A generalization of the Coefficient of variation with application to suppression of imprecise estimates

A.C. Singh, M. Westlake, and M. Feder
RTI International, NC

Abstract

In this paper we propose a generalization of the usual coefficient of variation (CV) to address some of the known problems when used in criteria developed to determine suppression of estimates. Some of the problems associated with CV include interpretation when the estimate is near zero, and the inconsistency in the interpretation about precision when computed for different one-to-one monotonic transformations. The proposed measure, termed discrimination coefficient of variation (DCV), generalizes CV using the hypothesis testing ideas involving length of the confidence interval (LCI) and the length of a discrimination interval (LDI). This discrimination interval is used to check whether the sample is large enough or the confidence interval is short enough to discriminate with certain power between the current value and postulated extreme values on either side of the current value. DCV allows for using an exact distribution for computing LCI such as the use of exact binomial when the estimated proportion is very small. This problem arose in the context of NSDUH (National Survey on Drug Use and Health), and the proposed measure is planned for application to NSDUH.

Key Words: Effective sample size; exact binomial; Length of confidence interval; Length of discrimination interval;

1. Introduction

Often as a result of processing a large data set, estimates in bulk are disseminated in the form of tables. To users, interpretation of these estimates could be misleading if no warning is given about their precision. Typically, standard errors are provided in a separate table if the user wants to look them up. However, even with information about $\hat{SE}$ (or se, denotes estimated standard error), some guidance is needed for the user to decide whether the estimate meets a suitable precision threshold. To this end, suppression rules are used by the data producer to decide whether certain estimates should be suppressed or not, or whether they should be published with some qualifiers. It is natural to say that an estimate $\hat{\theta}$ is imprecise if its se($\hat{\theta}$) is too large. How large is too large? To answer it, some standardization of se($\hat{\theta}$) is needed first so that it does not depend on the unit of measurement. One way to do it is to consider the length of the confidence interval (LCI) (which is generally based on se($\hat{\theta}$)) relative to the point estimate. The usual measure of coefficient of variation (CV), defined as $\hat{CV}(\hat{\theta}) = se(\hat{\theta})/\hat{\theta}$, assuming $\hat{\theta}$ >0, is a form of standardized LCI because considering a symmetric Normal two-sided ($1-\alpha$) confidence interval,

$$\hat{\theta} \pm z_{\alpha/2} se(\hat{\theta}) = \hat{\theta}(1 \pm z_{\alpha/2} \hat{CV}(\hat{\theta})) \qquad (1.1)$$

we see that $CV(\hat{\theta})$ is proportional to a version of the standardized LCI. There are several merits of CV as listed below.

(i) It is simple to understand.
(ii) It is proportional to the large sample relative CI length.
(iii) It provides a standardized (or scale free) measure of precision around the mean.
(iv) It allows for comparing two estimates with different means.
(v) It is useful for design or redesign of experiments and deciding about sample allocation.
(vi) It provides caution to users regarding precision of published estimates.

However, there are several concerns about the above standardization provided by CV as listed below. In the following, whenever estimates refer to a proportion, they will be denoted by $\hat{p}$, while $\hat{\theta}$ is used to signify a general parameter estimate

(i) It is not meaningful when $\hat{\theta}$ <0. This concern is only minor since we can easily redefine it as $CV(\hat{\theta}) = se(\hat{\theta})/|\hat{\theta}|$.

(ii) $CV(\hat{\theta})$ is not useful when $\hat{\theta}$ is zero or near zero. For example, if $\hat{\theta}$ denotes a difference between two estimates, it may be close to zero but its SE may not be small; see e.g. Kish (1965). In this case, CV could be high even for large samples. In the case of proportions, CV becomes extreme as $\hat{p}$ gets close to zero even for large but finite n, but the resulting CI may be precise enough to discriminate postulated changes from the current value of p. However, this is not what CV would imply.

(iii) If $\hat{p}$ is 0, then se is zero, and the CV is not defined. It is unable to capture the relevant information in the estimator. However, CI (such as one-sided CI based on exact binomial and not on SE) could still provide meaningful information.

