Chi-squared test, χ^2 test

Chi-squared test is used to determine whether two variables from the same sample are independent.

Chi-squared test examines the difference between the observed values we obtained from our sample and the expected values we have calculated.

Observed values

	Smoker	Non – smoker	Total
Male	350	150	500
Female	230	270	500
Total	580	420	1000

Expected values

	Smoker	Non – smoker	Total
Male			500
Female			500
Total	580	420	1000

Null hypothesis H_0: The two variables are independent.
Alternative hypothesis H_1: The two variables are not independent.

Expected value for each cell = $\frac{Row_{Total} \text{ x } Column_{Total}}{Total}$

$$\chi^2{}_{calc} = \sum \frac{(f_o - f_e)^2}{f_e}$$

where f_o is an observed frequency and f_e is an expected frequency.

Similar observed and expected values, small $(f_o - f_e)$, small ${\chi^2}_{calc}$.

Largely different between observed and expected values,
large $(f_o - f_e)$, large ${\chi^2}_{calc}$.

Reject H_0 when ${\chi^2}_{calc} >$ critical value.

Significance level
Significance level indicates the minimum acceptable probability that the variables are independent.

Degree of freedom = (Row – 1) (Column – 1)

Percentage Points of the Chi-Square Distribution

Degrees of Freedom	Probability of a larger value of x^2								
	0.99	0.95	0.90	0.75	0.50	0.25	0.10	0.05	0.01
1	0.000	0.004	0.016	0.102	0.455	1.32	2.71	3.84	6.63
2	0.020	0.103	0.211	0.575	1.386	2.77	4.61	5.99	9.21
3	0.115	0.352	0.584	1.212	2.366	4.11	6.25	7.81	11.34
4	0.297	0.711	1.064	1.923	3.357	5.39	7.78	9.49	13.28
5	0.554	1.145	1.610	2.675	4.351	6.63	9.24	11.07	15.09
6	0.872	1.635	2.204	3.455	5.348	7.84	10.64	12.59	16.81
7	1.239	2.167	2.833	4.255	6.346	9.04	12.02	14.07	18.48
8	1.647	2.733	3.490	5.071	7.344	10.22	13.36	15.51	20.09
9	2.088	3.325	4.168	5.899	8.343	11.39	14.68	16.92	21.67
10	2.558	3.940	4.865	6.737	9.342	12.55	15.99	18.31	23.21

Critical value of χ^2 is will be provided in exam now.

The p-value
p-value is provided when finding χ^2 on the calculator.
p-value is the probability of obtaining observed values as far or further from the expected values, assuming the variables are independent.
Reject H_0 when $p <$ significance level.

GDC skills

Casio

Menu → 2 Stat → F3 TEST → F3 CHI → F2 2WAY →
Choose Observed: Mat A → F3 DIM → m is row, n is column →
Enter data → Exit → Exit → EXE

→ Continue to find **expected value** → F6 Mat → Mat B → EXE

TI 84

2nd → Matrix → EDIT → [A] → put in row and column → Enter data
→ Stat → Tests → C: χ^2 test

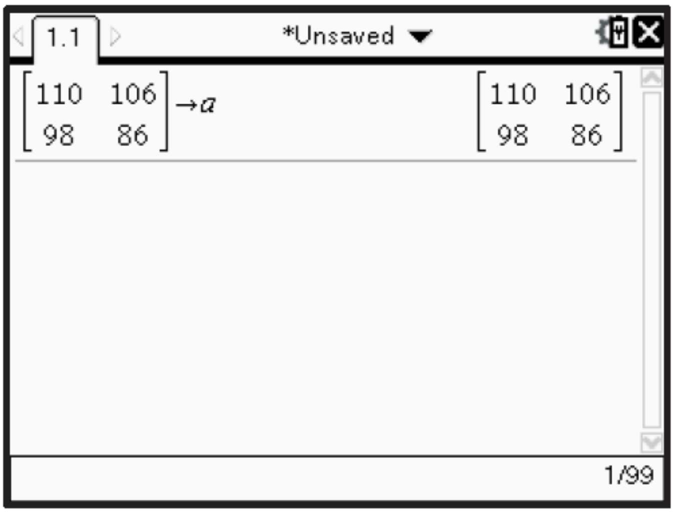
χ^2 – Test
Observed: [A]
Expected: [B]

→ Continue to find **expected value** → 2nd → Matrix → [B]

T-nspire

Menu → 7: Matrix → 1: Create → 1: Matrix → Put data → Ctrl Var → A → Enter

Then you will see the screen like below.



