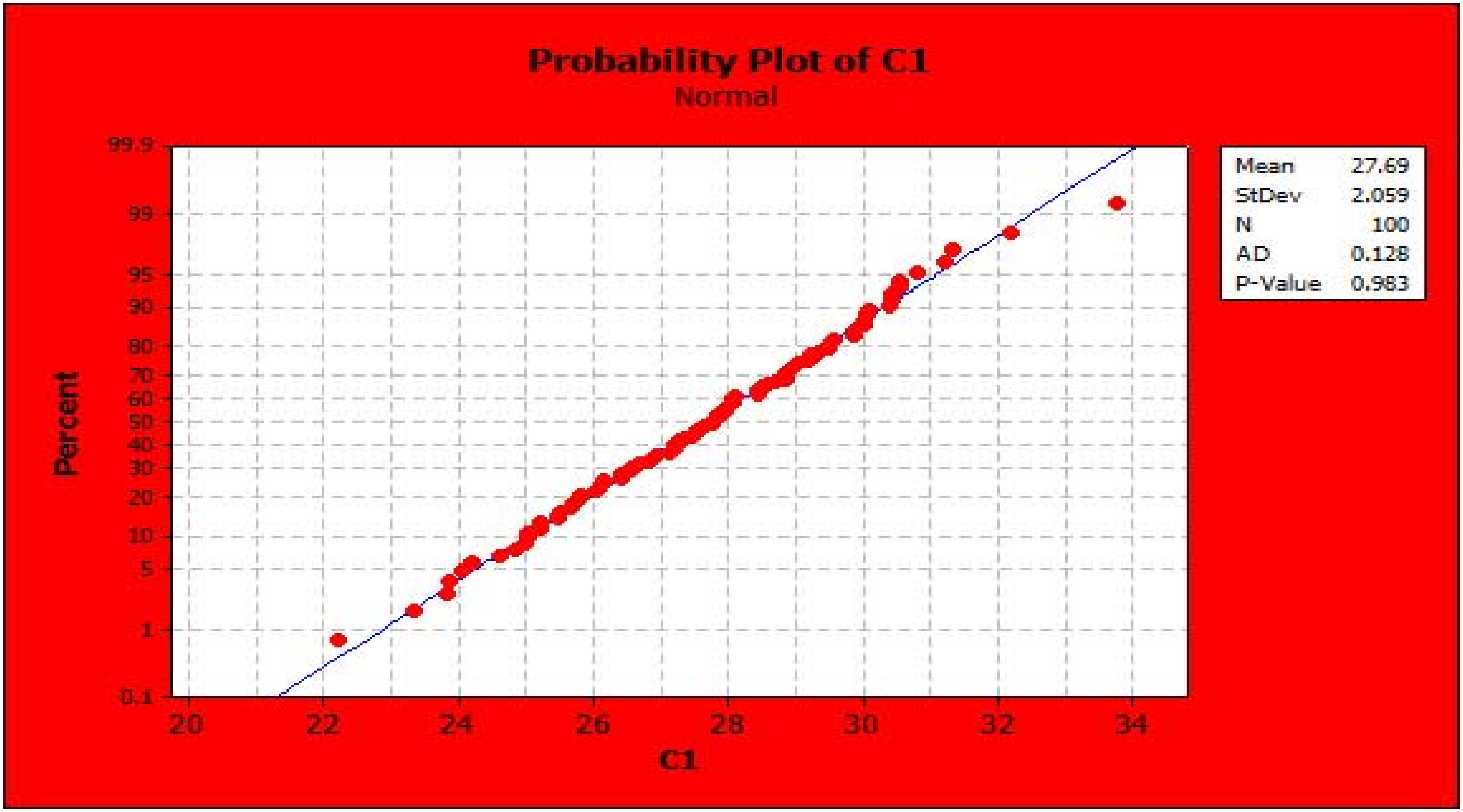
Tests for Normality (circa 1980) in order of increasing ? complexity

