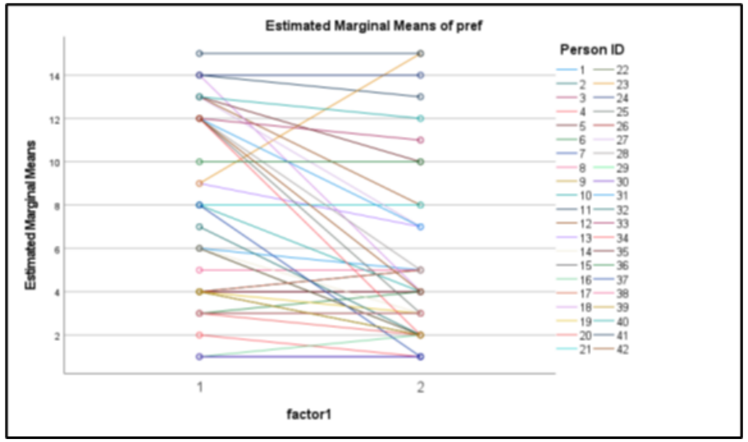
		Mean	N	Std. Deviation	Std. Error Mean
Pair 1	Glazed donuts overall	7.55	42	4.407	.680
	Glazed donuts snacks	5.33	42	4.046	.624

Paired Samples Correlations

		N	Correlation	Significance One-Sided p	Significance Two-Sided p
Pair 1	Glazed donuts overall & Glazed donuts snacks	42	.680	<.001	<.001

Paired Samples Test

		Paired Differences Mean	Paired Differences Std. Deviation	Paired Differences Std. Error Mean	95% Confidence Interval of the Difference Lower	95% Confidence Interval of the Difference Upper	t	df	Significance One-Sided p	Significance Two-Sided p
Pair 1	Glazed donuts overall - Glazed donuts snacks	2.214	3.397	.524	1.156	3.273	4.225	41	<.001	<.001



Yes, glazed donut preference was higher overall compared to snacks.

DaCCoTA
DAKOTA CANCER COLLABORATIVE ON TRANSLATIONAL ACTIVITY

UNIVERSITY OF NORTH DAKOTA

3. Paired t-test

Tests if the means of two different paired groups are different

#5) Is blueberry muffin preference higher in breakfasts compared to snacks?

Paired Samples Statistics

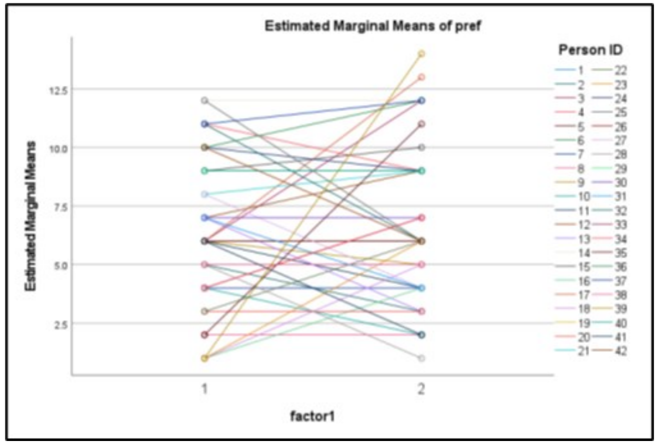
		Mean	N	Std. Deviation	Std. Error Mean
Pair 1	Blueberry breakfast	6.24	42	3.222	.497
	Blueberry snacks	6.69	42	3.432	.530

Paired Samples Correlations

		N	Correlation	Significance One-Sided p	Significance Two-Sided p
Pair 1	Blueberry breakfast & Blueberry snacks	42	.360	.010	.019

Paired Samples Test

		Paired Differences Mean	Paired Differences Std. Deviation	Paired Differences Std. Error Mean	95% Confidence Interval of the Difference Lower	95% Confidence Interval of the Difference Upper	t	df	Significance One-Sided p	Significance Two-Sided p
Pair 1	Blueberry breakfast - Blueberry snacks	-.452	3.769	.582	-1.627	.722	-.778	41	.221	.441



No, blueberry muffin preference was not different between breakfast and snacks.

#6) Is English muffin preference higher in cereal only breakfasts compared to bacon and eggs breakfast?

Paired Samples Statistics

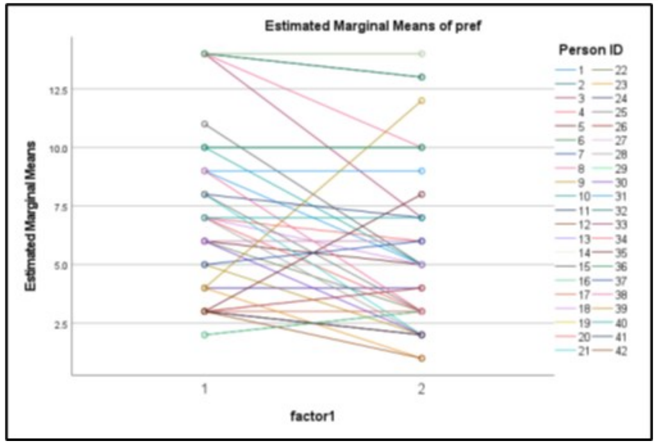
		Mean	N	Std. Deviation	Std. Error Mean
Pair 1	English bacon/eggs	5.62	42	3.715	.573
	English cereal	7.02	42	3.758	.580

Paired Samples Correlations

		N	Correlation	Significance One-Sided p	Significance Two-Sided p
Pair 1	English bacon/eggs & English cereal	42	.694	<.001	<.001

Paired Samples Test

		Paired Differences Mean	Paired Differences Std. Deviation	Paired Differences Std. Error Mean	95% Confidence Interval of the Difference Lower	95% Confidence Interval of the Difference Upper	t	df	Significance One-Sided p	Significance Two-Sided p
Pair 1	English bacon/eggs - English cereal	-1.405	2.922	.451	-2.315	-.494	-3.115	41	.002	.003



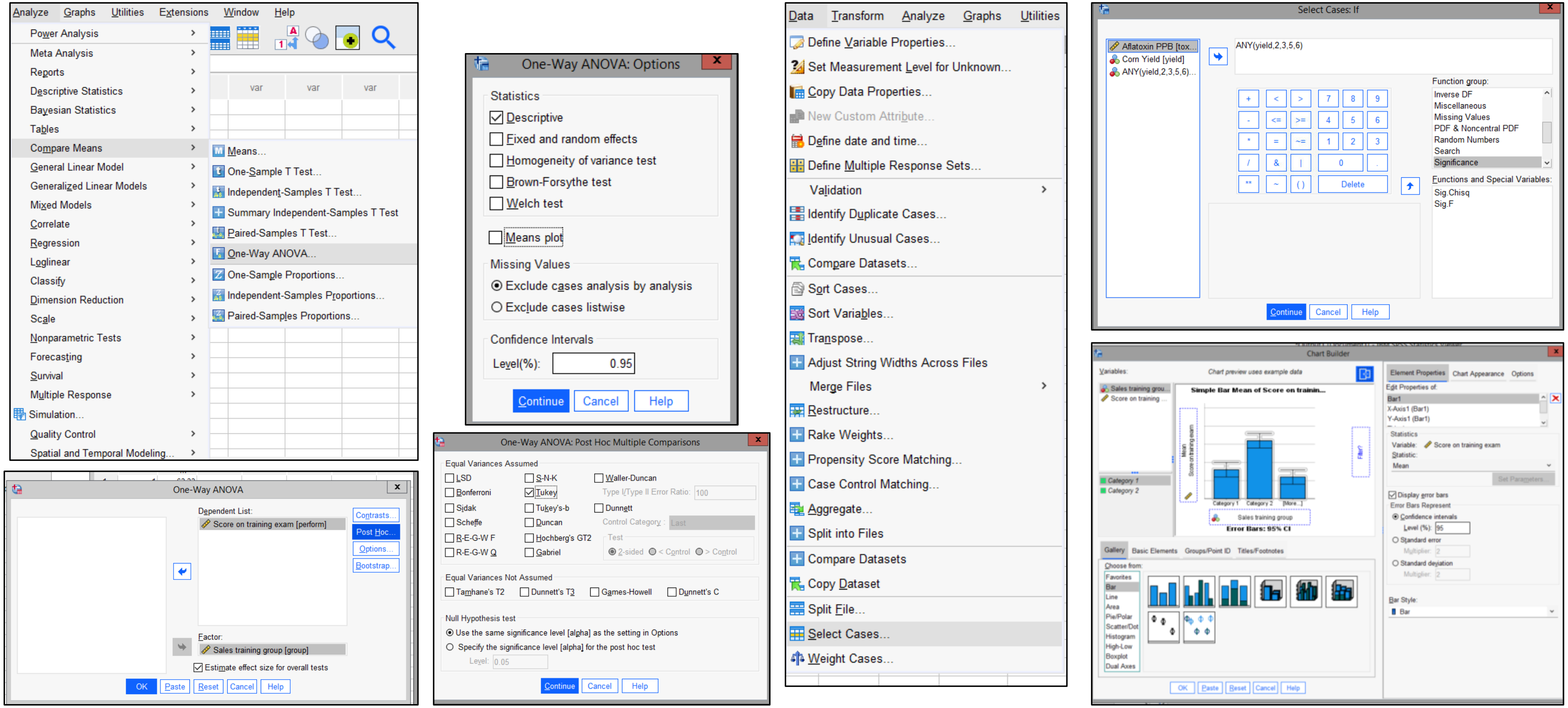
Yes, English muffin preference was significantly higher in cereal breakfasts.

DaCCoTA
DAKOTA CANCER COLLABORATIVE ON TRANSLATIONAL ACTIVITY

UND UNIVERSITY OF NORTH DAKOTA

4. One-way ANOVA

Tests if a variable's mean is different between a category with three or more groups



4. One-way ANOVA

DaCCoTA
DAKOTA CANCER COLLABORATIVE ON TRANSLATIONAL ACTIVITY

UNIVERSITY OF NORTH DAKOTA

Tests if a variable's mean is different between a category with three or more groups

#1) Are sales training course performance scores different across group?

ANOVA

Score on training exam

	Sum of Squares	df	Mean Square	F	Sig.
Between Groups	2525.691	2	1262.846	12.048	<.001
Within Groups	5974.724	57	104.820		
Total	8500.415	59			

	N	Mean	Std. Deviation
1	20	63.5798	13.50858
2	20	73.5677	10.60901
3	20	79.2792	4.40754
Total	60	72.1422	12.00312

Multiple Comparisons

Dependent Variable: Score on training exam
Tukey HSD

(I) Sales training group	(J) Sales training group	Mean Difference (I-J)	Std. Error	Sig.	95% Confidence Interval Lower Bound	95% Confidence Interval Upper Bound
1	2	-9.98789*	3.23759	.009	-17.7789	-2.1969
	3	-15.69947*	3.23759	<.001	-23.4905	-7.9085
2	1	9.98789*	3.23759	.009	2.1969	17.7789
	3	-5.71158	3.23759	.191	-13.5026	2.0794
3	1	15.69947*	3.23759	<.001	7.9085	23.4905
	2	5.71158	3.23759	.191	-2.0794	13.5026

*. The mean difference is significant at the 0.05 level.

Yes, training group 1 was significantly lower than 2 and 3, but 2 and 3 were not different.

#2) Are aflatoxin levels different across yields 1-4?

ANOVA

Aflatoxin PPB

	Sum of Squares	df	Mean Square	F	Sig.
Between Groups	4046.297	3	1348.766	19.264	<.001
Within Groups	4200.813	60	70.014		
Total	8247.109	63			

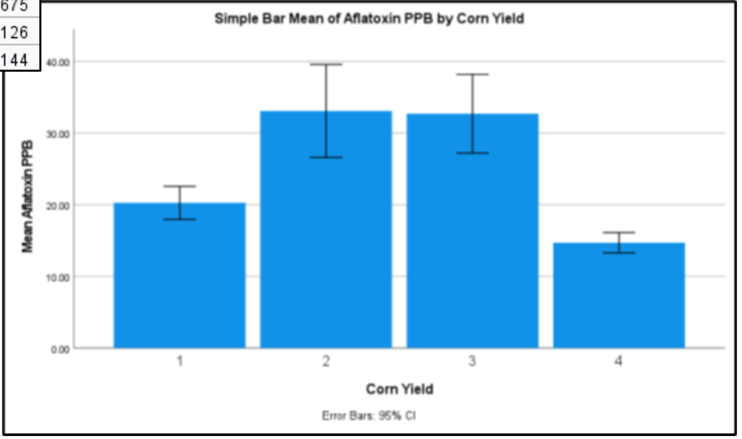
	N	Mean	Std. Deviation
1	16	20.2500	4.31277
2	16	33.0625	12.17357
3	16	32.6875	10.30675
4	16	14.6875	2.65126
Total	64	25.1719	11.44144

Multiple Comparisons

Dependent Variable: Aflatoxin PPB
Tukey HSD

(I) Corn Yield	(J) Corn Yield	Mean Difference (I-J)	Std. Error	Sig.	95% Confidence Interval Lower Bound	95% Confidence Interval Upper Bound
1	2	-12.81250*	2.95833	<.001	-20.6299	-4.9951
	3	-12.43750*	2.95833	<.001	-20.2549	-4.6201
	4	5.56250	2.95833	.247	-2.2549	13.3799
2	1	12.81250*	2.95833	<.001	4.9951	20.6299
	3	.37500	2.95833	.999	-7.4424	8.1924
	4	18.37500*	2.95833	<.001	10.5576	26.1924
3	1	12.43750*	2.95833	<.001	4.6201	20.2549
	2	-.37500	2.95833	.999	-8.1924	7.4424
	4	18.00000*	2.95833	<.001	10.1826	25.8174
4	1	-5.56250	2.95833	.247	-13.3799	2.2549
	2	-18.37500*	2.95833	<.001	-26.1924	-10.5576
	3	-18.00000*	2.95833	<.001	-25.8174	-10.1826

*. The mean difference is significant at the 0.05 level.



Yes, yields 2 and 3 were significantly higher than 1 and 4, but not from each other.

DaCCoTA
DAKOTA CANCER COLLABORATIVE ON TRANSLATIONAL ACTIVITY

UND UNIVERSITY OF NORTH DAKOTA

4. One-way ANOVA

Tests if a variable's mean is different between a category with three or more groups

#3) Are aflatoxin levels different across yields 2, 3, 5, and 6?

