Comments for MEDB 5501, Week 7

Two-sample t-test (Independent-samples t-test)

Speaker notes

Things will start to get really interesting now. The two sample t-test, also known as the independent samples t-test, is important in its own right, but also serves as a foundation for other methodologies, especially analysis of variance and analysis of covariance.

This test arises in a variety of settings but two of the most common are the randomized trial and the cohort design.

Assumptions, 1 of 3

- Group 1, 2 both normally distributed
  - Assessed with histograms, boxplots, Q-Q plots
    - Assess each group separately, or
    - Combine after subtracting means
  - Or rely on Central Limit Theorem

Speaker notes

You can assess normality in each group separately. If you combine the two groups, however, you might see a bimodal pattern if the means differ. It is better to center the data: subtract the group mean from each value in the group. Then you can combine the two groups together.

How concerned you should be about the normality assumption depends on several factors. First is the sample size. With larger sample sizes, you can rely on the Central Limit Theorem. Second is the degree of deviation from normality. A slight skewness, for example, is not that serious. Third, the type of deviation from normality makes a difference. Distributions that are highly skewed or heavy tailed are more of a concern than light tailed distributions.

Assumptions, 2 of 3

- Possibly different means, but same variance
  - Compare the box part of the box plots
    - Interquartile range
    - Look for large disparities only (2 or 3 fold)
  - Calculate and compare the standard deviations
    - Again, large disparities only
  - Levene's test (not recommended)
    - Too little power for small sample sizes and too much power for large sample sizes
    - Very sensitive to normality assumption
- The equal variance assumption is important mostly for unequal sample sizes.

Speaker notes

An important assumption is equality of variances.

Assumptions, 3 of 3

- Observations are independent
  - Between groups
    - No matching
    - No longitudinal measures
  - Within groups
    - No cluster effects
    - No infectious spread
  - Assessed qualitatively

Speaker notes

Independence is almost always assessed qualitatively. You have to assume indpendence between the two groups first. Data that is matched or paired violates this assumption (see the paired t-test later). You also can't use the same subjects again. If you want to compare the same subjects at different time points that would be a violation of independeance.

You also have to assume independence within groups. I was doing a study of breast feeding where the researchers recruited 84 infants. There were 72 singleton births and 6 sets of twins. I bet the researchers were excited to get these twins: two observations for the price of a single consent form. But the measurement of breast feeding duration is not going to be independent. If one baby in the twin pair stops breast feeding, it increases the chances that the other baby will stop at the same time. It doesn't happen all the time, but it happens enough that you can't say that the data are independent.

Any time you have a clustering effect, measurements in the same family, for example, you should probably assume that the outcomes are not independent. If you conduct studies at multiple clinics, the observations of two patients at the same clinic are not going to be independent because the standard of care, skill of the medical personnel, and patient referral patterns are going to influence the results.

It's okay if only a single clinic evaluates all of the patients. It's only when there are multiple clinics that this may become an issue.

If you have clusters in your data, you can still analyze the data, but you have to account for the clusters in the model. Otherwise, the confidence intervals and hypothesis tests will be distorted.

Housing data dictionary, 1 of 5

```
source:
  This file was found originally at a website
  DASL (Data And Story Library) that is no
  longer available.

description:
  The original source describes the data as
  "a random sample of records of resales of
  homes from Feb 15 to Apr 30, 1993 from the
  files maintained by the Albuquerque Board
  of Realtors. This type of data is
  collected by multiple listing agencies in
  many cities and is used by realtors as an
  information base."
```

Speaker notes

I’ve used this data in a lot of my classes, but it is no longer available on the web. There are lots of other similar data sets on housing resales.

Sales from 1993 are now three decades old, so some of the details may seem out of date.

Housing data dictionary, 2 of 5

```
copyright:
    Unknown. You should be able to use this data for
  individual educational purposes under the Fair Use
  guidelines of U.S. copyright law.

format:
  delimiter: space
  varnames: first row of data
  missing-value-code: *
  rows: 117
  columns: 8
```

Speaker notes

This data uses an asterisk to denote missing values. This will cause a bit of a hassle, but not too much during data import.

Housing data dictionary, 3 of 5

```
vars:
  Price:
    label: Selling price
    unit: dollars

  SquareFeet:
    label: Living space
    unit: square feet

  AgeYears:
    label: Age of home
    unit: years
```

Speaker notes

Here are the first three variables in the data. The key variable in all my analyses would be the selling price. This might be influenced by the size of the home and/or how old the home is.

Housing data dictionary, 4 of 5

```
NumberFeatures:
  label:
    Home features (dishwasher, refrigerator,
    microwave, disposer, washer, intercom,
    skylight(s), compactor, dryer, handicap
    fit, cable TV access)
  scale: count
  range: 0 to 11

Northeast:
  label: Located in northeast sector of city?
  values:
    Yes: 1
    No: 0
```

Speaker notes

This data is from 1993, and some of the features that were optional then are pretty much standard now. Location may or may not be important, and the data has an indicator of whether the house is located in the northeast sector of the city.

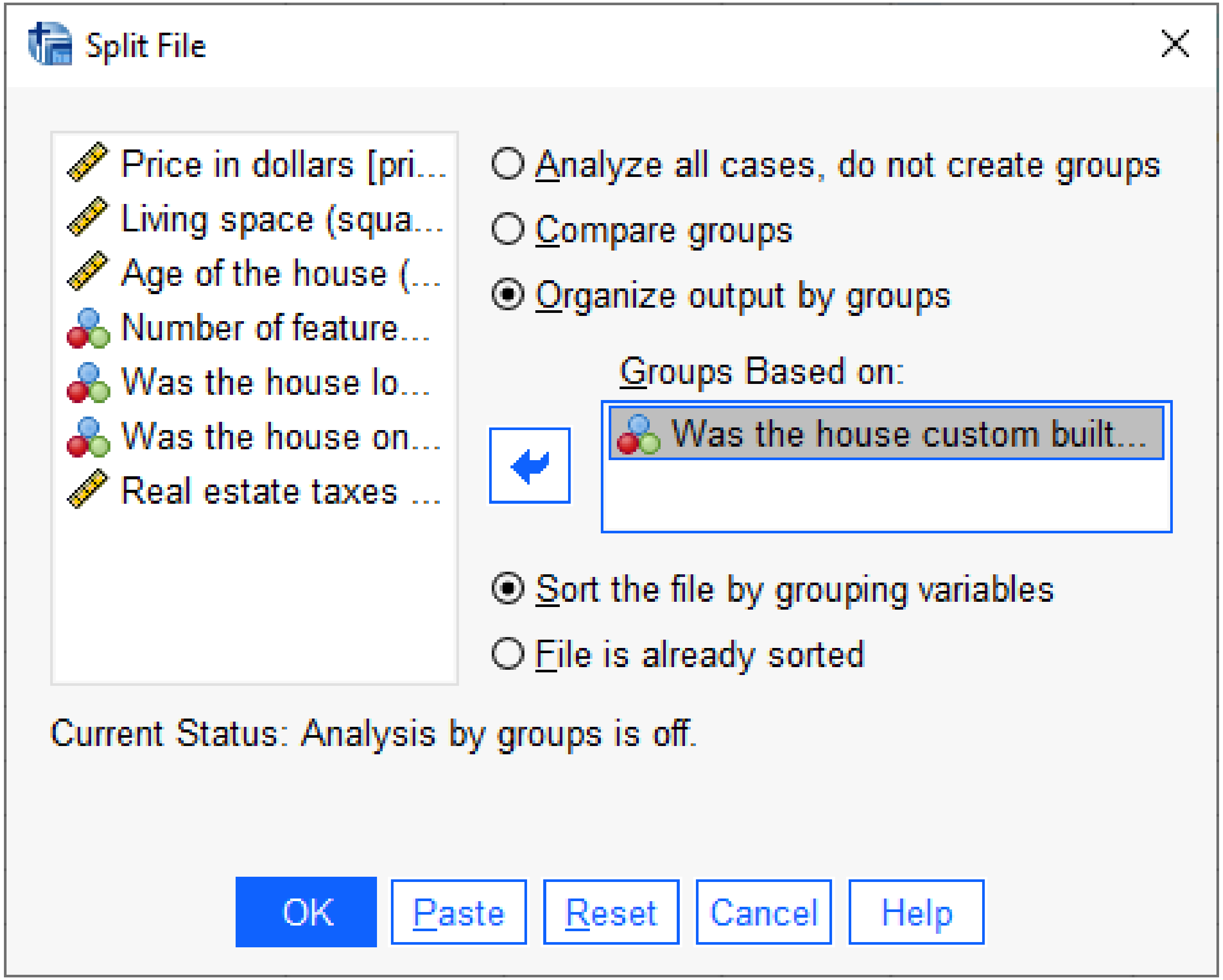
Housing data dictionary, 5 of 5

```
CustomBuild:
  label: Custom built?
  values:
    Yes: 1
    No: 0

CornerLot:
  label: Corner location?
  values:
    Yes: 1
    No: 0

Tax:
  label: Yearly property tax
  unit: dollars
```

Speaker notes

Here is information on the final three variables. Some houses are built to a pre-existing plan, but others are custom designed with an architect. Houses on a corner lot have some disadvantages, such as noise and less privacy. Finally, the yearly tax bill should be correlated with the price, though sometimes the government valuation of a property is out of date.

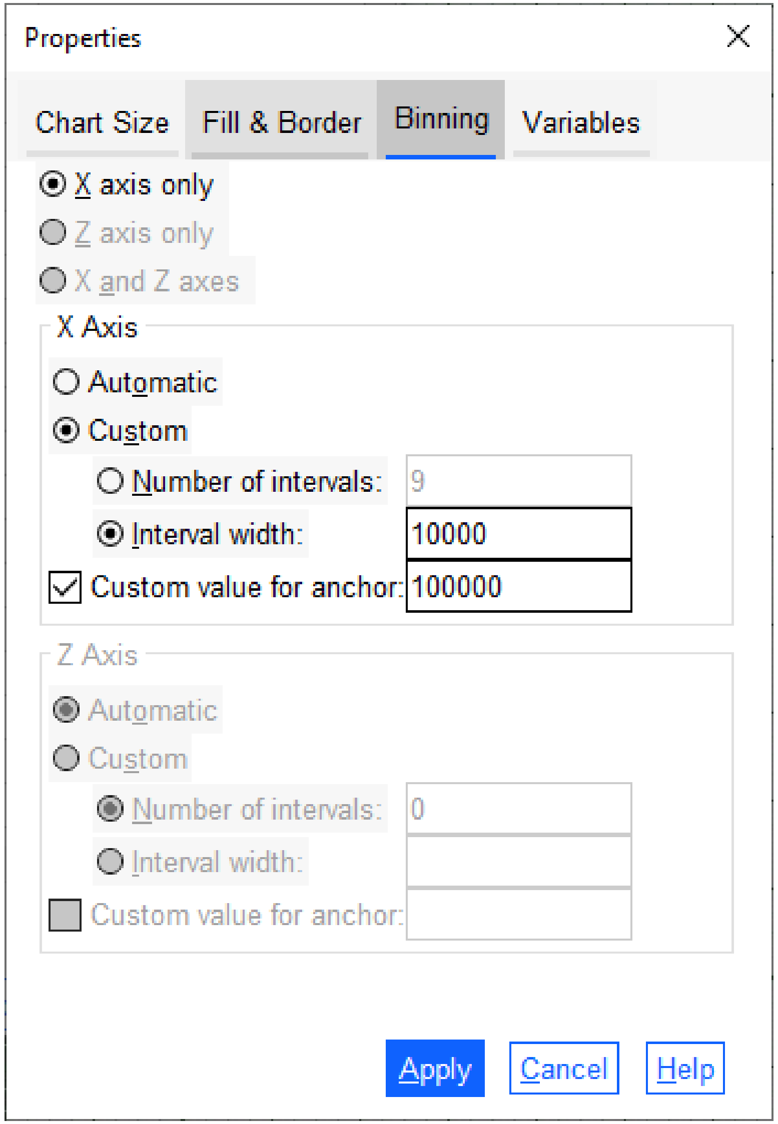
Split file dialog box



Speaker notes

Add note.

