9.4 & 3 | Hyp. Testing: Proportions and Means

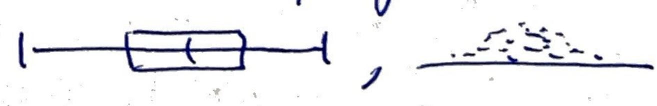
We cover both proportions and means in this document

	C. Intvl	section	H. Test	section
means, σ known	z-score	(8.1)	z	(9.2) skip
means, σ unkn.	t-score	(8.2)	t	(9.3)
proportions	z-score	(8.3)	z	(9.4)

See mymathmantra.com → stats reserve
(upper R.H. corner → "assumptions & formulas")

STEP 0:

Hypothesis Test (1-population) Conditions

- S. Random Sample: "stated" or "do it ourselves" (we)
- Not exceed 10% of pop.: $n \leq (0.10) \cdot N$ (Independence Cond.)
- Histogram to be unimodal & symmetric
  ↳ Box plot: 

  — OR —

  Sample size is greater than 30.

means 9.3

H. T. Conditions Cont.

9.4 proportions

- S. Random Sample
- Not to exceed 10% of population
- 10 success & 10 failures $\rightarrow np_0,\ n(1-p_0) \geq 10$

Notation

means 9.3

- μ = mean of the population
- μ_0 = mean from the claim of a pop.
- $\bar{X}$ = sample mean
- σ = pop. standard dev.
- S = sample std. dev.

Prop. 9.4

- P = population proportion
- P_0 = claim of the pop. proportion
- X = number of successes.
- $\hat{P}$ = sample's proportion $= x/n$

Comments

①

⊛ Hypotheses

- they come in pairs.

$\boxed{H_0}$ null hypothesis : claim, status-quo business as usuall, • no difference, no effect, no change

$\boxed{H_A}$ alternative hypothesis: "research hypoth we hope our parameter value is true.

EX Fair coin? $H_0: p = 1/2 \quad (P(H))$

$H_A: p \neq 1/2$

Test : 40 coin tosses & seeing 15 H's

Q: Is that usual or unusual

- The null hyp. gets the benefit of the doubt and is assumed to be true through-out the testing procedure

- If we find our data to be unusually extreme then, and only then will we reject the null hypothesis.

⊛ Level of Significance

- mistakes in H. Testing occur
- one mistake is to reject the null hypothesis when in fact it is true

ex Concluding a coin is not fair, but it is!

ex The defendant is "guilty", but he did not commit the crime.

ex Concluding a person has psychic abilities when she was guessing.

α

- The Level of Significance, α, is the probability of rejecting H_0 when H_0 is true.

— OR —

the probability of us making a mistake.

but NOT our fault —

- typical values are $\boxed{\alpha = 0.05}$, 5%

- If making an error against the claim has serious repercusions then use $\alpha = 0.01$

- $\alpha = 0.10$ is use when an error is less of issue {not a super bad deal if your alt. hyp. H_A is indeed true.

test statistic

Conf. Int'l

given, the pop.

Review:

8.1	• means	σ known	$SE = \sigma/\sqrt{n}$, $z = \dfrac{\bar{x} - \mu}{SE}$
8.2	• means	σ unknown	$SE = s/\sqrt{n}$, $t = \dfrac{\bar{x} - \mu}{SE}$
8.3	• proportions		$SE = \sqrt{\dfrac{\hat{p}(1-\hat{p})}{n}}$, $z = \dfrac{\hat{p} - p}{SE}$

Sample values

Hyp. Testing

9.2 (skip)	• means	σ known	$SD = \sigma/\sqrt{n}$, $z = \dfrac{\bar{x} - \mu_0}{SD}$
9.3	• means	σ - unknown	$SD = s/\sqrt{n}$, $t = \dfrac{\bar{x} - \mu_0}{SD}$
9.4	• proportions		$SD = \sqrt{\dfrac{p_0(1-p_0)}{n}}$, $z = \dfrac{\hat{p} - p_0}{SD}$

p_0 = claim

population values

- The test statistic tells us how unlikely that our data could occur given that the null hypothesis is true.

  $\boxed{\{\text{how many std. dev. is } \boxed{\text{our data}} \text{ from the claim}\}}$

- If the null is true then ideally the test statistic is near $\emptyset$.