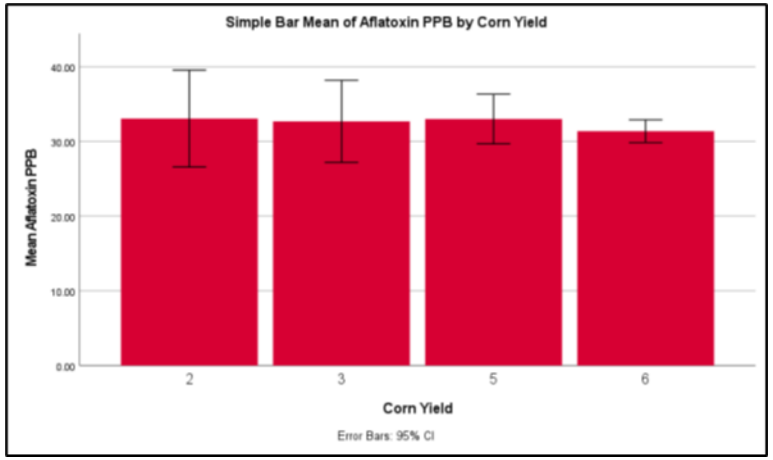
ANOVA

Aflatoxin PPB

	Sum of Squares	df	Mean Square	F	Sig.
Between Groups	29.813	3	9.938	.132	.941
Within Groups	4520.125	60	75.335		
Total	4549.938	63			

No, there was no difference in yields across groups.

	N	Mean	Std. Deviation
2	16	33.0625	12.17357
3	16	32.6875	10.30675
5	16	33.0000	6.22896
6	16	31.3750	2.84898
Total	64	32.5313	8.49831



#4) Are aflatoxin levels different across yields 1-8?

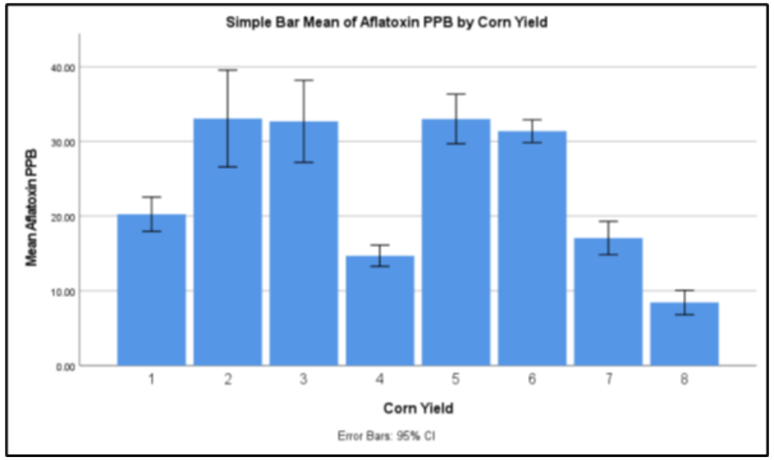
ANOVA

Aflatoxin PPB

	Sum of Squares	df	Mean Square	F	Sig.
Between Groups	10941.430	7	1563.061	35.327	<.001
Within Groups	5309.438	120	44.245		
Total	16250.867	127			

Yes, yields 2, 3, 5, and 6 were in highest clustering, while 1, 7 in middle, and 8 in lowest. Yield 4 was intermediate.

	N	Mean	Std. Deviation
1	16	20.2500	4.31277
2	16	33.0625	12.17357
3	16	32.6875	10.30675
4	16	14.6875	2.65126
5	16	33.0000	6.22896
6	16	31.3750	2.84898
7	16	17.0625	4.18678
8	16	8.4375	3.07612
Total	128	23.8203	11.31192



4. One-way ANOVA

Tests if a variable's mean is different between a category with three or more groups

#5) Are carpet cleaner preferences different across package type?

ANOVA

Preference

	Sum of Squares	df	Mean Square	F	Sig.
Between Groups	583.804	2	291.902	19.281	<.001
Within Groups	287.651	19	15.140		
Total	871.455	21			

	N	Mean	Std. Deviation
A*	9	17.56	3.358
B*	6	6.00	4.000
C*	7	8.29	4.424
Total	22	11.45	6.442

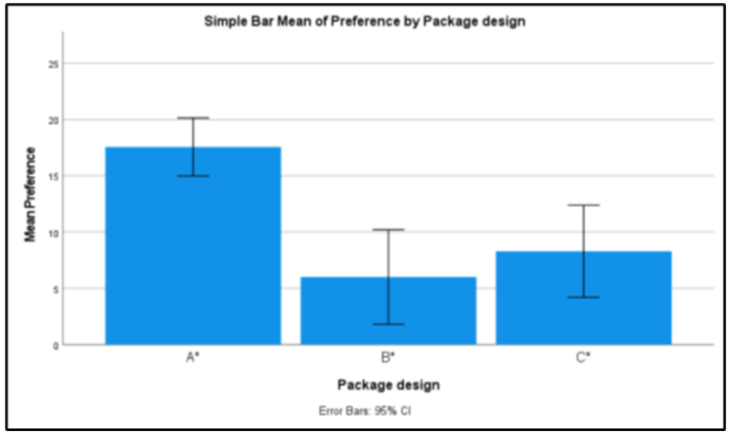
Multiple Comparisons

Dependent Variable: Preference

Tukey HSD

(I) Package design	(J) Package design	Mean Difference (I-J)	Std. Error	Sig.	95% Confidence Interval Lower Bound	Upper Bound
A*	B*	11.556*	2.051	<.001	6.35	16.77
	C*	9.270*	1.961	<.001	4.29	14.25
B*	A*	-11.556*	2.051	<.001	-16.77	-6.35
	C*	-2.286	2.165	.552	-7.79	3.21
C*	A*	-9.270*	1.961	<.001	-14.25	-4.29
	B*	2.286	2.165	.552	-3.21	7.79

*. The mean difference is significant at the 0.05 level.



Yes, package A had a significantly higher preference than B or C.

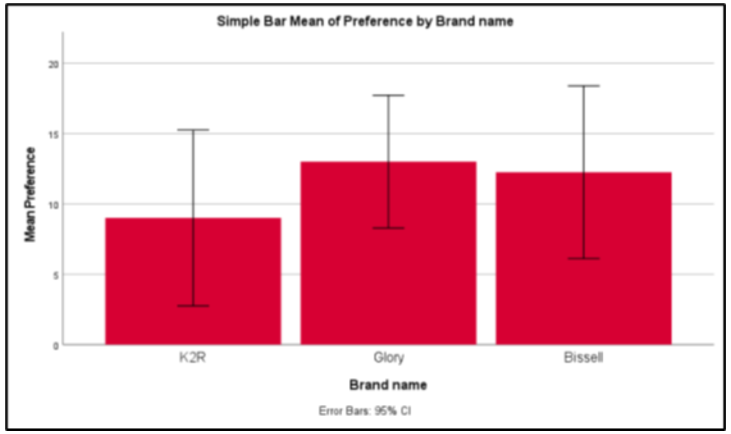
#6) Are carpet cleaner preferences different across brand?

ANOVA

Preference

	Sum of Squares	df	Mean Square	F	Sig.
Between Groups	63.955	2	31.977	.752	.485
Within Groups	807.500	19	42.500		
Total	871.455	21			

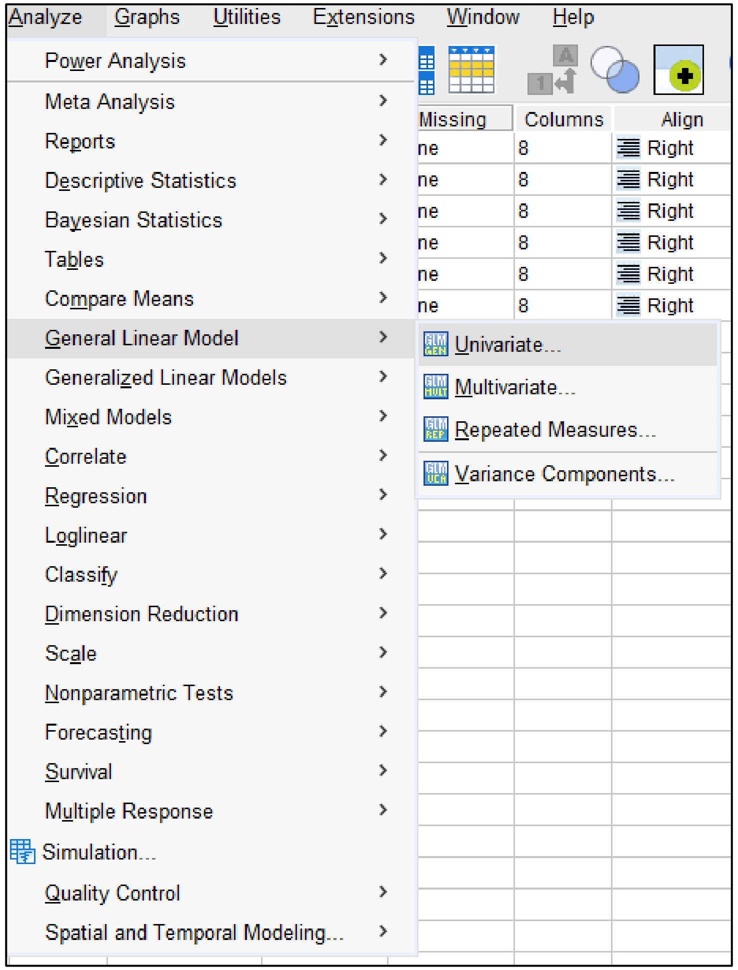
	N	Mean	Std. Deviation
K2R	7	9.00	6.758
Glory	7	13.00	5.099
Bissell	8	12.25	7.344
Total	22	11.45	6.442

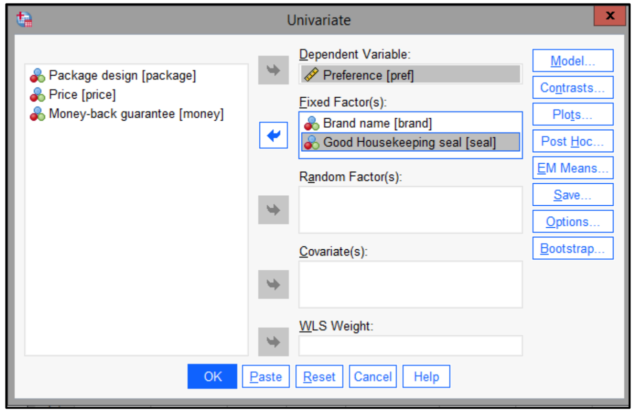


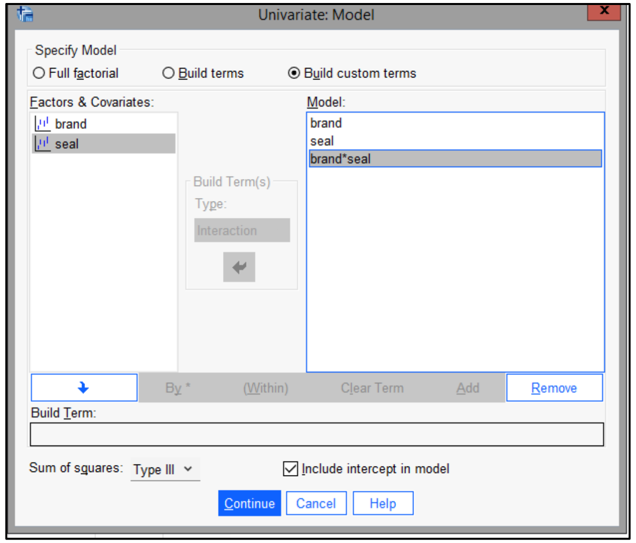
No, there was no difference across brand.

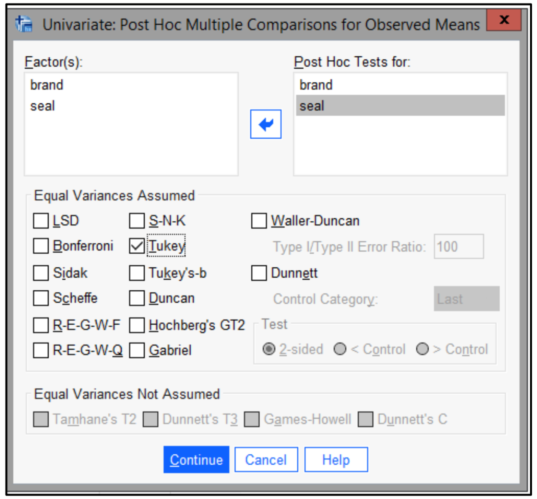
Two-way ANOVA

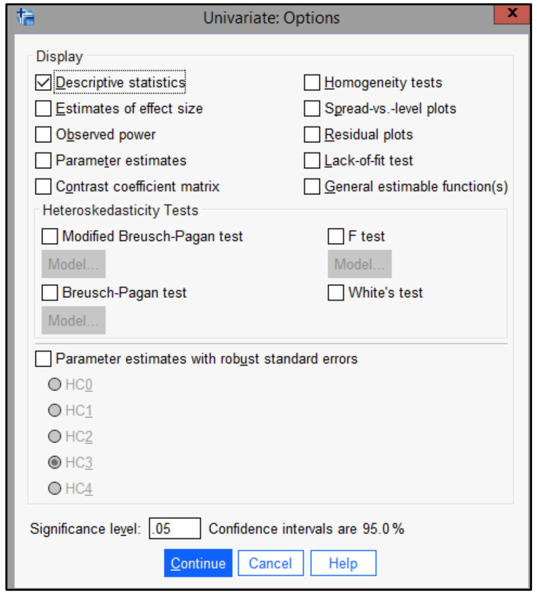
Tests if a variable's mean is different between a two categories with multiple groups each

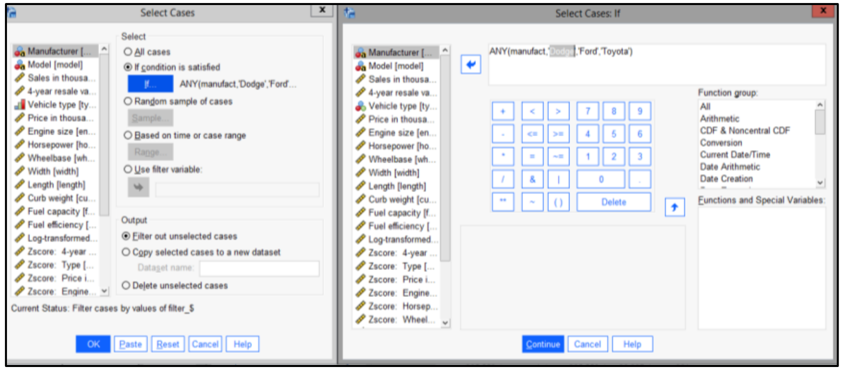












DaCCoTA

DAKOTA CANCER COLLABORATIVE ON TRANSLATIONAL ACTIVITY

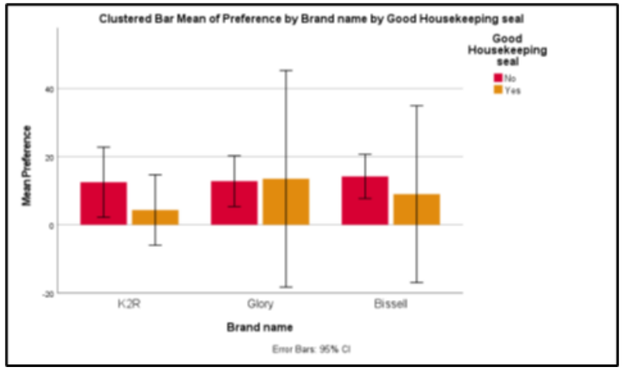
UNIVERSITY OF NORTH DAKOTA

5. Two-way ANOVA

Tests if a variable's mean is different between a two categories with multiple groups each

#1) Are carpet cleaner preferences different across brand and seal?

