APPENDIX N
Hypothesis Testing—Tests of Dispersion

N-1. Introduction. Many statistical tests make assumptions on the dispersion of data, as measured by variance. This Appendix considers some of the most commonly used statistical tests for equality of variance, a key assumption for the validity of a two-sample *t*-test and analysis of variance (ANOVA). More information on hypothesis tests on the variance can be found in EPA 600/R-96/084, QA/G-9.

N-2. *F*-Test for the Equality of Two Variances. An *F*-test may be used to see whether the true underlying variances of two populations are equal. Usually the *F*-test is employed as a preliminary test, before conducting the two-sample *t*-test for the equality of two means. The assumptions underlying the *F*-test are that the two samples are independent random samples from two underlying normal populations. The *F*-test for equality of variances is highly sensitive to departures from normality. (In the case of non-normality, Levene's test is recommended, Paragraph N-4.) Directions for implementing an *F*-test are given in Paragraph N-2.1, followed by an example in Paragraph N-2.2.

N-2.1. *Directions for an* F-*Test Comparing Two Variances.* Let $x_1, x_2,..., x_m$ represent the m data points from population 1 and $y_1, y_2,..., y_n$ represent n data points from population 2. To perform an *F*-test, proceed as follows.

N-2.1.1. Test the null hypothesis of equal variances:

$$H_0 : \sigma_x^2 = \sigma_y^2,\ H_A : \sigma_x^2 \neq \sigma_y^2\ .$$

N-2.1.2. Verify the assumption of normality using one of the methods described in Appendix F.

N-2.1.3. Calculate the sample variance, s_x^2 (for the X's) and s_y^2 (for the Y's) (Appendix D).

N-2.1.4. Calculate the variance ratios, $f_x = s_x^2 / s_y^2$ and $f_y = s_y^2 / s_x^2$.

N-2.1.5. Let f equal the larger of these two values.

N-2.1.5.1. If $f = f_x$, then let $k = m - 1$ and $q = n - 1$.

N-2.1.5.2. If $f = f_y$, then let $k = n - 1$ and $q = m - 1$.

N-2.1.6. Using Table B-7 of Appendix B of the *F*-distribution, we find the critical value,

$$U = F_{1-\alpha/2,k,q}$$

where k denotes the degrees of freedom in the numerator and q the degrees of freedom in the denominator for the ratio f.

N-2.1.6.1. If $f > U$, conclude that the variances of the two populations are not the same.

N-2.1.6.2. If $f \le U$, there is insufficient evidence to conclude the variances are different.

N-2.2. *Example of an* F-*Test Comparing Two Variances*. Consider the case where nickel concentrations in surface soil are compared between Site A and Background. The null and alternative hypotheses are:

$$H_0 : \sigma_x^2 = \sigma_y^2,\ H_A : \sigma_x^2 \ne \sigma_y^2 \ .$$

N-2.2.1. Nickel in surface soils at Site A (X) was detected at following concentrations (m = 6): 2.665, 3.610, 5.470, 7.150, 8.340, 7.960 mg/kg.

N-2.2.2. Nickel in surface background (bkgd) soils (Y) was detected at the following concentrations (n =10): 5.140, 7.460, 5.990, 3.360, 3.190, 2.870, 5.950, 1.720, 4.770, 5.605 mg/kg.

N-2.2.3. Verify the assumption of normality. For this case, the Shapiro-Wilk test is used.

N-2.2.4. Calculate the sample variance, s_x^2 (for the X's) and s_y^2 (for the Y's).

	Sample Mean	**Sample Variance**	**Sample Size**
Site data	5.87	5.53	6
Background data	4.61	3.12	10

N-2.2.5. Calculate the variance ratios:

$$f_x = s_x^2 / s_y^2 = \frac{5.53}{3.12} = 1.77$$

and

$$f_y = s_y^2 / s_x^2 = \frac{3.12}{5.53} = 0.56\ .$$

N-2.2.6. Therefore $f = 1.77$.

