STAT W/O MATH

Xiyue Wang & Mengxiao Zhu

中国科学技术大学
University of Science and Technology of China

1. Data Types

以下四个问题涵盖了我们在统计学中接触到的所有数据类型

你最喜欢喝的奶茶是什么牌子？

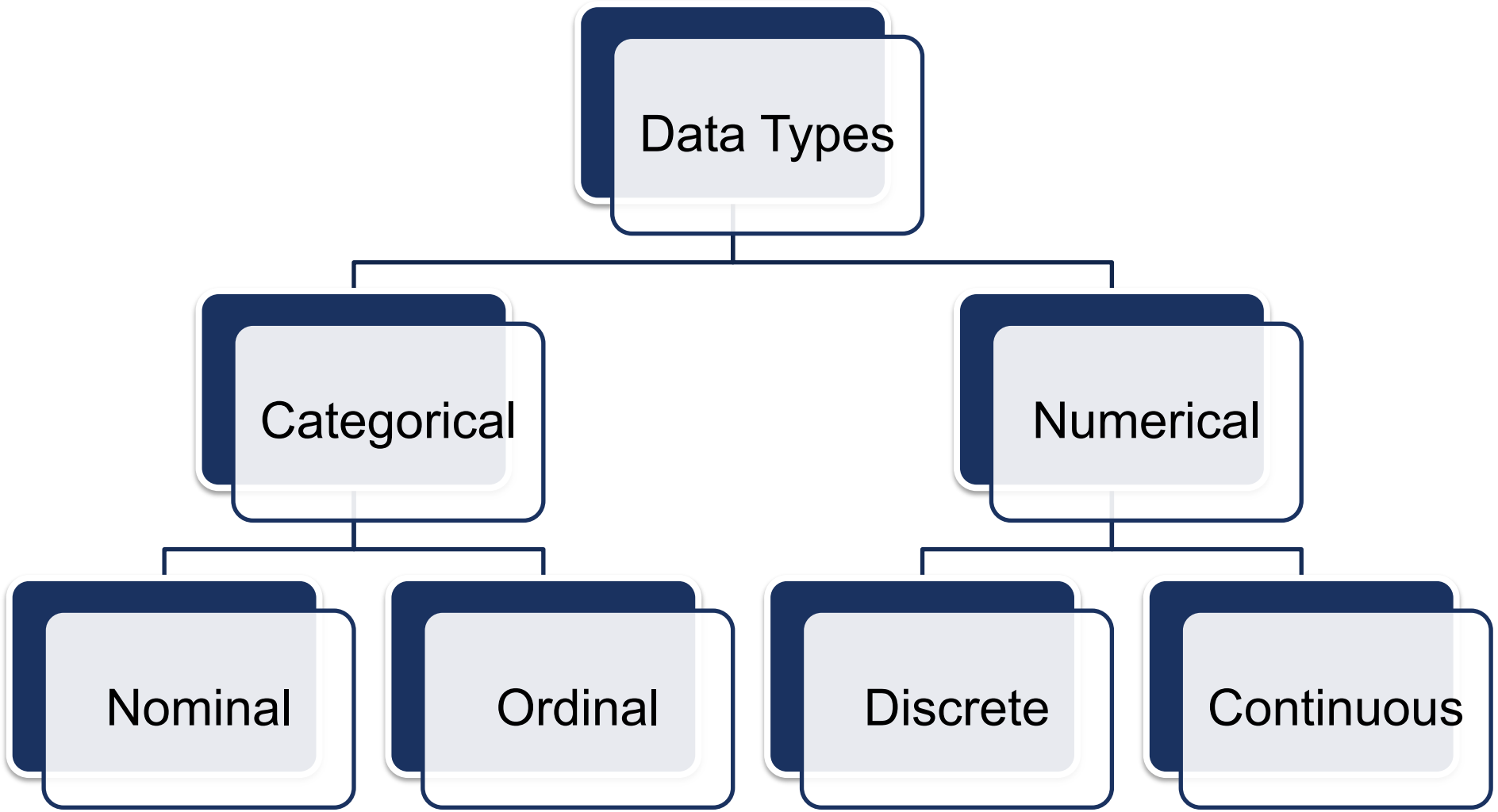
你的最高学历是？

本科生

才上大学的你

你一年发几篇paper？

你的身高是？


Data Types
Categorical
Numerical
Nominal
Ordinal
Discrete
Continuous

Categorical Data

Categorical Data

a. a form of information that can be stored and identified based on their names or labels
b. can be grouped into categories instead of being measured numerically
c. Numbers can sometimes represent it, but those numbers don't mean anything mathematically.
d. Nominal/ Ordinal

Categorical Data

Nominal Data

a. a type of data that consists of categories that can't be ordered or ranked
b. gender, place of birth, occupation

Ordinal Data

a. a category of data that has a natural order.
b. social class, the degree of education, clothing sizes

Numerical Data

Numerical Data

a. Numerical data refers to the data that is in the form of numbers, and not in any language or descriptive form.
b. Numerical data differentiates itself from other number form data types with its ability to carry out arithmetic operations with these numbers.
c. Discrete / Continuous

Numerical Data

Discrete Data

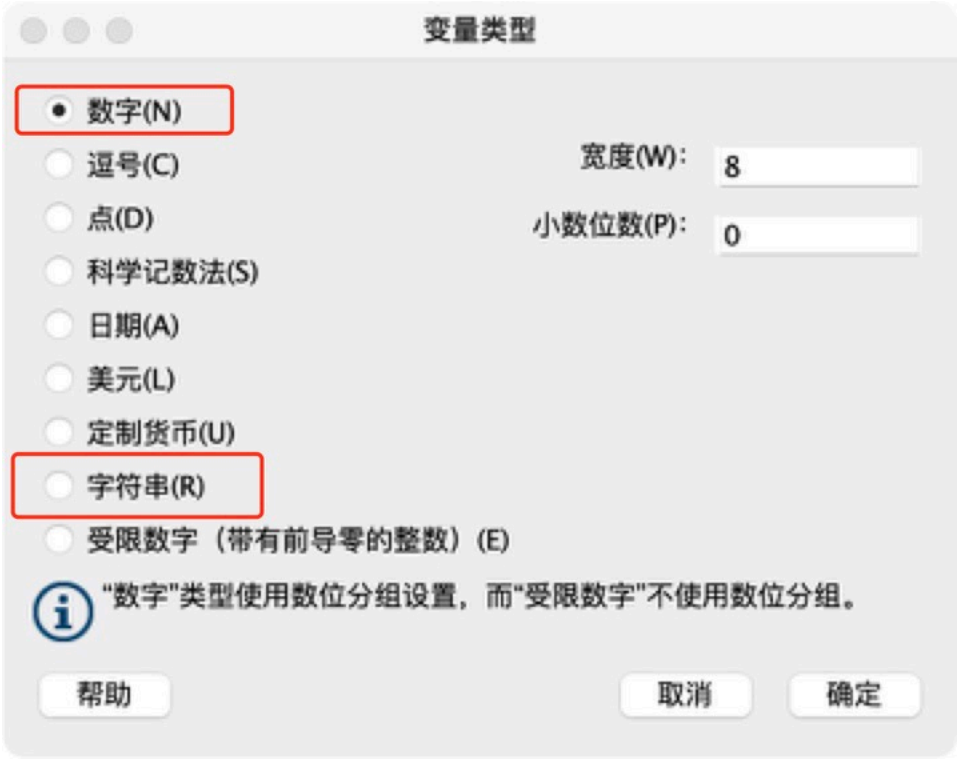
a. Discrete data is used to represent countable items. It can take both numerical and categorical forms and groups them into a list. This list can be finite or infinite too.