No, there was no difference across brand, seal, or the interaction.



Descriptive Statistics

Dependent Variable: Preference

Brand name	Good Housekeeping seal	Mean	Std. Deviation	N
K2R	No	12.50	6.455	4
	Yes	4.33	4.163	3
	Total	9.00	6.758	7
Glory	No	12.80	5.975	5
	Yes	13.50	3.536	2
	Total	13.00	5.099	7
Bissell	No	14.20	5.215	5
	Yes	9.00	10.440	3
	Total	12.25	7.344	8
Total	No	13.21	5.437	14
	Yes	8.38	7.249	8
	Total	11.45	6.442	22

Tests of Between-Subjects Effects

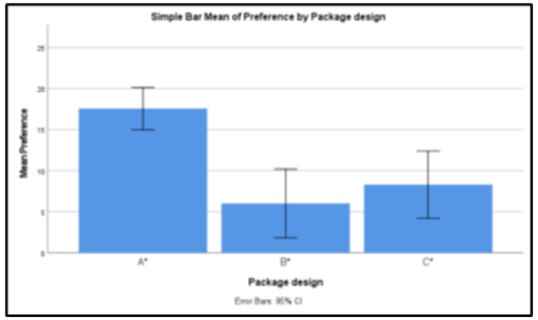
Dependent Variable: Preference

Source	Type III Sum of Squares	df	Mean Square	F	Sig.
Corrected Model	229.688[a]	5	45.938	1.145	.377
Intercept	2422.080	1	2422.080	60.385	<.001
brand	74.833	2	37.417	.933	.414
seal	88.318	1	88.318	2.202	.157
brand * seal	62.589	2	31.295	.780	.475
Error	641.767	16	40.110		
Total	3758.000	22			
Corrected Total	871.455	21			

a. R Squared = .264 (Adjusted R Squared = .033)

#2) Are carpet cleaner preferences different across brand and package?

Yes, package A was significantly higher than B and C. Brand and interaction were not significant.



Descriptive Statistics

Dependent Variable: Preference

Brand name	Package design	Mean	Std. Deviation	N
K2R	A*	17.50	3.536	2
	B*	3.00	2.828	2
	C*	7.33	3.786	3
	Total	9.00	6.758	7
Glory	A*	16.33	5.508	3
	B*	10.00	4.243	2
	C*	11.00	4.243	2
	Total	13.00	5.099	7
Bissell	A*	18.50	1.732	4
	B*	5.00	1.414	2
	C*	7.00	7.071	2
	Total	12.25	7.344	8
Total	A*	17.56	3.358	9
	B*	6.00	4.000	6
	C*	8.29	4.424	7
	Total	11.45	6.442	22

Tests of Between-Subjects Effects

Dependent Variable: Preference

Source	Type III Sum of Squares	df	Mean Square	F	Sig.
Corrected Model	664.621[a]	8	83.078	5.222	.004
Intercept	2336.709	1	2336.709	146.868	<.001
brand	36.108	2	18.054	1.135	.351
package	537.231	2	268.616	16.883	<.001
brand * package	55.229	4	13.807	.868	.509
Error	206.833	13	15.910		
Total	3758.000	22			
Corrected Total	871.455	21			

a. R Squared = .763 (Adjusted R Squared = .617)

Multiple Comparisons

Dependent Variable: Preference

Tukey HSD

(I) Package design	(J) Package design	Mean Difference (I-J)	Std. Error	Sig.	95% Confidence Interval Lower Bound	95% Confidence Interval Upper Bound
A*	B*	11.56*	2.102	<.001	6.00	17.11
	C*	9.27*	2.010	.001	3.96	14.58
B*	A*	-11.56*	2.102	<.001	-17.11	-6.00
	C*	-2.29	2.219	.572	-8.15	3.57
C*	A*	-9.27*	2.010	.001	-14.58	-3.96

5. Two-way ANOVA

DaCCoTA
DAKOTA CANCER COLLABORATIVE ON TRANSLATIONAL ACTIVITY

UNIVERSITY OF NORTH DAKOTA

Tests if a variable's mean is different between a two categories with multiple groups each

#3) Are car mpg numbers different across type and manufacturer (Dodge, Ford, Toyota)?

Yes, mpg was significantly higher in automobiles and mpg was significantly higher in Toyota compared to Dodge.

Tests of Between-Subjects Effects

Dependent Variable: Fuel efficiency

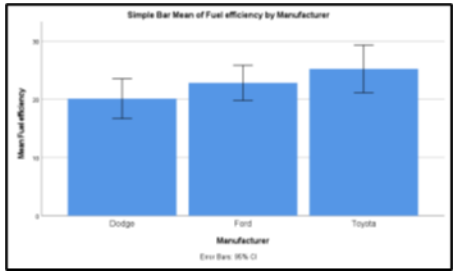
Source	Type III Sum of Squares	df	Mean Square	F	Sig.
Corrected Model	414.933[a]	5	82.987	5.823	.001
Intercept	15344.260	1	15344.260	1076.685	<.001
type	285.558	1	285.558	20.037	<.001
manufact	117.069	2	58.535	4.107	.029
type * manufact	4.621	2	2.310	.162	.851
Error	342.033	24	14.251		
Total	16125.000	30			
Corrected Total	756.967	29			

a. R Squared = .548 (Adjusted R Squared = .454)

Descriptive Statistics

Dependent Variable: Fuel efficiency

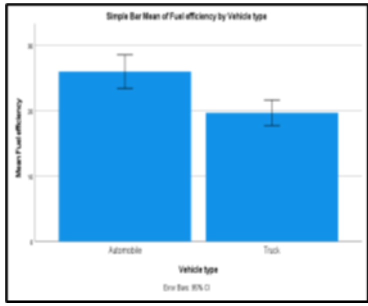
Vehicle type	Manufacturer	Mean	Std. Deviation	N
Automobile	Dodge	23.25	5.377	4
	Ford	25.67	3.615	6
	Toyota	29.25	3.304	4
	Total	26.00	4.455	14
Truck	Dodge	18.00	3.225	6
	Ford	19.40	2.702	5
	Toyota	22.00	4.359	5
	Total	19.69	3.665	16
Total	Dodge	20.10	4.771	10
	Ford	22.82	4.490	11
	Toyota	25.22	5.310	9
	Total	22.63	5.109	30



Multiple Comparisons

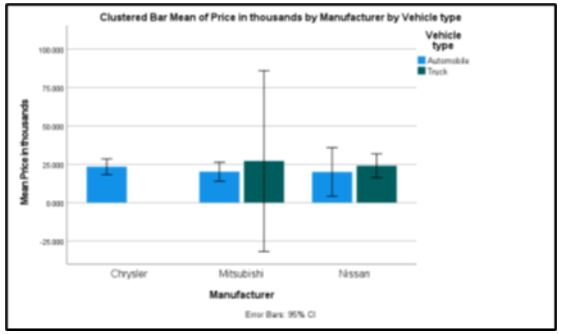
Dependent Variable: Fuel efficiency
Tukey HSD

(I) Manufacturer	(J) Manufacturer	Mean Difference (I-J)	Std. Error	Sig.
Dodge	Ford	-2.72	1.649	.246
	Toyota	-5.12*	1.735	.018
Ford	Dodge	2.72	1.649	.246
	Toyota	-2.40	1.697	.348
Toyota	Dodge	5.12*	1.735	.018



#4) Are car prices numbers different across type and manufacturer (Chrysler, Mitsubishi, Nissan)?

No, there was no difference across type and manufacturer.



Descriptive Statistics

Dependent Variable: Price in thousands

Vehicle type	Manufacturer	Mean	Std. Deviation	N
Automobile	Chrysler	23.43083	4.918479	6
	Mitsubishi	20.16760	4.964421	5
	Nissan	20.04600	6.381957	3
	Total	21.54007	5.103263	14
Truck	Mitsubishi	27.16700	6.561951	2
	Nissan	24.09675	4.918389	4
	Total	25.12017	5.063585	6
Total	Chrysler	23.43083	4.918479	6
	Mitsubishi	22.16743	5.938983	7
	Nissan	22.36071	5.509982	7
	Total	22.61410	5.234471	20

Tests of Between-Subjects Effects

Dependent Variable: Price in thousands

Source	Type III Sum of Squares	df	Mean Square	F	Sig.
Corrected Model	103.965[a]	4	25.991	.936	.470
Intercept	9030.404	1	9030.404	325.124	<.001
type	95.147	1	95.147	3.426	.084
manufact	41.575	2	20.787	.748	.490
type * manufact	6.775	1	6.775	.244	.629
Error	416.629	15	27.775		
Total	10748.544	20			
Corrected Total	520.594	19			

a. R Squared = .200 (Adjusted R Squared = -.014)

DaCCoTA
DAKOTA CANCER COLLABORATIVE ON TRANSLATIONAL ACTIVITY

UNIVERSITY OF NORTH DAKOTA

5. Two-way ANOVA

Tests if a variable's mean is different between a two categories with multiple groups each

#5) Are hourly nurse wages different across position and age range?

Yes, wages were significantly higher in hospitals compared to offices and wagers were highest in the top age range.

Descriptive Statistics

Dependent Variable: Hourly Salary

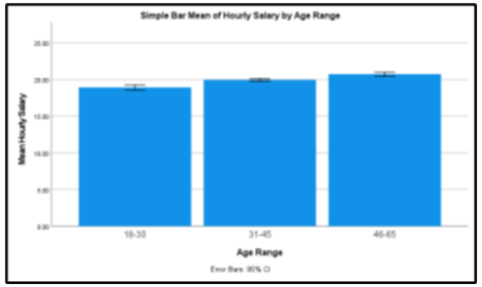
Nurse Type	Age Range	Mean	Std. Deviation	N
Hospital	18-30	19.7927	3.25954	314
	31-45	20.5993	3.44961	1057
	46-65	21.3018	3.59111	574
	Total	20.6764	3.49582	1945
Office	18-30	17.1317	4.59084	154
	31-45	18.6409	4.42222	525
	46-65	19.6023	4.66785	287
	Total	18.6859	4.58852	966
Total	18-30	18.9171	3.94875	468
	31-45	19.9494	3.90906	1582
	46-65	20.7353	4.05968	861
	Total	20.0159	4.00309	2911

Tests of Between-Subjects Effects

Dependent Variable: Hourly Salary

Source	Type III Sum of Squares	df	Mean Square	F	Sig.
Corrected Model	3647.268[a]	5	729.454	49.298	<.001
Intercept	771875.277	1	771875.277	52165.043	.000
position	2248.791	1	2248.791	151.978	<.001
agerange	1074.173	2	537.086	36.297	<.001
position * agerange	63.010	2	31.505	2.129	.119
Error	42984.680	2905	14.797		
Total	1212879.422	2911			
Corrected Total	46631.948	2910			

a. R Squared = .078 (Adjusted R Squared = .077)



Multiple Comparisons

Dependent Variable: Hourly Salary
Tukey HSD

(I) Age Range	(J) Age Range	Mean Difference (I-J)	Std. Error	Sig.	95% Confidence Interval Lower Bound	95% Confidence Interval Upper Bound
18-30	31-45	-1.0323*	.20241	<.001	-1.5069	-.5577
	46-65	-1.8182*	.22091	<.001	-2.3363	-1.3002
31-45	18-30	1.0323*	.20241	<.001	.5577	1.5069
	46-65	-.7859*	.16291	<.001	-1.1679	-.4039
46-65	18-30	1.8182*	.22091	<.001	1.3002	2.3363
	31-45	.7859*	.16291	<.001	.4039	1.1679

Based on observed means.
The error term is Mean Square(Error) = 14.797.
*. The mean difference is significant at the .05 level.

#6) Are hourly nurse wages different across age range and experience?

Yes, wages were different across experience, but not range or interaction. As experience increases, wage increases as well (immediately adjacent experience categories not significantly different).