N-2.2.7. Because $f = f_x$, $k = 6 - 1 = 5$ and $q = 10 - 1 = 9$.

$U = F_{1-\alpha/2,k,q} = f_{0.975,5,9} = 4.484$.

N-2.2.8. Because $f \le U$ $(1.77 \le 4.484)$, there is insufficient evidence to conclude the variances are different.

N-3. Bartlett's Test for the Equality of Two or More Variances. Bartlett's test, which is essentially a generalization of the *F*-test, is a way of testing whether two or more population variances of normal distributions are equal. In the case of only two variances, Bartlett's test is equivalent to the *F*-test. Directions for Bartlett's test are given in Paragraph N-3.1, followed by an example in Paragraph N-3.1. Like the *F*-test it is sensitive to deviations from normality.

N-3.1. *Directions for Bartlett's Test for Two or More Variances*. Let *K* represent the total number of populations to be compared. Let $n_1, n_2, \ldots, n_K$ represent the sample sizes of each of the *K* sample populations. Let *N* represent the total number of samples, $N = n_1 + n_2 .... + n_k$. Let the values from each population be represented by $x_{i,j}$, where $i = 1, 2, \ldots, K$ for the *K* groups and $j = 1, 2, \ldots, n_i$ for the observations in the i^{th} group.

N-3.1.1. $H_0 : \sigma_1^2 = \sigma_2^2 = \cdots = \sigma_K^2$ (no difference among the population variances).

N-3.1.2. H_A : at least one variance, σ_i^2, is different from one or more of the other variances.

N-3.1.3. For example, consider two wells, where four samples have been taken from Well 1 and three samples have been taken from Well 2. In this case, $K = 2$, $n_1 = 4$, $n_2 = 3$, and $N = 4 + 3 = 7$.

N-3.1.4. Verify the assumption of normality using one of the methods described in Appendix F. For each of the *K* groups, calculate the sample variances, $s_i^{\;2}$ (see Appendix D).

N-3.1.5. Compute the pooled variance using the *K* groups:

$$s_p^{\;2} = \frac{1}{(N-K)} \sum_{i=1}^{K} (n_i - 1) s_i^{\;2} \;.$$

N-3.1.6. Compute the test statistic (*TS*):

$$TS = (N - K)\text{Ln}(s_p{}^2) - \sum_{i=1}^{K}(n_i - 1)\text{Ln}(s_i{}^2)$$, where Ln is the natural logarithm.

N-3.1.7. Using a chi-square table (Table B-2 of Appendix B), find the critical value of the chi-squared distribution, $\chi^2_{1-\alpha,\nu}$, with $\nu = K - 1$ degrees of freedom and the $(1-\alpha)100\%$ level of confidence. For example, for a level of confidence of 95% (significance level α = 0.05) and ν = 5, $\chi^2_{0.95,5}$ = 11.1.

N-3.1.7.1. If $TS > \chi^2_{1-\alpha,\nu}$, reject H_0 (conclude that the variances are not all equal) at the $(1-\alpha)100\%$ level of confidence.

N-3.1.7.2. If $TS \le \chi^2_{1-\alpha,\nu}$, there is insufficient evidence to reject H_0.

N-3.2. *Example of Bartlett's Test for Two or More Variables*. Using chromium concentrations in subsurface site soil, the data are: 2.95, 5.17, 4.80, 4.53, 4.01, 5.91, 3.96, 4.81, 5.27, 5.99, 4.60, 5.51, 4.72, 3.56, 4.22, 3.91, 5.81, 4.48, 5.10, 4.94, 4.76, 4.62, 4.72, 4.73, 3.21, 4.14, 4.85, 4.25, 5.09, 3.68, 5.12, 6.60, 6.19, 3.15, 4.11, 2.80 mg/kg.

N-3.2.1. The chromium concentrations in subsurface background soil are: 4.60, 5.29, 4.26, 5.28, 4.53, 5.74, 5.86, 3.84 mg/kg.