(iv) The CV is not symmetric for estimated proportions.

(v) Often a threshold value c_0 such as .25 or .50 for CV is used to decide if the estimate is precise enough or not. For

example, if CV exceeds c_0, then the estimate is suppressed. For each estimate, setting CV equal to c_0 gives a minimum sample size required for the estimate to be publishable. (With complex surveys, we get effective sample size (neff) because SRS formulas are used for computation.) The threshold could be set so that the suppression rule yields a certain minimum sample size for a selected value of $\hat{p}$, e.g., n of 50 when $\hat{p}$ = 0.1. With this rule, which gives minimum n as $\hat{p}$ varies, the implied LCI for various $\hat{p}$ at the corresponding minimum n turns out to be very wide at .50 and very narrow near 0; see also Folsom (1991). In practice, it is desirable to have a suppression rule that has a direct control on the min n corresponding to low, medium, and high values of $\hat{p}$ such as 1%, 10% and 50%. Note that for the full range of $\hat{p}$, we only need to consider the interval [0, 0.5], as suppression rule for estimates in the complementary part can be derived by symmetry.

(vi) With one-to-one monotonic transformations of estimated proportions $\hat{p}$, such as log($\hat{p}$) or logit($\hat{p}$), the usual interpretation of CV becomes misleading. With the usual CV($\hat{p}$), the effective sample size (neff) as a function of $\hat{p}$ monotonically decreases as $\hat{p}$ increases from 0 to 0.5. This is as expected because CV($\hat{p}$) decreases as neff increases and so one needs a smaller sample size for p near .5 in order to achieve the same threshold for CV(p). However, with CV(log $\hat{p}$), the function neff first decreases and then increases indicating that one would need a bigger sample size near 0.5 which is contrary to what is expected under the usual CV($\hat{p}$); this problem was identified by Chromy (2001), see Fig 2(a). With CV(logit $\hat{p}$), it gets even more extreme because as $\hat{p}$ goes to 0.5, CV(logit $\hat{p}$) goes to infinity for fixed n. Thus different monotonic transformations of $\hat{p}$ give rise to conflicting interpretations of precision for the same sample size although they all should have the same level of precision. This suggests that there might be a problem with the definition of CV itself for the objective in mind when variance depends on the mean parameter.

2. Motivation for an Alternative Measure

In view of the concerns raised above, there is clearly a need for an alternative measure. First we note that CV is related to CI. However, with an ad hoc cut-off point, there is no control on the length of the CI supported by the suppression rule. Also note that the notion of large se for $\hat{\theta}$ to be imprecise is equivalent to the corresponding LCI($\hat{\theta}$) being large. However, the LCI measure which may or may not be based on se is more general than the se measure because there may be situations where LCI based on se is not very useful, for instance, when se is zero. So we prefer to work with LCI. Now to motivate the proposed measure, we introduce the notion of LCI relative to LDI (length of a discrimination interval) where LDI is the distance between two values deemed to be extreme in either direction relative to the current value. It is desired that the precision of the current estimator should be high enough such that it is able to discriminate parameter values at least as extreme as these. So, we should choose the cut-off point such that LCI is sufficiently smaller than the amount of any change we wish to detect. In other words, heuristically speaking, we should

$$\text{suppress } \hat{\theta} \text{ if } LCI(\hat{\theta}) > c \times LDI(\hat{\theta}) \qquad (2.1)$$

where c denotes a fraction to be chosen suitably. The smaller c is, the less is the error in discriminating the given extreme values from the current value, i.e., the rule is correct more often. This interpretation is related to the OC (operating characteristic) curve of a test procedure which turns out to be useful in specifying c more objectively; see Section 3.1. In general, LDI would depend on $\hat{\theta}$.

