Introductory Statistics
11: Comparison of Two Means: “t-test”

Richard Veale

Graduate School of Medicine
Kyoto University

https://youtu.be/6nRLTSiRBYI

Lecture Video at above link

Summary

1) Reminder: Metric Data vs. Categorical Data

2) What is the "t-distribution"?

3) How to test difference of means of metric data: <u>Student's t-test</u>

a) Independent data ("Lumped" t-test)

b) Dependent data ("Paired" t-test)

4) Non-parametric tests

Wilcoxon Signed-Rank Test, Mann-Whitney U-Test...

Metric versus Categorical Data

Categorical data:

χ^2-test, Fisher's exact test
Odds/risk ratios
z-test for proportions

Smoke?
Yes, No

Got CHD?
Yes, No

Color?
Blue, Green, Red

Metric (numerical) data:

Birth weight, height, air plane speed, etc.

e.g., 3201g 4300g 2900g 3430g …

Test?

Weight:
1906.3 g, 1906.4 g,
1906.354838 g

Height:
182.3 cm, 164.27 cm,
155.5 cm, …

Metric versus Categorical Data

Categorical data:

χ^2-test, Fisher's exact test
Odds/risk ratios
z-test for proportions

Smoke?
Yes, No

-No natural order
-Finite number of values for each variable ("categories")

Blue, Green, Red

Metric (numerical) data:

Birth weight, height, air plane spee

e.g., 3201g 4300g 2900g 3

Weight:
1906.3 g, 1906.4 g,

-Ordered!
-Infinite number of values for each variable

155.5 cm, ...

Test?

Recall Central Limit Theorem (CLT)

Population distribution

μ: population mean

σ: population standard deviation

Sampling distribution

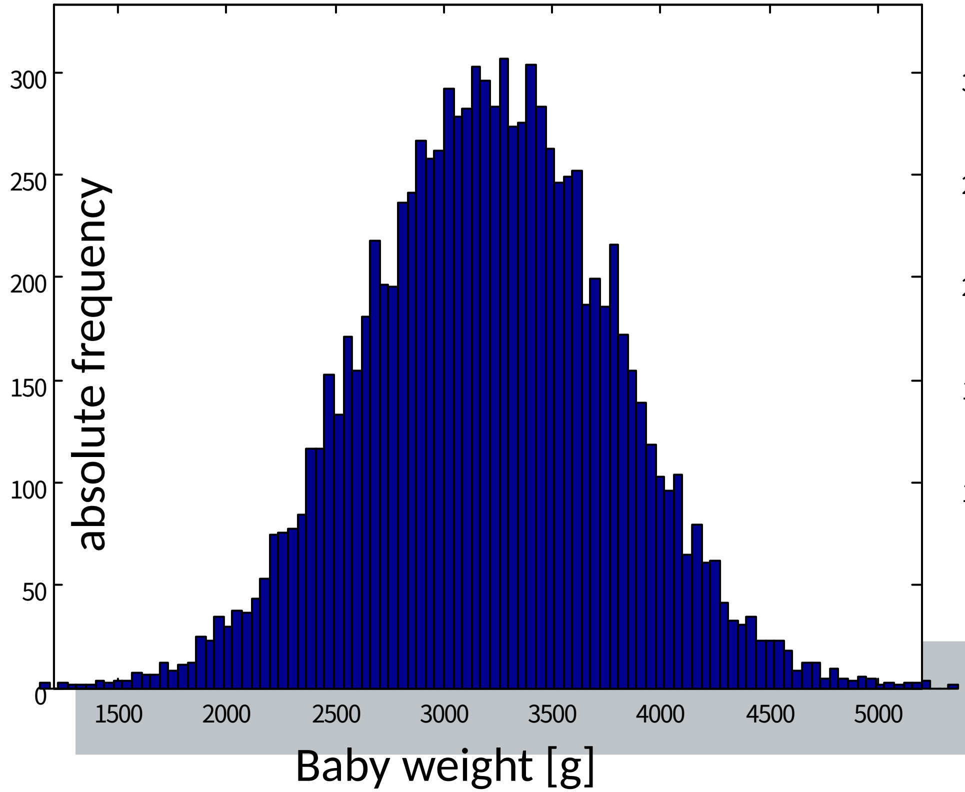
$\bar{X}$: sample mean

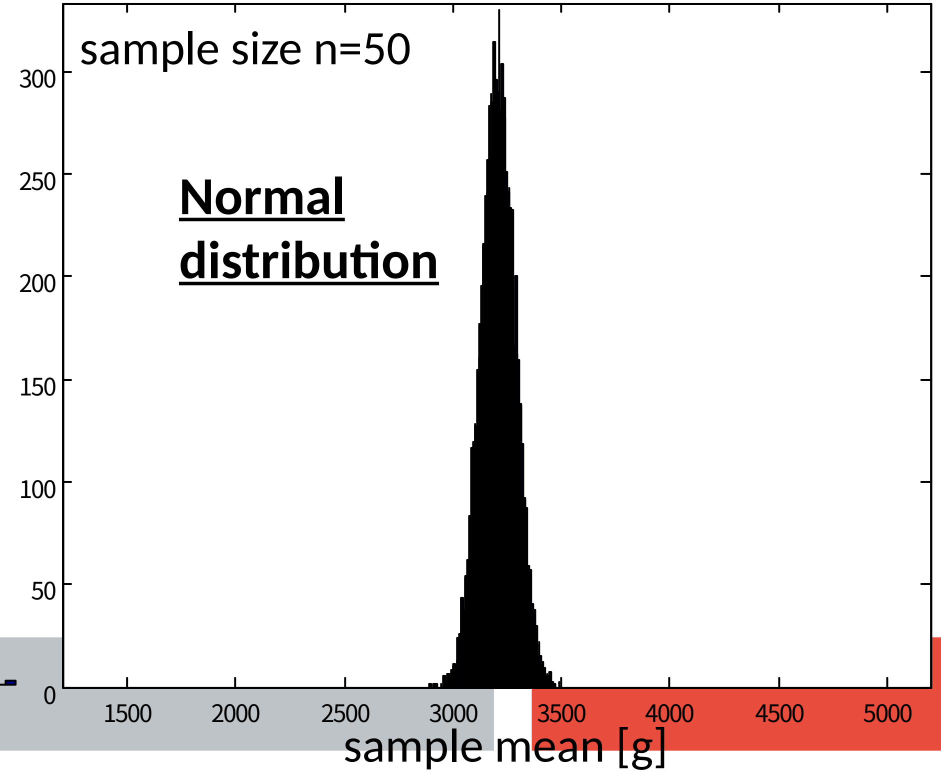
$s_x : s / \sqrt{n}$

S_x : standard error of mean (SEM)

S : sample standard deviation

n : sample size

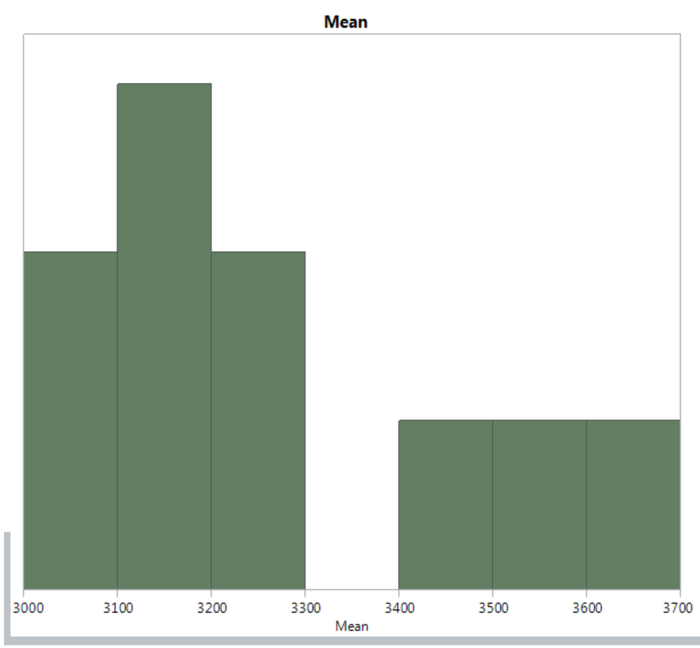




Requirement

But remember, the sample size should be large enough (**<u>n>30</u>**) or we should know the population variance σ^2 (otherwise we have to estimate it with s^2, the sample variance).

So, what do we do, if we don't know σ^2 (which is rare anyway) and we have a small sample size?



Example: distribution of birth weights in a small sample.

