Independent and Paired Samples T-Tests

The independent-samples t-test and the paired-samples t-test are two versions of the same statistical model. Both of them measure differences between two groups of subjects. For example: Do men and women differ in their salaries? Do athletes and non-athletes differ in life expectancy? Is the "freshman fifteen" real (i.e., do students entering their freshman year differ in bodyweight from students exiting their freshman year)? If you want to know whether there's a difference in your *sample* (the exact people you measured), just look at the mean values. That's all you need to know. The t-test tells you much more than this. It is an inferential statistic. It makes inferences about the larger population. Using only the data in your sample, what can you infer about differences in the larger population?

Independent and Dependent *Variables* in a T-Test

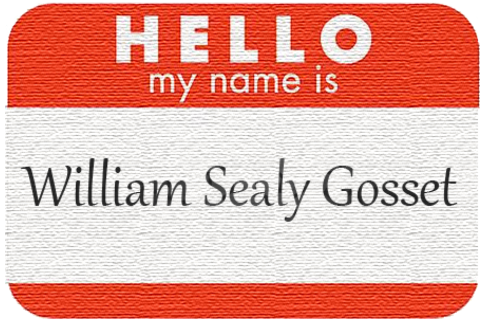
Your independent (or "grouping") variable must be dichotomous (categorical with exactly two categories, e.g., male or female). You can only measure the differences between two groups. If you want to compare more than two groups, you have to use ANOVA for that.

Your dependent variable must be continuous (can take on any number of values, e.g., salary or body fat percent or GPA). If you want to measure differences in categorical data, you have to use a chi-squared test for that.

Independent-Samples *vs.* Paired-Samples

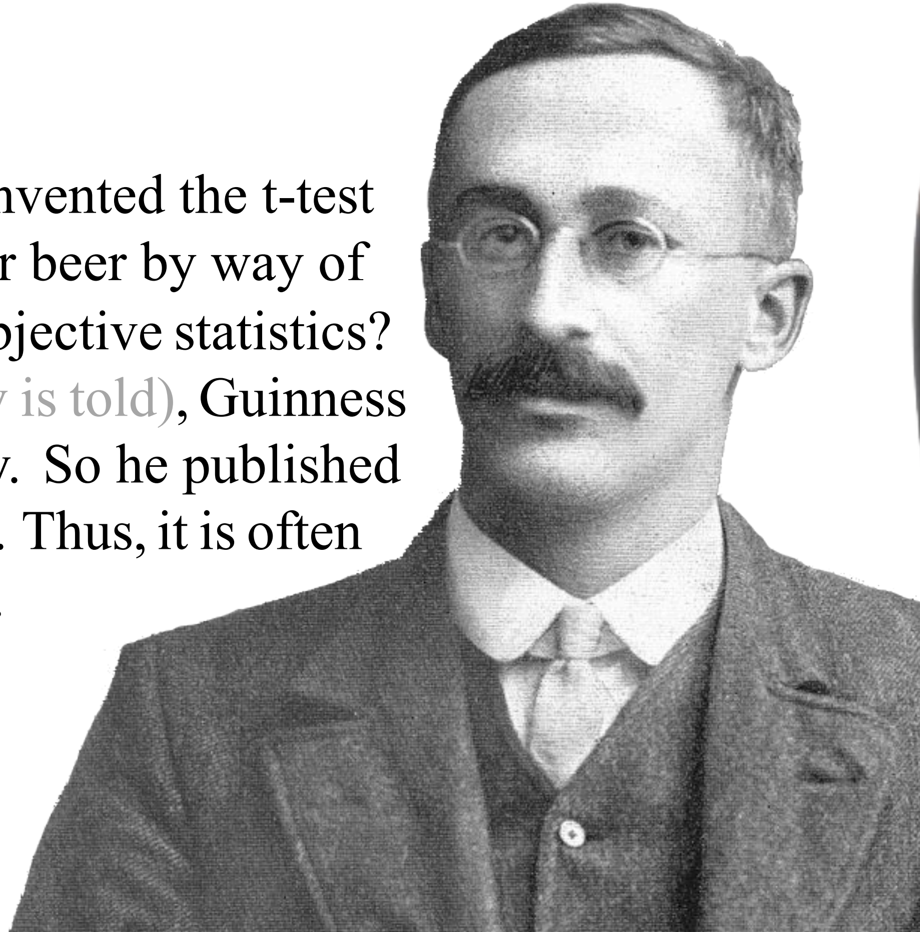
In an independent-samples t-test, you're evaluating two groups that are independent of each other. Males *vs*. females. Dogs *vs*. cats. People taking a drug *vs*. people taking a placebo. Males don't become females; dogs don't become cats; the placebo folks never take the drug. The samples you're comparing are independent. There's no overlap.

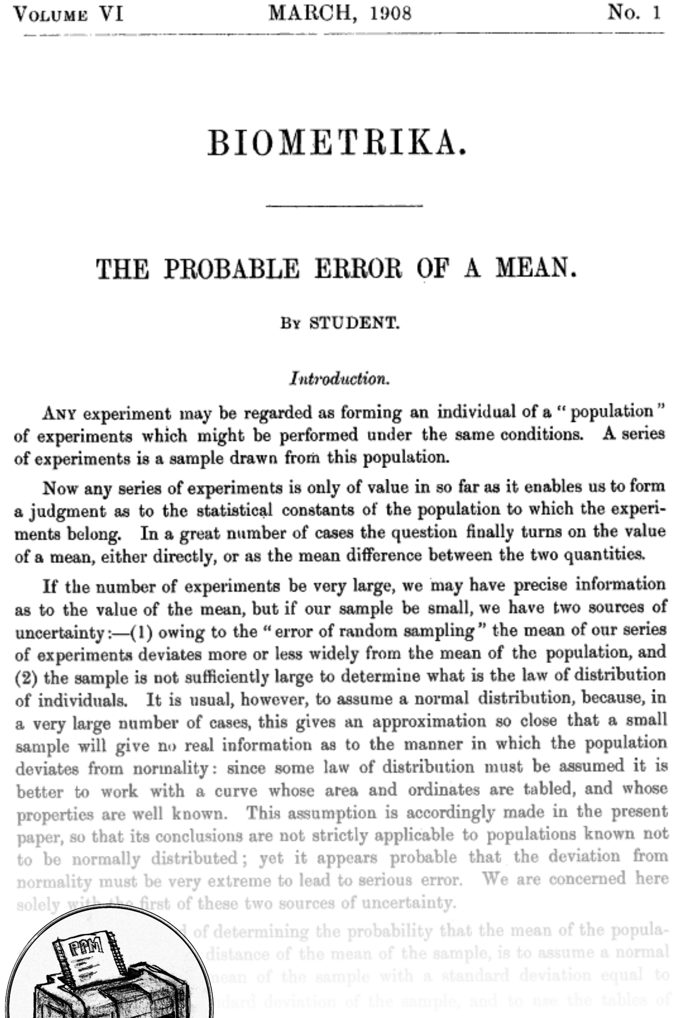
In a paired-samples t-test, you're comparing the same group of subjects at two different times. If it's a drug trial, you measure the subjects' blood levels of some biomarker before taking the drug and then measure it again afterward. Same subjects measured pre and post. Is there a difference? Or the freshman fifteen: same students measured pre and post.



The Origin of T-Tests

Gosset was a Guinness employee who invented the t-test to improve beer. Should we produce our beer by way of assumption or should we employ cold, objective statistics? Gosset went with math. But (as the story is told), Guinness had an “employees can’t publish” policy. So he published his formula under a pseudonym: Student. Thus, it is often called “Student’s t-test” in the literature.





VOLUME VI MARCH, 1908 No. 1

BIOMETRIKA.

THE PROBABLE ERROR OF A MEAN.

BY STUDENT.

Introduction.

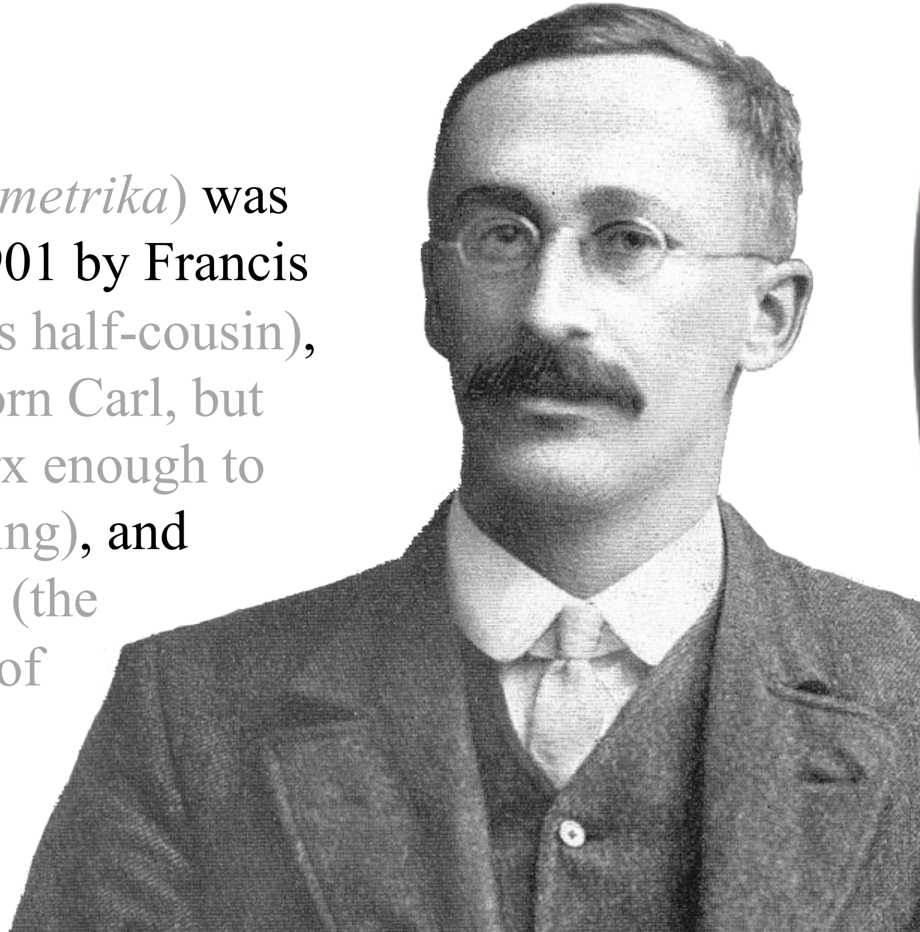
ANY experiment may be regarded as forming an individual of a "population" of experiments which might be performed under the same conditions. A series of experiments is a sample drawn from this population.

Now any series of experiments is only of value in so far as it enables us to form a judgment as to the statistical constants of the population to which the experiments belong. In a great number of cases the question finally turns on the value of a mean, either directly, or as the mean difference between the two quantities.

If the number of experiments be very large, we may have precise information as to the value of the mean, but if our sample be small, we have two sources of uncertainty:—(1) owing to the "error of random sampling" the mean of our series of experiments deviates more or less widely from the mean of the population, and (2) the sample is not sufficiently large to determine what is the law of distribution of individuals. It is usual, however, to assume a normal distribution, because, in a very large number of cases, this gives an approximation so close that a small sample will give no real information as to the manner in which the population deviates from normality: since some law of distribution must be assumed it is better to work with a curve whose area and ordinates are tabled, and whose properties are well known. This assumption is accordingly made in the present paper, so that its conclusions are not strictly applicable to populations known not to be normally distributed; yet it appears probable that the deviation from normality must be very extreme to lead to serious error. We are concerned here solely with the first of these two sources of uncertainty.