Tests of Between-Subjects Effects

Dependent Variable: Hourly Salary

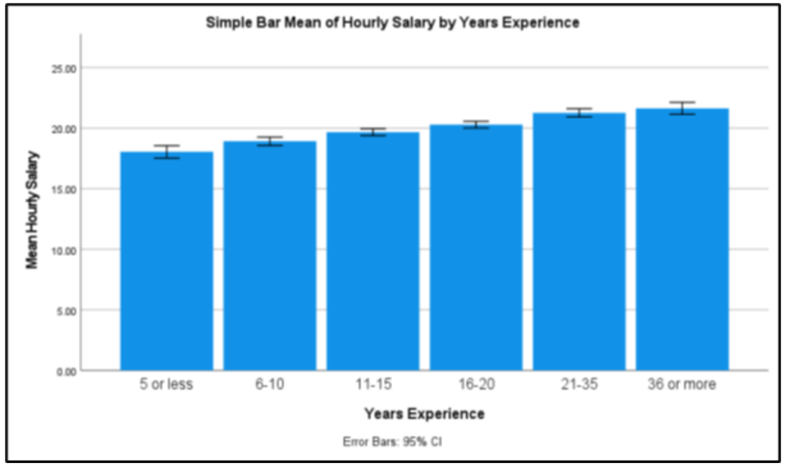
Source	Type III Sum of Squares	df	Mean Square	F	Sig.
Corrected Model	3086.856[a]	13	237.450	15.797	<.001
Intercept	336305.109	1	336305.109	22373.954	.000
agerange	3.863	2	1.931	.128	.879
yrsscale	628.261	5	125.652	8.359	<.001
agerange * yrsscale	109.518	6	18.253	1.214	.296
Error	43545.092	2897	15.031		
Total	1212879.422	2911			
Corrected Total	46631.948	2910			

a. R Squared = .066 (Adjusted R Squared = .062)

Descriptive Statistics

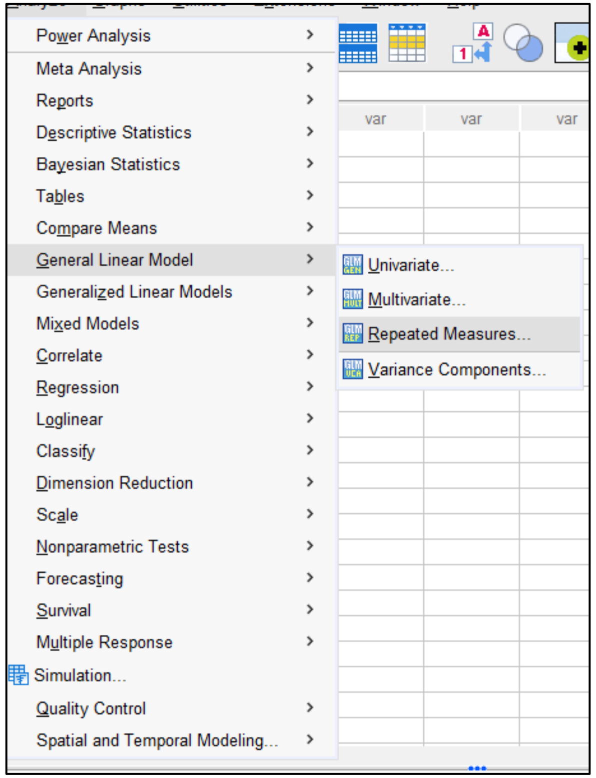
Dependent Variable: Hourly Salary

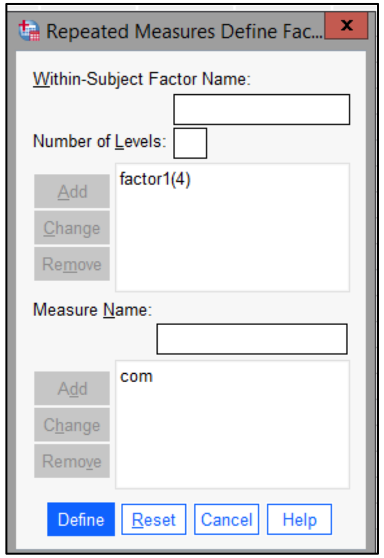
Age Range	Years Experience	Mean	Std. Deviation	N
18-30	5 or less	17.8515	3.91248	157
	6-10	18.8066	3.88162	186
	11-15	20.4199	3.64067	125
	Total	18.9171	3.94875	468
31-45	5 or less	18.5318	3.78045	62
	6-10	18.9972	3.72874	260
	11-15	19.5573	3.87657	525
	16-20	20.4038	3.76687	490
	21-35	21.2498	4.00631	245
	Total	19.9494	3.90906	1582
46-65	5 or less	17.7707	2.96022	2
	6-10	18.8936	3.50454	14
	11-15	19.2688	4.27079	102
	16-20	20.0495	3.94730	239
	21-35	21.2674	4.15928	294
	36 or more	21.6342	3.61826	210
	Total	20.7353	4.05968	861
Total	5 or less	18.0416	3.86667	221
	6-10	18.9169	3.77816	460
	11-15	19.6616	3.90528	752
	16-20	20.2876	3.82786	729
	21-35	21.2594	4.08669	539
	36 or more	21.6342	3.61826	210
	Total	20.0159	4.00309	2911

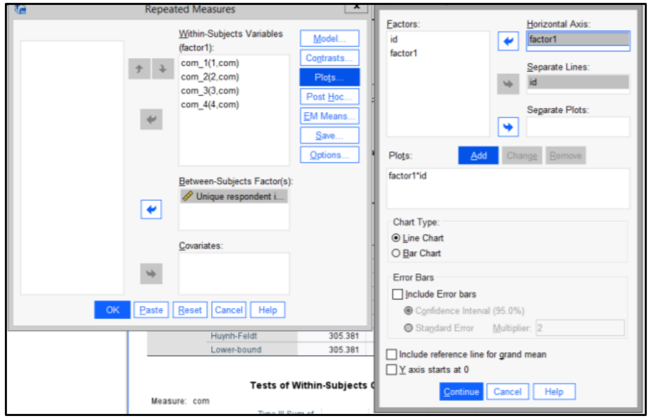


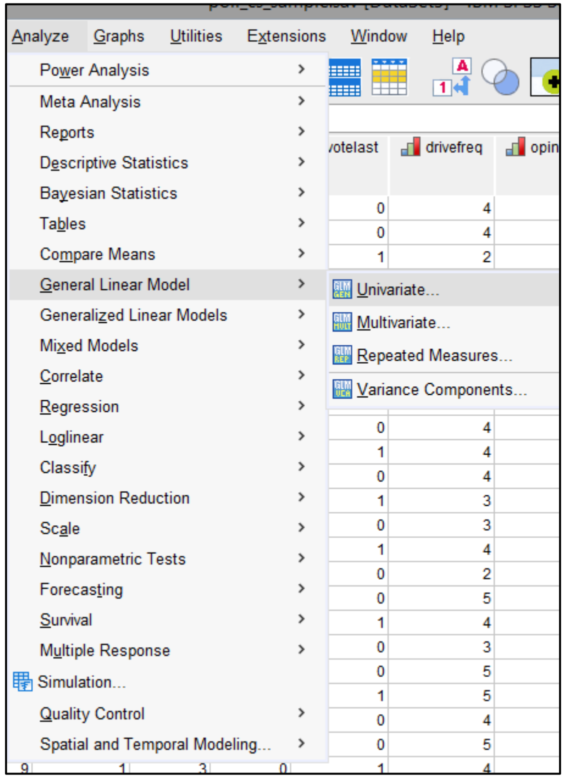
6. Advanced ANOVA

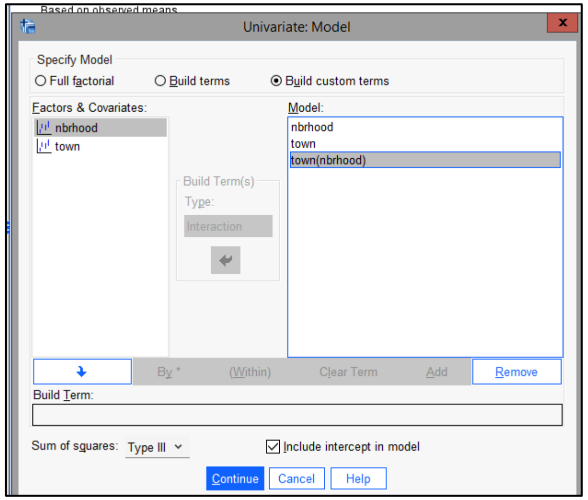
Tests if a variable's mean is different across categories while accounting for blocking, nesting, or repeated measures

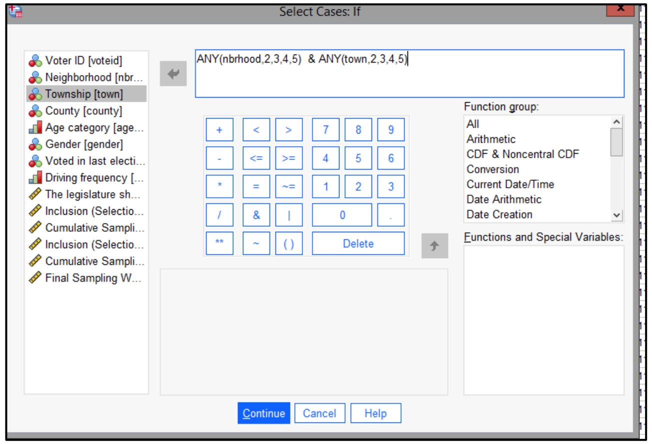












DaCCoTA
DAKOTA CANCER COLLABORATIVE ON TRANSLATIONAL ACTIVITY

UNIVERSITY OF NORTH DAKOTA

6. Advanced ANOVA

Tests if a variable's mean is different across categories while accounting for blocking, nesting, or repeated measures

#1) Is revenue different across product, while controlling for continent?

No, after controlling for continent, there was no difference between product type.

Report

Revenue

Product	Mean	N	Std. Deviation
Product A	669.8780	401	259.93054
Product B	791.8251	396	270.55511
Product C	785.8969	203	214.92935
Total	741.7209	1000	262.29698

Tests of Between-Subjects Effects

Dependent Variable: Revenue

Source	Type III Sum of Squares	df	Mean Square	F	Sig.
Corrected Model	42796103.27[a]	16	2674756.455	101.381	<.001
Intercept	281677478.93	1	281677478.93	10676.348	.000
Product	34179.829	2	17089.914	.648	.523
Continent	18557670.930	5	3711534.186	140.677	<.001
Product * Continent	85217.207	9	9468.579	.359	.954
Error	25934800.396	983	26383.317		
Total	618880764.49	1000			
Corrected Total	68730903.669	999			

a. R Squared = .623 (Adjusted R Squared = .617)

#2) Is shampoo pH different across shift, while controlling for batch?

Yes, after controlling for batch, afternoon shifts were significantly higher than morning and night shifts, and morning shifts were higher than night.

Report

pH level

Working shift	Mean	N	Std. Deviation
Night	4.8427	90	.21289
Morning	4.9848	78	.21952
Afternoon	5.1849	72	.21119
Total	4.9915	240	.25551

Tests of Between-Subjects Effects

Dependent Variable: pH level

Source		Type III Sum of Squares	df	Mean Square	F	Sig.
Intercept	Hypothesis	5959.005	1	5959.005	19159.746	<.001
	Error	1.555	5	.311[a]		
shift	Hypothesis	4.692	2	2.346	108.557	<.001
	Error	.216	10	.022[b]		
batch	Hypothesis	1.555	5	.311	14.340	<.001
	Error	.220	10.158	.022[c]		
shift * batch	Hypothesis	.216	10	.022	.527	.870
	Error	9.105	222	.041[d]		

a. MS(batch)
b. MS(shift * batch)
c. .996 MS(shift * batch) + .004 MS(Error)
d. MS(Error)

Multiple Comparisons

Dependent Variable: pH level
Tukey HSD

(I) Working shift	(J) Working shift	Mean Difference (I-J)	Std. Error	Sig.	95% Confidence Interval Lower Bound	95% Confidence Interval Upper Bound
Night	Morning	-.1422*	.03133	<.001	-.2161	-.0683
	Afternoon	-.3423*	.03202	<.001	-.4178	-.2667
Morning	Night	.1422*	.03133	<.001	.0683	.2161
	Afternoon	-.2001*	.03310	<.001	-.2782	-.1220
Afternoon	Night	.3423*	.03202	<.001	.2667	.4178
	Morning	.2001*	.03310	<.001	.1220	.2782

DaCCoTA
DAKOTA CANCER COLLABORATIVE ON TRANSLATIONAL ACTIVITY

UND UNIVERSITY OF NORTH DAKOTA

6. Advanced ANOVA

Tests if a variable's mean is different across categories while accounting for blocking, nesting, or repeated measures

#3) Is gas tax opinion different across neighborhood (2-5) when nested in township (2-5)?

Neighborhood * Township Crosstabulation

Count

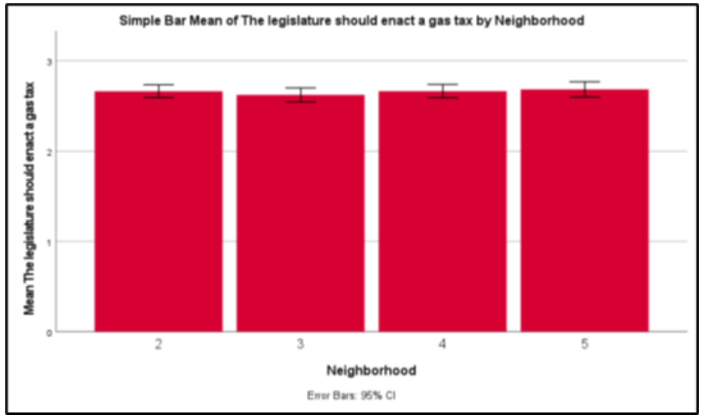
		Township 2	3	4	5	Total
Neighborhood	2	262	209	348	119	938
	3	257	252	173	61	743
	4	129	245	326	139	839
	5	144	224	154	125	647
Total		792	930	1001	444	3167

Tests of Between-Subjects Effects

Dependent Variable: The legislature should enact a gas tax

Source	Type III Sum of Squares	df	Mean Square	F	Sig.
Corrected Model	74.855[a]	15	4.990	4.226	<.001
Intercept	18910.720	1	18910.720	16014.547	.000
nbrhood	1.137	3	.379	.321	.810
town	25.497	3	8.499	7.197	<.001
town(nbrhood)	45.117	9	5.013	4.245	<.001
Error	3720.847	3151	1.181		
Total	26187.000	3167			
Corrected Total	3795.702	3166			

a. R Squared = .020 (Adjusted R Squared = .015)



No, there was no difference across neighborhood when nested in township.

#4) Is gas tax opinion different across neighborhood when nested in county?