N-3.2.2. Verify the assumption of normality. For this case, the Shapiro-Wilk test is used.

N-3.2.3. Let *N* represent the total number of samples. As the site data has $n_1 = 36$ samples and the background data has $n_2 = 8$ samples, $N = 44$ and $K = 2$.

N-3.2.4. For each of the *K* groups, calculate the sample variances, s_i^2: s_1^2 = 0.806 (site variance) and s_2^2 = 0.526 (background variance).

N-3.2.5. Compute the pooled variance:

$$s_p^2 = \frac{1}{(N-K)}\sum_{i-1}^{K}(n_i - 1)s_i^2$$

$$= \frac{1}{(44-2)}\left[(n_1-1)s_1^2+(n_2-1)s_2^2\right]$$
$$= \frac{1}{42}\left[(36-1)\,0.806 + (8-1)\,0.526\right] = 0.7593 \; .$$

N-3.2.6. Compute the test statistic *TS*:

$$TS = (N-K)\,\mathrm{Ln}(s_p^2) - \sum_{i=1}^{K}(n_i-1)\,\mathrm{Ln}(s_i^2)$$
$$= (44-2)\,\mathrm{Ln}(0.7593) - \left[(36-1)\,\mathrm{Ln}(0.806) + (8-1)\,\mathrm{Ln}(0.526)\right] = 0.4802 \; .$$

N-3.2.7. Using a chi-squared table (Table B-2 of Appendix B), find the critical value, $\chi^2_{1-\alpha,\nu}$. In this case, with a significance level of 5% and 1 degree of freedom, $\chi^2_{0.95,1}$ = 3.841. As $TS = 0.4802 \le 3.841$, there is insufficient evidence to conclude the variances are different at the $\alpha = 0.05$ significance level.

N-4. Levene's Test for the Equality of Two or More Variances. Levene's test is a non-parameter alternative to Bartlett's test for homogeneity of variance (testing for differences among the dispersions of several groups). Levene's test is less sensitive to departures from normality than Bartlett's test and has greater power than Bartlett's for non-normal data. In addition, Levene's test has power nearly as great as Bartlett's test for normally distributed data. However, Levene's test is more difficult to apply than Bartlett's test because it involves applying an ANOVA to the absolute deviations from the group means. Directions for Levene's test are given in Paragraph N-4.1, followed by an example in Paragraph N-4.2.

N-4.1. *Directions for Levene's Test for the Equality of Two or More Variances*. Let *K* represent the total number of populations to be compared. Let $n_1, n_2, \ldots, n_K$ represent the sample sizes of each of the *K* sample populations. Let *N* represent the total number of samples, $N = n_1 + n_2 .... + n_k$. Let the values from each population be represented by $x_{i,j}$ where $i = 1,2,\ldots,K$ for the *K* groups and $j = 1, 2, \ldots, n_i$ for the observations in the i^{th} group.

N-4.1.1. $H_0 : \sigma_1^2 = \sigma_2^2 = \cdots = \sigma_K^2$ (no difference among the population variances).

N-4.1.2. H_A : at least one variance, σ_i^2, is different from one or more of the other variances.

N-4.1.3. For example, consider two wells where four samples have been taken from well 1 and three samples have been taken from well 2. In this case, $K = 2,\ n_1 = 4,\ n_2 = 3,$ and $N = 4 + 3 = 7.$

N-4.1.4. Verify the assumption of normality using one of the methods described in Appendix F. For each of the *K* groups, calculate the group mean, $\bar{x}_i$:

$$\bar{x}_1 = \frac{1}{n_1}\sum_{j=1}^{n_1} x_{1,\,j}\ ,\quad \bar{x}_2 = \frac{1}{n_2}\sum_{j=1}^{n_2} x_{2,\,j}\ ,\ \ldots.\ ,\quad \bar{x}_K = \frac{1}{n_K}\sum_{j=1}^{n_K} x_{K,\,j}\ .$$