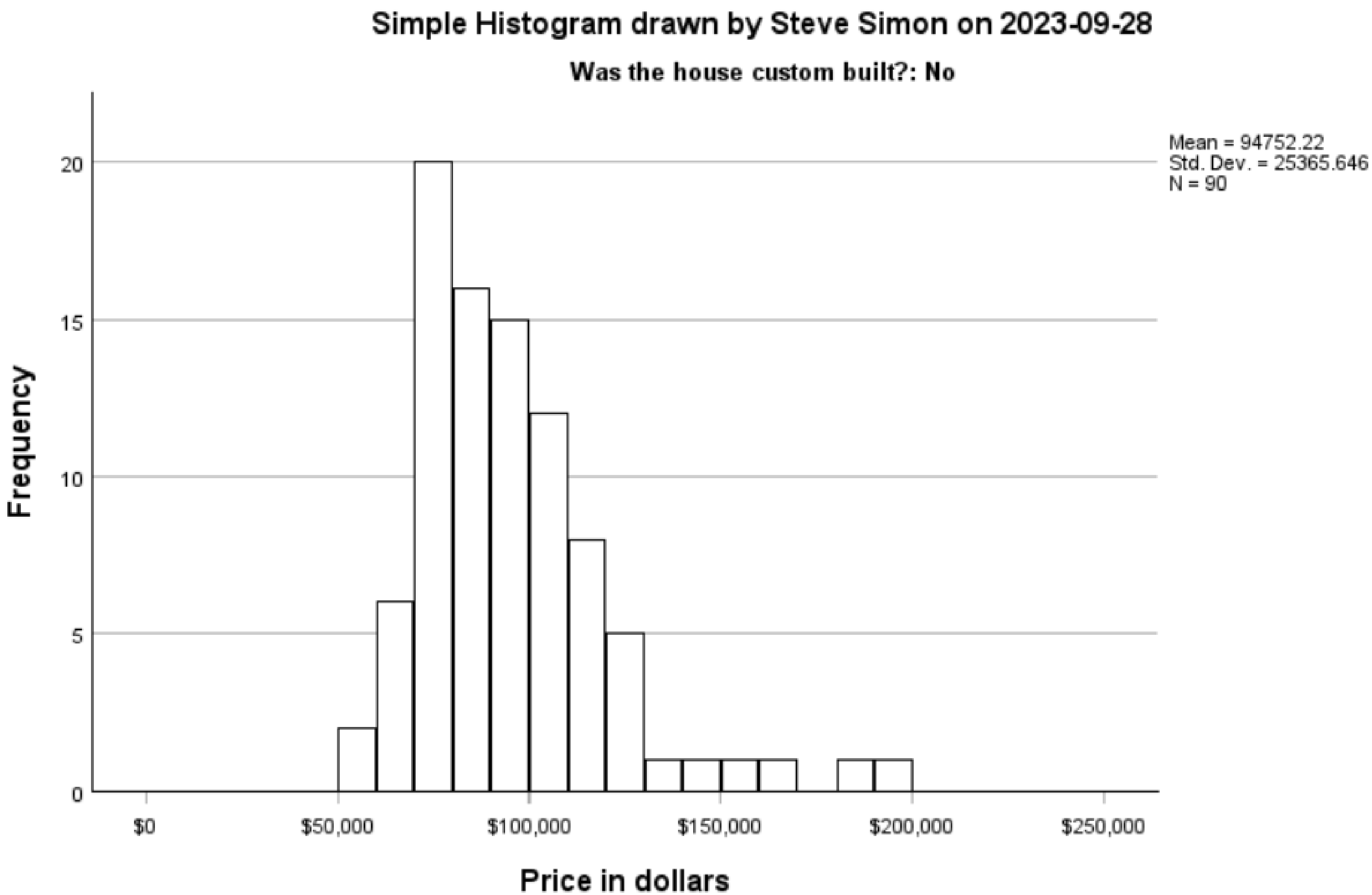
Housing analysis



Speaker notes

To get separate histograms for each subgroup, select Data | Split File from the SPSS menu.

Price histogram, 1 of 2

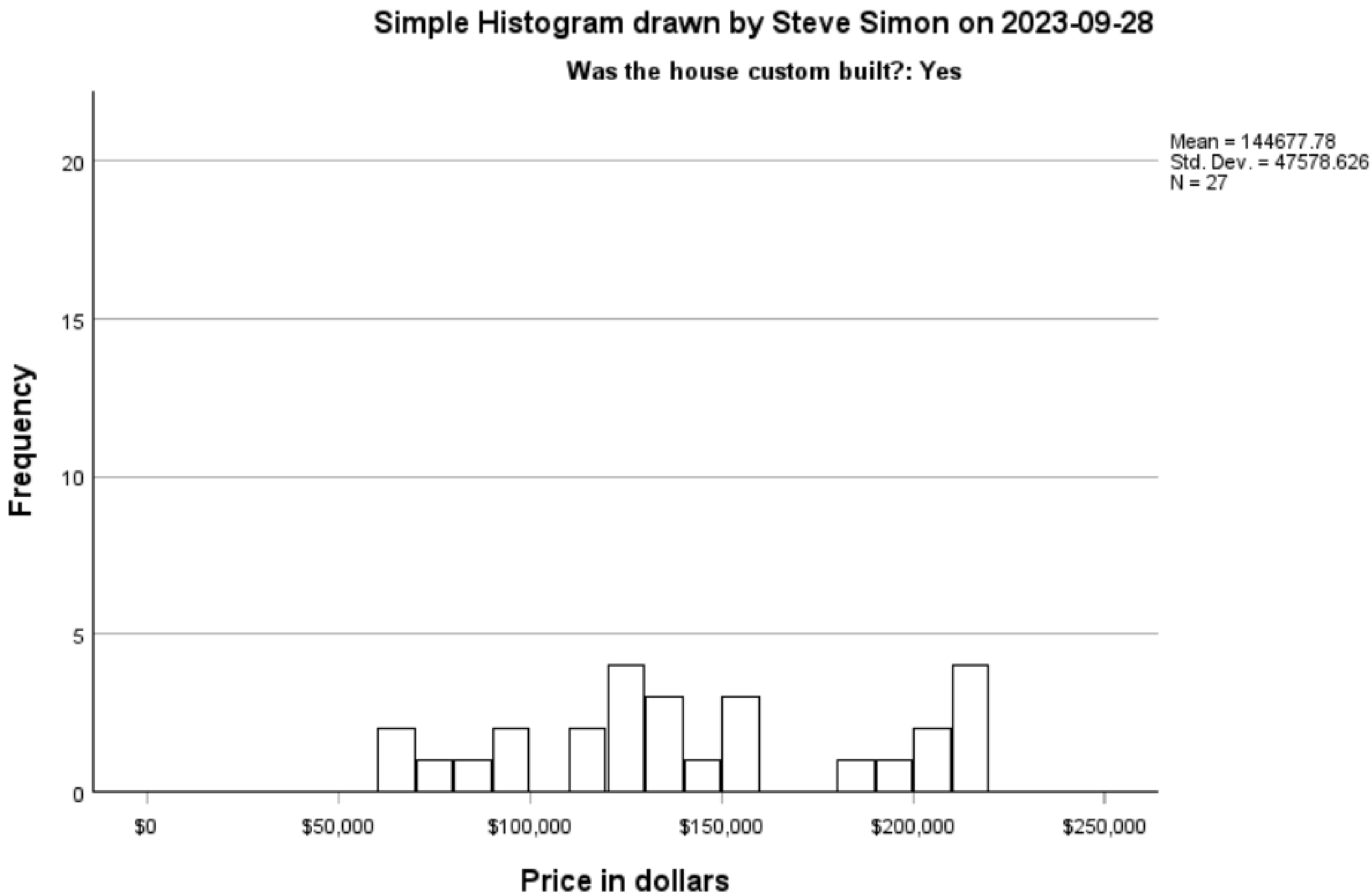


Speaker notes

Here is the histogram for the first group, not custom built. Note a few things.

1. Removal of the needless blue color
2. Change frequency to remove unneeded decimals
3. Use of comma and dollar signs for price.

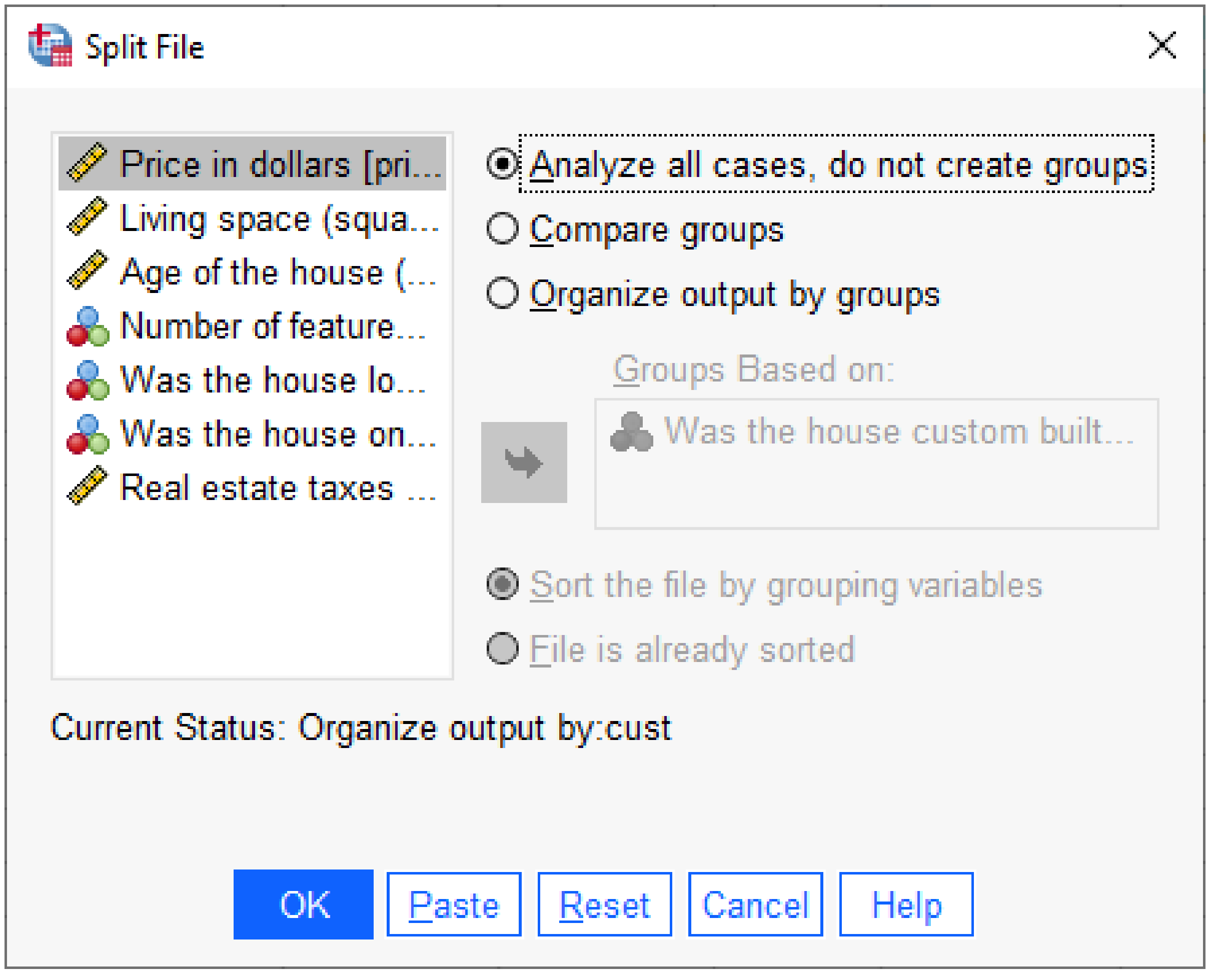
Price histogram, 2 of 2



Speaker notes

Here is the histogram for custom built=Yes. In addition to the changes mentioned on the previous slide, you should change the scaling on the X and Y axes so the two histograms are directly comparable.