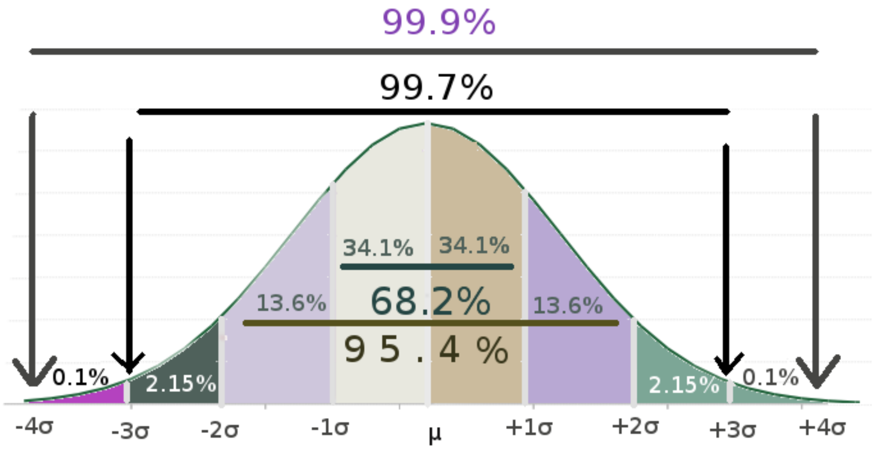
The Origin of T-Tests

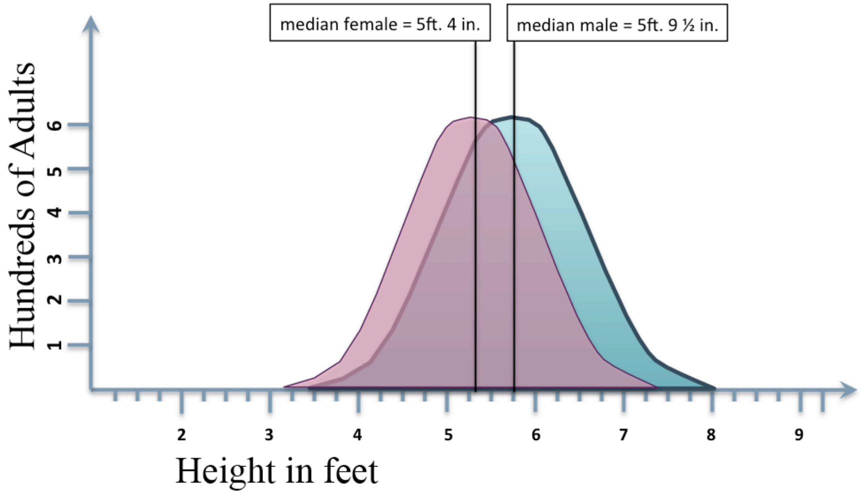
The journal (*Biometrika*) was established in 1901 by Francis Galton (Darwin's half-cousin), Karl Pearson (born Carl, but adored Karl Marx enough to change the spelling), and Raphael Weldon (the least interesting of the three).



The Value of a T-Test

It's a very crude test. Do two normally distributed data sets differ? That's all it tells you.





The Value of a T-Test

It's a very crude test. Do two normally distributed data sets differ? That's all it tells you. It controls for nothing. Sometimes that's good enough. Sometimes t-tests are your final statistical models. But usually those outputs are only used as introductory descriptions of your sample. When you don't control for any confounding variables, the only question you can ask is: Is there a difference between group A and group B?

The Value of a T-Test

Consider this phenomenon:

Group A: Lots of money.
Group B: Very little money.

T-Test: Group A lives longer.

Does money itself explain the difference in lifespan or is there more to the story?

The Gross Inequality of Death in America

THE NEW REPUBLIC By **ROGE KARMA** | May 10, 2019

The richest Americans live 10-15 years longer than the poorest Americans. It will take a lot more than Medicare for All to close the gap.

One of the most disquieting facts about life in the United States today is that the richest American men live 15 years longer than the poorest men, while for women it's 10 years. Put a different way, the life expectancy gap between rich and poor in the U.S. is wider than the gap between the average American and the average Yemeni or Ethiopian.

The Value of a T-Test

Consider this slightly-more sensitive phenomenon:

Is there a difference in wages between men and women: yes or no?

Is that a useful question to ask?

Equal Pay Act was signed by Kennedy on June 10, 1963. Beginning June 11, 1964, there must be differences in work to justify differences in pay.

PacificPaperMill.com

CNN BUSINESS

78 cents on the dollar: The facts about the gender wage gap

by Sara Ashley O'Brien @saraashleyo

April 14, 2015: 7:59 AM ET

What is the gender wage gap? 78 cents project

While exact statistics vary, it's been shown that when you compare the earnings of all full-time working women in the U.S. to all full-time working U.S. men, women earn 78 cents to every $1 that men earn.

Published Wed, Nov 28 2018 • 5:20 PM EST • Updated Tue, Dec 4 2018 • 4:35 PM EST

Women in America make just 78 cents for every dollar earned by American men. It's on all of us to end the wage gap. Are you ready to get to work?

CNBC make it

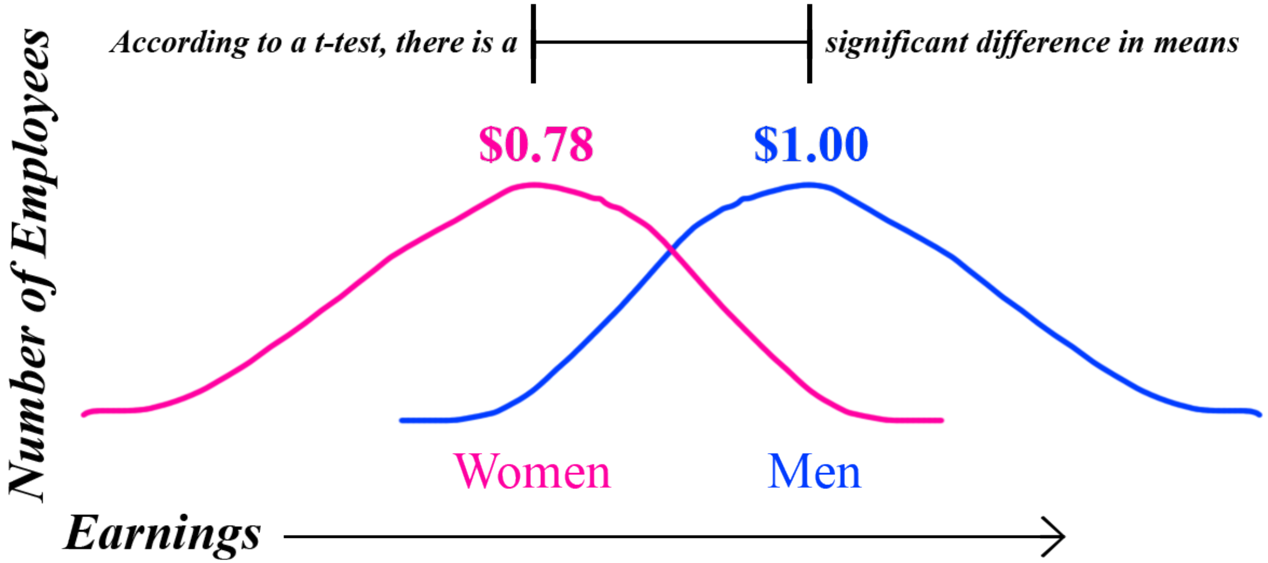
CLOSING THE GAP

A new study suggests women earn about half what men earn

NOVEMBER 28, 2018 | *Emma Newburger, NBR, CNBC.com*

The Value of a T-Test

According to a t-test, there's a difference, but what might the t-test fail to tell us?



The Value of a T-Test

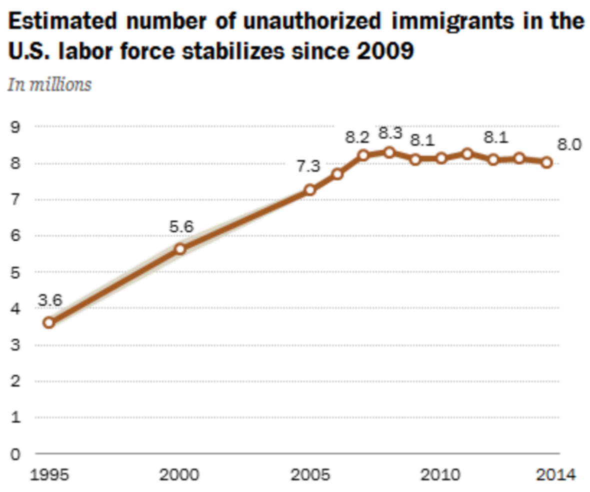
Consider: It is illegal (fines and risk of jail) to hire unauthorized immigrants and yet:

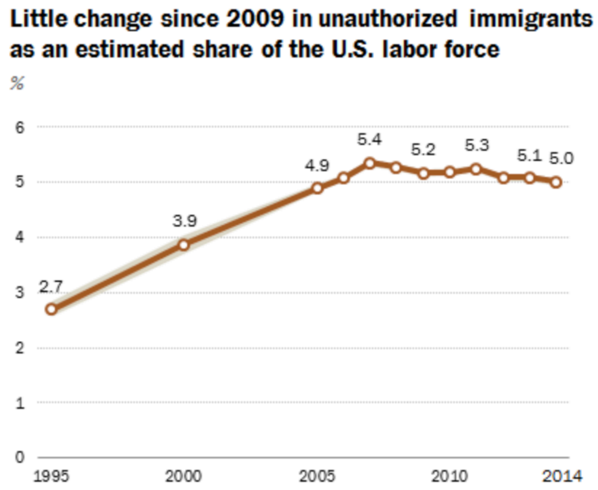
Pew Research Center *Hispanic Trends*

Size of U.S. Unauthorized Immigrant Workforce Stable After the Great Recession

NOVEMBER 3, 2016

BY JEFFREY S. PASSEL AND D'VERA COHN





Illegal workers account for ~6% (~8 million) of the American workforce.

The Value of a T-Test

Also consider: even marginal savings export jobs:

CNN Money

In China, outsourcing is no longer cheap

As China makes big moves to improve its environmental and labor conditions, U.S. companies that manufacture there face soaring costs.

China Trade, Outsourcing and Jobs

Growing U.S. trade deficit with China cost 3.2 million jobs between 2001 and 2013, with job losses in every state

Report • By **Will Kimball** and **Robert E. Scott** • December 11, 2014

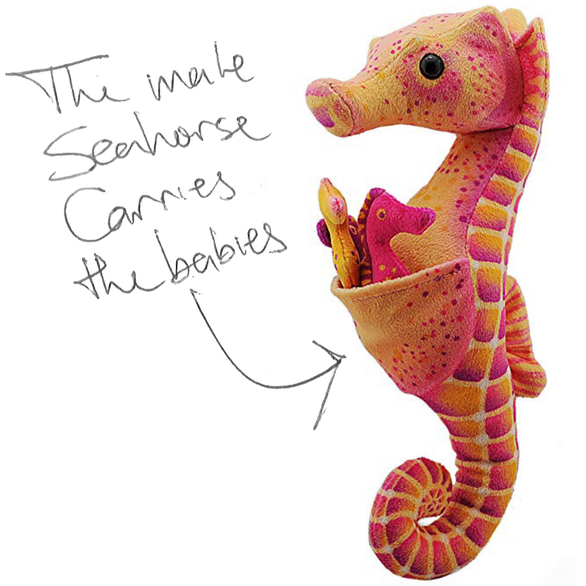
In percentage terms, the jobs lost or displaced due to the growing goods trade deficit with China in the 10 hardest-hit states ranged from 2.44 percent to 3.67 percent of the total state employment: Oregon (62,700 jobs lost or displaced, equal to 3.67 percent of total state employment), California (564,200 jobs, 3.43 percent), New Hampshire (22,700 jobs, 3.31 percent), Minnesota (83,300 jobs, 3.05 percent), Massachusetts (97,200 jobs, 2.96 percent), North Carolina (119,600 jobs, 2.85 percent), Texas (304,700 jobs, 2.66 percent), Rhode Island (13,200 jobs, 2.58 percent), Vermont (8,200 jobs, 2.51 percent), and Idaho (16,700 jobs, 2.44 percent).