→ Menu → 6: Statistics → 7: Stat Tests → 8: χ^2 2-way test → Observed Matrix: a

→ Continue to find **expected value** → Var → stat.expmatrix → enter

1. David Carried out a χ^2 test at the 5% significance level to determine whether a student's height impacts their chosen subject field: sport or music.
The following table shows the results of 400 students he surveyed.

	Sport	**Music**
> 180 cm	160	50
$\leq$ **180 cm**	40	150

(a) State the null hypothesis, H_0, for this test.
(b) Complete the expected values table below.

	Sport	**Music**
> 180 cm		
$\leq$ **180 cm**		

(c) Write down the number of degrees of freedom.
(d) Write down the $p - value$ for this test.
(e) State the result of the test. Give a reason for your answer.

2. Students in a class had their gender and food preference recorded from a choice of cake or chicken wings or fish ball. The table below shoes the results.

	Cake	Chicken wings	Fish ball
Boys	**20**	**40**	**30**
Girls	**60**	**10**	**20**

A χ^2 test was conducted at the 5% significance level.

(a) State the null hypothesis, H_0, for this test.
(b) Write down the number of degrees of freedom.
(c) Write down the χ^2 statistic.
The critical value for this test is 5.99.
(d) State the result of the test. Give a reason for your answer.

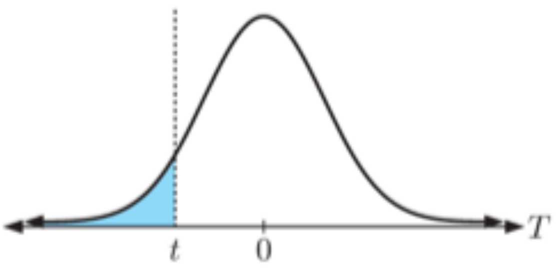
T–test

Null hypothesis H_0: $\mu = \mu_0$
Alternative hypothesis H_1:
H_1: $\mu > \mu_0$ (one-tailed hypothesis)
H_1: $\mu < \mu_0$ (one-tailed hypothesis)
H_1: $\mu \neq \mu_0$ (two-tailed hypothesis, $\mu \neq \mu_0$ means $\mu > \mu_0$ or $\mu < \mu_0$)

$$t = \frac{\bar{x} - \mu_0}{\frac{s}{\sqrt{n}}}$$

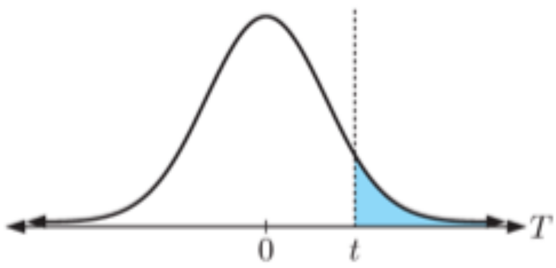
One-tailed T-test

- If H_1: $\mu < \mu_0$, we use the lower tail.



p-value $= P(T \leqslant t)$

- If H_1: $\mu > \mu_0$, we use the upper tail.



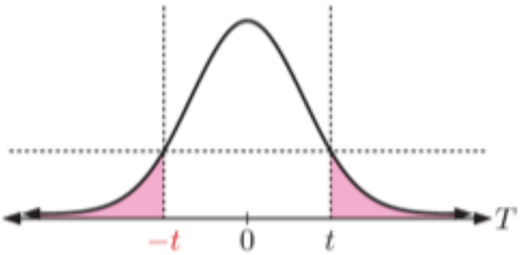
p-value $= P(T \geqslant t)$

Reject H_0 if p-value < α.