In specifying LDI, we consider the logit scale because it is easier to justify symmetry of the Normal CI which is convenient in constructing the suppression rule in practice. For example, for $\hat{p}$ =.50, we could take the lower and upper limits (LDL, UDL) for the discrimination interval as (25%, 75%), which corresponds approximately to 0 ± 1.1 in the logit scale, and for $\hat{p}$ =1%, we might want to use (0.1%, 10%) for (LDL,UDL) which corresponds to approximately -4.6 ± 2.3 in the logit scale. This leads to the question of choosing LDI as a function of $\hat{p}$ so that it can be easily specified for arbitrary $\hat{p}$. Clearly LDI should vary with $\hat{p}$. This suggests as a first approximation that LDI(logit $\hat{p}$) should have a term like $\gamma_1 |\log it(\hat{p})|$ where γ_1 denotes a fraction. Also at $\hat{p}$ =0.5, since logit $\hat{p}$ is zero, it suggests that LDI (logit $\hat{p}$) should have an offset or a constant term γ_0. We thus consider a linear function $\gamma_0 + \gamma_1 |\log it(\hat{p})|$ representing a location and scale transformation of logit $\hat{p}$. More generally, we can specify the form of LDI as

$$LDI(\hat{p}) = \gamma_0 + \gamma_1 |\log it(\hat{p})| + \gamma_2 (\log it(\hat{p}))^2 \qquad (2.2)$$

involving three parameters which allow for more flexibility in having a direct control on LDI for selected values (or anchor points) in the $\hat{p}$-scale . The choice of γ-parameters is considered in the next section. Also in the next section we consider how to choose the constant c objectively. It may be of interest to note that using approximately symmetric Normal CI for LCI, the proposed rule based on (2.1) becomes equivalent to the usual rule based on CV(logit $\hat{p}$) when $\gamma_0 = 0, \gamma_1 = -\gamma_2 = 1$ in (2.2). Thus the proposed rule can be viewed as a generalization of the usual CV. It may also be noted that for small $\hat{p}$, it may be better to use the exact binomial distribution to compute lower and upper bounds and then convert them to the logit scale to compute LCI (logit $\hat{p}$). Alternatively, we could work in the p-scale. Note that with the exact binomial, we will need to calculate effective n for simple random sampling which can, of course, be obtained by using deff (design effect) for the estimator under consideration.

3. Discrimination Coefficient of Variation (DCV)- the proposed measure

The DCV of a parameter estimator $\hat{\theta}$ is defined as

$$DCV(\hat{\theta}) = \frac{LCI(\hat{\theta})}{LDI(\hat{\theta})} \qquad (3.1a)$$

The estimator $\hat{\theta}$ is suppressed if DCV($\hat{\theta}$) $> c_0$. For the logit scale, $\hat{\theta}$ is replaced by logit($\hat{p}$). It is interesting to observe that the above definition can be seen as dual to the sample size rule in power analysis (see e.g., Kupper and Hafner, 1989).

3.1 Choice of c_0

For a given estimator $\hat{\theta}$, LCI depends on the confidence coefficient 1-α. However, for it to discriminate with power $1-\beta$ against two-sided alternatives which are at least as far away in either direction such that the total distance between them is given by LDI, it would depend on how small the LCI is as a fraction of LDI. The smaller this fraction is, the higher will be the power. Thus the fraction c can be set such that the estimator is precise enough to detect certain alternatives specified by LDI with power $1-\beta$. To get more insight, consider the usual CV($\hat{\theta}$). The LCI based on Normal CI is $2\, z_{\alpha/2}$ se($\hat{\theta}$), and taking LDI as $2\hat{\theta}$, we get DCV($\hat{\theta}$) as $z_{\alpha/2}$ CV($\hat{\theta}$). Now one way to specify c objectively is to consider the problem of detecting an alternative θ_1 against the current value $\theta_0 < \theta_1$ with power $1-\beta/2$. Using an approximate Normal test (using the statistic $(\hat{\theta}-\theta_0)/se(\hat{\theta})$), it can be shown that to achieve this, se($\hat{\theta}$) must be small enough such that

$$\frac{\text{se}(\hat{\theta})}{(\theta_1-\theta_0)} \le (z_{\alpha/2} + z_{\beta/2})^{-1} \qquad (3.2)$$