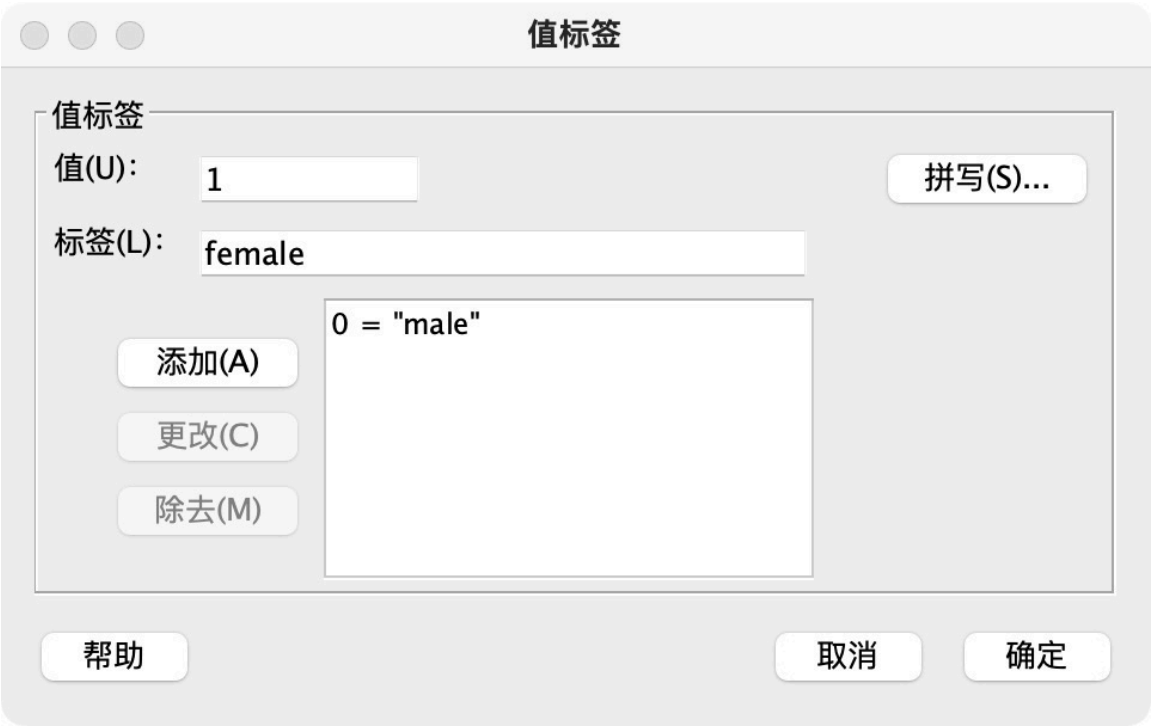
Continuous Data

a. This form has data in the form of intervals. Or simply said, ranges. Continuous numerical data represents measurements and their intervals fall on a number line. Hence, it doesn't involve taking counts of the items.

b. Continuous data is further divided into two categories: Interval and Ratio.

SPSS-数据类型


变量类型
数字(N)
逗号(C)
点(D)
科学记数法(S)
日期(A)
美元(L)
定制货币(U)
字符串(R)
受限数字（带有前导零的整数）(E)
宽度(W):
8
小数位数(P):
0
"数字"类型使用数位分组设置，而"受限数字"不使用数位分组。
帮助
取消
确定

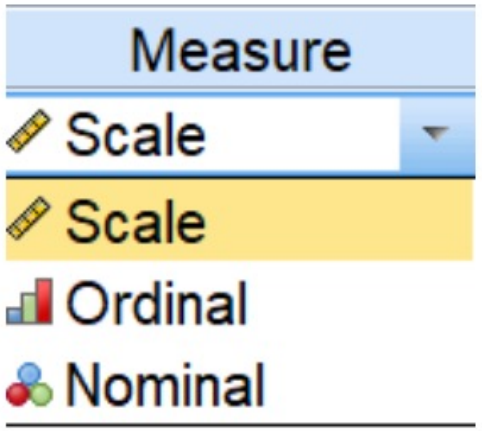

值标签
值标签
值(U):
1
拼写(S)...
标签(L):
female
添加(A)
更改(C)
除去(M)
0 = "male"
帮助
取消
确定

SPSS-测量尺度

Nominal —— 仅有名称、代号、类别组成的数据，不能对数据进行等级划分和排列。

Ordinal —— 适用于可以按照某一顺序进行排列的定性数据，对他们进行计算没有意义。

Scale —— Discrete Data/Continuous Data



2. Describing One Group

Key descriptive statistics

Describe one group	Mean, SD, Median, Mode, Inter-quartile range	Proportion

a. Distribution of variable:

frequencies, proportion

b. Central tendency of variable:

mean, median, mode, quartiles

c. Dispersion of variables:

variance, standard deviation (SD), inter-quartile range

Common central tendency statistics

Mean:

a. the arithmetical average of a distribution (i.e., the sum of all scores divided by N), which is the most frequently used

b. the mean takes into account all the values in the distribution, which makes it especially sensitive to extreme scores.

Median: the midpoint of a distribution with half of the scores above and half below it

Mode: the score(s) that occur(s) most frequently

Quartiles: Quartiles are the values that divide a list of numerical data into three quarters. There are three quartiles, first, second and third, denoted by Q1, Q2 and Q3. Here, Q2 is nothing but the median of the given data.

Common dispersion statistics

Range: In Statistics, the range is the smallest of all the measures of dispersion. It is the difference between the two extreme conclusions of the distribution. In other words, the range is the difference between the maximum and the minimum observation of the distribution.

Inter-quartile range: Inter-quartile range = Upper Quartile – Lower Quartile = Q3 – Q1

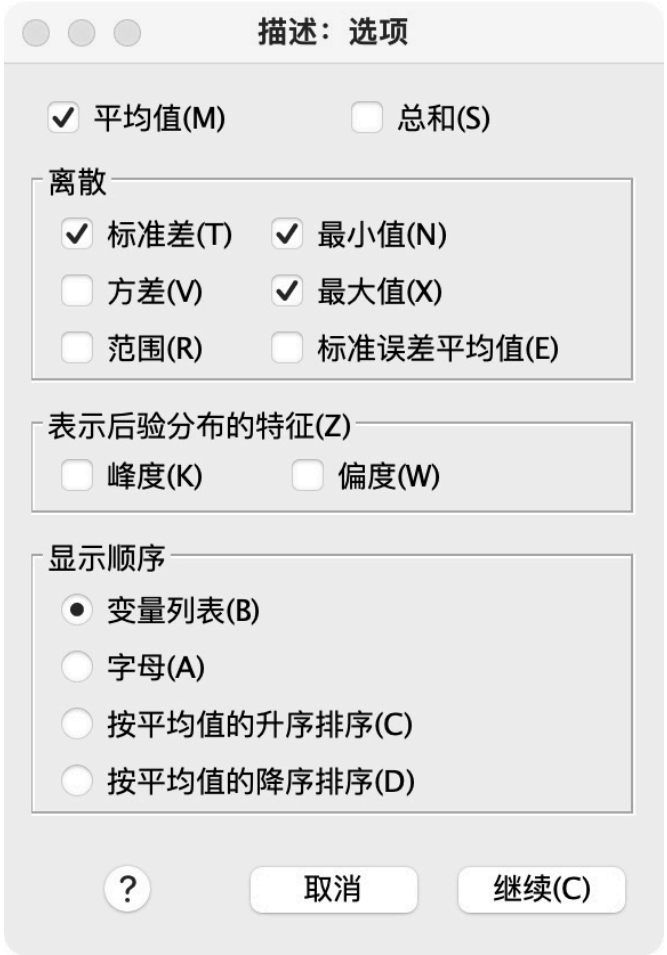
Common dispersion statistics

Variance: It refers to a statistical measurement of the spread between numbers in a data set. More specifically, variance measures how far each number in the set is from the mean (average), and thus from every other number in the set.

