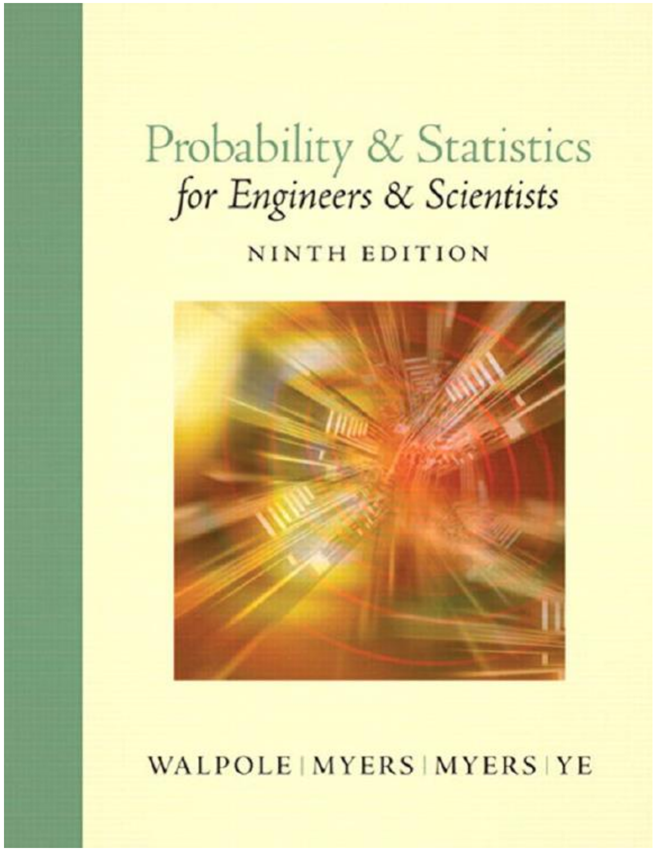
Statistical Analysis

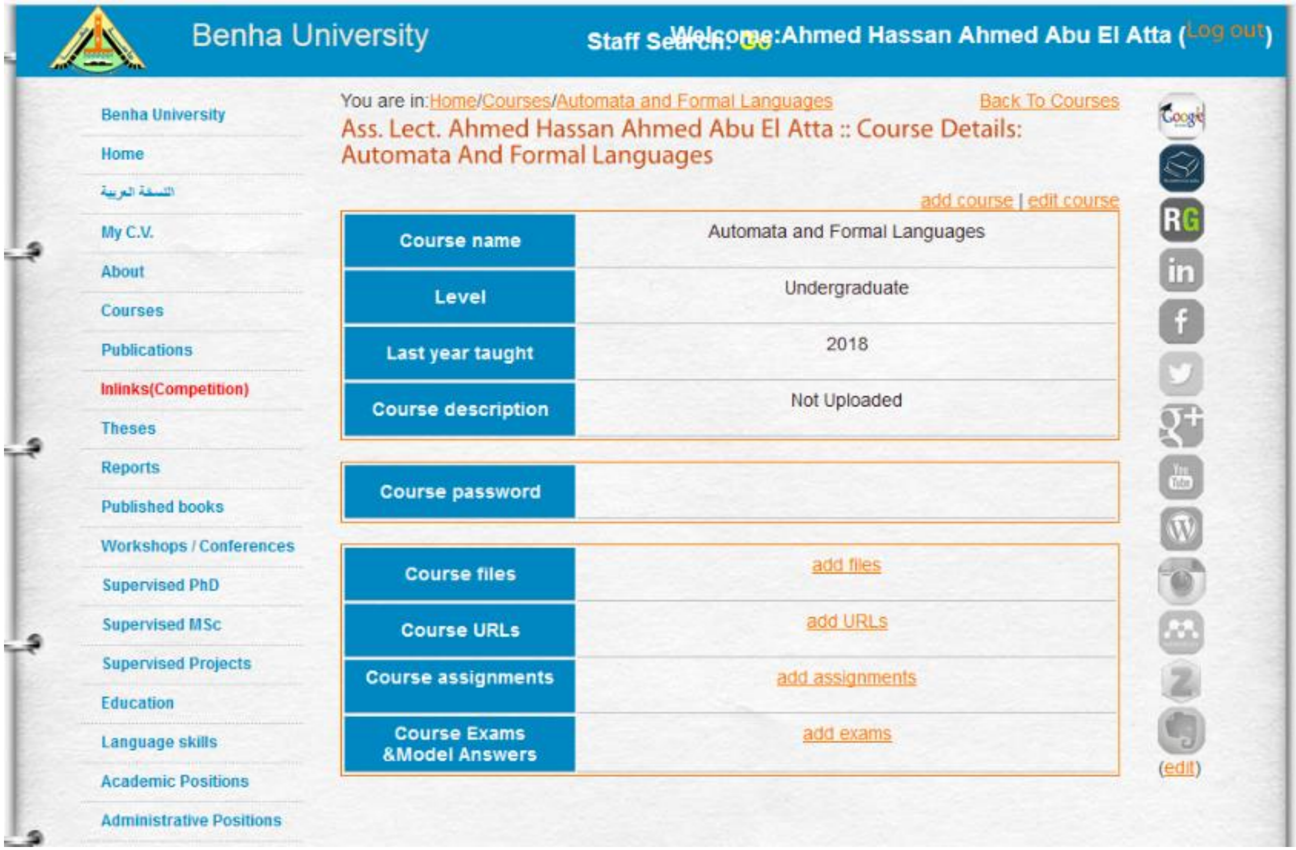
Lecture 02

Books


Probability & Statistics
for Engineers & Scientists
NINTH EDITION
WALPOLE | MYERS | MYERS | YE

PowerPoint

http://www.bu.edu.eg/staff/ahmedaboalatah14-courses/14767



Agenda

- Review "The Central Limit Theorem"
- Sampling Distribution of the Difference between Two Means
- Sampling Distribution of S^2

Sampling Distributions and Data Descriptions

CHAPTER 8

The Central Limit Theorem

Central Limit Theorem: If $\bar{X}$ is the mean of a random sample of size n taken from a population with mean μ and finite variance σ^2, then the limiting form of the distribution of

$$Z = \frac{\bar{X} - \mu}{\sigma/\sqrt{n}},$$

as $n \to \infty$, is the standard normal distribution $n(z; 0, 1)$.

Ex 1:

1. Suppose the grades in a mathematics class are Normally distributed with a mean of 75 and a standard deviation of 6.

What is the probability that the average grade for 9 randomly selected students was at least 80?

Solution :

In this case, $\mu = 75$ *and* $\sigma = 6$.

We need to calculate the probability P($\bar{X}$ > 80) with n = 9.

$$P(\overline{X} \geq 80) = P(\frac{\overline{X} - \mu}{\frac{\sigma}{\sqrt{n}}} \geq \frac{80 - 75}{\frac{6}{\sqrt{9}}}) = P(Z \geq 2.5)$$

$$P(Z \geq 2.5) = 1 - P(Z < 2.5) = P(Z \leq -2.5)$$

$$= 0.0062$$

Sampling Distribution of the Difference between Two Means