N-4.1.5. Compute the absolute residuals

$$z_{i,\,j} = \left| x_{i,\,j} - \bar{x}_i \right|$$

where $x_{i,\,j}$ represents the j^{th} value of the i^{th} group. For each of the *K* groups, calculate the means, $\bar{z}_i$,of these residuals:

$$\bar{z}_1 = \frac{1}{n_1}\sum_{j=1}^{n_2} z_{1,\,j}\ ,\quad \bar{z}_2 = \frac{1}{n_2}\sum_{j=1}^{n_2} z_{2,\,j}\ ,\ldots.,\quad \bar{z}_K = \frac{1}{n_K}\sum_{j=1}^{n_K} z_{K,\,j}\ .$$

N-4.1.6. Calculate the overall mean residual:

$$\bar{z} = \frac{1}{n}\sum_{i=1}^{K}\sum_{j=1}^{n_i} z_{i,\,j} = \frac{1}{n}\sum_{i=1}^{K} n_i \bar{z}_i\ .$$

N-4.1.7. Compute the following sums of squares for the absolute residuals:

$$SS_{TOTAL} = \sum_{i=l}^{K}\sum_{j=1}^{n_i} z_{i,\,j}^2 - n\,\bar{z}^2$$

$$SS_{GROUPS} = \sum_{i=l}^{K}\frac{\bar{z}_i^{\,2}}{n_i} - n\,\bar{z}^2$$

$$SS_{ERROR} = SS_{TOTAL} - SS_{GROUPS}\ .$$

N-4.1.8. Compute

$$f = \frac{SS_{GROUPS} /(K-1)}{SS_{ERROR} /(N-K)} \ .$$

N-4.1.9. Using Table B-7 of Appendix B, find $F_{1-\alpha, k-1, N-K}$, the critical value of the *F*-distribution with (K – 1) numerator degrees of freedom, $(N-K)$ denominator degrees of freedom, and the desired level of significance, α. For example, if $\alpha = 0.05$, the numerator degrees of freedom are 5, and the denominator degrees of freedom are 18, then using Table B-7, we find that $F_{0.95, 5, 18} = 2.77$.

N-4.1.10. If $f > F$, reject the assumption of equal variances.

N-4.2. *Example of Levene's Test for the Equality of Two or More Variables*. Consider the case where nickel concentrations in surface soil are compared between Site A and background (bkgd) using the test:

$$H_0 : \sigma_x^2 = \sigma_y^2, \ \ H_A : \sigma_x^2 \neq \sigma_y^2 \ .$$

N-4.2.1. Suppose data for nickel in surface site soil are: 2.665, 3.610, 5.470, 7.150, 8.340, 7.960 mg/kg. And suppose data for nickel in surface background are: 5.140, 7.460, 5.990, 3.360, 3.190, 2.870, 5.950, 1.720, 4.770, 5.605 mg/kg.

N-4.2.2. Verify the assumption of normality. For this case, the Shapiro-Wilk test is used.

Site mean = $\bar{x}_1 = \dfrac{1}{n_1}\sum_{j=1}^{n_1} x_{1,j} = 5.87$.

Background mean = $\bar{x}_2 = \dfrac{1}{n_2}\sum_{j=1}^{n_2} x_{2,j} = 4.61$.

Site A	$z_{i,j} = \lvert x_{i,j} - \bar{x}_i \rvert$	**Backgroundd**	$z_{i,j} = \lvert x_{i,j} - \bar{x}_i \rvert$
2.67	3.20	5.14	0.534
3.61	2.26	7.46	2.854
5.47	0.40	5.99	1.384
7.15	1.28	3.36	1.246
8.34	2.47	3.19	1.416
7.96	2.09	2.87	1.736
		5.95	1.344
		1.72	2.886
		4.77	0.164
		5.61	0.999

Mean of the site residuals = $\bar{z}_1 = \dfrac{1}{n_1}\sum_{j=1}^{n_1} z_{1,j} = 1.95$.

