

Inference for Distributions

IPS Chapter 7

- 7.1: Inference for the Mean of a Population
- 7.2: Comparing Two Means
- 7.3: Optional Topics in Comparing Distributions

Inference for Distributions

7.1 Inference for the Mean of a Population

Objectives

7.1 Inference for the mean of a population

- The t distributions
- The one-sample t confidence interval
- The one-sample t test
- Matched pairs t procedures
- Robustness
- Power of the t -test
- Inference for non-normal distributions

Sweetening colas



Cola manufacturers want to test how much the sweetness of a new cola drink is affected by storage. The sweetness loss due to storage was evaluated by 10 professional tasters (by comparing the sweetness before and after storage):

	Taster	Sweetness loss
■	1	2.0
■	2	0.4
■	3	0.7
■	4	2.0
■	5	-0.4
■	6	2.2
■	7	-1.3
■	8	1.2
■	9	1.1
■	10	2.3

Obviously, we want to test if storage results in a loss of sweetness, thus:

$$H_0: \mu = 0 \text{ versus } H_a: \mu > 0$$

This looks familiar. However, here we do not know the population parameter σ .

- The population of all cola drinkers is too large.
- Since this is a new cola recipe, we have no population data.

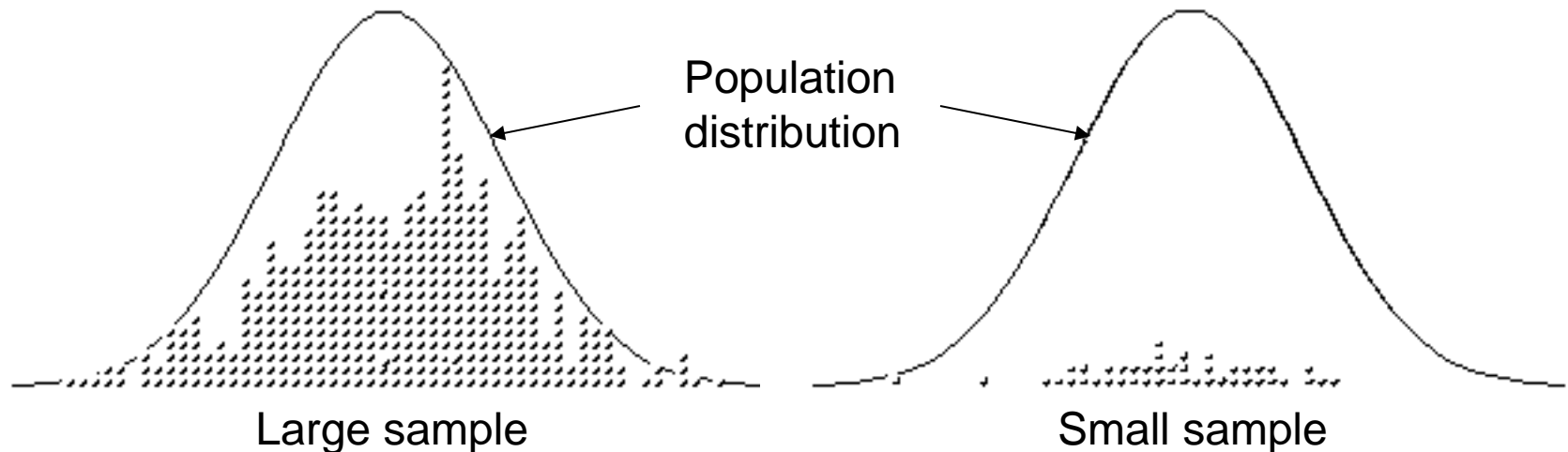
This situation is very common with real data.

When σ is unknown

The sample standard deviation s provides an estimate of the population standard deviation σ .

■ When the sample size is large, the sample is likely to contain elements representative of the whole population. Then s is a good estimate of σ .

■ But when the sample size is small, the sample contains only a few individuals. Then s is a mediocre estimate of σ .



Standard deviation s – standard error s/\sqrt{n}

For a sample of size n ,
the sample standard deviation s is:

$$s = \sqrt{\frac{1}{n-1} \sum (x_i - \bar{x})^2}$$

$n - 1$ is the “degrees of freedom.”

The value s/\sqrt{n} is called the standard error of the mean **SEM**.

Scientists often present sample results as mean \pm SEM.



A study examined the effect of a new medication on the seated systolic blood pressure. The results, presented as mean \pm SEM for 25 patients, are 113.5 ± 8.9 .

What is the standard deviation s of the sample data?

$$\text{SEM} = s/\sqrt{n} \quad \Leftrightarrow \quad s = \text{SEM} \cdot \sqrt{n}$$

$$s = 8.9 \cdot \sqrt{25} = 44.5$$

The t distributions

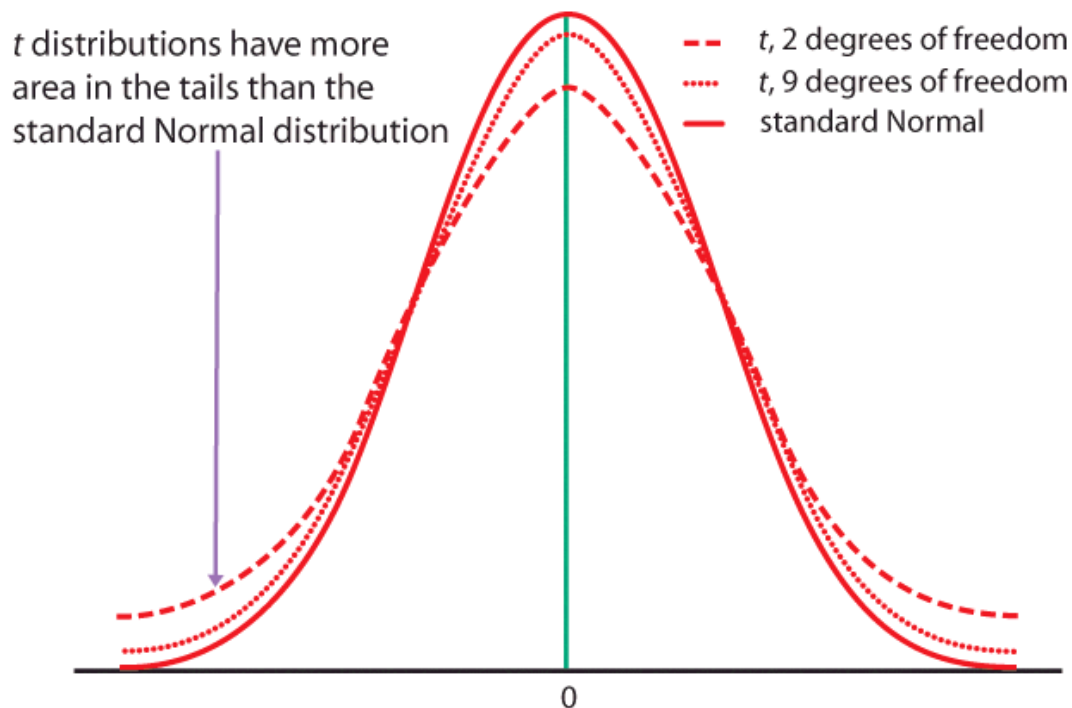
Suppose that an SRS of size n is drawn from an $N(\mu, \sigma)$ population.

- When σ is known, the sampling distribution is $N(\mu, \sigma/\sqrt{n})$.
- When σ is estimated from the sample standard deviation s , the sampling distribution follows a **t distribution $t(\mu, s/\sqrt{n})$ with degrees of freedom $n - 1$.**

$$t = \frac{\bar{x} - \mu}{s/\sqrt{n}} \quad \text{is the one-sample } t \text{ statistic.}$$

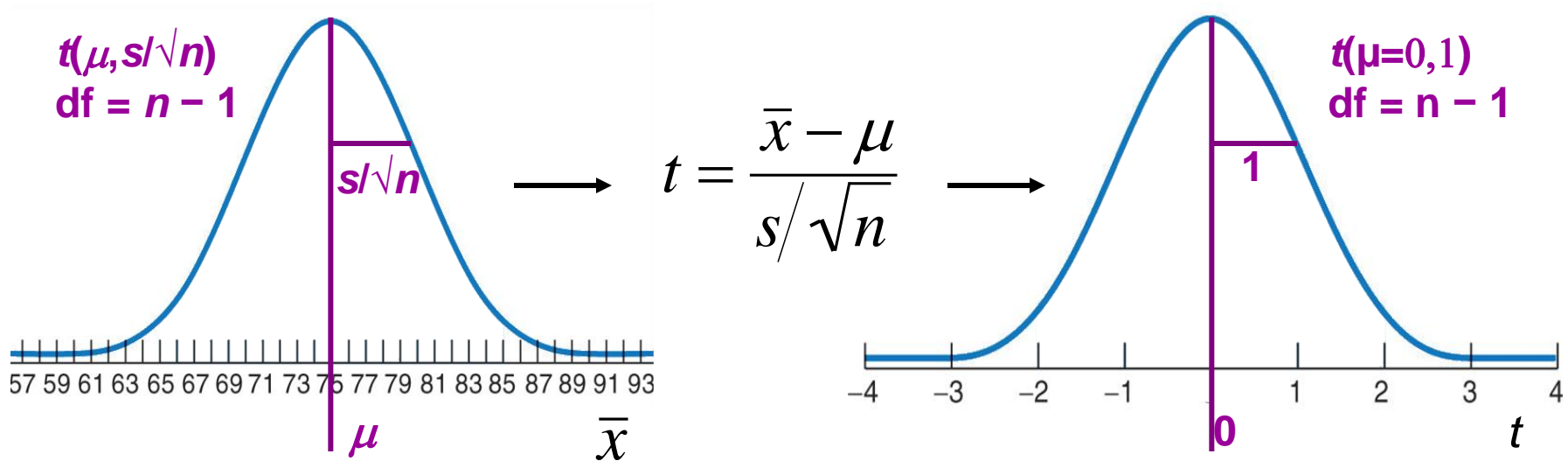
When n is very large, s is a very good estimate of σ , and the corresponding t distributions are very close to the normal distribution.

The t distributions become wider for smaller sample sizes, reflecting the lack of precision in estimating σ from s .



Standardizing the data before using Table D

As with the normal distribution, the first step is to standardize the data. Then we can use **Table D** to obtain the area under the curve.



Here, μ is the mean (center) of the sampling distribution, and the standard error of the mean s/\sqrt{n} is its standard deviation (width). You obtain s , the standard deviation of the sample, with your calculator.

Table D

When σ is unknown,
we use a t distribution
with “ $n-1$ ” degrees of
freedom (df).

Table D shows the
 z -values and t -values
corresponding to
landmark P-values/
confidence levels.

