

https://dspace.org/

A comparison of the jacknife and bootstrap estimators in linear models with reference to production models used by Sasol

Angus, Stuart Maxwell

2014-09-16
http://hdl.handle.net/10539/15551

Downloaded from a DSpace repository. https://dspace.org/

Table 5.5.9 : Bias, Sample Variance and Mean Square Error for a One Parameter Linear Model using the Bootstrap Method

(Errors are serially correlated - $\alpha=0.5$)

	Number of Bootstrap Replications					
	20	40	60	80	100	120
Bias	0.1286	0.1284	0.1292	0.1279	0.1286	0.1285
Sample Variance	9.77×10^{-3}	9.51×10^{-3}	9.21×10^{-3}	9.42×10^{-3}	9.30×10^{-3}	9.10×10^{-3}
Mean Square Error	2.63×10^{-2}	2.60×10^{-2}	2.59×10^{-2}	2.58×10^{-2}	2.58×10^{-2}	2.56×10^{-2}

Error Distribution

Exponential : Mean=0.5

Sample Size

N=30

Starting Value

$Y_0=0.1$

Mean

$\mu=0.7$

Table 5.5.10 : Sam e Confidence Intervals for a One Parameter Linear Model using the Bootstrap Method

(Errors are serially correlated - α=0.5)

		Number of Bootstrap Replications					
		20	40	60	80	100	120
Coverage	95%	81.9	85.4	87.1	87.3	88.4	89.6
	99%	94.2	96.3	96.7	97.3	97.4	97.5
Average Length	95%	4.59×10^{-1}	4.64×10^{-1}	4.65×10^{-1}	4.65×10^{-1}	4.67×10^{-1}	4.67×10^{-1}
	99%	6.03×10^{-1}	6.10×10^{-1}	6.11×10^{-1}	6.12×10^{-1}	6.14×10^{-1}	6.14×10^{-1}
Variance of Length	95%	1.09×10^{-2}	8.07×10^{-3}	6.39×10^{-3}	5.81×10^{-3}	5.78×10^{-3}	5.46×10^{-3}
	99%	1.88×10^{-2}	1.39×10^{-2}	1.10×10^{-2}	1.00×10^{-2}	9.98×10^{-3}	9.42×10^{-3}

Error Distribution

Exponential : Mean=0.5

Sample Size

N=30

Starting Value

Y_0=0.1

Mean

μ=0.7

Table 5.5.11 : Bias, Sample Variance and Mean Square Error for a One Parameter Linear Model using the Bootstrap Method

(Errors are serially correlated - $\alpha=0.5$)

	Number of Bootstrap Replications					
	20	40	60	80	100	120
Bias	0.1179	0.1180	0.1173	0.1172	0.1174	0.1176
Sample Variance	1.04×10^{-2}	1.02×10^{-2}	1.02×10^{-2}	1.01×10^{-2}	1.00×10^{-2}	1.02×10^{-2}
Mean Square Error	2.43×10^{-2}	2.41×10^{-2}	2.40×10^{-2}	2.38×10^{-2}	2.38×10^{-2}	2.40×10^{-2}

Error Distribution

Gamma : Mean=0.5 (k=2)

Starting Value

$Y_o=0.1$

Sample Size

N=30

Mean

$\mu=0.7$

TAble 5.5.12 : Sample Coufidence Intervals for a One Parameter Linear Model using the Bootstrap Method

(Errors are serially correlated - α=0.5)

		Number of Bootstrap Replications					
		20	40	60	80	100	120
Coverage	95%	69.9	71.3	71.8	73.9	72.3	72.3
	99%	86.0	88.6	90.2	90.3	91.9	92.0
Average Length	95%	3.94×10^{-1}	3.98×10^{-1}	5.23×10^{-1}	4.01×10^{-1}	3.99×10^{-1}	3.99×10^{-1}
	99%	5.18×10^{-1}	5.23×10^{-1}	1.17×10^{-1}	5.27×10^{-1}	5.23×10^{-1}	5.24×10^{-1}
Variance of Length	95%	9.90×10^{-3}	7.76×10^{-3}	6.75×10^{-3}	6.44×10^{-3}	5.97×10^{-3}	5.78×10^{-3}
	99%	1.71×10^{-2}	1.34×10^{-2}	1.17×10^{-2}	1.11×10^{-2}	1.03×10^{-2}	9.98×10^{-3}

Error Distribution

Gamma : Mean=0.5 (k=2)

Starting Value

Y_o=0.1

Sample Size

N=30

Mean

μ=0.7

Table 5.5.13 : Bias, Sample Variance and Mean Square Error for a One Parameter Linear Model using the Bootstrap Method

(Errors are serially correlated - α=0.5)

	Number of Bootstrap Replications					
	20	40	60	80	100	120
Bias	0.1352	0.1348	0.1359	0.1345	0.1347	0.1348
Sample Variance	9.24×10^{-3}	8.44×10^{-3}	8.21×10^{-3}	8.34×10^{-3}	8.21×10^{-3}	8.05×10^{-3}
Mean Square Error	2.75×10^{-2}	2.66×10^{-2}	2.67×10^{-2}	2.64×10^{-2}	2.63×10^{-2}	1.62×10^{-2}

Error Distribution

Long tailed h : Mean=0.5 (h=0.3)

Sample Size

N=30

Starting VAlue

Y_o=0.1

Mean

μ=0.7

Table 5.5.14 : Sample Confidence Intervals for a One Parameter Linear Model using the Bootstrap Method

(Errors are serially correlated - $\alpha=0.5$)

		Number of Bootstrap Replications					
		20	**40**	**60**	**80**	**100**	**120**
Coverage	95%	88.9	93.0	94.7	94.6	95.2	95.0
	99%	95.4	96.8	96.8	96.8	97.2	96.8
Average Length	95%	5.05×10^{-1}	5.14×10^{-1}	5.15×10^{-1}	5.16×10^{-1}	5.16×10^{-1}	5.16×10^{-1}
	99%	6.64×10^{-1}	6.76×10^{-1}	6.76×10^{-1}	6.78×10^{-1}	6.78×10^{-1}	6.78×10^{-1}
Variance of Length	95%	1.22×10^{-2}	9.07×10^{-3}	6.81×10^{-3}	6.29×10^{-3}	5.86×10^{-3}	5.61×10^{-3}
	99%	2.11×10^{-2}	1.57×10^{-2}	1.18×10^{-2}	1.09×10^{-2}	1.01×10^{-2}	9.69×10^{-3}

Error Distribution

Long tailed h : Mean=0.5 (h=0.3)

Sample Size

N=30

Starting Value

$Y_o=0.1$

Mean

$\mu=0.7$

The results of the percentile method are shown in Table 5.5.15. If these results are compared to the bootstrap results in Tables 5.5.8, 5.5.10, 5.5.12 and 5.5.14 respectively, it can be seen that the percentile method is better than the standard bootstrap method in terms of coverage. For the 95% confidence limit, the percentile method also produces shorter average lengths and smaller variances in length.

5.6 Summary of simulation results

The following conclusions can be made regarding the results presented in the previous section:

(i) the non-parametric jackknife and bootstrap methods do not generally improve on the least squares method when the error deviates are independent identically distributed. The only improvement was a reduction in the bias estimate using the jackknife method although this was at the expense of a higher mean square error.

(ii) the bootstrap method outperforms both the least squares and jackknife method when the error deviates are serially corrected. The bootstrap method not only produces better bias and mean square error results but also more robust confidence limits. Also, the percentile method gave very robust confidence limits and, generally, was an improvement on the standard bootstrap method

(iii) due to the inter-dependence between successive values in time-series models, the jackknife method was shown to be generally an unsatisfactory method in terms of parameter estimation and confidence limit approximation.

Table 5.5.15 Sample Confidence Intervals for a One Parameter Linear Model

(Errors are serially correlated - α=0.5)

			Bootstrap Percentile Method
A	Coverage	95%	77.3
		99%	85.7
	Average Length	95%	5.10×10^{-1}
		99%	6.91×10^{-1}
	Variance of Length	95%	2.16×10^{-2}
		99%	4.05×10^{-2}
B	Coverage	95%	94.1
		99%	98.9
	Average Length	95%	4.57×10^{-1}
		99%	6.23×10^{-1}
	Variance of Length	95%	5.32×10^{-3}
		99%	1.48×10^{-2}
C	Coverage	95%	79.4
		99%	95.4
	Average Length	95%	3.93×10^{-1}
		99%	5.30×10^{-1}
	Variance of Length	95%	6.37×10^{-3}
		99%	1.35×10^{-2}
D	Coverage	95%	95.9
		99%	97.6
	Average Length	95%	5.03×10^{-1}
		99%	6.99×10^{-1}
	Variance of Length	95%	5.69×10^{-3}
		99%	1.76×10^{-2}

Error Distributions

A Normal : Mean=0.', S.D.=0.5
B Exponential : Mean=0.5
C Gamma : Mean=0.5 (k=2)
D Long tailed h : Mean=0.5 (h=0.3)