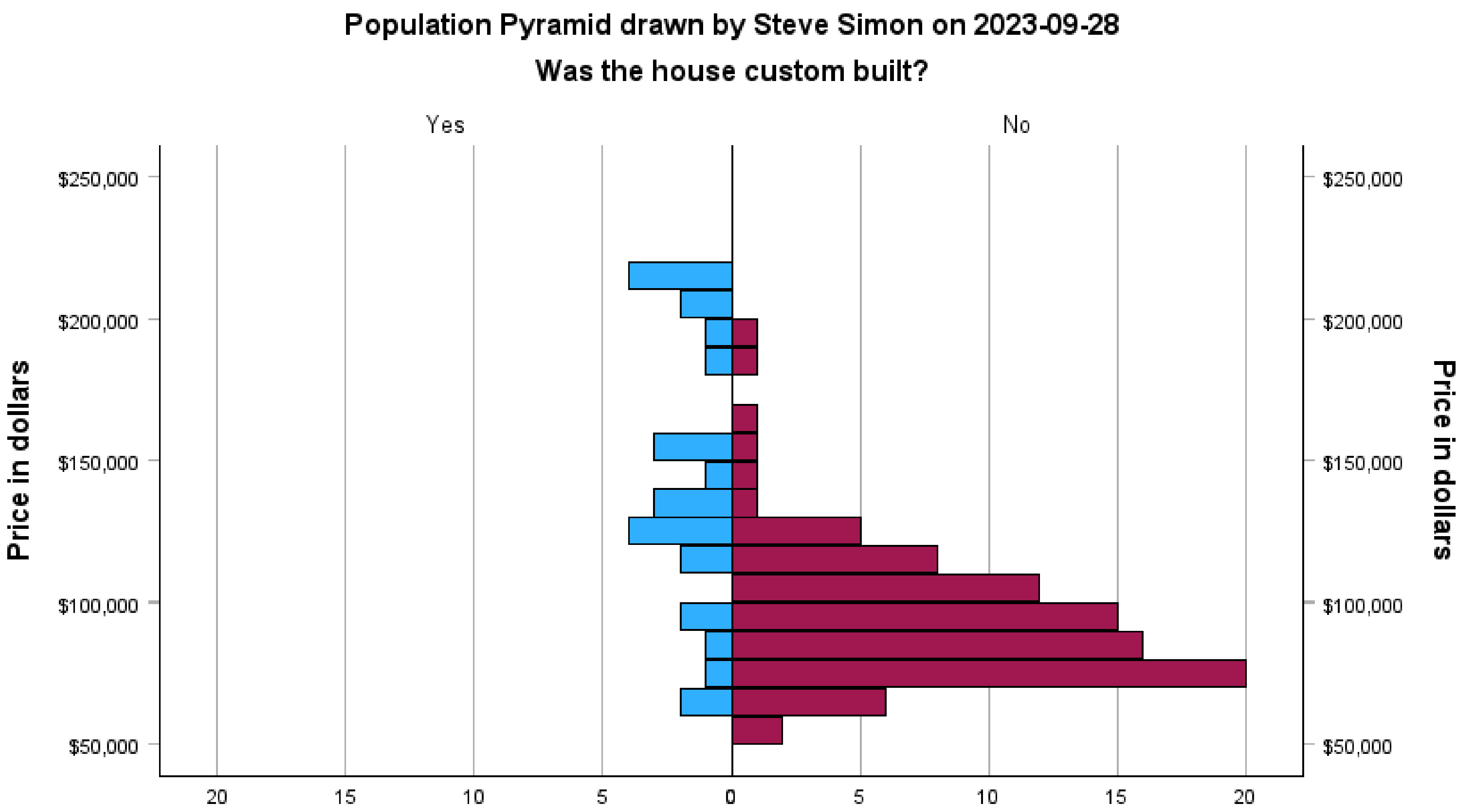
Stop grouping, using the Split file dialog box



Speaker notes

I always forget, but you should remember to turn off the split file option before doing any further analyses.

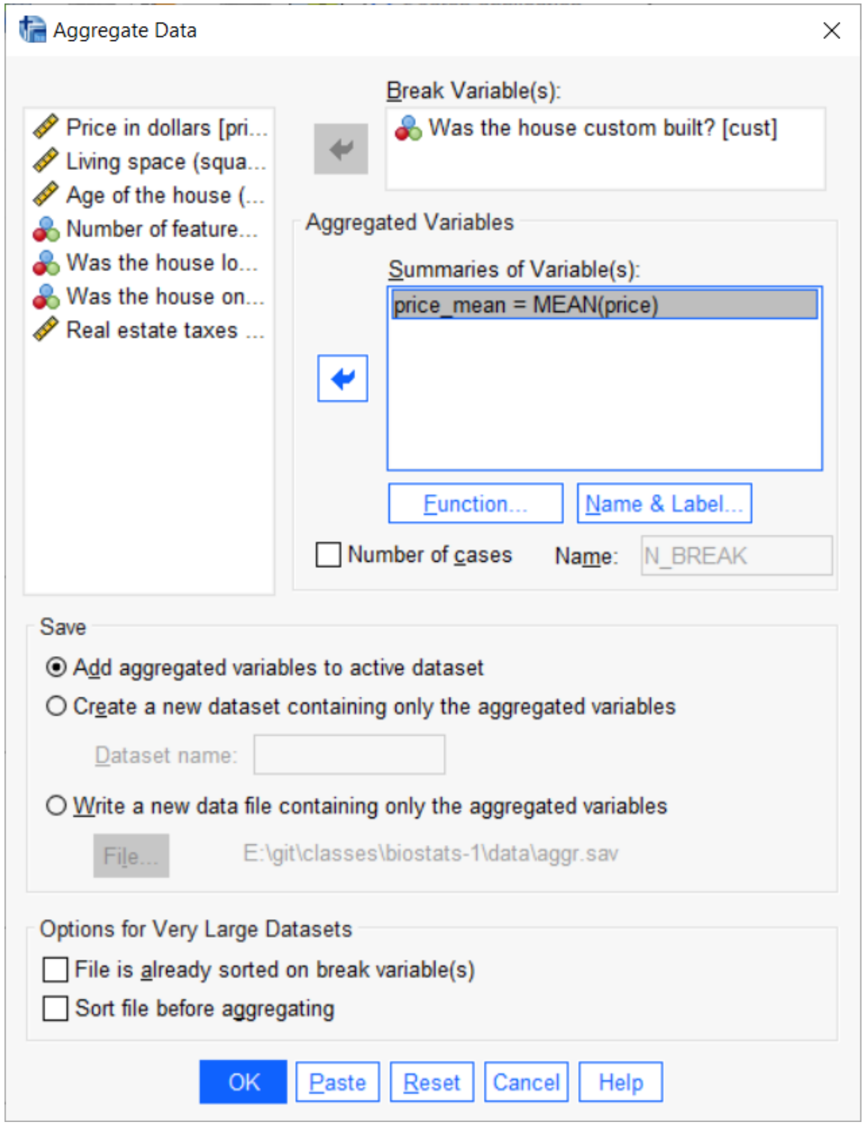
Population pyramid



Speaker notes

I’m not a big fan of the population pyramid, but it does pretty well right here. It has the advantage of not needing to use the split file command and it scales the X and y axes to a common appearance.

Agggregate data dialog box

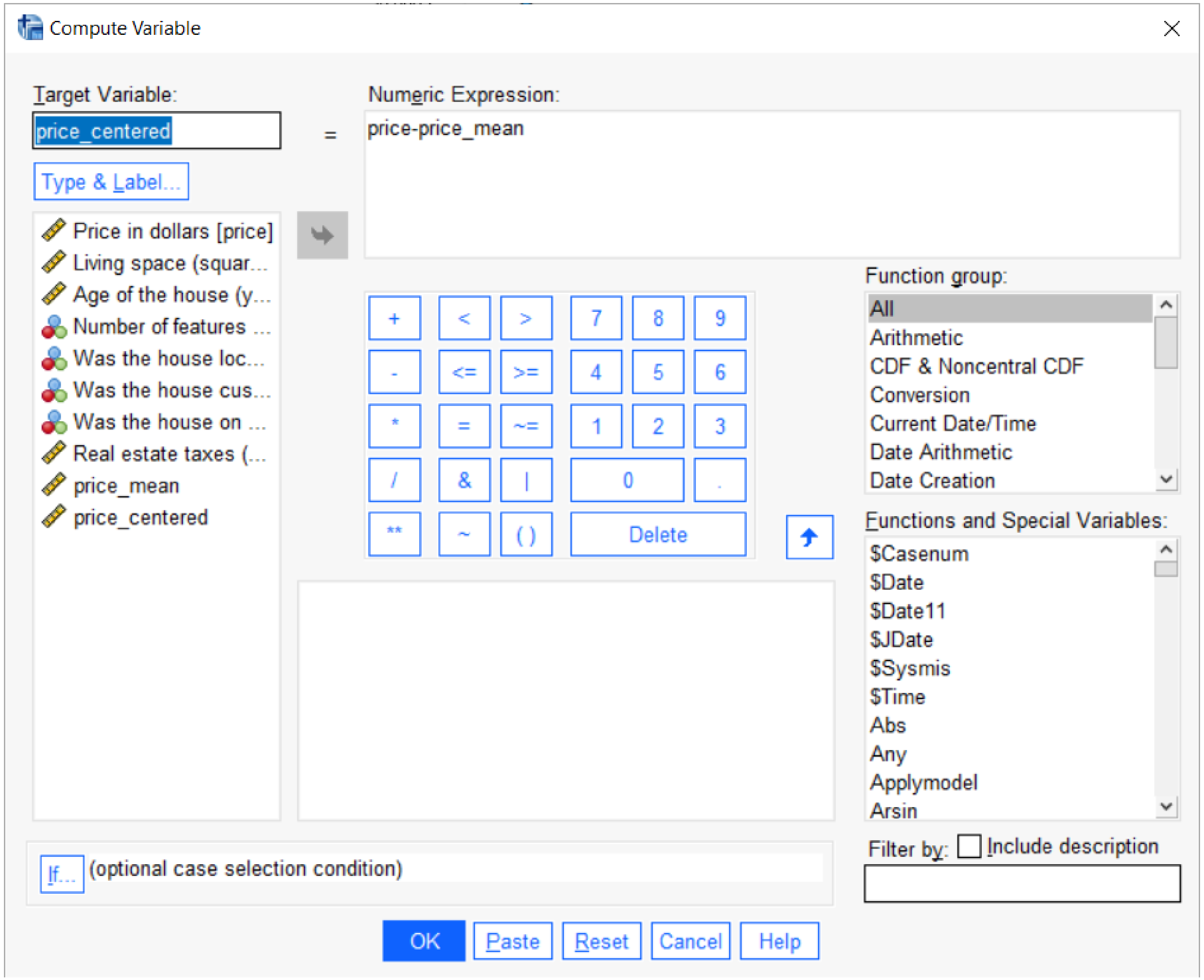


Speaker notes

If you want to draw a Q-Q plot, you could do it separately for each group, but I recommend that you combine them. Before you combine them, though, be sure to adjust the means of the two groups so they are both the same. This is a multi-step process.

First select Data | Aggregate from the menu.

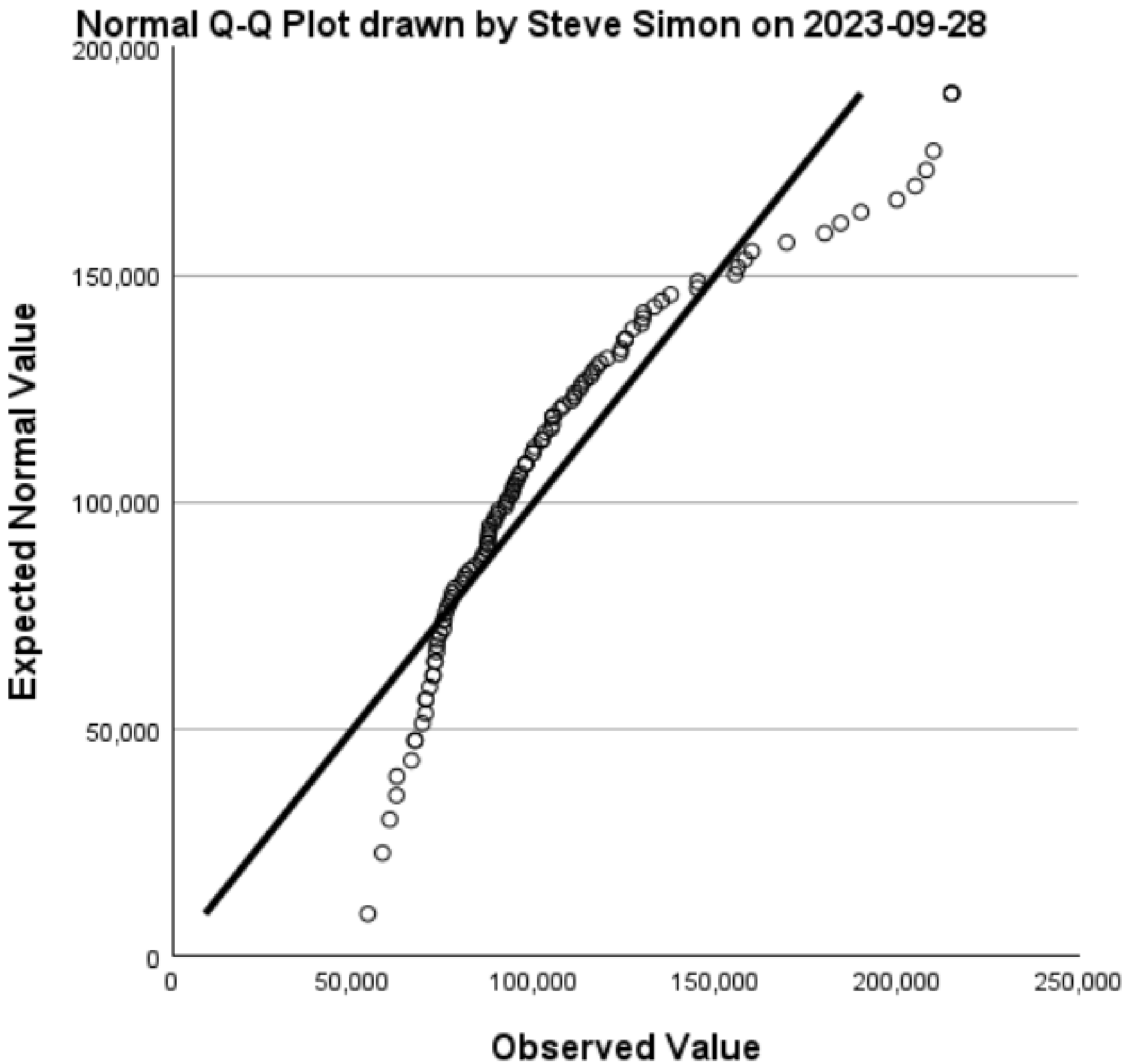
Compute variable dialog box



Speaker notes

Then subtract the means from each group using Transform | Compute from the SPSS menu.

