2) Test for the Difference Between Two Population Means: Given Two Large Samples (n_1 and $n_2 > 30$) and the Population Variances σ_1^2 and σ_2^2 are Unknown:

Suppose $\bar{x}_1$ and $\bar{x}_2$ are the means of two independent random samples of size n_1 and n_2 from populations with unknown means and unknown variances σ_1^2 and σ_2^2 respectively.

a) If $\sigma_1^2 = \sigma_2^2 = \sigma^2$

First, find s^2 as an estimate of σ^2.

$$s^2 = \frac{(n_1-1)\cdot s_1^2 + (n_2-1)\cdot s_2^2}{n_1+n_2-2} \equiv \frac{\sum(x-\bar{x}_1)^2 + \sum(x-\bar{x}_2)^2}{n_1+n_2-2}$$

$$z_{calc} = \frac{(\bar{x}_1 - \bar{x}_2)}{s\cdot\sqrt{\frac{1}{n_1}+\frac{1}{n_2}}}$$

Example: Two samples of electric light bulbs are treated for length of life. The results are given in the table below:

	Sample 1	Sample 2
# in sample	42	92
$\bar{x}$	2060	2040
$\sum(x - \bar{x})^2$	20100	18300

Are the samples from populations with different means ? Assume population variances are unknown and equal. Use α = 0.05 significance level.

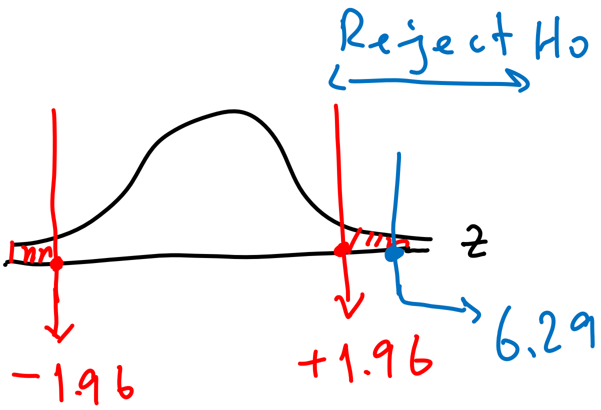
Solution: X $H_0: \mu_1 = \mu_2$

✓ $H_A: \mu_1 \neq \mu_2$

$$s^2 = \frac{20100 + 18300}{42+92-2} \cong 291$$

$$z_{\alpha/2} = z_{0.025} = \mp 1.96 = z_{table}$$

$$z_{calc} = \frac{2060 - 2040}{\sqrt{291 \cdot \left[\frac{1}{42} + \frac{1}{92}\right]}} = 6.29$$



Decision: Reject H_0

Conclusion: Samples are from populations with different means.

b) If $\sigma_1^2 \neq \sigma_2^2$

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

Example: Screws are made by two different machines A and B. Samples of size 120 and 150 are taken from the output of machines A and B respectively. Table shows the results of measuring the lengths of screws obtained.

	Machine A	Machine B
$\sum x$	492	690
$\sum(x - \bar{x})^2$	12	70

Test if the machines are set to give different lengths of screws. Assume population variances are unknown and unequal. Use α = 0.05 significance level.

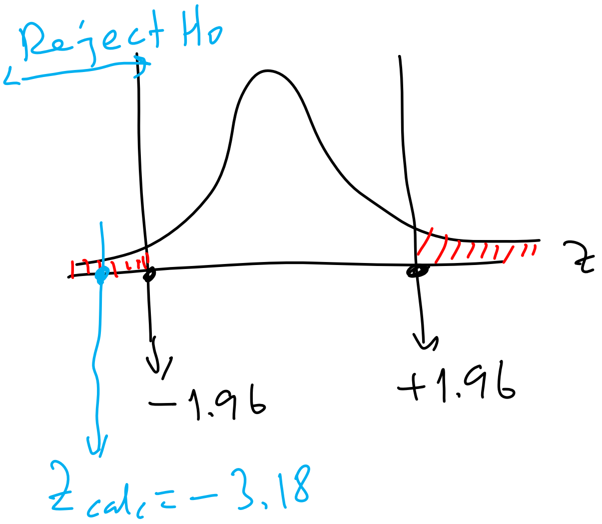
Solution:

$$\bar{X}_A = \frac{492}{120} = 4.1 , \quad \bar{X}_B = \frac{690}{150} = 4.6$$

$$S^2 = \frac{\sum(x-\bar{x})^2}{n-1} \Rightarrow S_A^2 = \frac{12}{120-1} = \frac{12}{119} , \quad S_B^2 = \frac{70}{149}$$

$$\left.\begin{array}{l} \times\ H_0: \mu_1 = \mu_2 \\ \checkmark\ H_A: \underset{A}{\mu_1} \neq \underset{B}{\mu_2} \end{array}\right\} \Rightarrow z_{calc} = \frac{4.1 - 4.6}{\sqrt{\frac{(12/119)}{120} + \frac{(70/149)}{150}}} = -3.18$$

$$z_{table} = z_{0,025} = \mp 1.96$$



Decision: Reject Ho

Conclusion: The machines are set to give different length of screws at 0,05 level.

3) Test for the Difference Between Two Population Means: Given Two Small Samples (n_1 and $n_2 < 30$) and the Population Variances σ_1^2 and σ_2^2 are Unknown:

a) if $\sigma_1^2 = \sigma_2^2 = \sigma^2$ and unknown.

Calculate S^2 as an estimate of σ^2.

$$S^2 = \frac{(n_1 - 1) \times S_1^2 + (n_2 - 1) \times S_2^2}{n_1 + n_2 - 2}$$

⊗ t- distribution is used since σ is unknown and $n < 30$.

$$t_{calc} = \frac{\bar{X}_1 - \bar{X}_2}{s_* \sqrt{\frac{1}{n_1} + \frac{1}{n_2}}}$$

# of degrees of freedom; $d.f. = \nu = n_1 + n_2 - 2$

Confidence Interval (CI):

$$(\bar{x}_1 - \bar{x}_2) - t_{\alpha/2}^{*} \left(s_* \sqrt{\frac{1}{n_1} + \frac{1}{n_2}}\right) < (\mu_1 - \mu_2) < (\bar{x}_1 - \bar{x}_2) + \cdots$$

Example: **Two samples of size 6 and 8 have respective means of $\bar{x}_1 = 4.25$, $\bar{x}_2 = 4.35$ and variances $s_1^2 = 0.0025$, $s_2^2 = 0.0016$. Test if the samples could come from the same population. Use $\alpha = 0.01$ significance level. Assume population variances are unknown and equal.**

Solution:

$\times\ H_0: \mu_1 = \mu_2$

$\checkmark\ H_A: \mu_1 \neq \mu_2$

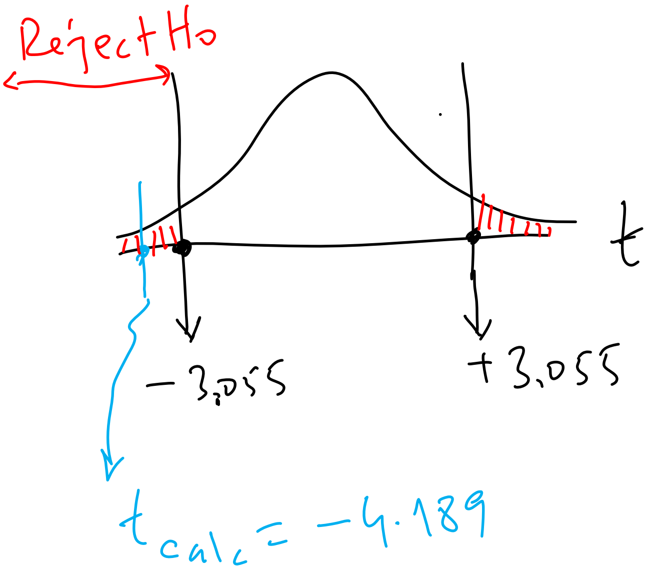
$n < 30$ and pop. variances are unknown $\Rightarrow$ t- distr.

$$s^2 = \frac{5 \times 0{,}0025 + 7 \times 0{,}0016}{6 + 8 - 2} = \frac{0{,}0237}{12}$$

$$t_{calc} = \frac{4.25 - 4.35}{\sqrt{\frac{0{,}0237}{12} \times \left(\frac{1}{6} + \frac{1}{8}\right)}} = -4.189$$

$t_{table} = ?$ $\quad \nu = 6 + 8 - 2 = 12 \Rightarrow$

$$t_{0.01/2}(12) = t_{0.005}(12) = 3.055$$



Decision: Reject Ho

Conclusion: Samples come from different populations.