Help! t-distribution

Enter “Student” from the Guinness brewery:
(“A student of statistics”)

William Sealy Gosset
(1876-1937)

t-distribution

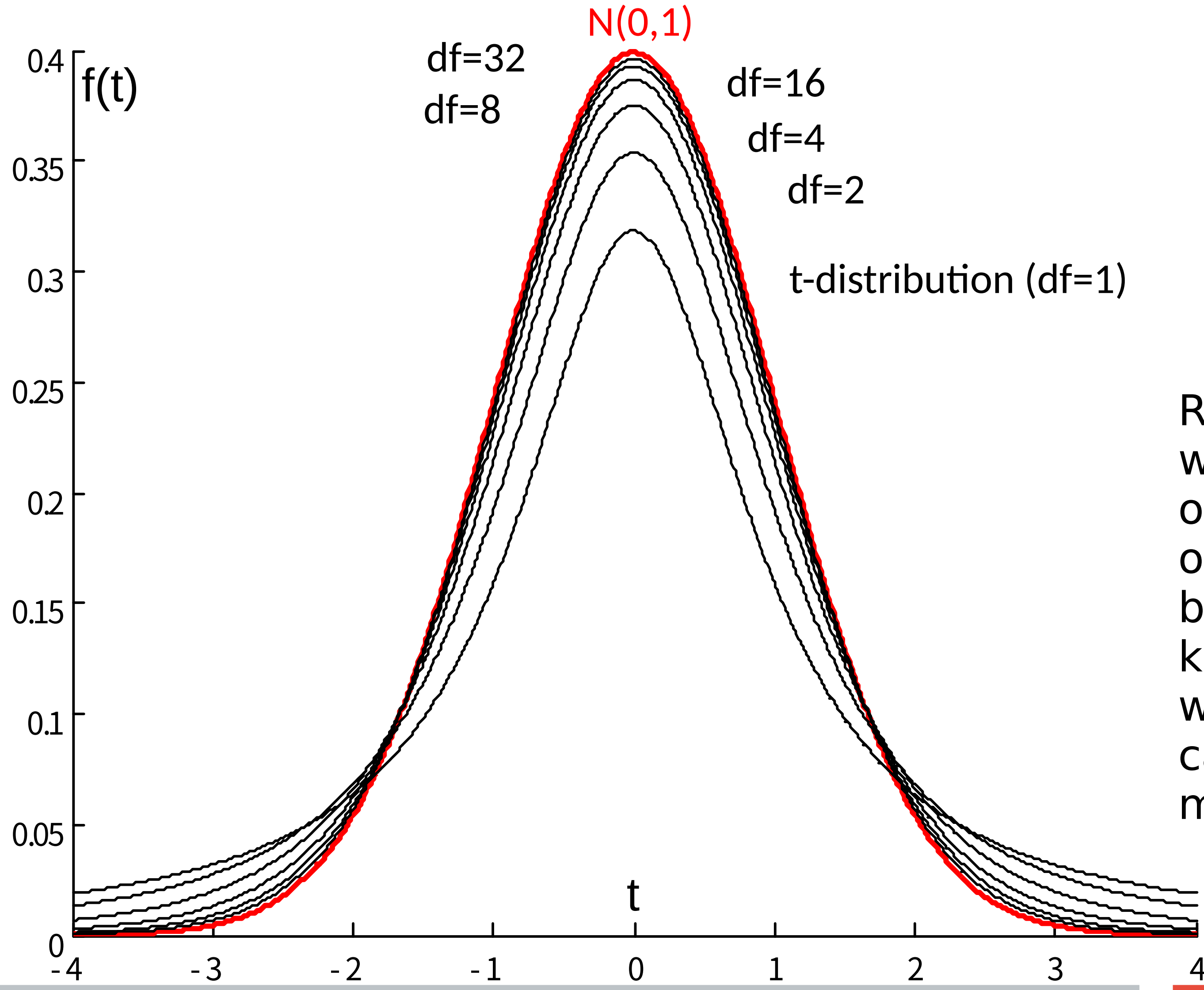
If x is a random variable that is normally distributed with mean μ and variance σ^2, then, if we take a sample of x,

$$t = \frac{\bar{x} - \mu}{s/\sqrt{n}}$$

t will be drawn from a t-distribution with n-1 degrees of freedom.

$\bar{x}$: sample mean (of our sample of x)
μ : population mean (x's true mean)
s : sample standard deviation
n : sample size

t-distribution for various df



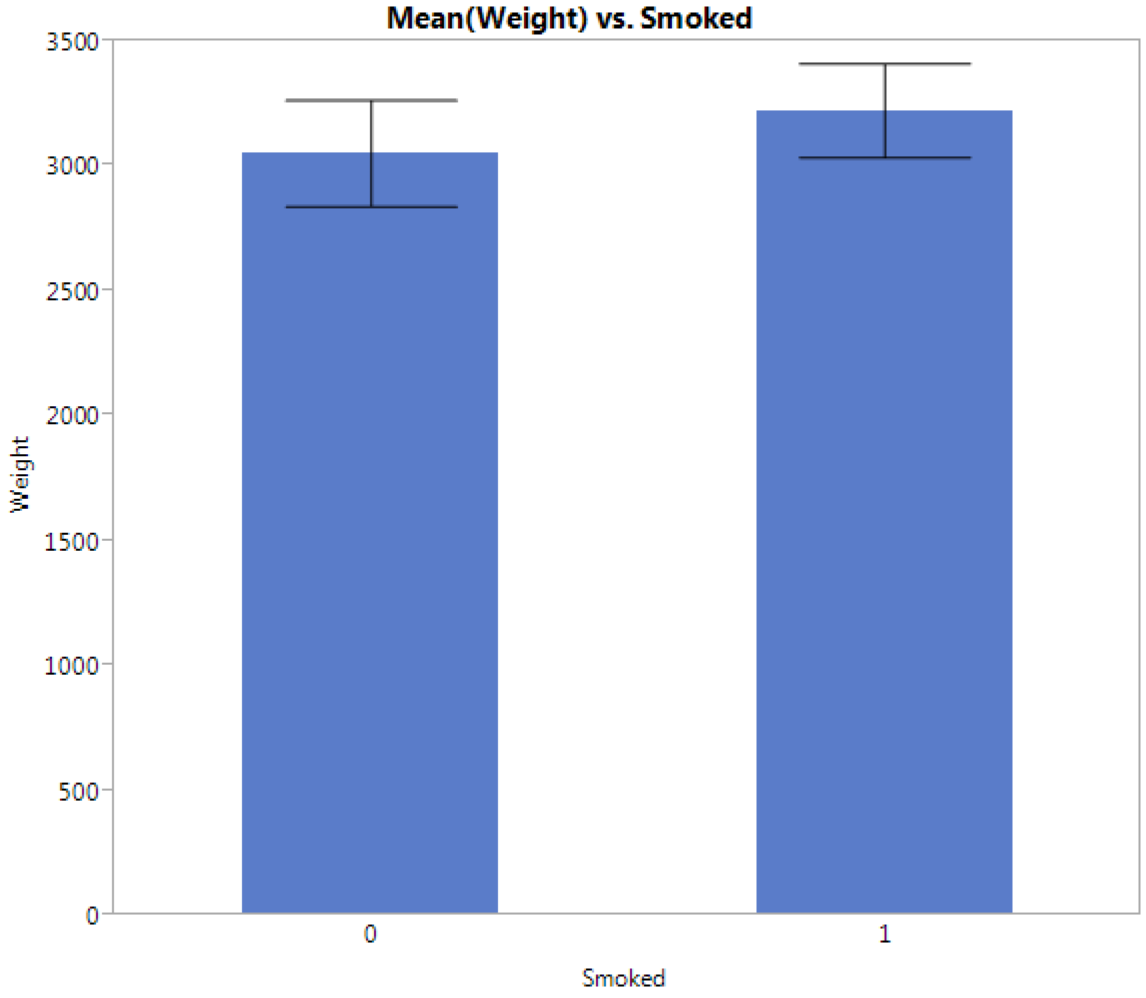
$$t = \frac{\overline{x} - \mu}{s/\sqrt{n}}$$

Remember, df will be n-1 if it is one group with one mean, because if we know the mean we can reverse-calculate one missing value!

For large n (df), the t-distribution approaches the normal distribution.

The t-test (two samples)

Remember assignment 1, problem 1:
You were asked to see whether there was a difference of babies' birth weight depending on whether mothers smoked or not:



The t-test (two samples)

We have two samples (one sample from each of our two groups):
One sample of n=10 birth weights from mothers who smoked,
One sample of n=10 birth weights from mothers who did not smoke.

<u>Smoked</u> during pregnancy:
2240 3050 4110 3740 3040 2920 2800 3090 4110 3130

$\bar{x}_1 = 3223$
$s_1 = 594$

<u>No smoking</u> during pregnancy:
3180 2560 2780 4550 2740 2940 1960 3460 3120 3220

$\bar{x}_2 = 3051$
$s_1 = 673$

H_0: null hypothesis: $\mu_1 = \mu_2$

H_a: alternative hypothesis: $\mu_1 \neq \mu_2$

Student's t-test (compare two samples)

$$t = \frac{\bar{x}_1 - \bar{x}_2}{s\sqrt{\frac{1}{m} + \frac{1}{n}}} \qquad s = \sqrt{\frac{s_1^{\,2}(m-1) + s_2^{\,2}(n-1)}{n+m-2}}$$

t ~ t-distribution with n+m-2 df.

$\bar{x}_1$: sample 1 mean

s_1: sample 1 standard deviation

m: sample 1 size

$\bar{x}_2$: sample 2 mean

s_2: sample 2 standard deviation

n: sample 2 size

We estimate a common variance by pooling the estimated sample variances, because we assume **equal** variances.