Starting Value	**Sample Size**	**Mean**
Y_0=0.1	N=30	μ=0.7

CHAPTER 6

PRACTICAL APPLICATIONS OF THE JACKKNIFE METHOD

6.1 Introduction

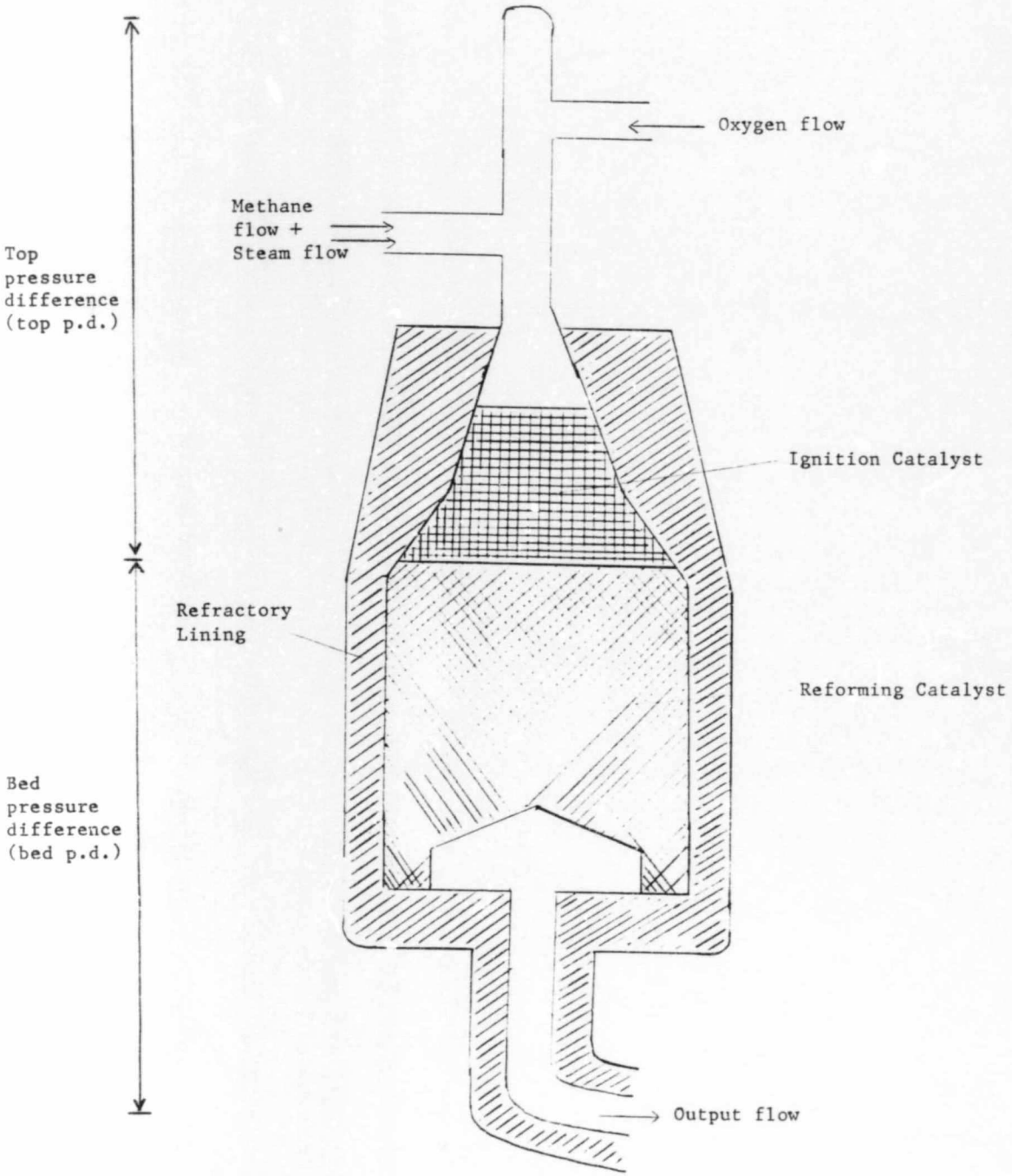
This chapter deals with the use of the jackknife method in practical problems at the oil from coal conversion plants at Sasol Two and Three at Secunda, South Africa, using actual plant data. In particular, it compares the jackknife to standard methods, with respect to parameter estimation. The bootstrap method, although producing almost unbiased variance estimates, does not generally improve on bias reduction, as indicated in Chapter 3, 4 and 5. Consequently, since in practice, greater emphasis is usually placed on obtaining good parameter estimates rather than accurate variance estimates or robust confidence intervals coverage, the bootstrap method was not evaluated.

The remainder of this chapter investigates the performance of the jackknife versus standard methods for four different case studies, which were undertaken at Sasol between January 1984 and December 1987. Due to the confidentiality of the data, for the purpose of this thesis, it has been necessary to 'generalise' on the description of some of the operating variables which were considered.

6.2 Case Study Evaluation

The first case study involves the performance of a methane gas reformer. The primary function of a reformer is to convert a methane rich gas steam into carbon monoxide which is a necessary constituent in the oil from coal process. Two other input gas flows, namely steam and oxygen, are necessary in the conversion process, which requires two different types of catalyst to promote the reaction. A simple schematic diagram of a gas reformer is shown below :

Figure 6.2.1 : Methane Gas Reformer



Of primary concern is the top pressure difference (top p.d.) and bed pressure difference (bed p.d.) - as figure 6.2.1. shows, which determines the volume of gas which flows through the reformer.

For planning purposes, it is very important to know what the maximum possible methane gas flow is, given the deteriorating state of the catalyst beds over time, which is reflected in increasing values of the top and bed p.d. The problem can be expresses mathematically as follows :

Let x_1 = actual methane gas flow (knm^3/h)

x_2 = actual oxygen flow (knm^3/h)

x_3 = actual steam flow (kg/h)

r = volumetric ratio between two given gas constituents

P_1 = actual top p.d. (kpA)

P_2 = actual bed p.d. (kpA)

P_3 = maximum top p.d. (kpA)

P_4 = maximum bed p.d. (kpA)

y = maximum methane gas flow (knm^3/h)

From chemical engineering,

$$\frac{\text{maximum mass flow}}{\text{actual mass flow}} = \sqrt{\frac{\text{maximum pressure difference}}{\text{actual pressure difference}}} \qquad (6.2.1)$$

Using the appropriate atomic weights for the gas streams, it has been suggested that the maximum methane gas flow y is dependent on x_1, x_2, x_3, r, P_1, P_2, P_3, P_4 through the non-linear models:

$$y = A_1^{k_1} B + e_1 \qquad (6.2.2)$$

or $$y = A_2^{k_2} B + e_2, \qquad (6.2.3)$$

where $A_1 = \left(\frac{P_3}{P_1}\right)$, $A_2 = \left(\frac{P_3 + P_4}{P_1 + P_2}\right)$,

$$B = \left(\frac{18x_1 + 32x_2 + 22.4x_3}{35.28 + 16.2r}\right)$$

and e_1 and e_2 are random errors which have mean 0 and variance σ_1^2 and σ_2^2 respectively.

A simp[illegible]re for estimating a parameter in a non-linear equation, [illegible]se the Hooke and Jeeves' method. This is a widely used direct search method which is concerned with the problem of minimising a function $f(x)$ of n real variables x_j under the assumption that no constraint is imposed on the values of the x_j.

Given an initial base point b_1 and step lengths h_j for the respective variables x_j, the procedure for an exploratory move about the point b_1 is as follows :

(i) evaluate $f(b_1+h_1e_1)$, where e_1 is the unit coordinate vector of variable 1. If the move from b_1 to $b_1+h_1e_1$ is a success, replace the base point b_1 by $b_1+h_1e_1$. If it is a failure, evaluate $f(b_1-h_1e_1)$. If this move is a success, replace b_1 by $b_1-h_1e_1$. If it is another failure, retain the original base point b_1.

(ii) repeat (i) for the variable x_2 by considering variations $\pm h_2e_2$ from the point which results from (i). Apply this procedure to each variable in turn, finally arriving at a new base point b_2 after (2n+1) function evaluations at most, including $f(b_1)$.

(iii) If $b_2=b_1$, halve each of the step lengths h_j and return to (i). The calculations terminate when the step lengths have been reduced to some prescribed level. If $b_2=b_1$, make a pattern move from b_2.

The procedure for a pattern move from b_2 is therefore as follows:

(iv) move from b_2 to $p_1=2b_2-b_1$ and continue with a new sequence of exploratory moves about p_1.

(v) if the lowest function value obtained during the pattern and exploratory moves of (iv) is less than $f(b_2)$, then a new base point b_3 has been reached. In this case, return to (iv) with all suffices increased by unity. Otherwise, abandon the pattern move from b_2 and continue with a new sequence of exploratory moves about b_2.

In their paper, Hooke and Jeeves (1961) report that their method has been used successfully to solve curve-fitting problems for which other methods failed.

In order to calculate the parameter estimates k_1 and k_2, data was collected over a period August - December 1982, using minutely averages from the Honeywell computer, which is the main computer for plant data at Sasol. One hundred and fifteen randomly assigned points were chosen for the database, and are shown in Table 6.2.1(a-d). For the purpose of the analysis, the terms 'actual' and 'maximum' refer to 'initial' and 'final' respectively, where an increase in the readings has taken place over a defined time period. The accuracy of each flow meter is based on random errors, where ε is random variable with mean zero and unknown constant variance σ^2. Each flow meter is 'zero-checked' on a daily basis for an off-set error and recalibrated as necessary.

In most practical situations, the parameter values are unknown. However, in this case, $k_1=k_2=0.5$. The computer program for the Hooke and Jeeves and jackknife methods was written in Basic on an IBM PC, and the results are shown in Table 6.2.2.

The results clearly show that the jackknife produces less biased estimates for k_1 and k_2, given that $k_1=k_2=0.5$ for the Hookes and Jeeves method. The jackknife also gives additional statistical information in terms of variance estimation and confidence limits for the parameters.

