ASSESSING LINEARIZATION VARIANCE ESTIMATORS

Phillip S. Kott
NASS/USDA; S-4801; Washington, DC 20250

Key words: Complex sampling design; model; bias; effective degrees of freedom

1. Introduction

The 1985 Continuing Survey of Food Intakes by Individuals (CSFII) is in many ways a typical complex multivariate survey. The results of a one-day survey of food intakes by women and children has been converted by the Human Nutrition Information Service (HNIS) into a data set that brakes down individual food consumption into measured intakes from among 60 food groups and 28 nutrients (not all mutually exclusive). This paper will focus on a estimated of mean intake (whether for a food group or a nutrient doesn't matter).

HNIS has calculated many different mean intake estimates from the 1985 CSFII. All of them take the form of the ratio estimator. Variances were estimated using the linearization (Taylor series) formula inherent in the SESUDAAN (Holt (1977)) and PC CARP (Fuller et al. (1986)) programs. Under mild conditions this formula returns asymptotically unbiased variance estimates when with replacement sampling is used at the first stage of random selection. Unfortunately, the number of primary sampling units (PSU's) in the CSFII is finite. As a result, the variance estimators are biased. Moreover, they themselves have variances that can be problematic. This paper proposes a means for assessing the bias and variance of a linearization variance estimator. A formula for determining the effective degrees of freedom is introduced that is more reasonable than the naive approach conventionally used. The theory is then applied to the 1985 CSFII. It turns out that the variance of the variance estimator is generally a more troublesome problem than the bias.

The literature contains many articles on variance estimators for complex samples. Rao and Wu (1985), for example, contains a very theoretical, design-based treatment of the bias issue. Empirical works, like Frankel (1971), also address the stability (variance of the variance estimator) and coverage issues.

This paper is constructed in such a way that it begins at a very basic level and does not build onto the previous literature. Nevertheless, the reader should be aware that such literature exists. Two good sources of many of the relevant articles are Rust (1987) and Rao and Wu (1987).

2. The Ratio Estimator

Suppose we are interested in estimating the per individual use of some food or nutrition item among a certain domain of individuals called a <u>cell</u>. Let $k=1, \ldots, K$ denote the strata, M_k be the population cell count in stratum k, and Y_k be total use of the item of interest by stratum k cell members. Let $Y = \Sigma Y_k$, and $M = \Sigma M_k$. The value $R = Y/M$ is what we want to estimate.

For simplicity, only the two sampled PSU's per stratum case will be treated here. It is not difficult, however, to extend the analysis to more general situations (see the appendix. Let y_{k1} and y_{k2} be the design unbiased estimates of Y_k derived from each of the two PSU's sampled from k, and let m_{k1} and m_{k2} be analogously defined estimates of M_k. The nearly design unbiased estimator for R that HNIS uses is the ratio:

$$r = \frac{\Sigma(y_{k1} + y_{k2})}{\Sigma(m_{k1} + m_{k2})} .$$

The design variance (more precisely, the design mean squared error) of r is (approximately)

$$E[(r-R)^2] = E\{[\frac{\sum^K \sum^2 (y_{kj} - Rm_{kj})}{\Sigma \Sigma m_{kj}}]^2\} \approx \frac{\sum^K [\mathrm{var}(d_{k1}) + \mathrm{var}(d_{k2})]}{4M^2}, \quad (1)$$

where $d_{kj} = y_{kj} - Rm_{kj}$. This makes use of the fact that all the PSU were independently drawn, so that all the d_{kj} are independent under the sampling design.

If M and R were known, then a design unbiased estimator for the right hand side of (1) would be

$$\frac{\sum^K (d_{k1} - d_{k2})^2}{4M^2} ,$$

since $E[(d_{k1} - d_{k2})^2] = \mathrm{var}(d_{k1}) + \mathrm{var}(d_{k2})$. Of course, M and R are not known and have to be estimated from the sample.

The estimators for M and R that are used in the linearization formula are $m = \Sigma \Sigma m_{kj}/2$ and r. The resultant variance estimator is

$$v_D = \frac{\sum^{K}(d_{k1}' - d_{k2}')^2}{4m^2}, \qquad (2)$$

where $d_{kj}' = y_{kj} - rm_{kj}$. It can be shown to be nearly design unbiased when the coefficients of variation of m and r are small.

