```
α = {0.05, 0.025, 0.010, 0.005}
cxm = {{1, 3.84, 5.02, 6.63, 7.88}, {2, 5.99, 7.38, 9.21, 10.60},
   {3, 7.81, 9.35, 11.34, 12.84}, {4, 9.49, 11.14, 13.28, 14.86},
   {5, 11.07, 12.83, 15.09, 16.75}, {6, 12.59, 14.45, 16.81, 18.55},
   {7, 14.07, 16.01, 18.48, 20.28}, {8, 15.51, 17.53, 20.09, 21.95},
   {9, 16.92, 19.02, 21.67, 23.59}, {10, 18.31, 20.48, 23.21, 25.19},
   {11, 19.68, 21.92, 24.72, 26.76}, {12, 21.03, 23.34, 26.22, 28.30},
   {13, 22.36, 24.74, 27.69, 29.82}, {14, 23.68, 26.12, 29.14, 31.32},
   {15, 25.00, 27.49, 30.58, 32.80}, {16, 26.30, 28.85, 32.00, 34.27},
   {17, 27.59, 30.19, 33.41, 35.72}, {18, 28.87, 31.53, 34.81, 37.16},
   {19, 30.14, 32.85, 36.19, 38.58}, {20, 31.41, 34.17, 37.57, 40.00},
   {21, 32.7, 35.5, 38.9, 41.4}, {22, 33.9, 36.8, 40.3, 42.8},
   {23, 35.2, 38.1, 41.6, 44.2}, {24, 36.4, 39.4, 43.0, 45.6},
   {25, 37.7, 40.6, 44.3, 46.9}, {26, 38.9, 41.9, 45.6, 48.3},
   {27, 40.1, 43.2, 47.0, 49.6}, {28, 41.3, 44.5, 48.3, 51.0},
   {29, 42.6, 45.7, 49.6, 52.3}, {30, 43.8, 47.0, 50.9, 53.7},
   {40, 55.8, 59.3, 63.7, 66.8}, {50, 67.5, 71.4, 76.2, 79.5},
   {60, 79.1, 83.3, 88.4, 92.0}, {70, 90.5, 95.0, 100.4, 104.2},
   {80, 101.9, 106.6, 112.3, 116.3}, {90, 113.1, 118.1, 124.1, 128.3},
   {100, 124.3, 129.6, 135.8, 140.2}, {200, 1/2 (Sqrt[199 - 1] + 1.64)^2,
    1/2 (Sqrt[199 - 1] + 1.96)^2, 1/2 (Sqrt[199 - 1] + 2.33)^2, 1/2 (Sqrt[199 - 1] + 2.58)^2}};
(*in case degrees of freedom goes above 199,
the applicable number can be substituted in to replace
 199 above in the last line, with the understanding
 that the values in the last line are approximate.*)

critCXM =
 Interpolation[Flatten[Table[{{cxm[[i, 1]], α[[j]]}, cxm[[i, j + 1]]},
    {j, 4}, {i, Length[cxm]}], 1]]

{0.05, 0.025, 0.01, 0.005}

InterpolatingFunction[ Domain: {{1., 200.}, {0.005, 0.05}}
                       Output: scalar ]
```

3. Test $\mu = 0$ against $\mu > 0$, assuming normality and using the sample 0, 1, -1, 3, -8, 6, 1 (deviations of the azimuth [multiples of 0.01 radian] in some revolution of a satellite). Choose $\alpha = 5\%$ as level of significance.

```
(*Clear["Global`*"]*)
```

Here the μ hypothesis is the null hypothesis. Under this hypothesis the deviations of azimuth in the satellite's orbit are due to random chance, and the assertion that the satellite is not experiencing a progressively dangerous orbital modification.