When σ is known, we
use the normal
distribution and the
standardized z -value.

df	Upper tail probability p											
	.25	.20	.15	.10	.05	.025	.02	.01	.005	.0025	.001	.0005
1	1.000	1.376	1.963	3.078	6.314	12.71	15.89	31.82	63.66	127.3	318.3	636.6
2	0.816	1.061	1.386	1.886	2.920	4.303	4.849	6.965	9.925	14.09	22.33	31.60
3	0.765	0.978	1.250	1.638	2.353	3.182	3.482	4.541	5.841	7.453	10.21	12.92
4	0.741	0.941	1.190	1.533	2.132	2.776	2.999	3.747	4.604	5.598	7.173	8.610
5	0.727	0.920	1.156	1.476	2.015	2.571	2.757	3.365	4.032	4.773	5.893	6.869
6	0.718	0.906	1.134	1.440	1.943	2.447	2.612	3.143	3.707	4.317	5.208	5.959
7	0.711	0.896	1.119	1.415	1.895	2.365	2.517	2.998	3.499	4.029	4.785	5.408
8	0.706	0.889	1.108	1.397	1.860	2.306	2.449	2.896	3.355	3.833	4.501	5.041
9	0.703	0.883	1.100	1.383	1.833	2.262	2.398	2.821	3.250	3.690	4.297	4.781
10	0.700	0.879	1.093	1.372	1.812	2.228	2.359	2.764	3.169	3.581	4.144	4.587
11	0.697	0.876	1.088	1.363	1.796	2.201	2.328	2.718	3.106	3.497	4.025	4.437
12	0.695	0.873	1.083	1.356	1.782	2.179	2.303	2.681	3.055	3.428	3.930	4.318
13	0.694	0.870	1.079	1.350	1.771	2.160	2.282	2.650	3.012	3.372	3.852	4.221
14	0.692	0.868	1.076	1.345	1.762	2.148	2.271	2.638	2.977	3.326	3.787	4.140
15	0.691	0.866	1.074	1.341	1.755	2.137	2.262	2.628	2.947	3.286	3.733	4.073
16	0.690	0.865	1.071	1.337	1.750	2.128	2.254	2.618	2.921	3.252	3.686	4.015
17	0.689	0.863	1.069	1.333	1.744	2.120	2.247	2.610	2.898	3.222	3.646	3.965
18	0.688	0.862	1.067	1.330	1.739	2.113	2.240	2.603	2.878	3.197	3.611	3.922
19	0.688	0.861	1.066	1.328	1.735	2.107	2.234	2.597	2.861	3.174	3.579	3.883
20	0.687	0.860	1.064	1.325	1.731	2.101	2.228	2.591	2.845	3.153	3.552	3.850
21	0.686	0.859	1.063	1.323	1.728	2.096	2.223	2.586	2.831	3.135	3.527	3.819
22	0.686	0.858	1.061	1.321	1.725	2.092	2.218	2.581	2.819	3.119	3.505	3.792
23	0.685	0.858	1.060	1.319	1.722	2.089	2.214	2.577	2.807	3.104	3.485	3.768
24	0.685	0.857	1.059	1.318	1.719	2.085	2.210	2.573	2.797	3.091	3.467	3.745
25	0.684	0.856	1.058	1.316	1.717	2.082	2.207	2.570	2.787	3.078	3.450	3.725
26	0.684	0.856	1.058	1.315	1.715	2.079	2.204	2.567	2.779	3.067	3.435	3.707
27	0.684	0.855	1.057	1.314	1.713	2.077	2.202	2.565	2.771	3.057	3.421	3.690
28	0.683	0.855	1.056	1.313	1.711	2.075	2.200	2.562	2.763	3.047	3.408	3.674
29	0.683	0.854	1.055	1.311	1.709	2.073	2.198	2.560	2.756	3.038	3.396	3.659
30	0.683	0.854	1.055	1.310	1.708	2.071	2.196	2.558	2.750	3.030	3.385	3.646
40	0.681	0.851	1.050	1.303	1.684	2.021	2.123	2.423	2.704	2.971	3.307	3.551
50	0.679	0.849	1.047	1.299	1.676	2.009	2.109	2.403	2.678	2.937	3.261	3.496
60	0.679	0.848	1.045	1.296	1.671	2.000	2.099	2.390	2.660	2.915	3.232	3.460
80	0.678	0.846	1.043	1.292	1.664	1.990	2.088	2.374	2.639	2.887	3.195	3.416
100	0.677	0.845	1.042	1.290	1.660	1.984	2.081	2.364	2.626	2.871	3.174	3.390
1000	0.675	0.842	1.037	1.282	1.646	1.962	2.056	2.330	2.581	2.813	3.098	3.300
z^*	0.674	0.841	1.036	1.282	1.645	1.960	2.054	2.326	2.576	2.807	3.091	3.291
Confidence level C												
	50%	60%	70%	80%	90%	95%	96%	98%	99%	99.5%	99.8%	99.9%

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

Table A vs. Table D

Table A gives the area to the LEFT of hundreds of z-values.

It should only be used for Normal distributions.

TABLE A Standard normal probabilities

z	.00	.01	.02	.03	.04	.05	.06	.07	.08	.09
-3.4	.0003	.0003	.0003	.0003	.0003	.0003	.0003	.0003	.0003	.0002
-3.3	.0005	.0005	.0005	.0004	.0004	.0004	.0004	.0004	.0004	.0003
-3.2	.0007	.0007	.0006	.0006	.0006	.0006	.0006	.0005	.0005	.0005
-3.1	.0010	.0009	.0009	.0009	.0008	.0008	.0008	.0008	.0007	.0007
-3.0	.0013	.0013	.0013	.0012	.0012	.0011	.0011	.0011	.0010	.0010
-2.9	.0019	.0018	.0018	.0017	.0016	.0016	.0015	.0015	.0014	.0014
-2.8	.0026	.0025	.0024	.0023	.0023	.0022	.0021	.0021	.0020	.0019
-2.7	.0035	.0034	.0033	.0032	.0031	.0030	.0029	.0028	.0027	.0026
-2.6	.0047	.0045	.0044	.0043	.0041	.0040	.0039	.0038	.0037	.0036
-2.5	.0062	.0060	.0059	.0057	.0055	.0054	.0052	.0051	.0049	.0048
-2.4	.0082	.0080	.0078	.0075	.0073	.0071	.0069	.0068	.0066	.0064
-2.3	.0107	.0104	.0102	.0099	.0096	.0094	.0091	.0089	.0087	.0084
-2.2	.0139	.0136	.0132	.0129	.0125	.0122	.0119	.0116	.0113	.0110

(...)

Table D *t* distribution critical values

df	Upper tail probability <i>p</i>											
	.25	.20	.15	.10	.05	.025	.02	.01	.005	.0025	.001	.0005
1	1.000	1.376	1.963	3.078	6.314	12.71	15.89	31.82	63.66	127.3	318.3	636.6
2	0.816	1.061	1.386	1.886	2.920	4.303	4.849	6.965	9.925	14.09	22.33	31.60
3	0.765	0.978	1.250	1.638	2.353	3.182	3.482	4.541	5.841	7.453	10.21	12.92
4	0.741	0.941	1.190	1.533	2.132	2.776	2.999	3.747	4.604	5.598	7.173	8.610
5	0.727	0.920	1.156	1.476	2.015	2.571	2.757	3.365	4.032	4.773	5.893	6.869
6	0.718	0.906	1.134	1.440	1.943	2.447	2.612	3.143	3.707	4.317	5.208	5.959
7	0.711	0.896	1.119	1.415	1.895	2.365	2.517	2.998	3.499	4.029	4.785	5.408
(...)												
100	0.677	0.845	1.042	1.290	1.660	1.984	2.081	2.364	2.626	2.871	3.174	3.390
1000	0.675	0.842	1.037	1.282	1.646	1.962	2.056	2.330	2.581	2.813	3.098	3.300
<i>z</i> *	0.674	0.841	1.036	1.282	1.645	1.960	2.054	2.326	2.576	2.807	3.091	3.291
50% 60% 70% 80% 90% 95% 96% 98% 99% 99.5% 99.8% 99.9%												
Confidence level <i>C</i>												

Table D gives the area to the RIGHT of a dozen *t* or z-values.

It can be used for *t* distributions of a given df and for the Normal distribution.

Table D also gives the middle area under a *t* or normal distribution comprised between the negative and positive value of *t* or z.

The one-sample t -confidence interval

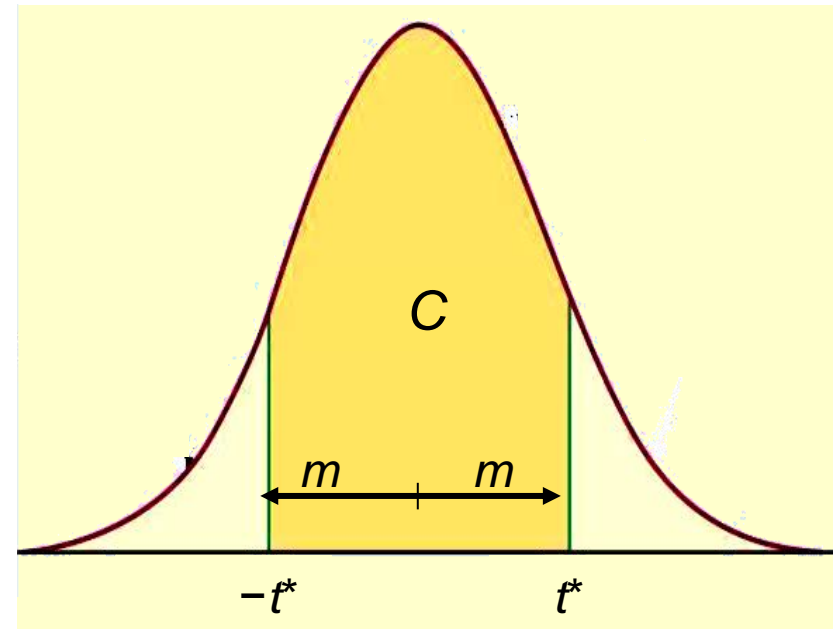
The **level C confidence interval** is an interval with probability C of containing the true population parameter.

We have a data set from a population with both μ and σ unknown. We use \bar{x} to estimate μ and s to estimate σ , using a t distribution (df $n-1$).

Practical use of t : t^*

- ▣ C is the area between $-t^*$ and t^* .
- ▣ We find t^* in the line of Table D for $df = n-1$ and confidence level C .
- ▣ The margin of error m is:

$$m = t^* s / \sqrt{n}$$



Red wine, in moderation

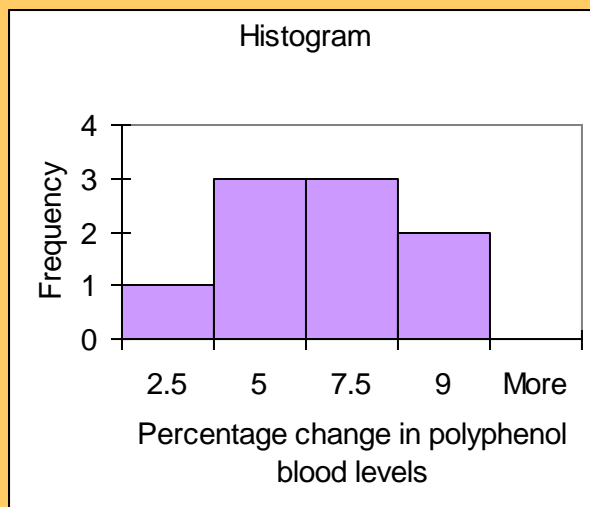
Drinking red wine in moderation may protect against heart attacks. The polyphenols it contains act on blood cholesterol, likely helping to prevent heart attacks.



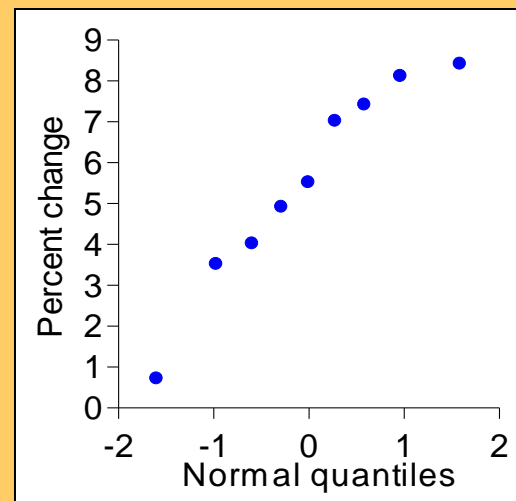
To see if moderate red wine consumption increases the average blood level of polyphenols, a group of nine randomly selected healthy men were assigned to drink half a bottle of red wine daily for two weeks. Their blood polyphenol levels were assessed before and after the study, and the percent change is presented here:

0.7 3.5 4 4.9 5.5 7 7.4 8.1 8.4

Firstly: Are the data approximately normal?



0	7
1	
2	
3	5
4	09
5	5
6	
7	04
8	14



There is a low value, but overall the data can be considered reasonably normal.



What is the 95% confidence interval for the average percent change?

Sample average = 5.5; $s = 2.517$; $df = n - 1 = 8$

8	0.706	0.889	1.108	1.397	1.860	2.306	2.449	2.896	3.355	3.833	4.501	5.041
(...)												
	50%	60%	70%	80%	90%	95%	96%	98%	99%	99.5%	99.8%	99.9%
	Confidence level C											

The sampling distribution is a t distribution with $n - 1$ degrees of freedom.