Standard deviation: Standard deviation is the square root of variance. It is sometimes more useful since taking the square root removes the units from the analysis. This allows for direct comparisons between different things that may have different units or different magnitudes. For instance, to say that increasing X by one unit increases Y by two standard deviations allows you to understand the relationship between X and Y regardless of what units they are expressed in.

Key descriptive statistics in SPSS

方法I ：分析 >> 描述统计 >> 描述



Key descriptive statistics in SPSS

方法2 ：分析 >> 描述统计 >> 频率

3. Comparing Two Groups

T-test — IV: Categorical DV: Numerical

Chi-square test — IV: Categorical DV: Categorical

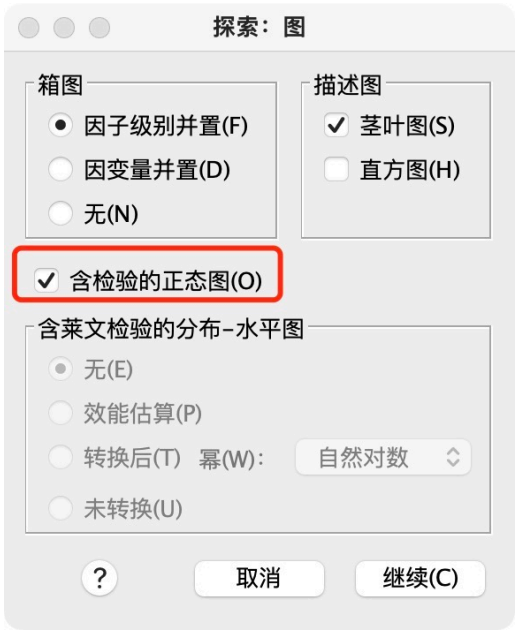
Normality test

Should I choose t-test or non-parametric test?

Small sample (SW)

Big sample (KS)

分析 >> 描述统计 >> 探索



Tests of Normality

	Kolmogorov-Smirnov[a]			Shapiro-Wilk		
	Statistic	df	Sig.	Statistic	df	Sig.
followers	.404	906	.000	.240	906	.000
total_income	.391	906	.000	.259	906	.000

a. Lilliefors Significance Correction

Sig. <0.05则为非正态,选择非参数检验

T-test

A t-test is a statistical test that is used to compare the means of two groups. It is often used in hypothesis testing to determine whether a process or treatment actually has an effect on the population of interest, or whether two groups are different from one another.

T-tests are used when the data sets follow a normal distribution and have unknown variances, like the data set recorded from flipping a coin 100 times.

T-statistic: the t-distribution values, and the degrees of freedom to determine statistical significance.

a. Degrees of freedom is related to your sample size, and shows how many 'free' data points are available in your test for making comparisons. The greater the degrees of freedom, the better your statistical test will work.

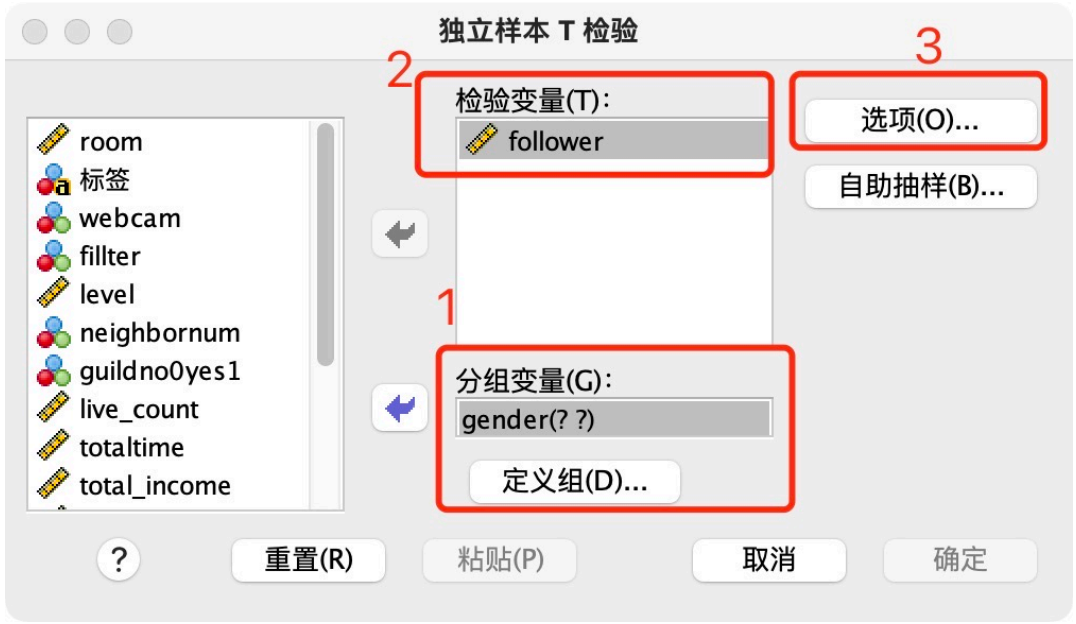
Compare two independent groups - Independent t test

The independent t-test compares the means of two independent groups in order to determine whether there is statistical evidence that the associated population means are significantly different.

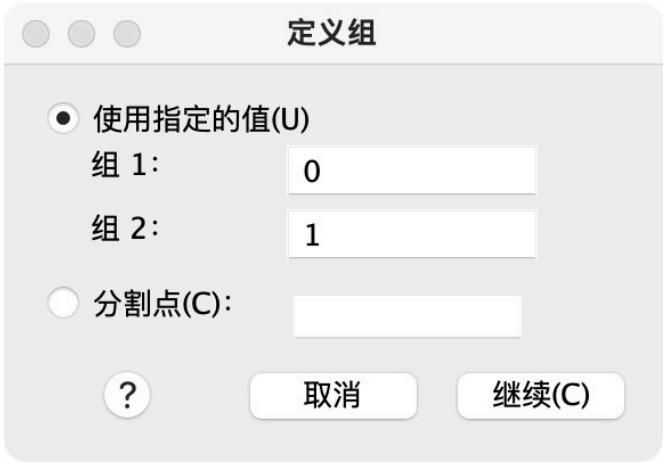
For example, suppose we are evaluating the effect of a medical treatment, and we enroll 100 subjects into our study, then randomly assign 50 subjects to the treatment group and 50 subjects to the control group. In this case, we have two independent samples and would use the unpaired form of the t-test.

Compare two independent groups - Independent t test - SPSS

分析 >> 比较平均值 >> 独立样本T检验



1 – 设置自变量（分组变量）



2 – 设置因变量

Compare two independent groups - Independent t test - SPSS

分析 >> 比较平均值 >> 独立样本T检验

- **T-检验**

组统计

	gender	个案数	平均值	标准 偏差	标准 误差平均值
follower	0	22	61396.3677	138496.195	29527.4879
	1	206	30417.1165	82441.3903	5743.96384

独立样本检验

		莱文方差等同性检验		平均值等同性 t 检验						
									差值 95% 置信区间	
		F	显著性	t	自由度	Sig.（双尾）	平均值差值	标准误差差值	下限	上限
follower	假定等方差	5.269	.023	1.549	226	.123	30979.2512	19995.5827	-8422.3694	70380.8718
	不假定等方差			1.030	22.616	.314	30979.2512	30080.9851	-31306.507	93265.0095

显著性>0.05则看第一行的sig，反之看第二行的sig. 如果sig<0.05，则两组之间有显著差异。

Compare two independent groups - Mann-Whitney test (non-parametric)

When the word "non-parametric" is used in stats, it doesn't quite mean that you know nothing about the population. It usually means that you know the population data does not have a normal distribution.

Mann-Whitney test (Wilcoxon Rank Sum Test) is a nonparametric test of the null hypothesis that, for randomly selected values X and Y from two populations, the probability of X being greater than Y is equal to the probability of Y being greater than X.