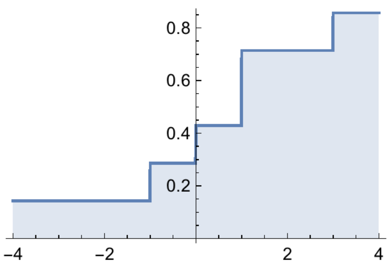
```
sam = {0, 1, -1, 3, -8, 6, 1}
{0, 1, -1, 3, -8, 6, 1}
```

```
emp = EmpiricalDistribution[sam]
DataDistribution[ Type Empirical
                  Datapoints 7 ]
```

```
N[Mean[sam]]
0.285714
```

```
N[StandardDeviation[sam]]
4.30946
```

```
DiscretePlot[CDF[emp, x], {x, -4, 4, .01}, ImageSize → 200]
```



The execution of this problem, from the point of view of the text, involves the *critical value*, what I understand is a somewhat dated evaluation device. I get the impression the p-value is now more commonly used, and one of its advantages is that it is not necessary to consult a table. The *MathWorld* article on Hypothesis Testing gives a basic routine which I use. The LocationTest in Mathematica provides what is necessary in a convenient tabular form, and all I have to do is to compare the level of significance with the p-value.

```
0.05 < 0.866
```

and the upshot of this is that the satellite is deemed not to be doomed, the null hypothesis is accepted.

```
h = LocationTest[sam, Automatic, {"TestDataTable", All}]
```

	Statistic	P-Value
PairedT	0.175412	0.866526
PairedZ	0.175412	0.860756
Sign	4	0.6875
SignedRank	13.	0.671566
T	0.175412	0.866526
Z	0.175412	0.860756

The T value in the above table agrees with the text answer within 2S.

The are seven values in the sample list, which means there are six degrees of freedom. And the α level of significance is 0.05. So

```
c = critCVM[6, 0.05]
```

```
1.94
```

The answer in the green cell above matches the answer in the text for the value of c. I should be prepared to work this out by hand according to the text examples. To that end I consider

$$t = \frac{\bar{x} - \mu_0}{s \Big/ \sqrt{n}} == \frac{0.286 - 0}{4.30946 \Big/ \sqrt{7}}$$

$$\frac{\sqrt{n}\ (\bar{x} - \mu_0)}{s} == 0.175587$$

For some reason the above is not precisely what is shown in the location table for T.

4. In one of his classical experiments, Buffon obtained 2048 heads in tossing a coin 4040 times. Was the coin fair?

5. Do the same test as in problem 4, using a result by K. Pearson, who obtained 6019 heads in 12,000 trials.

```
fin = Table[0, {i, 12 000 - 6019}];
```

```
fin1 = Table[1, {i, 6019}];
```

```
Total[fin1]
6019
```

```
gra = Join[fin, fin1];
```

```
grad = EmpiricalDistribution[gra]
```

```
DataDistribution[ Type Empirical
                  Datapoints 12000 ]
```

```
N[Mean[grad]]
0.501583
```

```
N[Variance[grad]]
0.249997
```

```
N[StandardDeviation[grad]]
0.499997
```

I'm going to illustrate something with confidence level. First, suppose I want to look at the set I made above "gra", containing the 12000 flips. If I want to have confidence level of 99% that the mean will be in an interval, that interval has to grow to include some slightly unusual possibilities that may be just around the corner.

```
MeanCI[gra, ConfidenceLevel → 0.99]
```

```
{0.489824, 0.513343}
```

On the other hand, if I don't need the confidence that the mean of the set will stay inside a certain boundary, upon reducing the required confidence level, the interval shrinks. So to observe the currently calculated mean, the confidence level can go way down, even as far as 0.1.

```
MeanCI[gra, ConfidenceLevel → .1]
```

```
{0.50101, 0.502157}
```

Let me get a number to judge the accuracy of the mean

```
Abs[0.5 - Mean[gra]]
```

```
0.00158333
```

Then let me throw around some Bernoulli pseudos.

```
b1 = RandomVariate[BernoulliDistribution[1/2], 10 000];
```

```
Abs[0.5 - Mean[b1]]
```

```
0.0069
```

```
b2 = RandomVariate[BernoulliDistribution[1/2], 12 000];
```

```
Abs[0.5 - Mean[b2]]
```

```
0.00175
```

```
N[Variance[b2]]
```

```
0.250018
```

Comparing the two yellow cells above, I see that Pearson's trial does at least as well as an instance of Mathematica trying to imitate a fair coin at the same sample size (yellow), which, since the sample size is sizable, seems like pretty good evidence of fairness to me.

