NOTES ON RATIO TEST FOR EQUAL-VARIANCE HYPOTHESIS

H.R.LAW

Notation

We are basically focus on the one-way analysis of variance model(ANOVA). Assume that there are n observations in all separated into ν groups with sizes r_i. Then we should make sure we know[1]:

$$\begin{aligned}\bar{x}_i &= \tfrac{1}{r_i}\textstyle\sum_{j=1}^{r_i} x_{ij}\\ \bar{x} &= \tfrac{1}{n}\textstyle\sum_{i=1}^{\nu}\sum_{j=1}^{r_i} x_{ij}\\ MSE &= \tfrac{1}{n-\nu}\textstyle\sum_{i=1}^{\nu}\sum_{j=1}^{r_i}(x_{ij}-\bar{x}_i)^2, dfe = n-\nu\\ MStot &= \tfrac{1}{n-1}\textstyle\sum_{i=1}^{\nu}\sum_{j=1}^{r_i}(x_{ij}-\bar{x})^2, dftot = n-1\\ (\nu-1)MST &= (n-1)\cdot MStot-(n-\nu)\cdot MSE, dft = \nu-1\end{aligned}$$

In current version of this note, we basically consider the equal-variance hypothesis about *normal* populations. For non-normal populations, we usually use nonparametric tests like Kruskal-Wallis test. But that makes no assumptions on variances.

I pointed out that [Tucker] gives a brief introduction and [Cochran] treated it from the perspective of sampling theory.

1. When Should We Care about Equal-variance Hypothesis?

1.1. Bartlett's Test for Equal-variance Hypothesis for Normal Populations. When we perform hypothesis test for the mean(or location) of two *normal* sample populations, a usual problem is to determine whether they have equal variances. When their variances are equal, we can either do Z-test or t-test depending on whether we know the value of their variances. When their variances are unequal, we have the only tool of performing a t-test using the Welch-Satterthwaite approximation. [2]. A natural question is how we know whether these two *normal* populations have equal variances or not. More generally, in the setting of *normal* multi-groups(or ANOVA), can we just look by eyes and say they have equal or unequal variances? A statistician's answer should always be: Let us do a hypothesis test.

[1] [Dean&Voss]

[2] This causes a confusion about the notion of degree of freedom, as I will point out in a future individual note.

The classical exact test[3] against the equal-variance hypothesis we need in the one-way layout with the assumptions $H_0 : \epsilon_{ij} \overset{i.i.d}{\sim} N(0,\sigma^2)$ is the Bartlett's test proposed by [Bartlett][4]:

$$\frac{(n-\nu)log(MSE) - \sum_{i=1}^{\nu}(r_i-1)log(s_i^2)}{1+\frac{1}{3(\nu-1)}(\sum_{i=1}^{\nu}(\frac{1}{r_i-1})-\frac{1}{n-\nu})} \sim \chi^2(\nu-1)$$

We seldom care about the equal-variance test when the sample size[5] is large, since we use central limit theorem in that case.

We care a bit about the equal-variance test when the design is balanced (i.e. the number of observations in each group is actually the same), since the t-test statistics of balanced design will be just the same whether the variances are equal or not, of course the degree of freedom will differ[6].

$$t\left(\frac{(\frac{s_1^2}{r}+\frac{s_2^2}{r})^2}{\left[(\frac{s_1^2}{r})^2+(\frac{s_2^2}{r})^2\right]\frac{1}{r-1}}\right) = t\,(Welch) \sim \frac{\bar{x_1}-\bar{x_2}}{\sqrt{\frac{s_1^2}{r}+\frac{s_2^2}{r}}} \overset{r_1=r_2=r}{\Longleftrightarrow}$$

$$\frac{\bar{x_1}-\bar{x_2}}{\sqrt{\frac{(r_1-1)s_1^2+(r_2-1)s_2^2}{r_1+r_2-2}}\sqrt{\frac{1}{r_1}+\frac{1}{r_2}}} \sim t(r_1+r_2-2)$$

In the case that the design is not balanced and the sample size is not large[7], we should worry about the hypothesis of equal-variance. However, as [Scheffé] pointed out[8], the Bartlett's test is highly sensitive to normality. Hence, Scheffe proposed an alternative ratio test of this equal-variance hypothesis.