Now letting LDI(θ) equal to 2($\theta_1-\theta_0$), (3.2) implies that with the suppression rule DCV($\hat{\theta}$)> $c(\alpha,\beta)$, the estimator is not suppressed if it can detect alternatives at least as far as θ_1 in either direction with power $1-\beta/2$ where

$$c(\alpha,\beta) = z_{\alpha/2}(z_{\alpha/2} + z_{\beta/2})^{-1} \qquad (3.3)$$

In practice, we don't know the true current value θ_0, and so we can estimate it by $\hat{\theta}$ and then a suitable choice of the function LDI($\hat{\theta}$) can be used to specify the distance $2(\theta_1-\theta_0)$ for discrimination. Note that the above specification of $c(\alpha,\beta)$ is quite general and can be applied to define DCV for any estimator under arbitrary but smooth scale transformation as long as the sample is large enough for approximate Normality. We also note that if $\alpha = \beta$, then $c(\alpha,\beta)$ =.50. In practice α is typically set to .05. If β is set as high as .50, then $c(\alpha,\beta) \approx .75$.

3.2. Choice of γ-parameters in LDI

For a given $c(\alpha,\beta)$, and any given $\hat{p}$, the choice of LDI($\hat{p}$) leads to a minimum effective sample size n such that the estimator can be published, i.e., it satisfies the equality LCI($\hat{p}$)= $c(\alpha,\beta)LDI(\hat{p})$. The reason for n being minimum is that LCI($\hat{p}$) is expected to have the natural property of being decreasing as n increases. For example, with normal CI, LCI(logit $\hat{p}$) is approximately $2z_{\alpha/2}se(\hat{p})/\hat{p}(1-\hat{p})$ which decreases with n. Now to specify the three γ-parameters, we need to choose three values of LDI corresponding to the three anchors for $\hat{p}$ (say, low, medium, and high) and then solve for γ's. This essentially sets the values of the effective sample size corresponding to the three chosen anchor points in the $\hat{p}$-scale. Now a simple model such as (2.2) can be used for specifying LDI for arbitrary $\hat{p}$. However, as mentioned earlier, the resulting effective n should have the

monotonicity property as $\hat{p}$ increases from 0 to .5. For this reason, the three anchors are carefully chosen in the range of $\hat{p}$ so that the effective n or neff as a function of $\hat{p}$ has the desirable monotonic decreasing behavior (this holds if the γ-parameters are chosen such that the derivative of $p(1-p)(\gamma_0 + \gamma_1 |\log it(p)| + \gamma_2 (\log it(p))^2)$ with respect to p is >0 over the range $0 < p < 0.5$) and that the resulting effective sample sizes are deemed to be reasonable in practice. If this is not the case, then one can first set effective sample sizes corresponding to anchor values of $\hat{p}$, compute the resulting LDI(logit $\hat{p}$), and then the corresponding (LDL, UDL) to see if they are acceptable extreme values against which the estimator should be able to discriminate with certain power. Thus the process of specifying LDI for selected anchors can be iterative. Note that the alternative of specifying LDI or the effective n separately for each $\hat{p}$ without modeling is not practical.

3.3 Properties of DCV

(i) The usual CV is a special case of DCV.
(ii) DCV is invariant to location and scale transformations.
(iii) DCV is approximately invariant to any 1-1 monotonic continuously differentiable transformation for small LDI.
(iv) It is symmetric for estimated proportions.
(v) Since LDI is bounded away from zero, there is no problem of the denominator being zero.
(vi) With suitable anchors in defining LDI, extreme behavior in tolerance of the suppression rule for proportions near end points can be avoided.
(vii) The conflicting interpretation with log otr logit transformations is no longer there because the there is control on the shape of the minimum neff curve accepted by the rule.

Figures 1(a,b), 2(a,b), and 3(a,b) illustrate respectively the behavior of CV(p), CV(logp), and DCV(logitp).

4. Summary

It was observed that the usual definition of CV was incomplete without an objective specification of the cut-off point in the suppression rule to define adequate precision. The proposed measure DCV corrects CV by considering both level and length of the confidence interval. An interesting , but rather obvious in retrospect, finding was that DCV is analogous to the sample size rule in power analysis.

