Appendix B

Proofs of Lemmas in Chapter 5 and Chapter 7

Proof of Lemma 5.2:

The contamination scheme starts within a treatment of ranks by changing in succession each 1 to a *t*, then each 2 to a *t*, etc, with *t* the highest possible rank. Let G_g be the set or group of all *m* that changes a *g* to *t*, *g* = 1, 2,…, *t*-1. Thus, G_1 = {*m*: *m* = 1, 2,…, *n*}, G_2 = {*m*: *m* = *n*+1, *n*+2,…, 2*n*}, etc. It is easy to see that for a given set G_g, the lowest *m* in the set is $m_l = n(g-1)+1$, and the highest *m* in the set is $m_u = ng$. So, we can say that

$$m_l \le m \le m_u \text{, or}$$

$$n(g-1)+1 \le m \le ng \text{ .}$$

Rearranging terms, we see that

$$ng \le m+n-1 \le ng+n-1 \text{, or}$$

$$g \le (m+n-1)/n \le g+1-1/n \text{ .}$$

Which implies that

$$g = [(m+n-1)/n]_{gif} \text{ ,}$$

since $(m+n-1)/n$ is between *g* and a number slightly less than *g*. (For *m* = 0, we define *g* = 1 and not *g* = 0 to make the formula correct when there exists no contamination).

Define $\overline{R}$ as the average rank sum, the rank sum for each treatment at *m* = 0. Let $R_{.1}$ correspond to the rank sum for the treatment being contaminated. For each contaminant *m*, $m \in G_g$, (*t*-*g*) is being added to $R_{.1}$. At m_u, $m_u \in G_g$, $R_{.1}$ has increased *n*(*t*-*g*). For $m \ne m_u$, $m \in G_g$, an adjustment of (*t*-*g*) needs to be made to $R_{.1}$ based on how far *m* is from $m_u = ng$, which can be written as $(ng-m)(t-g)$. Thus the squared rank sum can be written as

$$R_{.1}^2 = \left(\overline{R} + \sum_{i=1}^{g} n(t-i) - (t-g)(ng-m) \right)^2 \text{ .}$$

Now for each *m*, $m \in G_1$, (*t*-1) rank sums are decreasing by 1. For each *m*, $m \in G_2$, (*t*-2) rank sums are decreasing by 1. In general, within group *g*, (*t*-*g*) rank sums have decreased a total of *m* after *m* contaminations, which can be written as $(t-g)(\overline{R}-m)$, and the (*t*-*g*) squared rank sums are expressed as $R_{.2}^2 = (t-g)(\overline{R}-m)^2$. This leaves (*t*-1) - (*t*-*g*) = *g*-1 rank sums unaccounted. For each *m*, $m \in G_2$, one rank sum has decreased as much as possible and this occurred at $m_u, m_u \in G_1$. Thus, that rank sum decreased a total of ($\overline{R}-n$), since there are *n* contaminants within a group. At each *m*, $m \in G_3$, another rank sum has decreased as much as possible and this occurred at $m_u, m_u \in G_2$, and that rank sum decreased a total of ($\overline{R}-2n$). In general, for a given G_g, (*g*-1) rank sums have sequentially decreased to their limits of ($\overline{R}-n$), ($\overline{R}-2n$), …, ($\overline{R}-(g-1)n$). Those squared rank sums are expressed as $R_{.3}^2 = \sum_{i=1}^{g-1}(R-in)^2$.

Finally, $R_{.1}^2 + R_{.2}^2 + R_{.3}^2 = \sum_{j=1}^{t} R_{.j}$, and this completes the proof.