Two-tailed T-test

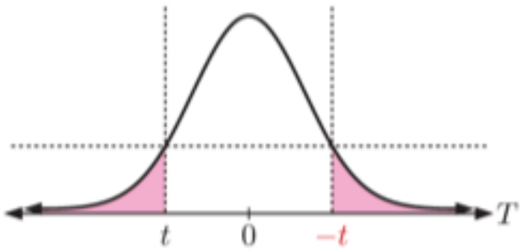
If $t \geqslant 0$,

$$p\text{-value} = \mathrm{P}(T \geqslant t \ \text{ or } \ T \leqslant -t)$$
$$= 2 \times \mathrm{P}(T \geqslant t) \qquad \{\text{symmetry}\}$$



If $t < 0$,

$$p\text{-value} = \mathrm{P}(T \geqslant -t \ \text{ or } \ T \leqslant t)$$
$$= 2 \times \mathrm{P}(T \geqslant -t) \qquad \{\text{symmetry}\}$$



So, for a two-tailed alternative hypothesis, $\boldsymbol{p\text{-value} = 2 \times \mathrm{P}(T \geqslant |\,t\,|)}$.

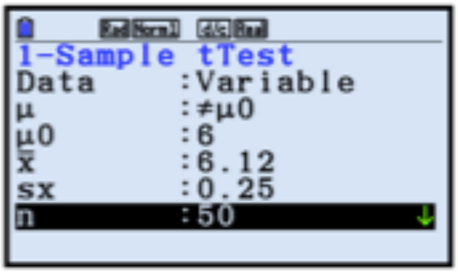
GDC Skills

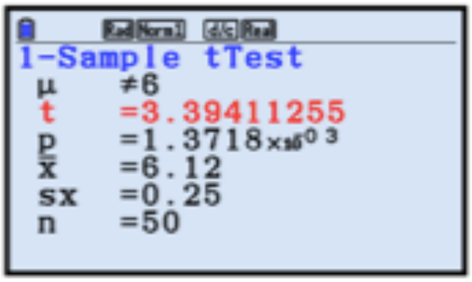
Casio

Calculate Test statistics (t)

Statistics input

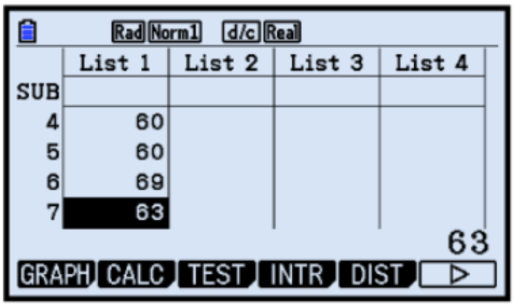
Menu → 2 Stat → F3 TEST → F2 t → F1 1-SAMPLE