Now I will bring in the way that *https://www.math.arizona.edu/~jwatkins/505d/Lesson_11.pdf* handles this coin-flipping situation, adapting from 1000 flips to 12000.

```
Clear[c]
```

I get my c value. 12000 flips is far enough above 200 that I will use the infinity value input, at 0.05 significance level.

```
c = critCVM[100 000, 0.05]
1.65
```

Getting my expected population size of heads, $\mu = 12\,000\ p_0$

$$\mu = 12\,000 \times \frac{1}{2}$$

6000

Getting my variance.

$$\text{sigmasq} == n\ p_0\ (1 - p_0) == 12\,000 \times \frac{1}{2}\ \frac{1}{2}$$

$$\text{sigmasq} == n\ (1 - p_0)\ p_0 == 3000$$

Getting my standard deviation.

$$\text{sigma} = (3000)^{0.5}$$

54.7723

Getting my low expectation.

μ - c sigma

5909.63

Getting my high expectation.

μ + c sigma

6090.37

So if I am interpreting this right, any number of heads between the yellow number and green number above indicates a fair coin, within the stated level of significance. Incidentally, the green number is the one that the text answer gives for c.

6. Assuming normality and known variance $\sigma^2 = 9$, test the hypothesis $\mu = 60.0$ against the alternative $\mu = 57.0$ using a sample of size 20 with mean $\bar{x} = 58.50$ and choosing $\alpha = 5\%$.

7. How does the result in problem 6 change if we use a smaller size, say, of size 5, the other data ($\bar{x} = 58.05$, $\alpha = 5\%$, etc.) remaining as before?

Problem 7 disagrees with problem 6 on the mean, although it is apparent they should agree. The discrepancy makes a significant difference in the statistics, so I show both versions.

First Comparison. (mean 58.50)

As I read the Mathematica docs for LocationTest, the μ items listed in the problem description cannot be done in the same test. A test checks the μ that is provided against a not-μ

condition. If I have this right, then each desired μ must be tested separately. Also, to make a decision I will look at the smaller p-value which is associated with T. In the first comparison, neither hypothesis looks attractive, but I will (would) choose $\mu = 60$, since it has the smaller p-value. I have included the location test that conforms to the current mean of the data, and in the tests I've tried, it has the worst p-value of all.

```
norm = RandomVariate[NormalDistribution[58.50, 3], 5]
{56.0948, 58.9339, 58.5426, 58.9813, 60.1389}
```

```
mn = Mean[norm]
58.5383
```

```
h = LocationTest[norm, 60.0, {"TestDataTable", All}]
```

	Statistic	P-Value
PairedT	−2.19294	0.093385
PairedZ	−2.19294	0.0283116
Sign	1	0.375
SignedRank	1.	0.105645
T	−2.19294	0.093385
Z	−2.19294	0.0283116

```
h = LocationTest[norm, 57.0, {"TestDataTable", All}]
```

	Statistic	P-Value
PairedT	2.30789	0.0822223
PairedZ	2.30789	0.021005
Sign	4	0.375
SignedRank	14.	0.105645
T	2.30789	0.0822223
Z	2.30789	0.021005

Second Comparison. (mean 58.05)

```
norm2 = RandomVariate[NormalDistribution[58.05, 3], 5]
{60.1996, 61.5539, 61.099, 58.5695, 52.5619}
```

```
mn = Mean[norm2]
58.7968
```

```
h = LocationTest[norm2, 60.0, {"TestDataTable", All}]
```

	Statistic	P-Value
PairedT	−0.733655	0.503841
PairedZ	−0.733655	0.463159
Sign	3	1.
SignedRank	7.	1.
T	−0.733655	0.503841
Z	−0.733655	0.463159

```
h = LocationTest[norm2, 57.0, {"TestDataTable", All}]
```

	Statistic	P-Value
PairedT	1.09557	0.334805
PairedZ	1.09557	0.273268
Sign	4	0.375
SignedRank	11.	0.418492
T	1.09557	0.334805
Z	1.09557	0.273268

In the second comparison, the p-value for T in $\mu = 57$ may actually be less than 0.05, (depending on the generated pseudo-values) so that μ would be my choice if it were available.