For $df = 8$ and $C = 95\%$, $t^* = 2.306$.

The margin of error m is: $m = t^* s / \sqrt{n} = 2.306 * 2.517 / \sqrt{9} \approx 1.93$.

With 95% confidence, the population average percent increase in polyphenol blood levels of healthy men drinking half a bottle of red wine daily is between 3.6% and 7.4%. Important: The confidence interval shows how large the increase is, but not if it can have an impact on men's health.

Excel

Menu: Tools/DataAnalysis: select “Descriptive statistics”

Descriptive Statistics

Input
Input Range:
Grouped By: ☒ Columns ☐ Rows
☒ Labels in first row

Output options
☐ Output Range:
☒ New Worksheet Ply:
☐ New Workbook
☒ Summary statistics
☒ Confidence Level for Mean: %
☐ Kth Largest:
☐ Kth Smallest:

OK
Cancel
Help

PercentChange	
Mean	5.5
Standard Error	0.838981 s/\sqrt{n}
Median	5.5
Mode	#N/A
Standard Deviation	2.516943
Sample Variance	6.335
Kurtosis	0.010884
Skewness	-0.7054
Range	7.7
Minimum	0.7
Maximum	8.4
Sum	49.5
Count	9
Confidence Level(95.0%)	1.934695 m

!!! Warning: do not use the function =CONFIDENCE(alpha, stdev, size)

This assumes a normal sampling distribution (stdev here refers to σ)

and uses z^* instead of t^* !!!

The one-sample t -test

As in the previous chapter, a test of hypotheses requires a few steps:

1. Stating the null and alternative hypotheses (H_0 versus H_a)
2. Deciding on a one-sided or two-sided test
3. Choosing a significance level α
4. Calculating t and its degrees of freedom
5. Finding the area under the curve with Table D
6. Stating the P-value and interpreting the result

The **P-value** is the probability, if H_0 is true, of randomly drawing a sample like the one obtained or more extreme, in the direction of H_a .

The P-value is calculated as the corresponding area under the curve, one-tailed or two-tailed depending on H_a :

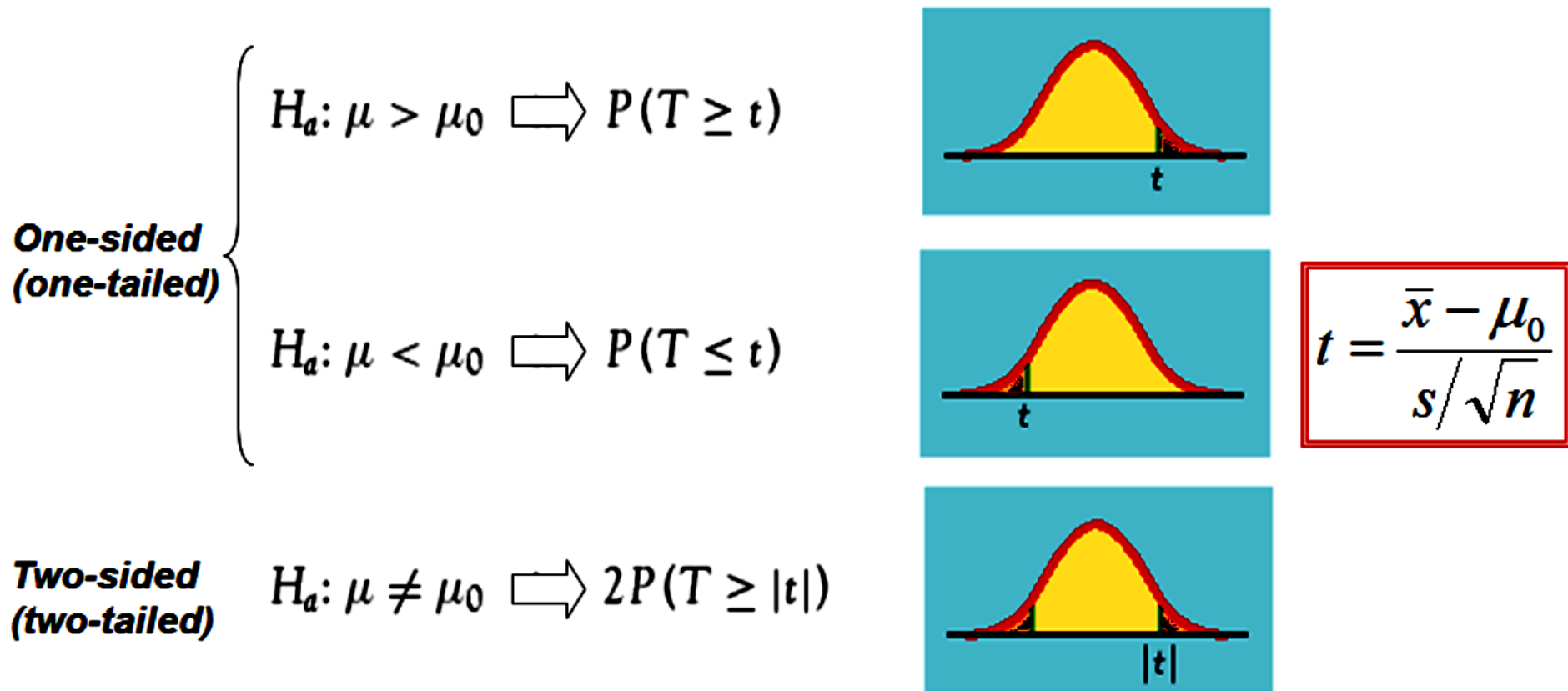


Table D

df	Upper tail probability p											
	.25	.20	.15	.10	.05	.025	.02	.01	.005	.0025	.001	.0005
1	1.000	1.376	1.963	3.078	6.314	12.71	15.89	31.82	63.66	127.3	318.3	636.6
2	0.816	1.061	1.386	1.886	2.920	4.303	4.849	6.965	9.925	14.09	22.33	31.60
3	0.765	0.978	1.250	1.638	2.353	3.182	3.482	4.541	5.841	7.453	10.21	12.92
4	0.741	0.941	1.190	1.533	2.132	2.776	2.999	3.747	4.604	5.598	7.173	8.610
5	0.727	0.920	1.156	1.476	2.015	2.571	2.757	3.365	4.032	4.773	5.893	6.869
6	0.718	0.906	1.134	1.440	1.943	2.447	2.612	3.143	3.707	4.317	5.208	5.959
7	0.711	0.896	1.119	1.415	1.895	2.365	2.517	2.998	3.499	4.029	4.785	5.408
8	0.706	0.889	1.108	1.397	1.860	2.306	2.449	2.896	3.355	3.833	4.501	5.041
9	0.703	0.883	1.100	1.383	1.833	2.262	2.398	2.821	3.250	3.690	4.297	4.781
10	0.700	0.879	1.093	1.372	1.812	2.228	2.359	2.764	3.169	3.581	4.144	4.587
11	0.697	0.876	1.088	1.363	1.796	2.201	2.328	2.718	3.106	3.497	4.025	4.437
12	0.695	0.873	1.083	1.356	1.782	2.179	2.303	2.681	3.055	3.428	3.930	4.318
13	0.694	0.870	1.079	1.350	1.771	2.160	2.282	2.650	3.012	3.372	3.852	4.221
14	0.692	0.868	1.076	1.345	1.761	2.145	2.264	2.624	2.977	3.326	3.787	4.140
15	0.691	0.866	1.074	1.341	1.753	2.131	2.249	2.602	2.947	3.286	3.733	4.073
16	0.690	0.865	1.071	1.337	1.746	2.120	2.235	2.583	2.921	3.252	3.686	4.015
17	0.689	0.863	1.069	1.333	1.740	2.110	2.224	2.567	2.899	3.223	3.646	3.965

For $df = 9$ we only look into the corresponding row.

t distribution,
9 degrees of freedom

P-value = 0.0123

$t = 2.70$

The calculated value of t is 2.7.
We find the 2 closest t values.

$$2.398 < t = 2.7 < 2.821$$

thus

$$0.02 > \text{upper tail } p > 0.01$$

For a one-sided H_a , this is the P-value (between 0.01 and 0.02);
for a two-sided H_a , the P-value is doubled (between 0.02 and 0.04).

TDIST(x, degrees_freedom, tails)

TDIST = $P(X > x)$ for a random variable X following the t distribution (x positive). Use it in place of Table C or to obtain the p-value for a positive t-value.

- ❑ X is the *standardized* value at which to evaluate the distribution (i.e., “t”).
- ❑ *Degrees_freedom* is an integer indicating the number of degrees of freedom.
- ❑ *Tails* specifies the number of distribution tails to return. If tails = 1, TDIST returns the one-tailed p-value. If tails = 2, TDIST returns the two-tailed p-value.

TINV(probability,degrees_freedom)

Gives the t-value (e.g., t^*) for a given probability and degrees of freedom.

- ❑ *Probability* is the probability associated with the two-tailed t distribution.
- ❑ *Degrees_freedom* is the number of degrees of freedom of the t distribution.

Function Arguments [?] [X]

TDIST

X = number

Deg_freedom = number

Tails = number

=

Returns the Student's t-distribution.

X is the numeric value at which to evaluate the distribution.

Formula result =

[Help on this function](#) [OK] [Cancel]

Function Arguments [?] [X]

TINV

Probability = number

Deg_freedom = number

=

Returns the inverse of the Student's t-distribution.

Probability is the probability associated with the two-tailed Student's t-distribution, a number between 0 and 1 inclusive.

Formula result =

[Help on this function](#) [OK] [Cancel]

Sweetening colas (*continued*)

Is there evidence that storage results in sweetness loss for the new cola recipe at the 0.05 level of significance ($\alpha = 5\%$)?



$H_0: \mu = 0$ versus $H_a: \mu > 0$ (one-sided test)

$$t = \frac{\bar{x} - \mu_0}{s/\sqrt{n}} = \frac{1.02 - 0}{1.196/\sqrt{10}} = 2.70$$

□ The critical value $t_\alpha = 1.833$.

$t > t_\alpha$ thus the result is significant.

□ $2.398 < t = 2.70 < 2.821$ thus $0.02 > p > 0.01$.

$p < \alpha$ thus the result is significant.

Taster	Sweetness loss
1	2.0
2	0.4
3	0.7
4	2.0
5	-0.4
6	2.2
7	-1.3
8	1.2
9	1.1
10	2.3
<hr/>	
Average	1.02
Standard deviation	1.196
Degrees of freedom	$n - 1 = 9$

	Upper tail probability p											
df	.25	.20	.15	.10	.05	.025	.02	.01	.005	.0025	.001	.0005
9	0.703	0.883	1.100	1.383	1.833	2.262	2.398	2.821	3.250	3.690	4.297	4.781

The t -test has a significant p -value. We reject H_0 .

There is a significant loss of sweetness, on average, following storage.