The Value of a T-Test

One wonders: If we could hire Americans and legally pay them 22% less, why would employers hire anyone who wasn't a woman? A t-test tells us this is the case and yet men still exist in the workplace. I'm confused.

The Value of a T-Test

The wage gap exists in every country:



Country	Gender gap (%)	Year
Australia	15.38	2014
Austria	17.73	2014
Belgium	5.91	2013
Canada	18.63	2015
Chile	16.67	2013
Czech Republic	16.46	2013
Denmark	6.77	2013
Estonia	26.60	2010
Finland	19.61	2014
France	13.67	2012
Germany	17.08	2014
Greece	9.09	2014
Hungary	9.52	2015
Iceland	13.59	2014
Ireland	15.17	2014
Israel	21.83	2011
Italy	5.56	2014
Japan	25.87	2014

Country	Gender gap (%)	Year
Korea	36.65	2014
Latvia	13.33	2010
Lithuania	6.96	2010
Luxembourg	4.97	2010
Mexico	16.67	2015
Netherlands	18.60	2010
New Zealand	6.08	2014
Norway	7.12	2015
Poland	11.07	2014
Portugal	18.88	2014
Slovakia	13.38	2015
Slovenia	11.63	2010
Spain	8.65	2012
Sweden	13.42	2013
Switzerland	14.50	2014
Turkey	20.06	2010
United Kingdom	16.93	2015
United States	18.88	2015

The Value of a T-Test

The Gender Pay Gap: A Cross-Country Analysis

Solomon W. Polachek
Department of Economics
State University of New York at Binghamton
Binghamton, New York 13902
607-777-6866
polachek@binghamton.edu

and

Jun (Jeff) Xiang
Department of Political Science
University of Rochester
Rochester, New York 14627

February 2006

The U.S. male-female wage gap is now about 78%, but an intriguing pattern emerges when examining this gender wage gap for different marital status groups. For single men and women the wage gap is generally less than 10%, implying single women on average earn over 90% what men earn. But married women earn *far* less than married men. Here the wage ratio is typically in the 60% to 70% range implying a 30-40% wage gap. Further deconstruction illustrates that children play a major role in the gender wage gap. Married women with children earn less than married women without children (Harkmess and Waldfogel, 2003). Married women who space their births widely apart receive even lower wages (Polachek, 1975b). Opposite patterns regarding marital status and family hold for men. Married men with children earn more, and spacing children at wide intervals is associated with even higher earnings (Polachek, 1975b). Thus the wage gap varies by marital status, children, and spacing of children. As it turns out, these demographic variables are more important predictors of the gender wage gap than any other explanatory factors.

There is now more than ample evidence of these family effects. Numerous studies corroborate this so-called "motherhood" penalty" For example, Korenman and Neumark (1992) find that cross-sectional ordinary least squares and first-difference estimates understate the negative effect of children on wages. Waldfogel (1998) shows that having children lowers a women's pay by about 10%, after controlling for age, education, experience, race, ethnicity and marital status. Budig and England (2001) find about a 7% wage penalty per child. Using the National Longitudinal Survey Panel, Baum (2002: 2) confirms the finding that "interrupting work to give birth has a negative effect on wages" but that "this negative effect is at least partially eliminated when [controlling for] whether the mother returns to work at her pre-childbirth job."

The Value of a T-Test

payscale.com/data/gender-pay-gap

WOMEN ARE STILL PAID LESS IN 2019

UNCONTROLLED GENDER PAY GAP

THIS MEASURES MEDIAN SALARY FOR ALL MEN AND ALL WOMEN

WOMEN EARN

79¢

FOR EVERY $1 EARNED BY MEN

CONTROLLED GENDER PAY GAP

THIS MEASURES MEDIAN SALARY FOR MEN AND WOMEN WITH THE SAME JOB AND QUALIFICATIONS.

WOMEN EARN

98¢

FOR EVERY $1 EARNED BY MEN

The Value of a T-Test

payscale.com/data/gender-pay-gap

Since we have started tracking the gender pay gap, the difference between the earnings of women and men has shrunk. But significant disparity in how men and women are paid still remains. The uncontrolled gender pay gap, which takes the ratio of median earnings of all women to all men, decreased by $0.05 since 2015. However, women still make only $0.79 for every dollar men make in 2019.

The controlled gender pay gap, which controls for a number of factors such as job title, years of experience, industry and location so that the only differentiation between workers is their gender, shrunk by just $0.008 since 2015. Women now make $0.98 for every dollar an equivalent man makes.

Why Do We Use T-Tests?

We need to note and define differences in our samples before using more precise statistics (e.g., regression) to *explain* those differences. The results of the t-tests will appear in Table 1. In the far left column are the values from frequencies and descriptives.

Table 1: *Subject baseline characteristics*

	Total	Men	Women	Sig.
N	45			
Age (years)	20.2 ± 0.7			
BMI (kg/m²)	26.5 ± 2.8			
GPA	2.9 ± 0.7			
Nightly sleep (hours)	7.5 ± 1.4			
Employed (%)	22.5%			
Weekly work (hours)	4.0 ± 5.2			
Academic Scholarship (%)	12.9%			
Athletic Scholarship (%)	11.5%			

Why Do We Use T-Tests?

The results of your t-tests will fill this geography in your Table 1.

It is a comparison of two subsamples (e.g., men and women). Do they differ in a continuous variable (e.g., age, BMI, GPA, etc.)?

Table 1: *Subject baseline characteristics*

	Total	Men	Women	Sig.
N	45			
Age (years)	20.2 ± 0.7			
BMI (kg/m²)	26.5 ± 2.8			
GPA	2.9 ± 0.7			
Nightly sleep (hours)	7.5 ± 1.4			
Employed (%)	22.5%			
Weekly work (hours)	4.0 ± 5.2			
Academic Scholarship (%)	12.9%			
Athletic Scholarship (%)	11.5%			

The Outputs From T-Tests

Mean

Standard Deviation

T statistic

P-Value

Test scores are: 67, 70, 78, 83, 86, 86, 89, 92, 94, 95

Mean = 840 ÷ 10 = 84.0

PacificPaperMill.com

The Outputs From T-Tests

Mean

Standard Deviation

T statistic

P-Value

Distance from 84: 17 14 6 1 2 2 5 8 10 11

Test scores are: 67, 70, 78, 83, 86, 86, 89, 92, 94, 95

Standard deviation is the square root of the variance of those scores.

Mean = 840 ÷ 10 = 84.0

Standard deviation is **not** the average distance from the mean, as in: 67 is 17 from 84, 70 is 14 from 84, 78 is 6 from 84, and so on. The average of those distances is 7.6. That's **not** the standard deviation. Rather, it is the square of the root of the variances.

The Outputs From T-Tests

Mean

Standard Deviation

T statistic

P-Value

Test scores are: 67, 70, 78, 83, 86, 86, 89, 92, 94, 95

Standard deviation is the square root of the variance of those scores.

Mean = 840 ÷ 10 = 84.0

$(67 - 84)^2 = 289$
$(70 - 84)^2 = 196$
$(78 - 84)^2 = 36$
$(83 - 84)^2 = 1$
$(86 - 84)^2 = 4$

$(86 - 84)^2 = 4$
$(89 - 84)^2 = 25$
$(92 - 84)^2 = 64$
$(94 - 84)^2 = 100$
$(95 - 84)^2 = 121$

Variance of the mean:

$$\frac{289 + 196 + 36 + 1 + 4 + 4 + 25 + 64 + 100 + 121}{10}$$

= 84.0 (same as the mean but it isn't always the same)

The Outputs From T-Tests

- Mean
- Standard Deviation →
- T statistic
- P-Value

Test scores are: 67, 70, 78, 83, 86, 86, 89, 92, 94, 95

Standard deviation is the square root of the variance of those scores.

Mean = 840 ÷ 10 = 84.0

$(67-84)^2 = 289$ $(86-84)^2 = 4$

$(70-84)^2 = 196$ $(89-84)^2 = 25$

$(78-84)^2 = 36$ $(92-84)^2 = 64$

$(94-84)^2 = 100$

$(86-84)^2 = 4$ $(95-84)^2 = 121$

Population standard deviation:

$\sqrt{84} = 9.2$

This is the number to use if you've tested everyone in the population.

Variance of the mean:

$$\frac{289 + 196 + 36 + 1 + 4 + 4 + 25 + 64 + 100 + 121}{10}$$

= **84.0** (same as the mean but it isn't always the same)

The Outputs From T-Tests

- Mean
- Standard Deviation
- T statistic
- P-Value

Test scores are: 67, 70, 78, 83, 86, 86, 89, 92, 94, 95

Standard deviation is the square root of the variance of those scores.

Mean = 840 ÷ 10 = 84.0

$(67-84)^2 = 289$ $(86-84)^2 = 4$

$(70-84)^2 = 196$ $(89-84)^2 = 25$

$(78-84)^2 = 36$ $(92-84)^2 = 64$

$(94-84)^2 = 100$

$(86-84)^2 = 4$ $(95-84)^2 = 121$

***Sample* standard deviation:**

$$\frac{289 + 196 + 36 + 1 + 4 + 4 + 25 + 64 + 100 + 121}{9}$$

= **83.333333333333**

Sample standard deviation:

$\sqrt{83.33} = 9.7$

This is the number to use if you've tested a sample of subjects from the population.

The Outputs From T-Tests

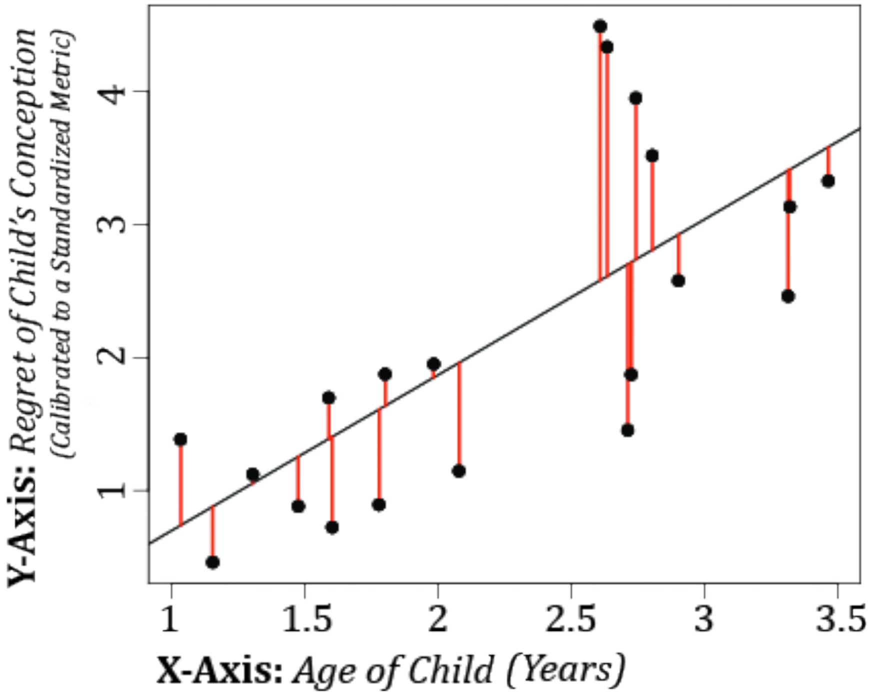
Mean

Standard Deviation

T statistic

P-Value

Don't memorize how it is calculated. Just understand that standard deviation is a measurement of the spread of values in a dataset from their mean value.



The Outputs From T-Tests

Mean

Standard Deviation

T statistic

P-Value

The t-statistic is a ratio of signal-to-noise. A value > 1 means you have more signal than noise; a value < 1 means you have more noise than signal.