It may be noted that if $\hat{p}$ is small, then exact binomial using neff can be used to calculate LCI and then the DCV-suppression rule can be defined as before; here for convenience we can use normality assumption to define $c(\alpha, \beta)$ as in (3.2). If $\hat{p}$ is zero, then the rule of three (see e.g., Jovanovic and Levy, 1997) can be used to get an upper bound. In this case DCV(logit $\hat{p}$) is not defined but DCV($\hat{p}$) is.

In many applications such as those for NSDUH, one is dealing with estimates of proportions such as drug use prevalence. However, the underlying distribution is not based on binomial because the data is not a simple random sample. In these situations, one can still use the DCV formulation for the cut-off under the working assumption of binomial. Note that the LCI here will be based on appropriate design-based standard error as it should be.

References

Chromy, J. (2001). Suppression rules for prevalence rates and other estimates. Memo to SAMHSA, May 14.

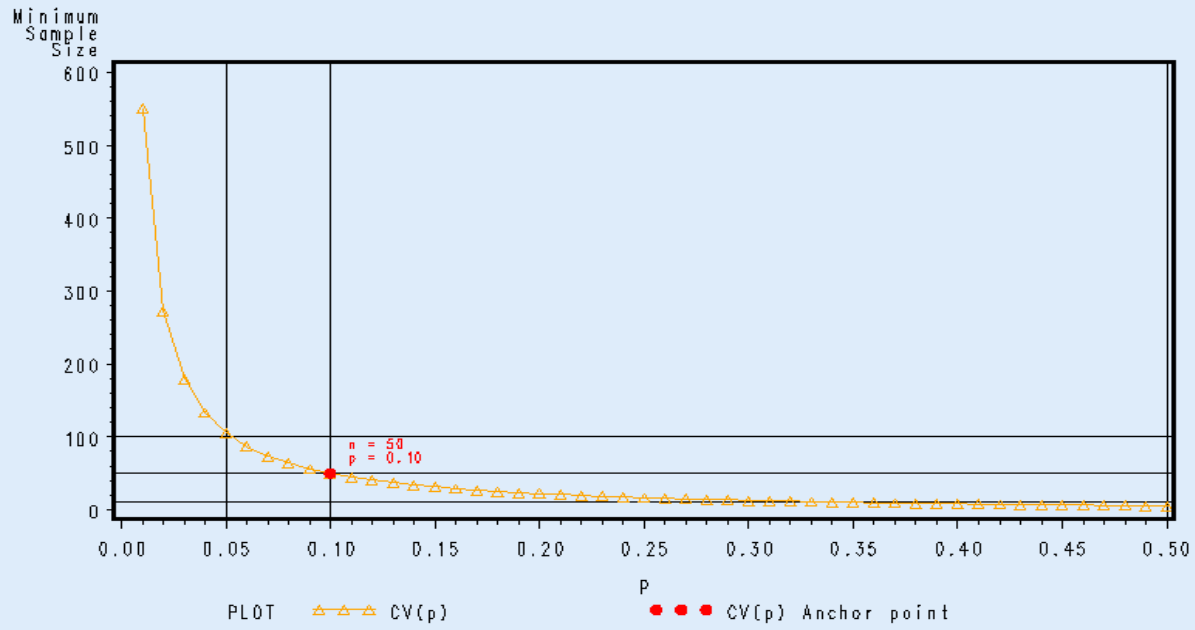
Folsom, R.E., Jr.(1991). A sensible alternative to the 50% suppression rule for NHSDA prevalence estimates. Memorandum to Elizabeth Lambert, SAMHSA, May 30.

Jovanovic, B.D., and Levy, P.S. (1997). A look at the rule of three. *American Statistician*, 51, 137-139.

Kish, L. (1965). *Survey Sampling*, John Wiley and Sons, NY.