- Graphical – Histograms, Stem and Leaf Plots
- Descriptive versus Theoretical Properties (68% of values within + or – 1 sigma, etc.)
- Normal Probability Plots
- Statistical Measures of Skewness and Kurtosis
- Chi-square Goodness of Fit test
- Statistical Tests like Anderson-Darling, Ryan-Joiner, Shapiro-Wilk, and Kolmogorov-Smirnov
- “preponderance of evidence”

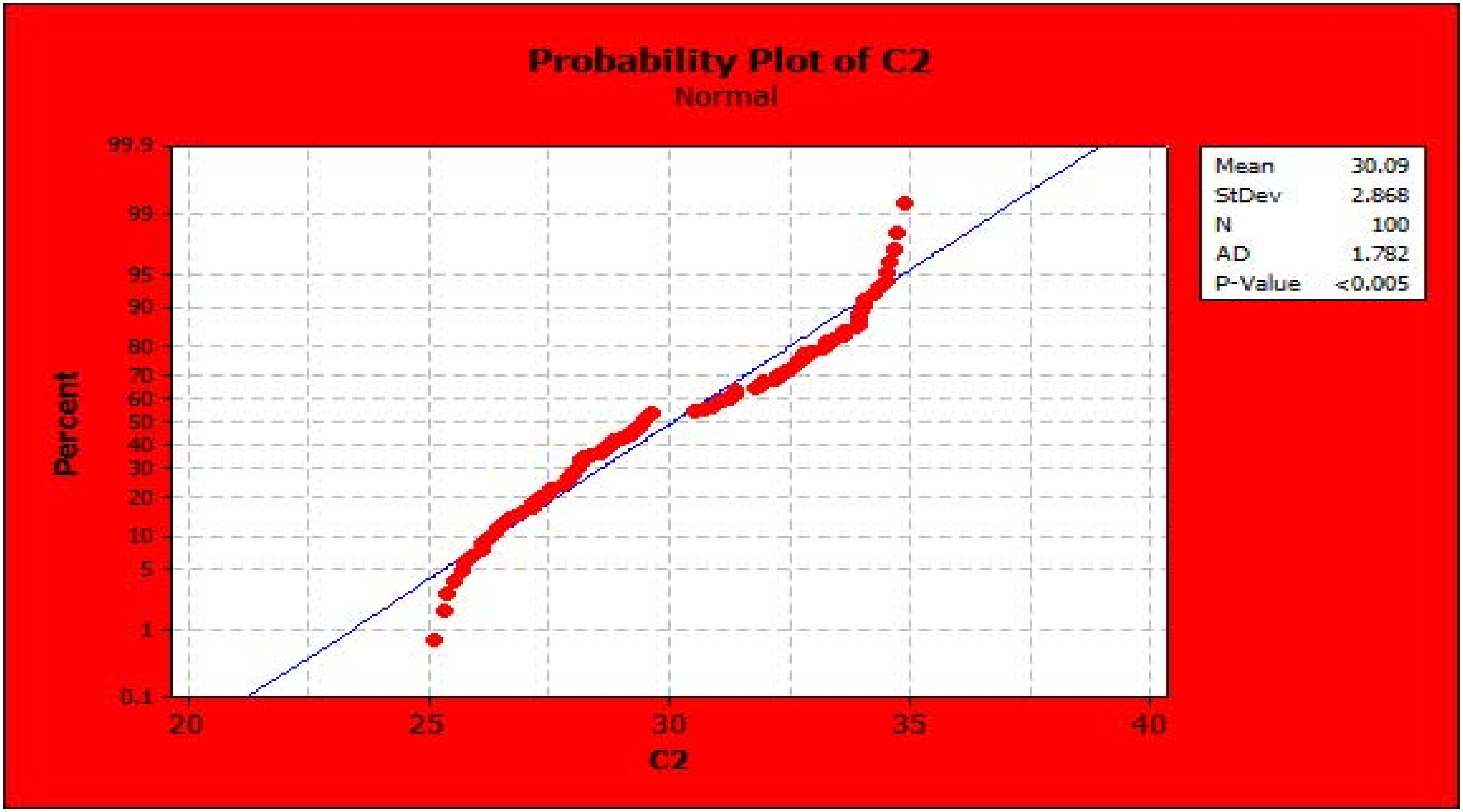
© 2014 ConteSolutions

Output from Minitab
Stat/Basic Statistics/Normality Test…



© 2014 ConteSolutions

Output from Minitab
Stat/Basic Statistics/Normality Test…



© 2014 ConteSolutions

Chi-square Goodness of Fit test

- GOF on Uniform distribution (Dog food preference)
- Sum the Difference between observed and expected values squared divided by the expected value on page 170
- Examples are only around an a priori and uniform distributions

© 2014 ConteSolutions

Chi-Square (χ^2)

- Chi-Square (χ^2): Used to test assumptions about single population variances, distributions, and independence of variables; create confidence intervals for population variances
- Commonly used is a test call a "goodness of fit"
- GOF for uniform, Poisson, binomial, normal, a-priori
- GOF for too much variation, too little variation

$$\chi^2_{calc} = \sum \frac{(\text{observed value} - \text{expected value})^2}{\text{expected value}}$$

© 2014 ConteSolutions

Example GOF Problem Normal Distribution

- Taken from Barron's *EZ 101 Study Keys Statistics, by Martin Sternstein, 2nd Edition, (Amazon $8.99)*
- *Page 193*

- Suppose that the assembly times for a sample of 300 units of an electronic product have mean $\mu = 84$, standard deviation $\sigma = 3$, and the following distribution:

Assembly Time (minutes)

	< 78	78–81	81–84	84–87	87–90	>90
Observed number of units	15	39	87	96	48	15