Student's t-test (compare two samples)

$$s = \sqrt{\frac{s_1^2(m-1)+s_2^2(n-1)}{n+m-2}} = \sqrt{\frac{594^2 \cdot 9 + 673^2 \cdot 9}{18}} \approx 634.46$$

$$t = \frac{\overline{x}_1 - \overline{x}_2}{s\sqrt{\frac{1}{m}+\frac{1}{n}}} = \frac{172}{634.46\sqrt{\frac{1}{10}+\frac{1}{10}}} \approx 0.6062$$

$\overline{x}_1 = 3223$
$s_1 = 594$
m=10

For α=0.05: t_{crit1} [18; 2.5%] = -2.101,
t_{crit2} [18;97.5%] = +2.101

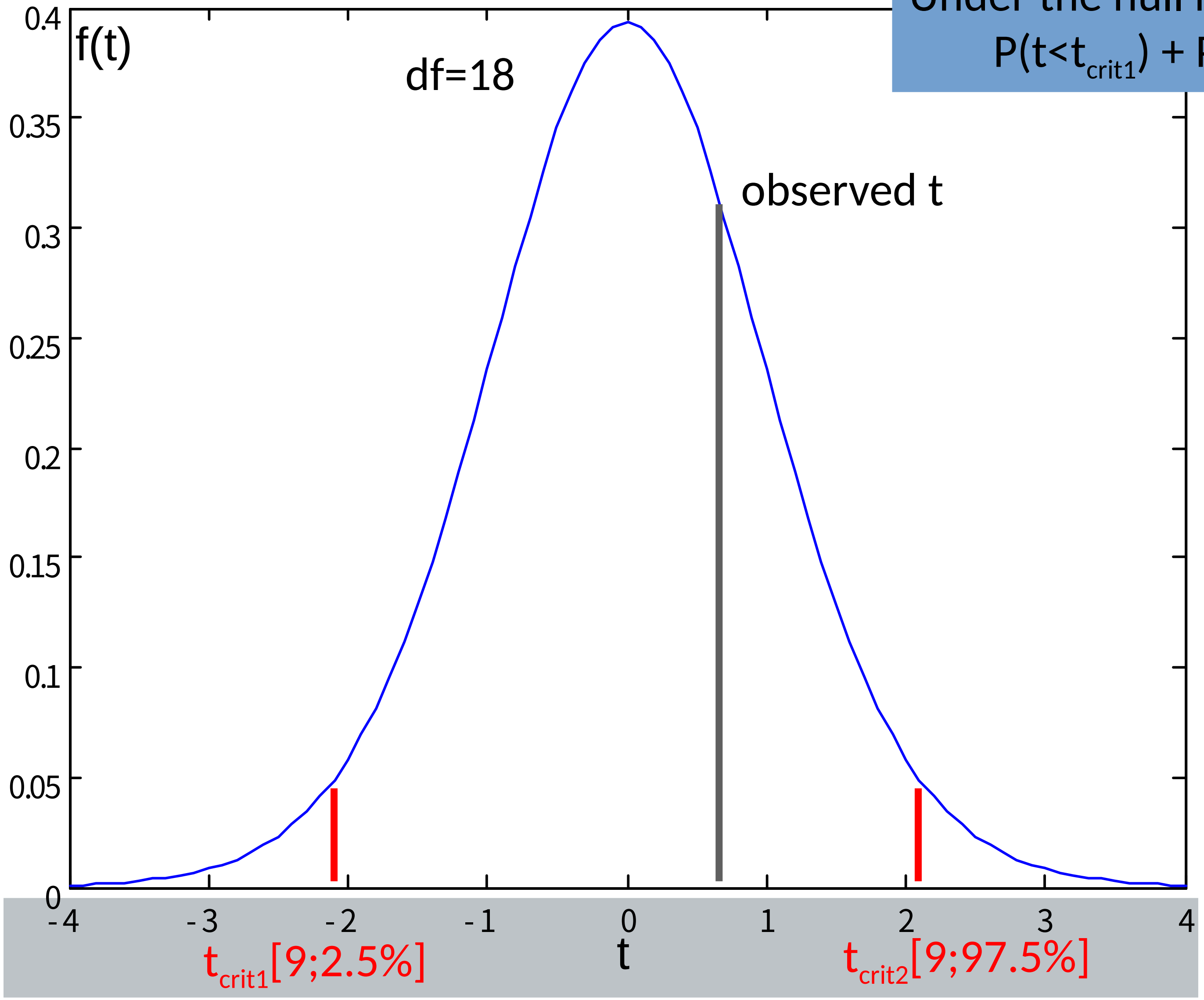
$\overline{x}_2 = 3051$
$s_2 = 673$
n=10

$t < t_{crit2}$

-> Our data suggests that we cannot reject the null hypothesis that

-> We found no significant difference between the two groups (two-tailed, two-sample Student's t-test: t[18] = 0.61, p = 0.55).

Student's t-test (compare two samples)



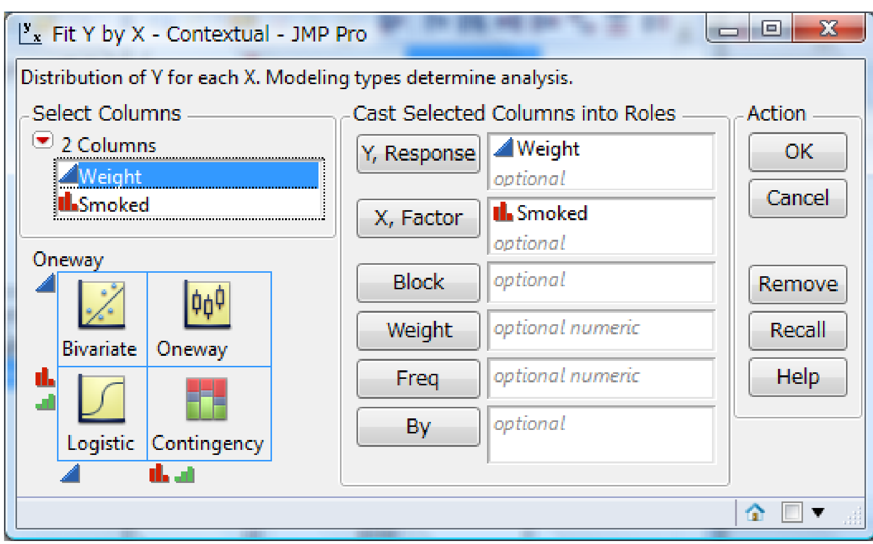
Under the null hypothesis:

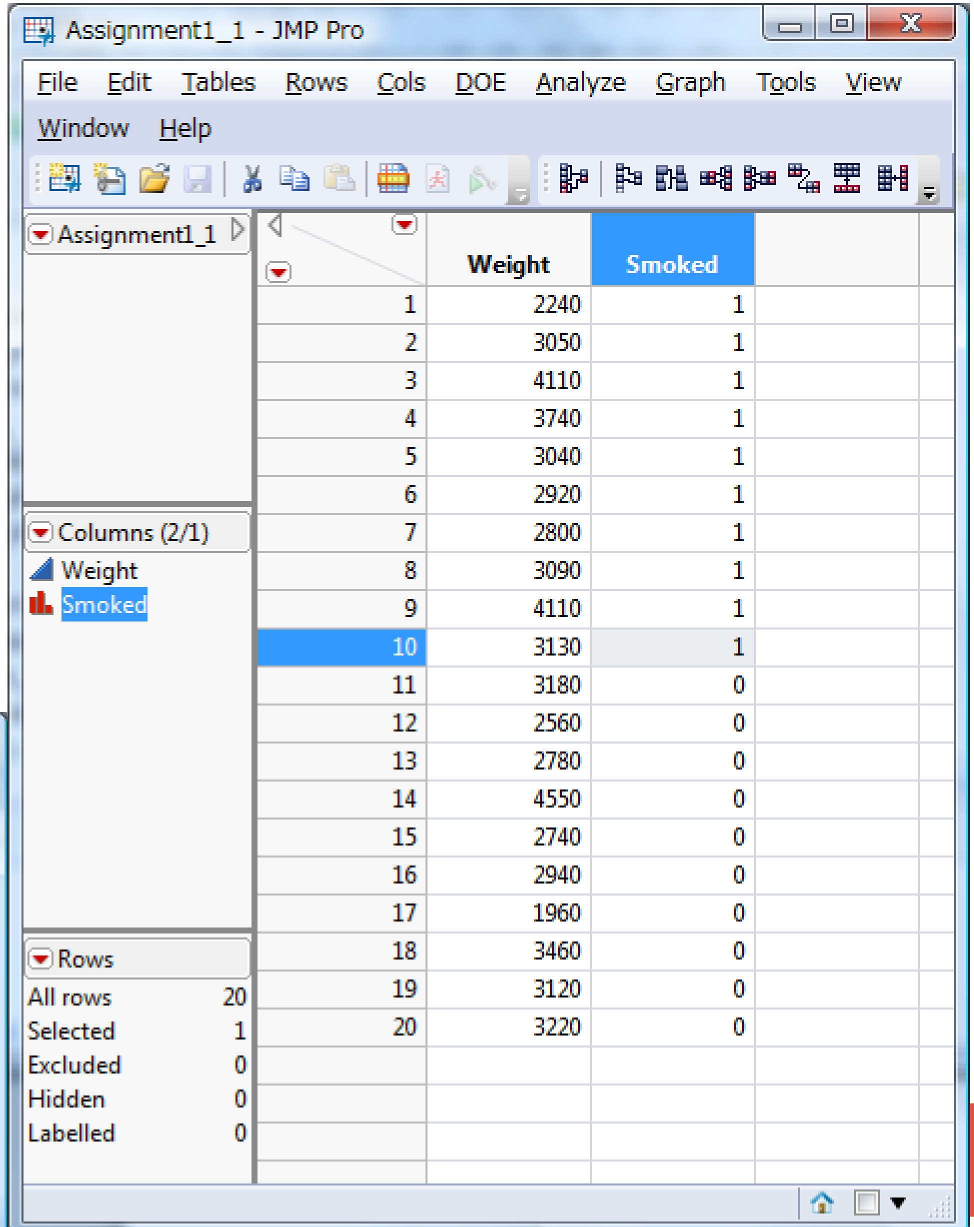
$P(t<t_{crit1}) + P(t>t_{crit2}) = \alpha = 0.05$