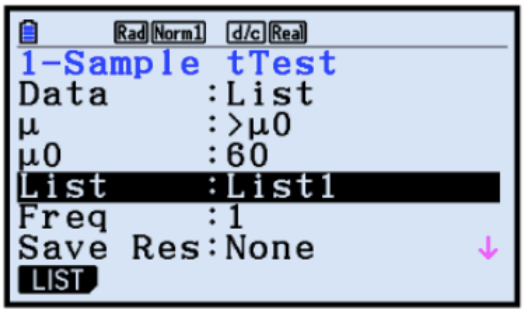


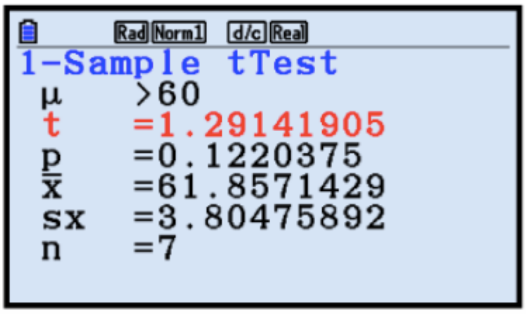
Data input

Menu → Statistics → enter data in List 1



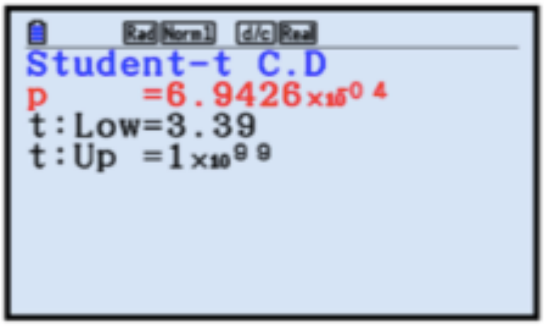
F3 Test → F2 t → F1 1-SAMPLE





Calculate the p-value

Menu → 2 Stat → 5 Dist → F2 t → F2 tcd

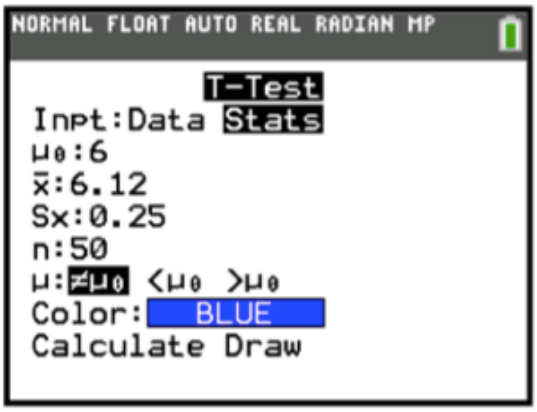


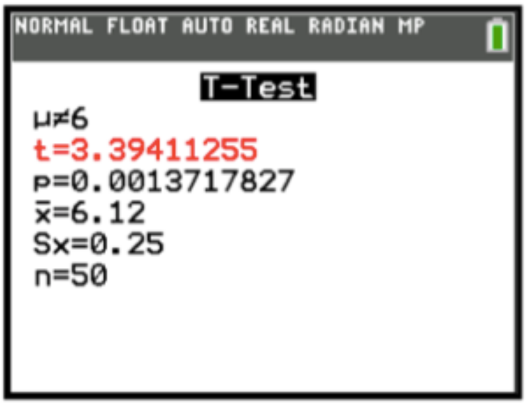
TI 84

Calculate Test statistics (t)