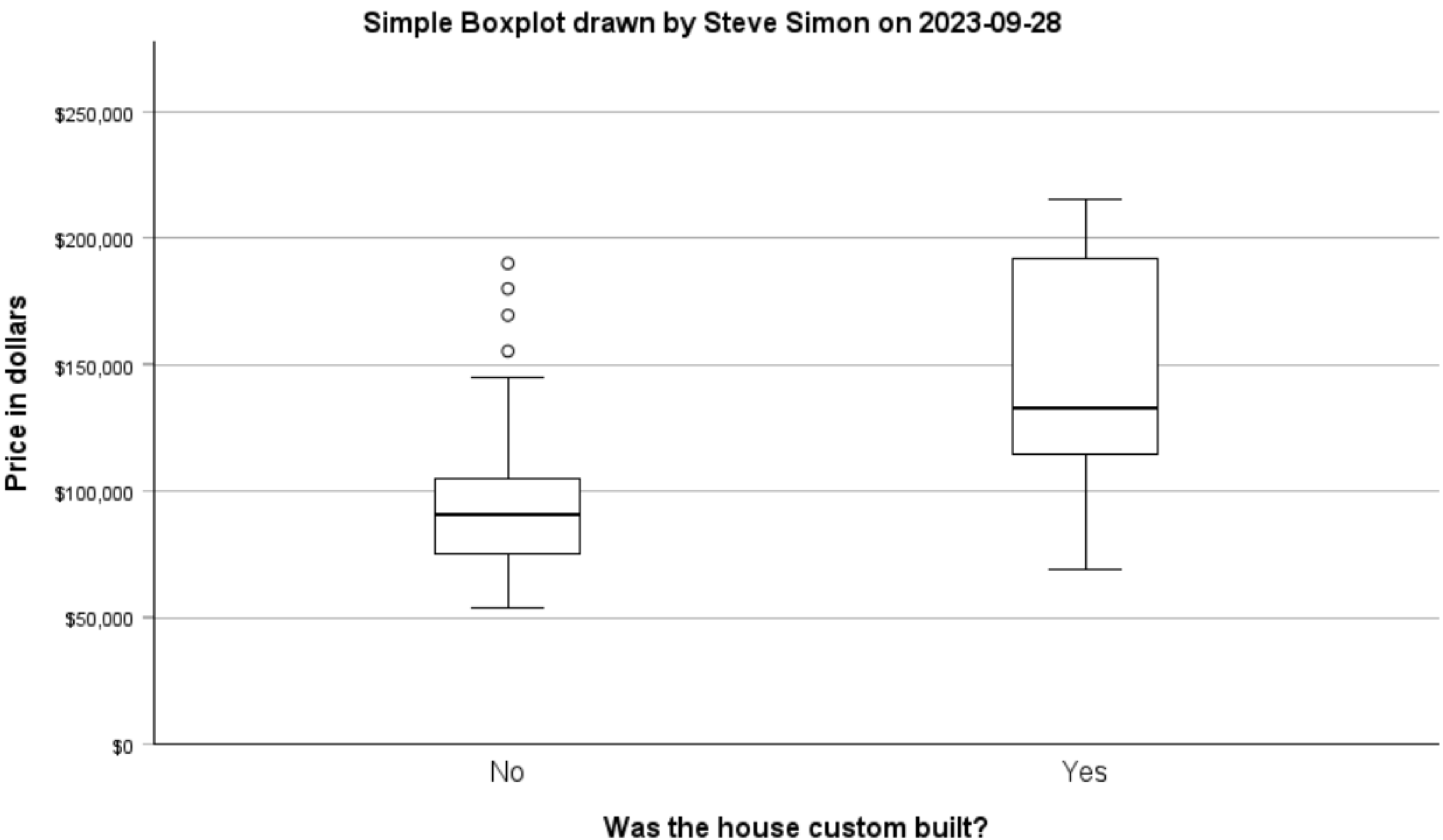
Q-Q plot for all of the data



Speaker notes

Add note.

Boxplots



Simple Boxplot drawn by Steve Simon on 2023-09-28
Price in dollars
$250,000
$200,000
$150,000
$100,000
$50,000
$0
No
Yes
Was the house custom built?

Speaker notes

The boxplot is very useful for examining normality and the equal variation assumption. For housing prices, it looks like the custom built houses have both a larger mean and a larger standard deviation. I generally worry a little bit about the equal variances assumption if the one interquartile range is at least twice as big than the other, and worry more when it is three times bigger.

Here is it a bit of a concern. Normality, though, is probably close enough to not be a concern.

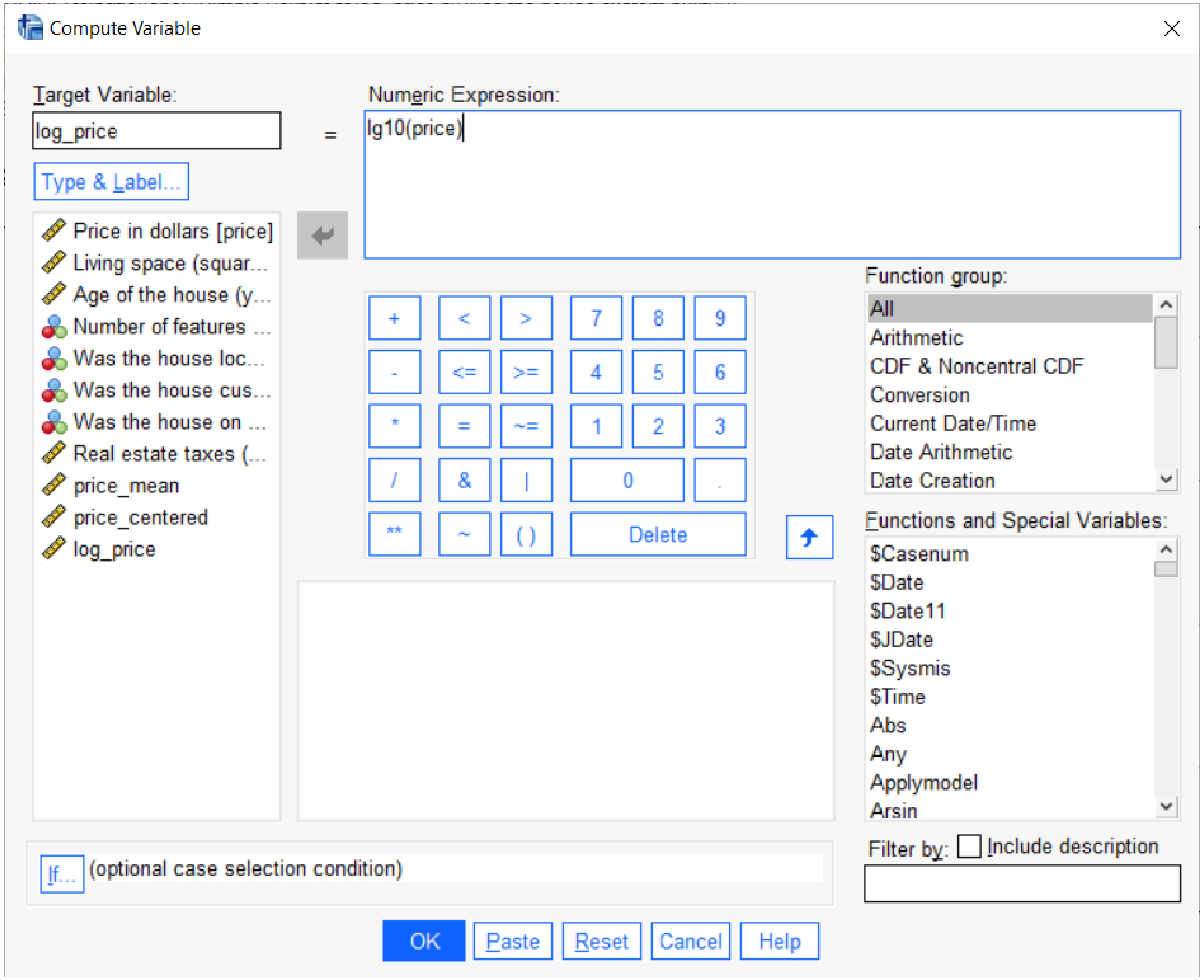
Should you consider a log transformation?

- Yes because
  - Data bounded below by zero
  - Some evidence of non-normality
  - Some evidence of unequal variation.
    - Unbalanced sample sizes
- No because
  - Some evidence of normality
    - Sample size is large
  - Differences in variation not too extreme

Speaker notes

The evidence is mixed but points to the possible use of a log transformation.

Compute Variable dialog box



Speaker notes

Select Transform | Compute from the SPSS menu to get a log transformation. Remember that SPSS uses “lg10” for the log transformation.

boxplots of the log transformed variables

Speaker notes

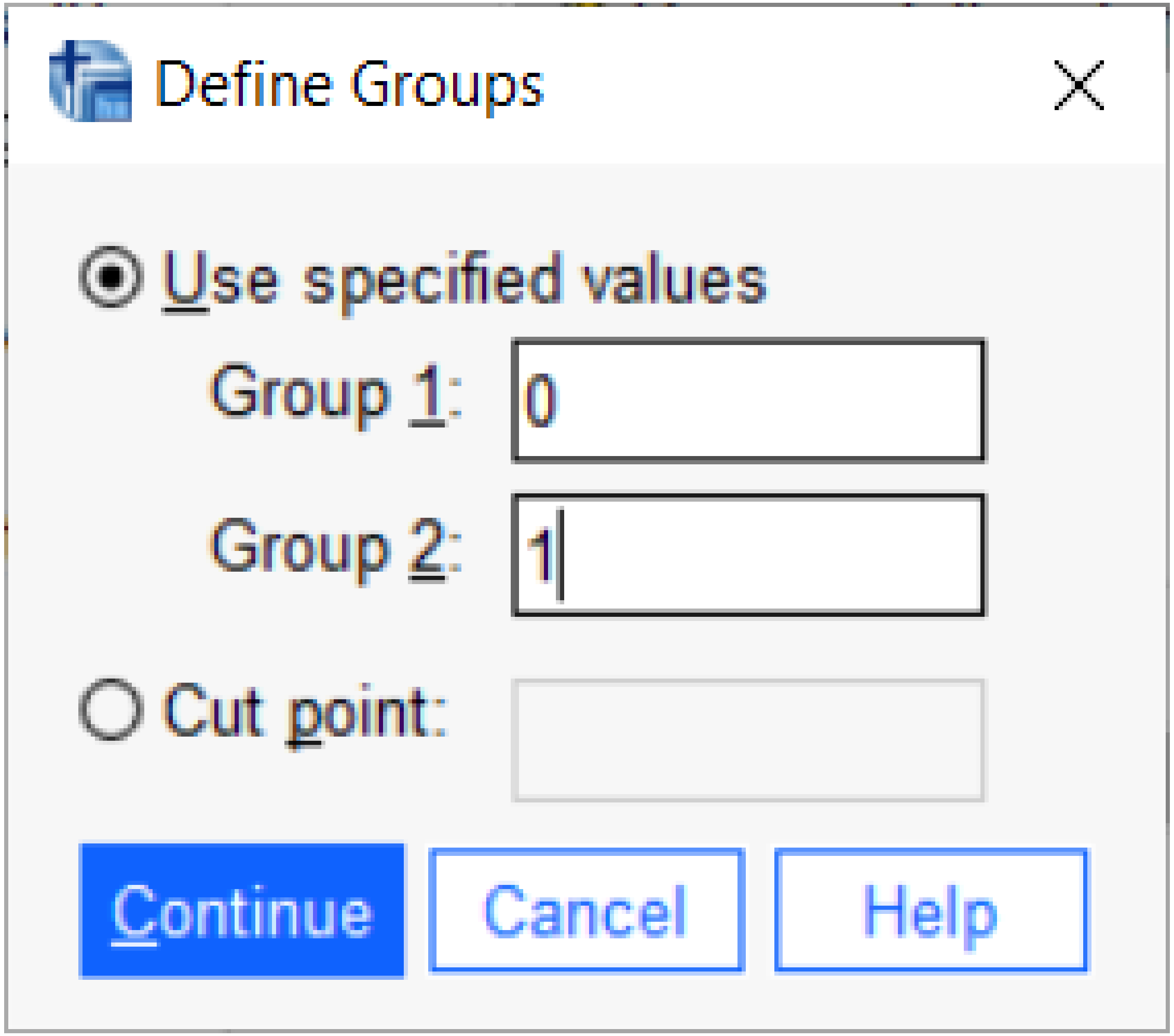
Here is the boxplot for the log transformed data.