Some researchers interpret this as comparing the medians between the two populations

Compare two independent groups - Mann-Whitney test - SPSS

分析 >> 非参数检验 >> 旧对话框 >> 2个独立样本

Compare two independent groups - Mann-Whitney test - SPSS

分析 >> 非参数检验 >> 旧对话框 >> 2个独立样本

- **NPar 检验**

曼-惠特尼检验

秩

	gender	个案数	秩平均值	秩的总和
follower	0	22	138.64	3050.00
	1	206	111.92	23056.00
	总计	228		

检验统计[a]

	follower
曼-惠特尼 U	1735.000
威尔科克森 W	23056.000
Z	-1.806
渐近显著性（双尾）	.071

a. 分组变量：gender

个案处理摘要

	个案					
	包括		排除		总计	
	个案数	百分比	个案数	百分比	个案数	百分比
follower * gender	228	100.0%	0	0.0%	228	100.0%

报告

中位数

gender	follower
0	13399.5446
1	10339.9375
总计	10485.6875

若需要看具体的值：分析 >> 比较平均值 >> 平均值 >> 选项 >> 中位数

Compare two dependent groups - Dependent t test

Dependent samples t-tests compares means from the same group at different times (say, one year apart).

For example:

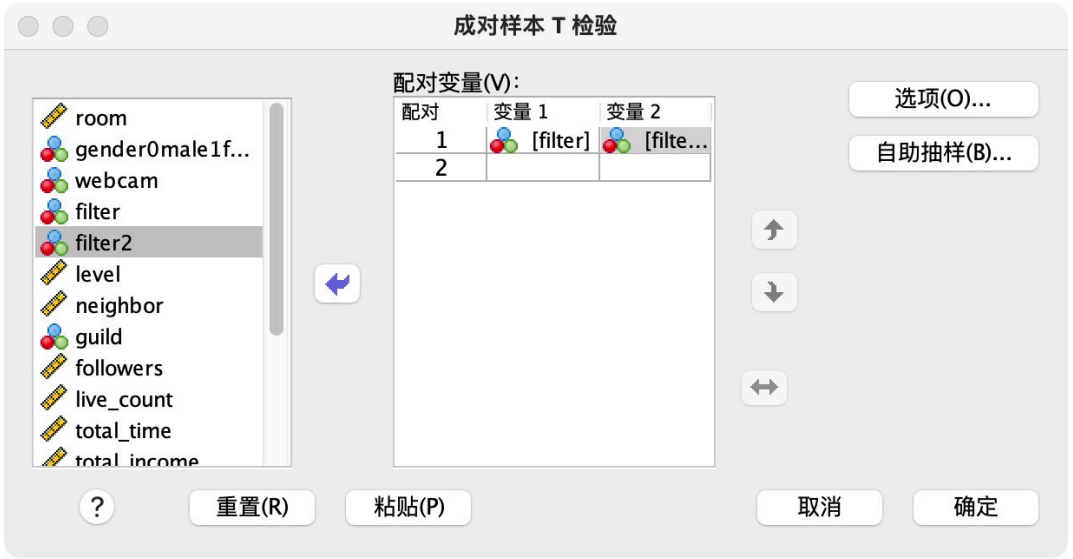
a. Knee MRI costs at two different hospitals,

b. Two tests on the same person before and after training,

c. Two blood pressure measurements on the same person using different equipment.

Compare two dependent groups - Dependent t test - SPSS

分析 >> 比较平均值 >> 成对样本T检验



Paired Samples Correlations

		N	Correlation	Sig.
Pair 1	filter & filter2	594	.990	.000

Paired Samples Test

		Paired Differences					t	df	Sig. (2-tailed)
		Mean	Std. Deviation	Std. Error Mean	95% Confidence Interval of the Difference Lower	Upper			
Pair 1	filter - filter2	-.003	.058	.002	-.008	.001	-1.415	593	.157

Compare two dependent groups - Wilcoxon test (non-parametric)

The Wilcoxon test : is a non-parametric test and is therefore subject to considerably fewer assumptions than its parametric counterpart, the t-test for dependent samples. Therefore, as soon as the boundary conditions for the t-test for dependent samples are no longer fulfilled, the Wilcoxon test is used.

When population means are not of interest. It is used either to test the location of a population based on a sample of data, or to compare the locations of two populations using two matched samples.

Compare two dependent groups - Wilcoxon test (non-parametric)

分析 >> 非参数检验 >> 旧对话框 >> 2个相关样本

Test Statistics[a]

	filter2 – filter
Z	-1.414[b]
Asymp. Sig. (2-tailed)	.157

a. Wilcoxon Signed Ranks Test

b. Based on negative ranks.

T test – APA Style

t (degrees of freedom) = the *t* statistic, *p* = *p* value.

Independent *T*-Test

The 25 participants who received the drug intervention (*M* = 480, *SD* = 34.5) compared to the 28 participants in the control group (*M* = 425, *SD* = 31) demonstrated significantly better peak flow scores, *t*(51) = 2.1, *p* = .04.

There was no significant effect for sex, *t*(38) = 1.7, *p* = .097, despite women (*M* = 55, *SD* = 8) attaining higher scores than men (M = 53, SD = 7.8).

Dependent *T*-Test

The results from the pre-test (*M* = 13.5, *SD* = 2.4) and post-test (*M* = 16.2, *SD* = 2.7) memory task indicate that the presence of caffeine in the bloodstream resulted in an improvement in memory recall, *t*(19) = 3.1, *p* = .006.

There was a significant increase in the volume of alcohol consumed in the week after the end of semester (*M* = 8.7, *SD* = 3.1) compared to the week before the end of semester (*M* = 3.2, *SD* = 1.5), *t*(52) = 4.8, *p* < .001.

T test - APA Style - Table

Results of Curve-Fitting Analysis Examining the Time Course of Fixations to the Target

Logistic parameter	9-year-olds		16-year-olds		*t*(40)	*p*	Cohen's *d*
	M	*SD*	*M*	*SD*			
Maximum asymptote, proportion	.843	.135	.877	.082	0.951	.347	0.302
Crossover, in ms	759	87	694	42	2.877	.006	0.840
Slope, as change in proportion per ms	.001	.0002	.002	.0002	2.635	.012	2.078

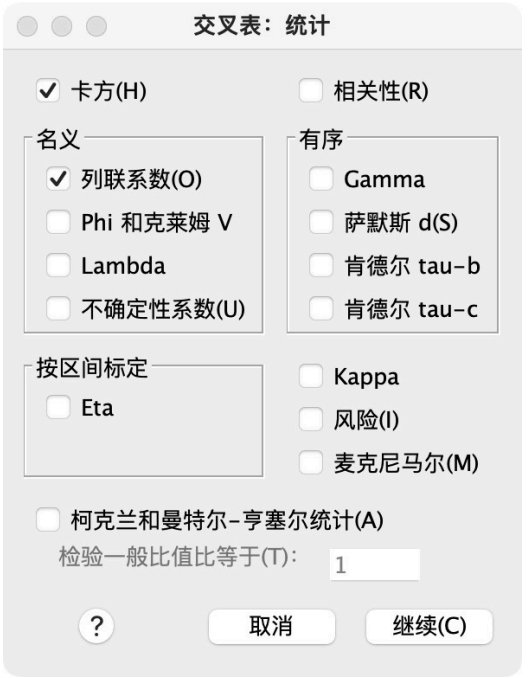
Note. For each subject, the logistic function was fit to target fixations separately. The maximum asymptote is the asymptotic degree of looking at the end of the time course of fixations. The crossover point is the point in time the function crosses the midway point between peak and baseline. The slope represents the rate of change in the function measured at the crossover. Mean parameter values for each of the analyses are shown for the 9-year-olds (n = 24) and 16-year-olds (n = 18), as well as the results of t tests (assuming unequal variance) comparing the parameter estimates between the two ages.

