Appendix D

Comparison of Quality Control and Acceptance Tests

"From Implementation Manual for Quality Assurance, Appendix F, Copyright 1996, by the American Association of State Highway and Transportation Officials, Washington, D.C. Used by permission."

APPENDIX

COMPARISON OF QUALITY CONTROL AND ACCEPTANCE TESTS

Purpose

The purpose of this procedure is to provide a method of comparing two different data sets of multiple test results - say contractor QC test results and Agency acceptance or verification test results to determine if the material tested came from the same population. The statistical tests used to make the comparisons are called *Hypothesis Tests* and are described in the following paragraphs.

Analysis

To compare two populations that are assumed normally distributed, you may compare their means (averages) and their variabilities (standard deviations or variances). A different test is used for each of these properties. The *F-test* provides a method for comparing the variance (standard deviation squared) of two sets of data. Possible differences in means are assessed by a ***t-test***.

The F-test is based on the ratio of the variances of two sets of data. In this case, the F-test is based on the ratio of the variances of the QC test results, s_c^2, and the acceptance test results, s_a^2. The t-test compares sample means, and in this case, is based on the means of the QC test results, $\bar{X}_c$ and the acceptance test results, $\bar{X}_a$.

Hypothesis tests, i.e., the F-test and t-test, are conducted at a selected level of significance, α. The level of significance is the probability of incorrectly deciding the data sets are different when they actually come from the same population. The value of α is typically selected as either 0.05 or 0.01. The following analysis is based on an α of 0.01 so as to minimize the likelihood of incorrectly concluding that the test results are different when they are not.

For the analysis to be meaningful, all of the samples must be obtained in a random manner, the two sets of test results must have been sampled over the same time period, and the same sampling and testing procedures must have been used for both QC and acceptance tests. If it is determined that a significant difference is likely between either the mean or the variance, the source of the difference should be identified. Although it is beyond the scope of the analysis presented here, a computer program could be developed that could identify the existence of significant differences once the test results are input.

If the analysis indicates there is no reason to believe the results came from different populations, then the mean and variance (or standard deviation) could be determined from the combined set of test results to provide a better estimate of the populations parameters than would be obtained from either of the sets individually.

For information on how the Operating Characteristics curves for these tests can be developed, the reader is referred to statistics text books such as Reference 2.

Procedure

F-test for the Sample Variances

Since the values used in the t-test are dependent upon whether or not the variances are equal for the two sets of data, it is necessary to test the variances of the test results before the means. The intent is to determine whether the difference in the variability of the contractor's QC tests and that of the State's acceptance tests is larger than might be expected from chance if they came from the same population. In this case, it does not matter which variance is larger. After comparing the test results, one of the following will be concluded.

- The two sets of data have different variances because the difference between the two sets of test results is greater than is likely to occur from chance if their variances are actually equal.

- There is no reason to believe the variances are different because the difference is not so great as to be unlikely to have occurred from chance if the variances are actually equal.

First, compute the variance (the standard deviation squared) for the QC tests, s_c^2, and the acceptance tests, s_a^2. Next, compute F, where $F = s_c^2/s_a^2$ or $F = s_a^2/s_c^2$. *Always use the larger of the two variances in the numerator.* Now, choose α, the level of significance for the test. As mentioned previously, the recommended α is 0.01. Next, a critical F value is determined from Table 1 using the degrees of freedom associated with each set of test results. The degrees of freedom for each set of results is the number of test results in the set, less one. If the number of QC tests is n_c and the number of acceptance test is n_a, then the degrees of freedom associated with s_c^2 is $(n_c\text{-}1)$ and the degrees of freedom associated with s_a^2 is $(n_c\text{-}1)$. The values in Table 1 are tabulated to test if there is a difference (either larger or smaller) between two variance estimates. This is known as a two-sided or two-tailed test. Care must be taken when using other tables of the F distribution, since they are usually based on a one-tailed test, i.e., testing specifically whether one variance is larger than another.

Once the value for F_{crit} is determined from Table 1 (be sure that the appropriate degrees of freedom for the numerator and denominator are used when obtaining the value from Table 1), if $F \geq F_{crit}$, then decide that the two sets of tests have significantly different variabilities. If $F < F_{crit}$ then decide that there is no reason to believe that the variabilities are significantly different.

t-test for Sample Means

Once the variances have been tested and been assumed to be either equal or not equal, the means of the test results can be tested to determine whether they differ from one another or can be assumed equal. The desire is to determine whether it is reasonable to assume that the QC tests came from the same population as the acceptance tests. A t-test is used to compare the sample means. Two approaches for the t-test are necessary. If the sample variances are assumed equal, then the t-test is conducted based on the two samples using a *pooled* estimate for the variance and the *pooled* degrees of freedom. This approach is ***Case 1*** described below. If the sample variances are assumed to be different, then the t-test is conducted using the individual sample variances, the individual sample sizes, and the *effective* degrees of freedom (estimated from the sample variances and sample sizes). This approach is ***Case 2*** presented below.