Proof of Lemma 5.4:

For $m \le g^2$,

Assume without loss of generality that the highest ranks are in the first treatment, the second highest in the second treatment, and so on. Also assume that the contamination scheme starts by contaminating the higher ranks, then the lower ranks. Within any set G_g, we are contaminating g treatments of the higher ranks and g treatments of the lower ranks, and the t-2g remaining rank sums are adjusted according which ranks are being contaminated (higher or lower). Here, we are contaminating the higher ranks. We define m^* in words as the ordered contaminant in set G_g, given $m \le g^2$. That is, for g = 1 and m = 1, m^*= 1. For g = 2 and m = 3, m^*= 1. For a general G_g, m^*= 1 when $m = g^2 - g + 1$. Replacing the 1 by m^* gives $m = g^2 - g + m^*$ or $m^* = g - g^2 + m$.

Starting with the first term, when contaminating the higher ranks within set G_g, g columns are adjusted downward one at a time by a total of t-g. It is easy to see that $g^2 - m$ is how many rank sums of the higher ranks that have not been adjusted down by t-g. Since they have not been adjusted, their sum is equal to $t[t-(g-1)]$ or $t^2 - tg + t$. Thus, those squared rank sums can be expressed as $(g^2 - m)(t^2 - tg + t)^2$.

Now, there exists m^* of the g treatments containing the higher rank sums being contaminated that have been adjusted down by t-g. Their previous rank sum was $t^2 - tg + t$, and after subtracting t-g, we get a rank sum equal to $t^2 - tg + g$. Those squared rank sums can now be expressed as $m^*(t^2 - tg + g)^2$.

For the t-2g treatments not being contaminated, their rank sums are all equal to their original starting value at $m = 0$. For each contamination of the higher ranks in G_g, an increase of one occurs in the rank sum. This also can be written as an increase of m^* for each rank sum from their starting value, which can be formulated as $\sum_{i=1}^{t-2g} t(g+i)+m^*$. Thus the squared rank sums for the non-contaminated treatments are expressed as $\sum_{i=1}^{t-2g}(t(g+i)+m^*)^2$.

For the last g treatments, their starting rank sum value in G_g, is tg. Again, an adjustment of one occurs for every contamination within G_g, and more specifically, an adjustment of m^*. Therefore the squared rank sums are represented by $g(tg+m^*)^2$. Adding all terms together gives the sum of the squared rank sums for the m^{th} contamination where $m \le g^2$.

For $m > g^2$,

Here, we are contaminating the treatments with the lower ranks. We define m^{**} in words as the number of contaminants remaining in set G_g, after a contaminant m. In general, for a group G_g, $m^{**} = g + g^2 - m$. Starting with the first term, we see that when contaminating the treatments with the lower ranks within set G_g, that the rank sums of g treatments are adjusted upward one at a time by a total of t-g. And within the set G_g, there exists $m - g^2$ rank sums of the lower ranks that have been adjusted up by t-g. Since they have been adjusted, their sum is equal to $t(g+1)$ or $t + tg$. Thus, those squared rank sums can be expressed as $(m-g^2)(t+tg)^2$.

Now, there remains m^{**} of the *g* treatments with lower ranks sums being contaminated in set G_g that have not been adjusted up by *t*-*g*. Their rank sum can be expressed as an adjusted rank sum of $t + tg$ minus the adjustment of $t - g$. After subtracting *t*-*g*, we get a rank sum equal to $tg + g$, and those squared rank sums can now be expressed as $m^{**}(tg + g)^2$.

For the *t*-2*g* treatments not being contaminated, their rank sums are all equal to their original starting value at $m = 0$. If there exists contaminants remaining in G_g, this will be indicated by m^{**}. Thus an increase of m^{**} is needed for each rank sum from their starting value, which can be formulated as $\sum_{i=1}^{t-2g}(t(g+i)+m^{**})$, and the squared rank sums for the non-contaminated treatments is $\sum_{i=1}^{t-2g}(t(g+i)+m^{**})^2$.

For the last *g* treatments, their ending rank sum value in set G_g is *t*(*t*-*g*). Again, an increase of one occurs for every remaining contaminant within set G_g, and more specifically, an increase of m^{**}. Therefore the squared rank sums are represented by $g(t(t-g)+m^{**})^2$.