Unlike Bartlett's likelihood ratio test, Scheffe's test is "based on an approximate test based on the analysis on variance of the logarithms of the sample variances[9]". So we have to stick to ANOVA when we referred to this ratio test. Before we reach the procedure of the ratio test, next two sections are basic results from sampling theory for readers unfamiliar with them.

---

[3] [Bartlett]:"By exact tests will be meant tests depending on a known probability distribution; that is, independent of irrelevant unknown parameters. It is assumed that no certain information is available on the range of these extra parameters, so that their complete elimination from our distributions is desirable."

[4] $s_i^2 = \frac{1}{r_i-1}\sum_{j=1}^{r_i}(x_{ij}-\bar{x_i})^2$

[5] What is more, we must make *all* group sizes tend to large at the same time. We cannot just let one group has many observations but leave other groups with small sizes. In that case H.Tucker pointed out that [Tucker] Sec3.3(pp.40) provided a nonparametric example of what happens in that case.

[6] But that will not differ greatly if the sample size is small. Actually we can see that the difference in the degrees of freedom is $\frac{(\frac{s_1^2}{r}+\frac{s_2^2}{r})^2}{\left[(\frac{s_1^2}{r})^2+(\frac{s_2^2}{r})^2\right]\frac{1}{r-1}} - (2r-2) = (r-1)\frac{(\frac{s_1^2}{r}+\frac{s_2^2}{r})^2-2\left[(\frac{s_1^2}{r})^2+(\frac{s_2^2}{r})^2\right]}{\left[(\frac{s_1^2}{r})^2+(\frac{s_2^2}{r})^2\right]} = (r-1)\frac{\frac{-(s_1^2-s_2^2)^2}{r^2}}{\left[(\frac{s_1^2}{r})^2+(\frac{s_2^2}{r})^2\right]}$, when the number of groups is small($r \geq 1$) and the difference $\left(s_1^2-s_2^2\right)$ is small, this should be near zero. And when the sample size is large, this should be near zero either since $\frac{r-1}{r^2} \to 0$. However, if the sample sizes are not equal, we should worry that this difference in degree of freedom is large.

[7] a.k.a The BehrensFisher problem for normal populations.

[8] [Scheffé]pp.83, footnote 28

[9] [Scheffé]pp.83

1.2. **Sample Mean and Variance in Finite Population Sampling.** Now we consider a sample $\{x_1, \cdots x_n\}$ of size n from the finite population of size $N \geq n$ with at least first to fourth moments finite.

Proposition 1. *([Tucker] pp.32-33)(Moments of sample mean for finite population)If $M = \frac{1}{n}\sum_k x_k, x_k\,(k = 1, \cdots, n)$ are independent with mean of $\mu_1 < \infty$ and variance of σ, then $E(M) = \mu_1, Var(M) = \frac{1}{n}\left(1 - \frac{n}{N}\right)\sigma^2$.*

Proof. ([Tucker]) (1)$E(M) = \frac{1}{n}\sum_{1\leq k_1<k_2<\cdots<k_n\leq N} \frac{\sum_i x_{k_i}}{\begin{pmatrix} N \\ n \end{pmatrix}}$ since every arrangement has the same probability to be drawn.

$$\begin{aligned}
E(M) &= \frac{1}{n}E\left(\sum_{1\leq k_1<k_2<\cdots<k_n\leq N}\begin{pmatrix} N-1 \\ n-1 \end{pmatrix}\frac{\sum_{i=1}^N x_i}{\begin{pmatrix} N \\ n \end{pmatrix}}\right) \\
&= \frac{1}{n}E\left(\frac{n}{N}\sum_i x_i\right) \\
&= \mu_1
\end{aligned}$$