I will (reluctantly) turn to WolframAlpha for a different view of this. (I don't like cells that cannot be refreshed off-line.) What I found was that both hypotheses, for 60 and for 57 mean, are rejected at the 5% level when 20 items make up the sample, but neither is rejected when only 5 items make up the sample.

```
WolframAlpha["t-test mu0=60, xbar=58.5, s=3, n=5"]
```

Input information

T-test for a population mean	
hypothesized mean	60
sample mean	58.5
sample standard deviation	3
sample size	5

Left-tailed test — Right-tailed test — Two-tailed test

Null hypothesis

$\mu = 60$

Alternative hypothesis

$\mu < 60$

Test statistic

-1.11803

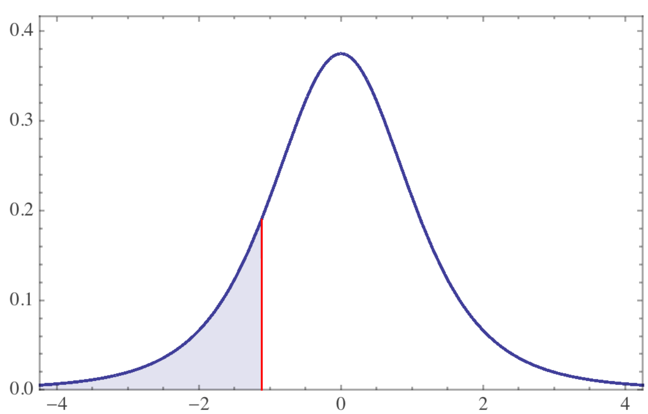
Degrees of freedom

4

p-value

0.1631

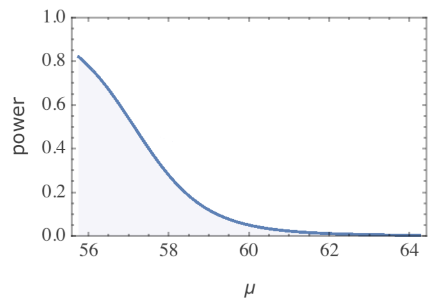
Sampling distribution of test statistic under the null hypothesis



min max

Power function

Significance level 5%



min max

Testconclusions
null hypothesis is not rejected at 1% significance level
null hypothesis is not rejected at 5% significance level
null hypothesis is not rejected at 10% significance level

(assuming a simple random sample; results are exact if the population
is normal and approximate for large samples from non-normal populations)

WolframAlpha

```
WolframAlpha["t-test mu0=57, xbar=58.5, s=3, n=5"]
```

Inputinformation

T-test for a population mean	
hypothesized mean	57
sample mean	58.5
sample standard deviation	3
sample size	5

Righttailedtest

Lefttailedtest Twotailedtest

Nullhypothesis
$\mu = 57$

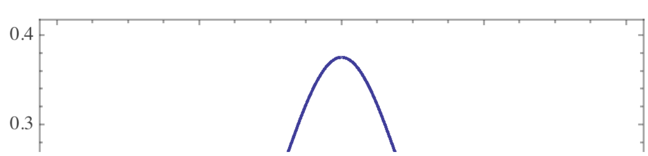
Alternativehypothesis
$\mu > 57$

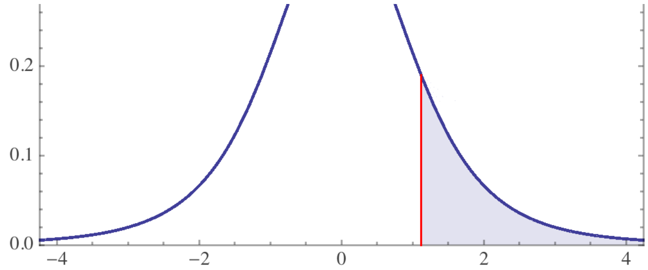
Teststatistic
1.11803

Degreesoffreedom
4

p-value
0.1631

Samplingdistributionofteststatisticunderthenullhypothesis

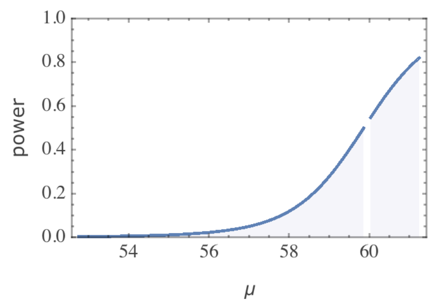




min max

Powerfunction

Significancelevel5%



min max

Testconclusions

null hypothesis **is not rejected** at **1% significance level**
null hypothesis **is not rejected** at **5% significance level**
null hypothesis **is not rejected** at **10% significance level**

(assuming a simple random sample; results are exact if the population is normal and approximate for large samples from non-normal populations)