Adding all terms together gives the sum of the squared rank sums for the m^{th} contamination where $m > g^2$, and this completes the proof.

Proof of Lemma 5.5:

It can be easily seen that the upper limit of the ordered contaminants, m_u, for a given set G_g is $m_u = n(g^2 + g)$, and the upper limit for the set G_{g-1}, m_l, is $m_l = n((g-1)^2 + (g-1)) = n(g^2 - g)$. So, we can say that

$$m_l < m \le m_u \text{, or } n(g^2 - g) < m \le n(g^2 + g)$$

$$\text{or } (g^2 - g) < m/n \le (g^2 + g).$$

By completing the square on each side of the inequality, we obtain

$$g^2 - g + .25 < m/n < g^2 + g + .25 \text{, or}$$

$$(g - .5)^2 < m/n < (g + .5)^2.$$

Note that by completing the square, the upper limit now becomes a strict inequality since $g \in Z^+$, and because $g \in Z^+$, the lower limit still stays a strict inequality. Taking the square root gives us

$$g - .5 < \sqrt{m/n} < g + .5 \text{, or}$$

$$g < \sqrt{m/n} + .5 < g + 1.$$

Therefore, we can say that

$$g = \left\lfloor \sqrt{m/n} + .5 \right\rfloor_{gif},$$

and this completes the proof.

Proof of Lemma 7.1:

For simplicity, we work with the sum of squared deviations, equal to $K^{-1}X^2_{BM}$. Assume that the ranks are in a Latin Square configuration. Define $R_{.j}$ as the sum of scores for the j^{th} treatment. At *m* = 0, $R_{.j} = \mu \, \forall j,\ j = 1,...,t$, where μ is the average sum of scores. The optimal way to increase the test statistic is to contaminate one treatment, say treatment *j*, and change all the ranks that correspond to a score $r_{R_{ij}} = 1$, say, to $r_{R_{ij}} = 0$. (Without loss of generality, assume that the sum of scores for the contaminated treatment is being decreased.) So, at each *m*, $R_{.j} - \mu = m$, and the contribution from treatment *j* to $K^{-1}X^2_{BM}$ is m^2 [1].

Now, for each contaminant in treatment *j* that changes $r_{R_{ij}} = 1$ to $r_{R_{ij}} = 0$, there exists a treatment j' such that $r_{R_{ij'}} = 0$ changes to $r_{R_{ij'}} = 1$ and these treatments differ for each contaminant *m* in treatment *j*. That is, there is a set of unique treatments $\{j'_i\}, i = 1,...m$, such that $R_{.j'_i} - \mu = 1$. So, at each *m*, the contribution from treatments j'_i to $K^{-1}X^2_{BM}$ is $\sum_{i=1}^{m}(R_{.j'_i} - \mu)^2 = \sum_{i=1}^{m} 1^2 = m$ [2].

Since the ranks are in a Latin square arrangement there exists one and only one treatment $j'' \in \{j'\}$ such that a contaminant placed in j'' will additionally increase its sum of scores and still decrease the sum of scores in treatment *j*. That is, instead of contaminating treatment *j* and obtaining an additional treatment j', such that $R_{.j'_i} - \mu = 1$, contaminate treatment j'' such that $R_{.j}$ is still decreasing and $R_{.j''} - \mu = 2$. This increase in the contribution of treatment j'' to $K^{-1}X^2_{BM}$ is $2^2 - 1^2 = 3$ and the loss of the contribution of an additional treatment j' = 1. This implies that the placement of the contaminant in treatment j'' adds an additional increase of 3 – 1 = 2 [3]. Adding [1] + [2] + [3] yields $K^{-1}X^2_{BM} = m^2 + m + 2$.

Finally, this equation is exact for all $t/2$ possible contaminants in treatment j. For $m > t/2$, this equation becomes an upper bound since it *assumes* that for $m > t/2$, the contamination continues in treatment j, which obviously it cannot. And this completes the proof.