Sweetening colas (*continued*)



Session

One-Sample T: Loss Minitab

Test of $\mu = 0$ vs $\mu > 0$

Variable	N	Mean	StDev	SE Mean
Loss	10	1.020	1.196	0.378

Variable	95.0% Lower Bound	T	P
Loss	0.327	2.70	0.012

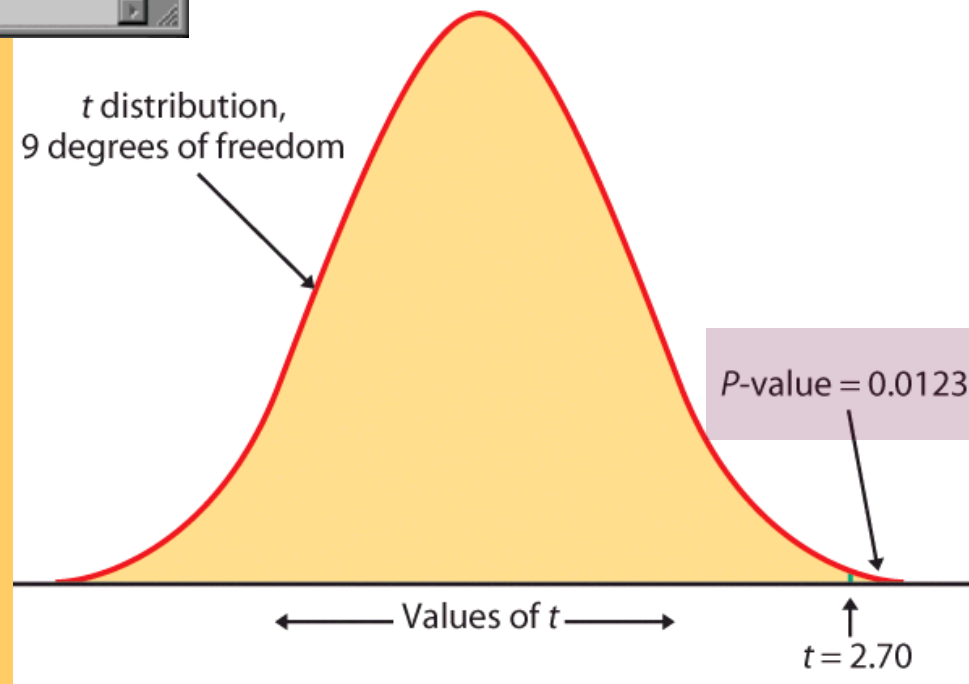
$$t = \frac{\bar{x} - \mu}{s/\sqrt{n}} = \frac{1.02 - 0}{1.196/\sqrt{10}} = 2.70$$

$$df = n - 1 = 9$$

In Excel, you can obtain the precise p-value once you have calculated t .

Use the function `dist(t , df , tails)`

`"=tdist(2.7, 9, 1),"` which gives 0.01226



Matched pairs *t* procedures

Sometimes we want to compare treatments or conditions at the individual level. These situations produce two samples that are not independent — they are related to each other. The members of one sample are identical to, or matched (paired) with, the members of the other sample.

- Example: Pre-test and post-test studies look at data collected on the same sample elements before and after some experiment is performed.
- Example: Twin studies often try to sort out the influence of genetic factors by comparing a variable between sets of twins.
- Example: Using people matched for age, sex, and education in social studies allows canceling out the effect of these potential lurking variables.

In these cases, we use the paired data to test the difference in the two population means. The variable studied becomes $X_{\text{difference}} = (X_1 - X_2)$, and

$$H_0: \mu_{\text{difference}} = 0 ; H_a: \mu_{\text{difference}} > 0 \text{ (or } < 0, \text{ or } \neq 0)$$

Conceptually, this is not different from tests on one population.

Sweetening colas (revisited)



The sweetness loss due to storage was evaluated by 10 professional tasters (comparing the sweetness **before and after** storage):

	Taster	Sweetness loss
□	1	2.0
□	2	0.4
□	3	0.7
□	4	2.0
□	5	-0.4
□	6	2.2
□	7	-1.3
□	8	1.2
□	9	1.1
□	10	2.3

We want to test if storage results in a loss of sweetness, thus:

$$H_0: \mu = 0 \text{ versus } H_a: \mu > 0$$

Although the text didn't mention it explicitly, this is a pre-/post-test design and the variable is the difference in cola sweetness before minus after storage.

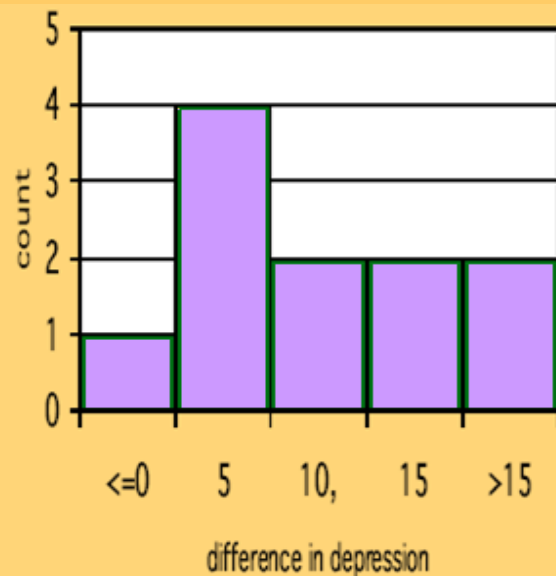
A matched pairs test of significance is indeed just like a one-sample test.

Does lack of caffeine increase depression?

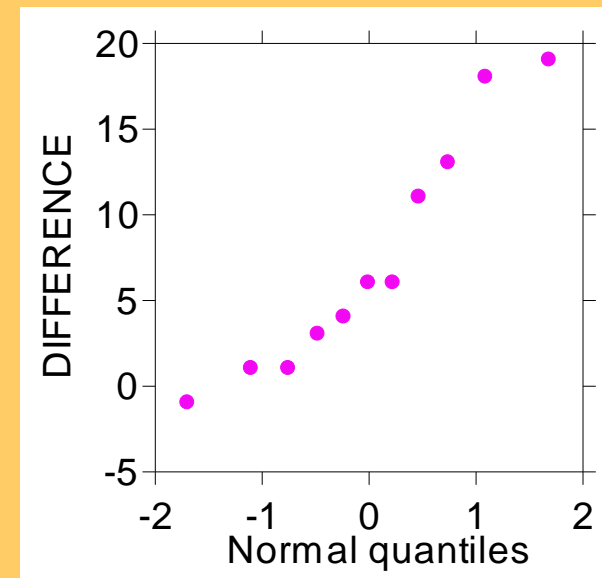
Individuals diagnosed as caffeine-dependent are deprived of caffeine-rich foods and assigned to receive daily pills. Sometimes, the pills contain caffeine and other times they contain a placebo. Depression was assessed.

Subject	Depression with Caffeine	Depression with Placebo	Placebo - Caffeine
1	5	16	11
2	5	23	18
3	4	5	1
4	3	7	4
5	8	14	6
6	5	24	19
7	0	6	6
8	0	3	3
9	2	15	13
10	11	12	1
11	1	0	-1

- There are 2 data points for each subject, but we'll only look at the difference.
- The sample distribution appears appropriate for a t -test.



11 “difference” data points.



Does lack of caffeine increase depression?

For each individual in the sample, we have calculated a difference in depression score (placebo minus caffeine).

There were 11 “difference” points, thus $df = n - 1 = 10$.

We calculate that $\bar{x} = 7.36$; $s = 6.92$

$$H_0: \mu_{\text{difference}} = 0 ; H_a: \mu_{\text{difference}} > 0$$

$$t = \frac{\bar{x} - 0}{s/\sqrt{n}} = \frac{7.36}{6.92/\sqrt{11}} = 3.53$$

Subject	Depression with Caffeine	Depression with Placebo	Placebo - Caffeine
1	5	16	11
2	5	23	18
3	4	5	1
4	3	7	4
5	8	14	6
6	5	24	19
7	0	6	6
8	0	3	3
9	2	15	13
10	11	12	1
11	1	0	-1

For $df = 10$, $3.169 < t = 3.53 < 3.581 \rightarrow 0.005 > p > 0.0025$

Caffeine deprivation causes a significant increase in depression.

SPSS statistical output for the caffeine study:

- Conducting a paired sample t -test on the raw data (caffeine and placebo)
- Conducting a one-sample t -test on difference (caffeine – placebo)

Paired Samples Test								
	Paired Differences					t	df	Sig. (2-tailed)
	Mean	Std. Deviation	Std. Error Mean	95% Confidence Interval of the Difference				
				Lower	Upper			
Placebo - Caffeine	7.364	6.918	2.086	2.716	12.011	3.530	10	.005

One-Sample Test						
	Test Value = 0					
	t	df	Sig. (2-tailed)	Mean Difference	95% Confidence Interval of the Difference	
					Lower	Upper
Difference	3.530	10	.005	7.364	2.72	12.01

Our alternative hypothesis was one-sided, thus our p-value is half of the two-tailed p-value provided in the software output (half of 0.005 = 0.0025).

Robustness

The t procedures are exactly correct when the population is distributed exactly normally. However, most real data are not exactly normal.

The t procedures are **robust** to small deviations from normality – the results will not be affected too much. Factors that strongly matter:

- ❑ **Random sampling.** The sample must be an SRS from the population.
- ❑ **Outliers and skewness.** They strongly influence the mean and therefore the t procedures. However, their impact diminishes as the sample size gets larger because of the Central Limit Theorem.

Specifically:

- ❑ When $n < 15$, the data must be close to normal and without outliers.
- ❑ When $15 > n > 40$, mild skewness is acceptable but not outliers.
- ❑ When $n > 40$, the t -statistic will be valid even with strong skewness.

Power of the t -test

The power of the one sample t -test for a specific alternative value of the population mean μ , assuming a fixed significance level α , is the probability that the test will reject the null hypothesis when the alternative value of the mean is true.

Calculation of the exact power of the t -test is a bit complex. But an approximate calculation that acts as if σ were known is almost always adequate for planning a study. This calculation is very much like that for the z -test.

When guessing σ , it is always better to err on the side of a standard deviation that is a little larger rather than smaller. We want to avoid failing to find an effect because we did not have enough data.

Does lack of caffeine increase depression?

Suppose that we wanted to perform a similar study but using subjects who regularly drink caffeinated tea instead of coffee. For each individual in the sample, we will calculate a difference in depression score (placebo minus caffeine). How many patients should we include in our new study?

In the previous study, we found that the average difference in depression level was 7.36 and the standard deviation 6.92.

We will use $\mu = 3.0$ as the alternative of interest. We are confident that the effect was larger than this in our previous study, and this increase in depression would still be considered important.

We will use $s = 7.0$ for our guessed standard deviation.

We can choose a one-sided alternative because, like in the previous study, we would expect caffeine deprivation to have negative psychological effects.

Does lack of caffeine increase depression?

How many subjects should we include in our new study? Would 16 subjects be enough? Let's compute the power of the t -test for

$$H_0: \mu_{\text{difference}} = 0 ; H_a: \mu_{\text{difference}} > 0$$

against the alternative $\mu = 3$. For a significance level α 5%, the t -test with n observations rejects H_0 if t exceeds the upper 5% significance point of $t(\text{df}:15) = 1.753$. For $n = 16$ and $s = 7$:

$$t = \frac{\bar{x} - 0}{s/\sqrt{n}} = \frac{\bar{x}}{7/\sqrt{16}} \geq 1.753 \Rightarrow \bar{x} \geq 3.06775$$

The power for $n = 16$ would be the probability that $\bar{x} \geq 3.068$ when $\mu = 3$, using $\sigma = 7$. Since we have σ , we can use the normal distribution here:

$$\begin{aligned} P(\bar{x} \geq 3.068 \text{ when } \mu = 3) &= P\left(z \geq \frac{3.068 - 3}{7/\sqrt{16}}\right) \\ &= P(z \geq 0.04) = 1 - P(z \leq 0.04) = 0.484 \end{aligned}$$

The power would be about 48 %.

Inference for non-normal distributions

What if the population is clearly non-normal and your sample is small?