The Outputs From T-Tests

Mean

Standard Deviation

T statistic →

P-Value

The t-statistic is a ratio of signal-to-noise. A value > 1 means you have more signal than noise; a value < 1 means you have more noise than signal. **Signal:** difference in means.

Mean 1 — Mean 2

The Outputs From T-Tests

Mean

Standard Deviation

T statistic

P-Value

The t-statistic is a ratio of signal-to-noise. A value > 1 means you have more signal than noise; a value < 1 means you have more noise than signal. **Noise:** how scattered the data are, i.e., what do the bell-shaped distributions look like?

$$\sqrt{\frac{\text{St.Dev.1}^2}{N_1} + \frac{\text{St.Dev.2}^2}{N_2}}$$

The Outputs From T-Tests

Mean

Standard Deviation

T statistic →

P-Value

The t-statistic is a ratio of signal-to-noise.

$$\frac{\text{Mean 1} - \text{Mean 2}}{\sqrt{\frac{\text{St.Dev.1}^2}{N_1} + \frac{\text{St.Dev.2}^2}{N_2}}}$$

The Outputs From T-Tests

Mean

Standard Deviation

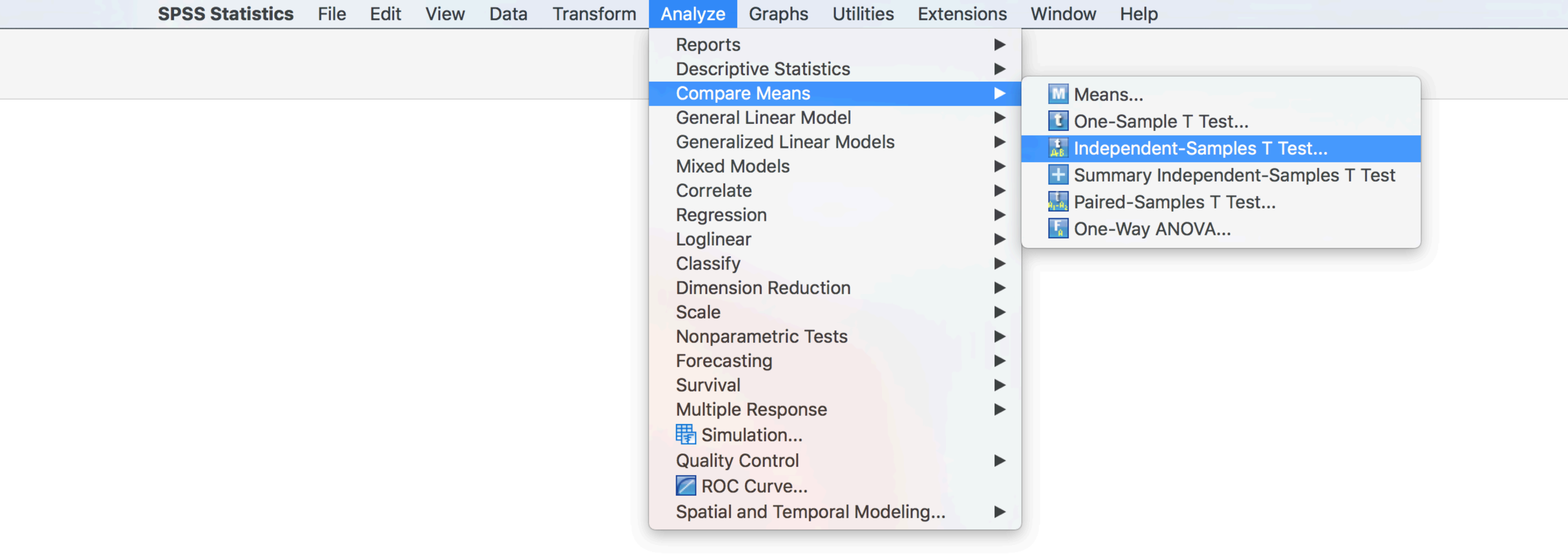
T statistic

P-Value

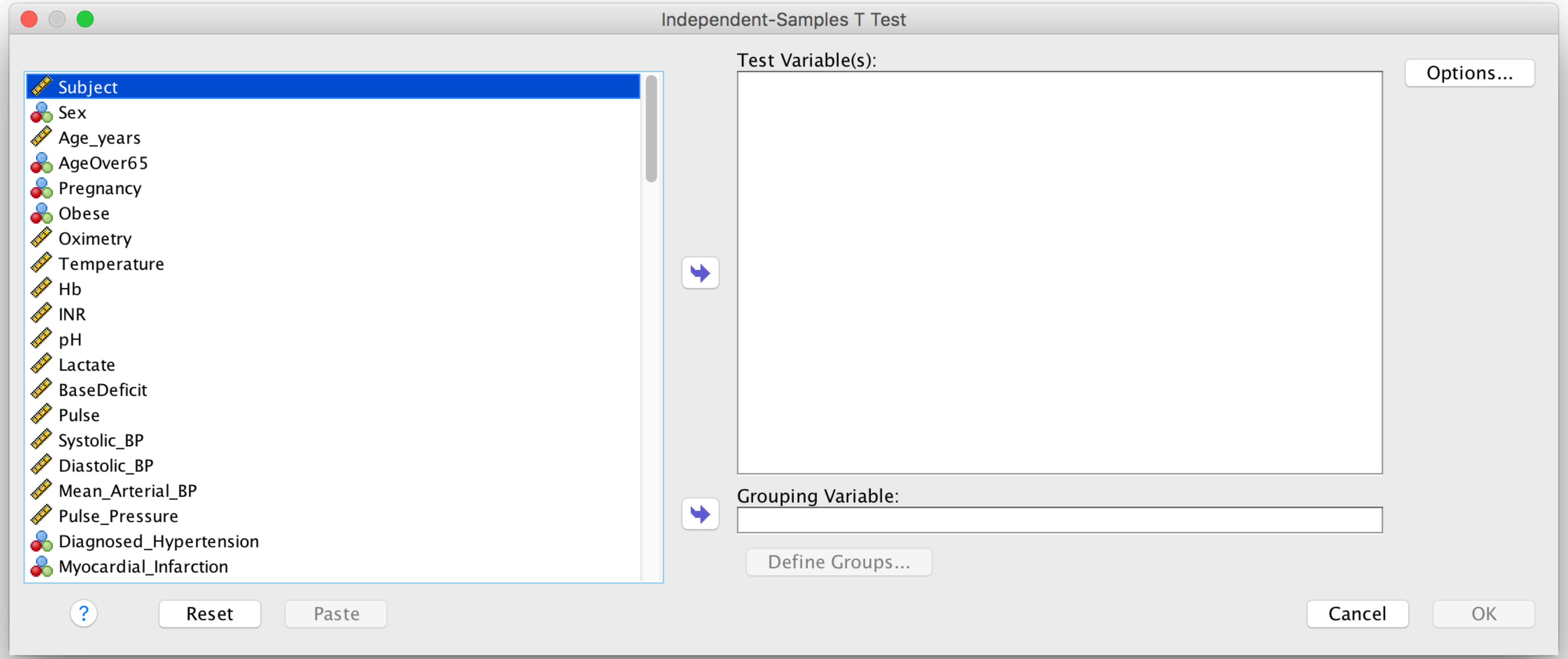
If the null hypothesis is true (no difference between the two groups), there is a p-value% chance you would have observed a difference of the magnitude you observed or something more extreme than that in a sample that size.

SPSS Statistics File Edit View Data Transform Analyze Graphs Utilities Extensions Window Help

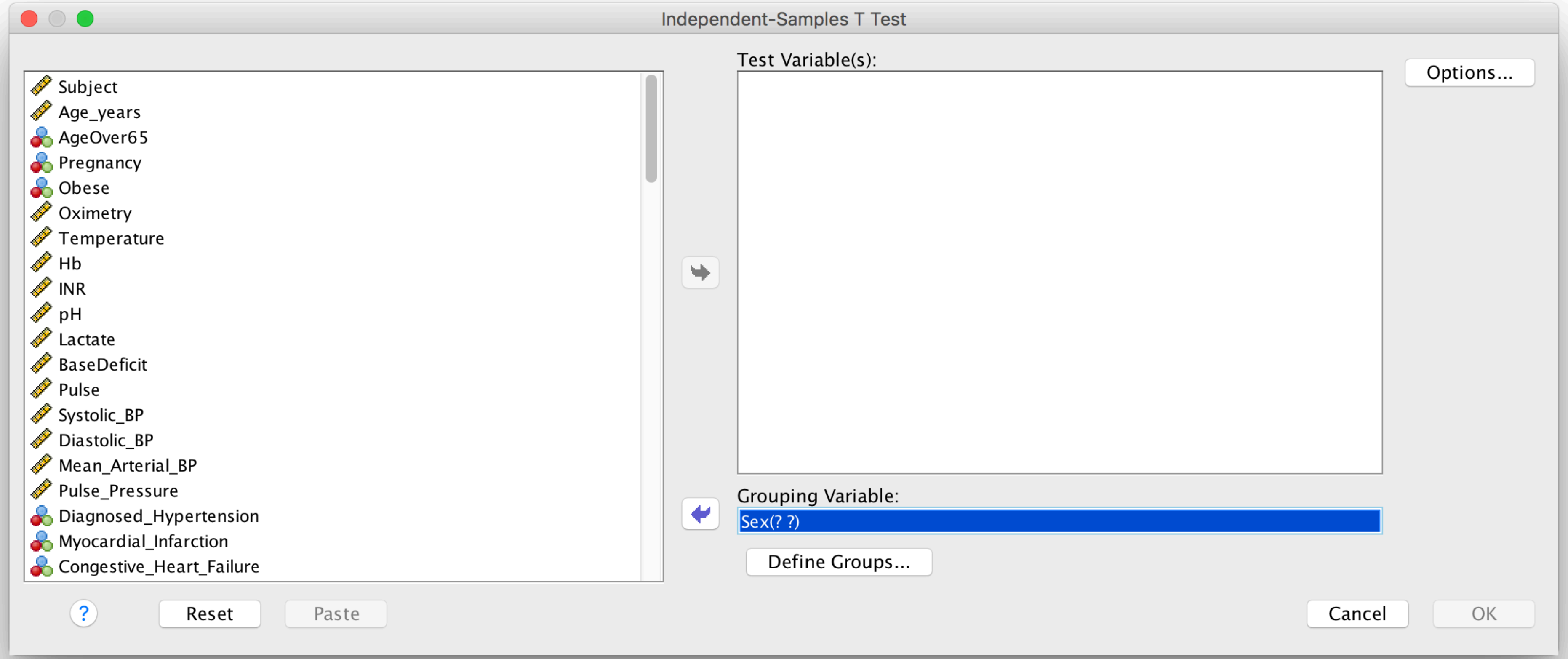
Conducting an Independent-Samples T-Test