Statistics input

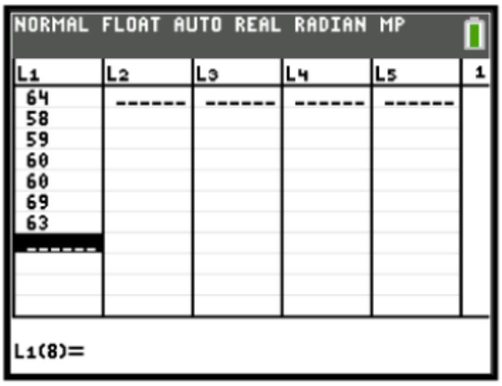
Stat → 2: T-Test



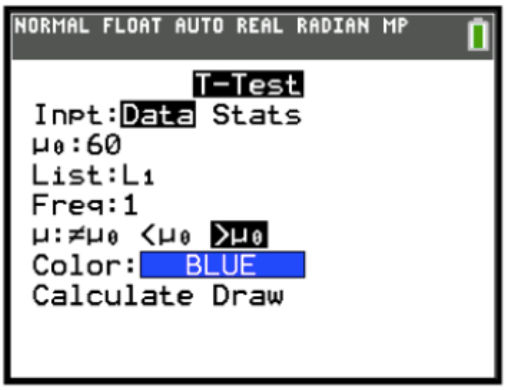


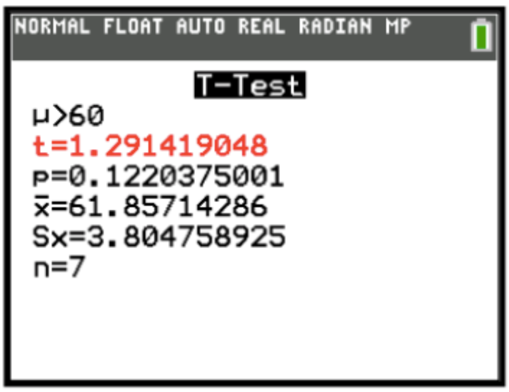
Data input

Stat→ 1: Edit → enter the data



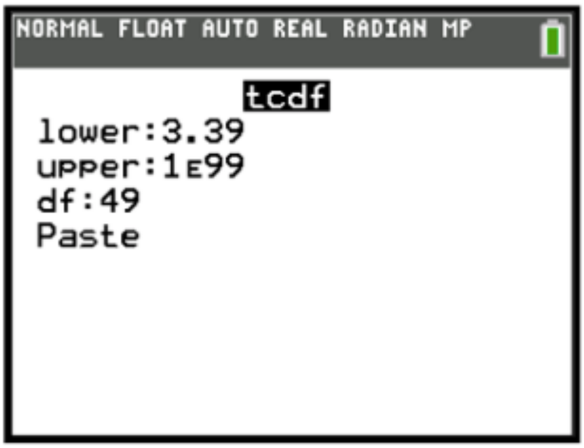
Stat → 2: T-test

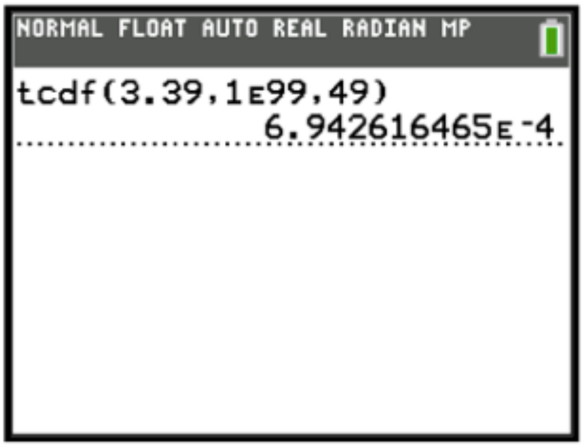




Calculate the p-value

2nd → Vars → 6: tcdf



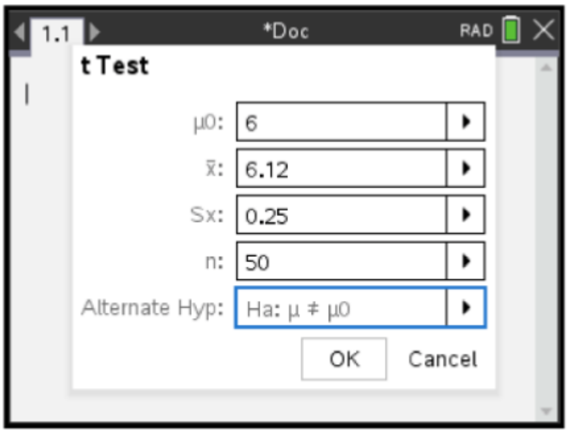


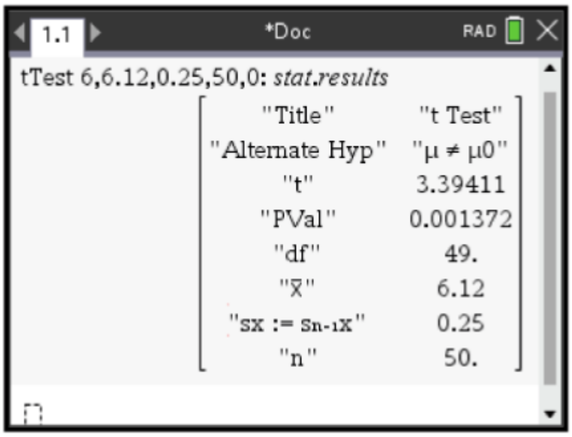
T-nspire

Calculate Test statistics (t)

Statistics input

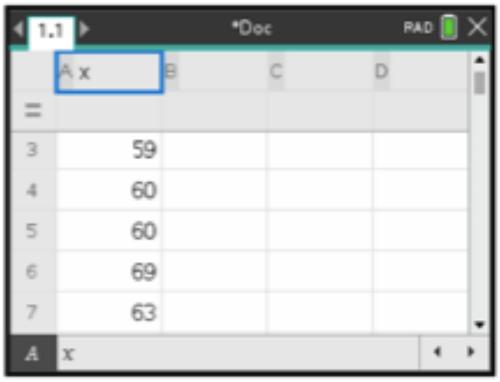
Menu → 6 Statistics → 7 Stat Test and 2 t Test



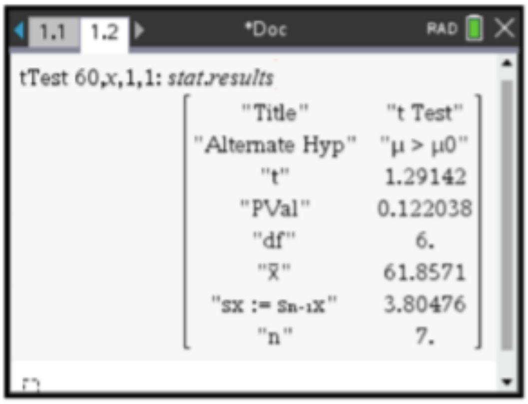


Data input

Press "On" → Add List & Spreadsheet → enter data in List A and name it x.