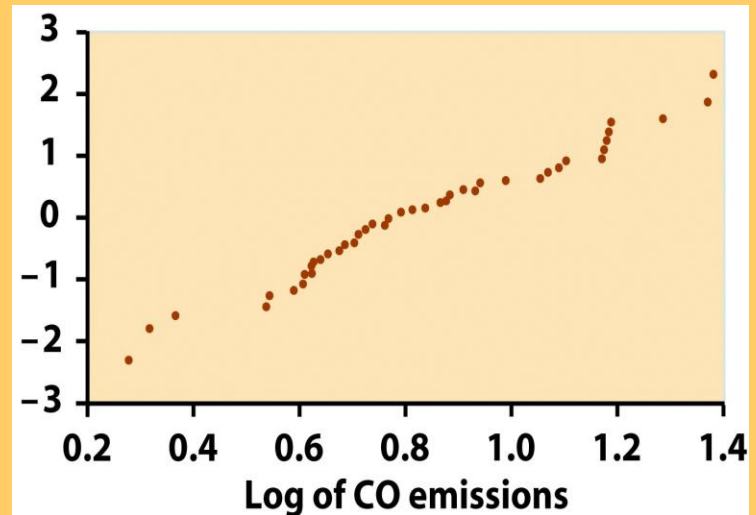
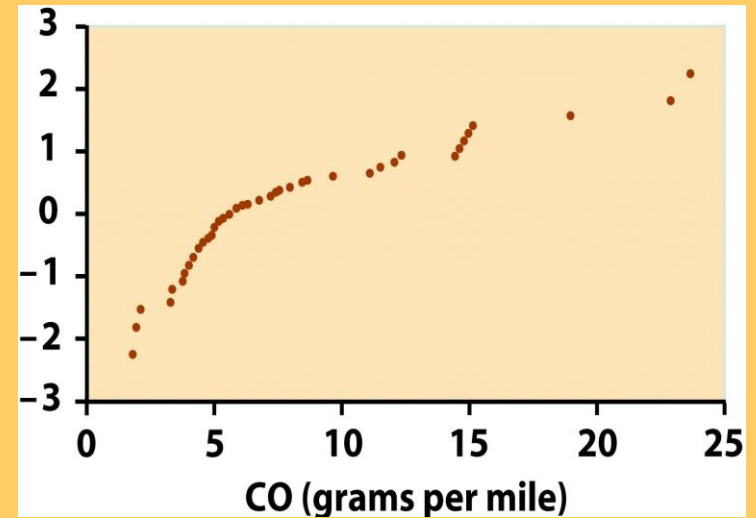
- If the data are skewed, you can attempt to **transform** the variable to bring it closer to normality (e.g., logarithm transformation). The t -procedures applied to transformed data are quite accurate for even moderate sample sizes.
- A distribution other than a normal distribution might describe your data well. Many non-normal models have been developed to provide inference procedures too.
- You can always use a **distribution-free (“nonparametric”)** inference procedure (see chapter 15) that does not assume any specific distribution for the population. But it is usually less powerful than distribution-driven tests (e.g., t test).

Transforming data

The most common transformation is the **logarithm (log)**, which tends to pull in the right tail of a distribution.

Instead of analyzing the original variable X , we first compute the logarithms and analyze the values of $\log X$.

However, we cannot simply use the confidence interval for the mean of the logs to deduce a confidence interval for the mean μ in the original scale.



**Normal quantile plots for
46 car CO emissions**

Nonparametric method: the sign test

A distribution-free test usually makes a statement of hypotheses about the median rather than the mean (e.g., “are the medians different”).

This makes sense when the distribution may be skewed.

$$H_0: \text{population median} = 0 \quad \text{vs.} \quad H_a: \text{population median} > 0$$

A simple distribution-free test is the **sign test for matched pairs**.

Calculate the matched difference for each individual in the sample.

Ignore pairs with difference 0.

The number of trials n is the count of the remaining pairs.

The test statistic is the count X of pairs with a positive difference.

P-values for X are based on the binomial $B(n, 1/2)$ distribution.

$$H_0: p = 1/2 \quad \text{vs.} \quad H_a: p > 1/2$$

Inference for Distributions

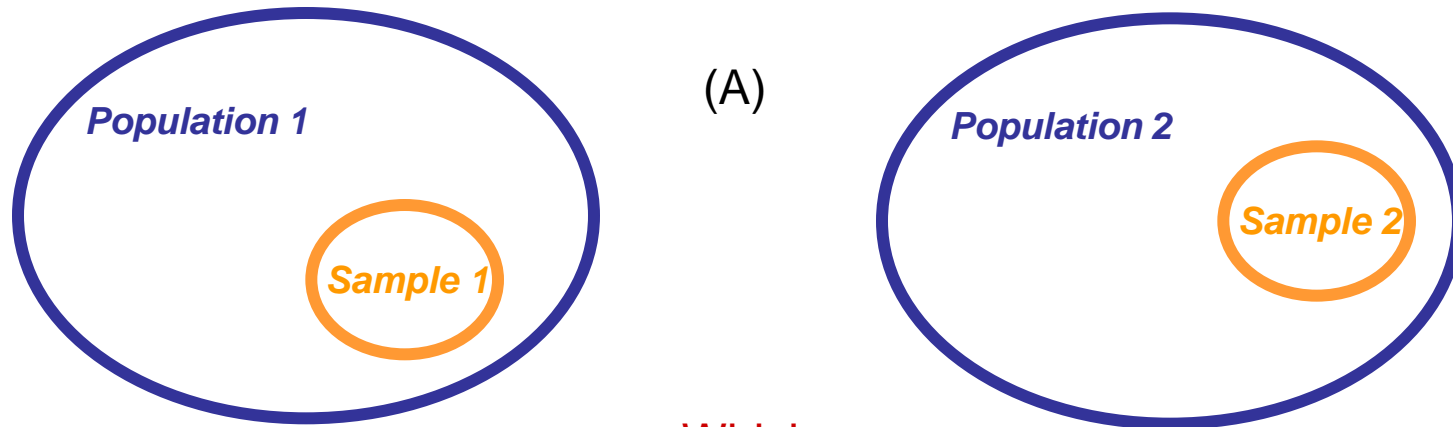
7.2 Comparing Two Means

Objectives

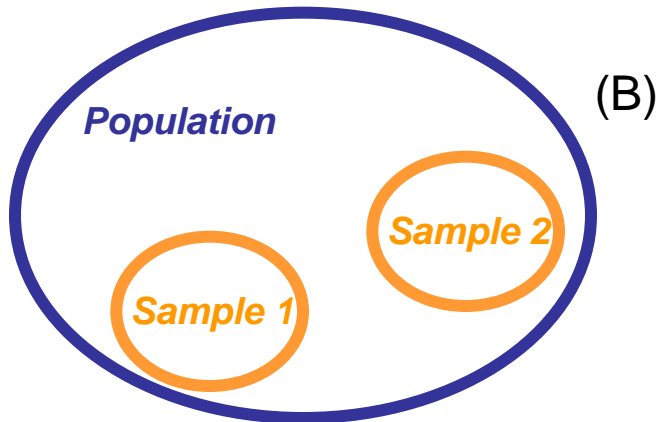
7.2 Comparing two means

- ▣ Two-sample z statistic
- ▣ Two-samples t procedures
- ▣ Two-sample t significance test
- ▣ Two-sample t confidence interval
- ▣ Robustness
- ▣ Details of the two-sample t procedures

Comparing two samples



Which
is it?



We often compare two treatments used on **independent** samples.

Is the difference between both treatments due only to variations from the random sampling, or does it reflect a true difference in population means ?

Independent samples: Subjects in one samples are completely unrelated to subjects in the other sample.

Two-sample z statistic

We have **two independent SRSs** (simple random samples) possibly coming from two distinct populations with (μ_1, σ_1) and (μ_2, σ_2) . We use \bar{x}_1 and \bar{x}_2 to estimate the unknown μ_1 and μ_2 .

When both populations are normal, the sampling distribution of $(\bar{x}_1 - \bar{x}_2)$

is also normal, with standard deviation :

$$\sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}$$

Then the **two-sample z statistic** has the standard normal $N(0, 1)$ sampling distribution.

$$z = \frac{(\bar{x}_1 - \bar{x}_2) - (\mu_1 - \mu_2)}{\sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}}$$

Two independent samples t distribution

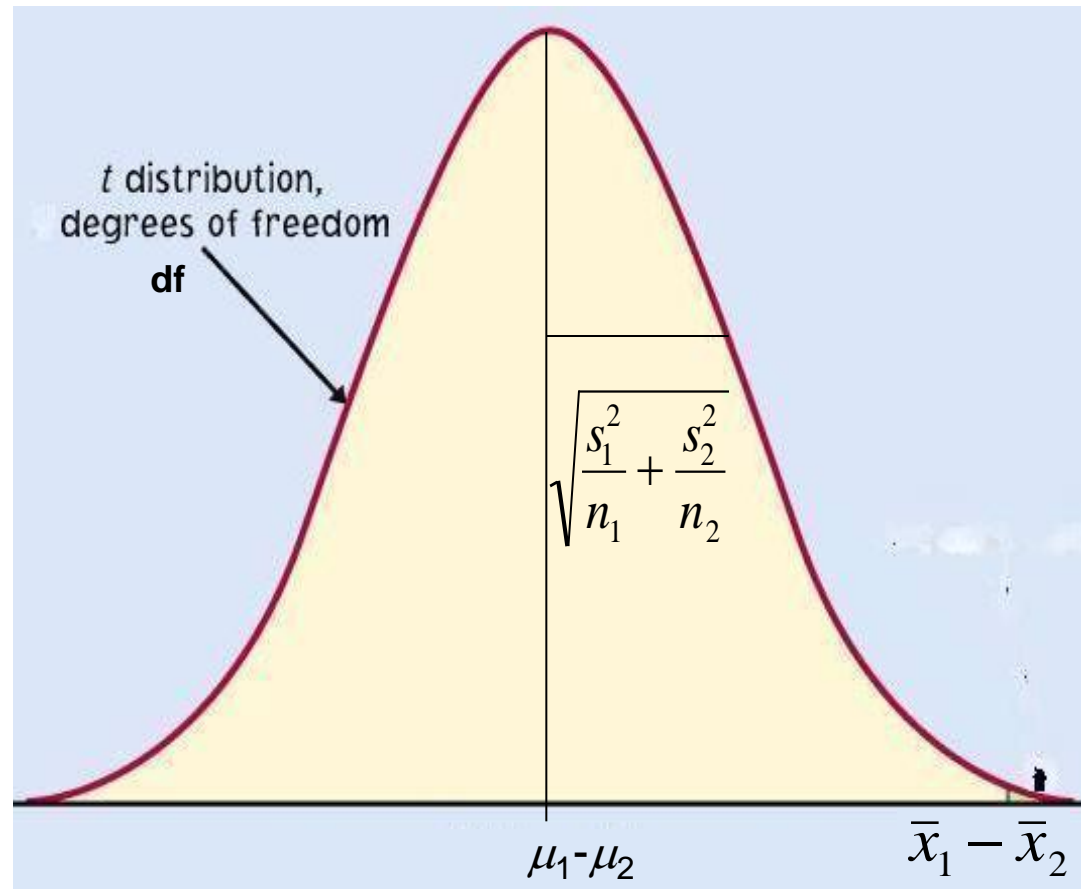
We have **two independent SRSs** (simple random samples) possibly coming from two distinct populations with (μ_1, σ_1) and (μ_2, σ_2) unknown. We use (\bar{x}_1, s_1) and (\bar{x}_2, s_2) to estimate (μ_1, σ_1) and (μ_2, σ_2) , respectively.

To compare the means, both populations should be normally distributed. However, in practice, it is enough that the two distributions have similar shapes and that the sample data contain no strong outliers.

The two-sample t statistic follows approximately the t distribution with a standard error SE (spread) reflecting variation from both samples:

$$SE = \sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}}$$

Conservatively, the degrees of freedom is equal to the smallest of $(n_1 - 1, n_2 - 1)$.



Two-sample t significance test

The null hypothesis is that both population means μ_1 and μ_2 are equal, thus their difference is equal to zero.

$$H_0: \mu_1 = \mu_2 \iff \mu_1 - \mu_2 = 0$$

with either a one-sided or a two-sided alternative hypothesis.

We find how many standard errors (SE) away from $(\mu_1 - \mu_2)$ is $(\bar{x}_1 - \bar{x}_2)$ by standardizing with t :

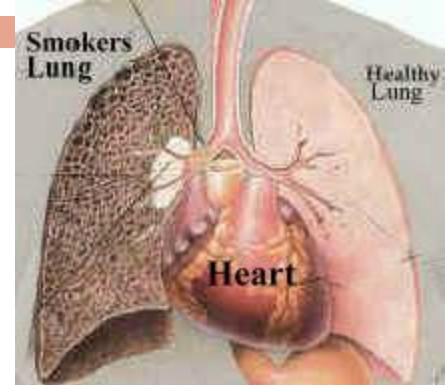
$$t = \frac{(\bar{x}_1 - \bar{x}_2) - (\mu_1 - \mu_2)}{SE}$$

Because in a two-sample test H_0 poses $(\mu_1 - \mu_2) = 0$, we simply use

With $df = \text{smallest}(n_1 - 1, n_2 - 1)$

$$t = \frac{\bar{x}_1 - \bar{x}_2}{\sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}}}$$

Does smoking damage the lungs of children exposed to parental smoking?