- $\boxed{\text{The further the value of the test statistic from } \emptyset \text{, then the more suspecious we become of the null}}$

OK - Lets put these together in an EXAMPLE

EX Some art instructors reasoned that exposure to music education would help students be more creative

9.4 example

- Sixty students we studied, 30 were given the treatment (music classes). A measure of creativity showed that 19 had higher scores.
  success

If the program had no effect we would expect a 50-50% outcome in scores.
{ 1/2 higher, $\frac{1}{2}$ lower }
success failure

Question: Test the hypothesis that the probability that a art child's creativity score will increase with music education.

- Use a 5% significance level

NOTE:
we would expect { 15 above creativity scores
15 below creativity scores.

if the music edution had no effect.

OK ... lets work this on the worksheet

→

EX Music classes and Creativity

CoC MATH 140 Statistics Manual Hypothesis Testing Name Statia

STEP 0: **(a) Type of problem and table to use**

- HT for a proportion $\hat{p}$: 1- pop or 2 pop (circle) then use a z-test statistic & z-table
- HT for means μ (σ unknown): 1- pop or 2 pop (circle) then use a t-test & t-table
- HT for matched pairs means μ (σ unknown): 1- pop or 2 pop (circle) then use a z-test
- goodness-of-fit test then use a χ^2-test statistic & χ^2-table
- contingency tests (independence or homogeneity) then use a χ^2-test & χ^2-table

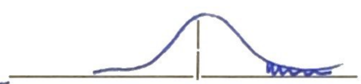
(b) Assumptions	**Justification**
SRS	not stated so assumed
Indep: n<10%	n=30 is < 10% pop.
10 succusses	19 successes ≥10 ✓
10 fails	30 - 19 = 11 fails ≥10 ✓

STEP 1: **State the Hypotheses and test-tail type (if appropriate)**

(a) H_0: p = 0.50 H_A: p >, <, ≠ 0.50 (circle)

(b) Tail: left | right | two-tail (circle)

(c) Sketch the tail(s):



STEP 2: **State the level of significance:** **α = 0. 05**

Now look up the critical value in the appropriate table { revealed in STEP 0 (a) }

z_c or t_c or X^2 (circle) = 1.645 (Last row of t-table)

STEP 3: **Compute the test statistic.** {for contingency tests Exp Val = (Row Total)(Col Total) / Grand Total }

(a) SE Formula $\sqrt{\frac{p_0 q_0}{n}}$ $\sqrt{\frac{\hat{p}\hat{q}}{n_1}+\frac{\hat{p}\hat{q}}{n_2}}$, $\hat{p}=\frac{y_1+y_2}{n_1+n_2}$ $\frac{s}{\sqrt{n}}$ $\sqrt{\frac{s_1^2}{n_1}+\frac{s_2^2}{n_2}}$ $\frac{s_d}{\sqrt{n}}$ (circle one):

SE = $\sqrt{\frac{(0.50)(1-0.50)}{30}} = 0.09129$

SE Value = $\boxed{0.0913}$

(b) test statistic = $\frac{\text{sample data} - \text{pop claim}}{\text{SE}}$, For tables use $\sum\frac{(Obs-Exp)^2}{Exp}$

z_c or t_c or X^2 (circle) = $\frac{\left(\frac{19}{30}\right)-0.50}{0.0913}$

test statistic = 1.46

STEP 4: **Compare the test statistic to the critical value:**

the test-statistic is < > (circle) than the critical value

STEP 5: **We therefore Reject | Fail-to-reject (circle) the claim**

STEP 6: **State a conclusion:**

19 of 30 is a statistical variation from the null of 15 of 30

"The study fails to show an increase in creativity in students who have music education"

Proportions (z)

C.L.Thm.

- **One sample** Assumptions — Justification

  1. Individuals are independent. — 1. SRS and $n < 10\%$ of the population.
  2. Sample is sufficiently large. — 2. Successes and failures each ≥ 10.

$\boxed{9.4}$

- **Two Groups**

  1. Groups are independent. — 1. (Think about how the data were collected.)
  2. Data in each group are independent. — 2. Both are SRSs and $n < 10\%$ of populations OR random allocation.
  3. Both samples are sufficiently large. — 3. Successes and failures each ≥ 10 for both groups.