Independent Samples T-Test dialog box

Speaker notes

Select Analyze | Compare Means | Independent Samples T-test from the SPSS menu.

Define Groups dialog box



Speaker notes

SPSS does not automatically determine the group codes from your data.

T-test output, 1 of 5

Group Statistics

	Was the house custom built?	N	Mean	Std. Deviation	Std. Error Mean
Price in dollars	No	90	94752.22	25365.646	2673.774
	Yes	27	144677.78	47578.626	9156.511

Speaker notes

SPSS produces way too much information in its output and you can’t really suppress most of it. The groups statistics are not bad. The standard error is not really that helpful, to be honest, but that is a nitpick.

T-test output, 2 of 5

		Levene's Test for Equality of Variances	
		F	Sig.
Price in dollars	Equal variances assumed	24.800	<.001
	Equal variances not assumed		

Speaker notes

I do not recommend Levene’s test, for reasons mentioned earlier. There is no easy way to keep this test from appearing in your output.

T-test output, 3 of 5

t-test for Equality of Means					
		Significance			
t	df	One-Sided p	Two-Sided p	Mean Difference	Std. Error Difference
-7.160	115	<.001	<.001	-49925.556	6972.621
-5.234	30.558	<.001	<.001	-49925.556	9538.908

Speaker notes

SPSS provides two t-tests. The first row (Equal variances assumed) is the traditional t-test. The second row (Equal variances not assumed) is a modification of the t-test to account for unequal variances.

SPSS also provides one-sided and two-sided p-values.

T-test output, 4 of 5

95% Confidence Interval of the Difference	
Lower	Upper
-63736.976	-36114.135
-69391.695	-30459.416

Speaker notes

SPSS also provides two different confidence intervals.

T-test output, 5 of 5

Independent Samples Effect Sizes

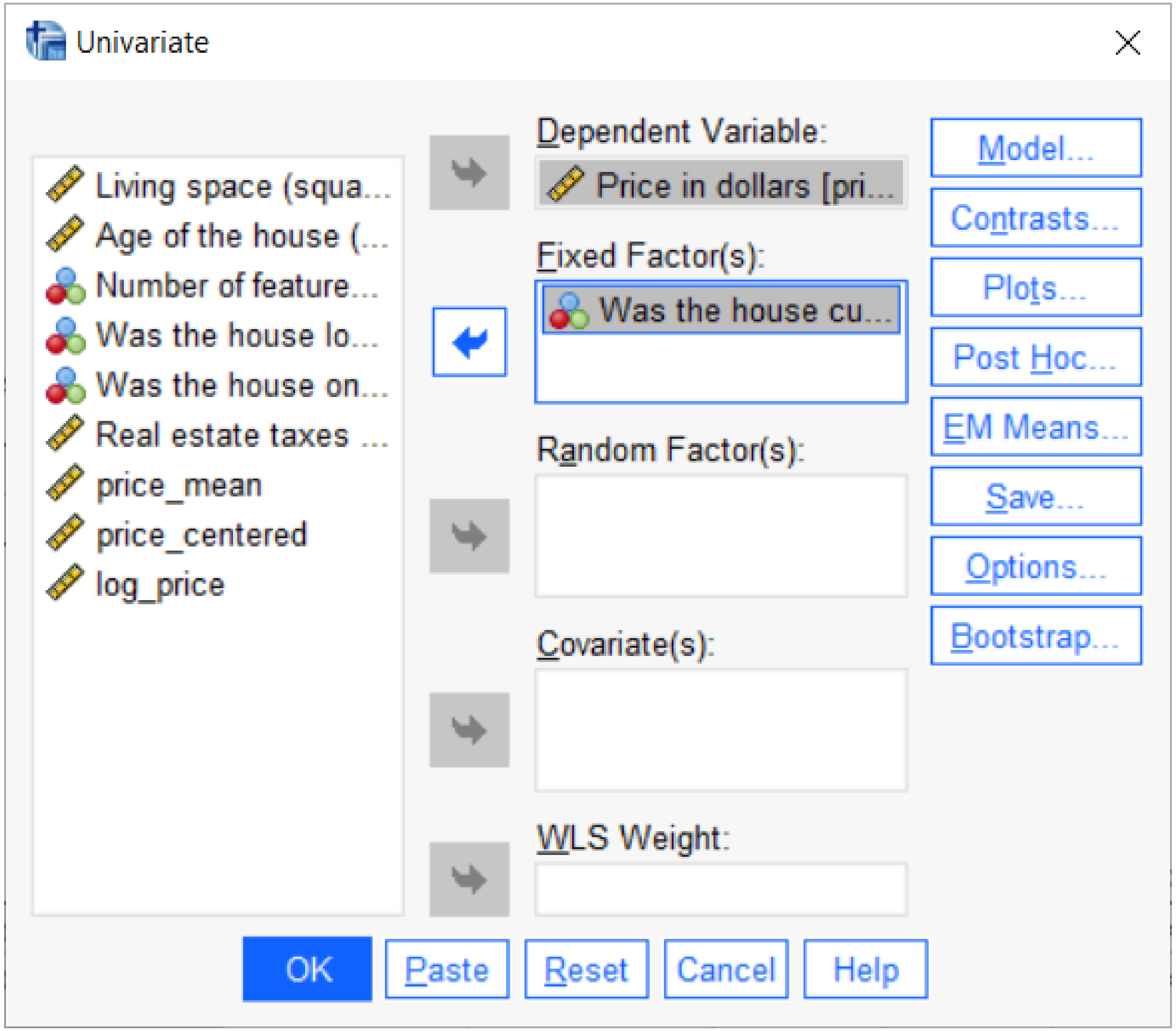
		Standardizer[a]	Point Estimate	95% Confidence Interval Lower	95% Confidence Interval Upper
Price in dollars	Cohen's d	31776.515	-1.571	-2.044	-1.093
	Hedges' correction	31985.647	-1.561	-2.030	-1.086
	Glass's delta	47578.626	-1.049	-1.557	-.527

a. The denominator used in estimating the effect sizes.
Cohen's d uses the pooled standard deviation.
Hedges' correction uses the pooled standard deviation, plus a correction factor.
Glass's delta uses the sample standard deviation of the control (i.e., the second) group.

Speaker notes

Finally, SPSS provides effect size measures. I do not like effect sizes and there is no easy way to suppress this.

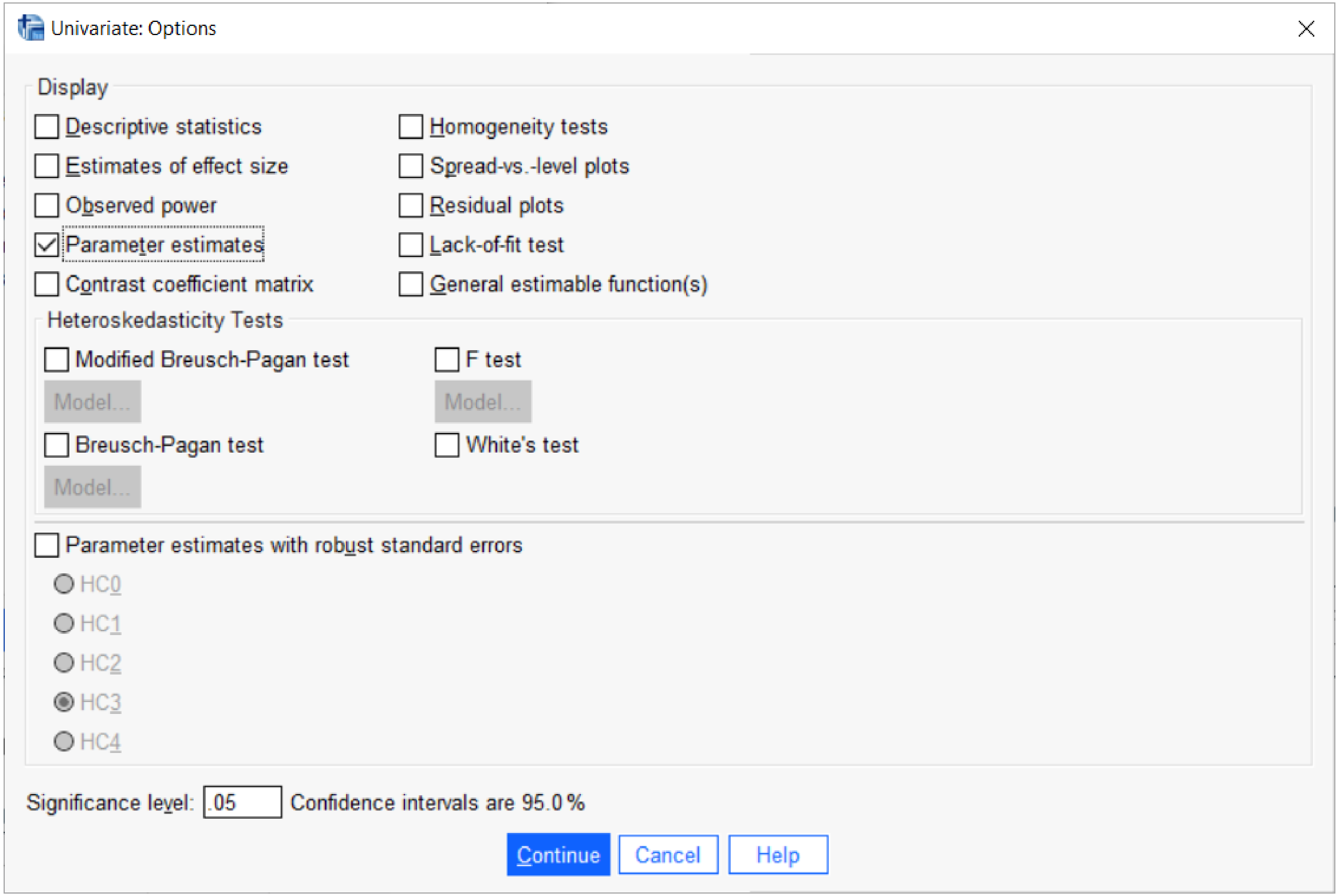
Univariate model dialog box



Speaker notes

I prefer to conduct a two-sample t-test using a regression model. Select Analyze | General Linear Model | Univariate from the menu.

Univariate model dialog box



Speaker notes

Add note.

General linear model output, 1 of 3

Between-Subjects Factors

		Value Label	N
Was the house custom built?	0	No	90
	1	Yes	27

Speaker notes

Add note.

General linear model output, 2 of 3

Tests of Between-Subjects Effects

Dependent Variable: Price in dollars

Source	Type III Sum of Squares	df	Mean Square	F	Sig.
Corrected Model	51768576641[a]	1	51768576641	51.269	<.001
Intercept	1.191E+12	1	1.191E+12	1179.139	<.001
cust	51768576641	1	51768576641	51.269	<.001
Error	1.161E+11	115	1009746880.2		
Total	1.489E+12	117			
Corrected Total	1.679E+11	116			

a. R Squared = .308 (Adjusted R Squared = .302)

Speaker notes

Add note.

General linear model output, 3 of 3

Parameter Estimates

Dependent Variable: Price in dollars

Parameter	B	Std. Error	t	Sig.	95% Confidence Interval Lower Bound	95% Confidence Interval Upper Bound
Intercept	144677.778	6115.393	23.658	<.001	132564.361	156791.195
[cust=0]	-49925.556	6972.621	-7.160	<.001	-63736.976	-36114.135
[cust=1]	0[a]	.	.	.	.	.

a. This parameter is set to zero because it is redundant.

Speaker notes

Add note.