Forced vital capacity (FVC) is the volume (in milliliters) of air that an individual can exhale in 6 seconds.

FVC was obtained for a sample of children not exposed to parental smoking and a group of children exposed to parental smoking.

Parental smoking	FVC \bar{x}	s	n
Yes	75.5	9.3	30
No	88.2	15.1	30

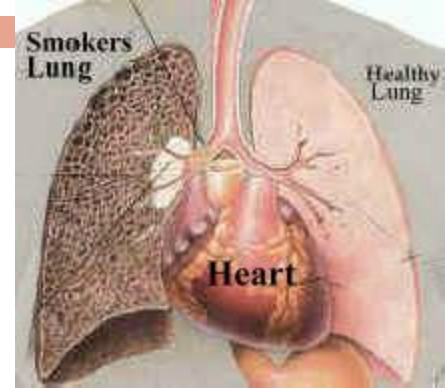


We want to know whether parental smoking decreases children's lung capacity as measured by the FVC test.

Is the mean FVC lower in the population of children exposed to parental smoking?

$$H_0: \mu_{\text{smoke}} = \mu_{\text{no}} \Leftrightarrow (\mu_{\text{smoke}} - \mu_{\text{no}}) = 0$$

$$H_a: \mu_{\text{smoke}} < \mu_{\text{no}} \Leftrightarrow (\mu_{\text{smoke}} - \mu_{\text{no}}) < 0 \text{ (one sided)}$$



The difference in sample averages follows approximately the t distribution: $t \left(0, \sqrt{\frac{s_{\text{smoke}}^2}{n_{\text{smoke}}} + \frac{s_{\text{no}}^2}{n_{\text{no}}}} \right), df = 29$

We calculate the t statistic:

$$t = \frac{\bar{x}_{\text{smoke}} - \bar{x}_{\text{no}}}{\sqrt{\frac{s_{\text{smoke}}^2}{n_{\text{smoke}}} + \frac{s_{\text{no}}^2}{n_{\text{no}}}}} = \frac{75.5 - 88.2}{\sqrt{\frac{9.3^2}{30} + \frac{15.1^2}{30}}}$$

Parental smoking	FVC \bar{x}	s	n
Yes	75.5	9.3	30
No	88.2	15.1	30

$$t = \frac{-12.7}{\sqrt{2.9 + 7.6}} \approx -3.9$$

In table D, for df 29 we find:

$$|t| > 3.659 \Rightarrow p < 0.0005 \text{ (one sided)}$$

It's a very significant difference, we reject H_0 .

Lung capacity is significantly impaired in children of smoking parents.

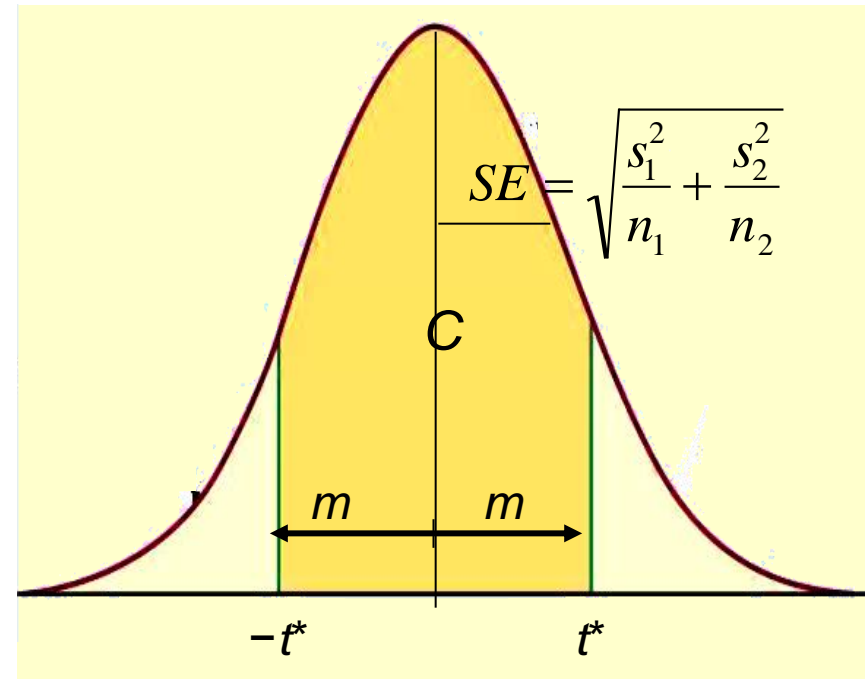
Two-sample t confidence interval

Because we have two independent samples we use the difference between both sample averages ($\bar{x}_1 - \bar{x}_2$) to estimate ($\mu_1 - \mu_2$).

Practical use of t : t^*

- ▣ C is the area between $-t^*$ and t^* .
- ▣ We find t^* in the line of Table D for $df = \text{smallest } (n_1 - 1; n_2 - 1)$ and the column for confidence level C .
- ▣ The margin of error m is:

$$m = t^* \sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}} = t^* SE$$



Common mistake !!!

A common mistake is to calculate a one-sample confidence interval for μ_1 and then check whether μ_2 falls within that confidence interval, or vice-versa.

This is WRONG because the variability in the sampling distribution for two independent samples is more complex and must take into account variability coming from both samples. Hence the more complex formula for the standard error.

$$SE = \sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}}$$

Can directed reading activities in the classroom help improve reading ability? A class of 21 third-graders participates in these activities for 8 weeks while a control classroom of 23 third-graders follows the same curriculum without the activities. After 8 weeks, all children take a reading test (scores in table).

Treatment group				Control group				Group	n	\bar{x}	s
24	61	59	46	42	33	46	37				
43	44	52	43	43	41	10	42	Treatment	21	51.48	11.01
58	67	62	57	55	19	17	55	Control	23	41.52	17.15
71	49	54		26	54	60	28				
43	53	57		62	20	53	48				
49	56	33		37	85	42					

95% confidence interval for $(\mu_1 - \mu_2)$, with $df = 20$ conservatively $\rightarrow t^* = 2.086$:

$$CI: (\bar{x}_1 - \bar{x}_2) \pm m; \quad m = t^* \sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}} = 2.086 * 4.31 \approx 8.99$$

With 95% confidence, $(\mu_1 - \mu_2)$, falls within 9.96 ± 8.99 or 1.0 to 18.9.

[illegible]

Robustness

The two-sample t procedures are more robust than the one-sample t procedures. They are the most robust when both sample sizes are equal and both sample distributions are similar. But even when we deviate from this, two-sample tests tend to remain quite robust.

→ When planning a two-sample study, choose equal sample sizes if you can.

As a guideline, a combined sample size ($n_1 + n_2$) of 40 or more will allow you to work with even the most skewed distributions.

Details of the two sample t procedures

The **true value of the degrees of freedom** for a two-sample t -distribution is quite lengthy to calculate. That's why we use an approximate value, $df = \text{smallest}(n_1 - 1, n_2 - 1)$, which errs on the conservative side (often smaller than the exact).

Computer software, though, gives the exact degrees of freedom—or the rounded value—for your sample data.

$$df = \frac{\left(\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2} \right)^2}{\frac{1}{n_1 - 1} \left(\frac{s_1^2}{n_1} \right)^2 + \frac{1}{n_2 - 1} \left(\frac{s_2^2}{n_2} \right)^2}$$

95% confidence interval for the reading ability study using the more precise degrees of freedom:

$$df = \frac{\left(\frac{11.01^2}{21} + \frac{17.15^2}{23}\right)^2}{\frac{1}{20} \left(\frac{11.01^2}{21}\right)^2 + \frac{1}{22} \left(\frac{17.15^2}{23}\right)^2}$$

$$= \frac{344.486}{9.099} = 37.86$$

$$m = t * \sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}}$$

$$m = 2.024 * 4.31 \approx 8.72$$

30	0.683	0.854	1.055	1.310	1.697	2.042	2.147	2.457
40	0.681	0.851	1.050	1.303	1.684	2.021	2.123	2.423
	50%	60%	70%	80%	90%	95%	96%	98%
Confidence level C								

t-Test: Two-Sample Assuming Unequal Variances

Excel

	<i>Treatment group</i>	<i>Control group</i>
Mean	51.476	41.522
Variance	121.162	294.079
Observations	21	23
Hypothesized Mean Difference	-	
df	38	
t Stat	2.311	
P(T<=t) one-tail	0.013	
t Critical one-tail	1.686	
P(T<=t) two-tail	0.026	
t Critical two-tail	2.024	

*t**

Independent Samples Test

		Levene's Test for Equality of Variances		t-test for Equality of Means						
		F	Sig.	t	df	Sig. (2-tailed)	Mean Difference	Std. Error Difference	95% Confidence Interval of the Difference	
									Lower	Upper
Reading Score	Equal variances assumed	2.362	.132	2.267	42	.029	9.95445	4.39189	1.09125	18.81765
	Equal variances not assumed			2.311	37.855	.026	9.95445	4.30763	1.23302	18.67588

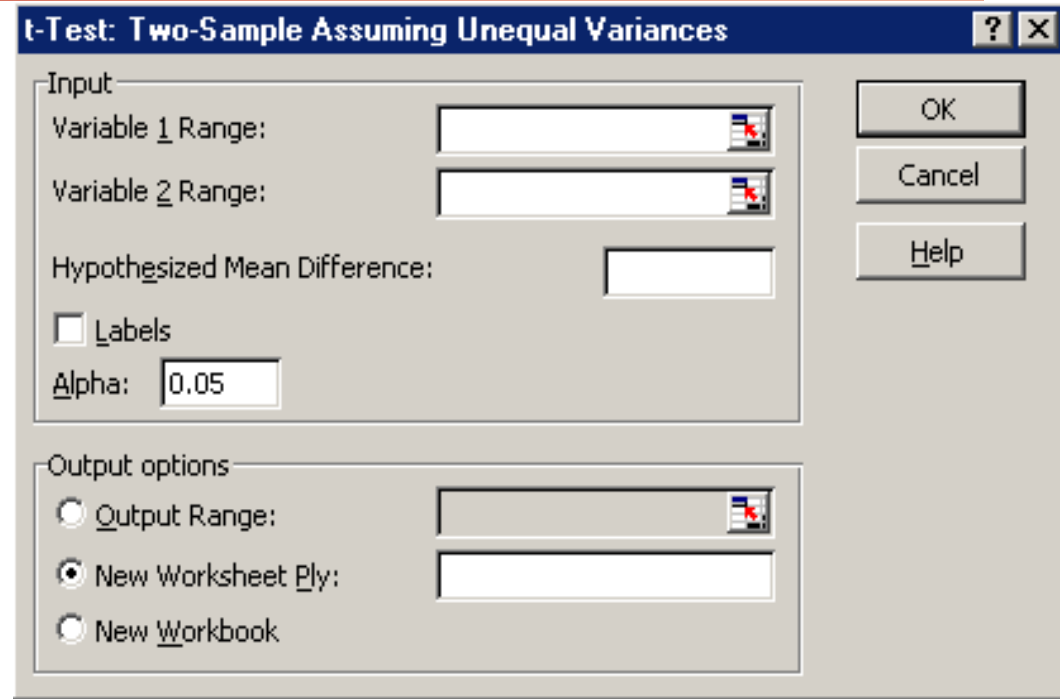
Excel

menu/tools/data_analysis →

or

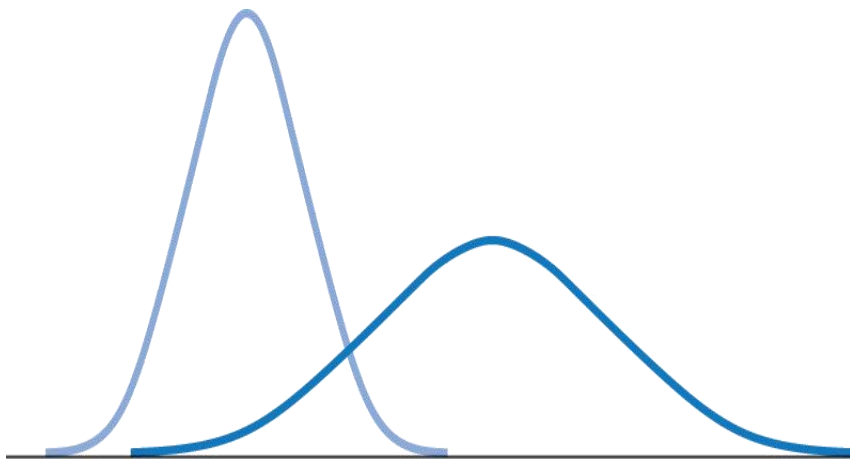
=TTEST(array1,array2,tails,type)

- *Array1* is the first data set.
- *Array2* is the second data set.
- *Tails* specifies the nature of the alternative hypothesis
(1: one-tailed; 2: two-tailed).
- *Type* is the kind of *t*-test to perform
(1: paired; 2: two-sample equal variance; 3: two-sample unequal variance).