Proof of Lemma 7.2:

For simplicity, we work with the sum of squared deviations, equal to $K^{-1}X^2_{BM}$. Define $R_{.j}$ as the sum of scores for the j^{th} treatment. At $m = 0$, $R_{.j} = \mu \; \forall j,\; j = 1,...,t$, where μ is the average sum of scores. The optimal way to increase the test statistic is to contaminate one treatment, say treatment *j*, and change all the ranks that correspond to a score $r_{R_{ij}} = 1$, say, to $r_{R_{ij}} = 0$. (Without loss of generality, assume that the sum of scores for the contaminated treatment is being decreased.) So, at each *m*, $R_{.j} - \mu = m$, and the contribution from treatment *j* to $K^{-1}X^2_{BM}$ is m^2.

Now, optimally, there exists a single treatment j' such that for *every* contaminant *m* in *j*, the scores from treatment j' are such that $r_{R_{ij'}} = 0$ changes to $r_{R_{ij'}} = 1$. So, at each *m*, $R_{.j'} - \mu = m$, and the contribution from treatment j' to $K^{-1}X^2_{BM}$ is m^2. Thus $K^{-1}X^2_{BM} = (R_{.j} - \mu)^2 + (R_{.j'} - \mu)^2 = m^2 + m^2 = 2m^2$. If there exists a treatment $j'' \neq j'$ such that $r_{R_{ij''}} = 0$ changes to $r_{R_{ij''}} = 1$, then $K^{-1}X^2_{BM} \leq 2m^2$ since this is not optimal. And this completes the proof.

Proof of Lemma 7.3:

For simplicity, we work with the sum of squared deviations, equal to $K^{-1}X^2_{BM}$. Define $\mu = nt/2$ as the mean sum of scores. Since all of the highest ranks are associated with one treatment, the next highest with another, etc., then there exists *t*/2 sums of scores equal to *nt* and *t*/2 sums of scores equal to zero. At each *m*, one of the *t*/2 higher sums of scores is decreasing by 1, and one of the *t*/2 lower sums of scores is increasing by 1, and it is the sums of scores farthest in magnitude from μ that are changing. Now, at $m = gt/2$, *g* = 1, 2,…, each sum of scores has increased or decreased by *g*. So, by rearranging terms, we can represent the total change from the original sum of scores for a given treatment as

$$g = \left[\frac{2m-1}{t}+1\right]_{gif},$$

where '*gif*' is the greatest integer function.

Define G_g as the set of contaminants that will yield a change in *g* from the original sums of scores. For values of $m \in G_g$, there exist $m^* = m-(g-1)t/2$ sums of scores that have been contaminated. Thus, their contribution to $K^{-1}X^2_{BM}$ can be expressed as $S_1 = m^*(nt-g-\mu)^2$ for the higher sums of scores and $S_2 = m^*(g-\mu)^2$ for the lower sums of scores. Also for values of $m \in G_g$, there exist $t/2-m^*$ sums of scores that have not been contaminated. Their contribution to $K^{-1}X^2_{BM}$ can be expressed as $S_3 = m^*(nt-g+1-\mu)^2$ for the higher sums of scores and $S_2 = m^*(g-1-\mu)^2$ for the lower sums of scores. Thus, the total sum contribution to $K^{-1}X^2_{BM}$ can be expressed as $S_1+S_2+S_3+S_4$. After substituting in the value of $\mu = nt/2$ and expanding on the quadratic terms, a little algebra will show that $K^{-1}X^2_{BM} = S_1 + S_2 + S_3 + S_4$ can be simplified to

$$K^{-1}X^2_{BM} = \left[\frac{t}{4}(nt - 2g + 2)^2 + m^*(4g - 2nt - 2)\right]$$

or

$$X^2_{BM} = K\left[\frac{t}{4}(nt - 2g + 2)^2 + m^*(4g - 2nt - 2)\right]$$

and this completes the proof.

