Technical Support Document

Calculating Anderson-Darling "By Hand" for Uncensored Data

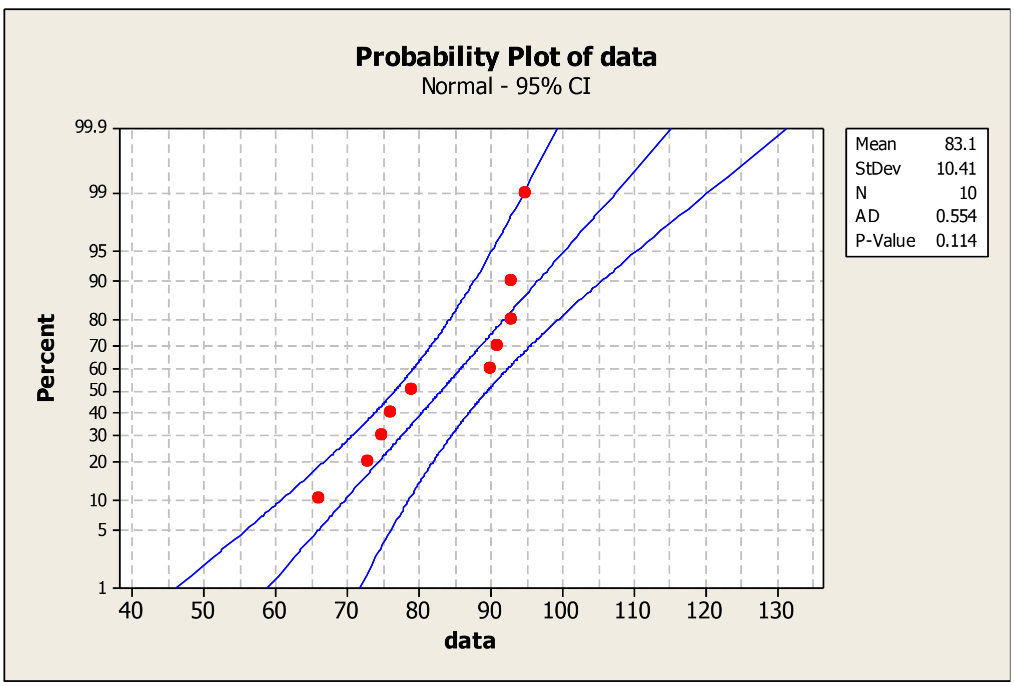
Store the following uncensored data in column C1:

C1
Data
95
93
90
91
93
76
79
73
75
66

You can calculate the Anderson-Darling (AD) statistic "by hand" using different functions in Minitab.

Using Probability Plot or Individual Distribution Identification

Using **Graph > Probability Plot** (or **Stat > Quality Tools > Individual Distribution Identification**) to check for normality, you obtain the following output:



Note: The plotted points on the probability plot may differ depending on the plot point method selected in **Tools > Options > Individual Graphs > Probability Plots**.

Calculating Anderson-Darling "By Hand" for Uncensored Data

To calculate the AD statistic "by hand," do the following:

1 Choose **Data > Sort** and sort the data in ascending order.

2 Choose **Calc > Probability Distributions > Normal**. Calculate the cumulative probabilities for a normal distribution with mean = 83.1 and standard deviation = 10.41 (from the output). Use 'data' as the Input column and store the results in 'z_i'. After the last row in 'z_i', enter 0.999999999999.

3 Choose **Calc > Calculator**. Store the result of the following expression in 'Fn(z_i)':

PARS(RCOUNT('data'))/COUNT('data')

Replace the value in the last row with the result of the following expression, where N is the total number of data points:

(N-0.1)/N (= 0.99 in this example)

After the last row in 'Fn(z_i)' enter 1.

4 Choose **Stat > Time Series > Lag**. Store lags of 1 from 'z_i' into 'z_(i-1)'. Replace the missing value in the first row with a zero.

5 Using **Stat > Time Series > Lag**, store lags of 1 from 'Fn(z_i)' into 'Fn(z_(i-1))'. Replace the missing value in the first row with a zero.

6 Choose **Calc > Calculator**. Store the result of the following expression in 'A_i':

-'z_i'-LN(1-'z_i')+'z_(i-1)'+LN(1-'z_(i-1)')

7 Using **Calc > Calculator**, store the result of the following expression in 'B_i':

2*'Fn(z_(i-1))'*LN(1-'z_i')-2*'Fn(z_(i-1))'*LN(1-'z_(i-1)')

8 Using **Calc > Calculator**, store the result of the following expression in 'C_i':

('Fn(z_(i-1))'**2)*LN('z_i')-('Fn(z_(i-1))'**2)*LN(1-'z_i')-('Fn(z_(i-1))'**2)*LN('z_(i-1)')+('Fn(z_(i-1))'**2)*LN(1-'z_(i-1)')

Replace the missing value in the first row of 'C_i' with a zero.