(2)$Var(M) = E(M^2) - (E(M))^2$

$$\begin{aligned}
Var(M) &= E(M^2) - \mu_1^2 \\
E(M^2) &= \frac{1}{n^2}E\left(\left(\sum_{i=1}^n x_i\right)^2\right) - \mu_1^2 \\
&= \frac{1}{n^2}\left[E\left(\left(\sum_{i=1}^n x_i^2\right) + \left(\sum_{i\neq j} x_i x_j\right)\right)\right] - \mu_1^2 \\
E\left(\sum_{i\neq j} x_i x_j\right) &= \sum_{1\leq k_1<k_2<\cdots<k_n\leq N}\frac{\sum_i x_{k_i}x_{k_j}}{\begin{pmatrix} N \\ n \end{pmatrix}} \\
&= \frac{2}{\begin{pmatrix} N \\ n \end{pmatrix}}\sum_{1\leq k_1<k_2<\cdots<k_n\leq N}\sum_{i<j} x_{k_i}x_{k_j} \\
&= \frac{2\begin{pmatrix} N-2 \\ n-2 \end{pmatrix}}{\begin{pmatrix} N \\ n \end{pmatrix}}\sum_{i<j} x_{k_i}x_{k_j} \\
&= \frac{n(n-1)}{N(N-1)}\sum_{i\neq j} x_i x_j \\
E\left(\sum_{i=1}^n x_i^2\right) &= \sum_{1\leq k_1<k_2<\cdots<k_n\leq N}\frac{\sum_i x_{k_i}^2}{\begin{pmatrix} N \\ n \end{pmatrix}} \\
&= \frac{\begin{pmatrix} N-1 \\ n-1 \end{pmatrix}}{\begin{pmatrix} N \\ n \end{pmatrix}}\sum_{j=1}^N x_j^2 \\
&= \frac{n}{N}\sum_{i=1}^N x_i^2 \\
Var(M) &= \frac{1}{n}E\left[\frac{n}{N}\sum_{i=1}^N x_i^2 + \frac{n(n-1)}{N(N-1)}\sum_{i\neq j} x_i x_j\right] - \mu_1^2 \\
&= \frac{1}{n}\left[\frac{n}{N}\sum_{i=1}^N E(x_i^2) + \frac{n(n-1)}{N(N-1)}\sum_{i\neq j} E(x_i x_j)\right] - \mu_1^2 \\
&= \frac{1}{n}\left(1 - \frac{n}{N}\right)\sigma^2
\end{aligned}$$

□

Lemma 2. *([Scheffé] pp.255)If $a = \sum_m a_m, S = \sum_k a_k(x_k - \sum_l \frac{a_l x_l}{a})^2, x_k$ are independent random variables with $E(x_k) = 0, Var(x_k) = \sigma_k^2 \geq 0$, then*

$E(S)=\sum_k a_k\sigma_k^2-\frac{1}{a}\sum_k a_k^2\sigma_k^2, Var(S)=\sum_k a_k^2\left(\gamma_{2,k}+2\right)\sigma_k^4-\frac{2}{a}\sum_k a_k^3\left(\gamma_{2,k}+2\right)\sigma_k^4+\frac{2}{a^2}\sum_k a_k\gamma_{2,k}\sigma_k^4+\frac{2}{a^2}\left(\sum_k a_k^2\sigma_k^2\right)^2$, where $\gamma_{2,k}=\frac{E(x_k^4)}{\sigma_k^4}-3$(kurtosis of x_k's distribution)[10]

Proof. The proof is a lengthy calculation, I did it for the interested readers. Those who are willing to accept this may just skip this like what Scheffe did in his book.

(1)$E(S)=\sum_k a_k\sigma_k^2-\frac{1}{a}\sum_k a_k^2\sigma_k^2$

$$\begin{aligned}
E(S) &= E\left(\sum_k a_k(x_k-\sum_l \frac{a_l x_l}{a})^2\right)\\
&= \sum_k a_k E(x_k-\sum_l \frac{a_l x_l}{a})^2\\
&= \sum_k a_k E(x_k^2)-2\sum_k a_k E(\sum_l \frac{a_l x_l x_k}{a})+\sum_k a_k E(\sum_l \frac{a_l x_l}{a})^2\\
&= \sum_k a_k\left(\sigma_k^2-0^2\right)-2\sum_l \frac{a_l^2(\sigma_l^2-0^2)}{a}+\sum_k a_k\left[Var(\sum_l \frac{a_l x_l}{a})+\left(E(\sum_l \frac{a_l x_l}{a})\right)^2\right]\\
&= \sum_k a_k\sigma_k^2-\frac{2}{a}\sum_k a_k^2\sigma_k^2+\sum_k a_k\left[\sum_l\left(\frac{a_l}{a}\right)^2\sigma_l^2+0^2\right]\\
&= \sum_k a_k\sigma_k^2-\frac{2}{a}\sum_k a_k^2\sigma_k^2+\frac{1}{a}\sum_k a_k^2\sigma_k^2\\
&= \sum_k a_k\sigma_k^2-\frac{1}{a}\sum_k a_k^2\sigma_k^2
\end{aligned}$$