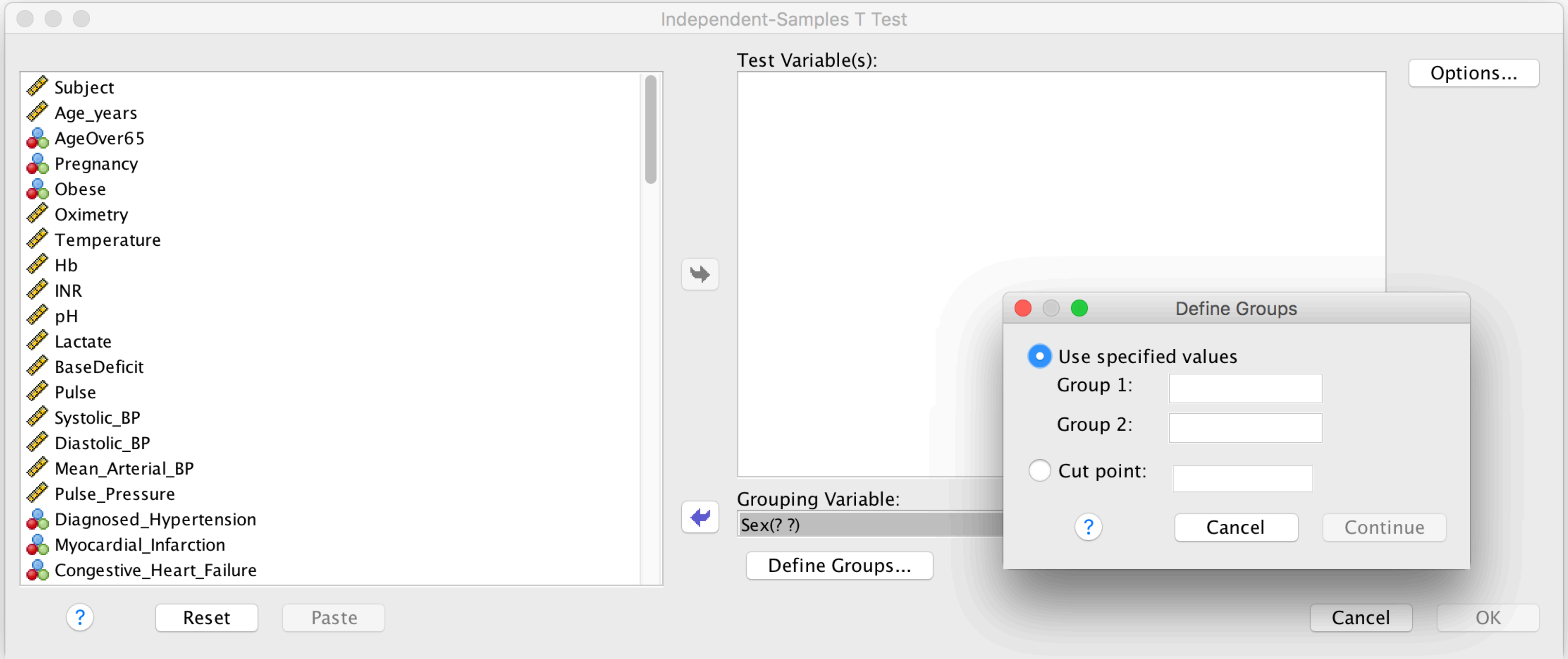
Select Independent-Samples T-Test from the menu bar.



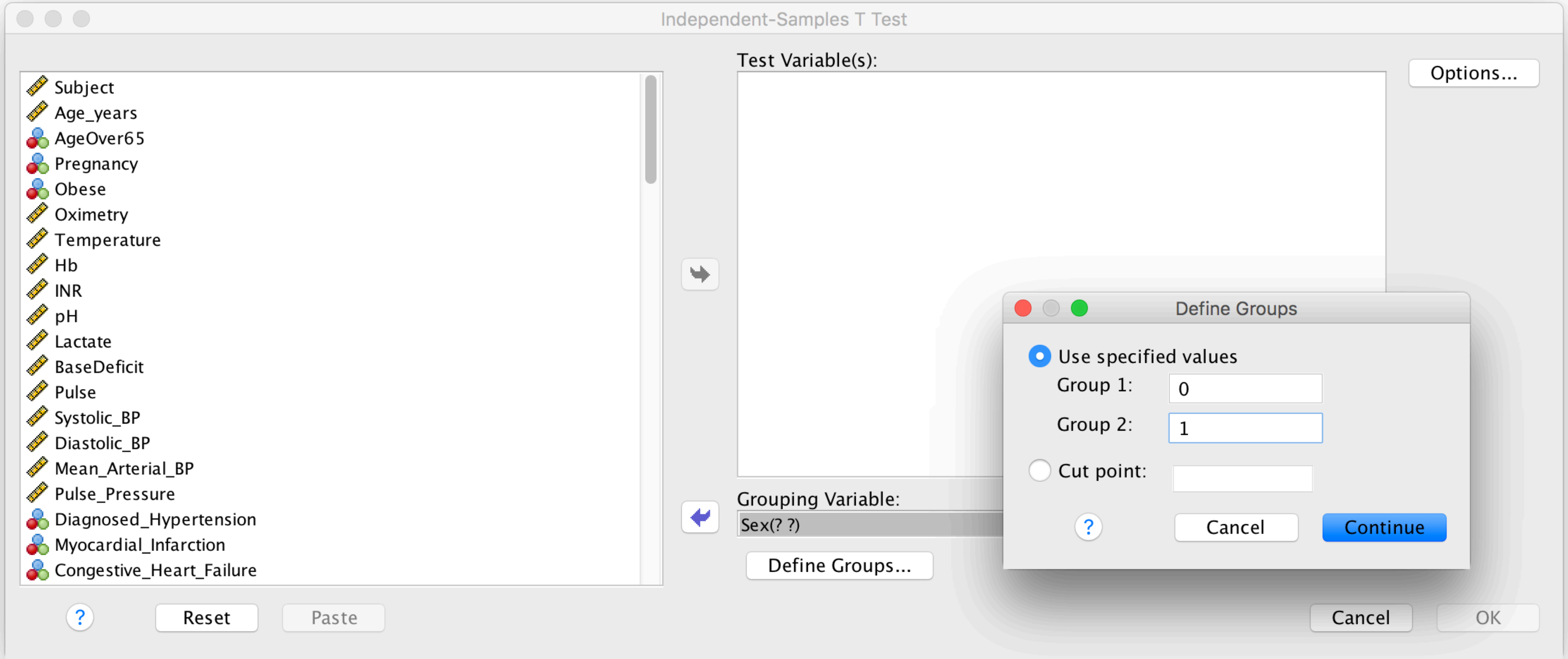
In the left box, you will find every variable in your dataset.



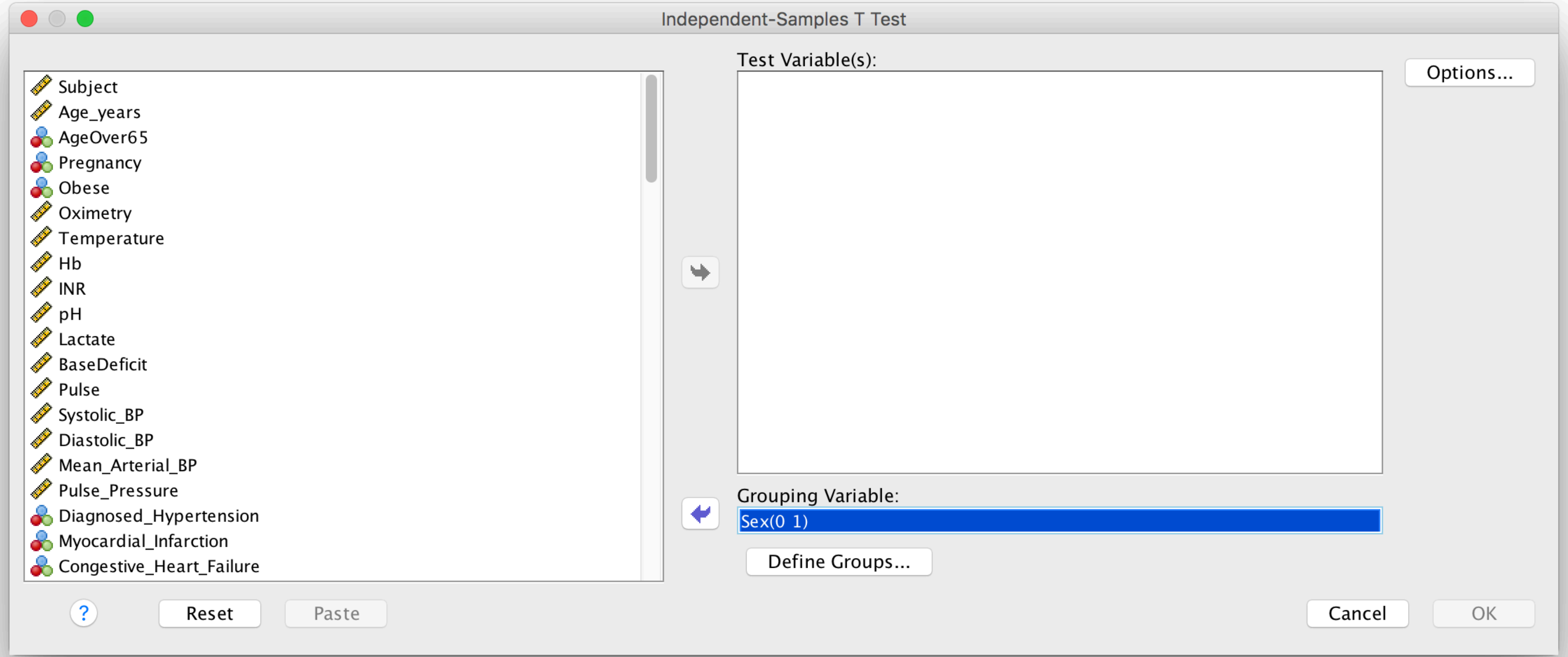
Select your "Grouping Variable" (dichotomous independent variable). This will separate your sample into two subsamples for comparison.



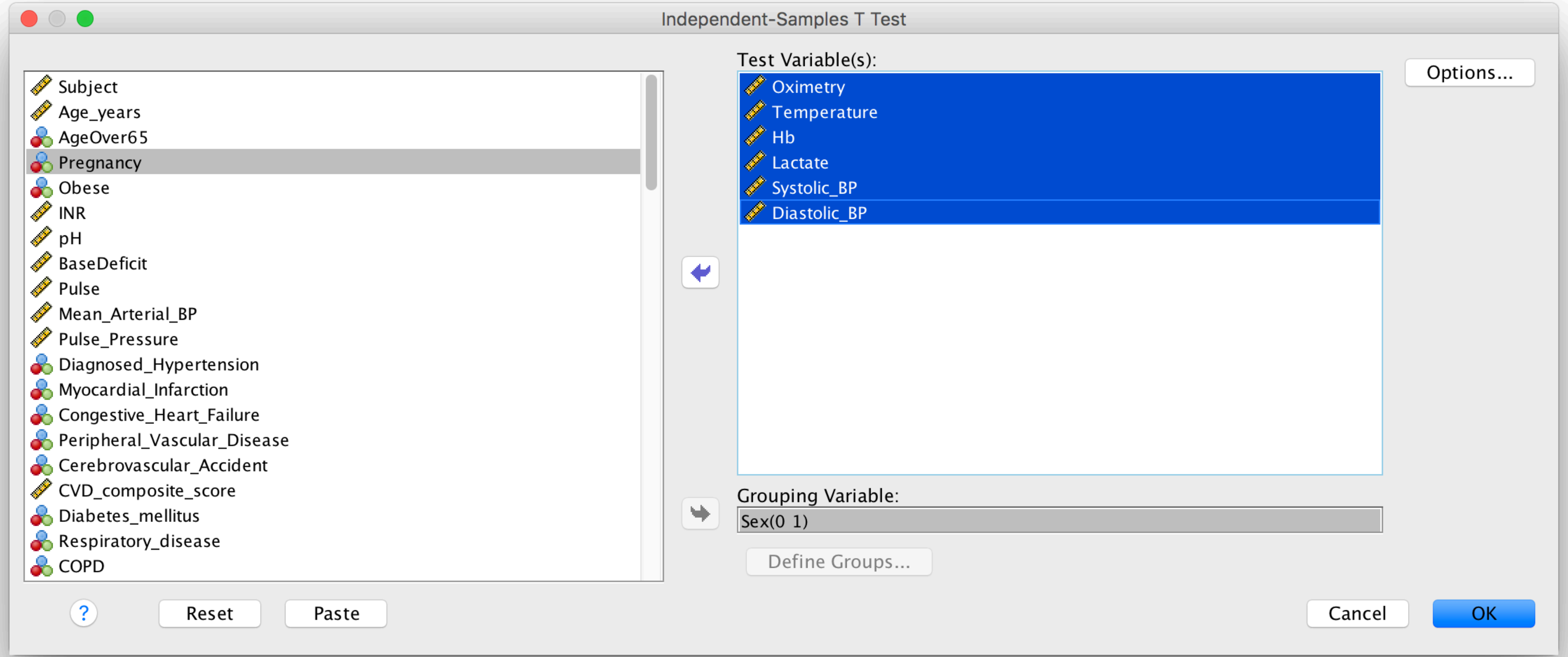
Click “Define Groups” and enter the coded values for each.