Chi-square Test

- A chi-square ($\chi 2$) statistic is a measure of the difference between the observed and expected frequencies of the outcomes of a set of events or variables.
- Chi-square is useful for analyzing such differences in categorical variables, especially those nominal in nature.
- The data used in calculating a chi-square statistic must be random, raw, mutually exclusive, drawn from independent variables, and drawn from a large enough sample. For example, the results of tossing a fair coin meet these criteria.
- $\chi 2$ depends on the size of the difference between actual and observed values, the degrees of freedom, and the sample size.
- Pearson's chi-squared test is used to determine whether there is a statistically significant difference between the expected frequencies and the observed frequencies in one or more categories of a contingency table. For contingency tables with smaller sample sizes, a Fisher's exact test is used instead.

Chi-square Test in SPSS

分析 >> 描述统计 >> 交叉表



Chi-square Test in SPSS

webcam * gender0male1female 交叉表

			gender0male1female 0	gender0male1female 1	总计
webcam	0	计数	228	84	312
		占 gender0male1female 的百分比	57.3%	16.5%	34.4%
	1	计数	170	424	594
		占 gender0male1female 的百分比	42.7%	83.5%	65.6%
总计		计数	398	508	906
		占 gender0male1female 的百分比	100.0%	100.0%	100.0%

对称测量

		值	渐进显著性
名义到名义	列联系数	.392	.000
有效个案数		906	

卡方检验

	值	自由度	渐进显著性（双侧）	精确显著性（双侧）	精确显著性（单侧）
皮尔逊卡方	164.139[a]	1	.000		
连续性修正[b]	162.339	1	.000		
似然比	167.852	1	.000		
费希尔精确检验				.000	.000
线性关联	163.957	1	.000		
有效个案数	906				

a. 0 个单元格 (0.0%) 的期望计数小于 5。最小期望计数为 137.06。

b. 仅针对 2x2 表进行计算

- Pearson卡方要求总样本量n≥40，且所有的期望频数E≥5。
- 当n≥40且存在任意一个1 ≤E<5，则使用连续性修正卡方。
- 当n<40，或E<1时，则用Fisher精确检验。

Chi-square Test – APA Style

X^2 (degress of freedom, N = sample size) = chi-square statistic value, p = p value.

Example

Imagine we conducted a study that looked at whether there is a link between gender and the ability to swim. We might report the results like this:

A chi-square test of independence was performed to examine the relation between gender and the ability to swim. The relation between these variables was significant, X^2 (1, N = 84) = 8.9, p = .0029. Women were more likely than men to be able to swim.

Other Examples

The proportion of subjects who reported being depressed did not differ by marriage, X^2 (1, N = 104) = 1.7, p > .05.

There is a significant relationship between the two variables. Hipsters are more likely than non-hipsters to own an IPhone, X^2 (1, N = 54) = 6.7, p < .01.

A chi-square test of independence showed that there was no significant association between gender and chocolate preference, X^2 (2, N = 88) = 2.1, p = .35.

Chi-square Test – APA Style — Table

Table 7.7 Sample Chi-Square Analysis Table

Table 1

Frequencies and Chi-Square Results for Belief Perseverance in Attitudes Toward Celebrities (N = 201)

Source	Do not believe		Unsure		Believe		$\chi^2(2)$
	n	%	n	%	n	%	
Media reports	17	8.46	140	69.65	44	21.89	124.75*
Family reports	47	23.38	106	52.74	48	23.88	34.06*
Friends' reports	42	20.90	112	55.72	47	23.38	45.52*
Caught by media	19	9.45	82	40.80	100	49.75	54.00*
Celebrity display of behavior	12	5.97	61	30.35	128	63.68	101.22*

*$p < .001$.

4. Compare three or more unmatched groups

One-way ANOVA — IV: Categorical DV: Numerical

Chi-square test — IV: Categorical DV: Categorical

Compare three or more unmatched groups - One-way ANOVA

One-way ANOVA compares the means of two or more independent groups in order to determine whether there is statistical evidence that the associated population means are significantly different.

Requirements:

a. Dependent variable that is continuous (i.e., interval or ratio level) and independent variable that is categorical (i.e., two or more groups)

b. Cases that have values on both the dependent and independent variables

c. Independent samples/groups (i.e., independence of observations)

d. There is no relationship between the subjects in each sample.

e. Normal distribution (approximately) of the dependent variable for each group

f. Homogeneity of variances (i.e., variances approximately equal across groups)

Compare three or more unmatched groups - One-way ANOVA

The test statistic for a One-Way ANOVA is denoted as F.

! If data are ordinal, a non-parametric alternative to this test should be used such as Kruskal–Wallis one-way analysis of variance.

One-way ANOVA in SPSS

分析 >> 比较平均值 >> 单因素ANOVA检验

One-way ANOVA in SPSS

分析 >> 比较平均值 >> 单因素ANOVA检验

One-way ANOVA in SPSS

分析 >> 比较平均值 >> 单因素ANOVA检验

Test of Homogeneity of Variances

		Levene Statistic	df1	df2	Sig.
followers	Based on Mean	20.841	2	903	.000
	Based on Median	7.511	2	903	.001
	Based on Median and with adjusted df	7.511	2	676.888	.001
	Based on trimmed mean	9.649	2	903	.000

先看方差齐性检验，若p<0.05，则方差不齐，则需要看

了解每两组的情况，则需要

Robust Tests of Equality of Means

followers

	Statistic[a]	df1	df2	Sig.
Welch	22.635	2	453.366	.000
Brown-Forsythe	9.159	2	677.636	.000

a. Asymptotically F distributed.

p<0.05，三组存在显著差异

Post Hoc Tests

Multiple Comparisons

Dependent Variable: followers

	(I) channel	(J) channel	Mean Difference (I-J)	Std. Error	Sig.	95% Confidence Interval Lower Bound	95% Confidence Interval Upper Bound
Bonferroni	0	1	.000000000	132053.250	1.000	-316722.99	316722.988
		2	515191.39*	147250.982	.001	162017.416	868365.367
	1	0	.000000000	132053.250	1.000	-316722.99	316722.988
		2	515191.39*	147250.982	.001	162017.416	868365.367
	2	0	-515191.4*	147250.982	.001	-868365.37	-162017.42
		1	-515191.4*	147250.982	.001	-868365.37	-162017.42
Tamhane	0	1	.000000000	152570.964	1.000	-365204.45	365204.447
		2	515191.39*	108046.386	.000	255927.879	774454.904
	1	0	.000000000	152570.964	1.000	-365204.45	365204.447
		2	515191.39*	108046.386	.000	255927.879	774454.904
	2	0	-515191.4*	108046.386	.000	-774454.90	-255927.88
		1	-515191.4*	108046.386	.000	-774454.90	-255927.88

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

One-way ANOVA in SPSS

分析 >> 比较平均值 >> 单因素ANOVA检验

Test of Homogeneity of Variances

		Levene Statistic	df1	df2	Sig.
followers	Based on Mean	20.841	2	903	.000
	Based on Median	7.511	2	903	.001
	Based on Median and with adjusted df	7.511	2	676.888	.001
	Based on trimmed mean	9.649	2	903	.000

先看方差齐性检验，若p>0.05，则方差齐性，则需要看

ANOVA

followers

	Sum of Squares	df	Mean Square	F	Sig.
Between Groups	4.529E+13	2	2.264E+13	7.661	.001
Within Groups	2.669E+15	903	2.956E+12		
Total	2.714E+15	905			

p<0.05，三组存在显著差异

了解每两组的情况，则需要

Post Hoc Tests

Multiple Comparisons

Dependent Variable: followers

	(I) channel	(J) channel	Mean Difference (I-J)	Std. Error	Sig.	95% Confidence Interval Lower Bound	95% Confidence Interval Upper Bound
Bonferroni	0	1	.000000000	132053.250	1.000	-316722.99	316722.988
		2	515191.39*	147250.982	.001	162017.416	868365.367
	1	0	.000000000	132053.250	1.000	-316722.99	316722.988
		2	515191.39*	147250.982	.001	162017.416	868365.367
	2	0	-515191.4*	147250.982	.001	-868365.37	-162017.42
		1	-515191.4*	147250.982	.001	-868365.37	-162017.42
Tamhane	0	1	.000000000	152570.964	1.000	-365204.45	365204.447
		2	515191.39*	108046.386	.000	255927.879	774454.904
	1	0	.000000000	152570.964	1.000	-365204.45	365204.447
		2	515191.39*	108046.386	.000	255927.879	774454.904
	2	0	-515191.4*	108046.386	.000	-774454.90	-255927.88
		1	-515191.4*	108046.386	.000	-774454.90	-255927.88

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

One-way ANOVA – APA Style

A one way ANOVA was performed to evaluate the relationship between [independent variable] and [dependent variable]. The means (1) and standard deviations (2) are presented in Table [table number] (3) below.