9 Using **Calc > Calculator**, store the result of the following expression in 'AD':

COUNT('data') * SUM('A_i' + 'B_i' + 'C_i')

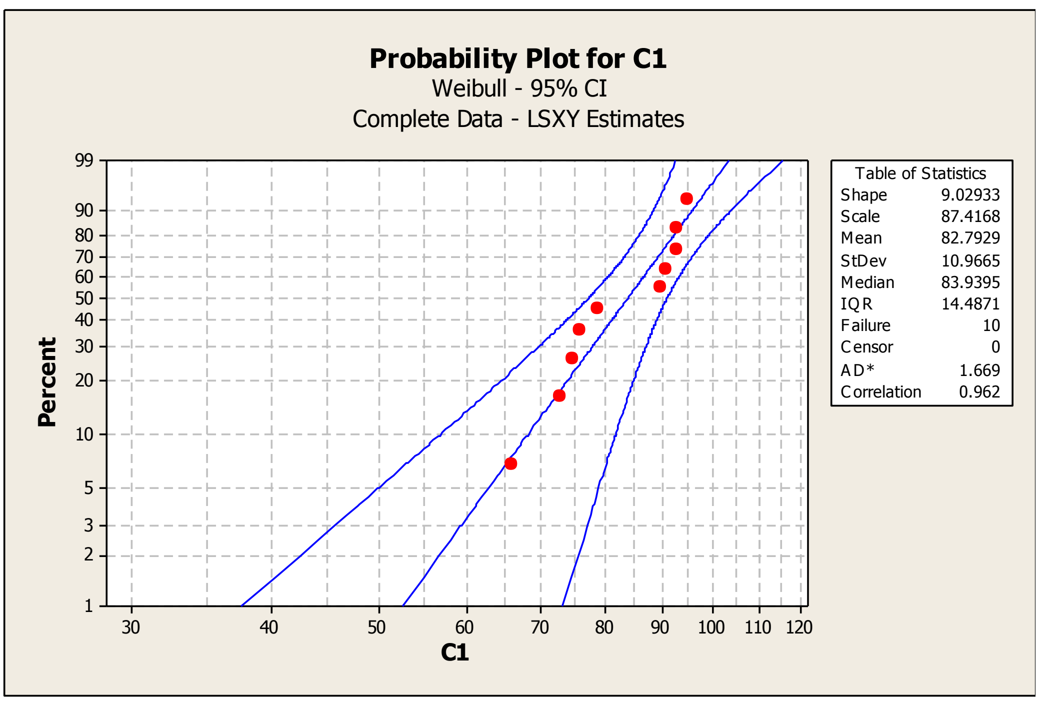
The result will match (within rounding) the AD statistic printed on the output:

Anderson-Darling ***

↓	C1	C2	C3	C4	C5	C6	C7	C8	C9
	data	z_i	Fn(z_i)	z_(i-1)	Fn(z_(i-1))	A_i	B_i	C_i	AD
1	66	0.05023	0.10	0.000000	0.00	0.0013	0.0000	0.0000	0.552051
2	73	0.16597	0.20	0.050228	0.10	0.0142	-0.0260	0.0133	
3	75	0.21826	0.30	0.165968	0.20	0.0125	-0.0259	0.0135	
4	76	0.24761	0.40	0.218256	0.30	0.0089	-0.0230	0.0148	
5	79	0.34685	0.50	0.247608	0.40	0.0422	-0.1132	0.0766	
6	90	0.74628	0.60	0.346845	0.50	0.5461	-0.9456	0.4279	
7	91	0.77604	0.70	0.746278	0.60	0.0950	-0.1497	0.0590	
8	93	0.82920	0.80	0.776040	0.70	0.2178	-0.3794	0.1652	
9	93	0.82920	0.90	0.829200	0.80	0.0000	0.0000	0.0000	
10	95	0.87351	0.99	0.829200	0.90	0.2560	-0.5406	0.2854	
11		1.00000	1.00	0.873508	0.99	25.4370	-50.6157	25.1873	

Using Reliability/Survival Commands

Using **Stat > Reliability/Survival > Distribution Analysis (Right Censoring) > Parametric Distribution Analysis** (or another **Reliability/Survival** command) with a Weibull distribution you obtain the output below:



Note: This output was created with the Median Rank plot point method selected in **Tools** > **Options** > **Individual Graphs** > **Probability Plots**. If other plot point methods are used, the probability plot and corresponding AD* value will change.