WolframAlpha

9. What is the rejection region in problem 6 in the case of a two-sided test with $\alpha = 5\%$?

The following grid shows the results of testing on Wolfram|Alpha.

```
Grid[{{57, Reject}, {58, Do not reject},
  {59, Do not reject}, {60, Reject}}, Frame → All]
```

57	Reject
58	Do not reject
59	Do not reject
60	Reject

11. A firm sells oil in cans containing 5000 g oil per can and is interested to know whether the mean weight differs significantly from 5000 g at the 5% level, in which case the filling machine has to be adjusted. Set up a hypothesis and an alternative and per-

form the test, assuming normality and using a sample of 50 fillings with mean 4990 g and standard deviation 20 g.

```
Clear["Global`*"]
```

```
oil = RandomVariate[NormalDistribution[5000, 20], 50];
Mean[oil]
```

```
5004.05
```

```
Variance[oil]
```

```
424.09
```

```
h = LocationTest[oil, 4990, {"TestDataTable", All}]
```

	Statistic	P-Value
PairedT	4.82445	0.0000140861
PairedZ	4.82445	1.40388×10^{-6}
Sign	35	0.00660045
SignedRank	1045.	0.0000853408
T	4.82445	0.0000140861
Z	4.82445	1.40388×10^{-6}

The p-value for T is so small that I think the alternative hypothesis based on 4990 must be accepted. I think this is what the text answer is saying also. This judgment is reinforced by W|A.

```
WolframAlpha["t-test mu0=5000, xbar=4990, s=20, n=50"]
```

Input information

T-test for a population mean	
hypothesized mean	5000
sample mean	4990
sample standard deviation	20
sample size	50

Left-tailed test   Right-tailed test   Two-tailed test

Null hypothesis

$\mu = 5000$

Alternative hypothesis

$\mu < 5000$

Test statistic

$-\frac{5}{\sqrt{2}} \approx -3.53553$

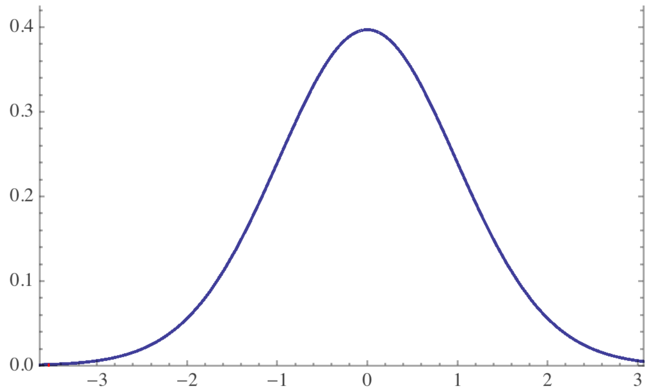
Degrees of freedom

49

p-value

4.5×10^{-4}

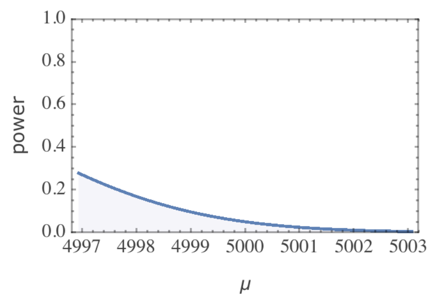
Sampling distribution of test statistic under the null hypothesis



min max

Powerfunction

Significancelevel5%



min max

Testconclusions

null hypothesis **is rejected** at **1% significance level**
null hypothesis **is rejected** at **5% significance level**
null hypothesis **is rejected** at **10% significance level**

(assuming a simple random sample; results are exact if the population is normal and approximate for large samples from non-normal populations)

WolframAlpha

Based on the W|A output above, it looks like the filling machine needs to be adjusted, because the sample data indicate significantly lower fill weight.

13. If simultaneous measurements of electric voltage by two different types of voltmeter yield the differences (in volts) 0.4, -0.6, 0.2, 0.0, 1.0, 1.4, 0.4, 1.6, can we assert at the 5% level that there is no significant difference in the calibration of the two types of instruments? Assume normality.