In either of the two cases discussed in the previous paragraph, one of the following decisions is made:

- The two sets of data have different means because the difference in the sample means is greater than is likely to occur from chance if their means are actually equal.
- There is no reason to believe the means are different because the difference in the sample means is not so great as to be unlikely to have occurred from chance if the means are actually equal.

Case 1: Sample Variances Assumed to Be Equal

To conduct the t-test when the sample variances are assumed equal, equation 1 is used to calculate the t value from which the decision is reached.

$$t=\frac{|\bar{X}_c-\bar{X}_a|}{\sqrt{\frac{s_p^2}{n_c}+\frac{s_p^2}{n_a}}} \tag{1}$$

Where:

$\bar{X}_c$	=	mean of QC tests
$\bar{X}_a$	=	mean of acceptance tests
s_p^2	=	pooled estimate for the variance (described below)
n_c	=	number of QC tests
n_a	=	number of acceptance tests

The pooled variance, which is the weighted average, using the degrees of freedom for each sample as the weighting factor, is computed from the sample variances using equation 2.

$$S_p^2 = \frac{S_c^2(n_c-1)+S_a^2(n_a-1)}{n_c+n_a-2} \qquad (2)$$

Where:

s_p^2 = pooled estimate for the variance
n_c = number of QC tests
n_a = number of acceptance tests
s_c^2 = variance of the QC tests
s_a^2 = variance of the acceptance tests

Once the pooled variance is estimated, the value of t is computed using equation 1.

To determine the critical t value against which to compare the computed t value, it is necessary to select the level of significance, α. As discussed above, a value of $\alpha = 0.01$ is recommended. Next, determine the critical t value, t_{crit}, from Table 2 for the pooled degrees of freedom. The pooled degrees of freedom for the case where the sample variances are assumed equal is ($n_c + n_a$ -2). If $t \geq t_{crit}$, then decide that the two sets of tests have significantly different means. If $t < t_{crit}$ then decide that there is no reason to believe that the means are significantly different.

Case 2: Sample Variances Assumed to Be Not Equal

If the sample variances are not assumed to be equal, then the individual sample variances, rather than the pooled variance, are used to calculate t, and the degrees of freedom used are an estimated effective degrees of freedom rather than the pooled degrees of freedom.To conduct the t-test when the sample variances are assumed not equal, equation 3 is used to calculate the t value from which the decision is reached.

$$t = \frac{|\bar{X}_c - \bar{X}_a|}{\sqrt{\frac{s_c^2}{n_c} + \frac{s_a^2}{n_a}}} \qquad (3)$$

Where:

$\bar{X}_c$ = mean of QC tests
$\bar{X}_a$ = mean of acceptance tests

s_c^2	=	variance of the QC tests
s_a^2	=	variance of the acceptance tests
n_c	=	number of QC tests
n_a	=	number of acceptance tests

To determine the critical t value against which to compare the computed t value, it is necessary to select the level of significance, α. As discussed above, a value of $\alpha = 0.01$ is recommended. Next, determine the critical t value, t_{crit}, from Table 2 for the effective degrees of freedom. The effective degrees of freedom, f', for the case where the sample variances are assumed not equal is determined from equation 4.

$$f' = \frac{\left(\frac{s_c^2}{n_c} + \frac{s_a^2}{n_a}\right)^2}{\left(\frac{\left(\frac{s_c^2}{n_c}\right)^2}{n_c+1} + \frac{\left(\frac{s_a^2}{n_a}\right)^2}{n_a+1}\right)} - 2 \tag{4}$$

Where all the symbols are as described previously.

If $t \geq t_{crit}$, then decide that the two sets of tests have significantly different means. If $t < t_{crit}$ then decide that there is no reason to believe that the means are significantly different.

Example Problem - Case 1.

A contractor has run 21 QC tests for asphalt content and the State highway agency (SHA) has run 8 acceptance tests over the same period of time for the same material property. The results are shown below. Is it likely that the tests came from the same population?