Means (t)

- **One Sample** (df $= n - 1$)

  1. Individuals are independent. — 1. SRS and $n < 10\%$ of the population.
  2. Population has a Normal model. — 2. Histogram is unimodal and symmetric.*

$\boxed{9.3}$

- **Matched pairs** (df $= n - 1$)

  1. Data are matched. — 1. (Think about the design.)
  2. Individuals are independent. — 2. SRS and $n < 10\%$ OR random allocation.
  3. Population of differences is Normal. — 3. Histogram of differences is unimodal and symmetric.* or n>30

- **Two independent samples** (df from technology)

  1. Groups are independent. — 1. (Think about the design.)
  2. Data in each group are independent. — 2. SRSs and $n < 10\%$ OR random allocation.
  3. Both populations are Normal. — 3. Both histograms are unimodal and symmetric.* or both n>30

Distributions/Association (χ^2)

- **Goodness of fit** (df = # of cells − 1; one variable, one sample compared with population model)

  1. Data are counts. — 1. (Are they?)
  2. Data in sample are independent. — 2. SRS and $n < 10\%$ of the population.
  3. Sample is sufficiently large. — 3. All expected counts ≥ 5.

- **Homogeneity** [df $= (r - 1)(c - 1)$; many groups compared on one variable]

  1. Data are counts. — 1. (Are they?)
  2. Data in groups are independent. — 2. SRSs and $n < 10\%$ OR random allocation.
  3. Groups are sufficiently large. — 3. All expected counts ≥ 5.

- **Independence** [df $= (r - 1)(c - 1)$; sample from one population classified on two variables]

  1. Data are counts. — 1. (Are they?)
  2. Data are independent. — 2. SRSs and $n < 10\%$ of the population.
  3. Sample is sufficiently large. — 3. All expected counts ≥ 5.

Regression (t, df $= n - 2$)

- **Association of each quantitative variable** ($\beta = 0$?)

  1. Form of relationship is linear. — 1. Scatterplot looks approximately linear.
  2. Errors are independent. — 2. No apparent pattern in residuals plot.
  3. Variability of errors is constant. — 3. Residuals plot has consistent spread.
  4. Errors have a Normal model. — 4. Histogram of residuals is approximately unimodal and symmetric, or Normal probability plot reasonably straight.*

(*less critical as n increases)

Ex music education and creativity in art 9.4

9

TABLE A-3 t Distribution: Critical t Values

Degrees of Freedom	0.005 (one tail) / 0.01 (two tails)	0.01 / 0.02	0.025 / 0.05	0.05 / 0.10	0.10 / 0.20
	Area in One Tail				
	0.005	0.01	0.025	0.05	0.10
	Area in Two Tails				
	0.01	0.02	0.05	0.10	0.20
1	63.657	31.821	12.706	6.314	3.078
2	9.925	6.965	4.303	2.920	1.886
3	5.841	4.541	3.182	2.353	1.638
4	4.604	3.747	2.776	2.132	1.533
5	4.032	3.365	2.571	2.015	1.476
6	3.707	3.143	2.447	1.943	1.440
7	3.499	2.998	2.365	1.895	1.415
8	3.355	2.896	2.306	1.860	1.397
9	3.250	2.821	2.262	1.833	1.383
10	3.169	2.764	2.228	1.812	1.372
11	3.106	2.718	2.201	1.796	1.363
12	3.055	2.681	2.179	1.782	1.356
13	3.012	2.650	2.160	1.771	1.350
14	2.977	2.624	2.145	1.761	1.345
15	2.947	2.602	2.131	1.753	1.341
16	2.921	2.583	2.120	1.746	1.337
17	2.898	2.567	2.110	1.740	1.333
18	2.878	2.552	2.101	1.734	1.330
19	2.861	2.539	2.093	1.729	1.328
20	2.845	2.528	2.086	1.725	1.325
21	2.831	2.518	2.080	1.721	1.323
22	2.819	2.508	2.074	1.717	1.321
23	2.807	2.500	2.069	1.714	1.319
24	2.797	2.492	2.064	1.711	1.318
25	2.787	2.485	2.060	1.708	1.316
26	2.779	2.479	2.056	1.706	1.315
27	2.771	2.473	2.052	1.703	1.314
28	2.763	2.467	2.048	1.701	1.313
29	2.756	2.462	2.045	1.699	1.311
30	2.750	2.457	2.042	1.697	1.310
31	2.744	2.453	2.040	1.696	1.309
32	2.738	2.449	2.037	1.694	1.309
34	2.728	2.441	2.032	1.691	1.307
36	2.719	2.434	2.028	1.688	1.306
38	2.712	2.429	2.024	1.686	1.304
40	2.704	2.423	2.021	1.684	1.303
45	2.690	2.412	2.014	1.679	1.301
50	2.678	2.403	2.009	1.676	1.299
55	2.668	2.396	2.004	1.673	1.297
60	2.660	2.390	2.000	1.671	1.296
65	2.654	2.385	1.997	1.669	1.295
70	2.648	2.381	1.994	1.667	1.294
75	2.643	2.377	1.992	1.665	1.293
80	2.639	2.374	1.990	1.664	1.292
90	2.632	2.368	1.987	1.662	1.291
100	2.626	2.364	1.984	1.660	1.290
200	2.601	2.345	1.972	1.653	1.286
300	2.592	2.339	1.968	1.650	1.284
400	2.588	2.336	1.966	1.649	1.284
500	2.586	2.334	1.965	1.648	1.283
750	2.582	2.331	1.963	1.647	1.283
1000	2.581	2.330	1.962	1.646	1.282
2000	2.578	2.328	1.961	1.646	1.282
Large	2.576	2.326	1.960	1.645	1.282