$$t = \frac{\overline{x} - \mu}{s/\sqrt{n}}$$

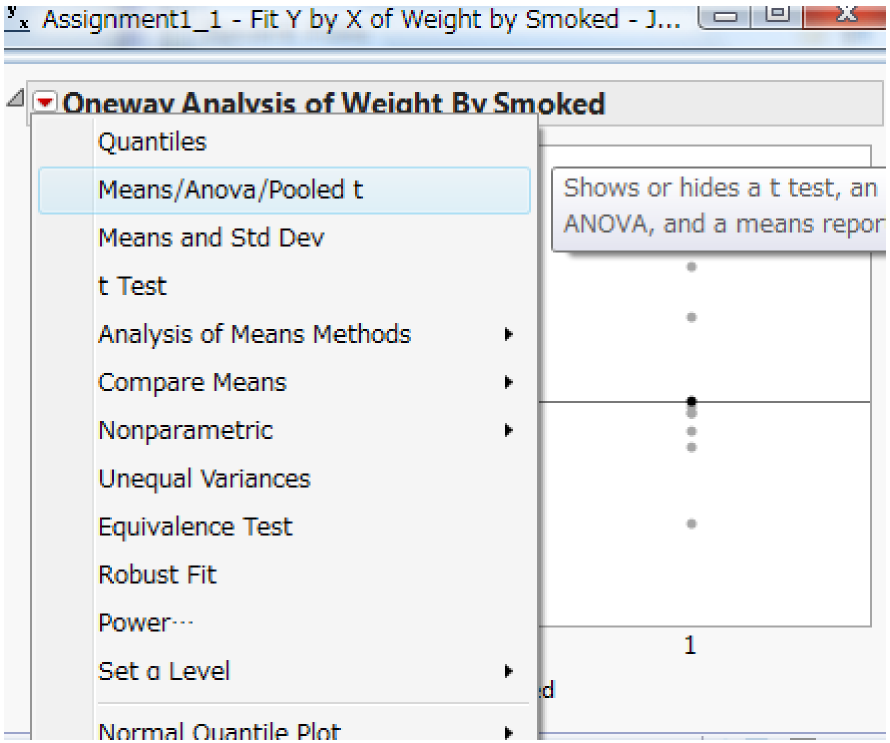
t-test in JMP

Important:
“Smoked” is to be a nominal (categorical), not a continuous (numerical) variable!