What belongs in an interpretation

"We designed the study to have 90% power to detect a 4 degree difference between the groups in the increased range of elbow flexion. Alpha was set at 0.05. Patients receiving electrical stimulation (n=26) increased their range of elbow flexion by a mean of 16 degrees with a standard deviation of 4.5, whereas patients in the control group (n=25) increased their range of flexion by a mean of only 6.5 degrees with a standard deviation of 3.4. This 9.5 degree difference was statistically significant (95% CI = 7.23 to 11.73, two-tailed Student's t test, t=8.43, P < 0.001)."

- Lang and Secic, How to Report Statistics in Medicine, p47.

Speaker notes

I have not stressed interpretation enough. Here is the interpretation found word for word in an excellent book by Lang and Secic.

Let me parse this word by word.

Power calculation, if possible

“**We designed the study to have 90% power to detect a 4 degree difference between the groups in the increased range of elbow flexion.** Alpha was set at 0.05. Patients receiving electrical stimulation (n=26) increased their range of elbow flexion by a mean of 16 degrees with a standard deviation of 4.5, whereas patients in the control group (n=25) increased their range of flexion by a mean of only 6.5 degrees with a standard deviation of 3.4. This 9.5 degree difference was statistically significant (95% CI = 7.23 to 11.73, two-tailed Student’s t test, t=8.43, P < 0.001).”

- Lang and Secic, How to Report Statistics in Medicine, p47.

Speaker notes

You may not have a power calculation conducted prior to the start of the experiment. If you do, be sure to brag about it.

Alpha level

“We designed the study to have 90% power to detect a 4 degree difference between the groups in the increased range of elbow flexion. **Alpha was set at 0.05.** Patients receiving electrical stimulation (n=26) increased their range of elbow flexion by a mean of 16 degrees with a standard deviation of 4.5, whereas patients in the control group (n=25) increased their range of flexion by a mean of only 6.5 degrees with a standard deviation of 3.4. This 9.5 degree difference was statistically significant (95% CI = 7.23 to 11.73, two-tailed Student’s t test, t=8.43, P < 0.001).”

- Lang and Secic, How to Report Statistics in Medicine, p47.

Speaker notes

Although almost everyone uses an alpha level of 0.05, it never hurts to restate it.

Sample statistics for each group

“We designed the study to have 90% power to detect a 4 degree difference between the groups in the increased range of elbow flexion. Alpha was set at 0.05. **Patients receiving electrical stimulation (n=26) increased their range of elbow flexion by a mean of 16 degrees with a standard deviation of 4.5, whereas patients in the control group (n=25) increased their range of flexion by a mean of** *only* **6.5 degrees with a standard deviation of 3.4.** This 9.5 degree difference was statistically significant (95% CI = 7.23 to 11.73, two-tailed Student’s t test, t=8.43, P < 0.001).”

- Lang and Secic, How to Report Statistics in Medicine, p47.

Speaker notes

State the sample statistics for each group (mean, standard deviation, and range).

The use of comparative words like “only” are strongly encouraged.

P-value AND confidence interval

“We designed the study to have 90% power to detect a 4 degree difference between the groups in the increased range of elbow flexion. Alpha was set at 0.05. Patients receiving electrical stimulation (n=26) increased their range of elbow flexion by a mean of 16 degrees with a standard deviation of 4.5, whereas patients in the control group (n=25) increased their range of flexion by a mean of only 6.5 degrees with a standard deviation of 3.4. This 9.5 degree difference was statistically significant (**95% CI = 7.23 to 11.73**, two-tailed Student’s t test, t=8.43, **P < 0.001**).”

- Lang and Secic, How to Report Statistics in Medicine, p47.

Speaker notes

State the p-value and the confidence interval. Both are important. I do not agree with reporting the t-statistic, but that is a minor quibble.

One or two sided test

"We designed the study to have 90% power to detect a 4 degree difference between the groups in the increased range of elbow flexion. Alpha was set at 0.05. Patients receiving electrical stimulation (n=26) increased their range of elbow flexion by a mean of 16 degrees with a standard deviation of 4.5, whereas patients in the control group (n=25) increased their range of flexion by a mean of only 6.5 degrees with a standard deviation of 3.4. This 9.5 degree difference was statistically significant (95% CI = 7.23 to 11.73, **two-tailed Student's t test**, t=8.43, P < 0.001)."

- Lang and Secic, How to Report Statistics in Medicine, p47.

Speaker notes

State the p-value and the confidence interval. Both are important. I do not agree with reporting the t-statistic, but that is a minor quibble.

My interpretation

- “Comparison of housing prices between groups used a two-sample t-test.”
- “All tests were conducted using a two-sided alpha level of 0.05.”
- “Houses that were custom built had an average price of 145 thousand dollars, which was 50 thousand dollars higher than the average price of normal houses. This difference was statistically significant (95% CI 36 thousand dollars to 64 thousand dollars, p=0.001).”

Speaker notes

There was no power calculation done prior to data collection, so don’t mention it at all. I like to state some of the details in the methods section rather than the results section.

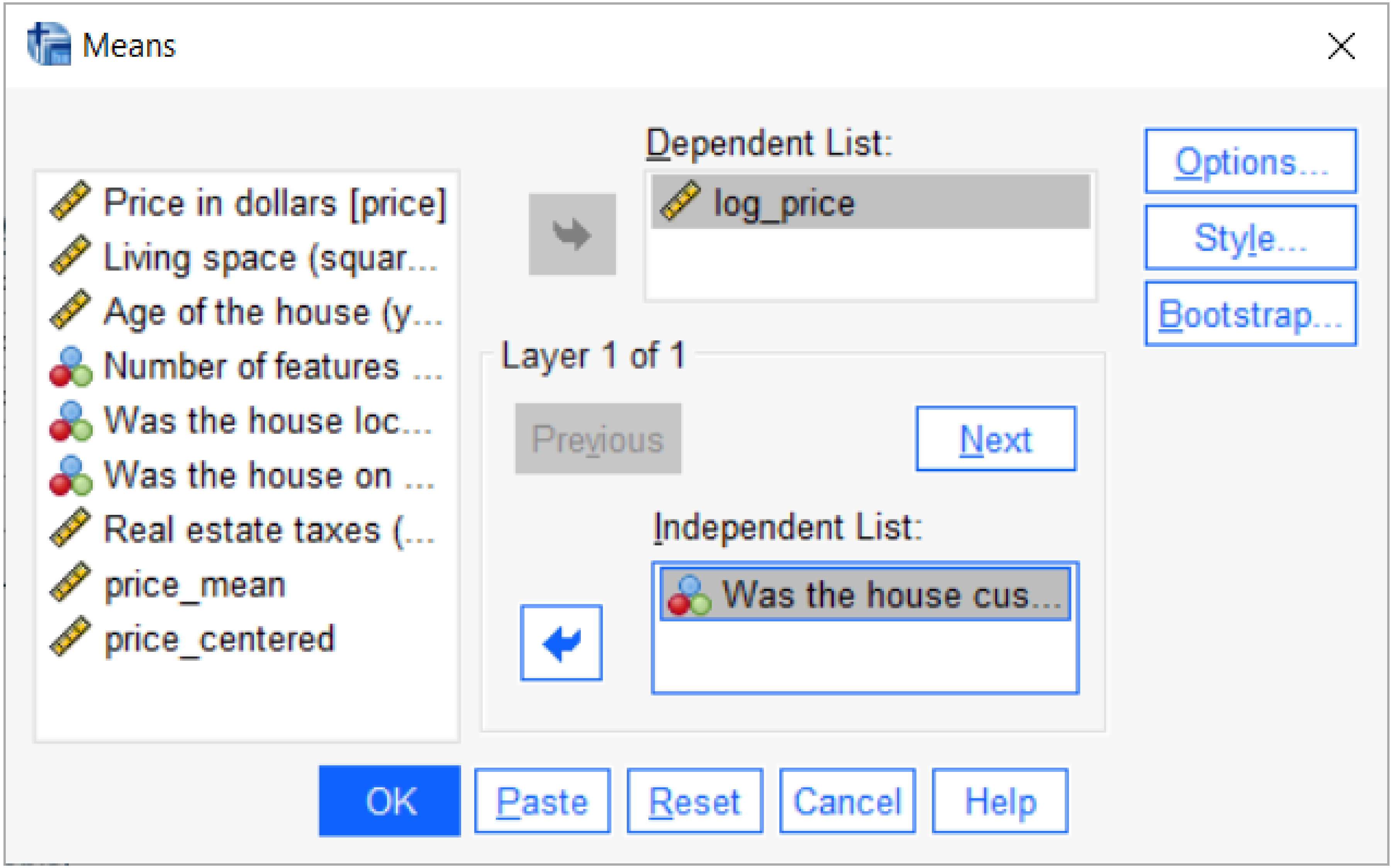
What I look for

- The test you are using
- Specify the number of tails
- Round
- Units of measure

Speaker notes

Here are the things I look for.

Compare means dialog box



Speaker notes

Add note.

Compare means output

log_price

Was the house custom built?	Mean	N	Std. Deviation
No	4.9630	90	.10686
Yes	5.1360	27	.15189
Total	5.0029	117	.13890

Speaker notes

Add note.

Geometric means


Calculator
Scientific
10^(4.963)
91,833.259648358087827572142241311
DEG
F-E
MC MR M+ M− MS M
Trigonometry
Function
2nd π e CE
x2 1/x |x| exp mod
2√x ( ) n! ÷
xy 7 8 9 ×
10x 4 5 6 −
log 1 2 3 +


Calculator
Scientific
10^(5.136)
136,772.88255958490835486987800108
DEG
F-E
MC MR M+ M− MS M
Trigonometry
Function
2nd π e CE
x2 1/x |x| exp mod
2√x ( ) n! ÷
xy 7 8 9 ×
10x 4 5 6 −
log 1 2 3 +

Speaker notes

Add note

Geometric standard deviations


Calculator
Scientific
10^(0.10686)
1.278968946820556391515949714107 2
DEG
F-E
MC
MR
M+
M−
MS
Trigonometry
Function
CE
exp
mod
log


Calculator
Scientific
10^(0.1389)
1.376892391200965370720425468529 3
DEG
F-E
MC
MR
M+
M−
MS
Trigonometry
Function
CE
exp
mod
log

Speaker notes

Add note

General linear model for log transformed data

Coefficients[a]

Model		Unstandardized Coefficients B	Unstandardized Coefficients Std. Error	Standardized Coefficients Beta	t	Sig.
1	(Constant)	4.963	.012		397.164	<.001
	Was the house custom built?	.173	.026	.527	6.652	<.001

a. Dependent Variable: log_price

Speaker notes

Add note.

Back-calculated statistics


Calculator
Scientific
10^(4.963)
91,833.259648358087827572142241311
DEG
F-E
MC
MR
M+
M−
MS
Trigonometry
Function
2nd
π
e
CE
mod
exp
n!
7
8
9
10^x
4
5
6
log
1
2
3


Calculator
Scientific
10^(0.173)
1.4893610777109152570247864696311
DEG
F-E
MC
MR
M+
M−
MS
Trigonometry
Function
2nd
π
e
CE
mod
exp
n!
7
8
9
10^x
4
5
6
log
1
2
3