t-table values

z-table values...

z critical

- stat disk.com
  - ↳ Analysis
    - ↳ Hyp. Test
      - ↳ prop. one sample

Alt hyp: $\boxed{\vee}$

$\boxed{\text{2) Pop. prop. > Claim Prop.}}$

Significance L: $\boxed{0.05}$

claimed Prop: $\boxed{0.50}$ (null)

sample size: $\boxed{30}$

successes: $\boxed{19}$

$\boxed{\text{evaluate}}$

- p-value 0.07206
- critical $\boxed{z_{cr} = 1.64485}$ *

Conclusion: p-value $> \alpha = 0.05$

-OR-

$z_{test} < z_{\text{critical Value}}$

So: we fail to reject the claim

→ * $\alpha = 0.05$ corresponds to $z_c = 1.645$

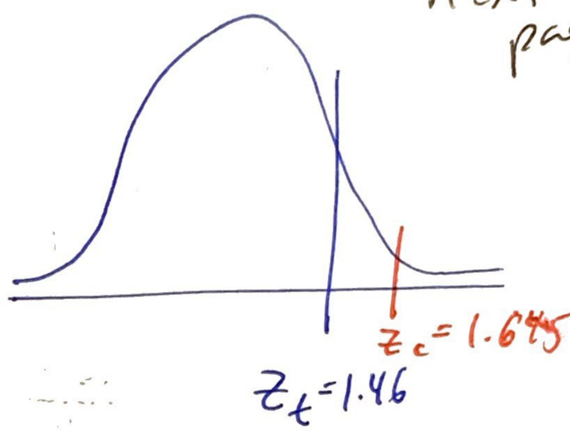
↳ use the last row of the t-table to · the values

Ex Via the (critical value method)

t-table

$\boxed{\alpha = 0.05 \text{ column}}$ $\Rightarrow$ $\boxed{\text{one-tail Header}}$ $\Rightarrow$ table on next page

$$\boxed{z_c = 1.645}$$



but $z_{test} = 1.46$

- since $|z_t| < |z_c|$ we "fail to reject" the null Hypothesis.

<u>Statdisk</u>:

Same as p-value method by click "plot"

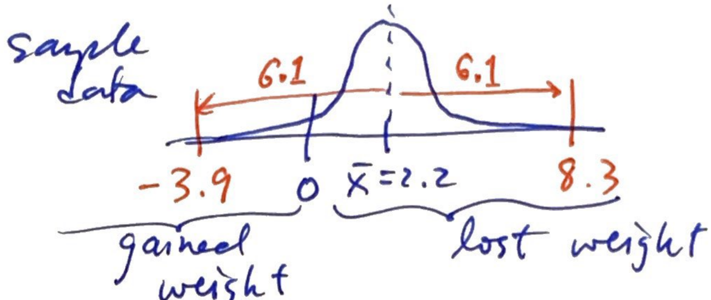
⊛ A 9.3 (means with σ unknown)

example

Ex A medical study shows that after a 12 month diet it was seen that a mean weight loss of 2.2 kg was recorded. Also the std. dev. of the sample was $s = 6.1$ kg. There were 76 patients in the study.