(2)$Var(S)=\sum_k a_k^2\left(\gamma_{2,k}+2\right)\sigma_k^4-\frac{2}{a}\sum_k a_k^3\left(\gamma_{2,k}+2\right)\sigma_k^4+\frac{2}{a^2}\sum_k a_k\gamma_{2,k}\sigma_k^4+\frac{2}{a^2}\left(\sum_k a_k^2\sigma_k^2\right)^2$

$$\begin{aligned}
Var(S) &= E(S^2)-\left[E(S)\right]^2\\
&= E\left[\sum_k a_k^2(x_k-\sum_l \frac{a_l x_l}{a})^4\right]-\left[\sum_k a_k\sigma_k^2-\frac{1}{a}\sum_k a_k^2\sigma_k^2\right]^2\\
&= \sum_k a_k^2 E(x_k-\sum_l \frac{a_l x_l}{a})^4-\left[\sum_k a_k\sigma_k^2-\frac{1}{a}\sum_k a_k^2\sigma_k^2\right]^2
\end{aligned}$$

It remains to calculate the term $\sum_k a_k^2 E(x_k-\sum_l \frac{a_l x_l}{a})^4$, we expand it into:

---

[10]Actually, this can be derived using combinatorial methods like Prop.1 too. Rewrite the sample variance into the form $S^2=\frac{1}{\binom{n}{2}}\sum_{\{i,j\}}\frac{1}{2}(X_i-X_j)^2, E[(X_i-X_j)^2/2]=\sigma^2$. Thus, the $Var(S^2)=\left(\frac{1}{\binom{n}{2}}\sum_{\{i,j\}}\left[\frac{1}{2}(X_i-X_j)^2-\sigma^2\right]\right)^2$ but we still have to calculate the term $\left[\frac{1}{2}(X_i-X_j)^2-\sigma^2\right]\left[\frac{1}{2}(X_k-X_\ell)^2-\sigma^2\right]$ using a similar discussion as I show below in the proof. i.e. Consider the $|\{i,j\}\cap\{k,l\}|$

If $|\{i,j\}\cap\{k,l\}|=0$, there're $\left(\begin{array}{c}n\\4\end{array}\right)$ such terms, and due to independence their expectations are all zeros.

If $|\{i,j\}\cap\{k,l\}|=1$, there're $\left(\begin{array}{c}4\\2\end{array}\right)\left(\begin{array}{c}n\\3\end{array}\right)$ such terms, and $\left[\frac{1}{2}(X_i-X_j)^2-\sigma^2\right]\left[\frac{1}{2}(X_k-X_j)^2-\sigma^2\right]=\frac{1}{4}(X_i-X_j)^2(X_k-X_j)^2-\sigma^2\left[\frac{1}{2}(X_i-X_j)^2+\frac{1}{2}(X_k-X_j)^2\right]+\sigma^4=\frac{1}{4}\left(3X_i^2X_j^2-4X_iX_j^3+X_j^4\right)-\sigma^2\left(2X_i^2-X_iX_j\right)+\sigma^4$(The key is that subscripts are irrelevant due to the independence.). Taking expectation over this expression gives $\frac{\mu_{4-\sigma^4}}{4}$.

If $|\{i,j\}\cap\{k,l\}|=2$, there're $\left(\begin{array}{c}n\\2\end{array}\right)$ such terms, and $\left[\frac{1}{2}(X_i-X_j)^2-\sigma^2\right]^2=\frac{1}{4}(X_i-X_j)^4-(X_i-X_j)^2\sigma^2+\sigma^4=\frac{1}{4}\left(2X_i^4-8X_i^3X_j+6X_i^2X_j^2\right)-\left(2X_i^2-2X_iX_j\right)\sigma^2+\sigma^4$Taking expectation over this expression gives $\frac{\mu_{4+\sigma^4}}{2}$.