Neighborhood * County Crosstabulation

Count

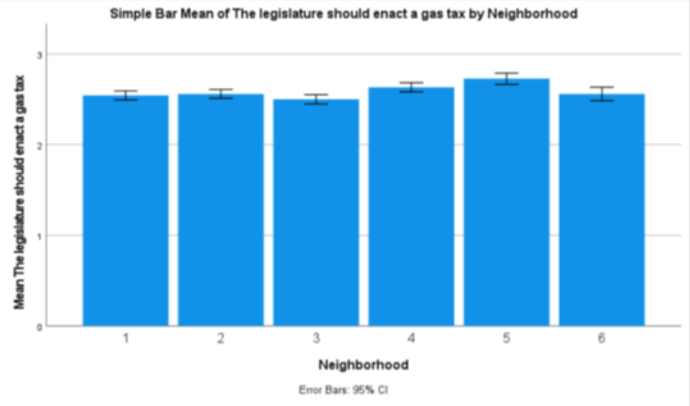
		County Eastern	Central	Western	Northern	Southern	Total
Neighborhood	1	408	425	199	516	309	1857
	2	401	358	339	562	361	2021
	3	436	320	239	459	328	1782
	4	420	242	370	452	275	1759
	5	295	221	287	243	150	1196
	6	399	80	90	141	124	834
Total		2359	1646	1524	2373	1547	9449

Tests of Between-Subjects Effects

Dependent Variable: The legislature should enact a gas tax

Source	Type III Sum of Squares	df	Mean Square	F	Sig.
Corrected Model	296.410[a]	29	10.221	8.716	<.001
Intercept	49609.058	1	49609.058	42303.280	.000
nbrhood	32.236	5	6.447	5.498	<.001
county	94.415	4	23.604	20.128	<.001
county(nbrhood)	107.858	20	5.393	4.599	<.001
Error	11045.661	9419	1.173		
Total	74293.000	9449			
Corrected Total	11342.072	9448			

a. R Squared = .026 (Adjusted R Squared = .023)



Yes, tax option was different across neighborhood when nested in county. Neighborhood 5 had the highest means, while 3 had the lowest.

6. Advanced ANOVA

DaCCoTA — Dakota Cancer Collaborative on Translational Activity

University of North Dakota

Tests if a variable's mean is different across categories while accounting for blocking, nesting, or repeated measures

#5) Are there differences between commercial ratings across 4 successive commercials?

Descriptive Statistics

	Mean	Std. Deviation	N
How appealing did you find the first commercial?	4.15	1.642	40
How appealing did you find the second commercial?	4.95	1.648	40
How appealing did you find the third commercial?	6.13	1.522	40
How appealing did you find the fourth commercial?	6.25	1.532	40

Mauchly's Test of Sphericity[a]

Measure: com

Within Subjects Effect	Mauchly's W	Approx. Chi-Square	df	Sig.	Epsilon[b] Greenhouse-Geisser	Huynh-Feldt	Lower-bound
factor1	.898	4.045	5	.543	.942	1.000	.333

Tests the null hypothesis that the error covariance matrix of the orthonormalized transformed dependent variables is proportional to an identity matrix.

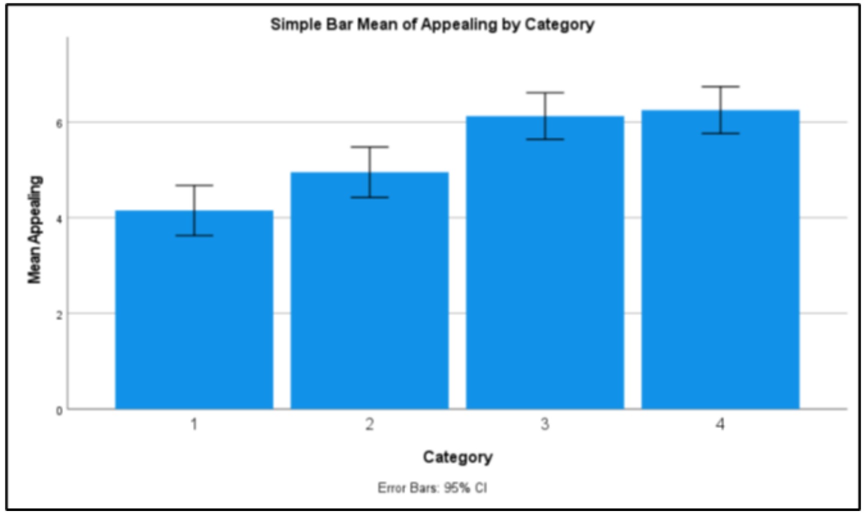
a. Design: Intercept
Within Subjects Design: factor1

b. May be used to adjust the degrees of freedom for the averaged tests of significance. Corrected tests are displayed in the Tests of Within-Subjects Effects table.

Tests of Within-Subjects Effects

Measure: com

Source		Type III Sum of Squares	df	Mean Square	F	Sig.
factor1	Sphericity Assumed	120.369	3	40.123	15.372	<.001
	Greenhouse-Geisser	120.369	2.827	42.572	15.372	<.001
	Huynh-Feldt	120.369	3.000	40.123	15.372	<.001
	Lower-bound	120.369	1.000	120.369	15.372	<.001
Error(factor1)	Sphericity Assumed	305.381	117	2.610		
	Greenhouse-Geisser	305.381	110.270	2.769		
	Huynh-Feldt	305.381	117.000	2.610		
	Lower-bound	305.381	39.000	7.830		



Yes, commercials 3 and 4 had significantly higher ratings than 1 and 2.

#6) Are the differences in shampoo pH across batches?

Descriptive Statistics

	Mean	Std. Deviation	N
ph_1	4.9183	.23747	40
ph_2	5.0104	.24267	40
ph_3	5.1464	.20504	40
ph_4	4.8991	.26423	40
ph_5	5.0146	.21793	40
ph_6	4.9604	.29096	40

Mauchly's Test of Sphericity[a]

Measure: bat

Within Subjects Effect	Mauchly's W	Approx. Chi-Square	df	Sig.	Epsilon[b] Greenhouse-Geisser	Huynh-Feldt	Lower-bound
factor1	.709	12.778	14	.545	.887	1.000	.200

Tests the null hypothesis that the error covariance matrix of the orthonormalized transformed dependent variables is proportional to an identity matrix.

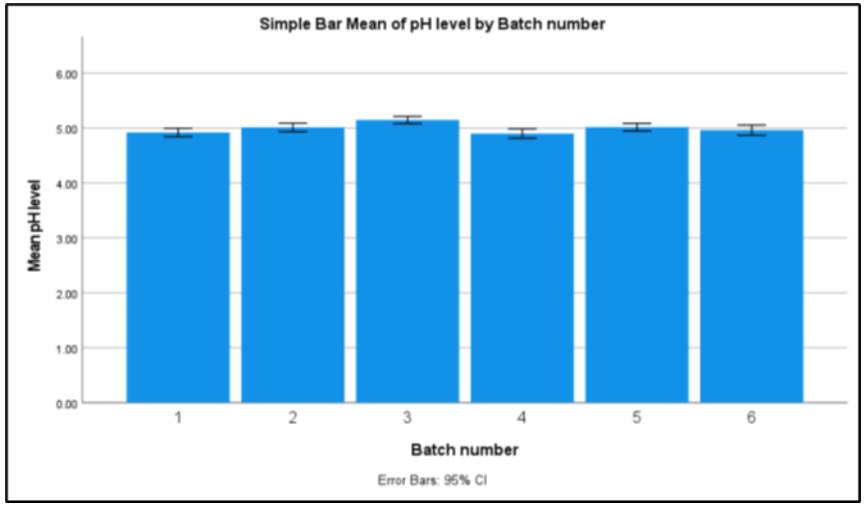
a. Design: Intercept
Within Subjects Design: factor1

b. May be used to adjust the degrees of freedom for the averaged tests of significance. Corrected tests are displayed in the Tests of Within-Subjects Effects table.

Tests of Within-Subjects Effects

Measure: bat

Source		Type III Sum of Squares	df	Mean Square	F	Sig.
factor1	Sphericity Assumed	1.590	5	.318	8.125	<.001
	Greenhouse-Geisser	1.590	4.433	.359	8.125	<.001
	Huynh-Feldt	1.590	5.000	.318	8.125	<.001
	Lower-bound	1.590	1.000	1.590	8.125	.007
Error(factor1)	Sphericity Assumed	7.633	195	.039		
	Greenhouse-Geisser	7.633	172.896	.044		
	Huynh-Feldt	7.633	195.000	.039		
	Lower-bound	7.633	39.000	.196		



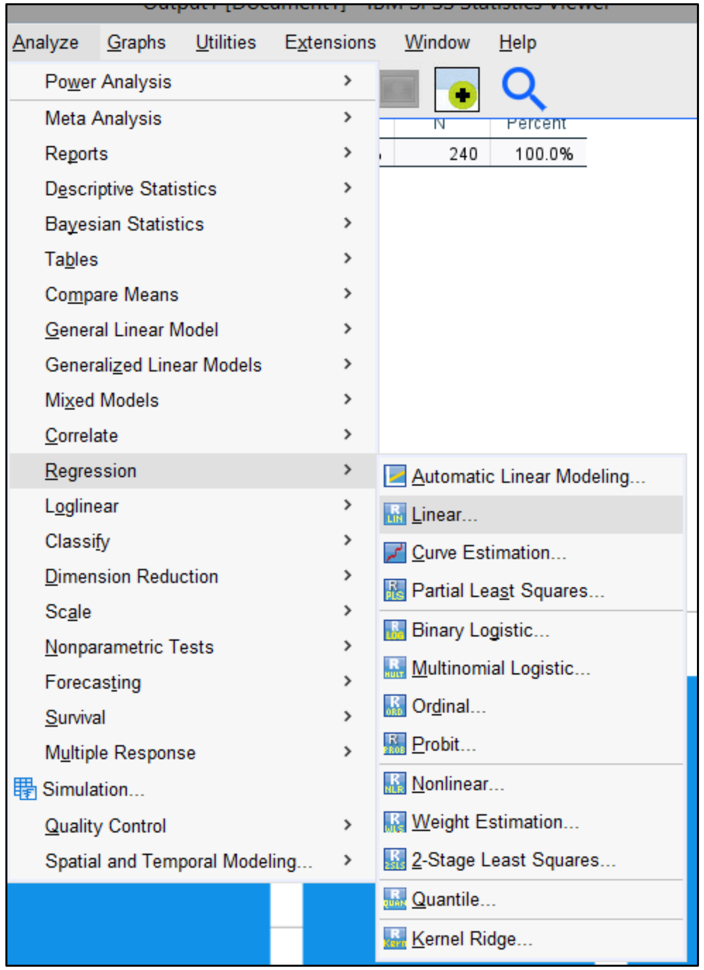
Yes, pH was significantly different across batch. Batch 3 had the highest pH.

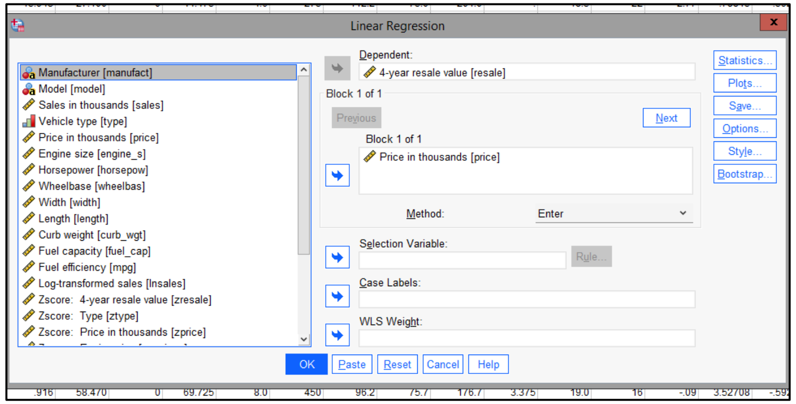
DaCCoTA
DAKOTA CANCER COLLABORATIVE ON TRANSLATIONAL ACTIVITY

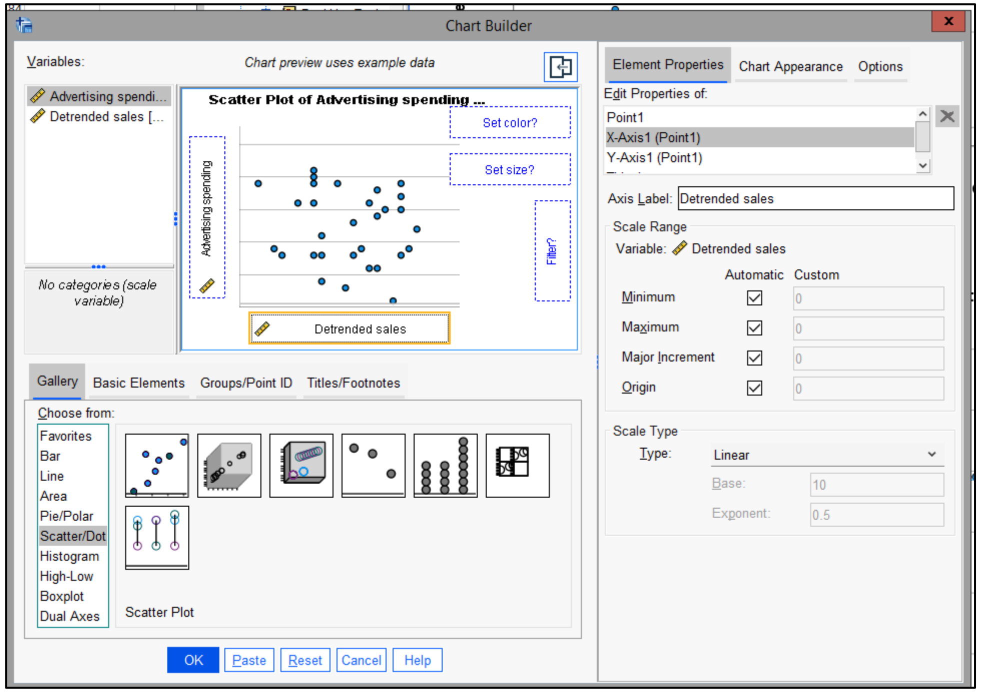
UND
UNIVERSITY OF NORTH DAKOTA

7. Simple Linear Regression

Tests if there is a relationship between a numerical response variable and one numerical predictor variable







DaCCoTA
DAKOTA CANCER COLLABORATIVE ON TRANSLATIONAL ACTIVITY

UNIVERSITY OF NORTH DAKOTA

7. Simple Linear Regression

Tests if there is a relationship between a numerical response variable and one numerical predictor variable

#1) Can sales be predicted by advertising spending?

Yes, there was a significant positive relationship. As advertising spending increased, sales increased.