Speaker notes

Add note

Back calculated confidence intervals, 1 of 2

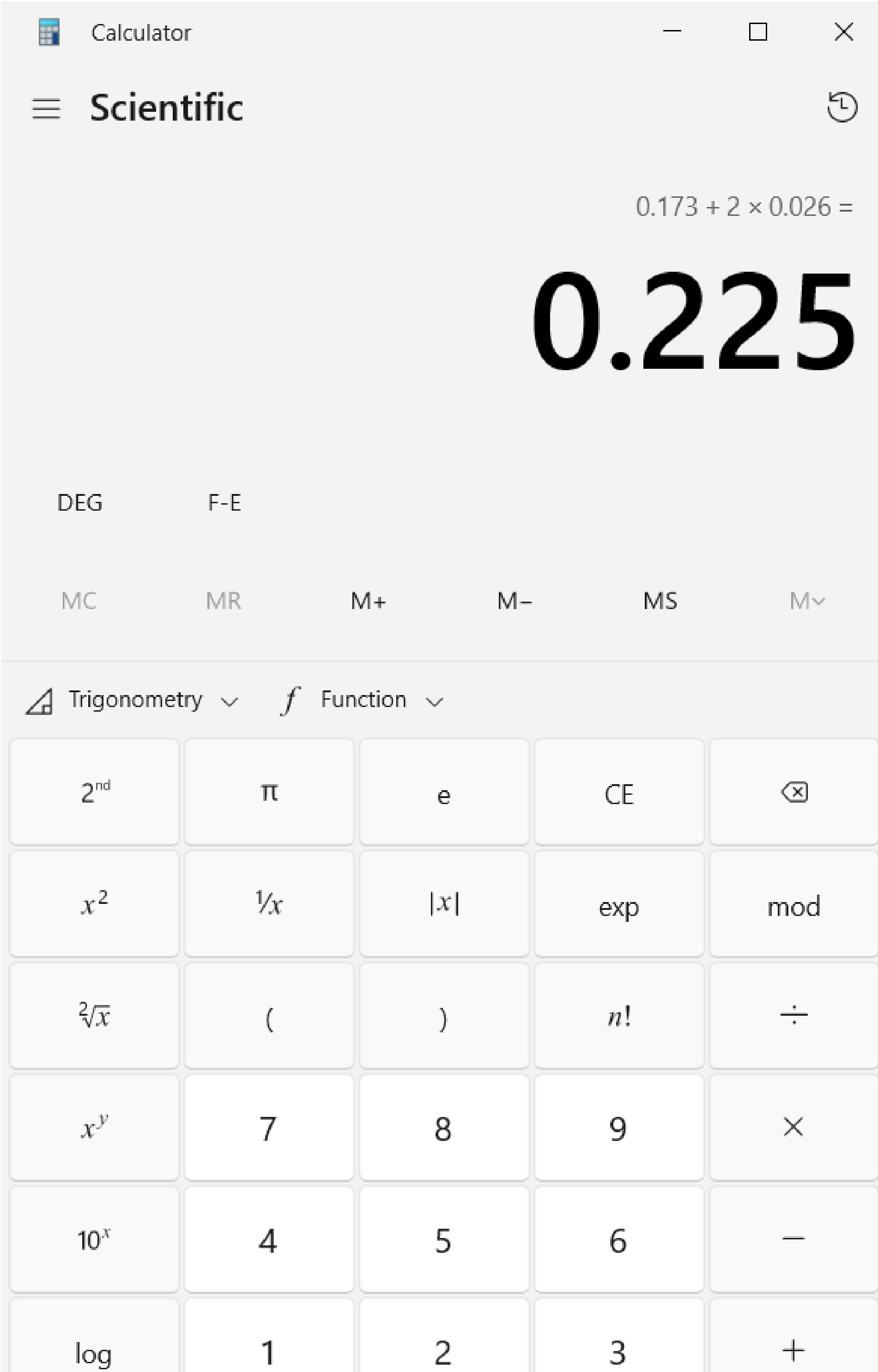

Calculator
Scientific
0.173 - 2 × 0.026 =
0.121
DEG
F-E
MC
MR
M+
M−
MS
M
Trigonometry
Function
2nd
π
e
CE
x2
1/x
|x|
exp
mod
2√x
(
)
n!
÷
xy
7
8
9
×
10x
4
5
6
−
log
1
2
3
+


Calculator
Scientific
10^(0.121)
1.3212956341865751642763000204162
DEG
F-E
MC
MR
M+
M−
MS
M
Trigonometry
Function
2nd
π
e
CE
x2
1/x
|x|
exp
mod
2√x
(
)
n!
÷
xy
7
8
9
×
10x
4
5
6
−
log
1
2
3
+

Speaker notes

Add note

Back calculated confidence intervals, 2 of 2


Calculator
Scientific
0.173 + 2 × 0.026 =
0.225
DEG
F-E
MC
MR
M+
M−
MS
Trigonometry
Function
π
e
CE
exp
mod
(
)
7
8
9
4
5
6
log
1
2
3


Calculator
Scientific
10^(0.225)
1.6788040181225603139719500804156
DEG
F-E
MC
MR
M+
M−
MS
Trigonometry
Function
π
e
CE
exp
mod
(
)
7
8
9
4
5
6
log
1
2
3

Speaker notes

Add note

Conceptual formula for sample size justification, 1 of 2

Review

Synthèse

Dr. Sackett is Director of the Trout Research & Education Centre at Irish Lake, Markdale, Ont.

This article has been peer reviewed.

CMAJ 2001;165(9):1226-37

Return to October 30, 2001
Table of Contents

Why randomized controlled trials fail but needn't: 2. Failure to employ physiological statistics, or the only formula a clinician-trialist is ever likely to need (or understand!)

David L. Sackett

> Because statistics has too often been presented as a bag of specialized computational tools, with morbid emphasis on calculation, it is no wonder that survivors of such courses regard their statistical tools as instruments of torture [rather] than as diagnostic aids in the art and science of data analysis.
>
> — George W. Cobb[1]

From the underside

It was the seventh time I had taken a course in basic biostatistics, and I vowed that this time I was actually going to understand it. The year was 1974, and we were on sabbatical in London. The Beatles had started to beat on each other,

Speaker notes

Add note.

Conceptual formula for sample size justification, 2 of 2

The "only formula" of physiological statistics

The formula is ridiculously simple, and looks like this

$$\text{Confidence} = \frac{\text{Signal}}{\text{Noise}} \times \sqrt{\text{Samplesize}}$$

(Equation 1):

Expressed in words, the *confidence* in the conclusion of an RCT is the ratio of the magnitude of the *signal* to the magnitude of the *noise* times the square root of the *sample size*.

Speaker notes

Add note.

Three things you need for a sample size justification

- Hypothesis
- Standard deviation
- Minimum clinically important difference

Speaker notes

If you work with a researcher who wants to justify a particular sample size, you need to get three things from them.

First, they need a research hypothesis. If they don’t have one, help them formulate one. Some research does not need a research hypothesis, of course. If you are conducting a descriptive study, try to find an important descriptive statistic and select a sample size so the confidence interval using that statistic is reasonably narrow. If your study has a qualitative goal, then you can use a qualitative justification for your sample size.

Second, they need to know how much noise there is in the data. This is measured by the standard deviation. They may have a standard deviation from a pilot study or from previous research. If they don’t have this, ask them for a paper that’s reasonably similar to theirs. It doesn’t have to be identical, but it should use the same outcome measure and the same types of patients. That paper should have a standard deviation in one of the tables.

If not, perhaps you can extrapolate from the range. The range divided by 4 (or maybe 6) is often a reasonable approximation to the standard deviation. You might be able to back-calculate the standard deviation from a confidence interval or p-value, but this is tricky. If all else fails, just use a reasonable guess.

Third, they need to know the minimum clinically important difference. This requires clinical judgement. How much of a change is large enough to get other clinicians to sit up and take notice?

Some people bypass the specification of the minimum clinically important difference and just say that they want to esimate a small, medium, or large effect size. This is a mistake. You cannot plan a study on the basis of a unitless measure.

Rule of 16

- n per group = 16 / ES^2
- Examples:
  - ES = 0.5, n per group = 64
  - ES = 0.1, n per group = 1,600

Speaker notes

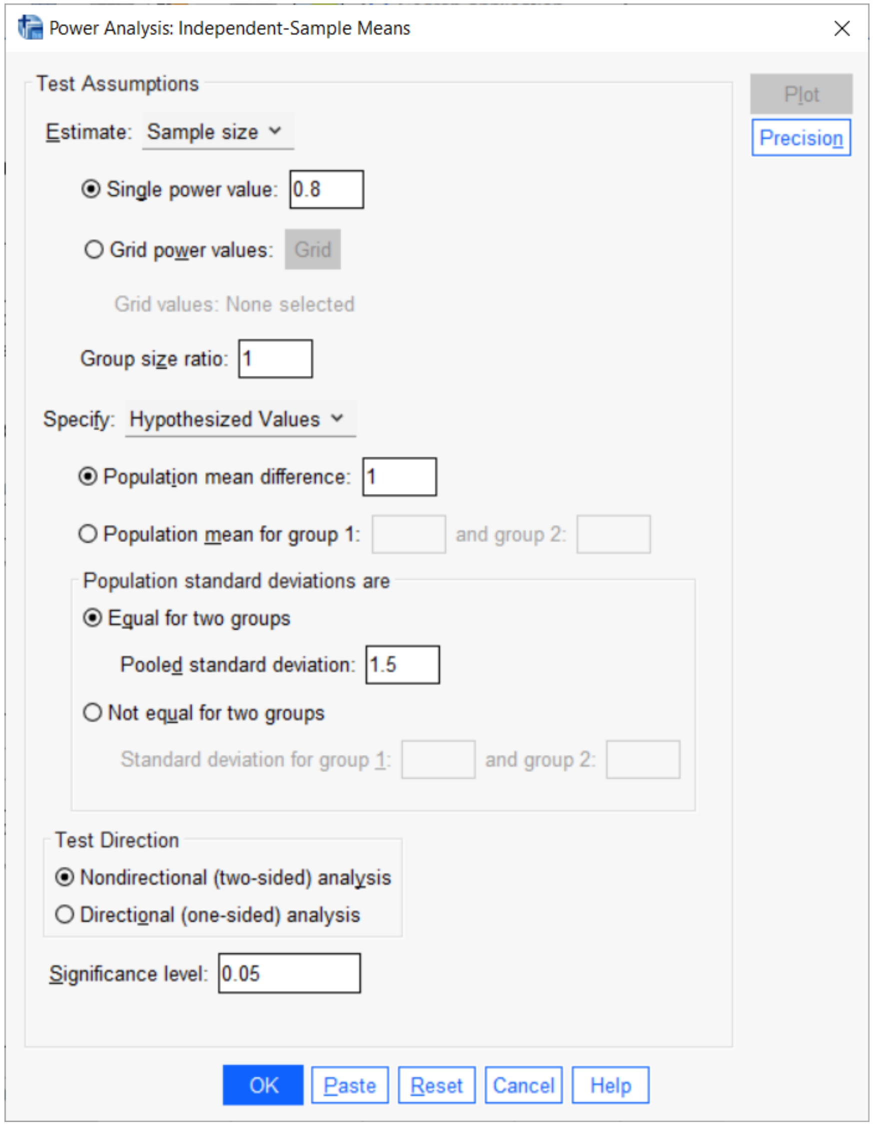
A simple rule for sample size justification that you can do without a lot of calculation is the rule of 16. Take 16 and divide it by the effect size squared.

If your minimum clinically important difference corresponds to half a standard deviation, then 16 divided by 1/2 squared is 64 patients per group.

This is not a formula you use in your grant proposal. It just helps during a brain storming session, or if you are reviewing someone else's work.

The approximation is based on a two-sided alpha level of 0.04, power of 0.80 and equal sample sizes per group.

Sample size calculation dialog box



Speaker notes

I worked on a study at Children's Mercy many years ago where the research team wanted to compare two different skin barrier methods for burn victims. The main outcome was pain after treatment as measured on the Oucher scale. The Oucher scale is a pictorial scale showing faces of patients in various levels of pain from 0 (a child's face showing no expression) to 10 (a face showing a screaming and crying child). Previous research showed that the Oucher scale had a standard deviation of 1.5. A difference of one unit on the Oucher scale was considered clinically significant.

Plug these values into the dialog box.

Sample size calculation output

Power Analysis Table

	N1	N2	Actual Power[b]	Test Assumptions			
				Power	Std. Dev.[c]	Effect Size	Sig.
Test for Mean Difference[a]	37	37	.808	.8	1.5	.667	.05

a. Two-sided test.

b. Based on noncentral t-distribution.

c. Group variances are assumed to be equal.

Speaker notes

This output shows that you need 37 patients per group.

Sample size calculation output for second scenario

Power Analysis Table

	N1	N2	Actual Power[b]	Test Assumptions			
				Power	Std. Dev.[c]	Effect Size	Sig.
Test for Mean Difference[a]	143	143	.802	.8	1.5	.333	.05

a. Two-sided test.

b. Based on noncentral t-distribution.

c. Group variances are assumed to be equal.

Speaker notes

Suppose you were told that your research budget would allow a larger sample size. Your research team decides that even a smaller change in the Oucher score, a 0.5 unit change, would still be clinically important. This represents an effect size of 0.33. Plug the numbers into SPSS and you can see that you need 143 patients per group.