Homework: An experiment was conducted to compare the mean time (in days) required to recover from a common cold for persons given a daily dose of 4 mg of Vitamin C versus those who were not given a vitamin supplement. Suppose that 35 adults were randomly selected for each treatment category and that the men recovery times and Standard deviations fort he two groups were as follows:

	No vitamin (2)	Vitamin (1)
Sample size	35	35
Sample mean	6.9	5.8
Sample Standard deviation	2.9	1.2

a) Suppose your research objective is to show that the use of vitamin C (group 1) reduces the mean time required to recover. State the null and alternative hypotheses. Is this a one-tailed test or two-tailed test ?

b) Conduct the statistical test of H_0 using $\alpha = 0.05$ significance level. What is your conclusion ? Assume population variances are unknown and unequal.

4) Test of Hypothesis for Paired Observations

$n_1 = n_2$ always.

Apply this test when,

1) pairs of samples from two populations are treated in the same way.

a) two types of metal specimens buried in the ground together in a variety of soil types to determine corrosion resistance

b) wear-rate test with two different types of tractor tires mounted in pairs of "n" tractors for a defined period of time.

2) Two types of measurements are made on the same unit.

a) blood pressure measurements made on the same individual before and after the administration of a stimulus.

b) smoothness determinations on the same film samples at two different testing laboratories.

A) For large paired samples (use z-statistic)

Suppose we have two independent samples of equal sizes "n" ($n_1 = n_2$) from populations with means μ_1 and μ_2 and variances σ_1^2 and σ_2^2, then,

$$z_{calc} = \frac{\bar{d}}{\sqrt{\frac{\sigma_1^2 + \sigma_2^2}{n}}}$$

B) For small paired samples (use t-statistic)

$$t_{calc} = \frac{\bar{d} - \cancelto{0}{\mu_D}}{s_d/\sqrt{n}}, \quad (\mu_D = \mu_1 - \mu_2 = 0)$$

$$t_{calc} = \frac{\bar{d}}{s_d / \sqrt{n}} \quad , \quad d.f. = \nu = n - 1$$

$\bar{d} = \frac{\sum d_i}{n}$: mean of differences

d_i: sample difference between the i^{th} pair of observations.

S_d : sample st. dev. of differences

μ_D : pop. mean of differences.

<u>Hypothesis:</u>

$H_0: \mu_D = 0$	$H_0: \mu_D = 0$	$H_0: \mu_D = 0$
$H_A: \mu_D \neq 0$	$H_A: \mu_D > 0$	$H_A: \mu_D < 0$
Two-tailed	RHS (one-tailed)	LHS (one-tailed)

<u>Example</u>: Pairs of pipes have been buried in 11 different locations to determine corrosion resistance for underground use. One type includes a lead-coated steel pipe (A) and the other a bare steel pipe (B). Test whether either type of pipe has a greater corrosion resistance than the other. Use α = 0.05 significance level.

Soil type	Corrosion values		d_i = difference
	Pipe A	Pipe B	
A	27.3	41.4	27.3 − 41.4 = −14.1
B	18.4	18.9	−0.5
C	11.9	21.7	−9.8
D	11.3	16.8	−5.5
E	14.8	9.0	5.8
F	20.8	19.3	1.5
G	17.9	32.1	−14.2
H	7.8	7.4	0.4
I	14.7	20.7	−6.0
J	19.0	34.4	−15.4
K	65.3	76.2	−10.9

$\sum d_i = -68.7$

Solution:

$$(\sum d_i)^2 = (-68.7)^2 = 4719.69$$

$\sum d_i^2 = 955.01 \leftarrow$ [i.e., $(-14.1)^2 + (-0.5)^2 + \dots$]

× $H_0: \mu_D = 0$ (i.e, no difference in the two types)

✓ $H_A: \mu_D \neq 0$ (i.e., $\mu_{\text{pipe A}} \neq \mu_{\text{pipe B}}$)