To get the AD* statistic "by hand", do the following:

1. Choose **Data** > **Sort**. Sort the data in ascending order.

2. Choose **Calc** > **Probability Distributions** > **Weibull**. Calculate the cumulative probabilities for a Weibull distribution with Shape = 9.02933 and Scale = 87.4168 (from the output). Use 'data' as an Input column and store the results in 'z_i'. After the last row in 'z_i', enter 0.999999999999.

3. Using **Calc** > **Calculator**, store the result of the following expression in 'Fn(z_i)':

   (PARSUM(RCOUNT('data'))-0.3)/(COUNT('data')+0.4)

   Note: This formula is used for the Median Rank method only. If a different method is selected in **Tools** > **Options** > **Individual Graphs** > **Probability Plots**, then the corresponding formula from this table should be used:

Method	**Formula**
Median Rank	(PARSUM(RCOUNT('data'))-0.3)/(COUNT('data')+0.4)
Mean Rank	PARSUM(RCOUNT('data'))/(COUNT('data')+1)
Modified K-M	(PARSUM(RCOUNT('data'))-0.5)/COUNT('data')
Kaplan-Meier	PARSUM(RCOUNT('data'))/COUNT('data')

   Note: If you use the Kaplan-Meier formula, replace the value in the last row with the result of the following expression, where N is the total number of data points:

   (N-0.1)/N (=0.99 in this example)

   After the last row in 'Fn(z_i)' enter 1.

4. Choose **Stat** > **Time Series** > **Lag**. Store lags of 1 from 'z_i' into 'z_(i-1)'. Replace the missing value in the first row with a zero.

5. Using **Stat** > **Time Series** > **Lag**, store lags of 1 from 'Fn(z_i)' into 'Fn(z_(i-1))'. Replace the missing value in the first row with a zero.

6. Choose **Calc** > **Calculator**. Store the result of the following expression in 'A_i':

   -'z_i'-LN(1-'z_i')+'z_(i-1)'+LN(1-'z_(i-1)')

7 Using **Calc > Calculator**, store the result of the following expression in 'B_i':

2*'Fn(z_(i-1))'*LN(1-'z_i')-2*'Fn(z_(i-1))'*LN(1-'z_(i-1)')

8 Using **Calc > Calculator**, store the result of the following expression in 'C_i':

('Fn(z_(i-1))'**2)*LN('z_i')-('Fn(z_(i-1))'**2)*LN(1-'z_i')-('Fn(z_(i-1))'**2)*LN('z_(i-1)')+('Fn(z_(i-1))'**2)*LN(1-'z_(i-1)')

Replace the missing value in the first row of 'C_i' with a zero.

9 Using **Calc > Calculator**, store the result of the following expression in 'AD':

COUNT('data') * SUM('A_i' + 'B_i' + 'C_i')

The result will match (within rounding) the AD* statistic printed on the output:

Anderson-Darling ***

↓	C1	C2	C3	C4	C5	C6	C7	C8	C9
	data	z_i	Fn(z_i)	z_(i-1)	Fn(z_(i-1))	A_i	B_i	C_i	AD
1	66	0.07602	0.06731	0.000000	0.000000	0.0030	0.0000	0.0000	1.66871
2	73	0.17836	0.16346	0.076016	0.067308	0.0150	-0.0158	0.0044	
3	75	0.22179	0.25962	0.178359	0.163462	0.0109	-0.0178	0.0073	
4	76	0.24619	0.35577	0.221785	0.259615	0.0075	-0.0165	0.0092	
5	79	0.33026	0.45192	0.246185	0.355769	0.0342	-0.0841	0.0522	
6	90	0.72768	0.54808	0.330261	0.451923	0.5025	-0.8134	0.3451	
7	91	0.76242	0.64423	0.727677	0.548077	0.1017	-0.1496	0.0550	
8	93	0.82605	0.74038	0.762418	0.644231	0.2481	-0.4016	0.1626	
9	93	0.82605	0.83654	0.826046	0.740385	0.0000	0.0000	0.0000	
10	95	0.87990	0.93269	0.826046	0.836538	0.3166	-0.6198	0.3034	
11		1.00000	1.00000	0.879899	0.932692	25.3915	-47.5890	22.3042	