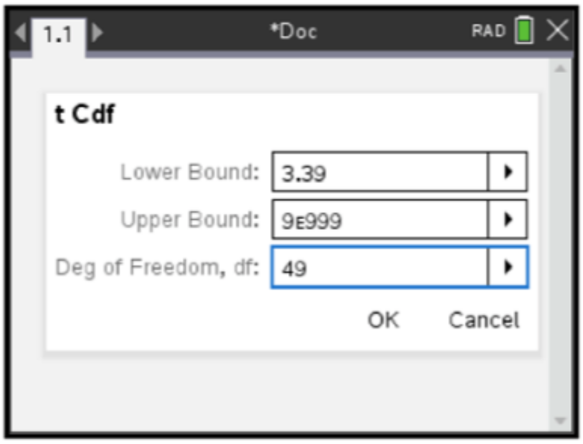


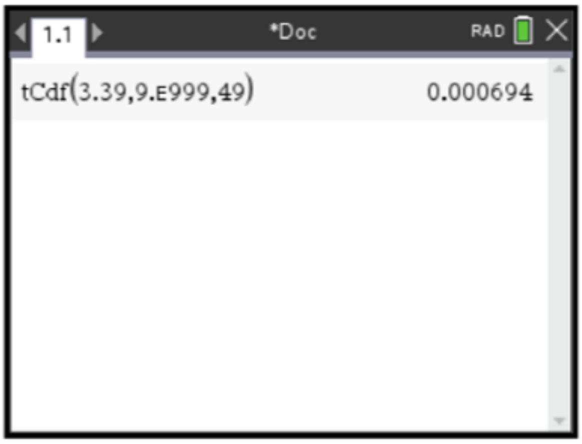
Press "On" → Add Calculator → Menu → 6 Statistics → 7 Stat test, 2 t Test



Calculate the p-value

Press "On" → Add Calculator → Menu → 5 Probability, 5 Distribution → 5 t Cdf





1. A population has known standard deviation σ = 3.97. A sample of size 36 is taken and the sample mean $\bar{x}$ = 23.75. We are required to test the hypothesis $\boldsymbol{H_0}$: μ = 25 against $\boldsymbol{H_1}$: μ < 25.

(a) Find:

(i) the test statistic

(ii) the p-value.

(b) What decision should be made at a 5% level?

2. A population has known standard deviation $\sigma = 4.23$. A sample of size 40 is taken and the sample mean $\bar{x} = 24.55$. We are required to test the hypothesis $\boldsymbol{H_0}: \mu = 21$ against $\boldsymbol{H_1}: \mu > 21$.

(a) Find:

(i) the test statistic

(ii) the p-value.

(b) What decision should be made at a 5% level?

Exercise

1. Abhinav carries out a χ^2 test at the 1% significance level to determine whether a person's gender impacts their chosen professional field: engineering, medicine or law. He surveyed 220 people and the results are shown in the table.

	Engineering	**Medicine**	**Law**
Male	55	30	25
Female	35	45	30

(a) State the null hypothesis, H_0, for this test.
(b) Calculate the expected number of male engineers.
(c) Find the $p-value$ for this test.

Abhinav rejects H_0.
(d) State a reason why Abhinav is incorrect in doing so.

2. On one day 180 flights arrived at a particular airport. The distance travelled and the arrival status for each incoming flight was recorded. The flight was then classified as on time, slightly delayed, or heavily delayed.

		Distance travelled			
		At most 500 km	Between 500 km and 5000 km	At least 5000 km	TOTAL
Arrival Status	On time	19	17	16	52
	Slightly delayed	13	18	14	45
	Heavily delayed	28	15	40	83
	TOTAL	60	50	70	180

A χ^2 test is carried out at the 10% significance level to determine whether the arrival status of incoming flights is independent of the distance travelled.

(a) State the alternative hypothesis.
(b) Calculate the expected frequency of flights travelling at most 500 km and arriving slightly delayed.

(c) Write down the number of degrees of freedom.
(d) Write down
(i) the χ^2 statistic;
(ii) the associated $p - value$.