Table 6.2.1(a) : Database for Operating Variables

Actual Methane Gas flow (knm^3/h)	Actual Oxygen flow (knm^3/h)	Actual Steam flow (kg/h)	Ratio between gas constituents	Actual top p.d (kpA)	Actual bed p.d (kpA)	Maximum top p.d (kpA)	Maximum bed p.d (kpA)	Maximum Methane Gas flow (knm^3/h)
26.41	12.92	51.61	2.49	310.3	140.9	359.6	163.8	28.93
28.91	14.34	54.06	2.37	361.1	165.0	397.8	182.8	30.62
26.51	13.31	51.36	2.55	284.6	131.7	302.4	137.4	27.80
26.75	13.42	51.71	2.57	291.0	134.3	357.4	164.9	29.99
27.64	13.79	52.73	2.54	313.9	143.3	354.7	163.3	29.95
27.98	14.01	48.92	2.33	290.2	145.2	339.0	170.1	30.56
30.44	15.27	53.75	2.28	337.7	169.2	376.7	198.6	32.79
27.98	14.01	48.92	2.28	290.2	145.2	360.4	182.4	32.04
30.29	15.14	55.42	2.39	310.1	162.0	359.9	181.2	31.71
31.64	15.17	57.86	2.37	372.0	215.0	378.4	218.4	32.22
27.31	13.33	49.89	2.42	249.7	141.6	306.6	175.9	29.39
28.95	13.94	53.33	2.39	279.2	160.0	317.8	186.3	30.83
28.82	14.36	54.60	2.47	315.0	181.6	332.8	192.7	29.56
29.19	14.61	56.46	2.40	328.8	186.9	357.2	201.5	30.56
29.31	14.52	55.60	2.43	357.9	211.2	361.2	218.2	30.29
27.58	13.75	54.96	2.55	314.1	138.7	322.1	198.5	28.83
26.48	13.17	52.46	2.57	291.2	176.3	311.8	185.4	27.42
27.36	13.64	54.07	2.45	320.6	192.2	364.3	223.1	29.65
28.53	13.41	55.40	2.49	296.6	185.4	325.9	203.1	29.91
29.27	14.66	55.67	2.38	325.5	218.1	347.9	235.7	31.63
29.53	14.84	54.78	2.38	319.5	215.5	359.6	238.1	31.00
24.36	12.09	48.45	2.60	444.9	119.2	458.0	127.5	25.02
25.30	12.78	50.81	2.36	442.5	149.2	420.7	152.6	27.30
26.46	12.99	54.47	2.54	427.2	111.2	429.2	127.1	28.48
28.43	13.87	52.88	2.48	415.9	125.0	438.9	132.5	29.33
26.35	12.33	47.04	2.38	405.0	100.5	456.3	115.1	27.71
26.53	13.09	53.53	2.51	275.0	221.6	329.5	256.6	29.68
23.02	11.89	47.19	2.46	441.6	257.5	446.2	267.9	24.19
29.20	14.74	54.96	2.38	322.5	217.8	359.6	238.1	31.00
27.66	13.80	53.30	2.35	330.2	188.2	362.0	209.2	29.82

Table 6.2.1(b) : Database for Operating Variables

Actual Methane Gas flow (knm^3/h)	Actual Oxygen flow (knm^3/h)	Actual Steam flow (kg/h)	Ratio between gas constituents	Actual top p.d (kpA)	Actual bed p.d (kpA)	Maximum top p.d (kpA)	Maximum bed p.d (kpA)	Maximum Methane Gas flow (knm^3/h)
29.43	14.83	54.21	2.38	341.8	213.2	356.2	221.3	30.6
25.38	13.41	51.13	2.50	364.8	269.1	342.7	300.8	27.9
26.71	14.05	53.04	2.51	329.2	286.2	347.4	316.5	27.7
27.43	14.19	52.79	2.43	326.3	301.1	334.7	308.4	28.4
26.21	13.65	53.08	2.50	308.6	284.0	318.0	298.7	27.4
26.15	13.89	51.09	2.39	315.7	286.0	343.5	312.3	28.0
25.79	13.36	51.03	2.58	321.0	271.0	329.6	281.0	26.9
26.38	13.68	52.65	2.51	330.1	260.8	343.3	273.2	27.2
28.77	14.47	52.14	2.35	269.6	261.8	278.4	270.9	29.6
26.48	13.69	46.33	2.17	409.1	257.1	428.3	283.3	29.0
26.48	13.69	46.33	2.36	409.1	257.1	424.3	271.1	27.6
28.43	14.49	50.11	2.20	273.6	248.1	300.0	276.9	30.4
30.36	15.50	50.93	2.21	33.0	276.9	311.6	283.3	31.0
28.34	14.47	51.76	2.23	261.0	255.1	279.8	275.1	29.9
28.27	13.78	48.68	2.11	333.6	244.0	379.3	275.7	30.9
27.86	14.57	53.72	2.47	359.0	105.0	387.0	114.9	29.8
30.38	15.82	55.88	2.34	400.2	119.2	445.0	128.0	32.0
26.68	13.88	49.74	2.50	385.9	415.0	413.0	428.5	27.2
26.97	14.51	50.08	2.38	421.8	4333.3	423.0	438.1	28.8
28.66	14.31	52.98	2.48	290.0	97.0	317.6	105.2	29.9
30.0	15.15	55.47	2.40	352.8	106.1	391.7	118.6	32.0
28.66	14.31	52.99	2.24	290.0	96.0	414.0	125.4	34.2
30.09	14.98	55.69	2.31	363.5	104.1	412.3	116.2	33.1
25.79	13.32	53.53	2.59	313.1	89.4	346.7	98.1	27.7
26.51	13.58	51.88	2.59	331.7	92.3	358.5	97.2	28.1
26.92	13.03	51.98	2.50	330.5	90.9	415.1	117.2	30.6
29.61	14.69	53.37	2.41	394.6	108.8	438.7	124.6	31.6
29.73	13.94	54.10	2.38	361.9	59.4	449.4	84.9	34.4
28.03	12.88	48.95	2.25	286.3	52.9	427.3	79.6	34.4
33.95	16.04	60.68	2.23	399.2	77.9	438.3	83.8	35.1

Table 6.2.1(c) : Database for Operating Variables

Actual Methane Gas flow (knm^3/h)	Actual Oxygen flow (knm^3/h)	Actual Steam flow (kg/h)	Ratio between gas constituents	Actual top p.d (kpA)	Actual bed p.d (kpA)	Maximum top p.d (kpA)	Maximum bed p.d (kpA)	Maximum Methane Gas flow (knm^3/h)
27.01	12.78	48.96	2.20	317.7	120.7	430.5	161.2	31.19
42.38	11.69	43.14	2.11	323.6	120.7	425.7	150.3	28.86
28.87	14.14	52.91	2.30	362.2	106.8	416.7	128.9	32.61
25.57	12.73	50.15	2.49	309.5	180.0	336.9	197.4	27.64
27.39	13.70	52.60	2.32	288.3	158.5	298.9	165.6	30.39
27.51	12.34	49.93	2.22	321.5	158.8	425.4	192.9	31.01
24.87	12.25	42.53	2.11	240.0	219.0	285.4	286.9	28.43
26.17	14.42	52.02	2.57	388.9	147.2	420.7	155.6	28.67
26.73	14.28	52.92	2.53	400.7	152.5	447.6	167.8	28.29
26.80	14.30	52.96	2.52	399.5	151.4	446.0	167.7	28.29
26.62	13.93	52.54	2.51	387.7	149.4	415.5	164.5	29.01
26.09	13.59	52.83	2.46	369.5	142.7	436.2	172.8	29.65
27.07	13.91	52.70	2.48	388.0	149.3	432.1	173.5	29.74
27.57	14.47	53.53	2.49	393.1	158.3	403.3	164.4	28.18
26.35	13.77	52.53	2.47	362.5	199.2	386.6	217.0	28.23
26.37	13.20	52.24	2.55	320.4	203.1	340.1	218.7	27.57
25.66	12.85	51.88	2.57	332.0	206.3	336.0	217.3	27.54
26.09	14.11	51.75	2.52	319.1	118.1	371.4	135.0	28.57
26.82	14.43	52.32	2.55	414.2	116.6	450.5	129.4	28.83
25.72	13.43	51.47	2.52	323.2	116.8	373.8	135.4	28.48
25.80	13.46	51.33	2.42	322.9	117.4	437.4	165.9	32.28
25.72	13.43	51.47	2.52	322.9	116.7	384.?	147.0	29.99
23.81	12.37	49.26	2.49	269.4	103.0	386.6	139.7	29.05
24.14	12.34	49.95	2.45	275.8	105.0	389.8	139.7	29.39
23.53	12.02	49.03	2.52	261.8	99.8	326.6	120.7	26.74
24.38	12.35	49.37	2.45	275.5	105.0	374.7	142.4	29.28
25.12	12.48	49.66	2.31	283.7	107.5	447.0	164.7	32.91
26.59	13.56	50.27	2.37	326.4	120.6	415.7	162.5	31.41
26.55	13.69	54.46	2.44	323.2	128.7	399.5	158.1	30.55
26.50	13.66	53.22	2.44	318.9	126.6	396.0	156.9	30.46