$n_1 = n_2 = 11$, $\nu = n - 1 = 10$

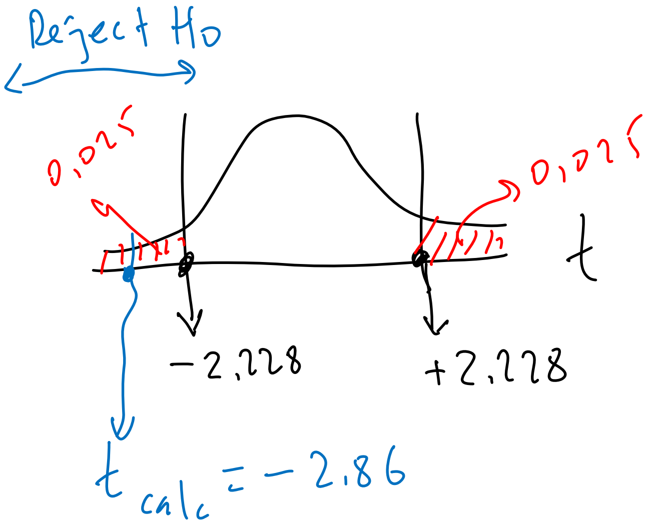
$t_{\text{table}} = ?$ $t_{0.025}(10) = 2.228$

$$\bar{d} = \frac{-68.7}{11} = -6.245$$

$$S_d^2 = \frac{n \cdot \sum d_i^2 - (\sum d_i)^2}{n \cdot (n-1)} = \frac{11 \times (955.01) - (4719.69)}{11 \times 10}$$

$S_d^2 = 52.56 \Rightarrow S_d = 7.25$

$$t_{calc} = \frac{-6.245}{7.25/\sqrt{11}} = -2.86$$



Decision: Reject Ho

Conclusion: One pipe has a greater corrosion resistance than the other.

Example: A stimulus was tested for its effect on blood pressure. 10 men were selected randomly, and their blood pressure was measured before and after the stimulus was administered. It was of interest to determine whether the stimulus had caused a significant increase in the blood pressure. Use α = 0.05 significance level.

Individual	Blood pressure		d_i = difference
	Before	After	
1	138	146	146 − 138 = 8
2	116	118	2
3	124	120	−4
4	128	136	8
5	155	174	19
6	129	133	4
7	130	129	−1
8	148	155	7
9	143	148	5
10	159	155	−4

$\sum d_i = 44$

Solution:

✗ $H_0: \mu_D = 0$ (i.e., $\mu_{after} = \mu_{before}$)

✓ $H_A: \mu_D > 0$ (i.e., $\mu_{after} > \mu_{before}$)

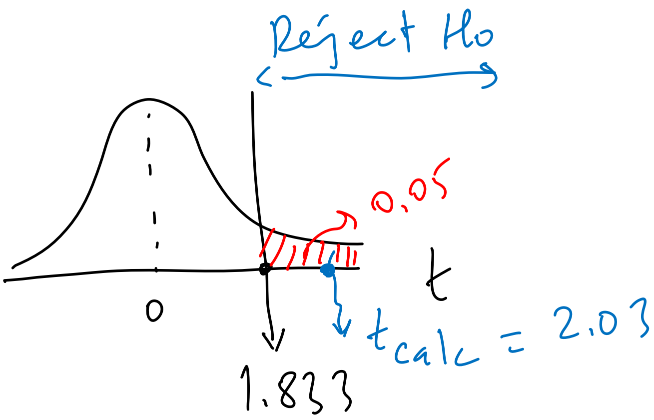
$n = 10 \Rightarrow d.f. = \nu = 10 - 1 = 9$

$t_\alpha(\nu) = t_{0.05}(9) = 1.833 = t_{table}$

Calculate $\bar{d}$ and $S_d \Rightarrow$

$\bar{d} = 4.4$, $S_d = 6.85$

$$t_{calc} = \frac{4.4}{6.85/\sqrt{10}} = 2.03$$



Decision: Reject Ho

Conclusion: The blood pressure has increased after the stimulus had been administered.

Confidence Interval (CI)

$$\bar{d} - t_{\alpha/2} \cdot \frac{S_d}{\sqrt{n}} < \mu_D < \bar{d} + t_{\alpha/2} \cdot \frac{S_d}{\sqrt{n}}$$

Example: Find 95 % confidence interval for the difference in mean wear for tire types A and B. Use α = 0.05 significance level.