3. A Model-Based Bias Analysis

One way to assess the damage caused by using r and m in place of M and R in the calculation of v_D is to assume a model and evaluate the model bias of v_D as an estimator of the model variance of r. The simplest model is

$$y_{kj} = \mu m_{kj} + e_{kj}, \qquad (3)$$

where the e_{kj} are independent random variables with mean zero. We will assume that $\epsilon(R) = \mu$ and $\epsilon[(R-\mu)^2] \approx 0$, where ϵ denotes model expectation. This means that the population is so large that the model parameter, μ, and the finite population value, R, are virtually identical. In addition, we will assume that the population of PSU's is so large than no PSU is sampled more than once.

The model in (3) is somewhat simplistic and may be doomed to failure (most critically, e_{k1} and e_{k2} can be correlated). Nevertheless, it is adequate for our purpose, which is presently to demonstrate the potential for negative bias in v_D.

Given the model in (3), r is unbiased. The model variance of r is (approximately)

$$\epsilon[(r-\mu)^2] = \epsilon\{[\frac{\sum^{K}\sum^{2}(\mu m_{kj} + e_{kj})}{\sum\sum m_{kj}} - \mu]^2\}$$

$$= \frac{1}{4m^2}\sum\sum\epsilon(e_{kj}^2). \qquad (4)$$

Before we can evaluate the v_D, we need to re-express d_{kj}' as

$$d_{kj}' = y_{kj} - rm_{kj}$$

$$= y_{kj} - \frac{\sum^{K}\sum^{2} y_{k'j'}}{\sum\sum m_{k'j'}} m_{kj}$$

$$= e_{kj} - \frac{\sum\sum e_{k'j'}}{\sum\sum m_{k'j'}} m_{kj}.$$

The model expectation of $(d_{k1}' - d_{k2}')^2$ is then (after much manipulation)

$$\epsilon[(d_{k1}'-d_{k2}')^2] = \epsilon(e_{k1}^2) + \epsilon(e_{k2}^2)$$

$$- [\epsilon(e_{k1}^2) - \epsilon(e_{k2}^2)](m_{k1} - m_{k2})/m \qquad (5)$$

$$+ \frac{1}{4m^2}\sum_{k'=1}^{K}\sum_{j'=1}^{2}\epsilon(e_{k'j'}^2)(m_{k1} - m_{k2})^2.$$

By putting together (2), (4), and (5), the model bias of v_D can be seen to equal

$$\epsilon\{v_D-\epsilon[(r-\mu)^2]\} =$$

$$- \frac{1}{4m^3}\sum^{K}[\epsilon(e_{k1}^2) - \epsilon(e_{k2}^2)](m_{k1} - m_{k2})$$

$$+ \frac{1}{16m^4}\sum_{k=1}^{K}\sum_{k'=1}^{K}\sum_{j'=1}^{2}\epsilon(e_{k'j'}^2)(m_{k1} - m_{k2})^2, \qquad (6)$$

which is no simple matter to evaluate. Assuming that the model variances of the e_{kj} are proportional to the m_{kj} helps simplify matters considerably (if all cell members had independent and identically distributed behavior and the sample were self weighting, this rather heroic assumption would be strictly true).

Let $\epsilon(e_{kj}^2) = cm_{kj}$. Consequently,

$$\epsilon\{v_D-\epsilon[(r-\mu)^2]\} = - \frac{c}{8m^3}\sum^{K}(m_{k1} - m_{k2})^2.$$

The relative model bias of v_D (since, from (4), $\epsilon[(r-\mu)^2] = c/(2m)$) is then

$$\text{RMB}(v_D) = - \frac{1}{4m^2}\sum^{K}(m_{k1} - m_{k2})^2$$

$$= - \frac{\sum(m_{k1}-m_{k2})^2}{4(\sum m_k)^2}$$

$$= - \frac{\sum \text{var}(m_k)}{(\sum m_k)^2}. \qquad (7)$$

where $m_k = (m_{k1} + m_{k2})/2$ is the estimated values of M_k, and $\text{var}(m_k) = \sum_j(m_{kj} - m_k)^2/2$ is the estimated value of the design variance of m_k. It can be shown that equation (7) also applies (with an appropriately redefined $\text{var}(m_k)$) when there are $n_k > 2$ sampled PSU's per stratum (see the appendix).