Table 6.2.1(d) : Database for Operating Variables

Actual Methane Gas flow (knm^3/h)	Actual Oxygen flow (knm^3/h)	Actual Steam flow (kg/h)	Ratio between gas constituents	Actual top p.d (kpA)	Actual bed p.d (kpA)	Maximum top p.d (kpA)	Maximum bed p.d (kpA)	Maximum Methane Gas flow (knm^3/h)
26.33	13.29	52.23	2.54	307.9	120.1	341.2	131.3	28.05
26.42	13.30	52.04	2.57	306.9	120.8	341.1	131.7	28.03
26.66	13.68	51.56	2.29	321.6	98.3	440.8	129.9	33.72
30.20	15.54	55.39	2.11	388.8	120.0	459.2	128.7	35.01
29.91	15.55	55.30	2.12	384.9	119.3	455.1	128.1	35.07
25.89	13.58	49.50	2.21	326.6	95.4	460.1	133.9	34.12
25.92	13.58	49.83	2.43	326.4	97.2	388.8	120.0	30.20
25.89	13.59	49.73	2.46	324.9	97.0	386.4	119.6	29.83
26.09	13.61	49.78	2.36	325.5	97.3	419.4	131.6	31.83
26.99	13.75	50.46	2.34	291.9	122.6	375.4	160.5	32.20
29.39	14.67	54.59	2.45	343.9	142.9	373.2	157.3	30.27
28.95	13.50	48.63	2.36	282.5	122.2	358.7	158.9	31.83
28.05	13.91	53.67	2.36	324.9	135.9	358.5	160.4	31.36
26.61	13.50	49.68	2.30	313.9	128.3	365.6	153.5	30.06
31.87	15.78	56.30	2.27	391.7	164.1	404.0	174.5	33.81
25.56	13.86	48.40	2.40	324.3	100.0	352.1	111.2	28.90
24.80	12.86	50.09	2.22	255.6	113.9	327.0	142.2	30.31
26.78	13.42	48.66	2.45	270.9	128.0	317.8	148.1	29.08
28.41	14.93	53.31	2.44	312.7	159.8	368.8	191.4	31.24
29.45	15.65	55.40	2.37	399.0	175.2	429.9	188.1	31.89
24.85	13.05	49.03	2.45	303.0	140.8	338.6	154.2	27.81
27.20	14.16	52.39	2.45	350.7	162.2	375.2	170.2	29.31
25.78	14.61	54.53	2.59	401.0	241.3	402.3	236.1	27.93
27.00	13.54	51.35	2.48	322.1	87.6	361.7	103.7	30.44
25.06	13.28	51.04	2.28	296.4	343.5	312.0	392.8	27.35

Table 6.2.2 : Parameter Estimation using the Hooke and Jeeves and Jackknife methods

		Hooke and Jeeves method	Jackknife method
Parameter Estimates	k_1	0.44	0.48
	k_2	0.44	0.47
Variance Estimates	k_1	-	2.42×10^{-4}
	k_2	-	1.95×10^{-4}
Confidence Limits (95% level)	k_1	-	0.48 ± 0.03
	k_2	-	0.47 ± 0.03

The parameter estimates in Table 6.2.2, for each method, were then used on a second independent database (see Table 6.2.3), to predict the 'maximum' methane gas flow. The difference between the predicted and actual values were then calculated in terms of 'sum of squares', and the results are shown in Table 6.2.4. In both cases, either using the estimates for k_1 or k_2, the jackknife method gives the better results in terms of residual error.

The second case study involves an investigation into the poor mass balances which were being attained at the Chemical Work-up unit (C.W.U.) at Sasol Three. A mass balance is a balance carried out across a unit such that the mass weight of the input flows should equate with the mass weight of the output flows. The mass balances were calculated on a daily basis and ideally, 97.5% of the mass flow into the unit should equal the mass flow out of the unit. In this case, there was an assumed 2.5% mass loss during the actual process which took place at the unit. For the C.W.U., the mass balance was based on the quantity of non-acid chemicals (NACS) in the reaction water flowing in and out of the unit.

There were several factors which were believed to contribute to the poor results ($\pm$ 20% error), primarily concerning suspected errors in the flow meter readings and laboratory analysis errors in the recorded composition of the flows. An examination of the November and December 1985 daily data suggested that the mass balance error might be related to the recorded % NACs in the reaction water into the unit (see Table 6.2.5). The input/output function i.e. 0.975I-O, was plotted against % NACs in the reaction water as is shown in Figure 6.2.2.

Using the database from Table 6.2.5, a least squares linear regression analysis was carried out on an IBM PC using the NWASTATPAK software package. Four different models were postulated, namely:

$$Y = A + BX + \varepsilon \qquad (6.2.4)$$

$$Y = Ae^{BX} + \varepsilon \qquad (6.2.5)$$

$$Y = A + B\log X + \varepsilon \qquad (6.2.6)$$

$$Y = AX^{B} + \varepsilon \qquad (6.2.7)$$

Table 6.2.3 : Database for Operating Variables

Actual Methane Gas flow (knm^3/h)	Actual Oxygen flow (knm^3/h)	Actual Steam flow (kg/h)	Ratio between gas constituents	Actual top p.d (kpA)	Actual bed p.d (kpA)	Maximum top p.d (kpA)	Maximum bed p.d (kpA)	Maximum Methane Gas flow (knm^3/h)
26.03	13.92	52.45	2.59	385.5	149.6	429.2	165.6	28.05
24.65	12.55	50.53	2.51	343.9	134.8	440.2	166.8	30.01
24.31	12.51	51.03	2.56	314.5	141.5	402.7	179.5	29.98
26.10	13.45	50.67	2.35	377.4	146.1	450.7	177.1	29.86
24.80	12.91	50.59	2.48	371.6	143.7	452.8	182.7	29.03
26.46	13.38	51.63	2.40	386.5	154.2	448.1	182.7	30.32
24.90	13.01	49.35	2.47	323.1	137.4	372.2	158.2	28.64
24.23	12.68	49.33	2.59	316.8	133.4	379.9	162.4	27.51
27.51	14.31	53.98	2.45	377.6	161.4	417.2	178.8	29.52
27.9[illegible]	14.28	53.94	2.29	338.6	156.9	394.8	182.6	33.49
26.98	15.14	54.35	2.50	401.4	194.1	450.2	216.5	30.29
24.32	13.01	50.13	2.38	341.0	160.5	440.8	209.6	30.12
26.39	14.17	51.97	2.51	404.2	188.6	449.1	209.9	28.16
26.63	14.01	52.96	2.50	362.2	178.4	407.0	200.4	28.57
24.70	13.35	49.66	2.47	425.7	196.1	452.7	214.8	28.21
24.70	13.08	48.77	2.50	306.7	159.7	389.8	200.0	29.01
25.56	13.93	51.88	2.51	362.0	172.9	431.6	209.3	28.36
24.34	13.65	50.80	2.55	358.1	171.8	416.6	196.4	28.06
23.66	13.53	51.02	2.63	363.2	187.1	400.1	201.3	26.78
27.01	14.43	52.93	2.48	382.3	224.9	427.7	253.4	29.32
24.26	13.3	50.78	2.59	321.4	182.3	384.9	224.0	27.57
24.74	14.03	51.06	2.63	368.0	222.3	430.3	260.6	27.45
26.83	14.44	52.92	2.36	289.7	111.9	357.3	133.8	31.85
26.60	14.30	51.56	2.41	279.4	105.1	295.5	113.1	29.29
25.86	13.79	53.66	2.48	330.7	93.5	347.6	107.1	29.40
29.83	15.26	55.96	2.39	390.2	106.1	438.5	121.7	32.46
23.58	11.84	46.82	2.45	264.0	81.6	440.4	123.3	30.75
27.30	13.69	48.76	1.87	352.3	102.5	455.4	135.2	35.30
26.35	13.01	53.46	2.49	351.2	131.5	414.9	156.8	30.83
27.16	14.05	52.71	2.47	360.0	139.1	442.1	171.9	30.04

Table 6.2.4 : Comparison of Residual Errors

		Hooke and Jeeves method	Jackknife method
Total sum of Squares	k_1	14.19	13.98
	k_2	14.18	13.96
Mean sum of Squares	k_1	4.73×10^{-2}	4.66×10^{-1}
	k_2	4.72×10^{-1}	4.65×10^{-1}