Table [X] (3)

Descriptive Statistics for [Dependent Variable]

[Independent Variable]	*M* (1)	*SD* (2)
[Group/level 1]	[12.34]	[1.11]
[Group/level 2]	[34.56]	[2.22]
[Group/level 3]	[67.89]	[3.33]

The ANOVA [was/was not] (4) significant at the [alpha level] (5) level, *F*([*df* between groups] (6), [*df* within groups] (7))= [*F* value] (8) , *p* = [*p* value] (9).

Note: Include the following section only if the ANOVA was significant and you have conducted a post hoc Tukey HSD test:

A post hoc Tukey HSD test (10) indicated that the mean [dependent variable] of the [group/level name] was significantly [higher/lower] than that of the [group/level name] (*p* = [*p* value]) (11). However, there were no significant differences between the mean [dependent variable] of the [group/level name] and [group/level name] (*p* = [*p* value]) (12).

One-way ANOVA – APA Style — Table

Table 1

Means, Standard Deviations, and One-Way Analyses of Variance in Psychological and Social Resources and Cognitive Appraisals

Measure	Urban		Rural		$F(1, 294)$	η^2
	M	*SD*	*M*	*SD*		
Self-esteem	2.91	0.49	3.35	0.35	68.87***	.19
Social support	4.22	1.50	5.56	1.20	62.60***	.17
Cognitive appraisals						
Threat	2.78	0.87	1.99	0.88	56.35***	.20
Challenge	2.48	0.88	2.83	1.20	7.87***	.03
Self-efficacy	2.65	0.79	3.53	0.92	56.35***	.16

***$p < .001$.

5. Quantify association between two variables

Quantify association between two variables - Pearson correlation

Correlation coefficient: how strong a relationship is between data. The formulas return a value between -1 and 1, where:

- 1 indicates a strong positive relationship.
- -1 indicates a strong negative relationship.
- A result of zero indicates no relationship at all.

Pearson correlation: it describes the strength and direction of the linear relationship between two quantitative variables.

Quantify association between two variables - Pearson correlation

When to use:

a. Both variables are quantitative.

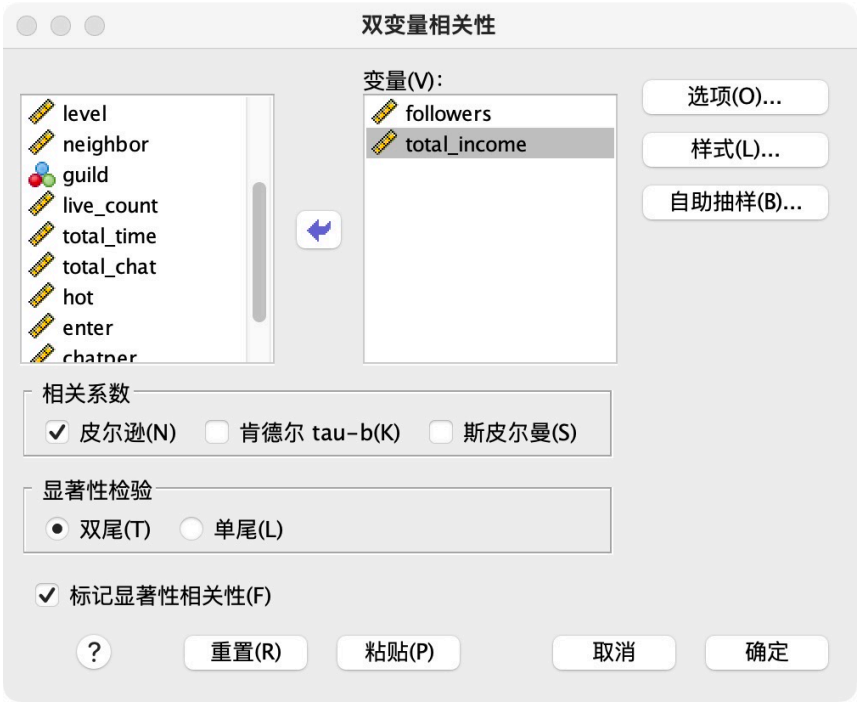
b. The variables are normally.

c. The data have no outliers.

d. The relationship is linear: “Linear” means that the relationship between the two variables can be described reasonably well by a straight line. You can use a scatterplot to check whether the relationship between two variables is linear.

Pearson correlation in SPSS

分析 >> 相关>> 双变量 >> 2个相关样本



相关性

相关性

		followers	total_income
followers	皮尔逊相关性	1	.727**
	Sig.（双尾）		.000
	个案数	906	906
total_income	皮尔逊相关性	.727**	1
	Sig.（双尾）	.000	
	个案数	906	906

**. 在 0.01 级别（双尾），相关性显著。

1）先看Sig.（显著性）<0.05 具有统计显著性之后 再看相关系数

2）看两个变量之间是否特别相关是看相关系数 而不是看星号

Pearson correlation — APA Style

A Pearson correlation coefficient was computed to assess the linear relationship between **[variable 1]** and **[variable 2]**.

There was a **[negative or positive]** correlation between the two variables, r(**df**) = **[r value]**, p = **[p-value]**.

Quantify association between two variables - Spearman correlation

Spearman's rank correlation coefficient is another widely used correlation coefficient. It's a better choice than the Pearson correlation coefficient when one or more of the following is true:

- The variables are ordinal.
- The variables aren't normally distributed.
- The data includes outliers.
- The relationship between the variables is non-linear and monotonic.

Spearman correlation in SPSS

分析 >> 相关>> 双变量 >> 2个相关样本

非参数相关性

相关性

			total_income	webcam
斯皮尔曼 Rho	total_income	相关系数	1.000	.110**
		Sig.（双尾）	.	.001
		N	906	906
	webcam	相关系数	.110**	1.000
		Sig.（双尾）	.001	.
		N	906	906

**. 在 0.01 级别（双尾），相关性显著。

Spearman correlation – APA Style

Spearman's rank correlation was computed to assess the relationship between **[variable 1]** and **[variable 2]**.

There was a **[negative or positive]** correlation between the two variables, r(**df**) = **[r value]**, p = **[p-value]**.

Correlation - APA Style - Table

Descriptive Statistics and Correlations for Study Variables

Variable	*n*	*M*	*SD*	1	2	3	4	5	6	7
1. Internal–external status [a]	3,697	0.43	0.49	—						
2. Manager job performance	2,134	3.14	0.62	−.08**	—					
3. Starting salary [b]	3,697	1.01	0.27	.45**	−.01	—				
4. Subsequent promotion	3,697	0.33	0.47	.08**	.07**	.04*	—			
5. Organizational tenure	3,697	6.45	6.62	−.29**	.09**	.01	.09**	—		
6. Unit service performance [c]	3,505	85.00	6.98	−.25**	−.39**	.24**	.08**	.01	—	
7. Unit financial performance [c]	694	42.61	5.86	.00	−.03	.12*	−.07	−.02	.16**	—

[a] 0 = internal hires and 1 = external hires.

