Statistical Research and Applications Branch, NCI, Technical Report #2009-02

AAPC for the Joinpoint Connect-the-Dots Scenario

Michelle Dunn and Joe Zou

When longitudinal data are first collected, there are not enough observations to accurately fit a regression model, much less a joinpoint model. However, presenting the data in an informative way is important. In the Cancer Trends Progress Report, when there are six or fewer observations, a "Connect-the-Dots" plot is presented. In the "Connect-the-Dots" scenario, rather than fitting a stochastic model, a deterministic one is fit which interpolates a line segment between each observation. In other words, for observations $r_1, r_2, \ldots, r_k, r_{k+1}$ at times $t_1, t_2, \ldots, t_k, t_{k+1}$ respectively, k line segments are fit to the data on a log scale, where the first line segment begins at $(t_1, log(r_1))$ and ends at $(t_2, log(r_2))$, the second one connects $(t_2, log(r_2))$ and $(t_3, log(r_3))$, etc.

The observations themselves are realizations of random variables, often from complicated surveys. Using complex survey methodology, an estimate of the variance of the random variable is computed. The observation r_i is a realization of the random variable R_i, a rate which is assumed to be approximately normal,

$$R_i \dot{\sim} Normal(\alpha_i, \gamma_i^2). \tag{1}$$

where $\dot{\sim}$ means "is approximately distributed." An estimate of the mean, $\hat{\alpha}_i$, is given, and the variance $\hat{\gamma}_i^2$ is assumed to be known.

Because the observations are not necessarily equally spaced, the lengths of the line segments will vary. Let weights w_i be proportional to the length of segment i, where

$$\sum_{i=1}^{k} w_i = 1.$$

Then by definition, the Average Annual Percentage Change (AAPC) is

$$\mu = exp\left(\sum_{i=1}^{k} w_i \beta_i\right) - 1 \tag{2}$$

where β_i is the slope for a line segment i. In the "connect-the-dots" scenario, the slope for a segment is

$$\beta_i = \frac{log(R_{i+1}) - log(R_i)}{t_{i+1} - t_i}. \tag{3}$$

A standard way to find a 95% confidence interval for μ is to find its distribution and then to remove 2.5% from each tail. Instead, to be consistent with the way that confidence intervals are found in joinpoint-based AAPCs, a confidence interval will be found for

$$\sum_{i=1}^{k} w_i \beta_i. \tag{4}$$

The confidence interval endpoints (L, U) will then be transformed by $exp(L) - 1$ and $exp(U) - 1$ respectively.

Using (1) and the Delta Method, the following approximation is derived

$$log(R_i) \dot{\sim} Normal(log(\alpha_i), \gamma_i^2/\alpha_i^2)$$

which together with (3) produces

$$\beta_i \dot{\sim} Normal \quad \frac{log(\alpha_{i+1}) - log(\alpha_i)}{t_{i+1} - t_i}, \frac{\gamma_{i+1}^2/\alpha_{i+1}^2 + \gamma_i^2/\alpha_i^2}{(t_{i+1} - t_i)^2} \quad .$$

To find the distribution of (4), first note that by substituting the definition of β_i from (3) and $w_i = (t_{i+1} - t_i)/(t_{k+1} - t_1)$,

$$\sum_{i=1}^{k} w_i \beta_i = \frac{log(R_{k+1}) – log(R_1)}{t_{k+1} - t_1}$$

and thus

$$\sum_{i=1}^{k} w_i \beta_i \dot{\sim} Normal \quad \frac{log(\alpha_{k+1}) - log(\alpha_1)}{t_{k+1} - t_1}, \frac{\gamma_{k+1}^2/\alpha_{k+1}^2 + \gamma_1^2/\alpha_1^2}{(t_{k+1} - t_1)^2} \quad .$$

Using this distribution, a 95% confidence interval for $\sum_{i=1}^{k} w_i \beta_i$ is

$$\left(\frac{log(\hat{\alpha}_{k+1}) - log(\hat{\alpha}_1)}{t_{k+1} - t_1} - 1.96\sqrt{\frac{\hat{\gamma}_{k+1}^2/\hat{\alpha}_{k+1}^2 + \hat{\gamma}_1^2/\hat{\alpha}_1^2}{(t_{k+1} - t_1)^2}}, \frac{log(\hat{\alpha}_{k+1}) - log(\hat{\alpha}_1)}{t_{k+1} - t_1} + 1.96\sqrt{\frac{\hat{\gamma}_{k+1}^2/\hat{\alpha}_{k+1}^2 + \hat{\gamma}_1^2/\hat{\alpha}_1^2}{(t_{k+1} - t_1)^2}} \right).$$

The lower and upper endpoints of this confidence interval are transformed using equation (2) to produce a 95% confidence interval for μ.