At the 1% significance level, test the null hypothesis that the distribution is normal.

© 2014 ConteSolutions

Classical Test of Hypothesis

- State the “null Hypothesis” in terms of a population parameter and an equal (=) sign.
- State the Alternative Hypothesis in terms of the same population parameter and one of three inequality signs.
- State the level of significance to which you wish to test the hypothesis.
- Identify the test statistic as either Z calculated, T calculated, Chi-squared calculated or F-calculated.
- Define the rejection criteria in terms of the calculated value and the criteria value.
- Evaluate the Test statistic and compare it to the criteria.
- Write out the conclusion.

© 2014 ConteSolutions

Classical Test of Hypothesis

- State the "null Hypothesis" in terms of a population parameter and an equal (=) sign. The data follows a normal distribution
- State the Alternative Hypothesis in terms of the same population parameter and one of three inequality signs. The data does not follow a normal distribution
- State the level of significance to which you wish to test the hypothesis. $\alpha = 0.01$
- Identify the test statistic as either Z calculated, T calculated, Chi-squared calculated or F-calculated.
- Define the rejection criteria in terms of the calculated value and the criteria value. Chi-squared calculated > Chi-square criteria
- Evaluate the Test statistic and compare it to the criteria.
- Write out the conclusion.

© 2014 ConteSolutions

Example GOF Problem Normal Distribution

- Taken from Barron's *EZ 101 Study Keys Statistics, by Martin Sternstein, 2nd Edition, (Amazon $8.99)*
- *Page 189*
- *What are the expected values for each of the cells?*

- Suppose that the assembly times for a sample of 300 units of an electronic product have mean $\mu = 84$, standard deviation $\sigma = 3$, and the following distribution:

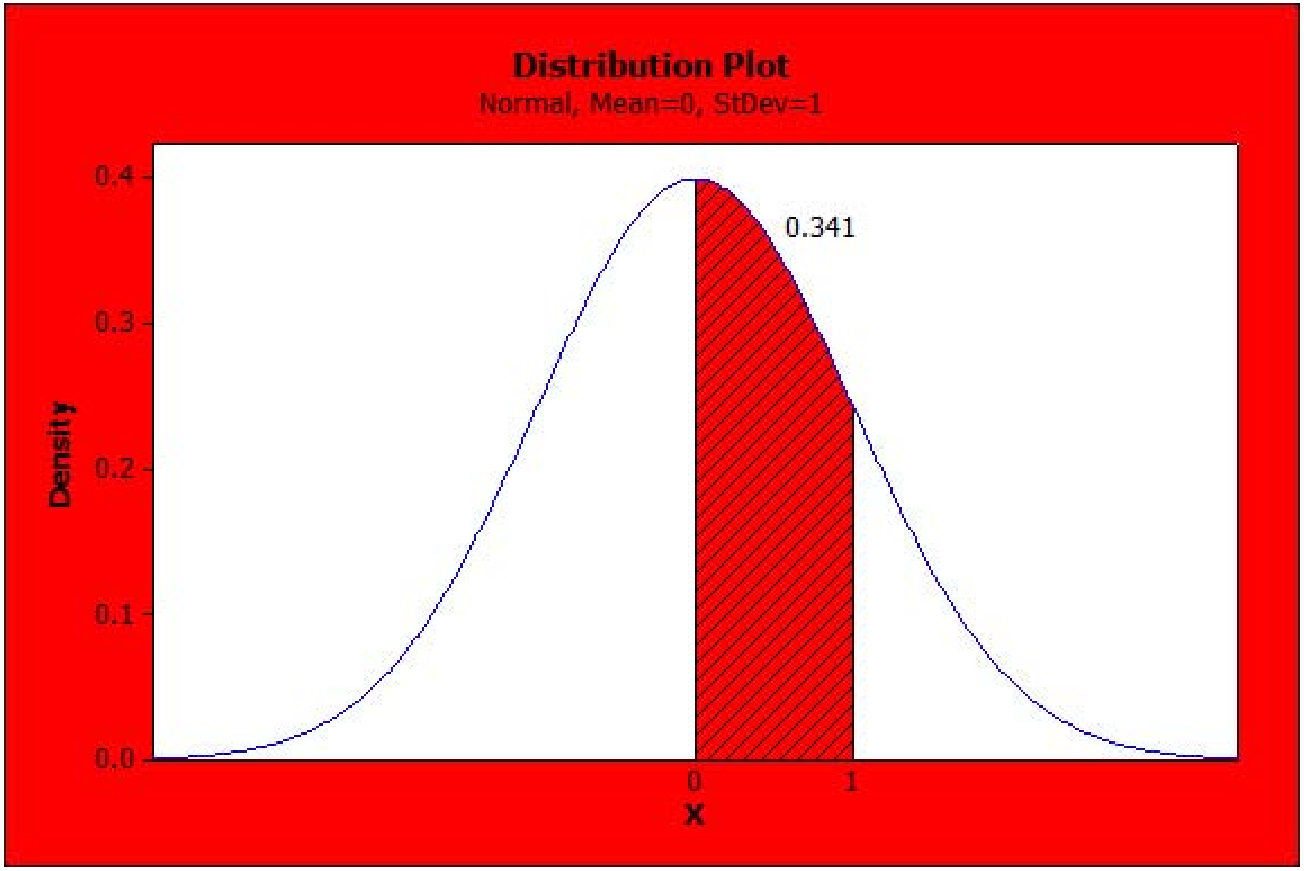
Assembly Time (minutes)

	< 78	78–81	81–84	84–87	87–90	>90
Observed number of units	15	39	87	96	48	15

At the 1% significance level, test the null hypothesis that the distribution is normal.