The right hand side of (7) can actually be calculated to determine the relative bias of v_D under the simple model in (3) with $\epsilon(e_{kj}^2) = cm_{kj}$. Even though this is may not be the true measure of the relative bias of v_D, one should nonetheless be wary of using v_D when $\text{RMB}(v_D)$ is not small -- say not 5% or less.

With or without $\epsilon(e_{kj}^2) = cm_{kj}$, equation (6) tells us that the model variance of v_D will have only a small model bias when either: (1) $m_{k1} - m_{k2}$ is relatively small for all k, or (2) there are a large number of PSU's.

4. Stability

We now turn to what in practice may be an even more important issue than bias -- the variance of v_D. Even if v_D were unbiased, it is possible that its instability could impair the usual coverage properties (that is, whether one has 95% confidence that the real R is in the range $r \pm 2\sqrt{v_D}$).

In this section, we will assume not only that the y_{kj} satisfy (3) and are independent with $\epsilon(e_{kj}^2) = cm_{kj}$, but also that they are normally distributed (which makes $\epsilon(e_{kj}^4) = 3c^2 m_{kj}^2$). Since the y_{kj} are inherently weighted aggregates, the normality assumption may not be too unreasonable.

We simplify the exposition further (and only marginally effect the analysis) by assuming $m_k = m_{k1} = m_{k2}$ for all k. As a result, v_D can be expressed as

$$v_D = \frac{\sum^{K} (e_{k1} - e_{k2})^2}{4(\Sigma m_k)^2}$$

The model expectation of v_D is (from (4)) $c/(2\Sigma\, m_k)$. Its model variance is then

$$\epsilon\{[v_D - \epsilon(v_D)]^2\} =$$

$$\epsilon\{[\frac{\sum^{K} (e_{k1}^2 - 2e_{k1}e_{k2} + e_{k2}^2) - \sum^{K} 2cm_k}{4(\Sigma m_k)^2}]^2\}$$

$$= \frac{\sum^{K} \{\sum^{2} \epsilon[(e_{kj}^2 - cm_k)^2] + 4\epsilon(e_{k1}^2 e_{k2}^2)\}}{16(\Sigma m_k)^4}$$

$$= \frac{c^2 \sum^{K} m_k^2}{2(\Sigma m_k)^4},$$

since $\epsilon[(e_{kj}^2 - cm_k)^2] = 2c^2 m_k^2$. As a result, the relative model variance of v_D is simply

$$RMV(v_D) = 2\,\Sigma\, m_k^2/(\Sigma\, m_k)^2. \quad (8)$$

It can be shown that when an n_k is greater than 2, the corresponding m_k^2 in the numerator of (8) gets divided by $n_k - 1$ (see the appendix).

If all the m_k were equal, then the RMV of v_D would be 2/K (this assumes two sampled PSU's per stratum). Note that the conventional variance estimator, S^2, for the mean of K+1 independent normal variates also has a relative variance of 2/K. Consequently, v_D is very similar to a chi-squared random variate with K degrees of freedom.

What if the m_k were not all equal, which is very likely to case? Since a chi-squared random variate with d degrees of freedom has a relative variance of 2/d, it seems reasonble to call 2/RMV the <u>effective degrees of freedom</u> for v_D (since RMV = 2/d implies d = 2/RMV). It can be shown that this number is never greater (but often less) than K.

When constructing confidence intervals based on $\sqrt{v_D}$, it obviously preferable to use a (perhaps interpolated) t-distribution with 2/RMV degrees of freedom rather than a standard normal distribution. This t-interval (in other contexts) has been called a Satterthwaite approximation.

A common, but naive, practice is to treat v_D as if it had K (rather than 2/RMV) degrees of freedom. Consequently, we will call K the <u>nominal</u> degrees of freedom for v_D (for a more general design, it is the number of PSU's minus the number of strata).

5. Application to 1985 CSFII Data

The indicators of the bias and variance of v_D proposed in the text (the right hand sides of (7) and (8), respectively, modified to handle strata with more than two sampled PSU's) have been calculated for published cells from the 1985 one-day CSFII. This was essentially a survey of women from 19 to 50 years of age and children from 1 to 5. There was a total of 84 cells:

six age groups (1-3, 4-5, 19-34, 35-50, 1-5, 19-50),

six age groups x four regions (Northeast, Midwest, South, West),

six age groups x three levels of urbanization (central cities, suburbs, nonmetropolitan),

six age groups x three races (white, black, other),

six age groups x three income levels, (under 131% of poverty, 131-300%, over 300%).