Summing up all these terms yields the same result. This method is pointed out by `ByronSchmuland(http://math.stackexchange.com/users/940/byron-schmuland)@http://math.stackexchange.com/q/73080(version:2011-10-23)`

$$
\begin{aligned}
& \textstyle\sum_k a_k^2 E(x_k - \sum_l \frac{a_l x_l}{a})^4 \\
= \quad & \textstyle\sum_k a_k^2 E[x_k^4 - 4x_k^3 \left(\sum_l \frac{a_l x_l}{a}\right) + \\
& \textstyle 6x_k^2 \left(\sum_l \frac{a_l x_l}{a}\right)^2 - 4x_k \left(\sum_l \frac{a_l x_l}{a}\right)^3 + \left(\sum_l \frac{a_l x_l}{a}\right)^4] \\
& \begin{cases} E(x_k^4) & = E(x_l^4) = (\gamma_{2,k}+3)\,\sigma_k^4 \\ E(x_k^3 x_l) & = E(x_k^3)E(x_l) = E(x_k^3)\cdot 0 = 0 \\ E(x_k^2 x_l^2) & = E(x_k^2)E(x_l^2) = \sigma_k^2\sigma_l^2 \\ E(x_k x_l^3) & = E(x_k)E(x_l^3) = 0\cdot E(x_l^3) = 0 \end{cases} \\
= \quad & \textstyle\sum_k a_k^2 [\left[(\gamma_{2,k}+3)\,\sigma_k^4\right] - [0] \\
& \textstyle +6\left[E(x_k^2 \sum_l \left(\frac{a_l}{a}\right)^2 x_l^2)\right] - 4\left[E(x_k \sum_l \begin{pmatrix} 3 \\ 2 \end{pmatrix} \frac{a_k}{a} x_k \cdot \begin{pmatrix} 1 \\ 1 \end{pmatrix} \left(\frac{a_l}{a}\right)^2 x_l^2)\right] \\
& \textstyle + \left[E\left(\sum_l \left(\frac{a_l}{a}\right)^4 x_l^4\right) + E(\sum_l \begin{pmatrix} 4 \\ 2 \end{pmatrix} \left(\frac{a_k}{a}\right)^2 x_k^2 \cdot \begin{pmatrix} 2 \\ 2 \end{pmatrix} \left(\frac{a_l}{a}\right)^2 x_l^2)\right]] \\
= \quad & \textstyle\sum_k a_k^2 [(\gamma_{2,k}+3)\,\sigma_k^4 + 6\cdot\sum_l \left(\frac{a_l}{a}\right)^2 \sigma_k^2\sigma_l^2 - 12\cdot\sum_l \left(\frac{a_k}{a}\right)\left(\frac{a_l}{a}\right)^2 \sigma_k^2\sigma_l^2 \\
& \textstyle + \sum_l \left(\frac{a_l}{a}\right)^4 (\gamma_{2,k}+3)\,\sigma_k^4 + 6\cdot\sum_l \left(\frac{a_k}{a}\right)^2\left(\frac{a_l}{a}\right)^2 \sigma_k^2\sigma_l^2] \\
& \begin{cases} \sum_k a_k^2[(\gamma_{2,k}+3)\,\sigma_k^4] = & \sum_k a_k^2 (\gamma_{2,k}+3)\,\sigma_k^4 \\ \sum_k a_k^2[6\cdot\sum_l \left(\frac{a_l}{a}\right)^2 \sigma_k^2\sigma_l^2] = & 6\sum_{k,l}\left(\frac{a_k a_l}{a}\right)^2 \sigma_k^2\sigma_l^2 \\ \sum_k a_k^2[-12\cdot\sum_l \left(\frac{a_k}{a}\right)\left(\frac{a_l}{a}\right)^2 \sigma_k^2\sigma_l^2] = & -12\sum_{k,l}\frac{a_k^3 a_l^2}{a^3}\sigma_k^2\sigma_l^2 \\ \sum_k a_k^2[\sum_l \left(\frac{a_l}{a}\right)^4 (\gamma_{2,k}+3)\,\sigma_k^4] = & \sum_{k,l}\frac{a_k^2 a_l^4}{a^4}(\gamma_{2,k}+3)\,\sigma_k^4 \\ \sum_k a_k^2[6\cdot\sum_l \left(\frac{a_k}{a}\right)^2\left(\frac{a_l}{a}\right)^2 \sigma_k^2\sigma_l^2] = & 6\sum_{k,l}\frac{a_k^4 a_l^2}{a^4}\sigma_k^2\sigma_l^2 \end{cases} \\
& \begin{cases} \sum_k a_k^2[(\gamma_{2,k}+3)\,\sigma_k^4 + = & \sum_k a_k^2 (\gamma_{2,k}+3)\,\sigma_k^4 + \\ \sum_l \left(\frac{a_l}{a}\right)^4 (\gamma_{2,k}+3)\,\sigma_k^4] & \sum_{k,l}\frac{a_k^2 a_l^4}{a^4}(\gamma_{2,k}+3)\,\sigma_k^4 \\ \sum_k a_k^2[6\cdot\sum_l \left(\frac{a_l}{a}\right)^2 \sigma_k^2\sigma_l^2 = & 6\sum_{k,l}\left(\frac{a_k a_l}{a}\right)^2 \sigma_k^2\sigma_l^2 \\ -12\cdot\sum_l \left(\frac{a_k}{a}\right)\left(\frac{a_l}{a}\right)^2 \sigma_k^2\sigma_l^2 & -12\sum_{k,l}\frac{a_k^3 a_l^2}{a^3}\sigma_k^2\sigma_l^2 \\ +6\cdot\sum_l \left(\frac{a_k}{a}\right)^2\left(\frac{a_l}{a}\right)^2 \sigma_k^2\sigma_l^2] & +6\sum_{k,l}\frac{a_k^4 a_l^2}{a^4}\sigma_k^2\sigma_l^2 \end{cases} \\
= \quad & \textstyle\sum_k a_k^2 (\gamma_{2,k}+2)\,\sigma_k^4 - \frac{2}{a}\sum_k a_k^3 (\gamma_{2,k}+2)\,\sigma_k^4 + \frac{2}{a^2}\sum_k a_k \gamma_{2,k}\sigma_k^4 + \frac{2}{a^2}\left(\sum_k a_k^2\sigma_k^2\right)^2
\end{aligned}
$$