I do not understand the problem. Would it not make a difference whether I am measuring 110 V or 10000 V? I suppose that the normality of the sample is the guiding factor. If the differences between the voltmeters is close to a normal curve, they are probably equivalent.

```
dtest = {0.4, -0.6, 0.2, 0.0, 1.0, 1.4, 0.4, 1.6}
{0.4, -0.6, 0.2, 0., 1., 1.4, 0.4, 1.6}
```

```
Mean[dtest]
0.55
```

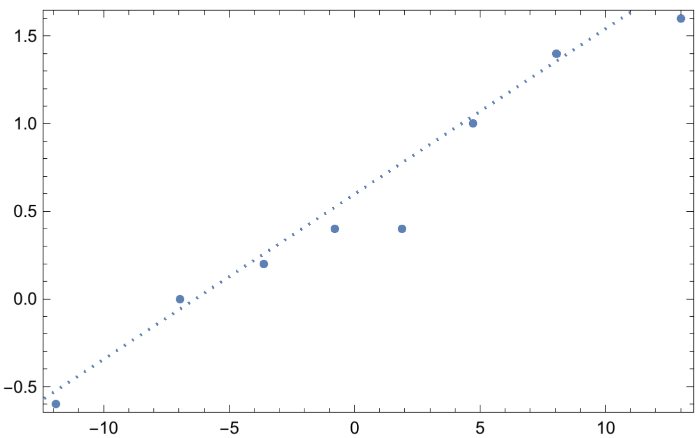
```
StandardDeviation[dtest]
0.738725
```

```
voltmeter = DistributionFitTest[dtest]
```

```
0.687289
```

```
d = NormalDistribution[0.55, 9]
NormalDistribution[0.55, 9]
```

```
QuantilePlot[dtest, d]
```



The score on the probabilty which is the result of the **DistributionFitTest** is pretty high, although I didn't find any numerical guidance on making a choice. I would call the meters equivalent.

15. Suppose that in the past the standard deviation of weights of certain 100.0-oz packages filled by a machine was 0.8 oz. Test the hypothesis $H_0 : \sigma = 0.8$ against the alternative $H_1 : \sigma > 0.8$ (an undesirable increase), using a sample of 20 packages with standard deviation 1.0 oz and assuming normality. Choose $\alpha = 5\%$.

```
Clear["Global`*"]
```

When fiddling with mean values, the Normal distribution is the common one. Sometimes, as in this problem, the focus is on standard deviation, and I get the impression that that is where the ChiSquare distribution is commonly used. I don't know if it's general, but in this case it is not necessary to construct a RandomVariate set.

```
c = critCXM[19, 0.05] (*degrees of freedom, level of significance*)
```

30.14

$$\text{DegreesOfFreedom}\left(\frac{\text{H1sigma}}{\text{H0sigma}}\right)^2$$

$$19\left(\frac{1.0}{0.8}\right)^2$$

29.6875

Since the calculated formula is less than the c value, the proposed hypothesis is accepted. That is, the standard deviation is considered to have risen above 0.8.

17. Brand A gasoline was used in 16 similar automobiles under identical conditions. The corresponding sample of 16 values (miles per gallon) had mean 19.6 and standard deviation 0.4. Under the same conditions, high-power brand B gasoline gave a sample of 16 values with mean 20.2 and standard deviation 0.6. Is the mileage of B significantly better than that of A? Test at the 5% level; assume normality. First guess. Then calculate.

According to example 5 on p. 1084, although the number of degrees of freedom is the sum of the two samples, it is necessary to subtract one from each sample, leaving, in this case, 30 total.

```
c = critCVM[30, 0.05]
```

1.7

Numbered line (12) on p. 1085 gives the formula for t_0 in a combined instance such as the present

$$t_0 = \sqrt{n}\ \frac{\bar{x}-\bar{y}}{\sqrt{s_x{}^2+s_y{}^2}}$$

I see that n is not the sum of the two samples, but rather the size of one complete sample. Inasmuch as the procedure is supposed to allow for samples of different sizes, there must be a rule for choosing n when the samples are not the same size, but I do not know it. The calculated set-up for t_0, shown below, appears in the text answer.

$$t_0 = \sqrt{16}\ \frac{20.2 - 19.6}{\sqrt{(0.4)^2 + (0.6)^2}}$$

3.3282

I notice that the formula for t_0 above contains the sense that $\bar{x}$ - $\bar{y}$ in the numerator makes x the H_0 hypothesis and y the H_a hypothesis. In this case H_0 is gasoline B. Since t_0 is larger than c at the 0.05 level, it means that when the alternative is ignored, B will be triumphant.