You've probably coded it as either 0 and 1 or as 1 and 2. Enter those values into the Group 1 and Group 2 boxes. Then click Continue.



Now select the dependent variables. As many as you want. But you cannot compare categorical variables; every dependent variable must be continuous.



Continuous variables are coded as “Scale” (icon of a ruler). Drag over every continuous variable of interest. Then click OK.

Independent-samples t-tests have two outputs. The first ("Group Statistics") provides means and standard deviations for the subgroups. In this case, men and women are compared across six health outcomes.

Group Statistics

	Sex	N	Mean	Std. Deviation	Std. Error Mean
Oximetry	Male	1261	97.191	2.6708	.0752
	Female	979	96.904	3.1857	.1018
Temperature	Male	1228	98.175	.7606	.0217
	Female	956	98.143	.6919	.0224
Hb	Male	1224	14.2276144	1.82961614	.052296096
	Female	982	12.6125255	1.66433964	.053111212
Lactate	Male	475	2.18656842	1.73361875	.079543884
	Female	256	1.69375000	.968672024	.060542002
Systolic_BP	Male	1273	139.164	21.2778	.5964
	Female	991	140.361	26.6703	.8472
Diastolic_BP	Male	1273	83.410	13.5908	.3809
	Female	991	79.949	14.4829	.4601

The second output contains the significance. There are two Sig. columns though. You report the second one.

Independent Samples Test

		Levene's Test for Equality of Variances		t-test for Equality of Means						
									95% Confidence Interval of the Difference	
		F	Sig.	t	df	Sig. (2-tailed)	Mean Difference	Std. Error Difference	Lower	Upper
Oximetry	Equal variances assumed	9.394	.002	2.323	2238	.020	.2876	.1238	.0448	.5305
	Equal variances not assumed			2.272	1897.904	.023	.2876	.1266	.0394	.5359
Temperature	Equal variances assumed	.252	.616	.986	2182	.324	.0311	.0315	-.0308	.0929
	Equal variances not assumed			.997	2129.911	.319	.0311	.0312	-.0300	.0922
Hb	Equal variances assumed	2.327	.127	21.445	2204	.000	1.61508892	.075312687	1.46739766	1.76278018
	Equal variances not assumed			21.668	2169.541	.000	1.61508892	.074536451	1.46891861	1.76125923
Lactate	Equal variances assumed	10.010	.002	4.207	729	.000	.492818421	.117134561	.262857105	.722779737
	Equal variances not assumed			4.930	728.074	.000	.492818421	.099962810	.296568673	.689068169
Systolic_BP	Equal variances assumed	46.988	.000	-1.188	2262	.235	-1.1971	1.0078	-3.1733	.7792
	Equal variances not assumed			-1.155	1858.934	.248	-1.1971	1.0361	-3.2290	.8349
Diastolic_BP	Equal variances assumed	2.155	.142	5.840	2262	.000	3.4605	.5926	2.2984	4.6226
	Equal variances not assumed			5.794	2059.356	.000	3.4605	.5973	2.2891	4.6318

The first "Sig." column ("Levene's Test for Equality of Variances") tells you whether the variances of your two populations are equal. This is called homoscedasticity. And a significant p-value means you *don't* have homoscedasticity.

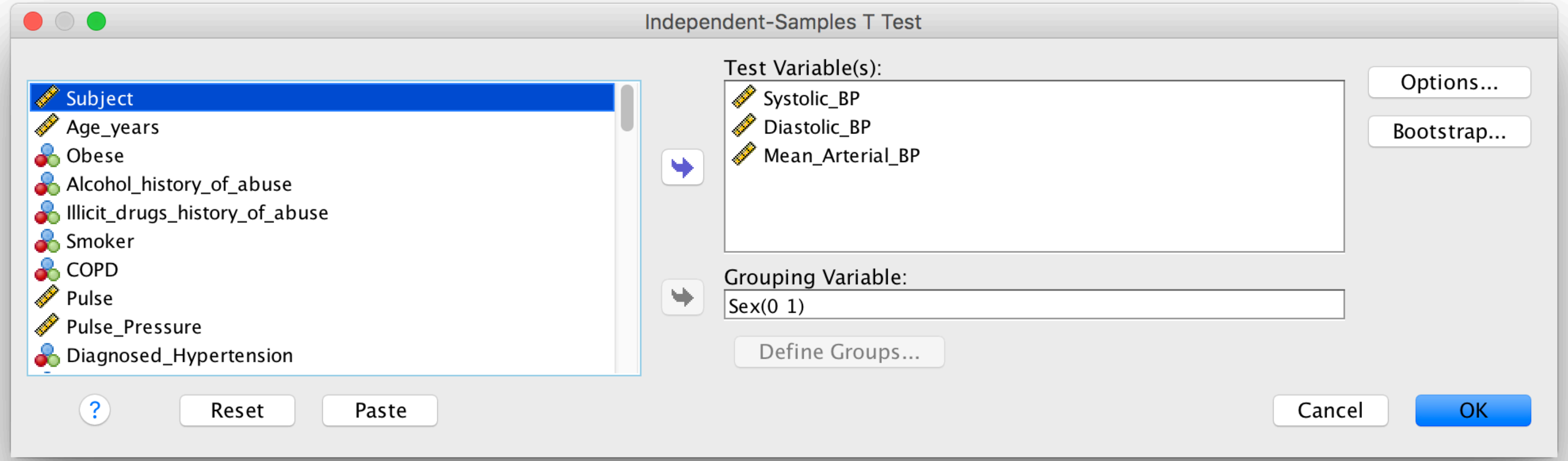
How to select the correct p-value:

Independent Samples Test

		Levene's Test for Equality of Variances		t-test for Equality of Means						
									95% Confidence Interval of the Difference	
		F	Sig.	t	df	Sig. (2-tailed)	Mean Difference	Std. Error Difference	Lower	Upper
Oximetry	Equal variances assumed	9.394	.002	2.323	2238	.020	.2876	.1238	.0448	.5305
	Equal variances not assumed			2.272	1897.904	.023	.2876	.1266	.0394	.5359
Temperature	Equal variances assumed	.252	.616	.986	2182	.324	.0311	.0315	-.0308	.0929
	Equal variances not assumed			.997	2129.911	.319	.0311	.0312	-.0300	.0922
Hb	Equal variances assumed	2.327	.127	21.445	2204	.000	1.61508892	.075312687	1.46739766	1.76278018
	Equal variances not assumed			21.668	2169.541	.000	1.61508892	.074536451	1.46891861	1.76125923
Lactate	Equal variances assumed	10.010	.002	4.207	729	.000	.492818421	.117134561	.262857105	.722779737
	Equal variances not assumed			4.930	728.074	.000	.492818421	.099962810	.296568673	.689068169
Systolic_BP	Equal variances assumed	46.988	.000	-1.188	2262	.235	-1.1971	1.0078	-3.1733	.7792
	Equal variances not assumed			-1.155	1858.934	.248	-1.1971	1.0361	-3.2290	.8349
Diastolic_BP	Equal variances assumed	2.155	.142	5.840	2262	.000	3.4605	.5926	2.2984	4.6226
	Equal variances not assumed			5.794	2059.356	.000	3.4605	.5973	2.2891	4.6318

If Levene's test is passed (i.e., p>0.05, i.e., you do have equality of variances), then you use the top "Sig. (2-tailed)" value. If Levene's test is *not* met, you use the bottom value. The reportable p-values are highlighted above.

Let’s look at another example. Here, we are comparing blood pressure outcomes between men and women:



Group Statistics

	Sex	N	Mean	Std. Deviation	Std. Error Mean
Systolic_BP	Male	1273	139.164	21.2778	.5964
	Female	991	140.361	26.6703	.8472
Diastolic_BP	Male	1273	83.410	13.5908	.3809
	Female	991	79.949	14.4829	.4601
Mean_Arterial_BP	Male	1291	100.572	18.7499	.5218
	Female	1015	97.720	22.3993	.7031

		Levene's Test for Equality of Variances		t-test for Equality of Means						
									95% Confidence Interval of the Difference	
		F	Sig.	t	df	Sig. (2-tailed)	Mean Difference	Std. Error Difference	Lower	Upper
Systolic_BP	Equal variances assumed	46.988	.000	-1.188	2262	.235	-1.1971	1.0078	-3.1733	.7792
	Equal variances not assumed			-1.155	1858.934	.248	-1.1971	1.0361	-3.2290	.8349
Diastolic_BP	Equal variances assumed	2.155	.142	5.840	2262	.000	3.4605	.5926	2.2984	4.6226
	Equal variances not assumed			5.794	2059.356	.000	3.4605	.5973	2.2891	4.6318
Mean_Arterial_BP	Equal variances assumed	12.672	.000	3.327	2304	.001	2.8525	.8573	1.1713	4.5337
	Equal variances not assumed			3.258	1969.190	.001	2.8525	.8756	1.1353	4.5696

The output tables show no significant difference between men and women in systolic pressure ($p=0.248$), but did find differences in diastolic pressure (men are 3.46 mmHg higher; $p<0.001$) and in mean arterial pressure (men are 2.85 mmHg higher; $p=0.001$).

Group Statistics

	Smoker	N	Mean	Std. Deviation	Std. Error Mean
Oximetry	0	1661	97.010	2.9966	.0735
	1	579	97.224	2.6414	.1098
Temperature	0	1614	98.160	.7271	.0181
	1	570	98.163	.7437	.0312
pH	0	174	7.3189	.45516	.03451
	1	74	7.3645	.06766	.00787

		Levene's Test for Equality of Variances		t-test for Equality of Means						
									95% Confidence Interval of the Difference	
		F	Sig.	t	df	Sig. (2-tailed)	Mean Difference	Std. Error Difference	Lower	Upper
Oximetry	Equal variances assumed	2.741	.098	-1.526	2238	.127	-.2142	.1404	-.4896	.0611
	Equal variances not assumed			-1.622	1133.512	.105	-.2142	.1321	-.4735	.0450
Temperature	Equal variances assumed	1.346	.246	-.072	2182	.942	-.0026	.0356	-.0725	.0673
	Equal variances not assumed			-.072	978.642	.943	-.0026	.0360	-.0733	.0681
pH	Equal variances assumed	1.491	.223	-.857	246	.392	-.04561	.05322	-.15043	.05922
	Equal variances not assumed			-1.289	190.230	.199	-.04561	.03539	-.11542	.02420

Here's another output. Smoking status. It is coded as either 0 (doesn't smoke) or 1 (does smoke). The labels for smoking status were not created in the variable view in SPSS, so they're displayed as actual values entered.