□

Corollary 3. *If* $a = \sum_m a_m, S = \sum_k a_k(x_k - \sum_l \frac{a_l x_l}{a})^2, x_k \sim N(0,\sigma^2)$, *then* $E(S) = \sum_k a_k\sigma^2 - \frac{1}{a}\sum_k a_k^2\sigma^2, Var(S) = 2\sum_k a_k^2\sigma^4 - \frac{4}{a}\sum_k a_k^3\sigma^4 + \frac{2}{a^2}\left(\sum_k a_k^2\sigma^2\right)^2$.

Corollary 4. *(Moments of sample variance for finite population[11])If* $S^2 = \frac{1}{n-1}\sum_k(x_k - \sum_l \frac{x_l}{n})^2, x_k\ (k = 1,\cdots,n)$ *are independent with mean of* $\mu_1 < \infty$ *and variance of* σ, *then* $E(S^2) = \sigma^2, Var(S^2) = \sigma^4(\frac{2}{n-1} + \frac{\gamma_2}{n}), \gamma_2 = \frac{\mu_4}{\sigma^4} - 3$.

Corollary 5. *If* $a = n, a_m = 1, S^2 = \frac{1}{n-1}\sum_k(x_k - \sum_l \frac{x_l}{n})^2, x_k$ *are independent with mean of* $\mu_1 < \infty$ *and variance of* σ, *then we let* $y = log(S^2), E(y) \sim log\sigma^2, Var(y) \sim (\frac{2}{n-1} + \frac{\gamma_2}{n}), \gamma_2 = \frac{\mu_4}{\sigma^4} - 3$.

Proof. Use the fact that under the transformation g if $X \sim g(Y)$, then $M_X \sim g(M_Y), S_X^2 \sim \left[g'(M_Y)S_Y^2\right]$ and the theorem above. The fact can be proven by following arguments:

[11]Let $a = n, a_m = 1$

TABLE 1. Comparison of Sample Means M from Sampling Models

	$E(M)$	$Var(M)$
Finite Population[15][16]	μ_1	$\frac{1}{n}\left(1-\frac{n}{N}\right)\sigma^2$
Infinite Population[17]	μ_1	$\frac{1}{n}\sigma^2$

TABLE 2. Comparison of Sample Variances S^2 from Sampling Models

	$E(S^2)$	$Var(S^2)$
Finite Population[18]	σ^2	$\sigma^4(\frac{2}{n-1}+\frac{\gamma_2}{n}), \gamma_2=\frac{\mu_4}{\sigma^4}-3$
Infinite Population[19]	σ^2	$\sigma^4(\frac{2}{n-1}+\frac{\gamma_2}{n}), \gamma_2=\frac{\mu_4}{\sigma^4}-3$