Proof of Lemma 7.4:

For simplicity, we work with the sum of squared deviations, equal to $K^{-1}X^2_{BM}$. Define $\mu = nt(t-1)/2t$ as the mean sum of scores. Since all of the highest ranks are associated with one treatment, the next highest with another, etc., then there exists (*t*-1)/2 sums of scores equal to *nt* and (*t*+1)/2 sums of scores equal to zero. At each *m*, one of the (*t*-1)/2 higher sums of scores is decreasing by 1, and one of the (*t*+1)/2 lower sums of scores is increasing by 1, and it is the sums of scores farthest in magnitude from μ that are changing. . Now, at $m = g_1(t+1)/2$, g_1 = 1, 2,…, each of the lower sum of scores has increased by g_1. Similarly, at $m = g_2(t-1)/2$, g_2 = 1, 2,…, each of the higher sum of scores has decreased by g_2. By rearranging terms, we can represent the total change from the original sum of scores for a given treatment as

$$g_1 = \left[\frac{2m-1}{t+1}+1\right]_{gif}, \text{ and } g_2 = \left[\frac{2m-1}{t-1}+1\right]_{gif},$$

where '*gif*' is the greatest integer function, g_1 is the total increase for each treatment with the lower sums of scores, and g_2 is the total decrease for each treatment of the higher sums of scores.

Define G_{g_1} as the set of contaminants that will yield a change of g_1 from the original lower sums of scores. For values of $m \in G_{g_1}$, there exist $m^* = m-(g-1)(t+1)/2$ sums of scores that have been contaminated. Thus, their contribution to $K^{-1}X^2_{BM}$ can be expressed as $S_1 = m^*(g_1-\mu)^2$ for the lower sums of scores. Also for $m \in G_{g_1}$, there are $t/2 - m^*$ lower sums of scores that have not been contaminated their contribution to $K^{-1}X^2_{BM}$ can be expressed as $S_2 = (t/2 - m^*)(g_1 - 1 - \mu)^2$.

Now, define G_{g_2} as the set of contaminants that will yield a change of g_2 from the original higher sums of scores. For values of $m \in G_{g_2}$, there exist $m^{**} = m - (g-1)(t-1)/2$ sums of scores that have been contaminated. Thus, their contribution to $K^{-1}X^2_{BM}$ can be expressed as $S_3 = m^{**}(nt - g_2 - \mu)^2$ for the higher sums of scores. Also for $m \in G_{g_2}$, there are $t/2 - m^{**}$ higher sums of scores that have not been contaminated their contribution to $K^{-1}X^2_{BM}$ can be expressed as $S_4 = (t/2 - m^{**})(nt - g_2 - 1 - \mu)^2$. Thus, the total sum contribution to $K^{-1}X^2_{BM}$ can be expressed as $S_1 + S_2 + S_3 + S_4$. After substituting in the value of $\mu = nt(t-1)/2t$ we obtain

$$K^{-1}X^2_{BM} = \begin{bmatrix} m^*\left(g_1 - n(t-1)/2\right)^2 + \left((t+1)/2 - m^*\right)\left(g_1 - 1 - n(t-1)/2\right)^2 \\ \\ + m^{**}\left(nt - g_2 - n(t-1)/2\right)^2 + \left((t-1)/2 - m^{**}\right)\left(nt - g_2 + 1 - n(t-1)/2\right)^2 \end{bmatrix}$$

or

$$X^2_{BM} = K\begin{bmatrix} m^*\left(g_1 - n(t-1)/2\right)^2 + \left((t+1)/2 - m^*\right)\left(g_1 - 1 - n(t-1)/2\right)^2 \\ \\ + m^{**}\left(nt - g_2 - n(t-1)/2\right)^2 + \left((t-1)/2 - m^{**}\right)\left(nt - g_2 + 1 - n(t-1)/2\right)^2 \end{bmatrix},$$

and this completes the proof.