ANOVA[a]

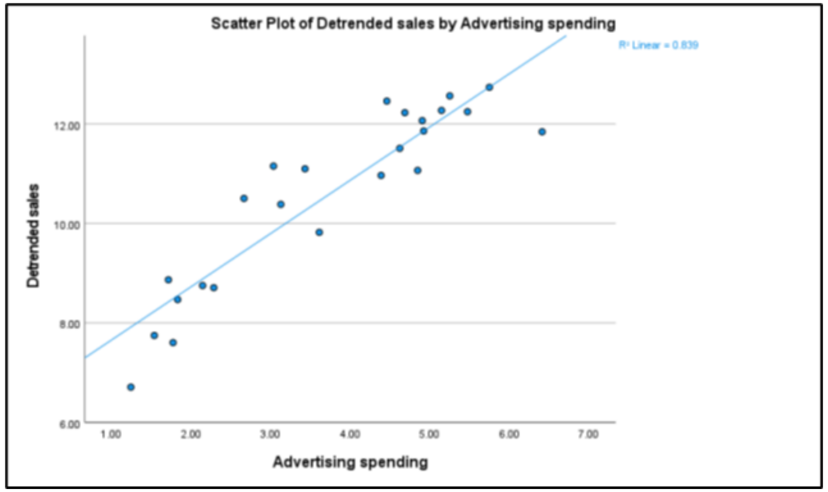
Model		Sum of Squares	df	Mean Square	F	Sig.
1	Regression	62.514	1	62.514	114.548	<.001[b]
	Residual	12.006	22	.546		
	Total	74.520	23			

a. Dependent Variable: Detrended sales
b. Predictors: (Constant), Advertising spending

Coefficients[a]

Model		Unstandardized Coefficients B	Unstandardized Coefficients Std. Error	Standardized Coefficients Beta	t	Sig.
1	(Constant)	6.584	.402		16.391	<.001
	Advertising spending	1.071	.100	.916	10.703	<.001

a. Dependent Variable: Detrended sales



#2) Can men's clothing sales be predicted by print advertising spending?

Yes, there was a significant positive relationship. As print advertising spending increased, men's clothing sales increased.

ANOVA[a]

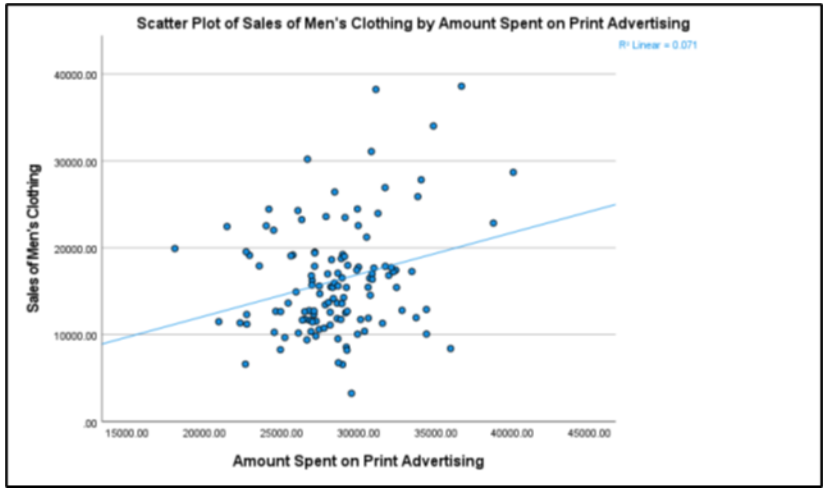
Model		Sum of Squares	df	Mean Square	F	Sig.
1	Regression	337502958.87	1	337502958.87	9.004	.003[b]
	Residual	4423148060.0	118	37484305.593		
	Total	4760651018.8	119			

a. Dependent Variable: Sales of Men's Clothing
b. Predictors: (Constant), Amount Spent on Print Advertising

Coefficients[a]

Model		Unstandardized Coefficients B	Unstandardized Coefficients Std. Error	Standardized Coefficients Beta	t	Sig.
1	(Constant)	2505.359	4612.162		.543	.588
	Amount Spent on Print Advertising	.482	.161	.266	3.001	.003

a. Dependent Variable: Sales of Men's Clothing



DaCCoTA
DAKOTA CANCER COLLABORATIVE ON TRANSLATIONAL ACTIVITY

7. Simple Linear Regression

UND UNIVERSITY OF NORTH DAKOTA

Tests if there is a relationship between a numerical response variable and one numerical predictor variable

#3) Can women's clothing sales be predicted by the number of customer service?

Yes, there was a significant positive relationship. As the number of sales representatives increased, women's clothing sales increased.

ANOVA[a]

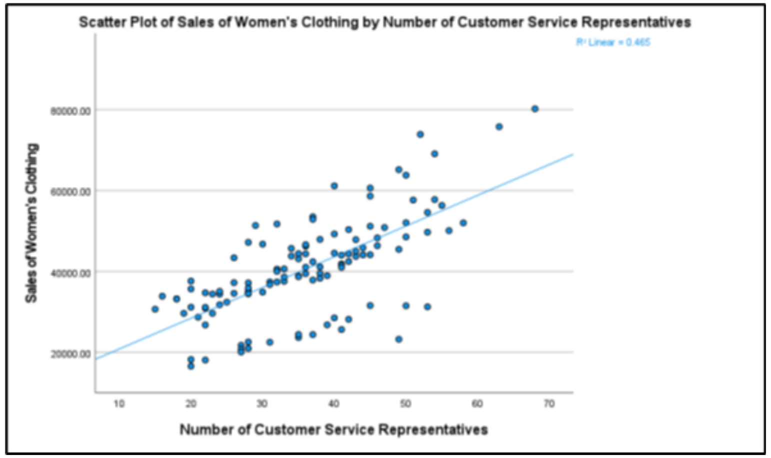
Model		Sum of Squares	df	Mean Square	F	Sig.
1	Regression	8232322474.5	1	8232322474.5	102.586	<.001[b]
	Residual	9469273576.2	118	80248081.154		
	Total	17701596051	119			

a. Dependent Variable: Sales of Women's Clothing
b. Predictors: (Constant), Number of Customer Service Representatives

Coefficients[a]

Model		Unstandardized Coefficients B	Std. Error	Standardized Coefficients Beta	t	Sig.
1	(Constant)	13244.439	2820.402		4.696	<.001
	Number of Customer Service Representatives	760.127	75.049	.682	10.128	<.001

a. Dependent Variable: Sales of Women's Clothing



#4) Can car price be predicted by mpg?

Yes, there was a significant negative relationship. As fuel efficiency (mpg) increased, car price decreased.

ANOVA[a]

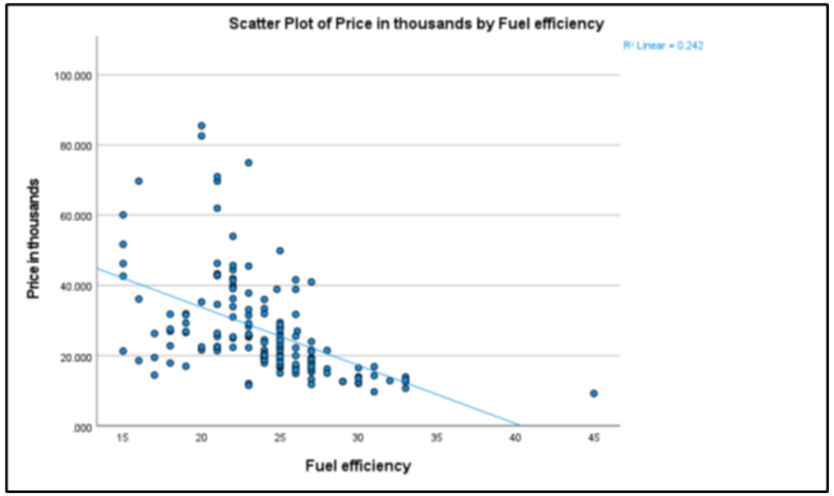
Model		Sum of Squares	df	Mean Square	F	Sig.
1	Regression	7682.281	1	7682.281	48.330	<.001[b]
	Residual	24002.287	151	158.956		
	Total	31684.567	152			

a. Dependent Variable: Price in thousands
b. Predictors: (Constant), Fuel efficiency

Coefficients[a]

Model		Unstandardized Coefficients B	Std. Error	Standardized Coefficients Beta	t	Sig.
1	(Constant)	66.904	5.767		11.601	<.001
	Fuel efficiency	-1.656	.238	-.492	-6.952	<.001

a. Dependent Variable: Price in thousands



7. Simple Linear Regression

Tests if there is a relationship between a numerical response variable and one numerical predictor variable

#5) Can car sales by predicted by fuel capacity?

No, there was no significant relationship between fuel capacity and car sales.

ANOVA[a]

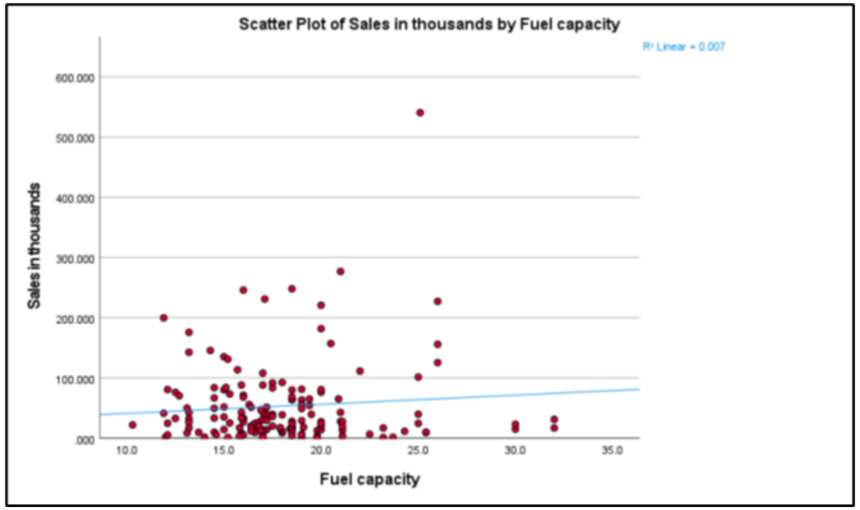
odel		Sum of Squares	df	Mean Square	F	Sig.
	Regression	5406.899	1	5406.899	1.162	.283[b]
	Residual	716561.220	154	4652.995		
	Total	721968.118	155			

a. Dependent Variable: Sales in thousands
b. Predictors: (Constant), Fuel capacity

Coefficients[a]

odel		Unstandardized Coefficients		Standardized Coefficients	t	Sig.
		B	Std. Error	Beta		
	(Constant)	25.724	25.881		.994	.322
	Fuel capacity	1.519	1.409	.087	1.078	.283

a. Dependent Variable: Sales in thousands



#5) Can car resale be predicted by price?

Yes, there was a significant positive relationship. As price increased, resale price increased.

ANOVA[a]

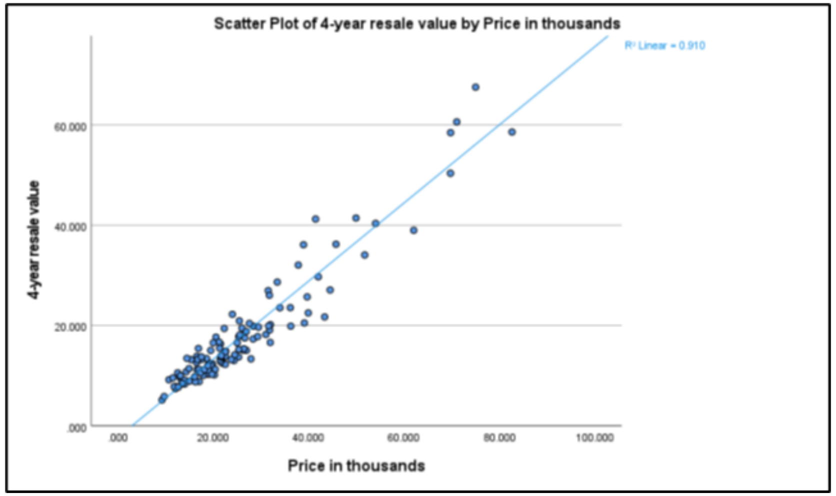
Model		Sum of Squares	df	Mean Square	F	Sig.
1	Regression	14319.890	1	14319.890	1180.282	<.001[b]
	Residual	1419.514	117	12.133		
	Total	15739.404	118			

a. Dependent Variable: 4-year resale value
b. Predictors: (Constant), Price in thousands

Coefficients[a]

Model		Unstandardized Coefficients		Standardized Coefficients	t	Sig.
		B	Std. Error	Beta		
1	(Constant)	-2.278	.673		-3.387	<.001
	Price in thousands	.779	.023	.954	34.355	<.001