© 2014 ConteSolutions

Minitab prob plot



© 2014 ConteSolutions

Expected value for 0.0 to +1.0 sigma cell

- Probability space times universe
- 0.3413 times 300 = 102.39

© 2014 ConteSolutions

Chi-squared calculated

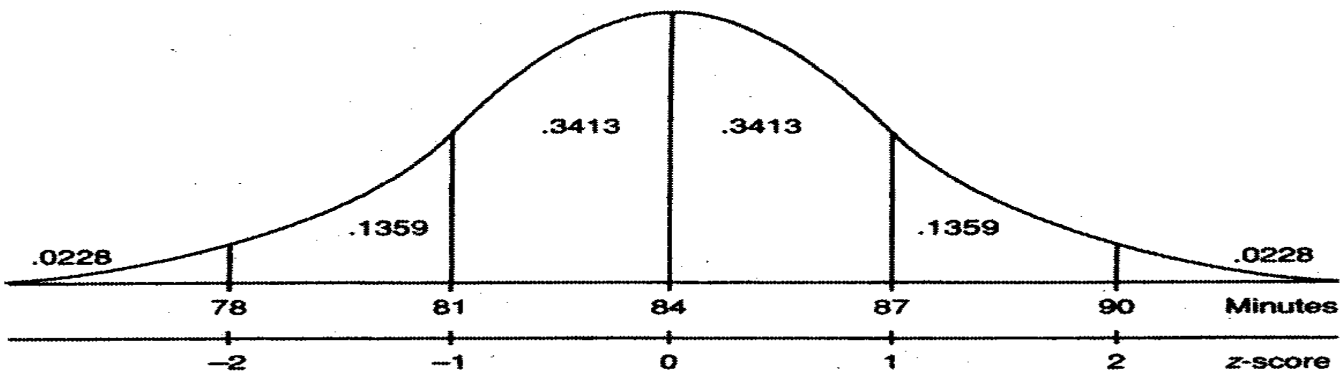
$$\chi^2_{calc} = \sum \frac{(\text{observed value} - \text{expected value})^2}{\text{expected value}}$$

© 2014 ConteSolutions

Example GOF Problem Normal Distribution

- Taken from Barron's *EZ 101 Study Keys Statistics, by Martin Sternstein, 2nd Edition, (Amazon $8.99)*
- *Page 193*

Answer: Here we must use the sample mean, 84, and the sample standard deviation, 3. The normal probability table gives:



Multiplying by 300 units yields the following expected numbers of units.

Assembly Time (minutes)

	< 78	78–81	81–84	84–87	87–90	>90
Expected number of units	6.84	40.77	102.39	102.39	40.7	6.84

© 2014 ConteSolutions

Example GOF Problem Normal Distribution

- Taken from Barron's *EZ 101 Study Keys Statistics, by Martin Sternstein, 2nd Edition, (Amazon $8.99)*
- *Page 193*

Thus

$$\chi^2 = \frac{(15-6.84)^2}{6.84} + \frac{(39-40.77)^2}{40.77} + \frac{(87-102.39)^2}{102.39} + \frac{(96-102.39)^2}{102.39}$$

$$+ \frac{(48-40.77)^2}{40.77} + \frac{(15-6.84)^2}{6.84}$$

$$= 23.540$$

© 2014 ConteSolutions

Chi Square GOF for normal distribution

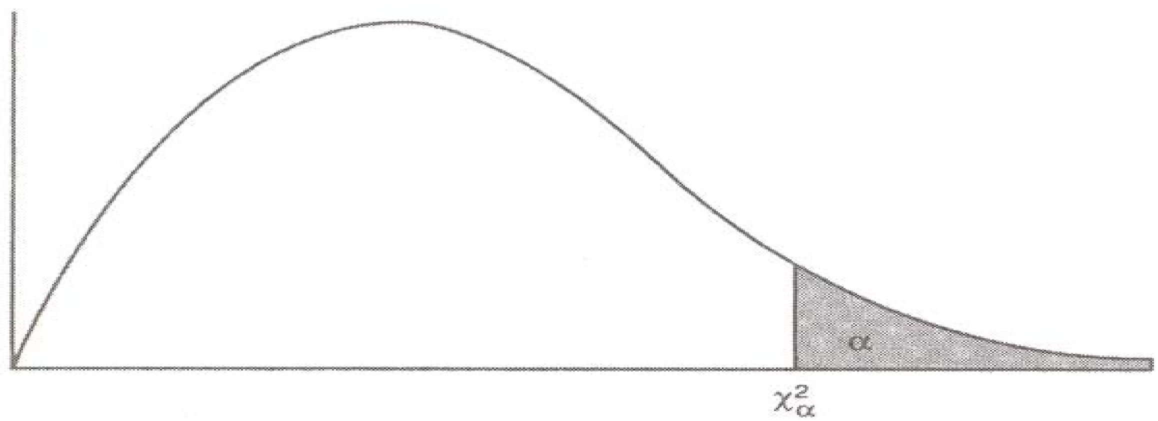
- Degrees of freedom
- Number of cells minus 3
- Technically k – c
- Where k is the number of cells
- And c is the number of population parameters plus 1
- Normal distribution has two population parameters, the mean and standard deviation (2), (2 + 1 = 3)
- k – 3 = 6 - 3 = 3
- Chi-square criteria for alpha = 0.01 and 3 d.f. is 11.34