Sampling Distribution of the Difference between Two Means

Suppose that we have two populations:

The first with mean μ_1 and variance $\sigma^2{}_1$, and the second with mean μ_2 and variance $\sigma^2{}_2$.

Let the statistic $\bar{X}_1$ represent the mean of a random sample of size n_1 selected from the first population, and

The statistic $\bar{X}_2$ represent the mean of a random sample of size n_2 selected from the second population, independent of the sample from the first population.

Sampling Distribution of the Difference between Two Means

According to Theorem 8.2, the variables X_1 and X_2
are both approximately normally distributed with means μ_1 and μ_2 and variances σ_1^2/n_1 and σ_2^2/n_2, respectively. This approximation improves as n_1 and n_2 increase.

Sampling Distribution of the Difference between Two Means

By choosing independent samples from the two populations we ensure that the variables $\bar{X}_1$ and $\bar{X}_2$ will be independent, and then using Theorem 7.11, with $a_1 = 1$ and $a_2 = -1$, we can conclude that $\bar{X}_1 - \bar{X}_2$ is approximately normally distributed with mean

$$\mu_{\bar{X}_1 - \bar{X}_2} = \mu_{\bar{X}_1} - \mu_{\bar{X}_2} = \mu_1 - \mu_2$$

and variance

$$\sigma^2_{\bar{X}_1 - \bar{X}_2} = \sigma^2_{\bar{X}_1} + \sigma^2_{\bar{X}_2} = \frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}.$$

Theorem 8.3:

If independent samples of size n_1 and n_2 are drawn at random from two populations, discrete or continuous, with means μ_1 and μ_2 and variances σ_1^2 and σ_2^2, respectively, then the sampling distribution of the differences of means, $\bar{X}_1 - \bar{X}_2$, is approximately normally distributed with mean and variance given by

$$\mu_{\bar{X}_1 - \bar{X}_2} = \mu_1 - \mu_2 \text{ and } \sigma^2_{\bar{X}_1 - \bar{X}_2} = \frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}.$$

Hence,

$$Z = \frac{(\bar{X}_1 - \bar{X}_2) - (\mu_1 - \mu_2)}{\sqrt{(\sigma_1^2/n_1) + (\sigma_2^2/n_2)}}$$

is approximately a standard normal variable.

Case Study 8.2: Paint Drying Time:

Paint Drying Time: Two independent experiments are run in which two different types of paint are compared. Eighteen specimens are painted using type A, and the drying time, in hours, is recorded for each. The same is done with type B. The population standard deviations are both known to be 1.0.

Assuming that the mean drying time is equal for the two types of paint, find $P(\bar{X}_A - \bar{X}_B > 1.0)$, where $\bar{X}_A$ and $\bar{X}_B$ are average drying times for samples of size $n_A = n_B = 18$.

***Solution*:** From the sampling distribution of $\bar{X}_A - \bar{X}_B$, we know that the distribution is approximately normal with mean

$$\mu_{\bar{X}_A - \bar{X}_B} = \mu_A - \mu_B = 0$$

and variance

$$\sigma^2_{\bar{X}_A - \bar{X}_B} = \frac{\sigma_A^2}{n_A} + \frac{\sigma_B^2}{n_B} = \frac{1}{18} + \frac{1}{18} = \frac{1}{9}.$$

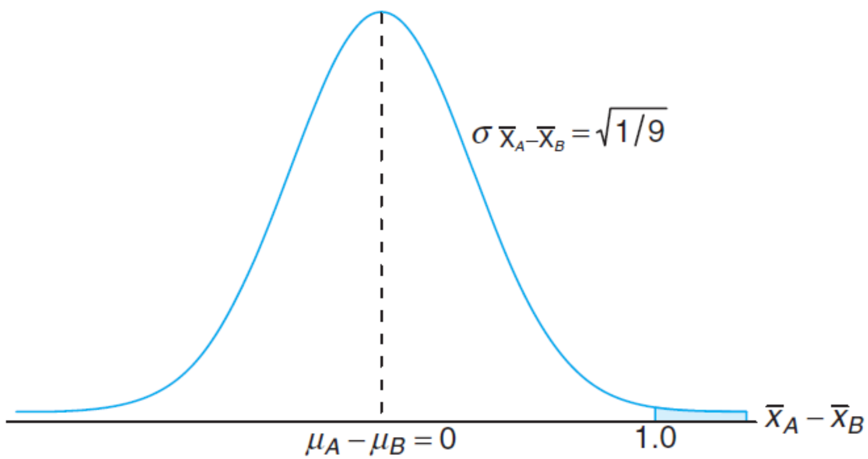


Figure 8.5: Area for Case Study 8.2.

The desired probability is given by the shaded region in Figure 8.5. Corresponding to the value $\bar{X}_A - \bar{X}_B = 1.0$, we have

$$z = \frac{1 - (\mu_A - \mu_B)}{\sqrt{1/9}} = \frac{1 - 0}{\sqrt{1/9}} = 3.0;$$

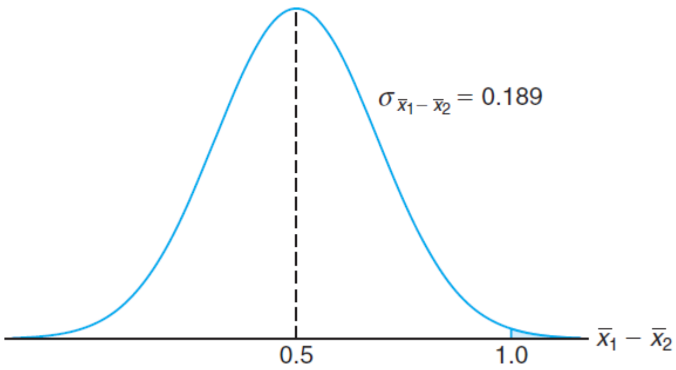
so

$$P(Z > 3.0) = 1 - P(Z < 3.0) = 1 - 0.9987 = 0.0013.$$

Example 8.6:

The television picture tubes of manufacturer A have a mean lifetime of 6.5 years and a standard deviation of 0.9 year, while those of manufacturer B have a mean lifetime of 6.0 years and a standard deviation of 0.8 year. What is the probability that a random sample of 36 tubes from manufacturer A will have a mean lifetime that is at least 1 year more than the mean lifetime of a sample of 49 tubes from manufacturer B?