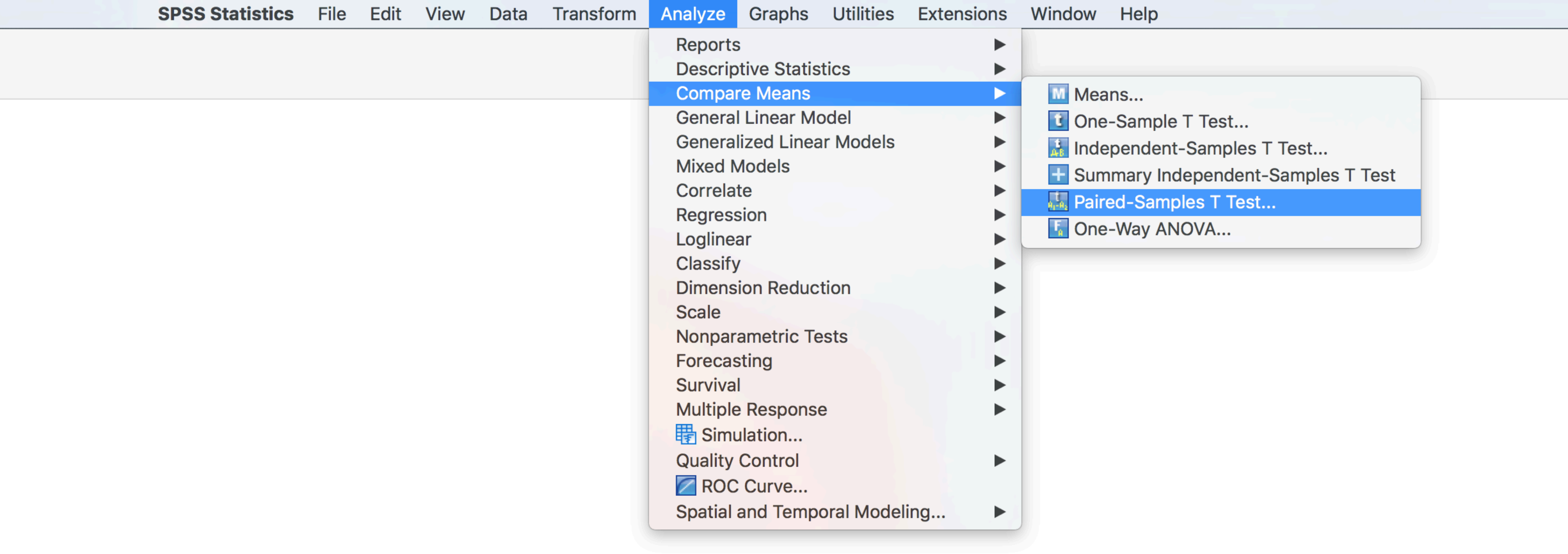
Table 1: *Subject baseline characteristics*

	Total	Men	Women	Sig.
N	45	23	22	
Age (years)	20.2 ± 0.7	21.0 ± 0.5	22.3 ± 0.9	p = 0.412
BMI (kg/m²)	26.5 ± 2.8	27.4 ± 0.8	26.0 ± 0.6	p = 0.518
GPA	2.9 ± 0.7	2.6 ± 1.0	3.2 ± 0.5	**p = 0.048**
Nightly sleep (hours)	7.5 ± 1.4	7.1 ± 1.2	7.6 ± 1.9	p = 0.525
Employed (%)	22.5%			
Weekly work (hours)	4.0 ± 5.2	8.5 ± 4.1	2.3 ± 6.6	p = 0.126
Academic Scholarship (%)	12.9%			
Athletic Scholarship (%)	11.5%			

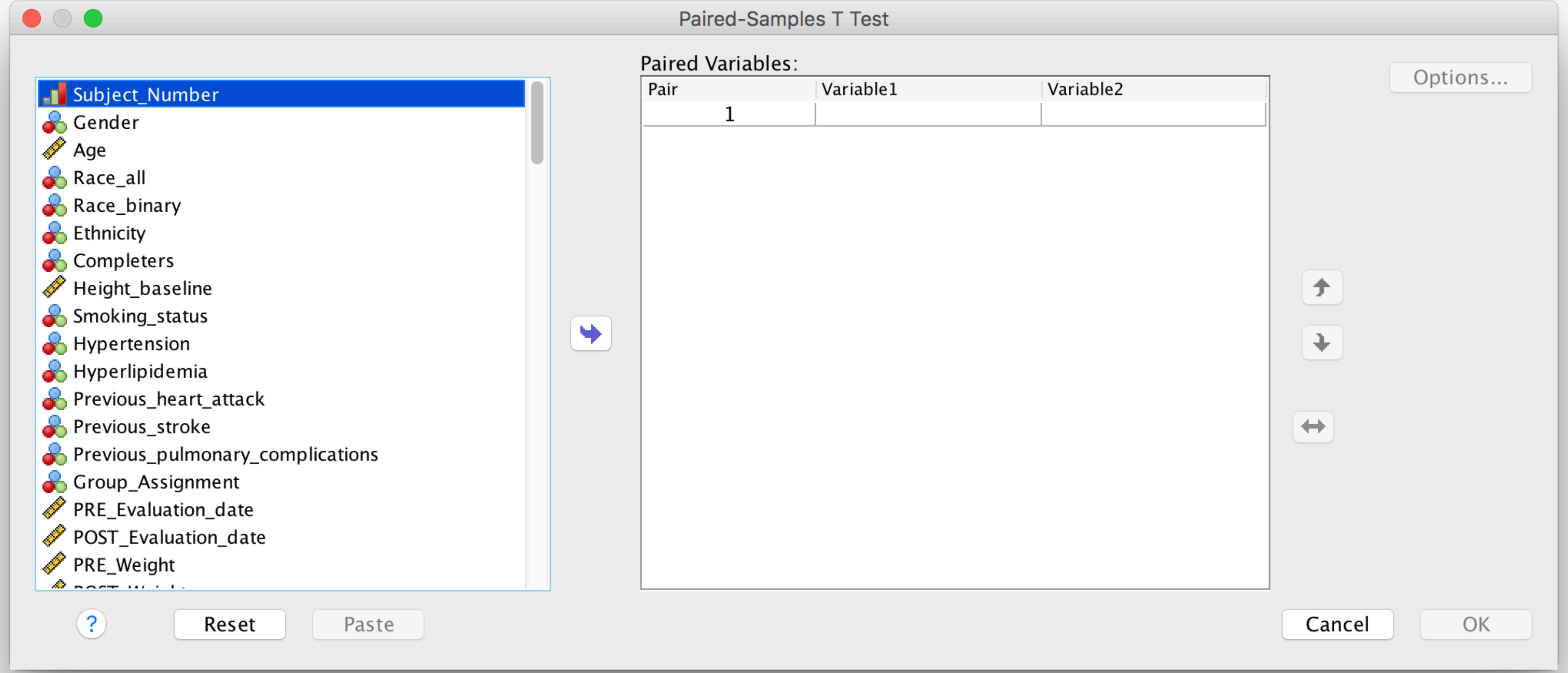
The means and standard deviations from the first output will go in the Men and Women columns (or whatever your two subsamples are) and the appropriate significance is placed in the Sig. column.

SPSS Statistics File Edit View Data Transform Analyze Graphs Utilities Extensions Window Help

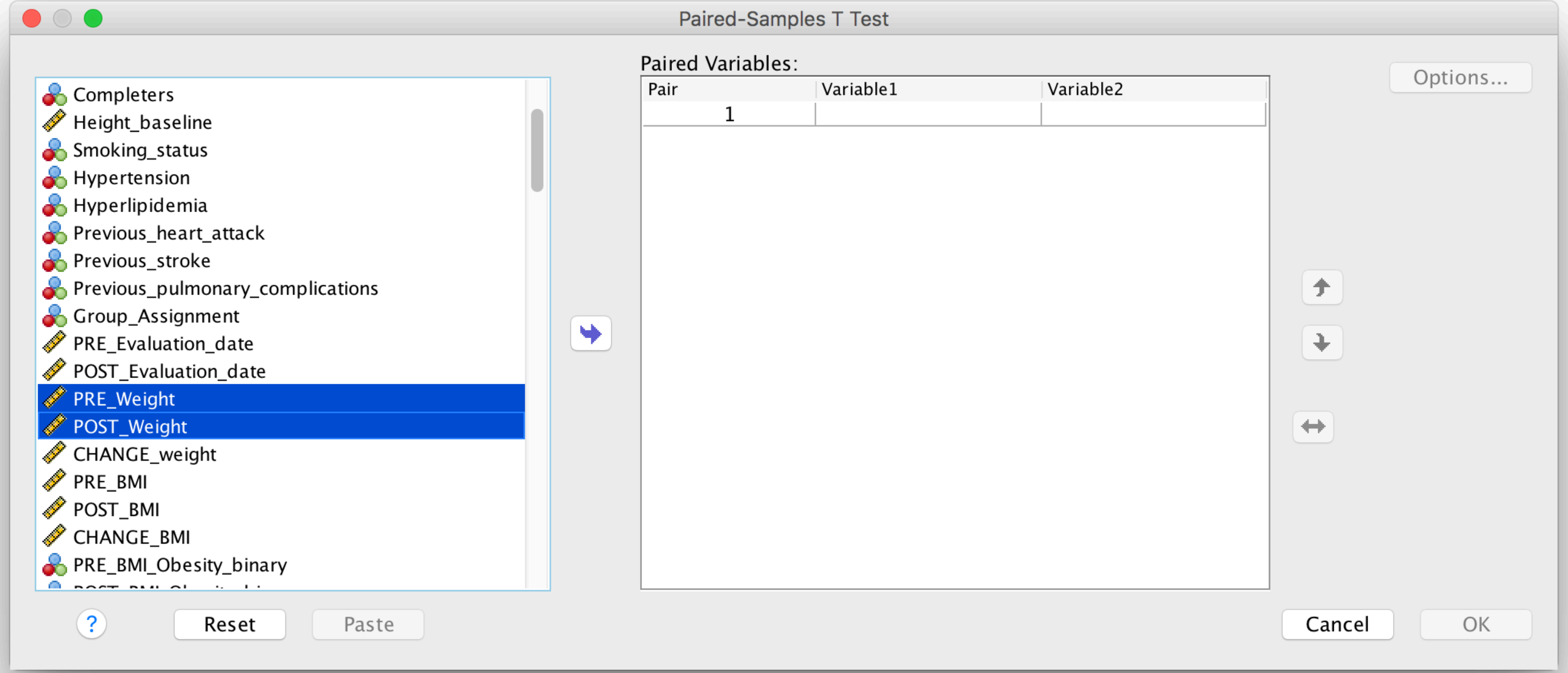
Conducting a Paired-Samples T-Test



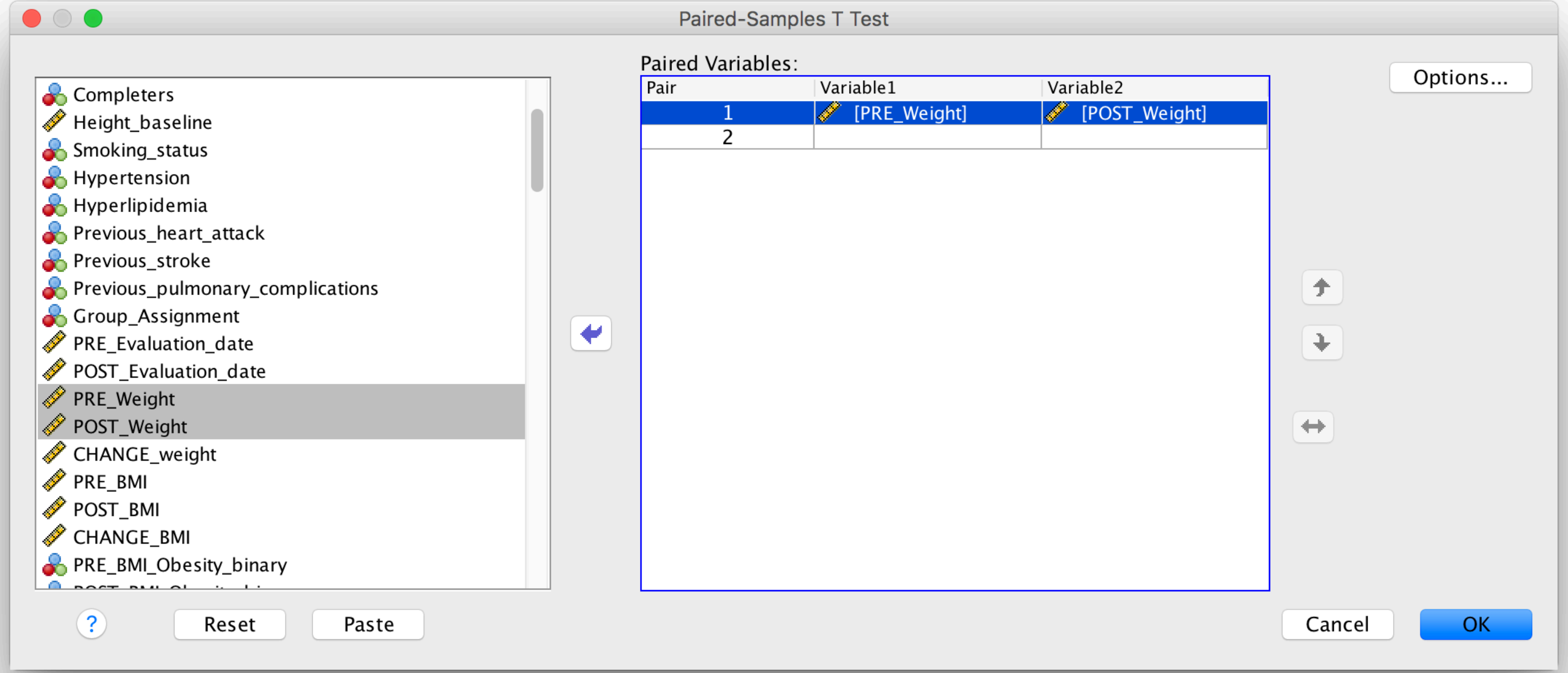
Select Paired-Samples T-Test from the menu bar.