Results of a tire wear test			
	Tire wear		
Automobile	X_A	X_B	$d_i = (X_A - X_B)$
1	10.6	10.2	0,4
2	9.8	9.4	0.4
3	12.3	11.8	0,5
4	9.7	9.1	0,6
5	8.8	8.3	0.5

$$\sum d_i = 2.4$$

Solution:

$$n = 5, \quad \nu = 5 - 1 = 4$$

$$S_d = \sqrt{\frac{\sum d_i^2 - (\sum d_i)^2/n}{n-1}}$$

$$\sum d_i^2 = (0,4)^2 + (0.4)^2 + \cdots + (0,5)^2 = 1.18$$

$$\bar{d_i} = \frac{2.4}{5} = 0,48$$

$$S_d = \sqrt{\frac{1.18 - (2.4)^2/5}{5-1}} \Rightarrow S_d = 0,0837$$

$$t_{\alpha/2}(\nu) = t_{0,025}(4) = 2.776$$

$$CI: 0,48 - (2.776) \times \frac{0,0837}{\sqrt{5}} < \mu_D < 0,48 + (2.776) \times \frac{0,0837}{\sqrt{5}}$$

$$0,376 < \mu_D < 0,58$$

χ^2 (Chi-Square) Distribution

- **Comparing two variances ⟹ F test is used**
- **Testing a single population variance (one-sample test) ⟹ χ^2 test is used.**

Applications of χ^2 distribution:

a) Testing a single population variance by χ^2 test

The sample (process) variance is compared with the population (specification that is set) variance, i.e., we want measurements involving products or processes fall inside specifications set by consumers/people etc.

The specifications are often met if the process <u>variance</u> is sufficiently <u>small</u>.

b) The χ^2 test of goodness of fit.

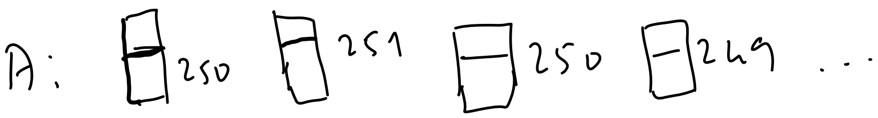
a) Test a Single Population Variance by χ^2 –Distribution

When random samples are drawn from a normal population of a known variance, the quantity $[(n-1).s^2/\sigma^2]$ possess a probability distribution that is known as the χ^2 distribution.

e.g., consider a fruit juice filling machine,

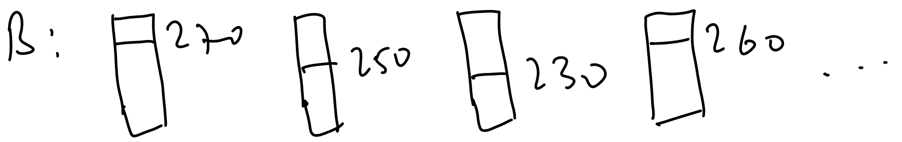
we want to fill the bottles on the average 250 mL capacity.

- if the variance is too small, then, no problem with the filling machine.



- if the variance of the sample is too large, then, many bottles are over filled and many are under filled.

Thus, the company will want to maintain as small a variance as possible.



(problem with the machine!!!)

e.g., consider a variance of 0.0004 to be acceptable:

when the sample variance becomes larger, then, the filling machine should be adjusted.

So, the decision will be made by use of hypothesis test procedure.

H_0: $\sigma^2 = 0.0004 \Rightarrow$ machine is not out of control.

H_A: $\sigma^2 > 0.0004 \Rightarrow$ " " out of control

Test statistic value; χ^2_{calc}

$$\chi^2_{calc} = \frac{(n-1) \cdot S^2}{\sigma^2}$$

$$d.f. = \nu = n-1$$

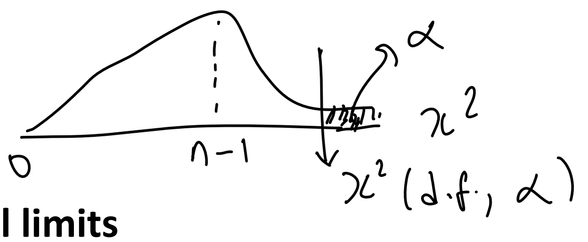
Properties of χ^2 –Distribution

1. χ^2 is nonnegative in value. Assumes zero or (+) values.