Solution :



We are given the following information:

Population 1	Population 2
$\mu_1 = 6.5$	$\mu_2 = 6.0$
$\sigma_1 = 0.9$	$\sigma_2 = 0.8$
$n_1 = 36$	$n_2 = 49$

If we use Theorem 8.3, the sampling distribution of $\bar{X}_1 - \bar{X}_2$ will be approximately normal and will have a mean and standard deviation

$$\mu_{\bar{X}_1 - \bar{X}_2} = 6.5 - 6.0 = 0.5 \quad \text{and} \quad \sigma_{\bar{X}_1 - \bar{X}_2} = \sqrt{\frac{0.81}{36} + \frac{0.64}{49}} = 0.189.$$

The probability that the mean lifetime for 36 tubes from manufacturer A will be at least 1 year longer than the mean lifetime for 49 tubes from manufacturer B is given by the area of the shaded region in Figure 8.6. Corresponding to the value $\bar{x}_1 - \bar{x}_2 = 1.0$, we find that

$$z = \frac{1.0 - 0.5}{0.189} = 2.65,$$

and hence

$$P(\bar{X}_1 - \bar{X}_2 \geq 1.0) = P(Z > 2.65) = 1 - P(Z < 2.65)$$
$$= 1 - 0.9960 = 0.0040.$$

Sampling Distribution of S^2

Sampling Distribution of S^2

If a random sample of size n is drawn from a normal population with mean μ and variance σ^2, and the sample variance is computed, we obtain a value of the statistic S^2. We shall proceed to consider the distribution of the statistic $(n-1)S^2/\sigma^2$.

Start From :

$$\sum_{i=1}^{n}(X_i - \mu)^2 :$$

Sampling Distribution of S^2

By the addition and subtraction of the sample mean $\bar{X}$, it is easy to see that

$$\sum_{i=1}^{n}(X_i-\mu)^2 = \sum_{i=1}^{n}[(X_i-\bar{X})+(\bar{X}-\mu)]^2$$

$$= \sum_{i=1}^{n}(X_i-\bar{X})^2 + \sum_{i=1}^{n}(\bar{X}-\mu)^2 + 2(\bar{X}-\mu)\sum_{i=1}^{n}(X_i-\bar{X})$$

$$= \sum_{i=1}^{n}(X_i-\bar{X})^2 + n(\bar{X}-\mu)^2.$$

Dividing each term of the equality by σ^2 and substituting $(n-1)S^2$ for $\sum_{i=1}^{n}(X_i-\bar{X})^2$,

we obtain

$$\frac{1}{\sigma^2}\sum_{i=1}^{n}(X_i-\mu)^2 = \frac{(n-1)S^2}{\sigma^2} + \frac{(\bar{X}-\mu)^2}{\sigma^2/n}.$$

Sampling Distribution of S^2

$$\sum_{i=1}^{n} \frac{(X_i - \mu)^2}{\sigma^2}$$

The First term is a chi-squared random variable with n degrees of freedom. (X_1 to X_n and no relation in the equation)

$$\frac{(\bar{X} - \mu)^2}{\sigma^2/n}$$

The second term on the right-hand side is Z^2, which is a chi-squared random variable with 1 degree of freedom. (only sample average)

$$\frac{(n-1)S^2}{\sigma^2}$$

The third term is a chi-squared random variable with n − 1 degree of freedom. (X_1 to X_n and relation in the equation between them (average))

Degrees of Freedom as a Measure of Sample Information

Example : suppose we have three numbers and we know their average is equal to zero then we have freedom to choose two of the three numbers

$X_1 = 5$, $X_2 = 10$, and average = 0

But for the third one

There is no freedom to choose X_3

$X_3 = 3 *$ average $- X_1 - X_2$

Degrees of Freedom as a Measure of Sample Information

as indicating that when μ is not known

considers the distribution of

$$\sum_{i=1}^{n} \frac{(X_i - \bar{X})^2}{\sigma^2},$$

there is **1 less degree of freedom**, or a degree of freedom is lost in the estimation of μ (i.e., when μ is replaced by $\bar{x}$). In other words, there are n degrees of freedom, or independent *pieces of information*, in the random sample from the normal distribution. When the data (the values in the sample) are used to compute the mean, there is 1 less degree of freedom in the information used to estimate σ^2.