[b] A linear transformation was performed on the starting salary values to maintain pay practice

6. Predict Value - Regression Models

Simple Linear Regeression

Linear regression is the most widely used statistical technique; it is a way to model a relationship between two sets of variables. The result is a linear regression equation that can be used to make predictions about data.

You can use it when you want to:

a. identify the strength of the effect that the independent variable(s) have on a dependent variable.

b. forecast effects or impact of change.

c. predict trends and future values.

Simple linear regression : 1 dependent variable (interval or ratio), 1 independent variable (interval or ratio or dichotomous), which is one type of linear regression.

Simple Linear Regeression

Types of Linear Regression

Simple linear regression
1 dependent variable (interval or ratio), 1 independent variable (interval or ratio or dichotomous)

Multiple linear regression
1 dependent variable (interval or ratio) , 2+ independent variables (interval or ratio or dichotomous)

Logistic regression
1 dependent variable (dichotomous), 2+ independent variable(s) (interval or ratio or dichotomous)

Ordinal regression
1 dependent variable (ordinal), 1+ independent variable(s) (nominal or dichotomous)

Multinomial regression
1 dependent variable (nominal), 1+ independent variable(s) (interval or ratio or dichotomous)

Discriminant analysis
1 dependent variable (nominal), 1+ independent variable(s) (interval or ratio)

Simple Linear Regression

key assumptions:

- Linear relationship: There must be a linear relationship between the outcome variable and the independent variables. Scatterplots can show whether there is a linear or curvilinear relationship.
- Multivariate Normality–Multiple regression assumes that the residuals are normally distributed.
- No Multicollinearity—Multiple regression assumes that the independent variables are not highly correlated with each other. This assumption is tested using Variance Inflation Factor (VIF) values.
- Homoscedasticity–This assumption states that the variance of error terms are similar across the values of the independent variables. A plot of standardized residuals versus predicted values can show whether points are equally distributed across all values of the independent variables.

Simple Linear Regression – APA Style

Simple linear regression was used to test if [predictor variable] significantly predicted [response variable].

The fitted regression model was: [fitted regression equation]

The overall regression was statistically significant (R^2 = [R2 value], $F(df_{regression}, df_{residual})$ = [F-value], p = [p-value]).

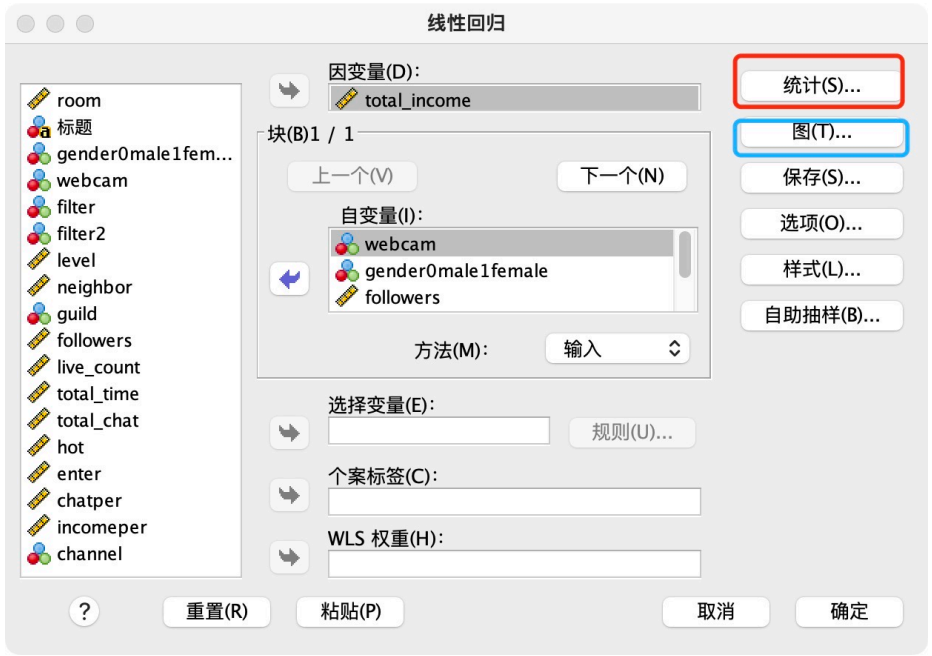
It was found that [predictor variable] significantly predicted [response variable] (β = [β-value], p = [p-value]).

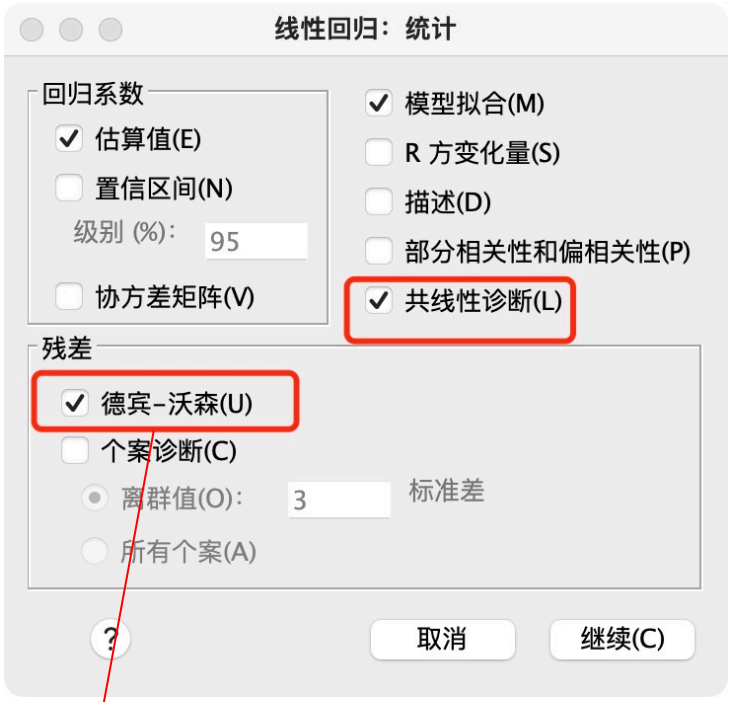
Multiple Linear Regression

Multiple linear regression is the most common form of linear regression analysis. As a predictive analysis, the multiple linear regression is used to explain the relationship between one continuous dependent variable and two or more independent variables. The independent variables can be continuous or categorical (dummy coded as appropriate).

Linear Regression in SPSS

分析 >> 回归 >> 线性





面板数据需要检测序列相关

Multiple Linear Regression in SPSS

Model Summary[b]

Model	R	R Square	Adjusted R Square	Std. Error of the Estimate	Durbin–Watson
1	.740[a]	.548	.546	325632.499	1.976

a. Predictors: (Constant), total_time, webcam, followers, gender0male1female, level

b. Dependent Variable: total_income

ANOVA[a]

Model		Sum of Squares	df	Mean Square	F	Sig.
1	Regression	1.157E+14	5	2.315E+13	218.290	.000[b]
	Residual	9.543E+13	900	1.060E+11		
	Total	2.112E+14	905			

a. Dependent Variable: total_income

b. Predictors: (Constant), total_time, webcam, followers, gender0male1female, level