QC Test Results	Acceptance Test Results
6.4	5.4
6.2	5.8
6.0	6.2
6.6	5.4
6.1	5.4
6.0	5.8
6.3	5.7
6.1	5.4
5.9	
5.8	
6.0	

5.7
6.3
6.5
6.4
6.0
6.2
6.5
6.0
5.9
6.3

First, use the F-test to determine whether or not to assume the variances of the QC tests differ from the acceptance tests.

Step 1. Compute the variance, s^2, for each set of tests.

$$s_c^2 = 0.0606 \qquad s_a^2 = 0.0855$$

Step 2. Compute F, using the largest s^2 in the numerator.

$$F = \frac{s_a^2}{s_c^2} = \frac{0.0855}{0.0606} = 1.41$$

Step 3. Determine F_{crit} from Table 1 being sure to use the correct degrees of freedom for the numerator (n_a-1=8-1=7) and the denominator (n_c-1=21-1=20). From Table 1, F_{crit} = 4.26.

Conclusion: Since $F < F_{crit}$ (i.e., 1.41 < 4.26), there is no reason to believe that the two sets of tests have different variabilities. That is, they could have come from the same population. Since we can assume that the variances are equal, we can use the pooled variance to calculate the t-test statistic, and the pooled degrees of freedom to determine the critical t value, t_{crit}.

Step 4. Compute the mean, $\bar{X}$, for each set of tests.

$$\bar{X}_c = 6.15 \qquad \bar{X}_a = 5.64$$

Step 5. Compute the pooled variance, s_p^2, using the sample variances from above.

$$S_p^2 = \frac{S_c^2(n_c-1)+S_a^2(n_a-1)}{n_c+n_a-2}$$

$$S_p^2 = \frac{(0.0606)(20)+(0.0855)(7)}{21+8-2} = 0.067$$

Step 6. Compute the t-test statistic, t.

$$t=\frac{|\bar{X}_c-\bar{X}_a|}{\sqrt{\frac{s_p^2}{n_c}+\frac{s_p^2}{n_a}}}$$

$$t=\frac{|6.15-5.64|}{\sqrt{\frac{0.067}{21}+\frac{0.067}{8}}}=\frac{0.51}{\sqrt{0.0116}}=4.735$$

Step 7. Determine the critical t value, t_{cru}, for the pooled degrees of freedom.

degrees of freedom $= (n_c+n_a-2)=(21+8-2)=27.$

From Table 2, for $\alpha=0.01$ and 27 degrees of freedom, $t_{cru} = 2.771$.

Conclusion: Since 4.735 > 2.771, we assume that the sample means are not equal. It is therefore probable that the two sets of tests did not come from the same population.

Example Problem - Case 2:

A contractor has run 25 QC tests and the SHA has run 10 acceptance tests over the same period of time for the same material property. The results are shown below. Is it likely that the test came from the same population?

QC Test Results	Acceptance Test Results
21.4	34.7
20.2	16.8
24.5	16.2
24.2	27.7
23.1	20.3
22.7	16.8
23.5	20.0
15.5	19.0
17.9	11.3
24.1	22.3
18.6	
15.9	
17.0	
20.0	
24.2	
14.6	
19.7	
16.0	
23.1	
20.8	
14.6	
16.4	
22.0	
18.7	
24.2	

First, use the F-test to determine whether or not to assume the variances of the QC tests differ from the acceptance tests.

Step 1. Compute the variance, S^2, for each set of tests.

$$S_c^2 = 11.50 \qquad S_a^2 = 43.30$$

Step 2. Compute F, using the largest S^2 in the numerator.

$$F = \frac{S_a^2}{S_c^2} = \frac{43.30}{11.50} = 3.76$$

Step 3. Determine F_{crit} from Table 1 being sure to use the correct degrees of freedom for the numerator (n_a-1=10-1=9) and the denominator (n_c-1=25-1=24). From Table 1, F_{crit} = 3.69.

Conclusion: Since $F > F_{crit}$ (i.e., 3.76 > 3.69), there is reason to believe that the two sets of tests have different variabilities. That is, it is likely that they came from populations with different variances. Since we assume that the variances are not equal, we use the individual sample variances to calculate the t-test statistic, and the approximate degrees of freedom to determine the critical t value, t_{crit}.