Theorem 8.4:

If S^2 is the variance of a random sample of size n taken from a normal population having the variance σ^2, then the statistic

$$\chi^2 = \frac{(n-1)S^2}{\sigma^2} = \sum_{i=1}^{n} \frac{(X_i - \bar{X})^2}{\sigma^2}$$

has a chi-squared distribution with $v = n - 1$ degrees of freedom.

χ^2_α represent the χ^2 value above which we find an area of α.

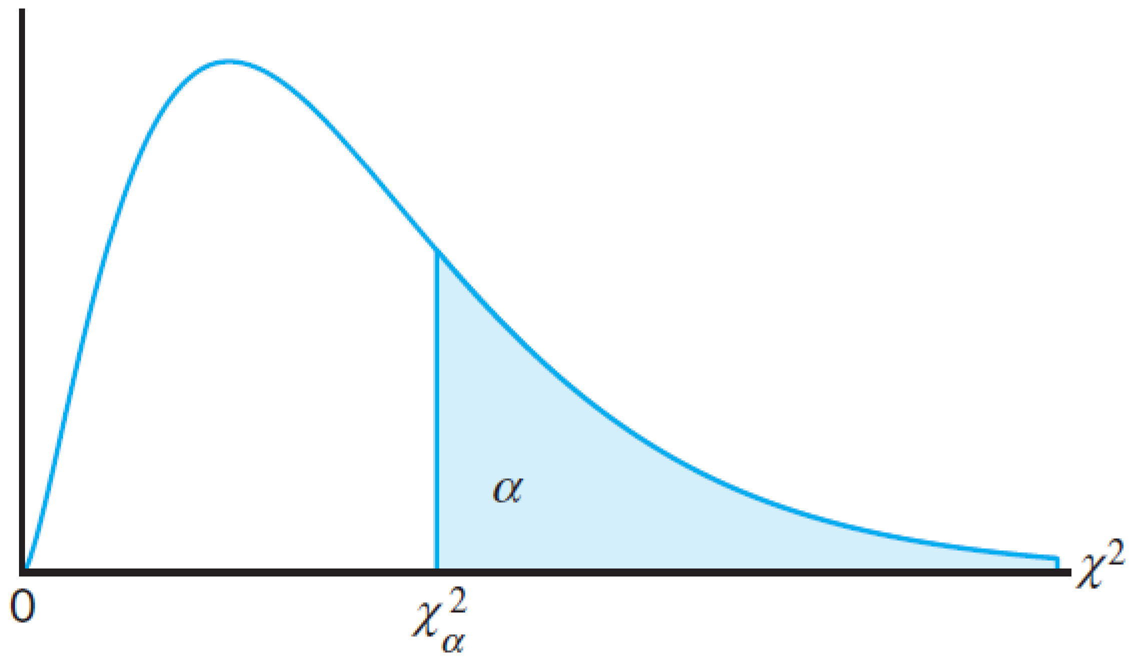


Table A.5 gives values of χ^2_α for various values of α and v. The areas, α, are the column headings; the degrees of freedom, v, are given in the left column; and the table entries are the χ^2 values. Hence, the χ^2 value with 7 degrees of freedom, leaving an area of 0.05 to the right, is $\chi^2_{0.05} = 14.067$. Owing to lack of symmetry, we must also use the tables to find $\chi^2_{0.95} = 2.167$ for $v = 7$.