a. Dependent Variable: 4-year resale value

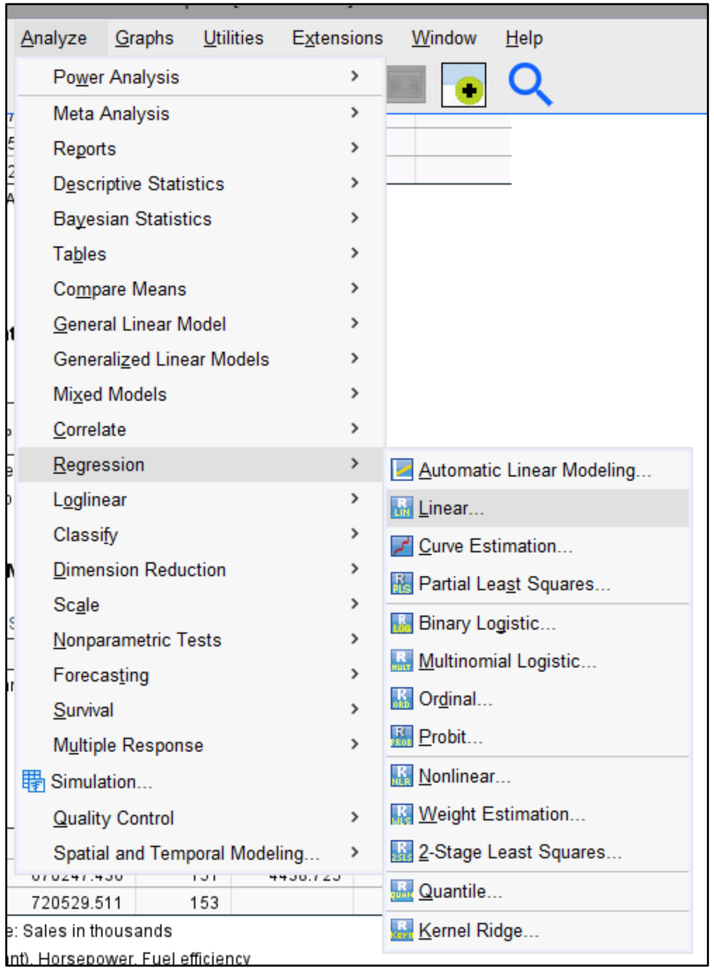


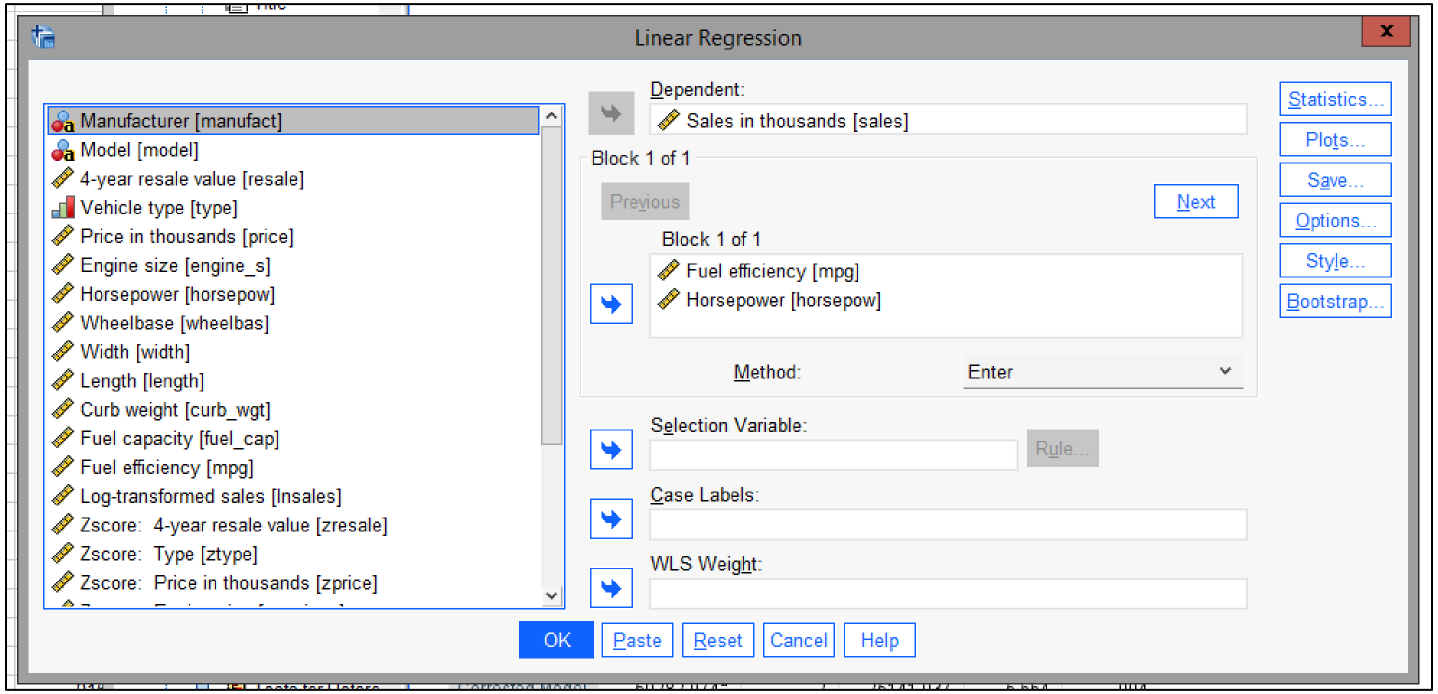
DaCCoTA
DAKOTA CANCER COLLABORATIVE ON TRANSLATIONAL ACTIVITY

UND
UNIVERSITY OF NORTH DAKOTA

8. Multiple Linear Regression

Tests if there is a relationship between a numerical response variable and multiple numerical predictor variables





8. Multiple Linear Regression

Tests if there is a relationship between a numerical response variable and multiple numerical predictor variables

#1) Can car sales be predicted by width and fuel capacity?

ANOVA[a]

Model		Sum of Squares	df	Mean Square	F	Sig.
1	Regression	14381.651	2	7190.826	1.555	.215[b]
	Residual	707586.467	153	4624.748		
	Total	721968.118	155			

a. Dependent Variable: Sales in thousands

b. Predictors: (Constant), Fuel capacity, Width

Coefficients[a]

Model		Unstandardized Coefficients B	Std. Error	Standardized Coefficients Beta	t	Sig.
1	(Constant)	-151.591	129.874		-1.167	.245
	Width	2.922	2.097	.148	1.393	.166
	Fuel capacity	-.183	1.862	-.010	-.098	.922

a. Dependent Variable: Sales in thousands

No, neither width nor fuel capacity significantly predicted car sales.

#2) can car mpg be predicted by horsepower, width, length, and curb weight?

ANOVA[a]

Model		Sum of Squares	df	Mean Square	F	Sig.
1	Regression	1961.099	4	490.275	86.201	<.001[b]
	Residual	841.758	148	5.688		
	Total	2802.858	152			

a. Dependent Variable: Fuel efficiency

b. Predictors: (Constant), Curb weight, Horsepower, Length, Width

Coefficients[a]

Model		Unstandardized Coefficients B	Std. Error	Standardized Coefficients Beta	t	Sig.
1	(Constant)	40.691	4.887		8.326	<.001
	Horsepower	-.013	.004	-.171	-2.953	.004
	Width	-.079	.093	-.063	-.842	.401
	Length	.045	.021	.141	2.149	.033
	Curb weight	-5.126	.495	-.757	-10.361	<.001

a. Dependent Variable: Fuel efficiency

Yes, horsepower, length and curb weight significantly predicted car mpg, while width did not. As horsepower and weight increased, mpg degreased. Conversely, when length increased, mpg increased.

DaCCoTA
DAKOTA CANCER COLLABORATIVE ON TRANSLATIONAL ACTIVITY

UNIVERSITY OF NORTH DAKOTA

8. Multiple Linear Regression

Tests if there is a relationship between a numerical response variable and multiple numerical predictor variables

#3) Can insurance claim costs be predicted by coverage and deductible?

ANOVA[a]

Model		Sum of Squares	df	Mean Square	F	Sig.
1	Regression	24142055.222	2	12071027.611	784.378	<.001[b]
	Residual	67897592.254	4412	15389.300		
	Total	92039647.477	4414			

a. Dependent Variable: Cost of claim in thousands

b. Predictors: (Constant), Deductible, Amount of coverage in thousands

Coefficients[a]

Model		Unstandardized Coefficients B	Std. Error	Standardized Coefficients Beta	t	Sig.
1	(Constant)	-.677	3.849		-.176	.860
	Amount of coverage in thousands	.222	.006	.502	34.951	<.001
	Deductible	.004	.002	.023	1.597	.110

a. Dependent Variable: Cost of claim in thousands

Yes, coverage significantly predicted insurance claim costs, while the deductible did not. As coverage increased, claim costs increased.

#4) Can insurance claim cost be predicted by coverage, income, and deductible?

ANOVA[a]

Model		Sum of Squares	df	Mean Square	F	Sig.
1	Regression	24270199.871	3	8090066.624	526.569	<.001[b]
	Residual	67769447.605	4411	15363.738		
	Total	92039647.477	4414			

a. Dependent Variable: Cost of claim in thousands

b. Predictors: (Constant), Household income in thousands, Deductible, Amount of coverage in thousands

Coefficients[a]

Model		Unstandardized Coefficients B	Std. Error	Standardized Coefficients Beta	t	Sig.
1	(Constant)	-3.822	3.997		-.956	.339
	Amount of coverage in thousands	.213	.007	.480	29.665	<.001
	Deductible	.004	.002	.023	1.584	.113
	Household income in thousands	.092	.032	.043	2.888	.004

a. Dependent Variable: Cost of claim in thousands

Yes, coverage and household income significantly predicted insurance claim costs, while the deductible did not. As coverage and income increased, claim costs increased.

DaCCoTA
DAKOTA CANCER COLLABORATIVE ON TRANSLATIONAL ACTIVITY

8. Multiple Linear Regression

UNIVERSITY OF NORTH DAKOTA

Tests if there is a relationship between a numerical response variable and multiple numerical predictor variables

#5) Can men's clothing sales be predicted by catalog, phone line, print advertising, and rep number?

ANOVA[a]

Model		Sum of Squares	df	Mean Square	F	Sig.
1	Regression	3742825209.2	4	935706302.31	105.722	<.001[b]
	Residual	1017825809.6	115	8850659.214		
	Total	4760651018.8	119			

a. Dependent Variable: Sales of Men's Clothing

b. Predictors: (Constant), Number of Customer Service Representatives, Amount Spent on Print Advertising, Number of Phone Lines Open for Ordering, Number of Catalogs Mailed

Coefficients[a]

Model		Unstandardized Coefficients B	Std. Error	Standardized Coefficients Beta	t	Sig.
1	(Constant)	-21399.415	2624.512		-8.154	<.001
	Number of Catalogs Mailed	2.000	.238	.537	8.416	<.001
	Number of Phone Lines Open for Ordering	361.274	44.888	.483	8.048	<.001
	Amount Spent on Print Advertising	.226	.080	.125	2.823	.006
	Number of Customer Service Representatives	-45.862	38.770	-.079	-1.183	.239

a. Dependent Variable: Sales of Men's Clothing

Yes, the number of catalogs mailed, number of phone lines open, and the amount spent on print advertising significantly predicted men's clothing; as each variable increased, sales increased. The number of customer service reps was not significant.

#6) Can women's clothing sales be predicted by catalog, phone line, print advertising, and rep number?

ANOVA[a]

Model		Sum of Squares	df	Mean Square	F	Sig.
1	Regression	11090126832	4	2772531708.0	48.225	<.001[b]
	Residual	6611469218.6	115	57491036.683		
	Total	17701596051	119			

a. Dependent Variable: Sales of Women's Clothing

b. Predictors: (Constant), Number of Customer Service Representatives, Amount Spent on Print Advertising, Number of Phone Lines Open for Ordering, Number of Catalogs Mailed

Coefficients[a]

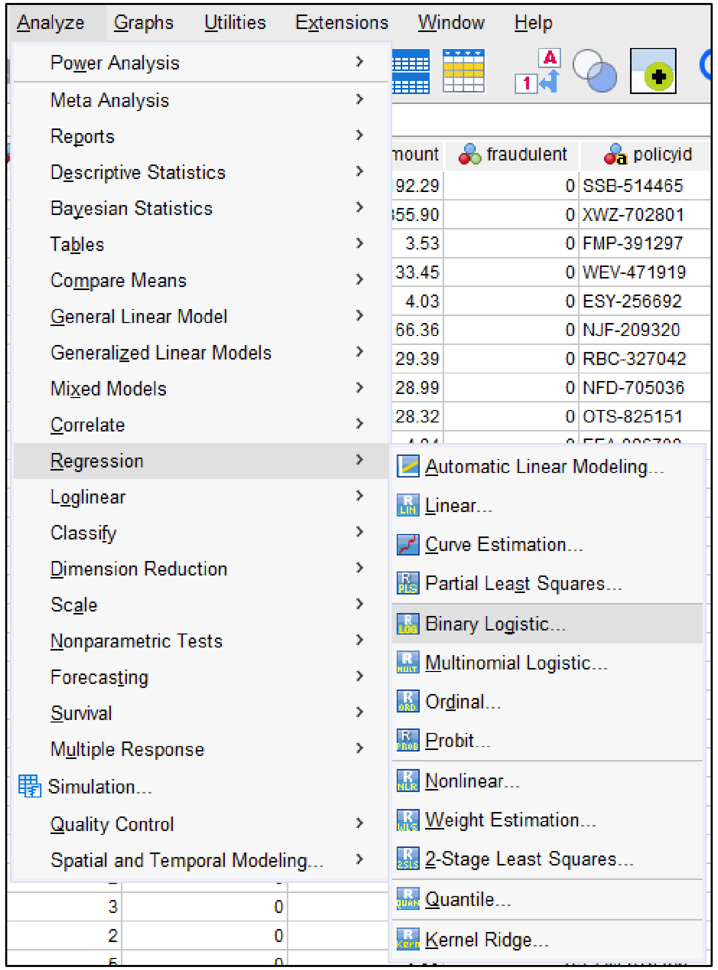
Model		Unstandardized Coefficients B	Std. Error	Standardized Coefficients Beta	t	Sig.
1	(Constant)	-29282.848	6688.994		-4.378	<.001
	Number of Catalogs Mailed	2.556	.606	.356	4.220	<.001
	Number of Phone Lines Open for Ordering	-13.701	114.406	-.010	-.120	.905
	Amount Spent on Print Advertising	1.020	.204	.292	4.996	<.001
	Number of Customer Service Representatives	427.210	98.812	.383	4.323	<.001

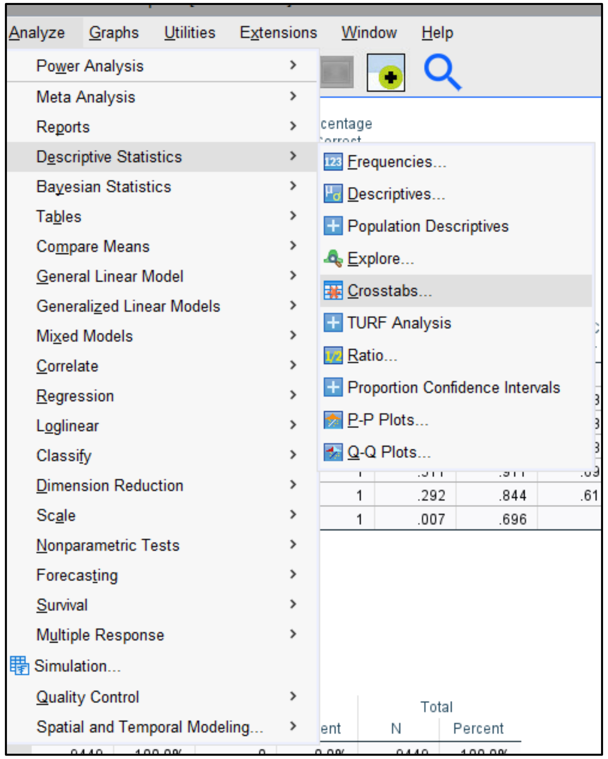
a. Dependent Variable: Sales of Women's Clothing

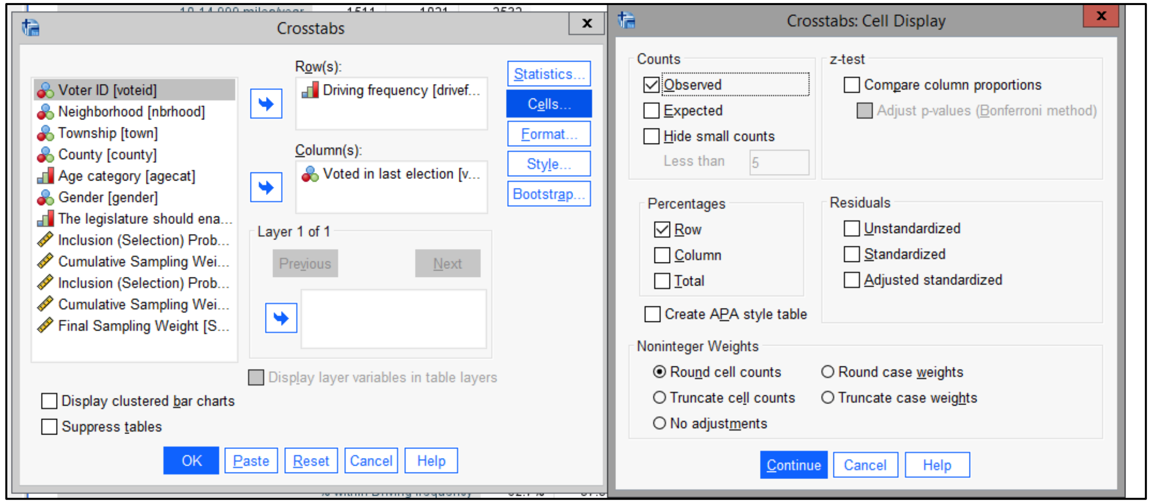
Yes, the number of catalogs mailed, the amount spent on print advertising, and the number of customer reps significantly predicted women's clothing; as each variable increased, sales increased. The number of phone lines open was not significant.