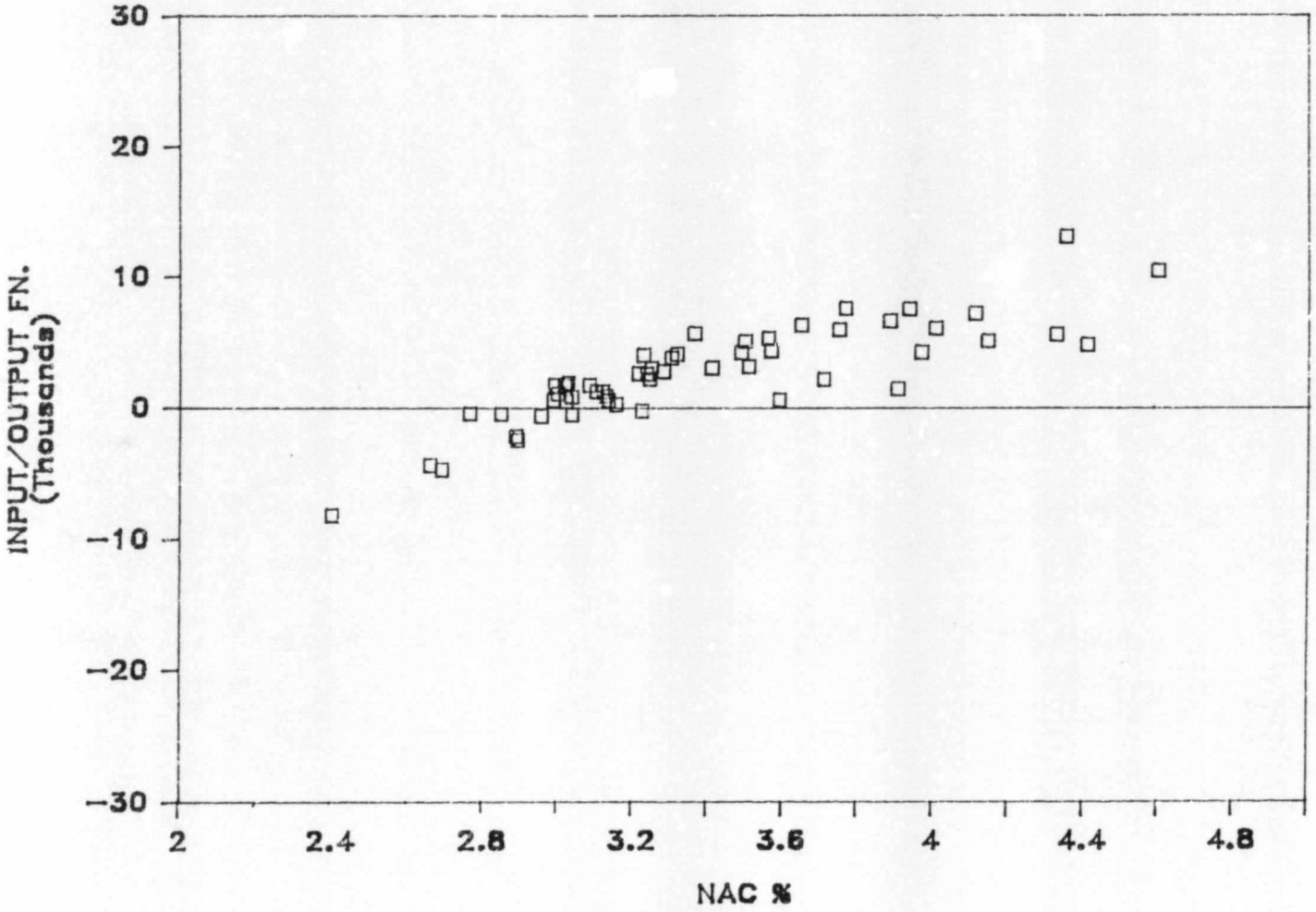
Table 6.2.5 : Mass Balance Database

% NAC in I	Mass Balance Error (0.975I-0)	% NAC in I	Mass Balance Error (0.975I-0)
3.59	772.60	2.96	-503.65
3.91	1583.38	2.89	-2327.87
3.97	43.75.53	3.02	645.47
4.14	5224.75	3.04	969.93
3.22	-113.82	2.99	748.60
2.69	-4562.77	3.94	7645.03
3.02	1811.93	3.56	5386.38
3.65	6422.53	3.75	6051.67
3.25	3160.15	3.51	3276.38
3.49	4361.97	3.16	361.35
4.35	13239.85	3.32	4220.88
3.12	1339.77	3.24	2331.40
3.49	5197.63	3.08	1855.70
3.03	2022.35	3.04	-413.95
2.99	1859.80	3.28	2891.38
3.57	4479.95	3.24	2710.67
3.13	998.57	3.13	695.90
3.37	5790.67	3.23	4129.78
4.41	4917.15	3.10	1339.18
2.41	-8029.95	3.30	3928.95
2.89	-2006.27	3.21	2730.58
2.66	-4107,45	3.41	3153.48
3.00	1201.13	3.89	6747.00
2.85	-348.55	3.77	7696.90
2.76	-312.05	4.01	6172.00
3.71	2283.20	4.33	5705.97
4.11	7306.03		

I - input flow

O - output flow

Figure 6.2.2 : Mass Balance Graph



where

Y = input/output function 0.975I-O

X = % NACs in reaction water

and ε is a random variable with mean zero and constant variance σ^2, unknown and independent of X.

In terms of the regression statistics, namely the coefficient of determination R^2 and the F-test, model (6.2.6) gave the best least squares estimates. The equation obtained was as follows :

$$\hat{Y} = -23939.4 + 22038.4 \log X \qquad (6.2.8)$$

The jackknife method was then applied in the normal manner to the statistics in (6.2.8) using the original data, to obtain the following equation:

$$\hat{Y} = -24075.1 + 22135.2 \log X \qquad (6.2.9)$$

Equations (6.2.8) and (6.2.9) suggest that a corrected % NAC figure of 2.96 and 2.97 respectively, would produce an exact mass balance i.e. 0.975I=O.

A second independent database consisting of the first three weeks of January 1986 daily data was corrected using the least squares and jackknife estimates in (6.2.8) and (6.2.9) respectively. (see Table 6.2.6). Apart from improving on the mass balance readings in both cases, the jackknife produced better results than the least squares method, in terms of the residual statistics.

The third case study involves using the jackknife method in principal component analysis. Beran and Srivastava (1984) considered bootstrap confidence regions and tests for eigenvalues and eigenvectors in both simple and multiple eigenvalue situations.

Table 6.2.6 : Comparison between Actual and Corrected Mass Balance Errors

Actual Mass Balance Error	'Corrected' Mass Balance Error	
	Least Squares Method	Jackknife Method
8788.03	-286.8	-280.0
5976.80	-519.1	-525.7
2865.27	-193.2	-212.0
5075.22	-1999.9	-2016.9
12620.50	-3092.3	-3031.0
4932.88	-26.1	-35.1
9236.43	-3465.8	-3471.0
8616.03	-1133.0	-1130.7
4692.55	-378.6	-390.2
7505.35	-652.6	-653.1
545.30	2966.4	2951.3
3638.42	1704.7	1697.6
8456.03	-415.4	-410.7
2906.60	730.7	715.2
2460.38	1503.1	1490.0
8196.97	-134.3	-129.5
15783.75	-2844.0	-2817.7
-7103.90	8352.3	8327.2
1918.18	3272.9	3265.1
4555.70	2061.2	2059.7
4007.00	3284.3	3285.8
Sum of Squared Errors	**146 101 983.6**	**145 358 423.8**
Mean Absolute Deviation	1857.9	1854.6

For planning purposes, it was very important to develop an equation which would accurately predict the top pressure difference (p.d.) across a catalyst bed in a gas reformer, as described in the first case study, for the first few weeks in a reformer's run when the catalyst was relatively undamaged. In this case, the top p.d. variable Y was dependent on five different gas flows, X_1, X_2, X_3, X_4 and X_5. Due to the operations in the unit, these flows were not independent variables and were adjusted in fixed ratios between each other from time to time. A database corresponding to these operating variables is shown in Tables 6.2.7 (a-d). One hundred randomly assigned data points, corresponding to minutely readings from the Honeywell computer, for the month of August 1985, were obtained for this purpose. The eigenvalues and eigenvectors of the correlation matrix obtained by a principal component analysis of the database is shown in Table 6.2.8. The eigenvalues and coefficient statistics were then jackknifed by deleting successive rows of the database and performing a principal component analysis in the remaining database, to obtain the corresponding jackknife pseudovalues. The jackknife estimates of the eigenvalues and coefficients were then calculated in the normal manner, together with the 95% confidence limits for the parameter values (see Table 6.2.9).

Estimates of the eigenvalues indicate that the first eigenvector accounts for $\pm$ 66% of the variation in both cases. The eigenvector is given by:

$$\hat{\tau}_1 = 0.53265X_1 - 0.34863X_2 - 0.23176X_3 - 0.73534X_4 + 0.01731X_5 \qquad (6.2.10)$$

or

$$\hat{\tau}_1 = 0.53271X_1 - 0.37735X_2 - 0.23454X_4 - 0.73178X_4 + 0.01974X_5 \qquad (6.2.11)$$

for the standard and jackknife methods respectively.

Table 6.2.7(a) : Database for Operating Variables

Operating Variables					
X_1	X_2	X_3	X_4	X_5	Y
30.40	54.42	6.378	15.12	19.57	367.3
30.07	54.31	6.369	15.29	19.51	369.6
30.54	54.64	6.432	15.22	19.59	373.5
30.98	54.36	6.617	15.63	19.47	380.1
31.59	55.67	6.807	15.82	19.47	393.6
31.63	55.71	6.659	15.68	19.51	389.7
32.61	56.50	7.392	16.18	19.41	406.9
32.53	56.84	7.341	16.41	19.47	411.2
32.77	57.59	7.463	16.44	20.03	414.6
32.79	57.26	7.422	16.50	20.07	414.6
32.95	57.07	7.355	16.59	20.02	418.7
35.15	57.45	7.604	14.90	20.03	420.6
33.01	55.77	7.453	15.09	19.90	380.0
33.15	55.96	7.422	15.13	20.02	385.1
32.88	56.49	7.448	15.19	20.03	387.4
32.80	56.10	7.261	15.44	19.94	385.9
33.08	56.57	7.400	15.68	19.97	390.3
32.56	55.97	7.292	15.72	20.01	390.3
32.54	56.11	7.423	15.82	19.96	395.6
32.68	55.86	7.430	16.04	19.98	394.8
32.49	56.03	7.359	16.25	19.97	399.5
32.89	56.41	7.499	16.66	20.02	401.6
32.56	56.69	7.467	16.44	19.97	401.4
32.84	56.26	7.433	16.71	19.96	400.3
32.90	56.36	7.305	16.57	20.05	401.4