© 2014 ConteSolutions

Chi-squared criteria

Table C

The χ^2-distribution



df	$\chi^2_{.995}$	$\chi^2_{.990}$	$\chi^2_{.975}$	$\chi^2_{.950}$	$\chi^2_{.900}$	$\chi^2_{.100}$	$\chi^2_{.050}$	$\chi^2_{.025}$	$\chi^2_{.010}$	$\chi^2_{.005}$
1	.0000	.0002	.0010	.0039	.0158	2.706	3.841	5.024	6.635	7.879
2	.0100	.0201	.0506	.1026	.2107	4.605	5.991	7.378	9.210	10.60
3	.0717	.1148	.2158	.3518	.5844	6.251	7.815	9.348	11.34	[illegible]
4	.2070	.2971	.4844	.7107	1.064	7.779	9.448	11.14	13.28	14.86
5	.4117	.5543	.8312	1.145	1.610	9.236	11.07	12.83	15.09	16.75
6	.6757	.8721	1.237	1.635	2.204	10.64	12.59	14.45	16.81	18.55
7	.9893	1.239	1.690	2.167	2.833	12.02	14.07	16.01	18.48	20.28
8	1.344	1.647	2.180	2.732	3.490	13.36	15.51	17.53	20.09	21.95
9	1.735	2.088	2.700	3.325	4.168	14.68	16.92	19.02	21.67	23.59
10	2.156	2.558	3.247	3.940	4.865	15.99	18.31	20.48	23.21	25.19

© 2014 ConteSolutions

Chi-squared calculated

$= 23.540$

Since we are using three measures from the sample (size, mean, and standard deviation), the number of degrees of freedom equals the number of classes minus 3, that is, df = 6 – 3 = 3. With α = .01, the critical χ^2-value is 11.34. Since 23.540 > 11.34, there *is* sufficient evidence to reject H_0 and to conclude that the distribution of assembly times is *not* normal.

© 2014 ConteSolutions

Goodness-of-fit Tests

The chi-square goodness-of-fit (GOF) test can be applied to any univariate distribution with a cumulative distribution function.

H_0: The data follow a specified distribution
H_1: The data do not follow the specified distribution

There observed frequency in each cell is O_i or f_o and the expected or theoretical frequency, E_i or f_e. Any cells which have an expected frequency of less than 5, are combined with an adjacent cell. Chi-square (χ^2) is then summed across all cells:

$$\chi^2 = \sum_{i=1}^{k} \frac{(O_i - E_i)^2}{E_i} \quad \text{or} \quad \chi^2 = \sum_{i=1}^{k} \frac{(f_o - f_e)^2}{f_e}$$

k is the number of cells after combining. c is the number of estimated population parameters for the distribution plus 1. The calculated chi-square is then compared to the chi-square critical value for the following appropriate degrees of freedom.

GOF Distribution	d.f. (k - c)
Weibull (3 parameter)	k - 4
Normal	k - 3
Poisson	k - 2
Binomial	k - 2
Uniform	k - 1

© 2014 ConteSolutions