9. Logistic Regression

Tests if there is a relationship between a binary response variable and one or more predictor variables







9. Logistic Regression

DaCCoTA
DAKOTA CANCER COLLABORATIVE ON TRANSLATIONAL ACTIVITY

UNIVERSITY OF NORTH DAKOTA

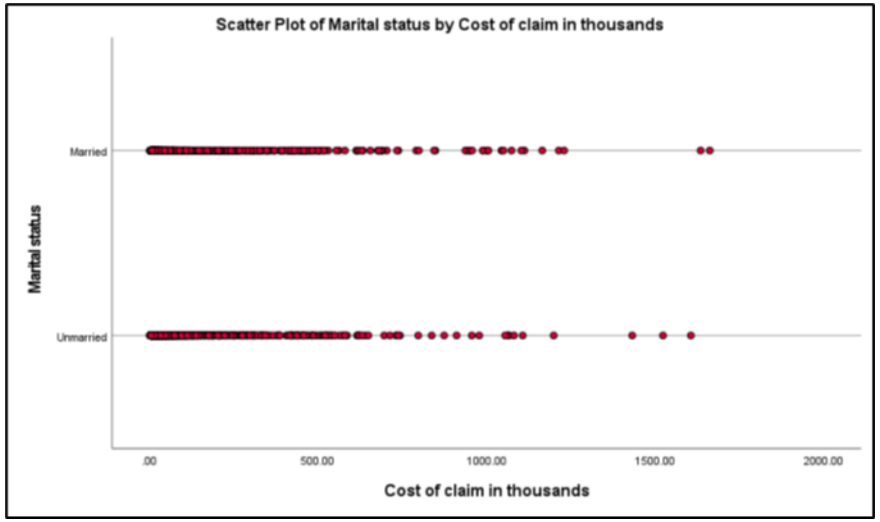
Tests if there is a relationship between a binary response variable and one or more predictor variables

#1) Can marital status of insurance claimants be predicted by claim cost?

Variables in the Equation

		B	S.E.	Wald	df	Sig.	Exp(B)
Step 1[a]	Cost of claim in thousands	.000	.000	.142	1	.707	1.000
	Constant	.145	.034	18.400	1	<.001	1.156

a. Variable(s) entered on step 1: Cost of claim in thousands.



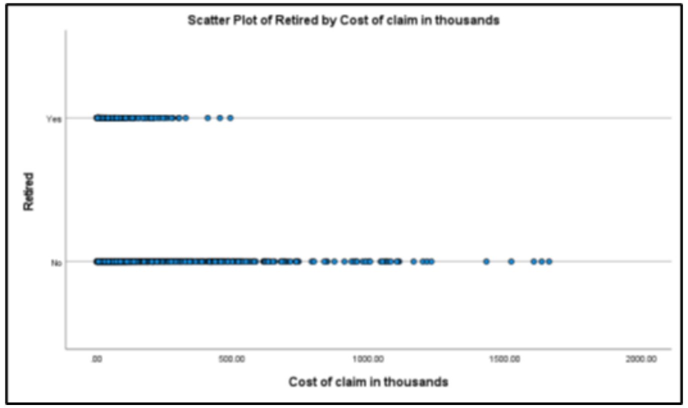
No, there was no significant relationship between claim cost and marital status.

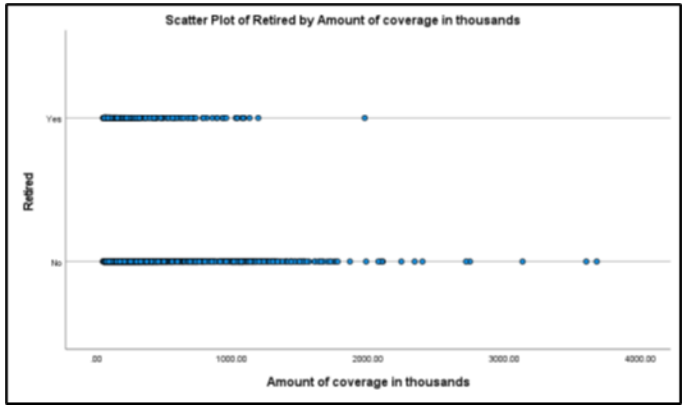
#2) Can retirement status of insurance claimants be predicted by claim cost and coverage?

Variables in the Equation

		B	S.E.	Wald	df	Sig.	Exp(B)
Step 1[a]	Cost of claim in thousands	-.002	.001	7.196	1	.007	.998
	Amount of coverage in thousands	-.002	.000	50.819	1	<.001	.998
	Constant	-1.294	.070	338.912	1	<.001	.274

a. Variable(s) entered on step 1: Cost of claim in thousands, Amount of coverage in thousands.





Yes, retirement status was significantly predicted by claim cost and coverage. As claim cost and coverage increased, the probability of being retired decreased.

DaCCoTA
DAKOTA CANCER COLLABORATIVE ON TRANSLATIONAL ACTIVITY

9. Logistic Regression

UNIVERSITY OF NORTH DAKOTA

Tests if there is a relationship between a binary response variable and one or more predictor variables

#3) Can HIV status be predicted by assay (numerical)?

Variables in the Equation

		B	S.E.	Wald	df	Sig.	Exp(B)
Step 1[a]	Assay result	1.414	.076	350.504	1	<.001	4.113
	Constant	-6.633	.438	229.091	1	<.001	.001

a. Variable(s) entered on step 1: Assay result.

Assay result * Actual state Crosstabulation

Count

		Actual state		Total
		HIV-negative	HIV-positive	
Assay result	1	942	3	945
	2	18	2	20
	3	16	2	18
	4	7	5	12
	5	2	10	12
	6	3	5	8
	7	9	8	17
	8	3	965	968
Total		1000	1000	2000

Yes, HIV status was significantly predicted by assay, when assay was treated as a numerical variable. As assay number increased, the probability of being HIV positive increased (or probability of being HIV negative decreased).

#4) Can HIV status be predicted by assay (categorical)?

Classification Table[a]

Observed			Predicted: Actual state HIV-negative	Predicted: Actual state HIV-positive	Percentage Correct
Step 1	Actual state	HIV-negative	992	8	99.2
		HIV-positive	20	980	98.0
	Overall Percentage				98.6

a. The cut value is .500

Variables in the Equation

		B	S.E.	Wald	df	Sig.	Exp(B)
Step 1[a]	Assay result			218.665	7	<.001	
	Assay result(1)	3.552	.943	14.178	1	<.001	34.889
	Assay result(2)	3.670	.947	15.017	1	<.001	39.250
	Assay result(3)	5.413	.823	43.263	1	<.001	224.286
	Assay result(4)	7.359	.967	57.955	1	<.001	1570.000
	Assay result(5)	6.260	.932	45.164	1	<.001	523.333
	Assay result(6)	5.632	.755	55.591	1	<.001	279.111
	Assay result(7)	11.523	.818	198.541	1	<.001	101003.333
	Constant	-5.749	.578	98.852	1	<.001	.003

a. Variable(s) entered on step 1: Assay result.

Yes, HIV status was significantly predicted by assay, when assay was treated as a categorical variable. Compared to assay category 1 (reference), the odds of being HIV positive was higher in all other categories and increased for each successive category.

DaCCoTA
DAKOTA CANCER COLLABORATIVE ON TRANSLATIONAL ACTIVITY

9. Logistic Regression

UNIVERSITY OF NORTH DAKOTA

Tests if there is a relationship between a binary response variable and one or more predictor variables

#5) Can past voting outcome be predicted by sex?

Gender * Voted in last election Crosstabulation

			Voted in last election No	Voted in last election Yes	Total
Gender	Male	Count	2983	1941	4924
		% within Gender	60.6%	39.4%	100.0%
	Female	Count	2786	1739	4525
		% within Gender	61.6%	38.4%	100.0%
Total		Count	5769	3680	9449
		% within Gender	61.1%	38.9%	100.0%

Classification Table[a]

Observed			Predicted: Voted in last election No	Predicted: Voted in last election Yes	Percentage Correct
Step 1	Voted in last election	No	5769	0	100.0
		Yes	3680	0	.0
	Overall Percentage				61.1

a. The cut value is .500

Variables in the Equation

		B	S.E.	Wald	df	Sig.	Exp(B)	95% C.I.for EXP(B) Lower	95% C.I.for EXP(B) Upper
Step 1[a]	Gender(1)	-.042	.042	.968	1	.325	.959	.883	1.042
	Constant	-.430	.029	217.142	1	<.001	.651		

a. Variable(s) entered on step 1: Gender.

No, there was no significant relationship between sex and voting outcome.

#6) Can past voting outcome be predicted by sex, age category, and county?

Age category * Voted in last election Crosstabulation

			Voted in last election No	Voted in last election Yes	Total
Age category	18-30	Count	1155	524	1679
		% within Age category	68.8%	31.2%	100.0%
	31-45	Count	1769	1044	2813
		% within Age category	62.9%	37.1%	100.0%
	46-60	Count	1665	1092	2757
		% within Age category	60.4%	39.6%	100.0%
	>60	Count	1180	1020	2200
		% within Age category	53.6%	46.4%	100.0%
Total		Count	5769	3680	9449
		% within Age category	61.1%	38.9%	100.0%

County * Voted in last election Crosstabulation

			Voted in last election No	Voted in last election Yes	Total
County	Eastern	Count	1375	984	2359
		% within County	58.3%	41.7%	100.0%
	Central	Count	1030	616	1646
		% within County	62.6%	37.4%	100.0%
	Western	Count	1000	524	1524
		% within County	65.6%	34.4%	100.0%
	Northern	Count	1412	961	2373
		% within County	59.5%	40.5%	100.0%
	Southern	Count	952	595	1547
		% within County	61.5%	38.5%	100.0%
Total		Count	5769	3680	9449
		% within County	61.1%	38.9%	100.0%

Observed			Voted in last election No	Voted in last election Yes	Percentage Correct
Step 1	Voted in last election	No	5769	0	100.0
		Yes	3680	0	.0
	Overall Percentage				61.1

a. The cut value is .500

Variables in the Equation

		B	S.E.	Wald	df	Sig.	Exp(B)	95% C.I.for EXP(B) Lower	95% C.I.for EXP(B) Upper
Step 1[a]	Gender(1)	-.038	.043	.777	1	.378	.963	.886	1.047
	Age category			99.015	3	<.001			
	Age category(1)	.275	.066	17.355	1	<.001	1.317	1.157	1.499
	Age category(2)	.389	.067	34.248	1	<.001	1.476	1.295	1.681
	Age category(3)	.668	.069	92.656	1	<.001	1.950	1.702	2.234
	County			26.784	4	<.001			
	County(1)	-.192	.066	8.385	1	.004	.825	.725	.940
	County(2)	-.318	.069	21.545	1	<.001	.727	.636	.832
	County(3)	-.079	.060	1.746	1	.186	.924	.822	1.039
	County(4)	-.031	.068	.211	1	.646	.969	.847	1.108
	Constant	-.679	.069	95.777	1	<.001	.507		

a. Variable(s) entered on step 1: Gender, Age category, County.

Yes, voting outcome was significantly predicted by age category and county, but not by sex. Members in older age categories were significantly more likely to have voted in the last election than the youngest age category. Members in the Central and Western counties were significantly less likely to have voted than the Eastern county (Northern and Southern not different).

DaCCoTA
DAKOTA CANCER COLLABORATIVE
ON TRANSLATIONAL ACTIVITY

UNIVERSITY OF NORTH DAKOTA

Acknowledgements

References

Title image:

- https://commons.wikimedia.org/wiki/File:Spiessgasse_Frundsberger_Kriegsbuch_Jost_Ammann_1525.JPG

Selected Resources:

- https://www.ibm.com/docs/en/spss-statistics/28.0.0?topic=tutorial-sample-files#data_files
- https://www.ibm.com/support/pages/how-specify-anova-model-spss-nested-repeated-measures-factors
- https://www.spss-tutorials.com/spss-repeated-measures-anova/
- https://courses.pbsci.ucsc.edu/eeb/bioe286/Lecture%20Handouts/nested%20%20and%20%20repeated%20measures%20anova%202016.pdf
- https://statistics.laerd.com/statistical-guides/repeated-measures-anova-statistical-guide-3.php
- https://www.spss-tutorials.com/spss-repeated-measures-anova/#run

DaCCoTA

The DaCCoTA is supported by the National Institute of General Medical Sciences of the National Institutes of Health under Award Number U54GM128729. For projects that use the Biostatistics, Epidemiology, and Research Design Core in any way, including this presentation, please acknowledge us for publications. ***"Research reported in this publication was supported by DaCCoTA (the National Institute of General Medical Sciences of the National Institutes of Health under Award Number U54GM128729)".***

DaCCoTA
DAKOTA CANCER COLLABORATIVE
ON TRANSLATIONAL ACTIVITY