Table 6.2.7(b) : Database for Operating Variables

Operating Variables					
X_1	X_2	X_3	X_4	X_5	Y
32.63	56.63	7.279	16.35	20.02	401.0
32.60	56.76	7.443	16.50	20.02	402.6
33.01	56.29	7.292	16.58	19.95	406.2
32.64	57.39	7.871	16.41	20.09	409.3
32.61	57.94	7.932	16.57	19.87	412.1
31.36	58.32	7.683	15.75	20.07	396.0
30.35	55.91	7.333	15.27	20.08	371.9
27.95	50.44	6.055	13.28	19.98	304.3
21.87	49.76	5.419	10.71	20.12	252.0
25.04	48.80	5.694	11.62	19.87	259.9
29.40	57.41	7.071	14.79	20.09	357.2
30.52	53.87	7.422	15.40	19.97	366.7
29.51	51.40	6.755	14.81	19.03	339.2
31.84	55.18	7.419	16.00	20.01	388.9
31.76	53.66	7.130	15.73	19.90	381.9
32.39	54.12	7.459	15.86	20.02	391.3
32.87	54.69	7.590	16.23	19.94	401.5
32.77	54.98	7.592	16.30	19.98	402.5
32.91	54.79	7.469	16.31	19.97	398.6
33.01	54.42	7.336	15.76	20.04	388.3
33.01	55.24	7.473	15.92	20.12	399.6
33.02	55.40	7.605	16.04	19.96	401.2
33.19	55.18	7.605	16.19	20.04	403.1
32.78	55.05	7.491	16.15	19.98	401.0
33.49	56.05	7.671	16.58	19.97	412.5

Table 6.2.7(c) : Database for Operating Variables

Operating Variables					
X_1	X_2	X_3	X_4	X_5	Y
33.00	57.04	7.932	16.66	20.51	421.3
30.74	55.53	7.435	15.55	20.47	384.9
29.33	55.16	7.212	14.85	20.44	359.7
28.09	50.71	6.733	13.95	20.54	329.4
28.69	51.62	6.756	14.40	20.53	343.7
28.27	51.51	6.881	14.36	20.49	339.5
28.82	52.42	7.013	14.57	20.52	348.6
28.77	53.18	6.990	14.15	20.48	346.1
28.70	52.44	6.973	14.14	19.98	343.5
28.58	52.35	6.712	14.10	20.00	345.0
28.74	52.38	6.891	14.44	20.02	347.4
28.59	52.26	6.710	14.20	19.99	348.2
27.62	52.11	6.671	13.76	20.00	332.9
25.75	49.52	6.052	12.88	19.97	295.9
24.77	48.75	5.750	12.12	19.98	285.0
24.59	48.85	5.783	12.25	19.98	281.8
24.61	49.36	5.717	12.70	19.93	238.8
24.57	50.63	5.708	12.69	20.00	298.5
24.58	51.12	5.923	12.72	20.05	299.6
24.53	50.87	5.720	12.84	20.04	300.0
22.43	49.01	5.272	12.42	20.14	273.4
22.30	48.58	5.242	11.40	20.07	264.5
24.49	50.84	5.734	12.86	19.99	300.0
24.48	50.79	5.955	12.84	20.05	300.0
24.51	49.60	5.722	12.96	20.01	299.2

Table 5.2.7(d) : Database for Operating Variables

Operating Variables					
X_1	X_2	X_3	X_4	X_5	Y
24.47	49.59	5.911	12.91	19.99	301.5
25.01	50.29	6.071	13.09	20.03	307.7
26.16	51.01	6.194	13.54	20.03	321.9
28.63	53.79	6.900	14.77	19.97	364.0
28.65	53.89	6.988	14.48	19.99	358.3
34.79	57.19	5.455	15.83	20.08	396.1
35.20	58.52	5.916	16.82	20.00	415.2
35.03	58.31	6.088	16.87	20.14	408.9
35.23	59.23	6.002	17.14	20.05	415.9
35.27	55.68	6.043	17.35	20.17	422.4
34.85	56.94	6.688	16.96	20.55	410.0
34.75	56.96	6.363	17.16	20.34	408.2
34.75	58.49	6.680	17.38	20.38	411.0
34.97	58.57	6.649	17.21	20.40	416.0
34.92	57.69	6.697	17.24	20.49	416.8
34.91	57.44	6.406	17.25	20.51	418.3
347.0	58.35	6.672	17.26	20.45	414.9
34.67	57.15	6.402	17.28	20.43	415.3
35.11	57.48	6.485	17.37	20.36	417.2
35.16	57.88	6.617	17.22	20.53	410.5
35.13	57.93	6.698	17.26	20.55	408.9
35.11	57.38	5.686	17.34	20.22	407.6
35.08	57.59	6.703	17.34	20.45	408.7
32.73	56.35	6.014	16.18	20.63	273.7
34.05	57.38	6.502	17.03	20.43	394.5

Table 6.2.8 : Eigenvalues of the Correlation Matrix of the Data

Eigenvalues	Coefficients				
	X_1	X_2	X_3	X_4	X_5
3.31226	0.53265	-0.34863	-0.23176	-0.73534	0.01731
0.08569	0.52554	0.81644	-0.22852	0.06621	0.02508
0.52150	0.37862	0.00343	0.88998	-0.00125	-0.38868
0.03979	0.53185	-0.46029	0.22261	0.67436	0.03076
1.04076	0.11778	0.00114	0.37276	-0.01104	0.92036

Table 6.2.9 : Jackknife Estimates of the Eigenvalues and Coefficients

Eigenvalues	95% Confidence Limits	Coefficients				
		X_1	X_2	X_3	X_4	X_5
3.30752	(3.51910,3.09595)	0.53271	-0.37735	-0.23454	-0.73178	0.01974
0.08671	(0.12450,0.04892)	0.52541	0.82784	-0.23274	0.04211	0.02983
0.52296	(0.70210,0.343829	0.38432	0.00764	0.84976	-0.00450	-0.39231
0.04119	(0.06594,0.01643)	0.53193	-0.44612	-0.22184	0.69760	0.03308
1.02036	(1.11306,0.92766)	0.11711	0.00067	0.37802	-0.01156	0.93082

Using the eigenvector as the independent variable X, linear regression equations were obtained for the dependent variable Y, using the least squares regression computer package NWASTATPAK

i.e.

$\hat{Y}$=-218.58957+10.13573X - standard method (6.2.12)

or

$\hat{Y}$=-218.53407+10.13104X - jackknife method (6.2.13)

The sum of squared errors between the actual and predicted values and the corresponding mean absolute deviation, was 6664.04 and 6.49 respectively, for the standard method. For the jackknife method, the corresponding statistics were 6638.88 and 6.47. These statistics are clearly better in the case of the jackknife method. The jackknife method also provided additional information in terms of establishing confidence limits for the eigenvalues.

The final case study involves a comparison of the quality of the coal which is transported from the mines to Sasol Two and Sasol Three. Of particular interest to Management is whether both plants receive coal of a similar quality on a monthly basis. The quality of the coal is defined in terms of particle size distribution with a range of 0.0 to 106.0 m.m. The homogeneity of the coal influences the oil from coal conversion process i.e. the greater the homogeneity of the coal, the higher the conversion. The variance estimate is the usual statistic calculated to assess the homogeneity of the coal.

Representative coal samples were taken, usually on a twice daily basis, for both plants during September 1987, and the results of the analyses are shown in Tables 6.2.10(a-d), 6.2.11 and Figure 6.2.3 respectively. Comparing the variances of the two populations by means of the F-test is clearly inappropriate in this case due to the non-normality of the data as depicted in Figure 6.2.3. Miller (1968), derived a test on the equality of the variances of two populations based on the jackknife method. A brief description of the test is as follows:

Table 6.2.10(a) : Percentage of Coal per Interval for each Sample - Sasol Two

COAL SIZE (M.M)							
0.00-3.35	3.35-6.70	6.70-9.50	9.50-13.20	13.20-26.50	26.50-53.00	53.00-75.00	75.00-106.00
24.6	13.6	10.7	9.4	21.4	13.8	4.0	2.5
25.8	12.0	10.4	8.8	19.1	13.1	6.2	4.6
34.9	15.6	11.0	8.0	18.2	10.6	1.2	0.5
24.9	13.5	9.9	8.5	19.8	15.7	5.8	1.9
23.1	9.3	7.6	7.0	16.8	22.0	11.0	3.2
25.3	10.6	9.4	8.6	23.8	16.7	4.4	1.2
21.5	12.2	9.9	8.6	21.3	18.4	5.2	2.9
25.7	12 2	10.0	8.8	21.5	14.6	5.2	2.0
27.0	12.1	9.9	8.7	21.1	15.6	4.3	1.3
24.5	12.5	10.3	8.4	20.3	16.7	3.9	3.4
18.7	9.0	8.4	7.6	21.2	19.7	8.7	6.7
22.2	11.4	9.8	8.4	22.2	17.7	4.5	3.8
25.5	11.9	9.4	7.7	18.9	16.0	6.1	4.5
17.7	7.4	7.0	8.4	29.5	22.0	7.2	0.8
25.2	12.1	9.7	8.4	19.5	15.2	6.5	3.4
23.1	11.7	10.2	9.2	20.4	16.6	3.6	5.2
29.1	11.9	9.3	7.9	17.9	14.0	4.9	5.0
22.8	10.5	8.9	8.3	22.2	17.6	7.3	2.4
30.3	14.7	9.9	8.3	18.2	13.7	3.1	1.8
24.6	12.5	10.8	9.7	19.6	16.9	4.6	1.3
28.4	14.5	10.7	8.0	15.7	12.5	5.9	4.3
18.9	7.8	6.8	6.7	23.8	21.0	10.6	4.4
24.2	10.2	8.6	7.8	19.3	18.2	8.0	3.7
25.7	11.2	9.4	7.8	19.8	15.5	6.0	4.6
22.5	10.2	9.1	8.0	22.1	17.7	5.8	4.6
19.8	8.2	8.0	6.7	19.7	21.3	10.2	6.1
22.1	9.3	8.7	6.9	19.0	17.6	8.9	7.5
22.2	11.4	8.8	8.3	20.6	21.1	5.5	2.1
19.9	10.0	8.6	7.3	22.4	21.3	7.6	2.9
31.9	13.7	8.8	9.0	16.8	12.5	4.6	2.7
31.8	13.4	9.6	7.9	16.4	12.7	3.3	2.9