Mean of the background residuals = $\bar{z}_2 = \dfrac{1}{n_2}\sum_{j=1}^{n_2} z_{2,j} = 1.46$.

Overall residual mean = $\bar{z} = \dfrac{1}{n}\sum_{i=1}^{K}\sum_{j=1}^{n_i} z_{i,j} = \dfrac{1}{n}\sum_{i=1}^{K} n_i\,\bar{z}_i = 1.64$.

$$SS_{TOTAL} = \sum_{i=1}^{K}\sum_{j=1}^{n_i} z_{i,j}^2 - n\,\bar{z} = 12.60.$$

$$SS_{GROUPS} = \sum_{i=l}^{K}\frac{\bar{z}_i^{\;2}}{n_i} - n\,\bar{z} = 0.92.$$

$$SS_{ERROR} = SS_{TOTAL} - SS_{GROUPS} = 11.68.$$

$$f = \frac{SS_{GROUPS}\,/(K-1)}{SS_{ERROR}\,/(N-K)} = \frac{0.9167\,/(2-1)}{11.68\,/(16-2)} = 1.098.$$

Numerator degrees of freedom: $(K - 1) = (2 - 1) = 1$.

Denominator degrees of freedom: $(N - K) = (16 - 2) = 14$.

N-4.2.3. Because $\alpha = 0.05$, the critical value $F_{0.95,1,14} = 4.611$. Comparing the calculated value (f) and the critical value, $F_{0.95,1,14}$, we see that $f \le F_{0.95,1,14}$, so do not reject H_0. Therefore, we can conclude that the variance for the surface soil site concentration of nickel is equal to the variance of the surface soil background concentrations of nickel.

N-5. Maximum *F*-Ratio Test for Equality of Two or More Variances. The maximum *F*-ratio tests whether three or more population variances from normal distributions are equal (Mason et al., 1989). The test also assumes that the sample sizes for the populations are equal. As this test is sensitive to departures from normality, it is recommended that normality tests be done before using it. Directions are given in Paragraph N-5.1, followed by an example in Paragraph N-5.2.

N-5.1. *Directions for the Maximum* F-*Ratio Test for Equality of Two or More Variances*. Let K represent the total number of populations to be compared. Let $n_1, n_2, \ldots, n_K$ represent the sample sizes of each of the K sample populations. Let N represent the total number of samples, $N = n_1 + n_2 .... + n_k$. Let the values from each population be represented by $x_{i,j}$ where $i = 1, 2, \ldots, K$ for the K groups and $j = 1, 2, \ldots, n_i$ for the observations in the i^{th} group.

N-5.1.1. $H_0 : \sigma_1^2 = \sigma_2^2 = \cdots = \sigma_K^2$ (no difference among the population variances).

N-5.1.2. H_A : at least one variance, σ_i^2, is different from one or more of the other variances.

N-5.1.3. Verify the assumption of normality using one of the methods described in Appendix F.

N-5.1.4. Calculate the sample standard deviation for each of the K data sets. Denote these standard deviations by s_i and the corresponding sample size by n_i, where $i = 1, 2, \ldots, K$. Identify the largest value of s_i, max(s_i), and the smallest value of s_i, min(s_i).

N-5.1.5. Calculate the ratio $f_{\max} = \left(\max(s_i) / \min(s_i)\right)^2$.

N-5.1.6. If $n_1 = n_2 = \ldots = n_k = \ldots = n$, use the critical values in Table B-27 of Appendix B, $F_{k,v,\alpha}$, where $k = K$ and $v = n - 1$ for the desired level of significance α, to determine whether to reject the hypothesis of equal standard deviations. If the n_i are unequal but not too different, use the "harmonic mean of the n_i", n':

$$v = n' - 1, \text{ where } n' = K \Big/ \sum_{i=1}^{K} \left(1/n_i\right).$$

N-5.1.7. If $f_{\max} > F_{K,v,\alpha}$, then conclude there is evidence that the variances are not equal.