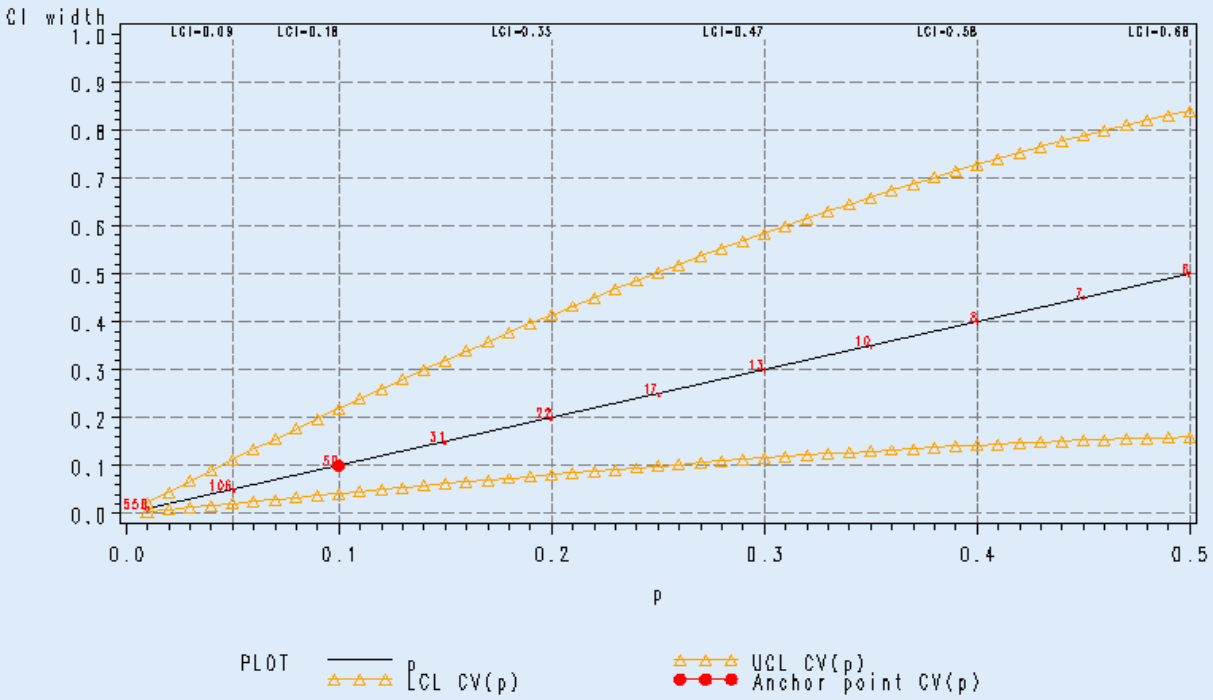
Kupper, L.L., and Hafner, K.B. (1989). How appropriate are popular sample size formulas?, *The American Statistician*, Vol. 43, No 2, 101-105.

Figure 1a: neff vs. p based on CV(p) Suppression Rule



neff based on minimum sample 50 at p=.10
*Note: p restricted to (.01, .50) for display

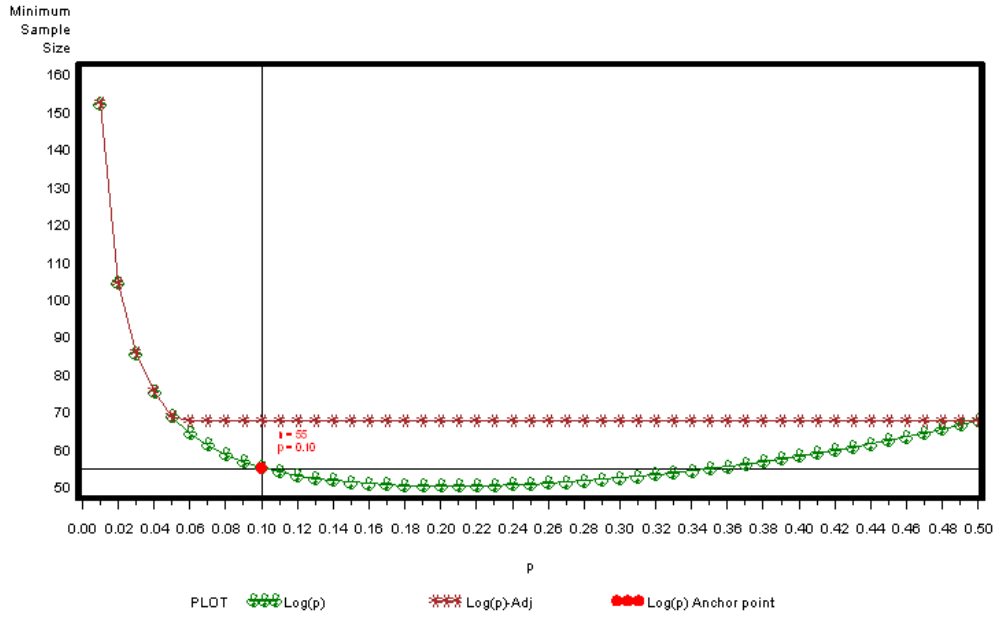
Figure 1b: Max 95% Confidence interval for p based on CV(p) Suppression Rule