Step 4. Compute the mean, $\bar{X}$, for each set of tests.

$$\bar{X}_c = 20.1 \qquad \bar{X}_a = 20.5$$

Step 5. Compute the t-test statistic, t.

$$t=\frac{|\bar{X}_c-\bar{X}_a|}{\sqrt{\frac{S_c^2}{n_c}+\frac{S_a^2}{n_a}}}$$

$$t=\frac{|20.5-20.1|}{\sqrt{\frac{11.50}{25}+\frac{43.30}{10}}}=\frac{0.4}{\sqrt{4.79}}=0.183$$

Step 6. Determine the critical t value, t_{crit}, for the approximate degrees of freedom, f'. Remember that the calculated effective degrees of freedom is rounded down to a whole number.

$$f'=\frac{\left(\frac{s_c^2}{n_c}+\frac{s_a^2}{n_a}\right)^2}{\left(\frac{\left(\frac{s_c^2}{n_c}\right)^2}{n_c+1}+\frac{\left(\frac{s_a^2}{n_a}\right)^2}{n_a+1}\right)}-2$$

$$f' = \frac{\left(\frac{11.50}{25} + \frac{43.30}{10}\right)^2}{\left(\frac{\left(\frac{11.50}{25}\right)^2}{26} + \frac{\left(\frac{43.30}{10}\right)^2}{11}\right)} - 2 = \frac{(4.79)^2}{1.713} - 2 = 11$$

From Table 2, for $\alpha = 0.01$ and 11 degrees of freedom, $t_{crit} = 3.106$.

Conclusion: Since $t < t_{crit}$, (i.e., $0.183 < 3.106$), there is no reason to assume that the sample means are not equal. It is therefore reasonable to assume that the sets of test results came from populations that had the same mean.

Table 1. Critical Values, F_{crit}, for the F-test for a Level of Significance, $\alpha = 0.01$*.

DEGREES OF FREEDOM FOR NUMERATOR

DEGREES OF FREEDOM FOR DENOMINATOR

	1	2	3	4	5	6	7	8	9	10	11	12
1	16200	20000	21600	22500	23100	23400	23700	23900	24100	24200	24300	24400
2	198	199	199	199	199	199	199	199	199	199	199	199
3	55.6	49.8	47.5	46.2	45.4	44.8	44.4	44.1	43.9	43.7	43.5	43.4
4	31.3	26.3	24.3	23.2	22.5	22.0	21.6	21.4	21.1	21.0	20.8	20.7
5	22.8	18.3	16.5	15.6	14.9	14.5	14.2	14.0	13.8	13.6	13.5	13.4
6	18.6	14.5	12.9	12.0	11.5	11.1	10.8	10.6	10.4	10.2	10.1	10.0
7	16.2	12.4	10.9	10.0	9.52	9.16	8.89	8.68	8.51	8.38	8.27	8.18
8	14.7	11.0	9.60	8.81	8.30	7.95	7.69	7.50	7.34	7.21	7.10	7.01
9	13.6	10.1	8.72	7.96	7.47	7.13	6.88	6.69	6.54	6.42	6.31	6.23
10	12.8	9.43	8.08	7.34	6.87	6.54	6.30	6.12	5.97	5.85	5.75	5.66
11	12.2	8.91	7.60	6.88	6.42	6.10	5.86	5.68	5.54	5.42	5.32	5.24
12	11.8	8.51	7.23	6.52	6.07	5.76	5.52	5.35	5.20	5.09	4.99	4.91
15	10.8	7.70	6.48	5.80	5.37	5.07	4.85	4.67	4.54	4.42	4.33	4.25
20	9.94	6.99	5.82	5.17	4.76	4.47	4.26	4.09	3.96	3.85	3.76	3.68
24	9.55	6.66	5.52	4.89	4.49	4.20	3.99	3.83	3.69	3.59	3.50	3.42
30	9.18	6.35	5.24	4.62	4.23	3.95	3.74	3.58	3.45	3.34	3.25	3.18
40	8.83	6.07	4.98	4.37	3.99	3.71	3.51	3.35	3.22	3.12	3.03	2.95
60	8.49	5.80	4.73	4.14	3.76	3.49	3.29	3.13	3.01	2.90	2.82	2.74
120	8.18	5.54	4.50	3.92	3.55	3.28	3.09	2.93	2.81	2.71	2.62	2.54
∞	7.88	5.30	4.28	3.72	3.35	3.09	2.90	2.74	2.62	2.52	2.43	2.36

* NOTE: This is for a two-tailed test with the null and alternate hypotheses shown below:

$$H_o: S_c^2 = S_a^2$$

$$H_a: S_c^2 \neq S_a^2$$

Table 1. Critical Values, F_{crit}, for the F-test for a Level of Significance, α = 0.01*. *(continued)*

