Recommend Approval: 5/22/12 Team Leader Date 5/31/12 Division Chief Date	Maryland Department of Transportation State Highway Administration Office of Materials Technology MARYLAND STANDARD METHOD OF TESTS	
Approved: 06/06/12 Director Date	**PROCEDURE FOR EVALUATING BITUMINOUS MATERIALS FOR STATISTICAL COMPLIANCE**	**MSMT 733**

SCOPE:

This procedure is used to compare the results of the Contractor's Quality Control (QC) test results and the Administration's acceptance 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 referred to as the F-test and the t-test. This procedure is used to determine if QC data can be combined with the acceptance data for pay factor determination.

The comparisons are made of the two groups of test results by analyzing each group's mean value (averages) and each group's variability (standard deviation or variances). The F-test is used first to compare the variance of the two groups of data and t-test is used next to assess possible differences in the means of the two groups of data. Each hypothesis test is conducted at a selected level of significance, α. The level of significance, α is the probability of incorrectly determining that the data for a lot for each group are different when they actually come from the same populations. For this standard, Maryland uses $\alpha = 0.01$.

REFERENCE:

R 9 Acceptance Sampling Plans for Highway Construction

MATERIALS AND EQUIPMENT:

Not applicable.

PROCEDURE:

Sampling and testing shall be performed as specified in MSMT 457, MSMT 458, MSMT 459 or other procedure or standard as required.

Sample Variance - F-test:

The F-test is a prerequisite to performing the t-test for acceptance. After the computations are complete for the F-test, one of two conclusions can be made:

1. The QC test data and the acceptance test data variances differ significantly or,

2. The QC test data and the acceptance test data have statistically similar variances.

Step 1: Compute the variance of the QC test, and the acceptance tests. The variance of the tests results is the square of the standard of deviation of each set of data. The standard deviation for each set of data can be found by completing steps 1 and 2 of or MSMT 735. **Note:** Round the mean to one decimal place more than the test values indicate and round the variance to two decimal places more than the test values indicate.

Step 2: Compute F statistic:

Note: Round F_{stat} to two decimal places.

$$F_{stat} = (S_c)^2/(S_a)^2 \text{ or } (S_a)^2/(S_c)^2$$

where:

$(S_c)^2$ = the sample variance of the QC test results

$(S_a)^2$ = the sample variance of the acceptance test results

Note: Always divide the larger variance by the smaller variance.

Step 3**:** Determine the degrees of freedom for each set of results:

Degrees of freedom for $(S_c)^2 = (N_c\text{-}1)$

Degrees of freedom for $(S_a)^2 = (N_a\text{-}1)$

where:

N_c = number of quality control tests

N_a = number of acceptance tests

Step 4**:** Determine critical F_{crit} from Table 1.

Step 5: Compare F_{stat} value derived from Step 2 above with F_{crit}:

- If F_{stat} is greater than or equal to F_{crit}, then two sets of tests have statistically different variances.
- If F_{stat} is less than F_{crit}, then the two sets of tests do not have statistically different variances.

Difference of Means, t-test:

Once the outcome of the F-test has been determined, the means of the test results can be compared to determine whether they differ statistically from one another or not. The t-test is used to compare the sample means. Two approaches for the t-test are necessary. If the sample variances are statistically similar, then conduct the t-test 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 conduct the t-test using the individual sample variances, the individual sample sizes, and the effective degrees of freedom. This approach is in Case 2 described below.

In either of the two cases discussed above, one of the following decisions is made:

- The two sets of data have statistically different means.
- The two sets of data do not have statistically different means.

CASE 1: Where the sample variances are assumed to be equal.

Step 1: Calculate the pooled estimate of variance:

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

Note: Round S^2 to two decimal places more than the data used to calculate it.

where:

S_p^2 = pooled estimate of variance

n_c = number of QC tests

n_a = number of QA tests

S_c^2 = variance of the QC tests

S_a^2 = variance of the QA tests

Step 2: Calculate the t statistic using the following equation for equal variances:

$$t_{stat} = \frac{\left|\overline{X}_c - \overline{X}_a\right|}{\sqrt{\frac{S_p^2}{n_c} + \frac{S_p^2}{n_a}}}$$

Note: Round t_{stat} to three decimal places.

where:

$\overline{X_c}$ = mean of QC tests

$\overline{X_a}$ = mean of QA tests

S_p^2 = variance of the pooled QC and QA tests

n_c = number of QC tests

n_a = number of QA tests

Step 3: Determine the pooled degrees of freedom:

Pooled degrees of freedom = $(n_c + n_a - 2)$

Step 4: Determine the critical t value (t_{crit}) from Table 2 (pg. 10) for the pooled degrees of freedom.