Pooled two-sample procedures

There are two versions of the two-sample t -test: one **assuming equal variance** (“**pooled 2-sample test**”) and one **not assuming equal variance** (“**unequal**” variance, as we have studied) for the two populations. They have slightly different formulas and degrees of freedom.



Two normally distributed populations with unequal variances

The pooled (equal variance) two-sample t -test was often used before computers because it has exactly the t distribution for degrees of freedom $n_1 + n_2 - 2$.

However, the assumption of equal variance is hard to check, and thus the unequal variance test is safer.

When both population have the same standard deviation, the **pooled estimator of σ^2** is:

$$s_p^2 = \frac{(n_1 - 1)s_1^2 + (n_2 - 1)s_2^2}{(n_1 + n_2 - 2)}$$

The sampling distribution for $(x_1 - x_2)$ has exactly the t distribution with **$(n_1 + n_2 - 2)$ degrees of freedom**.

A level C confidence interval for $\mu_1 - \mu_2$ is $(\bar{x}_1 - \bar{x}_2) \pm t^* \sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}}$
(with area C between $-t^*$ and t^*)

$$t = \frac{\bar{x}_1 - \bar{x}_2}{\sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}}}$$

To test the hypothesis $H_0: \mu_1 = \mu_2$ against a one-sided or a two-sided alternative, compute the pooled two-sample t statistic for the $t(n_1 + n_2 - 2)$ distribution.

Which type of test? One sample, paired samples, two samples?

- Comparing vitamin content of bread immediately after baking vs. 3 days later (the same loaves are used on day one and 3 days later).
- Comparing vitamin content of bread immediately after baking vs. 3 days later (tests made on independent loaves).
- Average fuel efficiency for 2005 vehicles is 21 miles per gallon. Is average fuel efficiency higher in the new generation “green vehicles”?
- Is blood pressure altered by use of an oral contraceptive? Comparing a group of women not using an oral contraceptive with a group taking it.
- Review insurance records for dollar amount paid after fire damage in houses equipped with a fire extinguisher vs. houses without one. Was there a difference in the average dollar amount paid?

Inference for Distributions

7.3 Optional Topics in Comparing Distributions

Objectives

7.3 Optional topics in comparing distributions

- Inference for population spread
- The F test
- Power of the two-sample t -test

Inference for population spread

It is also possible to compare two population standard deviations σ_1 and σ_2 by comparing the standard deviations of two SRSs. However, these procedures are **not robust at all against deviations from normality**.

When s_1^2 and s_2^2 are sample variances from independent SRSs of sizes n_1 and n_2 drawn from normal populations, the F statistic

$$F = s_1^2 / s_2^2$$

has the F distribution with $n_1 - 1$ and $n_2 - 1$ degrees of freedom when $H_0: \sigma_1 = \sigma_2$ is true.

The F distributions are right-skewed and cannot take negative values.

- The peak of the F density curve is near 1 when both population standard deviations are equal.
- Values of F far from 1 in either direction provide evidence against the hypothesis of equal standard deviations.

Table E in the back of the book gives critical F -values for upper p of 0.10, 0.05, 0.025, 0.01, and 0.001. We compare the F statistic calculated from our data set with these critical values for a one-side alternative; the p -value is doubled for a two-sided alternative.

$$F \text{ has } \frac{Df_{\text{numerator}} : n_1 - 1}{Df_{\text{denom}} : n_2 - 1}$$

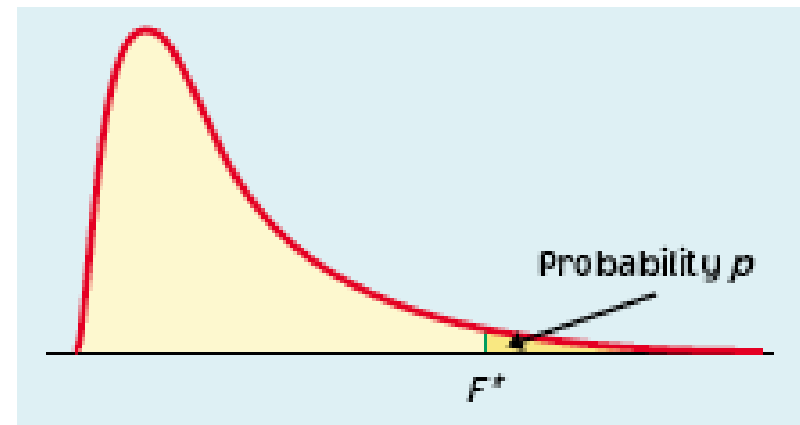
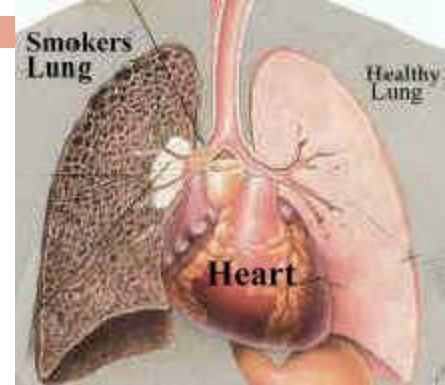


Table E F distribution critical values

$$df_{\text{num}} = n_1 - 1$$

		Degrees of freedom in the numerator							
p		1	2	3	4	5	6	7	8
1	0.100	39.86	49.50	53.59	55.83	57.24	58.20	58.91	59.44
	0.050	161.45	199.50	215.71	224.58	230.16	233.99	236.77	238.88
	0.025	647.79	799.50	864.16	899.58	921.85	937.11	948.22	956.66
	0.010	4052.2	4999.5	5403.4	5624.6	5763.6	5859	5928.4	5981.1
	0.001	405284	500000	540379	562500	576405	585937	592873	598144
2	0.100	8.53	9.00	9.16	9.24	9.29	9.33	9.35	9.37
	0.050	18.51	19.00	19.16	19.25	19.30	19.33	19.35	19.37
	0.025	38.51	39.00	39.17	39.25	39.30	39.33	39.36	39.37
	0.010	98.50	99.00	99.17	99.25	99.30	99.33	99.36	99.37
	0.001	998.50	999.00	999.17	999.25	999.30	999.33	999.36	999.37
3	0.100	5.54	5.46	5.39	5.34	5.31	5.28	5.27	5.25
	0.050	10.13	9.55	9.28	9.12	9.01	8.94	8.89	8.85
	0.025	17.44	16.04	15.44	15.10	14.88	14.73	14.62	14.54
	0.010	34.12	30.82	29.46	28.71	28.24	27.91	27.67	27.49
	0.001	167.03	148.50	141.11	137.10	134.58	132.85	131.58	130.62
4	0.100	4.54	4.32	4.19	4.11	4.05	4.01	3.98	3.95
	0.050	7.71	6.94	6.59	6.39	6.26	6.16	6.09	6.04
	0.025	12.22	10.65	9.98	9.60	9.36	9.20	9.07	8.98
	0.010	21.20	18.00	16.69	15.98	15.52	15.21	14.98	14.80
	0.001	74.14	61.25	56.18	53.44	51.71	50.53	49.66	49.00
5	0.100	4.06	3.78	3.62	3.52	3.45	3.40	3.37	3.34
	0.050	6.61	5.79	5.41	5.19	5.05	4.95	4.88	4.82
	0.025	10.01	8.43	7.76	7.39	7.15	6.98	6.85	6.76
	0.010	16.26	13.27	12.06	11.39	10.97	10.67	10.46	10.29
	0.001	47.18	37.12	33.20	31.09	29.75	28.83	28.16	27.65
6	0.100	3.78	3.46	3.29	3.18	3.11	3.05	3.01	2.98
	0.050	5.99	5.14	4.76	4.53	4.39	4.28	4.21	4.15
	0.025	8.81	7.26	6.60	6.23	5.99	5.82	5.70	5.60
	0.010	13.75	10.92	9.78	9.15	8.75	8.47	8.26	8.10
	0.001	35.51	27.00	23.70	21.92	20.80	20.03	19.46	19.03

$$df_{\text{den}} = n_2 - 1$$



Does parental smoking damage the lungs of children?

Forced vital capacity (FVC) was obtained for a sample of children not exposed to parental smoking and a group of children exposed to parental smoking.

Parental smoking	FVC \bar{x}	s	n
Yes	75.5	9.3	30
No	88.2	15.1	30

$$H_0: \sigma^2_{\text{smoke}} = \sigma^2_{\text{no}}$$

$$H_a: \sigma^2_{\text{smoke}} \neq \sigma^2_{\text{no}} \text{ (two sided)}$$

$$F = \frac{\text{larger } s^2}{\text{smaller } s^2} = \frac{15.1^2}{9.3^2} \approx 2.64$$

The degrees of freedom are 29 and 29, which can be rounded to the closest values in Table E: 30 for the numerator and 25 for the denominator.

$$2.54 < F(30,25) = 2.64 < 3.52$$

$$\Rightarrow 0.01 > 1\text{-sided } p > 0.001$$

$$\Rightarrow 0.02 > 2\text{-sided } p > 0.002$$

F*	Proba	Degrees of freedom (Df) in the numerator												
		1	2	3	4	5	6	7	8	9	10	15	20	30
25	0.100	2.92	2.53	2.32	2.18	2.09	2.02	1.97	1.93	1.89	1.87	1.77	1.72	1.66
	0.050	4.24	3.39	2.99	2.76	2.6	2.49	2.4	2.34	2.28	2.24	2.09	2.01	1.92
	0.025	5.69	4.29	3.69	3.35	3.13	2.97	2.85	2.75	2.68	2.61	2.41	2.3	2.18
	0.010	7.77	5.57	4.68	4.18	3.85	3.63	3.46	3.32	3.22	3.13	2.85	2.7	2.54
	0.001	13.88	9.22	7.45	6.49	5.89	5.46	5.15	4.91	4.71	4.56	4.06	3.79	3.52
50	0.100	2.81	2.41	2.2	2.06	1.97	1.9	1.84	1.8	1.76	1.73	1.63	1.57	1.5
	0.050	4.03	3.18	2.79	2.56	2.4	2.29	2.2	2.13	2.07	2.03	1.87	1.78	1.69
	0.025	5.34	3.97	3.39	3.05	2.83	2.67	2.55	2.46	2.38	2.32	2.11	1.99	1.87
	0.010	7.17	5.06	4.2	3.72	3.41	3.19	3.02	2.89	2.78	2.7	2.42	2.27	2.1
	0.001	12.22	7.96	6.34	5.46	4.9	4.51	4.22	4	3.82	3.67	3.2	2.95	2.68

Cautions about Inference for population spread

- Unlike t procedures, the F test and other inference procedures about variances are **not robust at all against deviations from normality**.
- The F test for comparing standard deviations or variances performs poorly if populations are not close to normal. Consequently, statisticians recommend that *these procedures are not used*.

Power of the two-sample t -test

The power of the two-sample t -test for a specific alternative value of the difference in population means ($\mu_1 - \mu_2$), assuming a fixed significance level α , is the probability that the test will reject the null hypothesis when the alternative is true.

The basic concept is similar to that for the one-sample t -test. The exact method involves the **noncentral t distribution**. Calculations are carried out with software.