Consider the case of a linear transformation g first. $M_X = \frac{1}{n}(\sum_i X_i) = \frac{1}{n}(\sum_i g(Y_i)) = g(\frac{1}{n}\sum_i Y_i) = g(M_Y)$. $S_X^2 = \frac{1}{n-1}(\sum_i (X_i - M_X)^2) = \frac{1}{n-1}(\sum_i (g(Y_i) - g(M_Y))^2) \approx \frac{1}{n-1} g'(M_Y)^2 \sum_i (Y_i - M_Y)^2 = g'(M_Y)^2 S_Y^2$. And use the derivative transformation to approximate the transformation g leaving a Peano residue. □

If we want to test the hypothesis $H : \sigma_1^2 = \cdots = \sigma_\nu^2 = \sigma^2 \iff H' : log\sigma_1^2 = \cdots = log\sigma_\nu^2 = log\sigma^2$, then we must make the assumption that $Var(S^2) = (\frac{2}{n-1} + \frac{\gamma_2}{n})$ is approximately a constant θ since it does not vary much with the sample size n[12]. Scheffe derived an F-distribution using this approximate assumption.

So loosely speaking, the Scheffe's F-test is actually testing variances using the approximately non-sensitivity of Var(Variance) in a finite sampling. And we can also see that the sample variance is a less sensitive statistic than the sample mean. This phenomena is not occasional, the higher moment we are investigating, the less sensitive with respect to the sample size and outliers it is. Therefore we might tend to use such a statistic when we want to catch some extreme unusual characteristic like unequal variances.

1.3. **Sample Mean and Variance in Infinite Population Sampling.** Although the lemma is proven in the setting of finite population sampling, the readers might sometimes interested in the result in For completeness, I also provide the infinite population sampling version of the lemma below[13][14]:

2. THE MAX-MIN RATIO TEST IN ANOVA

2.1. **Scheffe's F-test on the Variances.** Scheffe's assumption is imposed for two-way-like in pp. 84 of [Scheffé]. It compared different populations ' variances and take the inner treatment group structure in each of the population into consideration.

A proper way of thinking Scheffe's test is $y_{ij} = logS_{ij}^2$ where S_{ij}^2 is the sample variance of the sample from the jth population in the ith sets, inside these sets the sample variances are thought to be equal. Which is to say, now we move to a higher

[12] A reasonable understanding is that this decreasing approximately like $o(\frac{1}{n})$, which is approximately mild when $n \geq 3$(This number is chosen by looking at the derivative of $\frac{1}{n}$, which is $-\frac{1}{n^2}$). So actually we might want a larger group size in order to make this ratio test robust.

[13] All these could be found in [Cochran] Sec 2.4, in a more systematic but less elementary way.

[14] Such a comparison also highlighted the fact that the sample variance is somehow less sensitive to the size of the sampling population N.(Not the sample size)

level and treat populations(to be compared) as samples. We proceed the F-test on this new population. Following illustration might be helpful.

$$\begin{array}{lllll} Pop_1: & x_{11} & \cdots & x_{1r_1} & s_1^2 \\ Pop_2: & x_{21} & \cdots & x_{2r_2} & s_2^2 \\ Pop_3: & x_{31} & \cdots & x_{3r_3} & s_3^2 \\ Pop_4: & x_{41} & \cdots & x_{4r_4} & s_4^2 \\ Pop_5: & x_{51} & \cdots & x_{5r_5} & s_5^2 \end{array} \quad If \; \overset{H_0}{\Longrightarrow} TRUE \quad \begin{array}{lll} Group1: & Pop_1 & Pop_2 \\ & s_{1,1}^2 = s_1^2 & s_{1,2}^2 = s_2^2 \\ Group2: & Pop_3 & Pop_4 \\ & s_{2,1}^2 = s_3^2 & s_{2,2}^2 = s_4^2 \\ Group3: & Pop_5 & \\ & s_{3,1}^2 = s_5^2 & \end{array}$$

$$\Longrightarrow \begin{array}{lll} Pop_1^* & s_{1,1}^2 & s_{1,2}^2 \\ Pop_2^* & s_{2,1}^2 & s_{2,2}^2 \\ Pop_3^* & s_{3,1}^2 & \end{array} \Longrightarrow \begin{array}{lll} Pop_1' & log_{1,1}^2 & logs_{1,2}^2 \\ Pop_2' & logs_{2,1}^2 & logs_{2,2}^2 \\ Pop_3' & logs_{3,1}^2 & \end{array}$$