Table 6.7.10(b) : Percentage of Coal per Interval for each Sample - Sasol Two

COAL SIZE (MM)							
0.00-3.35	3.35-6.70	6.70-9.50	9.50-13.20	13.20-26.50	26.50-53.00	53.00-75.00	75.00-106.00
28.6	13.5	11.0	8.6	21.2	13.5	2.4	1.2
30.8	15.9	11.8	8.5	18.5	10.1	2.0	2.4
24.9	11.1	9.5	9.1	22.3	15.6	5.1	2.4
24.3	10.0	9.1	8.4	23.2	16.5	5.9	2.6
26.1	14.8	10.9	9.2	19.9	13.3	4.2	1.6
26.6	14.4	9.9	7.9	18.0	13.7	6.2	3.3
23.3	11.9	9.6	8.2	20.3	18.2	5.6	2.9
25.4	11.8	10.0	8.4	21.4	15.6	4.9	2.5
35.1	16.6	12.6	9.7	17.9	6.6	1.5	0.7
24.1	14.8	13.2	11.4	21.5	11.7	2.6	2.4
30.5	15.8	11.7	9.3	18.5	9.3	2.5	3.3
28.3	13.7	10.9	8.5	18.6	12.7	4.0	2.5
21.4	9.8	9.7	9.4	25.7	17.1	4.4	2.2
25.4	12.0	9.9	9.0	21.3	14.6	5.6	0.7
26.6	13.2	11.9	9.9	21.5	12.4	3.8	2.6
26.8	12.9	10.7	9.0	20.6	12.3	5.1	1.3
31.0	14.0	11.4	9.6	20.1	10.8	1.8	1.3
28.9	15.0	10.7	9.1	19.6	11.1	4.3	0.6
16.2	7.0	6.4	7.6	24.1	25.6	12.5	1.5
34.1	15.0	10.9	8.8	16.4	9.1	4.2	5.3
17.3	7.9	7.3	7.2	21.6	25.4	8.0	1.7
22.0	9.9	9.3	7.9	21.8	20.7	6.7	1.1
11.6	5.5	6.5	9.4	31.2	26.0	8.7	0.6
21.6	9.7	9.0	7.8	24.2	22.2	4.9	1.5
22.9	9.9	8.8	8.7	23.9	18.9	5.4	2.2
20.8	9.8	8.6	8.1	22.0	20.3	8.2	2.7
18.2	9.9	8.7	8.7	23.4	21.8	6.6	0.0

Table 6.2.10(c) : Percentage of Coal per Interval for each Sample - Sasol Three

COAL SIZE (M.M)							
0.00-3.35	3.35-6.70	6.70-9.50	9.50-13.20	13.20-26.50	26.50-53.00	53.00-75.00	75.00-106.00
16.4	9.2	8.2	7.9	22.7	22.8	11.1	1.7
17.7	6.9	6.8	7.4	22.6	24.6	8.2	5.8
17.4	10.3	9.5	9.1	24.1	20.2	4.9	4.5
23.5	10.3	9.0	8.5	22.2	17.2	5.8	3.5
27.8	12.3	10.9	9.0	20.8	13.3	5.3	0.6
28.3	12.7	10.5	8.9	20.8	14.0	3.9	0.9
22.2	11.6	10.1	8.9	22.6	16.9	5.7	2.0
31.0	16.8	13.5	10.5	19.0	6.6	2.6	0.0
21.5	7.8	6.7	6.2	17.8	24.0	9.2	6.8
19.8	10.7	9.2	8.3	22.5	19.4	6.7	3.4
26.7	15.4	12.6	10.2	19.2	10.0	3.9	2.0
21.8	12.5	11.0	9.9	21.1	16.4	5.0	2.3
26.4	14.8	11.6	9.1	19.4	12.4	3.6	2.7
19.5	10.5	10.5	9.5	23.9	17.4	7.2	1.5
25.0	13.2	10.3	9.0	19.8	14.1	6.9	1.7
20.7	9.8	8.6	8.4	22.9	19.5	7.1	3.0
21.3	9.7	8.3	7.7	19.4	19.2	8.1	6.3
25.1	11.8	9.4	8.9	21.3	16.6	4.4	2.5
24.4	11.6	10.7	10.4	27.0	14.0	1.9	0.0
28.5	12.8	10.0	8.5	19.3	15.3	5.0	0.6
22.0	9.1	8.3	6.9	19.2	16.4	8.7	9.4
27.6	11.5	9.7	9.0	21.1	15.6	4.6	0.9
22.0	10.7	9.4	9.0	23.0	19.8	5.5	0.6
24.5	11.2	9.8	9.2	21.6	17.0	5.7	1.0
22.7	11.3	9.8	8.0	21.8	16.0	8.0	2.4
19.6	8.8	7.6	6.8	19.2	22.8	9.6	5.6
24.9	12.2	11.5	10.6	23.5	14.5	2.5	0.3
26.7	12.2	9.4	8.3	18.9	16.3	5.7	2.5
26.3	12.6	10.3	8.9	22.1	16.1	2.8	0.9
30.6	10.3	9.5	7.8	19.0	16.3	5.6	0.9
21.8	11.0	8.9	8.5	23.2	20.0	4.7	1.9

Table 6.2.10(d) : Percentage of Coal per Interval for each Sample - Sasol Three

COAL SIZE (M.M)							
0.00-3.35	3.35-6.70	6.70-9.50	9.50-13.20	13.20-26.50	26.50-53.00	53.00-75.00	75.00-106.00
20.9	8.0	6.9	7.2	25.5	25.1	5.5	0.9
28.4	11.9	9.5	8.3	20.9	15.1	4.9	1.0
24.7	10.6	9.4	8.4	22.8	17.3	5.4	1.4
15.6	6.0	5.6	6.3	21.6	30.3	10.6	4.0
30.2	13.4	10.8	8.8	19.4	13.4	3.2	0.8
23.3	11.3	9.4	8.3	21.0	19.2	5.7	1.8
22.4	11.4	9.2	8.4	21.8	16.1	6.7	4.2
20.9	11.2	9.0	7.9	22.2	17.9	7.9	3.0
21.9	11.7	9.3	8.5	21.2	18.1	7.5	1.8
22.3	12.4	9.3	8.6	20.7	17.1	6.6	3.0
23.3	13.0	10.3	9.2	21.1	16.1	4.6	2.4
21.9	9.6	8.3	7.8	20.2	19.2	7.2	5.8
18.5	9.2	7.7	7.5	20.6	20.6	10.0	5.9
25.4	11.6	9.4	8.7	20.6	17.3	3.9	3.1
23.9	11.7	11.0	10.7	24.2	14.2	2.6	1.7
26.6	13.4	10.6	9.2	19.0	15.2	5.2	0.8
22.6	10.2	8.5	8.1	20.1	20.7	8.1	1.7
20.8	9.6	8.2	7.8	21.7	21.2	8.4	2.3
27.0	13.3	10.6	9.3	22.0	12.6	3.4	1.8
27.2	12.2	9.6	8.0	20.0	16.4	6.2	0.4
21.7	9.9	8.1	7.6	19.0	18.0	13.0	2.7
20.5	11.3	8.7	8.6	21.7	18.3	8.7	2.2
22.2	9.3	8.1	7.3	21.5	22.8	7.3	1.5
22.5	10.8	9.8	10.3	25.9	15.3	3.4	2.0
26.1	11.3	10.0	9.0	21.2	14.1	3.7	4.6
22.0	9.8	8.8	8.0	24.0	20.4	6.4	0.6
25.2	11.0	9.9	8.2	21.8	17.1	5.2	1.6
26.1	11.5	10.0	9.2	22.0	15.2	4.8	1.2
22.2	9.5	8.5	8.0	21.7	20.6	6.8	2.7
21.7	11.2	8.4	7.8	21.5	17.8	8.7	2.9