Moon data dictionary, 1 of 4


```
---
data_dictionary: moon.txt

source:
  This data file is part of OzDASL, an archive of various data sets
  useful for teaching. The maintainers of the data archive are from
  Australia, but this particular data set is not specific to that part
  of the world. The entire archive is at
  https://dasl.datadescription.com/
```

Moon data dictionary, 2 of 4

```
description:
  This data set shows a perceptual experiment where subjects were asked
  to estimate a size ratio with their head level to the ground and then
  with their head elevated (in other words, looking upward). Although
  the objects being compared were the same size, almost all subjects
  overestimated the relative sizes. The hypothesis to be tested is
  whether the overestimation is greater with eyes level than with eyes
  elevated. A more detailed description is available at
  https://gksmyth.github.io/ozdasl/general/moon.html

download:
  https://gksmyth.github.io/ozdasl/general/moon.txt
```

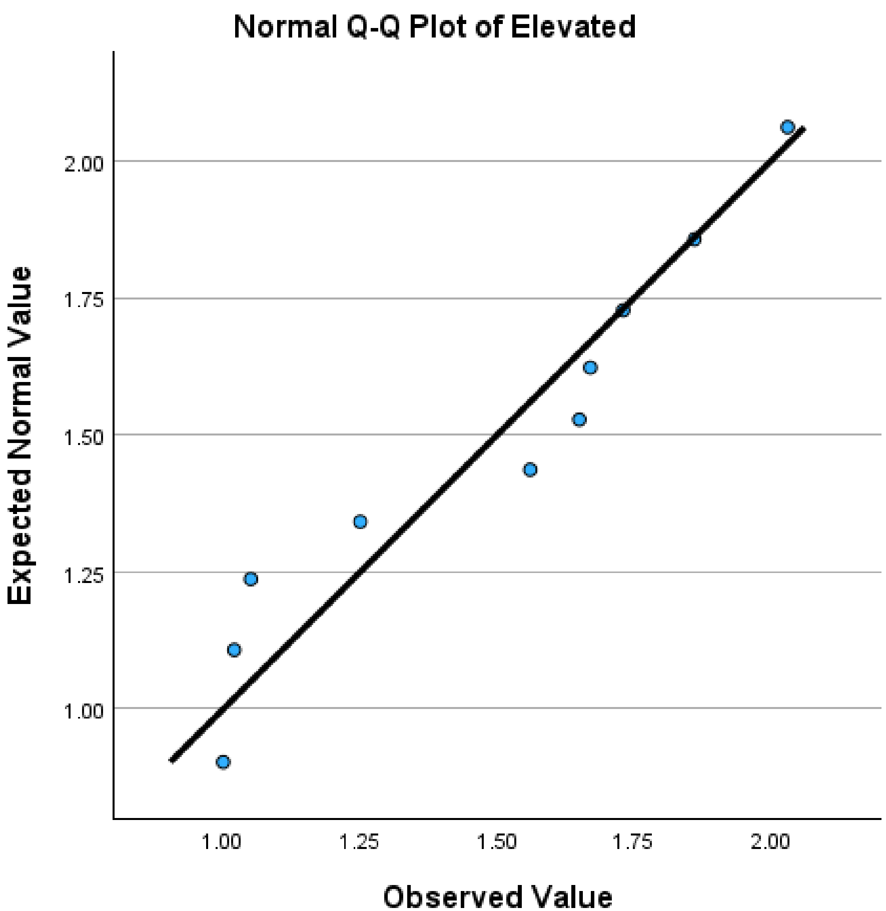
Moon data dictionary, 3 of 4

```
copyright:
  Unknown. You should be able to use this data for individual
  educational purposes under the Fair Use guidelines of U.S.
  copyright law.

format:
  delimiter: tab
  varnames: first row of data
  missing-value-code: not needed
  rows: 10
  columns: 3
```

Moon data dictionary, 4 of 4


```
vars:
  Subject:
    label: Subject number
    format: numeric

  Elevated:
    label: Perceived ratio with eyes elevated
    format: numeric

  Level:
    label: Perceived ratio with eyes level
    format: numeric
---
```

QQ plot of ratio under elevated eye condition



Speaker notes

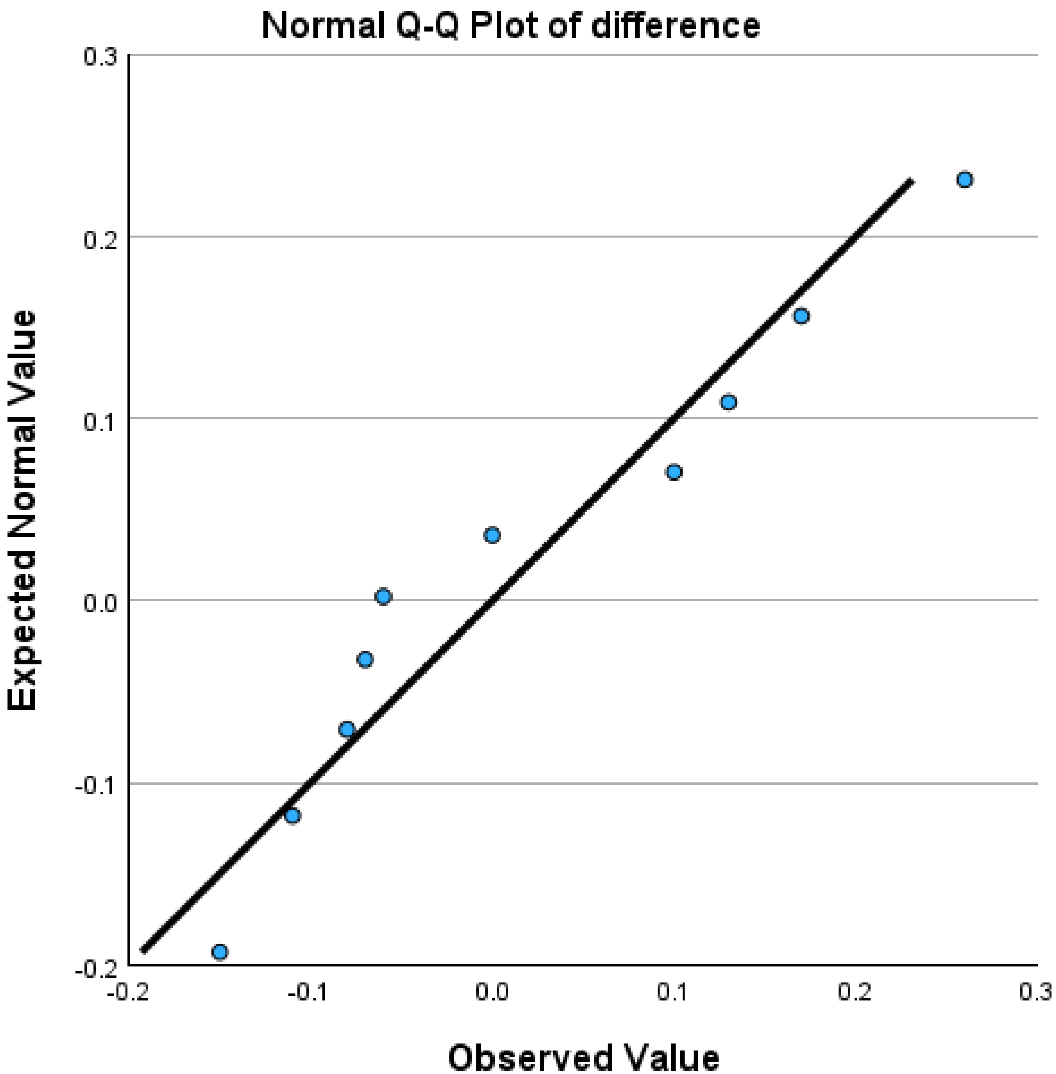
Add note.

QQ plot of ratio under level eye condition

Speaker notes

Add note.

QQ plot of differences



Speaker notes

Add note.

Moon data, descriptive statistics

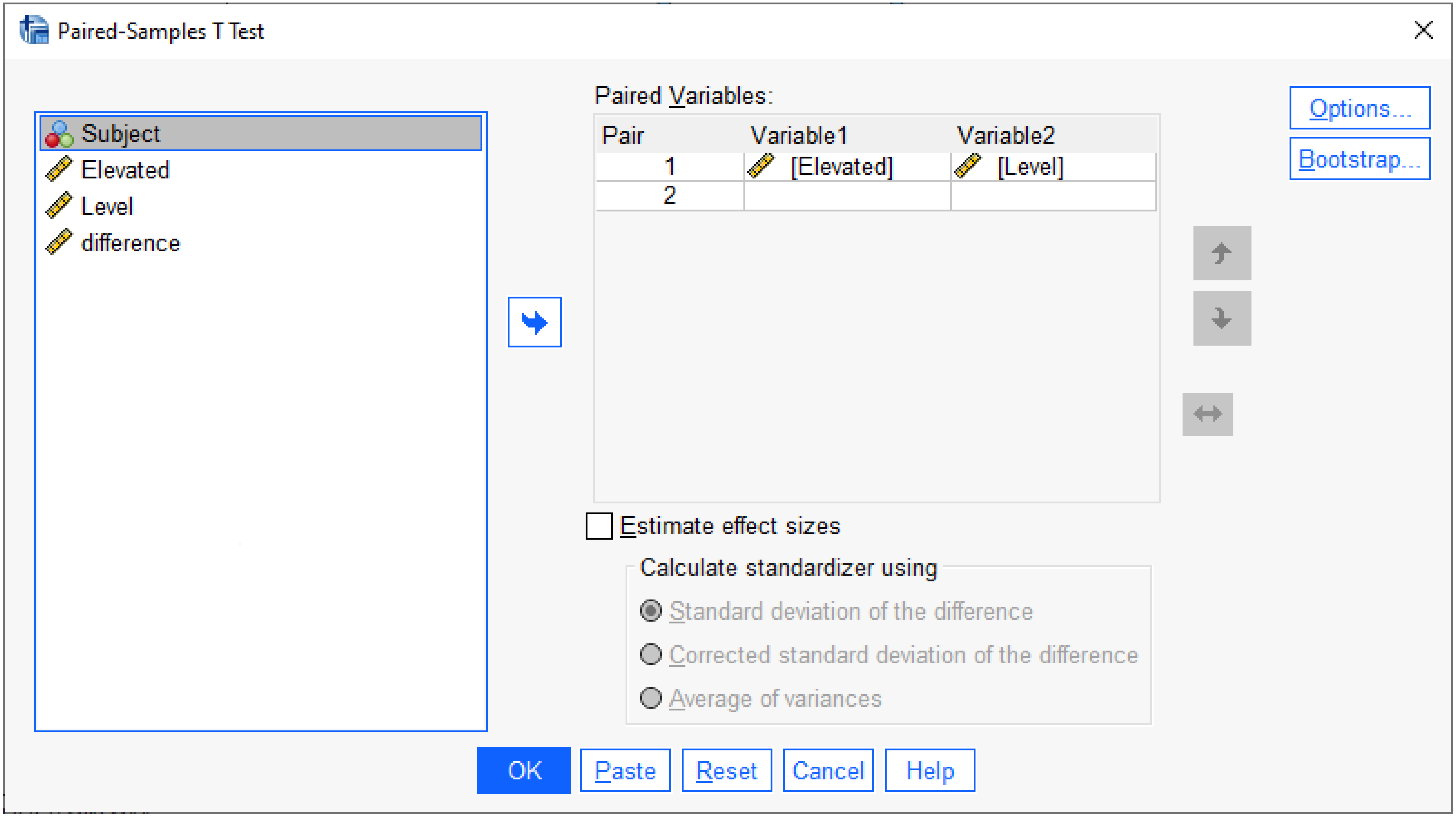
Descriptive Statistics

	N	Minimum	Maximum	Mean	Std. Deviation
Elevated	10	1.00	2.03	1.4820	.37425
Level	10	.95	2.03	1.4630	.34069
difference	10	-.15	.26	.0190	.13715
Valid N (listwise)	10				

Speaker notes

Add note.

Paired Samples T Test dialog box



Speaker notes

Add note.

Paired T Test output, 1 of 5

Paired Samples Statistics

		Mean	N	Std. Deviation	Std. Error Mean
Pair 1	Elevated	1.4820	10	.37425	.11835
	Level	1.4630	10	.34069	.10773

Speaker notes

Add note.

Paired T Test output, 2 of 5

Paired Samples Correlations

		N	Correlation	Significance	
				One-Sided p	Two-Sided p
Pair 1	Elevated & Level	10	.931	<.001	<.001

Speaker notes

Add note.

Paired T Test output, 3 of 5

		Mean	Std. Deviation	Std. Error Mean
Pair 1	Elevated - Level	.01900	.13715	.04337

Speaker notes

Add note.

Paired T Test output, 4 of 5

95% Confidence Interval of the Difference	
Lower	Upper
-.07911	.11711

Speaker notes

Add note.

Paired T Test output, 5 of 5

t	df	Significance: One-Sided p	Significance: Two-Sided p
.438	9	.336	.672

Speaker notes

Add note.

My interpretation

- "Comparison of viewing ratios between the eyes level and eyes elevated conditions used a paired t-test."
- "All tests were conducted using a two-sided alpha level of 0.05."
- "Viewing ratios under both conditions had comaparable sample means and standard deviations. The mean difference was 0.02 with a standard deviation of 0.14. This difference was not statistically significant (95% CI: -0.08 to 0.12, p=0.67).

Speaker notes

Again, there was no power calculation done prior to data collection, so don’t mention it at all. The first two sentences might belong better in the methods section.