Every mean intake estimate in a particular cell whether for a food group or a nutrient uses the same variance estimation formula and thus has an identical model relative bias and variance (see equations (7) and (8)). Most of the cells have acceptably low indicators of relative bias -- that is, no more than 5%. The 14 exceptions are all six (unpublished) "other race" cells, the three children cells both among blacks and in the Northeast region, and the 4-5 year old children cells in the West and in central cities.

The indicators of the stability of the variance estimators tell another story. With 18 effective degrees of freedom, v_D would have a model coefficient of variation ($\sqrt{RMV}$) of 33% and a t-based 95% confidence interval would be 5% thicker than a normal-based one. Only 36 out of 84 cells have no less than 18 effective degrees of freedom. These are all six age cells both nationally and among whites, all 18 income level cells (three levels of income cross six age groups), five out of six suburban age groups (excluding the 4-5 years old cell), and women from the South in the 19-50 age group.

The cell with the lowest effective degrees of freedom (roughly five) is "other race" ages 4-5. Among published cells, blacks ages 1-3 has roughly six degrees of freedom. Finally, among cells with acceptable biases, Northeasterners ages 19-34 has effectively 7.8 degrees of freedom.

Table 1 displayes the relative model biases (in absolute terms), effective degrees of freedom, and nominal degrees of freedom for linearization variance estimators for all 84 cells.

6. Discussion

In Kott (1989a, 1989b), I proposed adjusting the linearization variance estimator v_D so that it would be unbiased under a model. With the model under discussion, this means replacing v_D by

$$v_D' = \frac{v_D}{1 + RMB(v_D)} . \qquad (9)$$

The adjusted variance estimator in (9) has the same asymptotic design-based property as v_D -- it is a consistent estimator for the design mean sqared error of r. Moreover, unlike v_D, it is a model unbiased estimator for the model variance of r under <u>some</u>, albeit simplistic, model.

The bias (model and otherwise) of v_D has received extensive attention in the literature, as the many references in Kott (1989a) can attest. If we are to take confidence intervals and hypothesis tests for r based on $\sqrt{v_D}$ with any seriousness, however, it is clear that the stability of v_D merits more active theoretical consideration. The empirical shortfallings of confidence intervals based on linearization variance estimators is already well known (see Rust (1987, pp. 42-43)).

My own view is that our ability to construct meaningful confidence intervals based on data from complex samples is questionable (other forms of direct variance estimation face the same stability problem as linearization). If one insists on constructing them, however, then it is far more reasonable to produce Satterthwaite approximate t-intervals using the model-driven effective degrees of freedom developed here rather than t-intervals based on the nominal degrees or freedom, or worse (but certainly not unheard of), z-intervals based on the normal distribution.

Table 1. Relative Model Bias and Effective Degrees of Freedom for Particular Cells

Cell: Ages	Rel. Bias (%)	Effec. DOF	Nominal DOF
All 1-3	0.8	42.6	61
All 4-5	1.4	36.3	61
All 1-5	0.8	42.7	61
All 19-34	0.4	46.7	61
All 35-50	0.3	45.4	61
All 19-50	0.2	49.8	61
Race			
Whites 1-3	0.8	39.9	61
Whites 4-5	1.4	34.5	61
Whites 1-5	0.8	40.5	61
Whites 19-34	0.5	44.5	61
Whites 35-50	0.5	43.5	61
Whites 19-50	0.3	48.1	61
Blacks 1-3	5.9	6.1	61
Blacks 4-5	7.5	8.6	61
Blacks 1-5	5.2	8.0	61
Blacks 19-34	2.9	8.7	61
Blacks 35-50	3.9	11.4	61
Blacks 19-50	2.6	10.3	61
Others 1-3	11.3	11.5	61
Others 4-5	23.8	4.9	61
Others 1-5	8.7	13.4	61
Others 19-34	8.8	8.8	61
Others 35-50	9.0	9.7	61
Others 19-50	7.5	10.8	61
Income			
Low 1-3	2.3	23.9	61
Low 4-5	3.4	19.1	61
Low 1-5	2.3	24.8	61
Low 19-34	1.5	27.1	61
Low 35-50	1.2	28.9	61
Low 19-50	1.0	31.1	61
Middle 1-3	1.8	33.9	61
Middle 4-5	2.7	25.9	61
Middle 1-5	1.6	35.5	61
Middle 19-34	0.8	41.9	61
Middle 35-50	0.8	35.6	61
Middle 19-50	0.6	43.9	61
High 1-3	1.7	20.9	61
High 4-5	4.9	19.1	61
High 1-5	1.8	26.2	61
High 19-34	1.1	34.5	61
High 35-50	0.9	32.0	61
High 19-50	0.6	38.7	61