Step 5: Compare the t_{stat} value derived from Step 2 above with t_{crit}:

- If t_{stat} is greater than or equal to t_{crit}, then two sets of tests have statistically different means.
- If t_{stat} is less than t_{crit}, then there is no reason to believe that the means are statistically different.

CASE 2: Where the sample variances are not assumed to be equal.

Step 1**:** Calculate the t statistic using the following equation:

$$t_{stat} = \frac{\left|\overline{X}_c - \overline{X}_a\right|}{\sqrt{\frac{S_c^2}{n_c} + \frac{S_a^2}{n_a}}}$$

Note: Round t_{stat} to three decimal places.

where:

$\overline{X_c}$ = mean of QC tests

$\overline{X_a}$ = mean of QA tests

S_c^2 = variance of QC tests

S_a^2 = variance of acceptance tests

n_c = number of QC tests

n_a = number of QA tests

Step 2: Determine the effective degrees of freedom:

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

Note: Round f' to a whole number.

where: all symbols are defined in previous step.

Step 3: Determine the critical t value (t_{crit}) from Table 2 for the effective degrees of freedom.

Step 4: Compare the t_{stat} value derived from Step 1 above with t_{crit}:

- If t_{stat} is greater than or equal to t_{crit}, then two sets of tests have statistically different means.
- If t_{stat} is less than t_{crit}, then there is no reason to believe that the means are statistically different.

Example Problem for Case 1:

A Contractor has performed 21 QC tests for asphalt content and the Administration has performed 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 populations?

Contractor QC Test Results	Administration Acceptance Test Results
6.4, 6.2, 6.0, 6.6, 6.1, 6.0, 6.3, 6.1, 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	5.4, 5.8, 6.2, 5.4, 5.4, 5.8, 5.7, 5.4

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.

$$S_a^2 = 0.086 \quad S_c^2 = 0.061$$

Note: Round S^2 to two decimal places more than the data used to calculate it.

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

$$F_{stat} = \frac{S_a^2}{S_c^2} = \frac{0.086}{0.061} = 1.41$$

Note: Round F_{stat} to two decimal places.

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_{stat} is less than F_{crit} (i.e., 1.41 is less than 4.26), there is no reason to believe that the two sets of tests have different variances. 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, $\overline{X}$, for each set of tests.

> **Note:** Round $\bar{X}$ to one decimal place more than the data used to calculate it.

$$\overline{X_c} = 6.15 \qquad \overline{X_a} = 5.64$$

Step 5: Compute the pooled variance,**Error! Bookmark not defined.Error! Bookmark not defined.** S_p^2 using the sample variances from above.

> **Note:** Round S^2 to two decimal places more than the data used to calculate it.

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

$$S_p^2 = \frac{(0.061)(20) + (0.086)(7)}{21 + 8 - 2} = 0.067$$

Step 6: Compute the t-test statistic, t_{stat}

> **Note:** Round t to three decimal places.

$$t_{stat} = \frac{\left|\overline{x}_c - \overline{x}_a\right|}{\sqrt{\frac{S_p^2}{n_c} + \frac{S_p^2}{n_a}}}$$

$$t_{stat} = \frac{\left|6.15 - 5.64\right|}{\sqrt{\frac{0.067}{21} + \frac{0.067}{8}}} = 4.742$$

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

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

From Table 2, for a = 0.01 and 27 degrees of freedom, t_{crit} = 2.771

Conclusion: Since t_{stat} is greater than t_{crit} (i.e., 4.742 is greater than 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 for Case 2:

A Contractor has performed 25 QC tests and the Administration has performed 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?

Contractor QC Test Results	Administration Acceptance Test Results
21.4, 20.2, 24.5, 24.2, 23.1, 22.7, 23.5, 15.5, 17.9, 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, 24.1, 18.6	34.7, 16.8, 16.2, 27.7, 20.3, 16.8, 20.0, 19.0, 11.3, 22.3

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.

$$S_c^2 = 11.522 \qquad S_a^2 = 43.308$$

> **Note:** Round S^2 to two decimal places more than the data used to calculate it.

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

$$F_{stat} = \frac{S_a^2}{S_c^2} = \frac{43.308}{11.522} = 3.76$$

> **Note:** Round F_{stat} to two decimal places.

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_{stat} is greater than F_{crit} (i.e., 3.76 is greater than 3.69), there is reason to believe that the two sets of tests have different variability's. 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, $\overline{X}$, for each set of tests.

$$\overline{X}_c = 20.12 \qquad \overline{X}_a = 20.51$$

> **Note:** Round $\bar{X}$ to one decimal place more than the data used to calculate it.