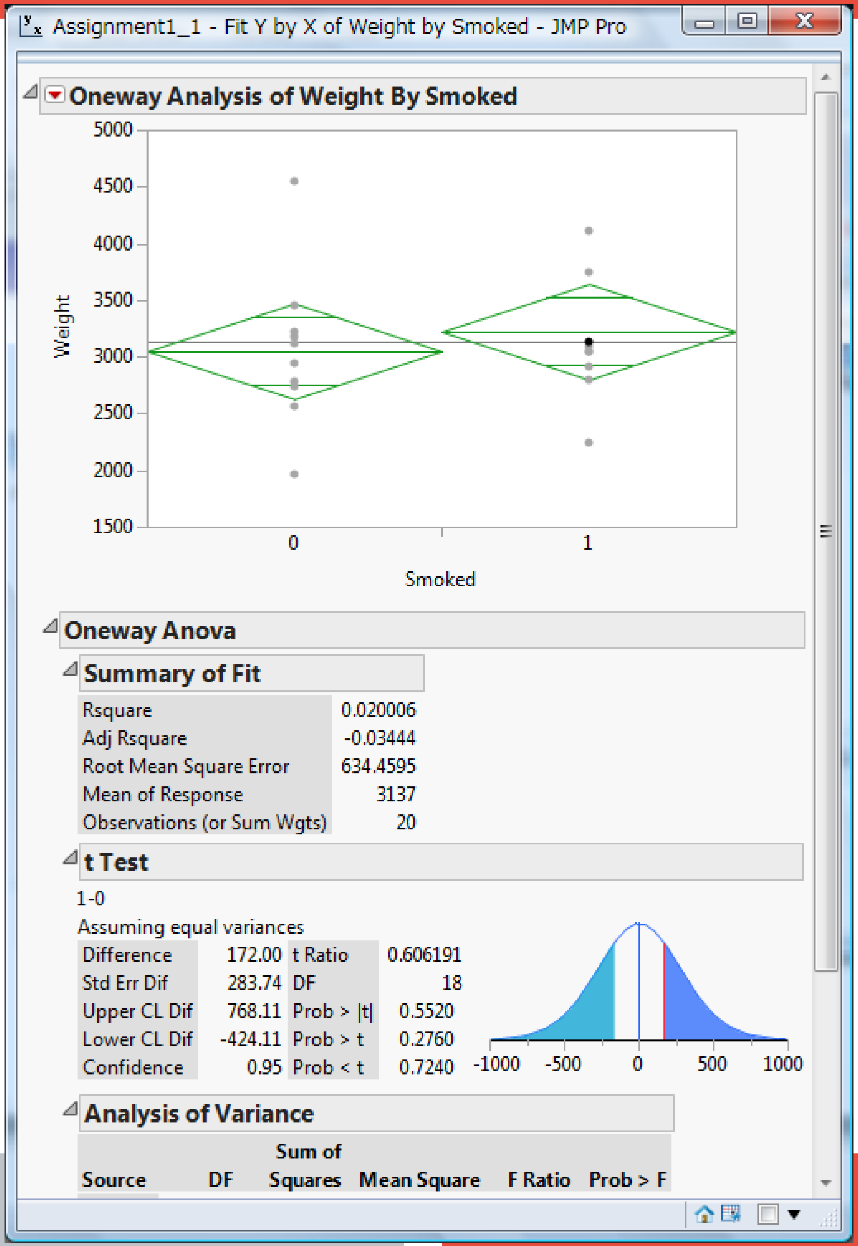




t-test in JMP



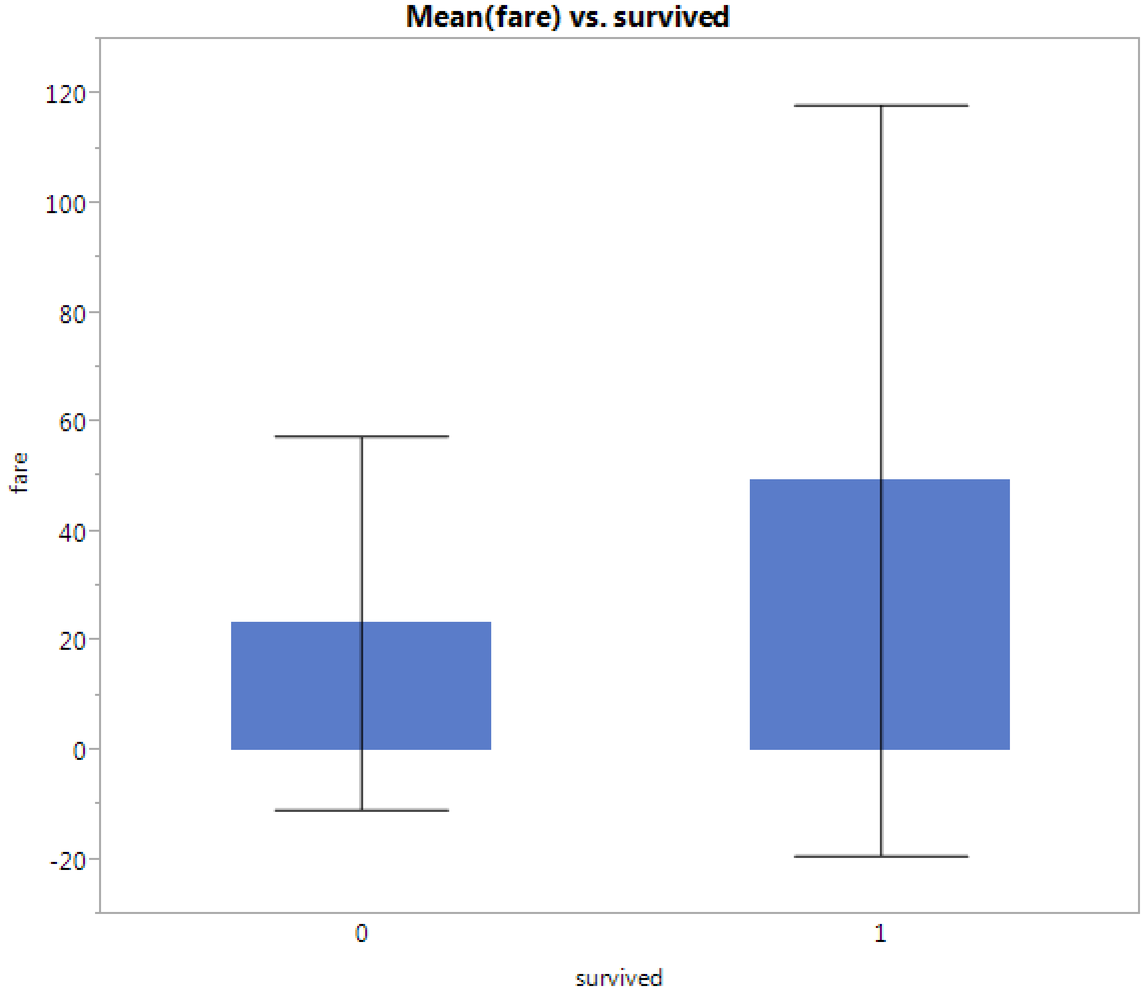
Assignment1_1 - Fit Y by X of Weight by Smoked - J...
Oneway Analysis of Weight By Smoked
Quantiles
Means/Anova/Pooled t
Means and Std Dev
t Test
Analysis of Means Methods
Compare Means
Nonparametric
Unequal Variances
Equivalence Test
Robust Fit
Power...
Set α Level
Normal Quantile Plot
Shows or hides a t test, an
ANOVA, and a means repor
1



Assignment1_1 - Fit Y by X of Weight by Smoked - JMP Pro
Oneway Analysis of Weight By Smoked
Weight
Smoked
Oneway Anova
Summary of Fit
Rsquare 0.020006
Adj Rsquare -0.03444
Root Mean Square Error 634.4595
Mean of Response 3137
Observations (or Sum Wgts) 20
t Test
1-0
Assuming equal variances
Difference 172.00 t Ratio 0.606191
Std Err Dif 283.74 DF 18
Upper CL Dif 768.11 Prob > |t| 0.5520
Lower CL Dif -424.11 Prob > t 0.2760
Confidence 0.95 Prob < t 0.7240
Analysis of Variance
Source DF Sum of Squares Mean Square F Ratio Prob > F

Unequal Variances

We cannot always assume equal variances.



No Yes

Think about the Titanic passengers: those that survived and those that did not survive might have been from very different social groups.

T-test, unequal variances

$$t = \frac{\bar{x}_1 - \bar{x}_2}{\sqrt{\frac{s_1^{\ 2}}{m} + \frac{s_2^{\ 2}}{n}}}$$

t ~ t-distribution with v degrees of freedom.

$$v = \frac{\left(\frac{s_1^2}{m} + \frac{s_2^2}{n}\right)^2}{\frac{(s_1^{\ 2}/m)^2}{m-1} + \frac{(s_2^{\ 2}/n)^2}{n-1}}$$

Satterthwaite, 1941
Let the software do it!

$\bar{x}_1$: sample 1 mean
s_1: sample 1 standard deviation
m: sample 1 size

$\bar{x}_2$: sample 2 mean
s_2: sample 2 standard deviation
n: sample 2 size

T-test, unequal variances (example)

$$t = \frac{\overline{x}_1 - \overline{x}_2}{\sqrt{\frac{s_1^2}{m} + \frac{s_2^2}{n}}} = \frac{-26.01}{\sqrt{\frac{34.15\;2}{808} + \frac{68.45\;2}{500}}} \approx \frac{-26.01}{\sqrt{10.81}} \approx -7.91$$

died:

$\overline{x}_1 = 23.35$
$S_1 = 34.15$
$m = 808$

$$v = 654.002$$

survived:

$\overline{x}_2 = 49.36$
$s_2 = 68.45$
$n = 500$

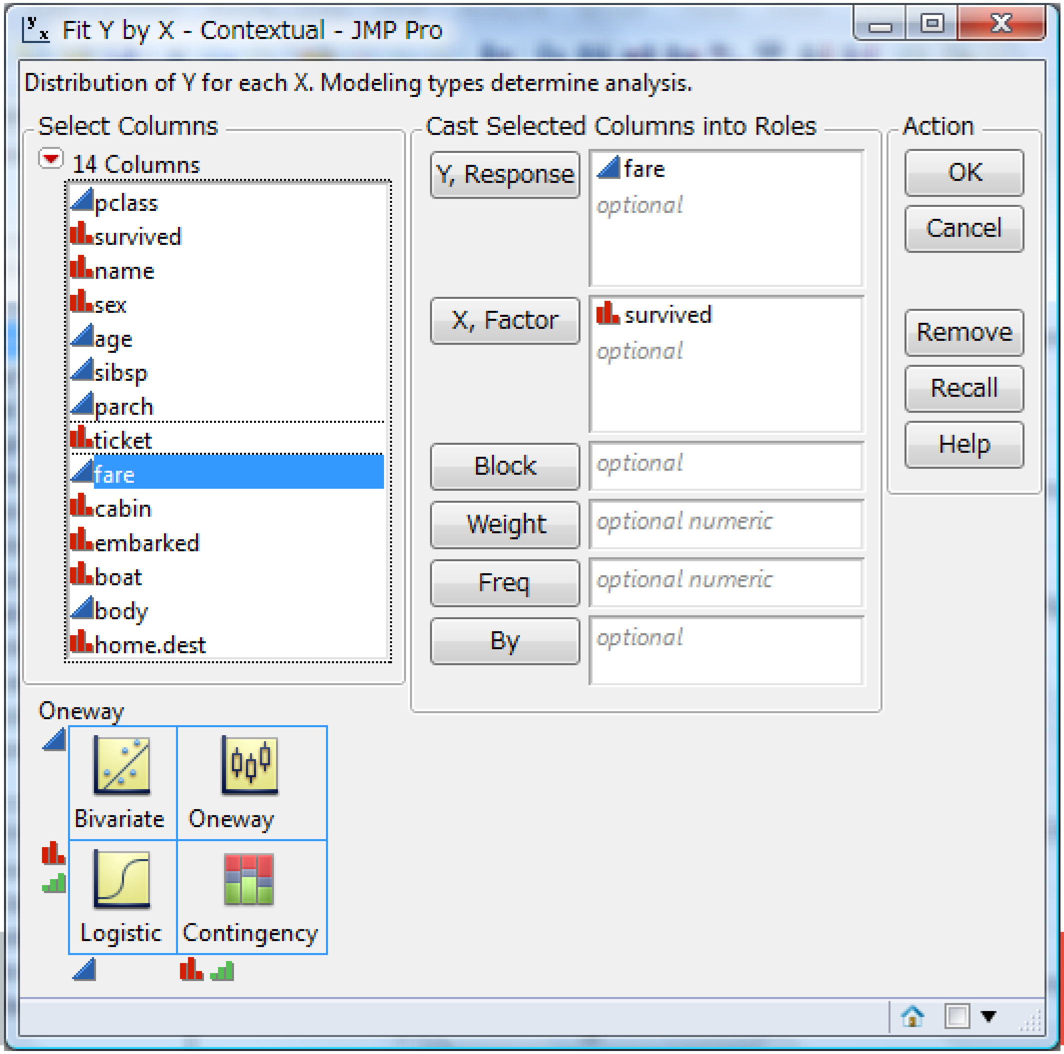
For α=0.05: t_{crit1}[654.002; 2.5%] = -1.9636
t_{crit2}[654.002;97.5%] = +1.9636