DEGREES OF FREEDOM FOR NUMERATOR

DEGREES OF FREEDOM FOR DENOMINATOR

	15	20	24	30	40	50	60	100	120	200	500	∞
1	24600	24800	24900	25000	25100	25200	25300	25300	25400	25400	25400	25500
2	199	199	199	199	199	199	199	199	199	199	199	200
3	43.1	42.8	42.6	42.5	42.3	42.2	42.1	42.0	42.0	41.9	41.9	41.8
4	20.4	20.2	20.0	19.9	19.8	19.7	19.6	19.5	19.5	19.4	19.4	19.3
5	13.1	12.9	12.8	12.7	12.5	12.5	12.4	12.3	12.3	12.2	12.2	12.1
6	9.81	9.59	9.47	9.36	9.24	9.17	9.12	9.03	9.00	8.95	8.91	8.88
7	7.97	7.75	7.65	7.53	7.42	7.35	7.31	7.22	7.19	7.15	7.10	7.08
8	6.81	6.61	6.50	6.40	6.29	6.22	6.18	6.09	6.06	6.02	5.98	5.95
9	6.03	5.83	5.73	5.62	5.52	5.45	5.41	5.32	5.30	5.26	5.21	5.19
10	5.47	5.27	5.17	5.07	4.97	4.90	4.86	4.77	4.75	4.71	4.67	4.64
11	5.05	4.86	4.76	4.65	4.55	4.49	4.45	4.36	4.34	4.29	4.25	4.23
12	4.72	4.53	4.43	4.33	4.23	4.17	4.12	4.04	4.01	3.97	3.93	3.90
15	4.07	3.88	3.79	3.69	3.59	3.52	3.48	3.39	3.37	3.33	3.29	3.26
20	3.50	3.32	3.22	3.12	3.02	2.96	2.92	2.83	2.81	2.76	2.72	2.69
24	3.25	3.06	2.97	2.87	2.77	2.70	2.66	2.57	2.55	2.50	2.46	2.43
30	3.01	2.82	2.73	2.63	2.52	2.46	2.42	2.32	2.30	2.25	2.21	2.18
40	2.78	2.60	2.50	2.40	2.30	2.23	2.18	2.09	2.06	2.01	1.96	1.93
60	2.57	2.39	2.29	2.19	2.08	2.01	1.96	1.86	1.83	1.78	1.73	1.69
120	2.37	2.19	2.09	1.98	1.87	1.80	1.75	1.64	1.61	1.54	1.48	1.43
∞	2.19	2.00	1.90	1.79	1.67	1.59	1.53	1.40	1.36	1.28	1.17	1.00

* NOTE: This is for a two-tailed test with null and alternate hypotheses shown below:

$$H_o: S_c^2 = S_a^2$$

$$H_a: S_c^2 \neq S_a^2$$

Table 2. Critical Values, t_{crit}, for the t-test* for Various Levels of Significance.

degrees of freedom	$\alpha = 0.01$	$\alpha = 0.05$	$\alpha = 0.10$
1	63.657	12.706	6.314
2	9.925	4.303	2.920
3	5.841	3.182	2.353
4	4.604	2.776	2.132
5	4.032	2.571	2.015
6	3.707	2.447	1.943
7	3.499	2.365	1.895
8	3.355	2.306	1.860
9	3.250	2.262	1.833
10	3.169	2.228	1.812
11	3.106	2.201	1.796
12	3.055	2.179	1.782
13	3.012	2.160	1.771
14	2.977	2.145	1.761
15	2.947	2.131	1.753
16	2.921	2.120	1.746
17	2.898	2.110	1.740
18	2.878	2.101	1.734
19	2.861	2.093	1.729
20	2.845	2.086	1.725
21	2.831	2.080	1.721
22	2.819	2.074	1.717
23	2.807	2.069	1.714
24	2.797	2.064	1.711
25	2.787	2.060	1.708
26	2.779	2.056	1.706
27	2.771	2.052	1.703
28	2.763	2.048	1.701
29	2.756	2.045	1.699
30	2.750	2.042	1.697
40	2.704	2.021	1.684
60	2.660	2.000	1.671
120	2.617	1.980	1.658
∞	2.576	1.960	1.645

*** NOTE:** This is for a two-tailed test with the null and alternate hypotheses shown below:

$$H_0: \bar{X}_c = \bar{X}_a$$
$$H_a: \bar{X}_c \neq \bar{X}_a$$