Can we conclude that the mean weight loss was positive? I.E $\mu > 0$. Use a sig. level of $\alpha = 0.05$

Lets get a handle on the $\bar{x} = 2.2$, $s = 6.1$ kg



so some people lost and some gained. Since $\bar{x} = 2.2 > 0$ we want to test the hypothesis that their was more weight loosers than weight gainers

---

Step ∅: Assumptions ⇒ see next pages for Conditions

- SRS assumed
- $n < 10\%$ yes
- unimodal and symmetric? $n \geq 30$ ✓

BTW: means so use t-table

Step 1: $H_0 : \mu = 0$ kg lost {diet did not work}

$H_A : \mu > 0$ kg lost {diet did work!}

↑ right tail test

Step 2: $\alpha = 0.05$, $n = 76$ so use row 75 $t_{crit} = \boxed{1.665}$ ⇒ next pages

Step 3: $t_{test} = \dfrac{\bar{x} - \mu_0}{SE} = \dfrac{2.2 - 0.0}{6.1/\sqrt{76}} = \boxed{3.144}$

Step 4: critical method: we see $|t_{test}| > |t_{crit}|$ so we reject the null hyp of "no weight loss"

Step 5: There is sufficient data to conclude that the diet resulted in weight loss

Assumptions for Inference	And the Conditions That Support or Override Them

Proportions (z)

- **One sample**

1. Individuals are independent.	1. SRS and $n < 10\%$ of the population.
2. Sample is sufficiently large.	2. Successes and failures each ≥ 10.

- **Two Groups**

1. Groups are independent.	1. (Think about how the data were collected.)
2. Data in each group are independent.	2. Both are SRSs and $n < 10\%$ of populations OR random allocation.
3. Both samples are sufficiently large.	3. Successes and failures each ≥ 10 for both groups.

Means (t)

9.3

- **One Sample** (df $= n - 1$)

conditions	Justifications
1. Individuals are independent.	1. SRS and $n < 10\%$ of the population.
2. Population has a Normal model.	2. Histogram is unimodal and symmetric.*

- **Matched pairs** (df $= n - 1$)

1. Data are matched.	1. (Think about the design.)
2. Individuals are independent.	2. SRS and $n < 10\%$ OR random allocation.
3. Population of differences is Normal.	3. Histogram of differences is unimodal and symmetric.* or n>30

- **Two independent samples** (df from technology)

1. Groups are independent.	1. (Think about the design.)
2. Data in each group are independent.	2. SRSs and $n < 10\%$ OR random allocation.
3. Both populations are Normal.	3. Both histograms are unimodal and symmetric.* or both n>30

Distributions/Association (χ^2)

- **Goodness of fit** (df = # of cells $- 1$; one variable, one sample compared with population model)

1. Data are counts.	1. (Are they?)
2. Data in sample are independent.	2. SRS and $n < 10\%$ of the population.
3. Sample is sufficiently large.	3. All expected counts ≥ 5.

- **Homogeneity** [df $= (r - 1)(c - 1)$; many groups compared on one variable]

1. Data are counts.	1. (Are they?)
2. Data in groups are independent.	2. SRSs and $n < 10\%$ OR random allocation.
3. Groups are sufficiently large.	3. All expected counts ≥ 5.

- **Independence** [df $= (r - 1)(c - 1)$; sample from one population classified on two variables]

1. Data are counts.	1. (Are they?)
2. Data are independent.	2. SRSs and $n < 10\%$ of the population.
3. Sample is sufficiently large.	3. All expected counts ≥ 5.

Regression (t, df $= n - 2$)

- **Association of each quantitative variable** ($\beta = 0$?)

1. Form of relationship is linear.	1. Scatterplot looks approximately linear.
2. Errors are independent.	2. No apparent pattern in residuals plot.
3. Variability of errors is constant.	3. Residuals plot has consistent spread.
4. Errors have a Normal model.	4. Histogram of residuals is approximately unimodal and symmetric, or Normal probability plot reasonably straight.*

(*less critical as n increases)