$t < t_{crit1}$

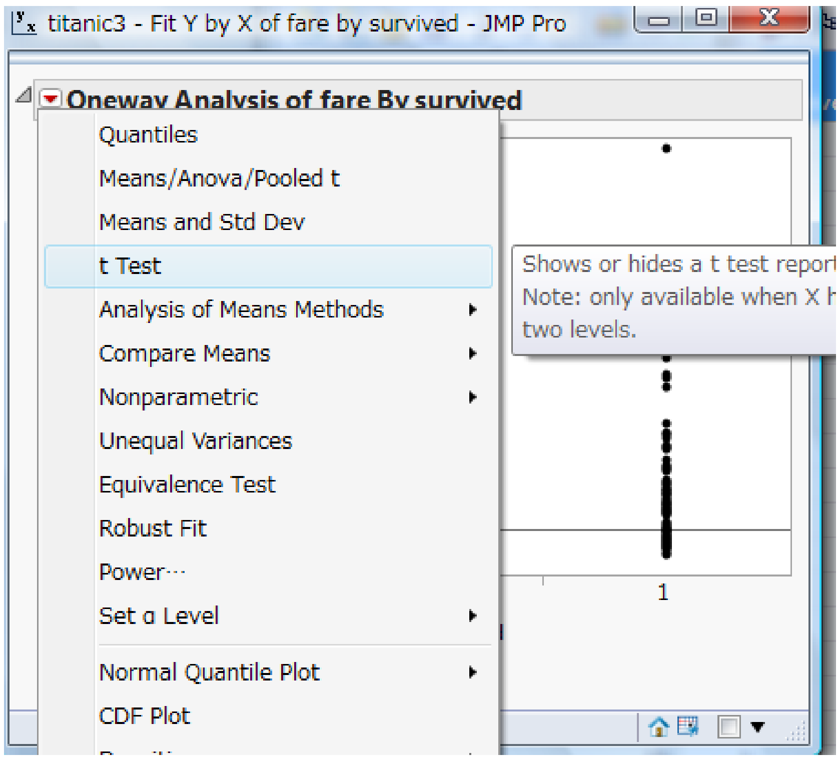
-> Based on our data we can reject the null hypothesis that -> We found a statistically significant difference of ticket price between survivors and victims of the Titanic disaster (two-tailed, two-sample t-test assuming unequal variances: t[654.002]=7.91, p<0.0001). Survivors paid on average higher prices (M=49.36, SD=68.45 Pound Sterling) than victims (M=23.35, SD=34.15 Pound Sterling).

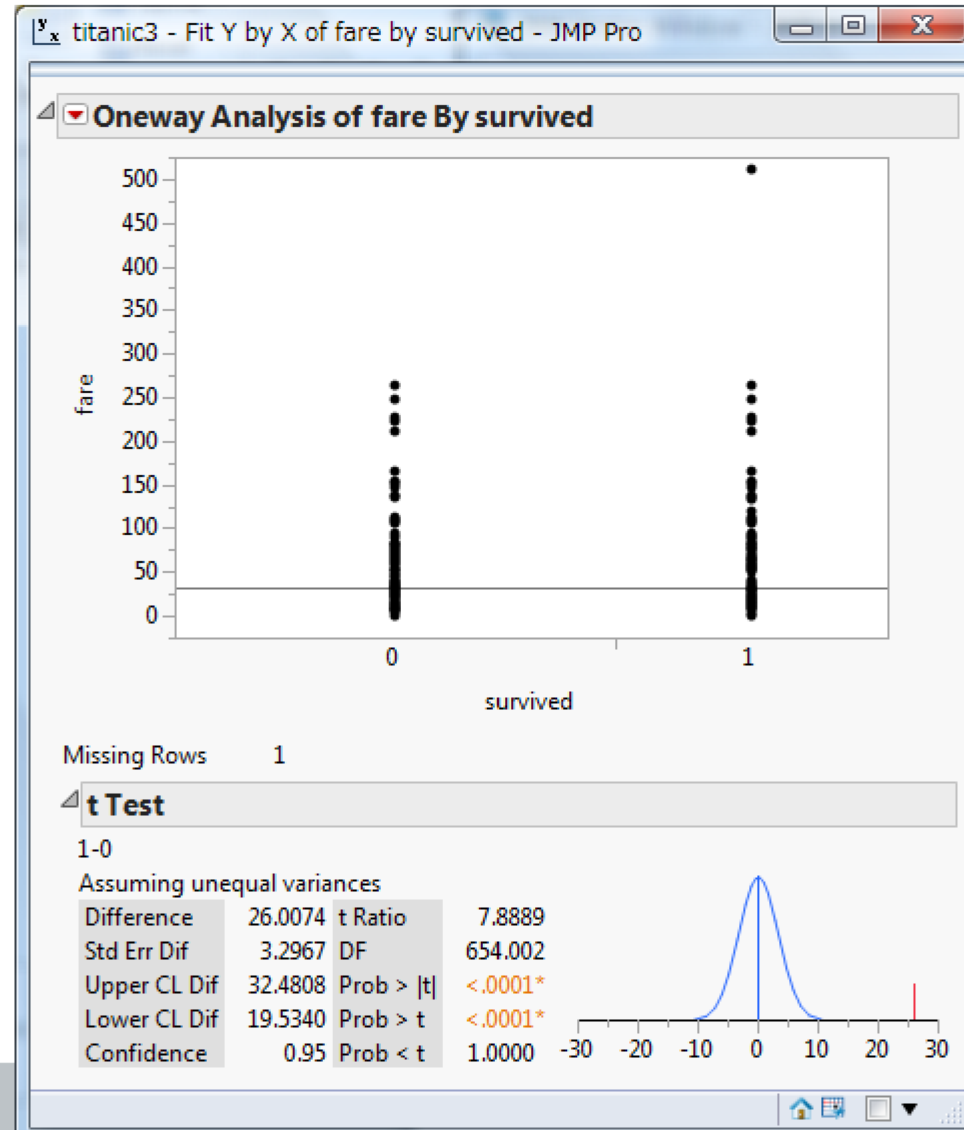
T-test, unequal variances in JMP

Important:
“Survived” has to be a nominal, not a continuous (numerical) variable!



T-test (unequal variances) in JMP





"t Test" in JMP *automatically assumes* unequal variances,

→ "Pooled t" assumes equal variances.

Paired t-test

So far, we have considered independent samples, but what can we do if samples are paired? For example, when the same patients are tested with two different treatments?

Additional hours of sleep after taking a drug.

Patient	Hyoscyamine	Hyoscine	Difference
1	+0.7	+1.9	+1.2
2	-1.6	+0.8	+2.4
3	-0.2	+1.1	+1.3
4	-1.2	+0.1	+1.3
5	-0.1	-0.1	0
6	+3.4	+4.4	+1.0
7	+3.7	+5.5	+1.8
8	+0.8	+1.6	+0.8
9	0	+4.6	+4.6
10	+2.0	+3.4	+1.4
Mean	+0.75	+2.33	+1.58

Paired t-test

So far, we have considered independent samples, but what can we do if samples are paired? For example, when the same patients are tested with two different treatments?

Additional hours of sleep after taking a drug.

Patient	Hyoscyamine	Hyoscine	Difference
1	+0.7	+1.9	+1.2
2	-1.6	+0.8	+2.4
3	-0.2	+1.1	+1.3
4	-1.2	+0.1	+1.3
5	-0.1	-0.1	0
6	+3[illegible]	+[illegible]	+1.0
7	[illegible]	[illegible]	+1.8
8	[illegible]	[illegible]6	+0.8
9	0	+4.6	+4.6
10	+2.0	+3.4	+1.4
Mean	+0.75	+2.33	+1.58

Sample 1?

Sample 2?

Paired t-test

So far, we have considered independent samples, but what can we do if samples are <u>paired</u>? For example, when the same patients are tested with two different treatments?

Additional hours of sleep after taking a drug.

Patient	Hyoscyamine	Hyoscine	Difference
1	+0.7	+1.9	+1.2
2	-1.6	+0.8	+2.4
3	[illegible]	[illegible]	+1.3
4	[illegible]	[illegible]	+1.3
5	[illegible]	[illegible]	[illegible]
6	[illegible]	[illegible]	+1.0
7	[illegible]	[illegible]	+1.8
8	[illegible]	[illegible]	+0.8
9	[illegible]	[illegible]	+4.6
10	+2.0	+3.4	+1.4
<u>Mean</u>	+0.75	+2.33	+1.58

But, it's the same person in both groups! So it's not two separate samples!

(he just took two different medicines on different days)

Paired t-test

So far, we have considered independent samples, but what can we do if samples are paired? For example, when the same patients are tested with two different treatments?

Additional hours of sleep after taking a drug.

Patient	Hyoscyamine	Hyoscine	Difference
1	+0.7	+1.9	+1.2
2	-1.6	+0.8	+2.4
3	-0.2	+1.1	+1.3
4	[illegible]	[illegible]	+1.3
5	[illegible]	[illegible]	0
6	+3.4	+4.4	+1.0
7	+3.7	+5.5	+1.8
8	+0.8	+1.6	+0.8
9	0	+4.6	+4.6
10	+2.0	+3.4	+1.4
Mean	+0.75	+2.33	+1.58

So, we take the *difference*

Paired t-test

$$t = \frac{\overline{x}_D}{s_D / \sqrt{n}}$$

t ~ t-distribution with n-1 degrees of freedom.