neff based on minimum sample 50 at p=.10

Figure 2a: n-eff v/s p (Suppression Rule based on CV(Log(p)) and CV(Log(p))-Adj

where alpha=.05, beta=.50



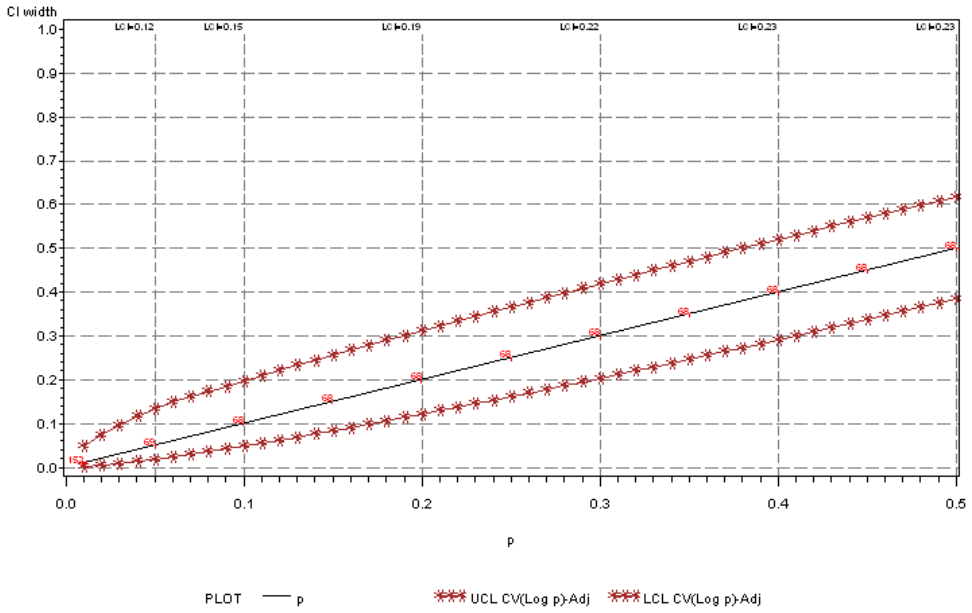
n-eff based on minimum sample of 55 at p=.10