Scheffe's method is to carry out F-test on the Pop_I^*. Under the null hypothesis, we are actually testing $H': s_1^2 = s_2^2, s_3^2 = s_4^2, s_5^2$. But Scheffe's F-test simultaneously gives us a multi-comparison:

> "We now see that the origin lies inside the confidence ellipsoid, and hence the F-test will accept H, if and only if $|h'\hat{\psi}| \le \sqrt{h'M^{-1}h}$ is satisfied for all h, because the origin must then lie between all pairs of parallel planes of support...Whenever a hypothesis H is rejected by the F-test we can investigate the different estimable functions in L[20] to find out which ones are responsible for rejecting H."[Scheffé]pp.70-71

2.2. **The Ratio Test of Equal-variance Hypothesis.** The ratio test is a simplified version for Scheffe's F-test for equal-variance hypothesis. The ratio test is based on the fact that $|\hat{y_i}. - \hat{y_j}.| \le \lambda \Longleftrightarrow exp(-\lambda) \le \frac{\sigma_i^2}{\sigma_j^2} \le exp(\lambda)$. Suppose now that we are still testing for one-way layout with sample variances σ_k^2, the null hypothesis is of course $H: \sigma_k^2 = \sigma^2, k = 1, 2, \cdots I$. In the above reasoning, we can see that we are considering I hypotheses and thus separate all the variances into I new populations. A usual ratio test[21] is proposed by rejecting H if $\frac{max_i s_i^2}{min_i s_i^2} \ge 3$. The cut-off number $C = 3$ is sometimes replaced by $e \approx 2.7$ for numerical simplification reasons.

Theorem 6. *The maximal ratio* $\frac{max_i \sigma_i^2}{min_i \sigma_i^2} \le C$ *means* $max_{i,j}|y_i - y_j| \le logC, \forall i, j$, *which means the probability of occurrence of such a sample while* H *is true is no more than* $F_{I-1,n-I}(f \le (\frac{logC}{MSE \cdot \frac{I}{2}}))$.

Proof. This means that $msd \le logC$, but the $msd = F_{\alpha*,I-1,n-I} MSE \cdot \frac{I}{2}$(F is a lower quantile). Hence it is easy to see that $F_{\alpha*,I-1,n-I} \le (\frac{logC}{MSE \cdot \frac{I}{2}}), \alpha* \le F_{I-1,n-I}(f \le (\frac{logC}{MSE \cdot \frac{I}{2}}))$. If the value of $I \cdot MSE$ is actually large, much larger than 2, then the upper bound is close to zero. □

[20] L is the linear space spanned by estimable functions, in this specific case, it is $span\{s_1^2, \cdots, s_5^2\}$.See[Scheffé] pp.70

[21] like [Dean&Voss] pp. 112-113

For the special case $\frac{max_i s_i^2}{min_i s_i^2} \geq 3$, we know that $F_{\nu-1,n-\nu}(f \geq (\frac{log3}{MSE \cdot \frac{\nu}{2}}))$ for ν treatment groups in an one-way layout. So the ratio test is somehow an upper bound for the Scheffe's F-test carried out on general hypothesis.

2.3. **Ratio Test in R.** The typical grammar should be

```
#Suppose there are 4 treatment groups
#Max-min Variance Ratio
vars<-c(var(trmt.0),var(trmt.1),var(trmt.2),var(trmt.3))
ratio<-max(vars)/min(vars)
ratio
#Bartlett's Test
bartlett.test(trmt.0, trmt.1)
bartlett.test(trmt.0, trmt.2)
bartlett.test(trmt.0, trmt.3)
bartlett.test(trmt.1, trmt.2)
bartlett.test(trmt.1, trmt.3)
bartlett.test(trmt.2, trmt.3)
```

References

[Dean&Voss] A.Dean, D.Voss, Design and Analysis of Experiments, Springer, 1999

[Scheffé] H.Scheffé, The Analysis of Variance, John Wiley & Sons, 1959

[Bartlett] M.S.Bartlett, Properties of sufficiency and statistical tests, Proceedings of the Royal Statistical Society (Series A), Vol. 160, No. 901 (May, 1937), pp. 268282

[Tucker] H.G.Tucker, A Short Course in Nonparametric Inference(Lecture Notes), University of California (Irvine), 1984

[Cochran] W.Cochran, Sampling Techniques, 3ed, John Wiley&Sons, 1977

[Casella&Berger] G.Casella and O.Berger, Statistical Inference, 2ed, John Wiley&Sons, 2003