N-5.2. *Example of the Maximum* F-*Ratio Test for Equality of Three or More Variances*. Manganese concentrations in groundwater are compared between seven wells from Site A using the test:

N-5.2.1 $H_0 : \sigma_i^2 = \sigma_j^2$ for all i and j.

N-5.2.2. $H_A : \sigma_i^2 \neq \sigma_j^2$ for some $i \neq j$.

N-5.2.3. The data (Table N-1) were tested for equal variances using Bartlett's test (Paragraph N-3). The data were also tested for normality using the Shapiro-Wilk test. Because the data were not normal, they were transformed so that residuals would follow a normal distribution.

Well	69-2-02	69-2-04	69-2-05	69-2-06A	69-2-06B	69-2-07	69-2-08
s_i	0.254	0.203	0.182	0.103	0.378	0.283	0.301.

$$f_{\max} = \left(\max(s_i)/\min(s_i)\right)^2 = (0.301/0.103)^2 = 8.54\,.$$

Table N-1.
Data for Example N-5.2

Well location	Result (mg/L)	Log result	Well location	Result (mg/L)	Log result
69-2-02	0.432	–0.839	69-2-06A	0.294	–1.224
69-2-02	0.44	–0.821	69-2-06A	0.301	–1.201
69-2-02	0.513	–0.667	69-2-06A	0.379	–0.970
69-2-02	0.704	–0.351	69-2-06A	0.352	–1.044
69-2-02	0.327	–1.118	69-2-06A	0.346	–1.061
69-2-02	0.316	–1.152	69-2-06B	0.13	–2.040
69-2-02	0.454	–0.790	69-2-06B	0.184	–1.693
69-2-02	0.401	–0.914	69-2-06B	0.209	–1.565
69-2-04	0.0504	–2.988	69-2-06B	0.2	–1.609
69-2-04	0.0502	–2.992	69-2-06B	0.0739	–2.605
69-2-04	0.054	–2.919	69-2-06B	0.0876	–2.435
69-2-04	0.0523	–2.951	69-2-06B	0.126	–2.071
69-2-04	0.0923	–2.383	69-2-06B	0.129	–2.048
69-2-04	0.0556	–2.890	69-2-07	0.0137	–4.290
69-2-04	0.0534	–2.930	69-2-07	0.019	–3.963
69-2-04	0.0517	–2.962	69-2-07	0.0163	–4.117
69-2-05	0.00684	–4.985	69-2-07	0.0195	–3.937
69-2-05	0.00639	–5.053	69-2-07	0.0112	–4.492
69-2-05	0.00631	–5.066	69-2-07	0.0112	–4.492
69-2-05	0.00813	–4.812	69-2-07	0.0102	–4.585
69-2-05	0.00747	–4.897	69-2-07	0.00946	–4.661
69-2-05	0.00679	–4.992	69-2-08	0.563	–0.574
69-2-05	0.00731	–4.919	69-2-08	0.512	–0.669
69-2-05	0.00444	–5.417	69-2-08	0.475	–0.744
69-2-06A	0.3	–1.204	69-2-08	0.546	–0.605
69-2-06A	0.286	–1.252	69-2-08	0.276	–1.287
69-2-06A	0.303	–1.194	69-2-08	0.383	–0.960
			69-2-08	0.33	–1.109
			69-2-08	0.27	–1.309

N-5.2.4. Because $n_1 = n_2 = \ldots = n_k = \ldots = n$, use the critical values in Table B-27 of Appendix B with $v = n - 1 = 8 - 1 = 7$. So, $F_{K,v,\alpha} = F_{7,7,0.05} = 11.80$.

N-5.2.5. Compare the calculated value (8.54) to the critical value (11.80); because the calculated value f_{max} is not greater than the critical value, H_0 cannot be rejected (i.e., there is evidence that the variances are equal).