note: Log(p) adjusted = Log(p) except that min n-eff = 68

*Note: p restricted to (.01, .50) for display

Figure 2b: Confidence interval for CV(Log p)-Adj

where alpha=.05, beta=.50

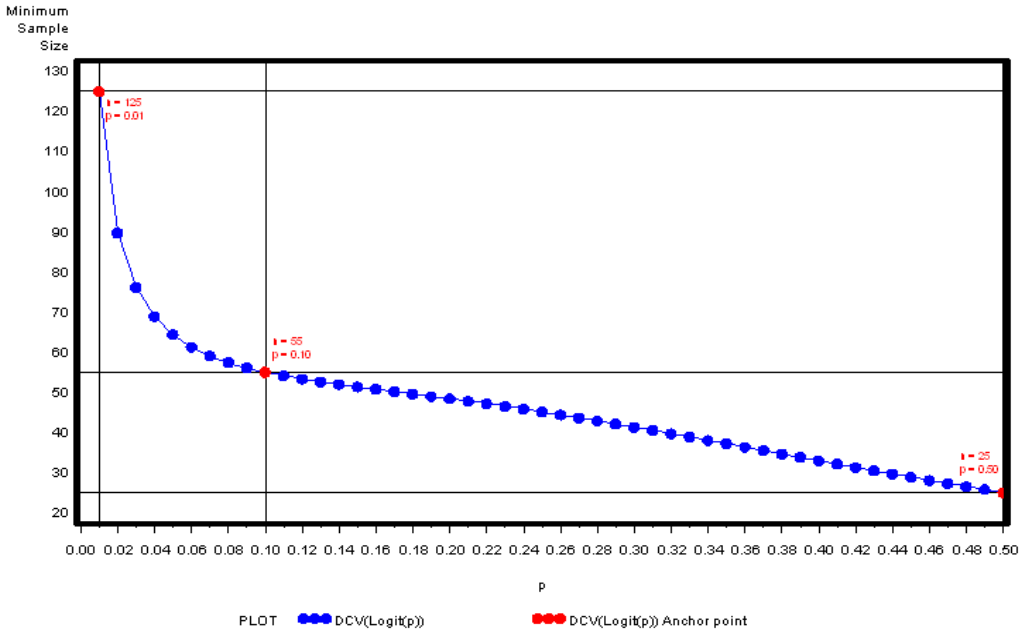


Red label = n-eff for a given value of p

n-eff based on minimum sample 55 at p=.10

Figure 3a: n Effective v/s p Based on DCV(Logit(p))

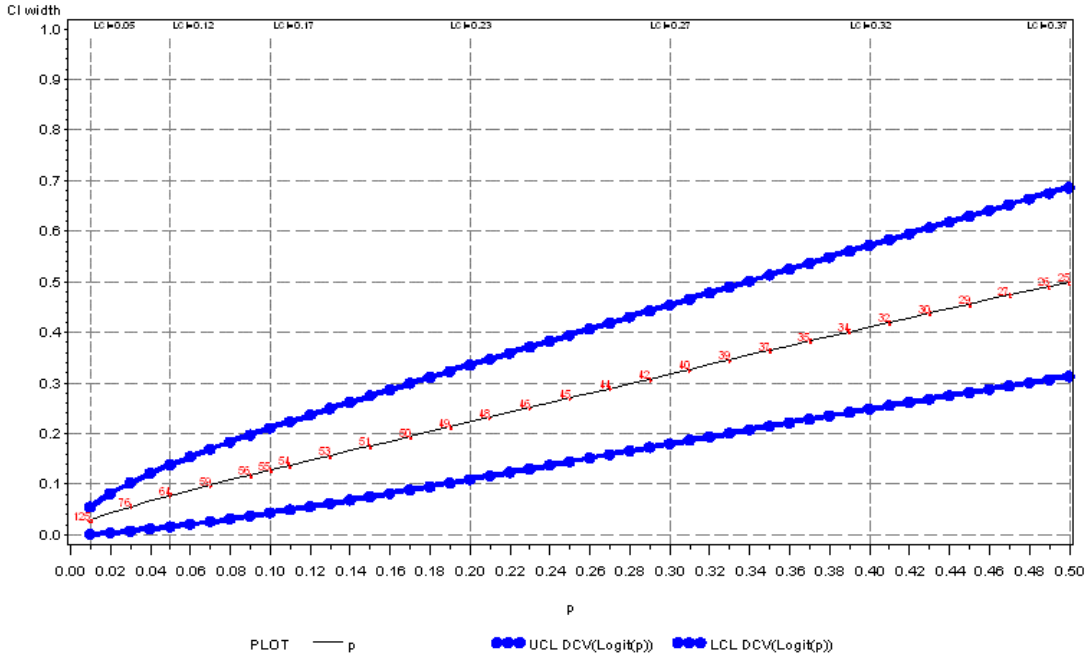
where alpha=.05, beta=.50



Neff based on minimum sample 125 at p=.01, 55 at p=.10, 25 at p=.50 under tightest CI

Figure 3b: Confidence interval for DCV(Logit(p)), alpha = .05, beta = .50

Red label = neff for a given value of p



Neff based on minimum sample 100 at p=.01, 55 at p=.10, 10 at p=.50, under tightest CI