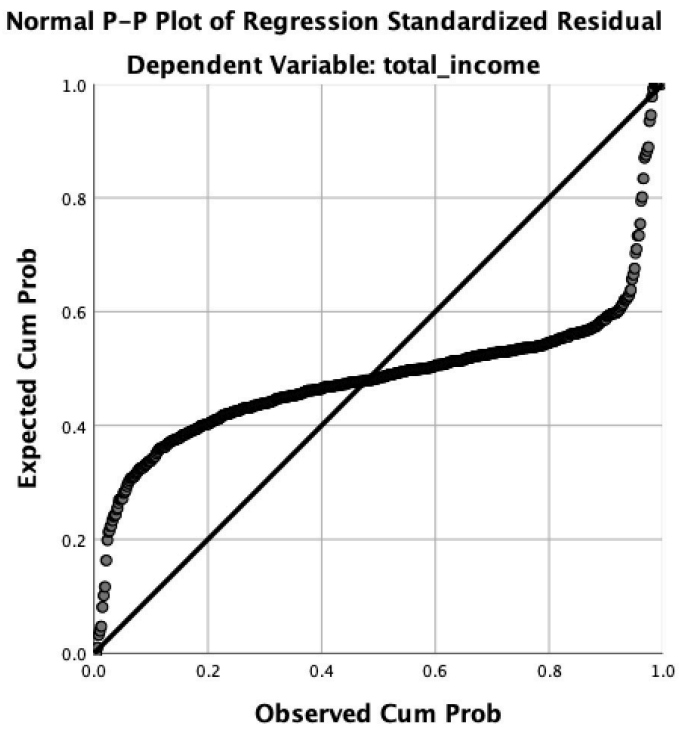
ANOVA表中的“Sig.”检验R

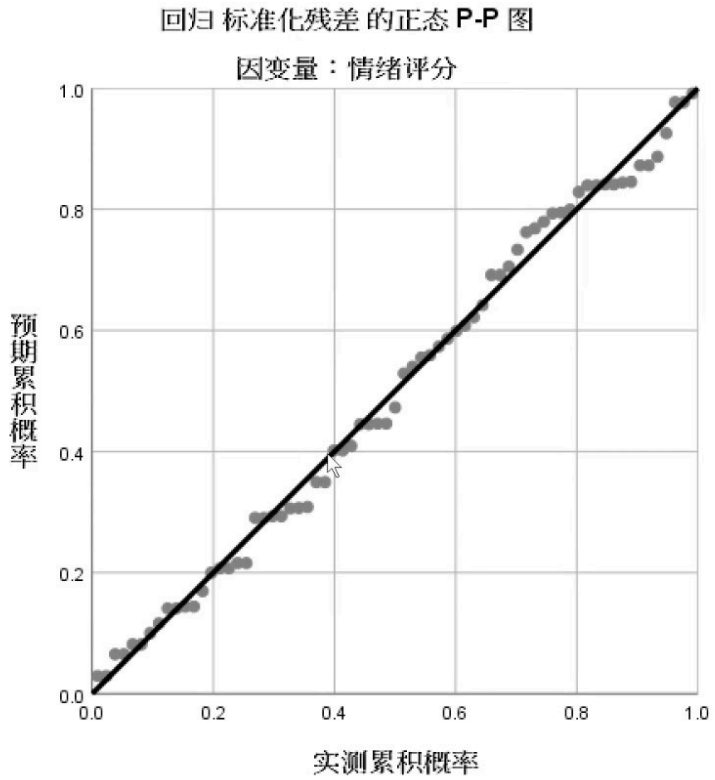
Coefficients[a]

Model		Unstandardized Coefficients B	Unstandardized Coefficients Std. Error	Standardized Coefficients Beta	t	Sig.	Collinearity Statistics Tolerance	Collinearity Statistics VIF
1	(Constant)	-152253.37	38336.452		-3.972	.000		
	webcam	1746.758	26343.983	.002	.066	.947	.747	1.339
	gender0male1female	8135.094	25751.277	.008	.316	.752	.717	1.396
	followers	.184	.007	.660	25.494	.000	.750	1.333
	level	3311.734	707.132	.134	4.683	.000	.610	1.638
	total_time	133.202	94.121	.036	1.415	.157	.756	1.322

a. Dependent Variable: total_income

Multiple Linear Regression in SPSS





残差是否符合正态分布？在P-P Plot上点基本分布在直线上才符合残差正态分布，如右图

Multiple Linear Regression – APA Style

Multiple linear regression was used to test if [predictor variable 1], [predictor variable 2], … significantly predicted [response variable].

The fitted regression model was: [fitted regression equation]

The overall regression was statistically significant (R^2 = [R2 value], $F(df_{regression}, df_{residual})$ = [F-value], p = [p-value]).

It was found that [predictor variable 1] significantly predicted [response variable] (β = [β-value], p = [p-value]).

It was found that [predictor variable 2] did not significantly predict [response variable] (β = [β-value], p = [p-value]).

Multiple Linear Regression – APA Style - Table

Table 2

Regression Coefficients of Leader Sleep on Charismatic Leadership

Variable	Model 1			Model 2		
	B	β	*SE*	*B*	β	*SE*
Constant	2.65**		.31	2.76		
Leader gender[a]	−.11	−.07	.16	−.09	−.06	.15
Leader sleep condition[b]				−.36**	−.24	.15
R^2	.09				.14	
ΔR^2					.05*	

use of specific notes

Note. $N = 88$. We examined the impact of leader sleep condition (control vs. sleep deprived) on ratings of charismatic leadership. In Model 1, we entered the control variables of gender and video length to predict leader charisma. In Model 2, we entered sleep condition as a predictor.

[a]Male = 1, female = 2. [b]Control condition = 0, sleep-deprived condition = 1.

*$p < .05$. **$p < .01$.

Table 7.16 Sample Regression Table, With Confidence Intervals in Brackets

Table 4

Regressions of Associations Between Marital Satisfaction and Average Levels of Marital Behavior

Variable	*B*	*SE*	*t*	*p*	95% CI
			Angry behavior		
Actor					
H → H	−98.90	40.20	−2.46	.016	[− 179.1, −18.7]
W → W	−.87.11	30.87	−2.82	.006	[−148.7, −25.6]
Partner					
W → H	−76.18	39.43	−1.93	.057	[−154.8, 2.4]
H → W	−91.80	38.16	−2.41	.019	[−167.9, −15.7]
			Disregard		
Actor					
H → H	−38.62	27.86	−1.39	.170	[−94.2, 16.9]
W → W	−47.54	26.99	−1.76	.082	[−101.4, 6.3]
Partner					
W → H	−82.81	32.01	−2.59	.012	[−146.6, −19.0]
H → W	−79.36	27.16	−2.92	.005	[−133.5, −25.2]
			Distancing		
Actor					
H → H	−47.42	24.72	−1.92	.059	[−96.7, 1.9]
W → W	3.04	23.48	0.13	.897	[−43.8, 49.8]
Partner					
W → H	−0.05	23.91	0.00	.998	[−47.7, 47.6]
H → W	−53.50	24.47	−2.19	.032	[−102.3, −4.7]

square brackets around confidence intervals

Note. CI = confidence interval; H → H = husband-as-actor effect on the husband's own marital satisfaction; W → W = wife-as-actor effect on the wife's own marital satisfaction; W → H = wife-as-partner effect on the husband's satisfaction; H → W = husband-as-partner effect on the wife's satisfaction.

Multiple Linear Regression – APA Style - Table

confidence intervals in separate columns

Table 3

Moderator Analysis: Types of Measurement and Study Year

Effect	Estimate	*SE*	95% CI		*p*
			LL	*UL*	
Fixed effects					
Intercept	.119	.040	.041	.198	.003
Creativity measurement[a]	.097	.028	.042	.153	.001
Academic achievement measurement[b]	−.039	.018	−.074	−.004	.03
Study year[c]	.0002	.001	−.001	.002	.76
Goal[d]	−.003	.029	−.060	.054	.91
Published[e]	.054	.030	−.005	.114	.07
Random effects					
Within-study variance	.009	.001	.008	.011	<.001
Between-study variance	.018	.003	.012	.023	<.001

Note. Number of studies = 120, number of effects = 782, total *N* = 52,578. CI = confidence interval; *LL* = lower limit; *UL* = upper limit.

[a]0 = self-report, 1 = test. [b]0 = test, 1 = grade point average. [c]Study year was grand centered.

[d]0 = other, 1 = yes. [e]0 = no, 1 = yes.