Ex weight loss diet

TABLE A-3 t Distribution: Critical t Values

← one tail test

Degrees of Freedom	Area in One Tail 0.005	0.01	0.025	0.05	0.10
	Area in Two Tails 0.01	0.02	0.05	0.10	0.20
1	63.657	31.821	12.706	6.314	3.078
2	9.925	6.965	4.303	2.920	1.886
3	5.841	4.541	3.182	2.353	1.638
4	4.604	3.747	2.776	2.132	1.533
5	4.032	3.365	2.571	2.015	1.476
6	3.707	3.143	2.447	1.943	1.440
7	3.499	2.998	2.365	1.895	1.415
8	3.355	2.896	2.306	1.860	1.397
9	3.250	2.821	2.262	1.833	1.383
10	3.169	2.764	2.228	1.812	1.372
11	3.106	2.718	2.201	1.796	1.363
12	3.055	2.681	2.179	1.782	1.356
13	3.012	2.650	2.160	1.771	1.350
14	2.977	2.624	2.145	1.761	1.345
15	2.947	2.602	2.131	1.753	1.341
16	2.921	2.583	2.120	1.746	1.337
17	2.898	2.567	2.110	1.740	1.333
18	2.878	2.552	2.101	1.734	1.330
19	2.861	2.539	2.093	1.729	1.328
20	2.845	2.528	2.086	1.725	1.325
21	2.831	2.518	2.080	1.721	1.323
22	2.819	2.508	2.074	1.717	1.321
23	2.807	2.500	2.069	1.714	1.319
24	2.797	2.492	2.064	1.711	1.318
25	2.787	2.485	2.060	1.708	1.316
26	2.779	2.479	2.056	1.706	1.315
27	2.771	2.473	2.052	1.703	1.314
28	2.763	2.467	2.048	1.701	1.313
29	2.756	2.462	2.045	1.699	1.311
30	2.750	2.457	2.042	1.697	1.310
31	2.744	2.453	2.040	1.696	1.309
32	2.738	2.449	2.037	1.694	1.309
34	2.728	2.441	2.032	1.691	1.307
36	2.719	2.434	2.028	1.688	1.306
38	2.712	2.429	2.024	1.686	1.304
40	2.704	2.423	2.021	1.684	1.303
45	2.690	2.412	2.014	1.679	1.301
50	2.678	2.403	2.009	1.676	1.299
55	2.668	2.396	2.004	1.673	1.297
60	2.660	2.390	2.000	1.671	1.296
65	2.654	2.385	1.997	1.669	1.295
70	2.648	2.381	1.994	1.667	1.294
75	2.643	2.377	1.992	1.665	1.293 ← tcrit
80	2.639	2.374	1.990	1.664	1.292
90	2.632	2.368	1.987	1.662	1.291
100	2.626	2.364	1.984	1.660	1.290
200	2.601	2.345	1.972	1.653	1.286
300	2.592	2.339	1.968	1.650	1.284
400	2.588	2.336	1.966	1.649	1.284
500	2.586	2.334	1.965	1.648	1.283
750	2.582	2.331	1.963	1.647	1.283
1000	2.581	2.330	1.962	1.646	1.282
2000	2.578	2.328	1.961	1.646	1.282
Large	2.576	2.326	1.960	1.645	1.282

Ex cont. Statdisk.com

analysis → hypoth Testing → mean one sample

Alt Hyp

3) Pop mean > claimed mean   right tail test

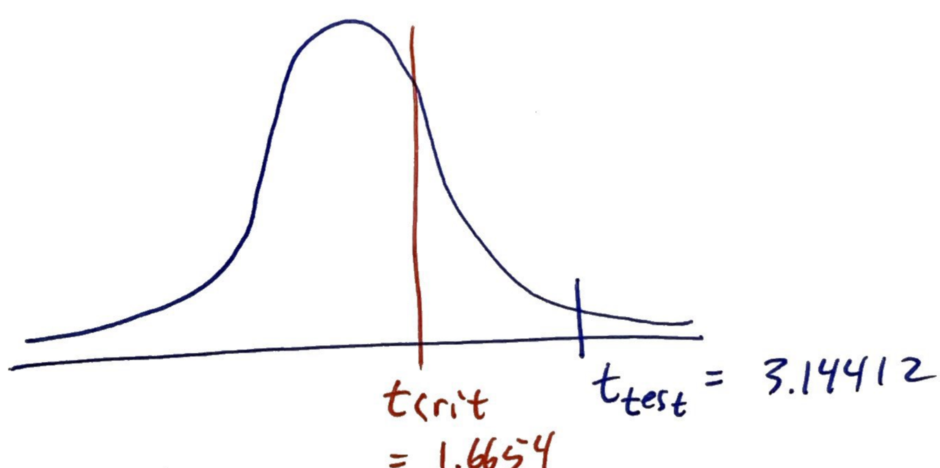
points to the right

significance : α = 0.05
claimed mean : 0
pop. std.dev : leave blank
sample size : n = 76
sample mean : $\bar{x}$ = 2.2
sample s.Dev : s = 6.1

Evaluate

Results :

$t_{crit} = 1.66542$
p-value = 0.00119
→ click on "plot"



Since our test statistic is beyond the critical value we reject the null of "no weight loss"

Our data clearly shows the mean weight loss of the group was positive