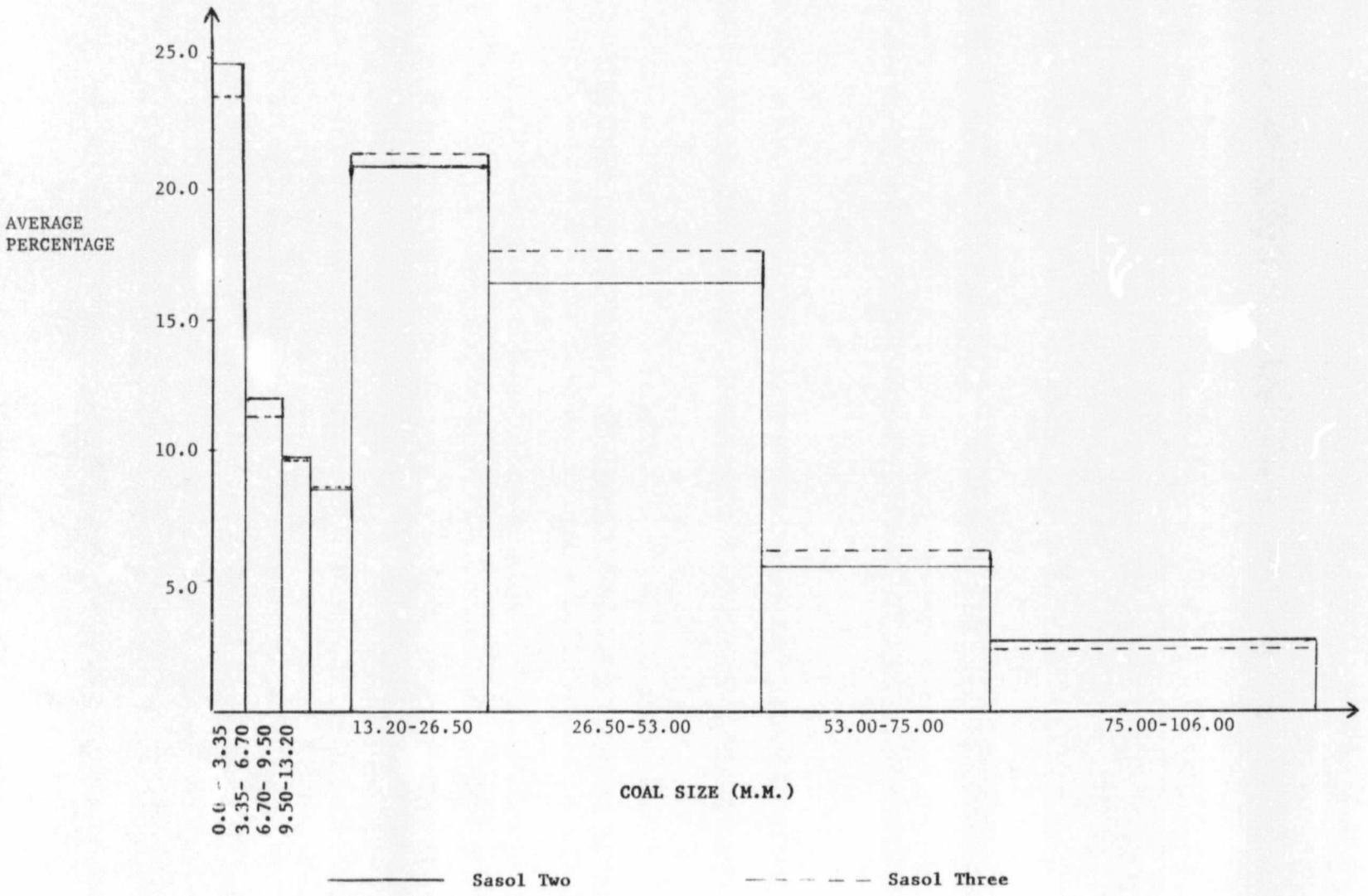
Table 6.2.11 : Average Coal Size Distribution for Sasol Two and Sasol Three

		AVERAGE PERCENTAGE	
		Sasol Two	Sasol Three
COAL SIZE (M.M)	0.00-3.35	24.8	23.5
	3.35-6.70	11.8	11.1
	6.70-9.50	9.6	9.4
	9.50-13.20	8.5	8.5
	13.20-26.50	20.8	21.4
	26.50-53.00	16.3	17.5
	53.00-75.00	5.5	6.1
	75.00-106.00	2.7	2.4

Number of coal samples = 58 - for Sasol Two

= 61 - for Sasol Three

Figure 6.2.3 : Average Coal Size Distribution for Sasol Two and Sasol Three



Consider the data to consist of $N=m+n$ observations $X_1,\ldots,X_m$, and $Y_1,\ldots,Y_n$ where it is assumed:

$$X_i=\sigma_1 e_i+u_1 \ , \ i=1,\ldots,m \tag{6.2.14}$$

$$Y_j=\sigma_2 e_{m+j}+u_2, \ j=1,\ldots,n \tag{6.2.15}$$

and

$e_1,\ldots,e_{m+n}$ are unobservable random variables.

The quantities u_1 and u_2 are the unknown medians of the X and Y populations respectively, and the parameter of interest is the unknown ratio of the parameters $v=\sigma_2/\sigma_1$. The N r.v.s e_i are identically independent and each e_i comes from the same continuous population having a median of zero and finite fourth moment.

The following steps are required to test the null hypothesis, $H_o: v^2=1$, namely

(i) Select positive integers k_1 and k_2 such that m/k_1 and n/k_2 are integers. Divide randomly the X and Y observations into m' and n' groups respectively, of size k_1 and k_2 respectively.

(ii) Set $d_1=m-k_1$ and $d_2=n-k_2$. For $i=1,\ldots,m'$, let $X_{i1},\ldots,X_{id_1}$ denote the d_1 observations obtained by deleting the i^{th} group of k_1 X observations. Similarly, for $j=1,\ldots,n'$, let $Y_{j1},\ldots,Y_{jd_2}$ denote the d_2 observations obtained by deleting the j^{th} group of k_2 Y observations.

(iii) Define $S_1,\ldots,S_{m'}$ by

$$S_i = \ln \frac{\sum_{s=1}^{d_1} (X_{is}-\bar{X}_i)^2}{d_1-1} \tag{6.2.16}$$

where

$$\bar{X}_i = \frac{\sum_{s=1}^{d_1} X_{is}}{d_1} \tag{6.2.17}$$

Similarly, define $T_1,\ldots,T_n$, by

$$T_j = \ln \frac{\sum_{t=1}^{d_2} (Y_{jt}-\bar{Y}_j)^2}{d_2-1} \qquad (6.2.18)$$

where

$$\bar{Y}_j = \frac{\sum_{t=1}^{d_2} Y_{jt}}{d_2} \qquad (6.2.19)$$

(iv) Next, define

$$S_o = \ln \frac{\sum_{i=1}^{m} (X_i-\bar{X})^2}{m-1} \qquad (6.2.20)$$

and

$$T_o = \ln \frac{\sum_{j=i}^{n} (X_j-\bar{Y})^2}{n-1} \qquad (6.2.21)$$

where

$$\bar{X} = \frac{\sum_{i=1}^{m} X_i}{m} \quad \text{and} \quad \bar{Y} = \frac{\sum_{j=1}^{n} Y_j}{n} \qquad (6.2.22)$$

(v) Compute the jackknife pseudovalues:

$$A_i = m'S_o - (m'-1)S_i \text{ , for } i=1,\ldots,m' \qquad (6.2.23)$$

and

$$B_j = n'T_o - (n'-1)T_j \text{ , for } j=1,\ldots,n' \qquad (6.2.24)$$

(vi) Set

$$\bar{A} = \frac{\sum_{i=1}^{m'} A_i}{m'} \quad \text{and} \quad \bar{B} = \frac{\sum_{j=1}^{n'} B_j}{n'} \tag{6.2.25}$$

and compute

$$V_1 = \frac{\sum_{i=1}^{m'} (A_i - \bar{A})^2}{m'(m'-1)} \quad \text{and} \quad V_2 = \frac{\sum_{j=1}^{n'} (B_j - \bar{B})^2}{n'(n'-1)} \tag{6.2.26}$$

(vii) Calculate the statistic

$$Q = \frac{\bar{B} - \bar{A}}{(V_1 + V_2)^{\frac{1}{2}}}$$

(viii) For a two sided test of H_o versus the alternative $v^2 \neq 1$, at the (approximate) α level of significance:

reject H_o , if $Q > Z_{(\alpha_2)}$ or $Q < -Z_{(\alpha_1)}$

accept H_o, if $-Z_{(\alpha_1)} < Q < Z_{(\alpha_2)}$

where $\alpha = \alpha_1 + \alpha_2$

When $k_1 = k_2$ and m' and n' are small and equal, the approximate α-level test above can be improved by replacing the normal quantiles $Z_{(\alpha_1)}$ and $Z_{(\alpha_2)}$ by Student t quantiles $t_{(m'+n'-2, \alpha_1)}$ and $t_{(m'+n'-2, \alpha_2)}$ respectively.

This test, based on the jackknife method to compare variances of two populations, was then used on the coal sampling data, shown in Table 6.2.10(a-d). For given values of m=58, n=61 and $k_1 = k_2 = 1$, values of the following statistics were obtained:

S_o = 6.074783

T_o = 6.064524

A = 6.075567

B = 6.065479

V_1 = 0.000979

V_2 = 0.0012469

The computed value of the test statistic Q=-0.213855. Since -1.96<-0.213855<1.96, the null hypothesis is accepted at the 5% level. i.e. there is no significant difference in the homogeneity of the coal between Sasol Two and Sasol Three. This final case study has shown an important application of the jackknife method in comparing variances of two populations, when the non-normality of the data is such that the standard parametric methods are inappropriate.

Author Angus Stuart Maxwell
Name of thesis A Comparison Of The Jackknife And Bootstrap Estimators In Linear Models With Reference To Production Models Used By Sasol. 1988

PUBLISHER:
University of the Witwatersrand, Johannesburg
©2013

LEGAL NOTICES:

Copyright Notice: All materials on the University of the Witwatersrand, Johannesburg Library website are protected by South African copyright law and may not be distributed, transmitted, displayed, or otherwise published in any format, without the prior written permission of the copyright owner.

Disclaimer and Terms of Use: Provided that you maintain all copyright and other notices contained therein, you may download material (one machine readable copy and one print copy per page) for your personal and/or educational non-commercial use only.

The University of the Witwatersrand, Johannesburg, is not responsible for any errors or omissions and excludes any and all liability for any errors in or omissions from the information on the Library website.