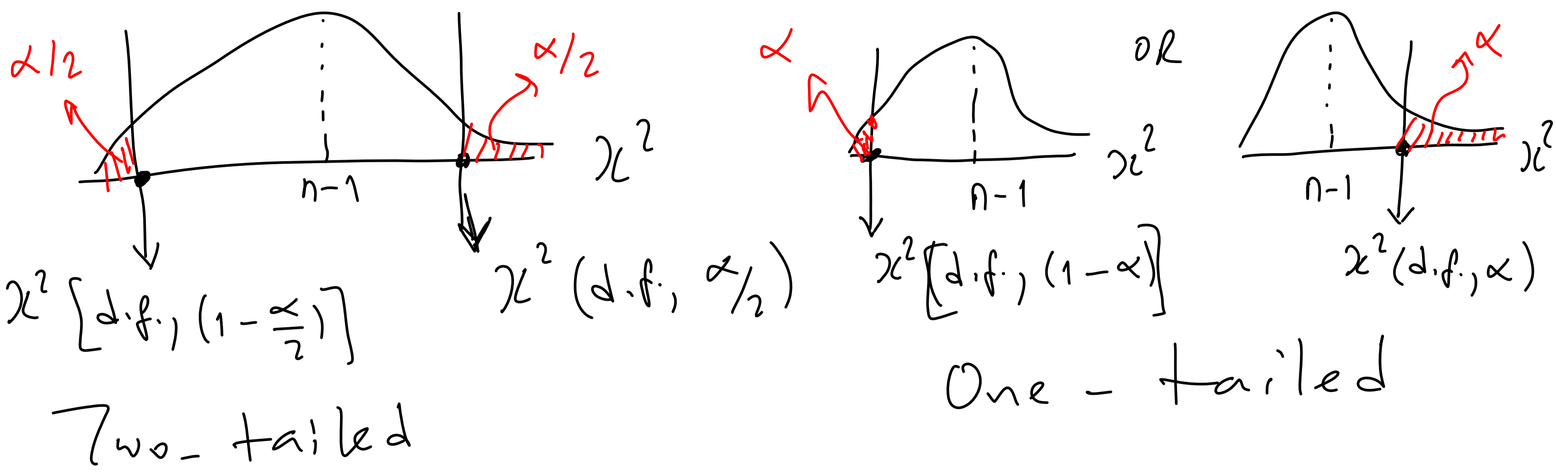
2. The mean value of χ^2 –Distribution is (n-1).

3. χ^2 is not symmetrical.

4. χ^2_α denotes the χ^2 –value having area α to its right under χ^2 –curve.



5. Critical limits

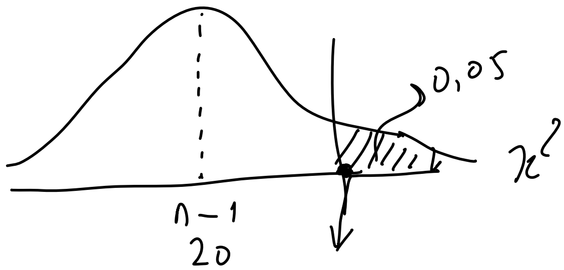


Example: Find χ^2 (20, 0.05) value.

Solution: $\underbrace{20}_{\nu}$, $\underbrace{0.05}_{\alpha}$

χ^2 - table $\Rightarrow$

d.f. \ α	0,995	...	0,05	...	0,005
⋮					
20			31.4		
⋮					

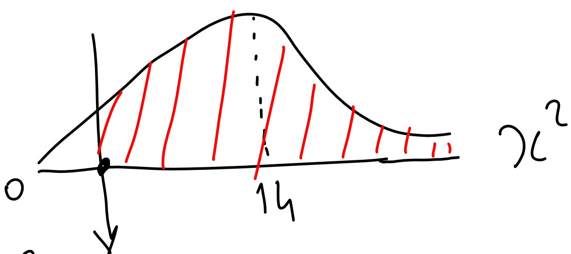


31.4

$\chi^2(20,\ 0.05)$

Example: Find χ^2 (14, 0.90) value.

Solution:



$\chi^2(14, 0.90) = 7.79$

d.f.	...	0,90	...
⋮			
14		7.79	
⋮			