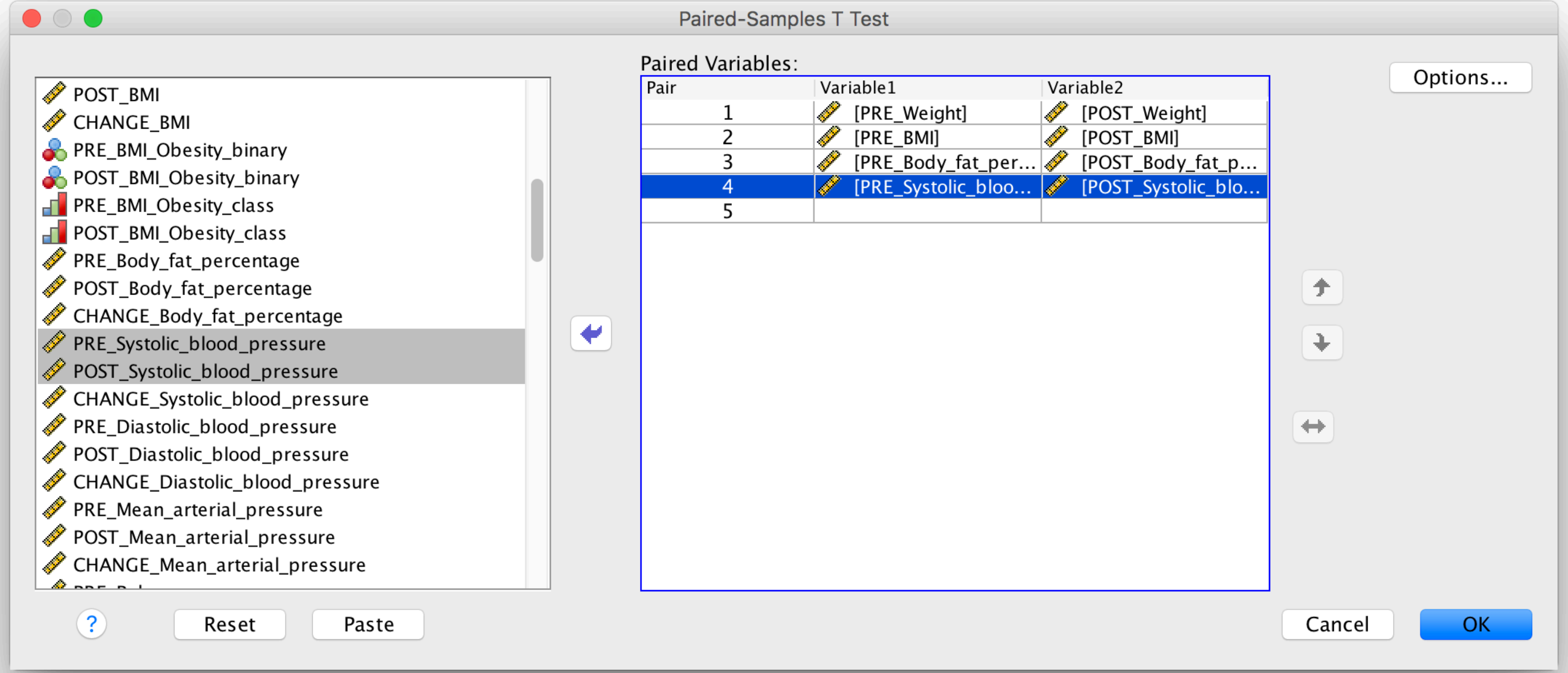
This menu will appear. In the left box is every variable in your dataset. In the right box, there is no grouping variable. You must select two variables from the left column to compare.



Example: pre-intervention bodyweight and post-intervention bodyweight.



Move both of those variables to the right. This is how to compare the subjects' mean bodyweight before the intervention to the mean bodyweight afterward (same exact subjects).



Move over every variable of interest, but make sure you also move its corresponding variable. Pre-BMI compared to post-BMI. Pre-fat% compared to post-fat%. Etc.

The top right output provides the means and standard deviations of each condition (pre and post).

The Top left output has bivariate correlations of the paired samples.

The bottom output has the significance of the difference.

Paired Samples Correlations

		N	Correlation	Sig.
Pair 1	PRE_Weight & POST_Weight	39	.993	.000
Pair 2	PRE_BMI & POST_BMI	39	.991	.000
Pair 3	PRE_Body_fat_percentage & POST_Body_fat_percentage	38	.957	.000
Pair 4	PRE_Systolic_blood_pressure & POST_Systolic_blood_pressure	38	.604	.000

Paired Samples Statistics

		Mean	N	Std. Deviation	Std. Error Mean
Pair 1	PRE_Weight	194.756	39	44.1789	7.0743
	POST_Weight	192.769	39	42.6824	6.8347
Pair 2	PRE_BMI	31.6157	39	5.88722	.94271
	POST_BMI	31.2998	39	5.73948	.91905
Pair 3	PRE_Body_fat_percentage	38.224	38	7.2948	1.1834
	POST_Body_fat_percentage	37.884	38	6.9464	1.1269
Pair 4	PRE_Systolic_blood_pressure	129.105	38	12.8332	2.0818
	POST_Systolic_blood_pressure	124.947	38	11.2345	1.8225

Paired Samples Test

		Paired Differences					t	df	Sig. (2-tailed)
					95% Confidence Interval of the Difference				
		Mean	Std. Deviation	Std. Error Mean	Lower	Upper			
Pair 1	PRE_Weight - POST_Weight	1.9872	5.3784	.8612	.2437	3.7307	2.307	38	.027
Pair 2	PRE_BMI - POST_BMI	.31590	.80899	.12954	.05366	.57814	2.439	38	.020
Pair 3	PRE_Body_fat_percentage - POST_Body_fat_percentage	.3395	2.1223	.3443	-.3581	1.0370	.986	37	.331
Pair 4	PRE_Systolic_blood_pressure - POST_Systolic_blood_pressure	4.1579	10.7992	1.7519	.6083	7.7075	2.373	37	.023

Results: from pretest to posttest, BMI decreased 0.32 points (p=0.020), weight decreased 1.99lb (p=0.027), and body fat percent was unchanged (p=0.331).

Paired Samples Correlations

		N	Correlation	Sig.
Pair 1	PRE_Weight & POST_Weight	39	.993	.000
Pair 2	PRE_BMI & POST_BMI	39	.991	.000
Pair 3	PRE_Body_fat_percentage & POST_Body_fat_percentage	38	.957	.000
Pair 4	PRE_Systolic_blood_pressure & POST_Systolic_blood_pressure	38	.604	.000

Paired Samples Statistics

		Mean	N	Std. Deviation	Std. Error Mean
Pair 1	PRE_Weight	194.756	39	44.1789	7.0743
	POST_Weight	192.769	39	42.6824	6.8347
Pair 2	PRE_BMI	31.6157	39	5.88722	.94271
	POST_BMI	31.2998	39	5.73948	.91905
Pair 3	PRE_Body_fat_percentage	38.224	38	7.2948	1.1834
	POST_Body_fat_percentage	37.884	38	6.9464	1.1269
Pair 4	PRE_Systolic_blood_pressure	129.105	38	12.8332	2.0818
	POST_Systolic_blood_pressure	124.947	38	11.2345	1.8225

Paired Samples Test

		Paired Differences					t	df	Sig. (2-tailed)
					95% Confidence Interval of the Difference				
		Mean	Std. Deviation	Std. Error Mean	Lower	Upper			
Pair 1	PRE_Weight - POST_Weight	1.9872	5.3784	.8612	.2437	3.7307	2.307	38	.027
Pair 2	PRE_BMI - POST_BMI	.31590	.80899	.12954	.05366	.57814	2.439	38	.020
Pair 3	PRE_Body_fat_percentage - POST_Body_fat_percentage	.3395	2.1223	.3443	-.3581	1.0370	.986	37	.331
Pair 4	PRE_Systolic_blood_pressure - POST_Systolic_blood_pressure	4.1579	10.7992	1.7519	.6083	7.7075	2.373	37	.023

In the outputs for both types of t-test, there is a 95% confidence interval (CI). Being inferential statistics, we are making inferences about the population the tested subjects were sampled from. The 95% CI is a range of values that you are 95% sure the true population value falls within. Independent-samples example: We're 95% sure men have higher diastolic blood pressure by 1.14 to 4.57 mmHg in the larger population. Paired-samples example: We're 95% sure the exercise program tested elicits a loss of 0.24 to 3.73 lb from beginning to end.

Independent-samples t-test:

		Levene's Test for Equality of Variances		t-test for Equality of Means						
									95% Confidence Interval of the Difference	
		F	Sig.	t	df	Sig. (2-tailed)	Mean Difference	Std. Error Difference	Lower	Upper
Systolic_BP	Equal variances assumed	46.988	.000	-1.188	2262	.235	-1.1971	1.0078	-3.1733	.7792
	Equal variances not assumed			-1.155	1858.934	.248	-1.1971	1.0361	-3.2290	.8349
Diastolic_BP	Equal variances assumed	2.155	.142	5.840	2262	.000	3.4605	.5926	2.2984	4.6226
	Equal variances not assumed			5.794	2059.356	.000	3.4605	.5973	2.2891	4.6318
Mean_Arterial_BP	Equal variances assumed	12.672	.000	3.327	2304	.001	2.8525	.8573	1.1713	4.5337
	Equal variances not assumed			3.258	1969.190	.001	2.8525	.8756	1.1353	4.5696

Paired-samples t-test:

		Paired Differences							
					95% Confidence Interval of the Difference				
		Mean	Std. Deviation	Std. Error Mean	Lower	Upper	t	df	Sig. (2-tailed)
Pair 1	PRE_Weight - POST_Weight	1.9872	5.3784	.8612	.2437	3.7307	2.307	38	.027
Pair 2	PRE_BMI - POST_BMI	.31590	.80899	.12954	.05366	.57814	2.439	38	.020
Pair 3	PRE_Body_fat_percentage - POST_Body_fat_percentage	.3395	2.1223	.3443	-.3581	1.0370	.986	37	.331
Pair 4	PRE_Systolic_blood_pressure - POST_Systolic_blood_pressure	4.1579	10.7992	1.7519	.6083	7.7075	2.373	37	.023

That's everything you need to know for independent-samples and paired-samples t-tests.