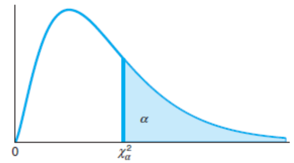


Table A.5 Critical Values of the Chi-Squared Distribution

	α									
v	**0.995**	**0.99**	**0.98**	**0.975**	**0.95**	**0.90**	**0.80**	**0.75**	**0.70**	**0.50**
1	0.0^4393	0.0^3157	0.0^3628	0.0^3982	0.00393	0.0158	0.0642	0.102	0.148	0.455
2	0.0100	0.0201	0.0404	0.0506	0.103	0.211	0.446	0.575	0.713	1.386
3	0.0717	0.115	0.185	0.216	0.352	0.584	1.005	1.213	1.424	2.366
4	0.207	0.297	0.429	0.484	0.711	1.064	1.649	1.923	2.195	3.357
5	0.412	0.554	0.752	0.831	1.145	1.610	2.343	2.675	3.000	4.351
6	0.676	0.872	1.134	1.237	1.635	2.204	3.070	3.455	3.828	5.348
7	0.989	1.239	1.564	1.690	2.167	2.833	3.822	4.255	4.671	6.346
8	1.344	1.647	2.032	2.180	2.733	3.490	4.594	5.071	5.527	7.344
9	1.735	2.088	2.532	2.700	3.325	4.168	5.380	5.899	6.393	8.343
10	2.156	2.558	3.059	3.247	3.940	4.865	6.179	6.737	7.267	9.342
11	2.603	3.053	3.609	3.816	4.575	5.578	6.989	7.584	8.148	10.341
12	3.074	3.571	4.178	4.404	5.226	6.304	7.807	8.438	9.034	11.340
13	3.565	4.107	4.765	5.009	5.892	7.041	8.634	9.299	9.926	12.340
14	4.075	4.660	5.368	5.629	6.571	7.790	9.467	10.165	10.821	13.339
15	4.601	5.229	5.985	6.262	7.261	8.547	10.307	11.037	11.721	14.339
16	5.142	5.812	6.614	6.908	7.962	9.312	11.152	11.912	12.624	15.338
17	5.697	6.408	7.255	7.564	8.672	10.085	12.002	12.792	13.531	16.338
18	6.265	7.015	7.906	8.231	9.390	10.865	12.857	13.675	14.440	17.338
19	6.844	7.633	8.567	8.907	10.117	11.651	13.716	14.562	15.352	18.338
20	7.434	8.260	9.237	9.591	10.851	12.443	14.578	15.452	16.266	19.337
21	8.034	8.897	9.915	10.283	11.591	13.240	15.445	16.344	17.182	20.337
22	8.643	9.542	10.600	10.982	12.338	14.041	16.314	17.240	18.101	21.337
23	9.260	10.196	11.293	11.689	13.091	14.848	17.187	18.137	19.021	22.337
24	9.886	10.856	11.992	12.401	13.848	15.659	18.062	19.037	19.943	23.337
25	10.520	11.524	12.697	13.120	14.611	16.473	18.940	19.939	20.867	24.337
26	11.160	12.198	13.409	13.844	15.379	17.292	19.820	20.843	21.792	25.336
27	11.808	12.878	14.125	14.573	16.151	18.114	20.703	21.749	22.719	26.336
28	12.461	13.565	14.847	15.308	16.928	18.939	21.588	22.657	23.647	27.336
29	13.121	14.256	15.574	16.047	17.708	19.768	22.475	23.567	24.577	28.336
30	13.787	14.953	16.306	16.791	18.493	20.599	23.364	24.478	25.508	29.336
40	20.707	22.164	23.838	24.433	26.509	29.051	32.345	33.66	34.872	39.335
50	27.991	29.707	31.664	32.357	34.764	37.689	41.449	42.942	44.313	49.335
60	35.534	37.485	39.699	40.482	43.188	46.459	50.641	52.294	53.809	59.335