You need information from a pilot study or previous research to calculate an expected power for your t -test and this allows you to plan your study smartly.

Power calculations using a noncentral t distribution

For the pooled two-sample t -test, with parameters μ_1 , μ_2 , and the common standard deviation σ we need to specify:

- ▣ An alternative that would be important to detect (i.e., a value for $\mu_1 - \mu_2$)
- ▣ The sample sizes, n_1 and n_2
- ▣ The Type I error for a fixed significance level, α
- ▣ A guess for the standard deviation σ

We find the degrees of freedom $df = n_1 + n_2 - 2$ and the value of t^* that will lead to rejection of $H_0: \mu_1 - \mu_2 = 0$

Then we calculate the **non-centrality parameter** δ

$$\delta = \frac{|\mu_1 - \mu_2|}{\sigma \sqrt{\frac{1}{n_1} + \frac{1}{n_2}}}$$

Lastly, we find the power as the probability that a noncentral t random variable with degrees of freedom df and noncentrality parameter δ will be less than t^* :

- ❑ In SAS this is 1-PROBT(t^* , df , δ). There are also several free online tools that calculate power.
- ❑ Without access to software, we can approximate the power as the probability that a standard normal random variable is greater than $t^* - \delta$, that is, $P(z > t^* - \delta)$, and use Table A.

For a ***test with unequal variances*** we can simply use the conservative degrees of freedom, but we need to guess both standard deviations and combine them for the guessed standard error:

$$\sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}$$



Java applets for power and sample size

Select the analysis to be used in your study

- CI for one proportion
- Test of one proportion
- Test comparing two proportions
- CI for one mean
- One-sample t test (or paired t)
- Two-sample t test (pooled or Satterthwaite)**
- Linear regression
- Balanced ANOVA (any model)
- Generic chi-square test
- Generic Poisson test

Run

This software is intended to be useful in planning statistical studies. It is not intended to be used for

Two-sample t test (general case)

Options Help

signal = 1

sigma2 = 1

☒ **Equal sigmas**

n1 = 25

n2 = 25

Allocation Equal

☒ **Two-tailed** Alpha .05

☐ **Equivalence**

Degrees of freedom = 48













True difference of means = .5

Power = .4101

Solve for Sample size

Power Calculator

Choose a Model and Push a Button. [Disclaimer](#).

NORMAL	Power for a given Sample Size	Sample Size for a given Power
1 Sample		
2 Sample, Equal Variances		
2 Sample, Unequal Variances		
Lognormal		
EXPONENTIAL	Power for a given Sample Size	Sample Size for a given Power
1 Sample		
2 Sample		

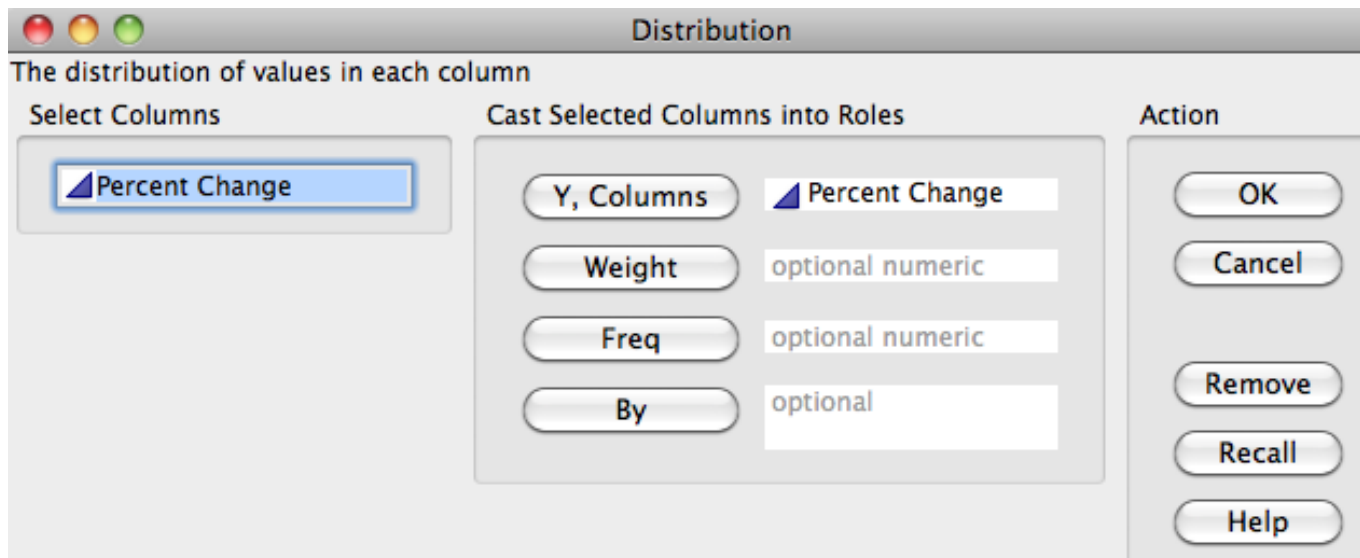
BINO
1 Sample
1 Sample Arcsin
2 Sample Arcsin
2 Sample Median
Fisher's Exact T
Proportion Res
Case Control
POIS
1 Sample
2 Sample
CORRELATION
1 Sample

Enter a "?" for the item to be calculated. Entering "?"s in positions 3 and 4 will calculate equal sample sizes for both groups.	
μ_1 The Mean of Population 1	<input type="text"/>
μ_2 The Mean of Population 2	<input type="text"/>
N_1 The Sample Size from Population 1	<input type="text"/>
N_2 The Sample Size from Population 2	<input type="text"/>
σ_1 Standard Deviation of Group 1	<input type="text"/>
σ_2 Standard Deviation of Group 2	<input type="text"/>
Significance Level The Significance Level of the test or Prob (reject null hypothesis ($H_0 : \mu_1 = \mu_2$) given it is true)	<input type="text"/>
Power The Power desired for the test or Prob (reject H_0 given that H_a is true)	<input type="text"/>
Number of Sides Specifies Alternative Hypothesis. One sided and $\mu_1 > \mu_2 \Rightarrow H_1 : \mu_1 > \mu_2$ One sided and $\mu_1 < \mu_2 \Rightarrow H_1 : \mu_1 < \mu_2$ Two sided $\Rightarrow H_1 : \mu_1 \text{ not equal } \mu_2$	<input checked="" type="radio"/> 1 Side <input type="radio"/> 2 Sides
<input type="button" value="Calculate"/>	

Alternate Slides

The following slides offer alternate software output data and examples for this presentation.

Menu: Analyze/Distribution: select “Percent Change” as your “Y, Columns” OK



A 95% confidence interval for the mean is included in “Moments”.

For a different confidence level, click the red triangle next to Percent Change

Confidence Interval:

Choose desired confidence level

OK

Percent Change

Moments

Mean	
Std Dev	2.5169426
Std Err Mean	0.8389809
Upper 95% Mean	7.4346933
Lower 95% Mean	3.5653067
N	9

s/\sqrt{n}

Confidence Intervals

Parameter	Estimate	Lower CI	Upper CI	1-Alpha
Mean	5.5	3.565307	7.434693	0.950
Std Dev	2.516943	1.700087	4.821885	0.950

Sweetening colas (*continued*)



Test Mean=value

Hypothesized Value	0
Actual Estimate	1.02
df	9
Std Dev	1.1961

JMP

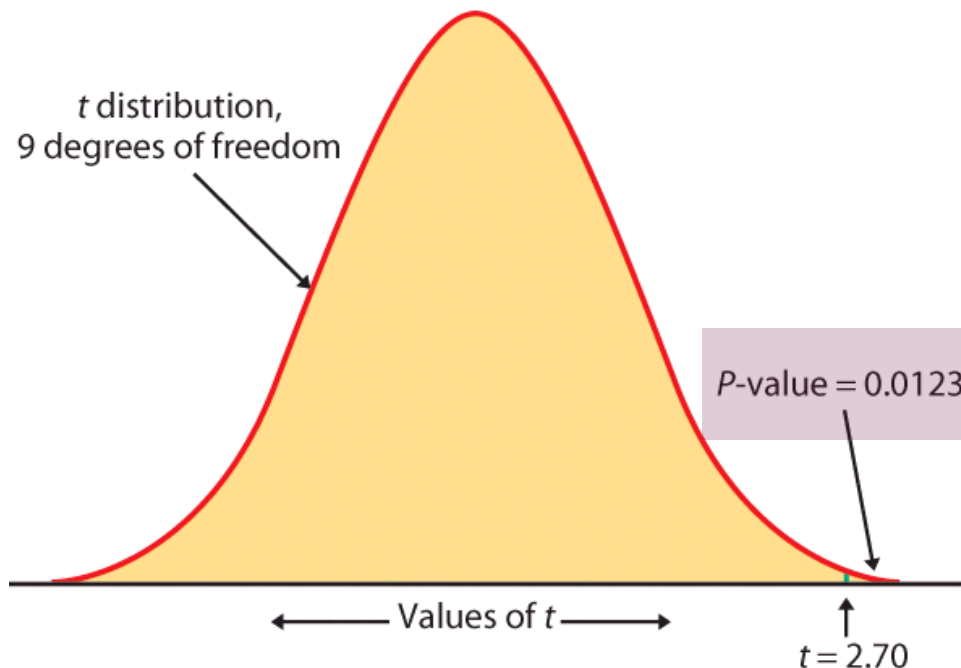
	t Test
Test Statistic	2.70
Prob > t	0.0245*
Prob > t	0.0123*
Prob < t	0.9877

$$t = \frac{\bar{x} - \mu}{s/\sqrt{n}} = \frac{1.02 - 0}{1.196/\sqrt{10}} = 2.70$$

$$df = n - 1 = 9$$

JMP provides the precise P-value

0.0123



JMP statistical output for the caffeine study:

- Conducting a paired sample t -test on the raw data (caffeine and placebo), using **Analyze: Matched Pairs**
- Conducting a one-sample t -test on difference (caffeine – placebo), using **Analyze: Distribution** for paired differences.

Difference: Depression with Placebo–Depression with Caffeine			
Depression with Placebo	11.3636	t-Ratio	3.530426
Depression with Caffeine	4	DF	10
Mean Difference	7.36364	Prob > t	0.0054*
Std Error	2.08576	Prob > t	0.0027*
Upper95%	12.011	Prob < t	0.9973
Lower95%	2.71626		
N	11		
Correlation	0.4872		

▼ **Distributions**

▼ **Placebo – Caffeine**

▼ **Test Mean=value**

Hypothesized Value

0

Actual Estimate

7.36364

df

10

Std Dev

6.9177

t Test

Test Statistic

3.5304

Prob > |t|

0.0054*

Prob > t

0.0027*

Prob < t

0.9973

Our alternative hypothesis is one-sided, thus our P-value is 0.0027, provided in the screenshot output.

Our alternative hypothesis was one-sided, thus our P-value is 0.0027, provided in the software output.

95% confidence interval for the reading ability study using the more precise degrees of freedom:

$$df = \frac{\left(\frac{11.01^2}{21} + \frac{17.15^2}{23}\right)^2}{\frac{1}{20} \left(\frac{11.01^2}{21}\right)^2 + \frac{1}{22} \left(\frac{17.15^2}{23}\right)^2}$$

$$= \frac{344.486}{9.099} = 37.86$$

30	0.683	0.854	1.055	1.310	1.697	2.042	2.147	2.457
40	0.681	0.851	1.050	1.303	1.684	2.021	2.123	2.423
	50%	60%	70%	80%	90%	95%	96%	98%
	Confidence level C							

JMP

Analyze: Fit Y by X

Enter Group as “X, Factor”, Test Score as “Y, Response”

Click ▼, select **t Test**

t Test

Treatment-Control

Assuming unequal variances

Difference	9.9545	t Ratio	2.310889
Std Err Dif	4.3076	DF	37.8554
Upper CL Dif	18.6759	Prob > t	0.0264*
Lower CL Dif	1.2330	Prob > t	0.0132*
Confidence	0.95	Prob < t	0.9868