Table 1. (continued)

Cell: Ages	Rel. Bias (%)	Effec. DOF	Nominal DOF
Region			
NE 1-3	6.3	7.6	12
NE 4-5	11.2	6.6	12
NE 1-5	7.4	7.4	12
NE 19-34	2.5	7.8	12
NE 35-50	1.5	9.1	12
NE 19-50	1.5	8.6	12
MW 1-3	1.7	12.5	15
MW 4-5	3.2	9.7	15
MW 1-5	1.5	12.2	15
MW 19-34	1.0	12.6	15
MW 35-50	2.0	11.3	15
MW 19-50	0.9	12.4	15
SO 1-3	2.5	16.2	21
SO 4-5	2.2	14.4	21
SO 1-5	1.4	17.1	21
SO 19-34	1.0	18.5	21
SO 35-50	0.9	17.6	21
SO 19-50	0.6	19.3	21
WE 1-3	4.2	8.2	13
WE 4-5	7.9	7.2	13
WE 1-5	5.0	8.3	13
WE 19-34	2.0	9.9	13
WE 35-50	1.2	8.6	13
WE 19-50	0.9	10.7	13
Level of Urbanization			
City 1-3	3.5	10.7	17
City 4-5	7.0	9.1	17
City 1-5	3.7	10.4	17
City 19-34	1.2	13.2	17
City 35-50	0.7	10.3	17
City 19-50	0.6	13.7	17
Suburb 1-3	1.6	21.9	29
Suburb 4-5	2.4	18.8	29
Suburb 1-5	1.7	22.2	29
Suburb 19-34	0.9	22.7	29
Suburb 35-50	0.8	23.5	29
Suburb 19-50	0.5	24.3	29
Nonmetr 1-3	2.9	10.8	15
Nonmetr 4-5	3.5	10.4	15
Nonmetr 1-5	1.8	11.6	15
Nonmetr 19-34	1.3	11.3	15
Nonmetr 35-50	1.4	12.9	15
Nonmetr 19-50	1.0	12.5	15

7. References

Holt M.M. (1977): SURREGR: Standard Errors of Regression Coefficients from Sample Survey Data. Research Triangle Institute, Research Triangle Park.

Frankel, M.R. (1971): Inference from Sample Surveys: An Empirical Investigation. Institute for Social Research, University of Michigan, Ann Arbor Michigan.

Fuller, W.A., Kennedy, W., Schnell, D., Sullivan, G., and Park, H. J. (1986): PC CARP. Statistical Laboratory, Iowa State University, Ames, IA.

Kott, P.S. (1989a): Estimating the Conditional Variance of a Design Consistent Regression Estimator. Journal of Statistical Planning and Inference, forthcoming.

Kott, P.S. (1989b): Robust Small Domain Estimation Using Random Effects Modeling. Survey Methodology, forthcoming.

Mood, A.M., Graybill, F.A., and Boes, D.C. (1974): Introduction to the Theory of Statistics (3'rd edition). McGraw-Hill, New York.

Neter, J., Wasserman, W., and Kutner M.H. (1985): Applied Linear Statistical Models (2'nd edition). Irwin, Homewood, IL.

Rao, J.N.K. and Wu, C.F.J. (1985): Inferences from Stratified Samples: Second-Order Analysis of Three Methods for Nonlinear Statistics. Journal of the American Statistical Association, 80, pp. 620-630.

Rao, J.N.K. and Wu, C.F.J. (1987): Methods for Standard Errors and Confidence Intervals from Sample Survey Data: Some Recent Work. Bulletin of the International Statistical Institute, Proceedings of the 46th Session, Volume LII, Book 3, pp. 5-22.