$\overline{x}_D$: sample mean of the difference

s_D: sample standard deviation of the difference

n: sample size

Null hypothesis: mean difference is 0 (no difference between days)
Alternate hypothesis: mean difference is not 0 (different between days)

Paired t-test (example)

$$t = \frac{\overline{x}_D}{s_D / \sqrt{n}} = \frac{1.58}{1.23/\sqrt{10}} \approx 4.06$$

For α=0.05: $t_{crit1}[9;\ 2.5\%] = -2.262$

$t_{crit2}[9;97.5\%] = +2.262$

$t > t_{crit2}$

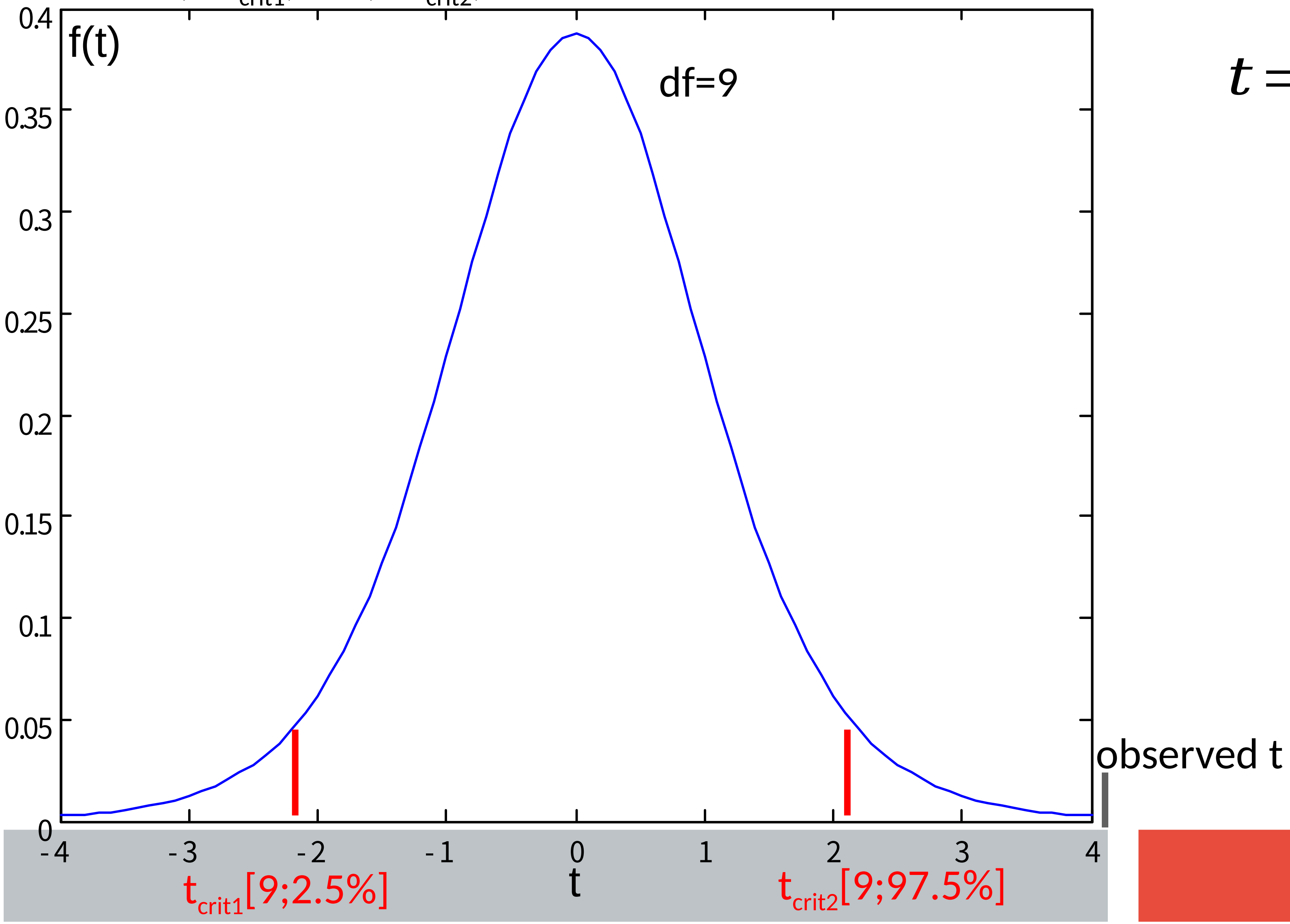
-> Based on our data we can reject the null hypothesis that

-> On average, patients could sleep longer when taking hyoscine (M=2.33, SD=1.79 hours) compared to hyoscyamine (M=0.75, SD=2 hours). This difference was significant (paired t-test: t[9]=4.06, p=0.0028).

Paired t-test (example)

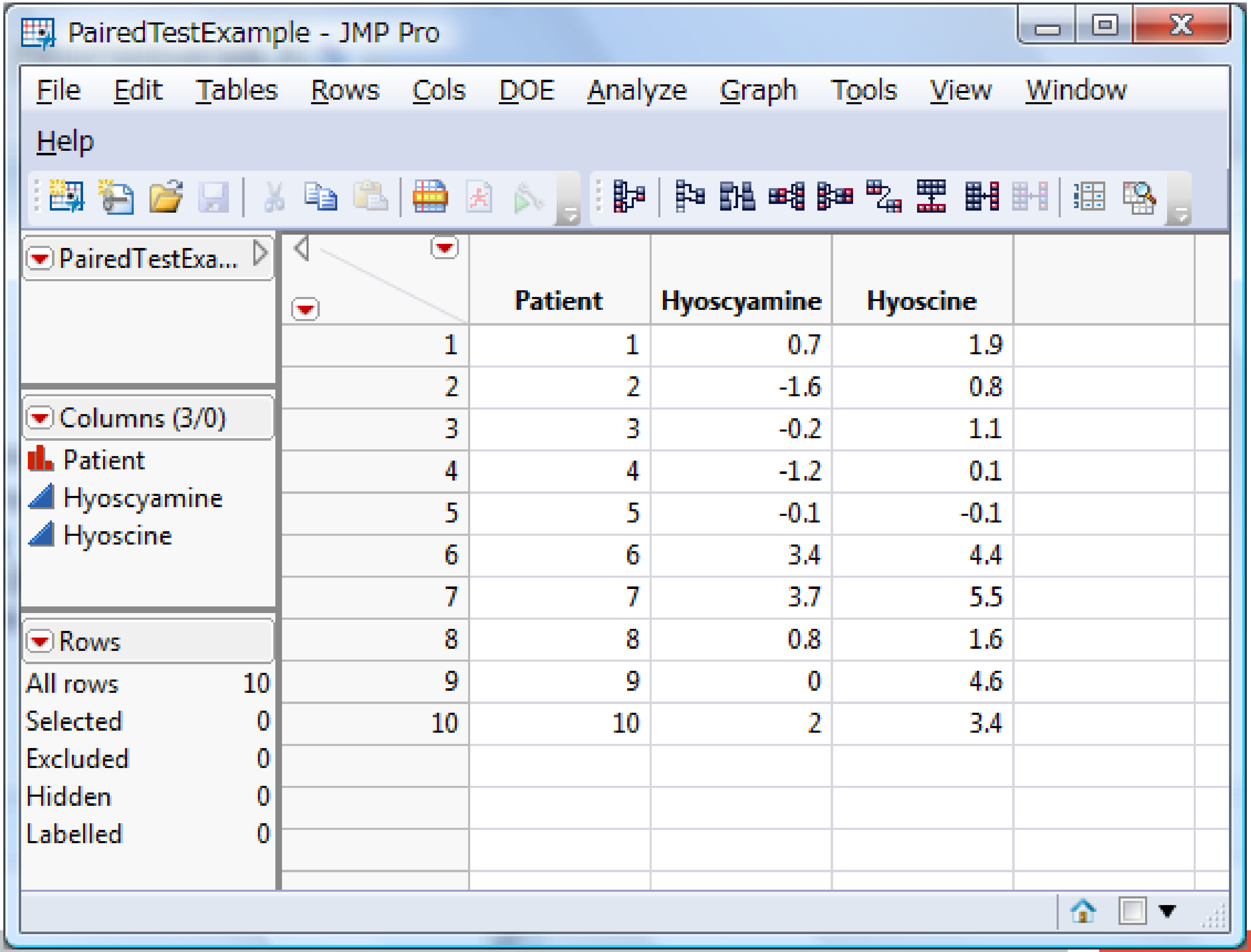
Under the null hypothesis:

$P(t<t_{crit1}) + P(t>t_{crit2}) = \alpha = 0.05$



$$t = \frac{\overline{x}_D}{s_D / \sqrt{n}}$$

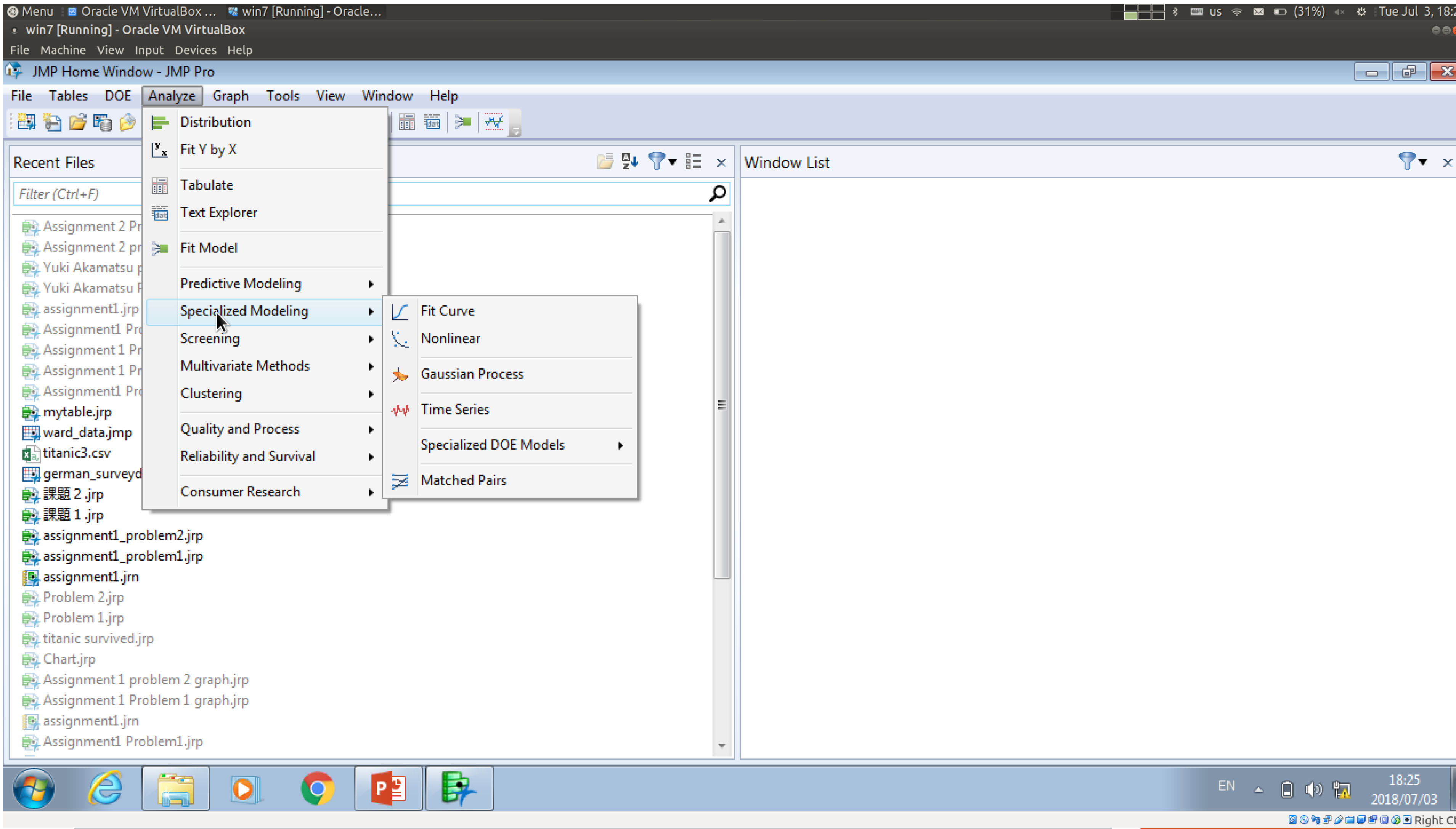
Paired t-test (JMP)



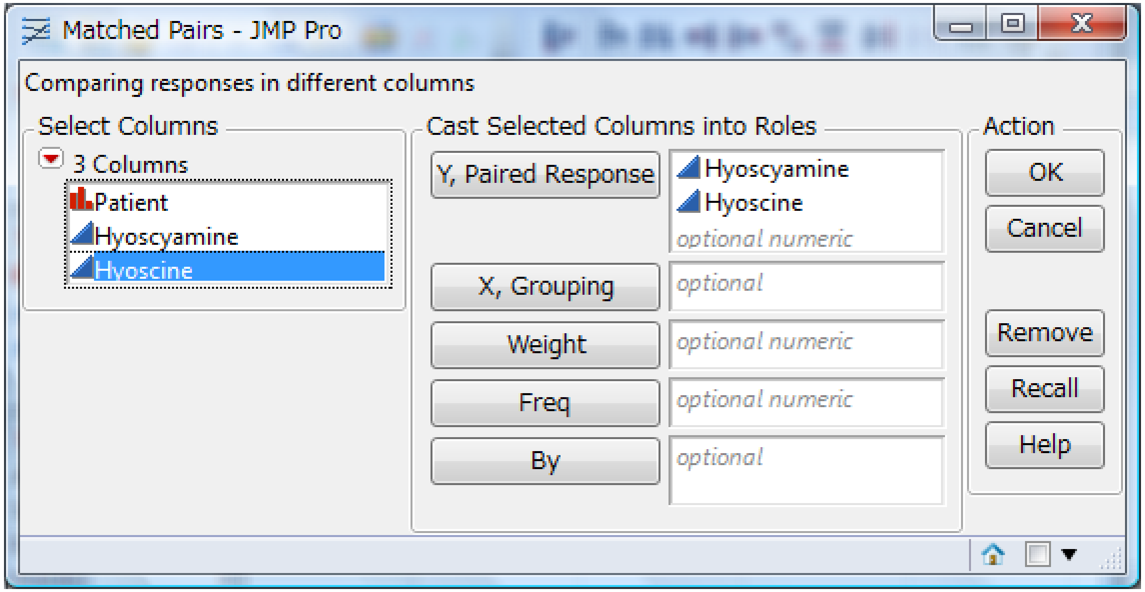
	Patient	Hyoscyamine	Hyoscine
1	1	0.7	1.9
2	2	-1.6	0.8
3	3	-0.2	1.1
4	4	-1.2	0.1
5	5	-0.1	-0.1
6	6	3.4	4.4
7	7	3.7	5.5
8	8	0.8	1.6
9	9	0	4.6
10	10	2	3.4

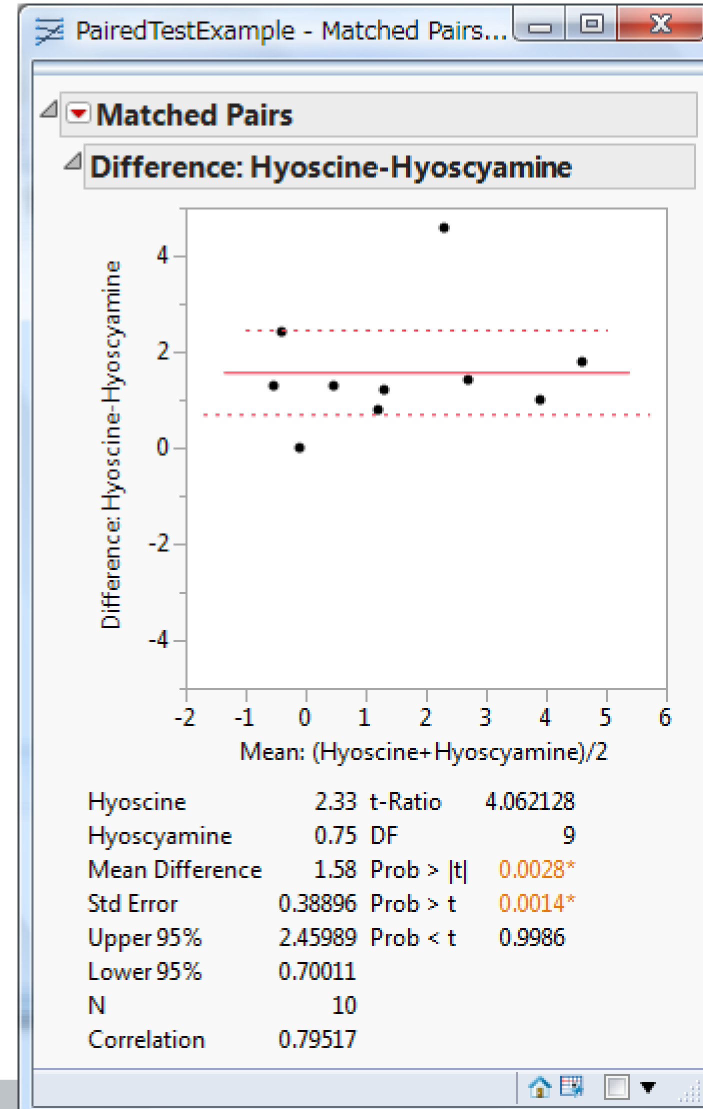
Paired t-test (JMP)

In JMP 13+, it is under: Analyze->Specialized Modeling->Matched Pairs



Paired t-test (JMP)





The report shows descriptive statistics plus t-ratio, DF=degrees of freedom,
and p-values for two-tailed and one-tailed tests.

Summary: t-tests

- Two means can be compared with the t-test when

a) the two random variables are normally distributed and independent.

or b) the difference of two paired variables is normally distributed.

- Depending on whether the two random variables are dependent or independent, or have equal or unequal variance, we can derive the test statistic t which follows a t-distribution.

- We compare the observed t with the critical t-value(s) given the null hypothesis (usually no difference), the degrees of freedom, and the α level.

- We reject the null hypothesis if the observed t is more extreme than the critical t-value(s).

Other tests

T-test is called a "parametric test" because it *assumes* something about the distribution of observations (that it is normal)

What do we do, if **sample size is small**, but **normal distribution cannot be assumed**?

→ "Non-parametric" tests:

Mann-Whitney U test (a.k.a Wilcoxon rank-sum test)
(independent conditions)

Wilcoxon signed rank test
(paired conditions)

What do we do, if we have **more than two groups (means)**?

-> ANOVA (Analysis of variance)