Table A.5 (continued) Critical Values of the Chi-Squared Distribution

	α									
v	0.30	0.25	0.20	0.10	0.05	0.025	0.02	0.01	0.005	0.001
1	1.074	1.323	1.642	2.706	3.841	5.024	5.412	6.635	7.879	10.827
2	2.408	2.773	3.219	4.605	5.991	7.378	7.824	9.210	10.597	13.815
3	3.665	4.108	4.642	6.251	7.815	9.348	9.837	11.345	12.838	16.266
4	4.878	5.385	5.989	7.779	9.488	11.143	11.668	13.277	14.860	18.466
5	6.064	6.626	7.289	9.236	11.070	12.832	13.388	15.086	16.750	20.515
6	7.231	7.841	8.558	10.645	12.592	14.449	15.033	16.812	18.548	22.457
7	8.383	9.037	9.803	12.017	14.067	16.013	16.622	18.475	20.278	24.321
8	9.524	10.219	11.030	13.362	15.507	17.535	18.168	20.090	21.955	26.124
9	10.656	11.389	12.242	14.684	16.919	19.023	19.679	21.666	23.589	27.877
10	11.781	12.549	13.442	15.987	18.307	20.483	21.161	23.209	25.188	29.588
11	12.899	13.701	14.631	17.275	19.675	21.920	22.618	24.725	26.757	31.264
12	14.011	14.845	15.812	18.549	21.026	23.337	24.054	26.217	28.300	32.909
13	15.119	15.984	16.985	19.812	22.362	24.736	25.471	27.688	29.819	34.527
14	16.222	17.117	18.151	21.064	23.685	26.119	26.873	29.141	31.319	36.124
15	17.322	18.245	19.311	22.307	24.996	27.488	28.259	30.578	32.801	37.698
16	18.418	19.369	20.465	23.542	26.296	28.845	29.633	32.000	34.267	39.252
17	19.511	20.489	21.615	24.769	27.587	30.191	30.995	33.409	35.718	40.791
18	20.601	21.605	22.760	25.989	28.869	31.526	32.346	34.805	37.156	42.312
19	21.689	22.718	23.900	27.204	30.144	32.852	33.687	36.191	38.582	43.819
20	22.775	23.828	25.038	28.412	31.410	34.170	35.020	37.566	39.997	45.314
21	23.858	24.935	26.171	29.615	32.671	35.479	36.343	38.932	41.401	46.796
22	24.939	26.039	27.301	30.813	33.924	36.781	37.659	40.289	42.796	48.268
23	26.018	27.141	28.429	32.007	35.172	38.076	38.968	41.638	44.181	49.728
24	27.096	28.241	29.553	33.196	36.415	39.364	40.270	42.980	45.558	51.179
25	28.172	29.339	30.675	34.382	37.652	40.646	41.566	44.314	46.928	52.619
26	29.246	30.435	31.795	35.563	38.885	41.923	42.856	45.642	48.290	54.051
27	30.319	31.528	32.912	36.741	40.113	43.195	44.140	46.963	49.645	55.475
28	31.391	32.620	34.027	37.916	41.337	44.461	45.419	48.278	50.994	56.892
29	32.461	33.711	35.139	39.087	42.557	45.722	46.693	49.588	52.335	58.301
30	33.530	34.800	36.250	40.256	43.773	46.979	47.962	50.892	53.672	59.702
40	44.165	45.616	47.269	51.805	55.758	59.342	60.436	63.691	66.766	73.403
50	54.723	56.334	58.164	63.167	67.505	71.420	72.613	76.154	79.490	86.660
60	65.226	66.981	68.972	74.397	79.082	83.298	84.58	88.379	91.952	99.608

Exactly 95% of a chi-squared distribution lies between $\chi^2_{0.975}$ and $\chi^2_{0.025}$. A χ^2 value falling to the right of $\chi^2_{0.025}$ is not likely to occur unless our assumed value of σ^2 is too small. Similarly, a χ^2 value falling to the left of $\chi^2_{0.975}$ is unlikely unless our assumed value of σ^2 is too large. In other words, it is possible to have a χ^2 value to the left of $\chi^2_{0.975}$ or to the right of $\chi^2_{0.025}$ when σ^2 is correct, but if this should occur, it is more probable that the assumed value of σ^2 is in error.

Example 8.7:

A manufacturer of car batteries guarantees that the batteries will last, on average, 3 years with a standard deviation of 1 year. If five of these batteries have lifetimes of 1.9, 2.4, 3.0, 3.5, and 4.2 years, should the manufacturer still be convinced that the batteries have a standard deviation of 1 year? Assume that the battery lifetime follows a normal distribution.

Solution :

We first find the sample variance using Theorem 8.1,

$$s^2 = \frac{(5)(48.26) - (15)^2}{(5)(4)} = 0.815.$$

Then

$$\chi^2 = \frac{(4)(0.815)}{1} = 3.26$$

is a value from a chi-squared distribution with 4 degrees of freedom. Since 95% of the χ^2 values with 4 degrees of freedom fall between 0.484 and 11.143, the computed value with $\sigma^2 = 1$ is reasonable, and therefore the manufacturer has no reason to suspect that the standard deviation is other than 1 year.