Rust, K. (1987): Practical Problems in Sampling Error Estimation. Bulletin of the International Statistical Institute, Proceedings of the 46th Session, Volume LII, Book 3, pp. 39-56.

APPENDIX

Let stratum k (which may be a collapsed variance stratum) contain $n_k \geq 2$ distinct PSU's. The ratio estimator for R is

$$r = \frac{\Sigma\, y_k}{\Sigma\, m_k},$$

where $z_k = \Sigma_j\, z_{kj}/n_k$ (z can be either y or m).

Assume that $e_{kj} = y_{kj} - \mu m_{kj}$ ($k = 1, \ldots, K$; $j = 1, \ldots, n_k$) are independent random variables (in the model-based sense) satisfying $\epsilon(e_{kj}) = 0$, $\epsilon(e_{kj}^2) = \sigma_{kj}^2$, and $\epsilon(e_{kj}^4) = 3\sigma_{kj}^4$, and $R \approx \mu$. The model variance of r is thus (approximately)

$$\epsilon[(r-\mu)^2] = \frac{\sum^{K} \frac{\sigma_{kj}^2}{n_k^2}}{(\Sigma\, m_k)^2} . \quad \text{(A1)}$$

So that

$$\epsilon[(r-\mu)^2] = \frac{c'}{\Sigma\, m_k}$$

when $\sigma_{kj}^2 = c'n_k m_{kj}$.

The linearization variance estimator is

$$v_D = \frac{\sum_{k=1}^{K} \left\{ \sum_{j=1}^{n_k} \frac{d_{kj}'^2}{n_k(n_k-1)} - \frac{(\sum^{n_k} d_{kj}')^2}{n_k^2(n_k-1)} \right\}}{(\Sigma\, m_k)^2},$$

where $d_{kj}' = y_{kj} - rm_{kj}$. Deriving equation (7) with $\mathrm{var}(\bar{m}_k) = \Sigma_j\, (m_{kj} - m_k)^2/[n_k(n_k-1)]$ and $\sigma_{kj}^2 = c'n_k m_{kj}$ is cumbersome but straightforward.

Now assume

$$v_D \approx \frac{\sum_{k=1}^{K} \left\{ \sum_{j=1}^{n_k} \frac{e_{kj}^2}{n_k(n_k-1)} - \frac{(\sum^{n_k} e_{kj})^2}{n_k^2(n_k-1)} \right\}}{(\Sigma\, m_k)^2}, \qquad \text{(A2)}$$

Equation (9) would follow from (A1), (A2), and $\sigma_{kj}^2 = c'n_k m_{kj}$ if the following lemma is true:

<u>Lemma</u> If $x_1, \ldots x_n$ are independent random variables satisfying $E(x_i) = 0$, $E(x_i^2) = \sigma_i^2$, and $E(x_i^4) = 3\sigma_i^4$, then

$$E\left\{ \left[\frac{\sum^n x_i^2}{n-1} - \frac{(\sum^n x_i)^2}{n(n-1)} - \frac{\sum^n \sigma_i^2}{n} \right]^2 \right\} =$$

$$\frac{2}{(n-1)^2} \left[\left(1 - \frac{2}{n}\right) \sum^n \sigma_i^4 + \frac{1}{n^2} \left(\sum^n \sigma_i^2\right)^2 \right].$$

<u>Proof</u> The key step in the proof of this lemma is observing that

$$\frac{\sum^n x_i^2}{n-1} - \frac{(\sum^n x_i)^2}{n(n-1)} - \frac{\sum^n \sigma_i^2}{n} =$$

$$\frac{\sum_{i=1}^{n} (x_i^2 - \sigma_i^2)}{n} - \frac{2 \sum_{i>j}^{n} x_i x_j}{n(n-1)} . \qquad \text{(A3)}$$

The expectation of the square of the first term on the right hand side of (A3) is $2\Sigma\sigma_i^4/n^2$; the expection of the square of the second term is

$$2\, \frac{(\sum^n \sigma_i^2)^2 - \sum^n \sigma_i^4}{n^2(n-1)^2} ;$$

the cross term has an expectation of zero. Since $(1 - 1/(n-1)^2) = (n^2 - 2n)/(n-1)^2$, the lemma soon follows after rearranging terms.