Step 5: Compute the t-test statistic, t_{stat}

$$t_{stat} = \frac{\left|\overline{x}_c - \overline{x}_a\right|}{\sqrt{\dfrac{S_c^2}{n_c} + \dfrac{S_a^2}{n_a}}}$$

Note: Round t to three decimal places.

$$t_{stat} = \frac{20.12-20.51}{\sqrt{\frac{11.519}{25}+\frac{43.310}{10}}} = \frac{0.39}{\sqrt{4.792}} = 0.178$$

Step 6: Determine the critical t value, t_{crit}, for the approximate degrees of freedom f'

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

Note: Round f' to a whole number.

$$f' = \frac{\left(\dfrac{11.519}{25} + \dfrac{43.310}{10}\right)^2}{\left(\dfrac{\left(\dfrac{11.519}{25}\right)^2}{26} + \dfrac{\left(\dfrac{43.310}{10}\right)^2}{11}\right)} - 2 = \frac{(4.792)^2}{1.713} - 2 = 11$$

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

Conclusion: Since t_{stat} is less than t_{crit} (i.e., 0.178 is less than 3.106), there is no reason to assume that the sample means are not statistically similar. It is therefore reasonable to assume that the sets of test results came from similar populations.

Table 1

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

DEGREE OF FREEDOM FOR NUMERATOR

DEGREE OF FREEDOM FOR DENOMINATOR	1	2	3	4	5	6	7	8	9	10	11	12	15	20	24	30	40	50	60	100	120	200	500	∞
1	16200	20000	21600	22500	23100	23400	23700	24100	24100	24200	24300	24400	24600	24800	24900	25000	25100	25200	25300	25300	25400	25400	25400	
2	198	199	199	199	199	199	199	199	199	199	199	199	199	199	199	199	199	199	199	199	199	199	199	200
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	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	31.3	26.3	24.3	23.2	22.5	22.0	21.6	21.4	21.1	21.0	20.8	20.7	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	22.8	18.3	16.5	15.6	14.9	14.5	14.2	14.0	13.8	13.6	13.5	13.4	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	18.6	14.5	12.9	12.0	11.5	11.1	10.8	10.6	10.4	10.2	10.1	10.0	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	16.2	12.4	10.9	10.0	9.52	9.16	8.89	8.68	8.51	8.38	8.27	8.18	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	14.7	11.0	9.60	8.81	8.30	7.95	7.69	7.50	7.34	7.21	7.10	7.01	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	13.6	10.1	8.72	7.96	7.47	7.13	6.88	6.69	6.54	6.42	6.31	6.23	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	12.8	9.43	8.08	7.34	6.87	6.54	6.30	6.12	5.97	5.85	5.75	5.66	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	12.2	8.91	7.6	6.88	6.42	6.10	5.86	5.68	5.54	5.42	5.32	5.24	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	11.8	8.51	7.23	6.52	6.07	5.76	5.52	5.35	5.20	5.09	4.99	4.91	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	10.8	7.70	6.48	5.80	5.37	5.07	4.85	4.67	4.54	4.42	4.33	4.25	4.07	3.88	3.79	3.69	3.59	3.52	3.48	3.36	3.37	3.33	3.29	3.26
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	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	9.55	6.66	5.52	4.89	4.49	4.20	3.99	3.83	3.69	3.59	3.50	3.42	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	9.18	6.35	5.24	4.62	4.23	3.95	3.74	3.58	3.45	3.34	3.25	3.18	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	8.83	6.07	4.98	4.37	3.99	3.71	3.51	3.35	3.22	3.12	3.03	2.95	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	8.49	5.80	4.73	4.14	3.76	3.49	3.29	3.13	3.01	2.90	2.82	2.74	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	8.18	5.54	4.5	3.92	3.55	3.28	3.09	2.93	2.81	2.71	2.62	2.54	2.37	2.19	2.09	1.98	1.87	1.80	1.75	1.64	1.61	1.54	1.48	1.43
∞	7.88	5.3	4.28	3.72	3.35	3.09	2.90	2.74	2.62	2.52	2.43	2.36	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 the null and alternate hypothesis 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	$\alpha = 0.01$
1	63.657
2	9.925
3	5.841
4	4.604
5	4.032
6	3.707
7	3.499
8	3.355
9	3.250
10	3.169
11	3.106
12	3.055
13	3.012
14	2.977
15	2.947
16	2.921
17	2.898
18	2.878
19	2.861
20	2.845
21	2.831
22	2.819
23	2.807
24	2.797
25	2.787
26	2.779
27	2.771
28	2.763
29	2.756
30	2.750
40	2.704
60	2.660
120	2.617
∞	2.576

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

$$H_0 : \overline{X}_c = \overline{X}_a$$

$$H_a : \overline{X}_c \neq \overline{X}_a$$