Q2 EXAM (2010) ANSWER KEY
QUESTION # 1

1. A study was designed to identify factors related to violent and non-violent infractions among inmates in Swiss prisons. The factors examined include demographic variables, information about the index offense that led to imprisonment, criminal history as well as in-prison behavior.

The sample consists of 318 male prisoners in Switzerland.

The code sheet for these data is:

	Variable	Label	Coding
Demographic	id	Identification number	
	age	Age in years (entry in prison)	Continuous number
	swiss	Swiss national	0 = no; 1 = yes
	civilstatus	Marital status	1 = single 2 = married 3 = divorced 4 = widowed
	illegal	Illegal residance status in Switzerland	0 = no; 1 = yes
Criminal history	nr_con	Number of convictions prior to the index offense	Continuous number
	prior_con	At least one prior conviction prior to the index offense	0 = no; 1 = yes
	viol_rec	Conviction for a violent offense prior to the index offense	0 = no; 1 = yes
Index offense	index_viol	Index offense that let to imprisonment was a violent offense (assault, murder, robbery)	0 = no; 1 = yes
	index_abuse	Index offense was child abuse	0 = no; 1 = yes
	index_sex	Index offense was a sex offense (adult victim) (e.g. rape)	0 = no; 1 = yes
	index_prop	Index offense was property offense	0 = no; 1 = yes
	index_drug	Index offense was drug offense	0 = no; 1 = yes
Imprisonment	time	Time spent in prison at the time of the investigation (month)	Continuous number
	infrac	Number of nonviolent infractions	Continuous number
	violence	Violent infractions during imprisonment	0 = no; 1 = yes

Of interest is to determine the factors that are predictive of which prisoners commit violent infractions during their imprisonment.

The 2×2 table of the outcome variable "violence" vs the categorical variable "swiss" is as follows:

```
. tab violence swiss

   violent |         swiss
 infracton |         0          1 |     Total
-----------+----------------------+----------
         0 |       144         91 |       235
         1 |        60         23 |        83
-----------+----------------------+----------
     Total |       204        114 |       318
```

(a) [5 points] Is there an association between being a Swiss national and committing a violent infraction while incarcerated? What would the coefficient be for the variable "swiss" in a logistic regression model? What would the standard error of the coefficient be? Using your point estimate and standard error, compute a 95% confidence interval for the log odds ratio of the association between "swiss" and "violence".

Ho: No association between being a Swiss national and committing a violent infraction while incarcerated
Ha: There is an association

$$\chi^2 = \frac{n(ad-bc)^2}{(a+c)(b+d)(c+d)(a+b)} = \frac{318(144\times 23-91\times 60)^2}{(204)(114)(83)(235)} = 3.23.\ \textit{[1 point]}$$

Fail to reject Ho. $\chi^2 < \chi^2_{.95}(1) = 3.84$

$$\widehat{OR} = \frac{144\times 23}{91\times 60} = 0.6066$$

$$\hat{\beta}_1 = \ln\left(\widehat{OR}\right) = \ln(0.6066) = -0.4999 \qquad \text{[1 point]}$$

$$\widehat{Var}\left(\hat{\beta}_1\right) = \widehat{Var}\left(\ln(OR)\right) = \frac{1}{144}+\frac{1}{91}+\frac{1}{60}+\frac{1}{23} = .0781 \qquad \text{[1 point]}$$

$$\widehat{SE}\left(\hat{\beta}_1\right) = \widehat{SE}\left(\ln(OR)\right) = \sqrt{.0781} = .2794 \qquad \text{[1 point]}$$

$$95\%\ \text{CI} = -0.4999 \pm 1.96\times .2794 = (-1.0475, .0478) \qquad \text{[1 point]}$$

(b) Assuming violence is the response, carefully write the likelihood function that underlies the logistic regression model corresponding to the 2×2 table shown earlier.

$$l(\beta) = \left[\frac{e^{\beta_0+\beta_1}}{1+e^{\beta_0+\beta_1}}\right]^{23}\left[\frac{1}{1+e^{\beta_0+\beta_1}}\right]^{91}\left[\frac{e^{\beta_0}}{1+e^{\beta_0}}\right]^{60}\left[\frac{1}{1+e^{\beta_0}}\right]^{144} \qquad \text{[3 points]}$$

A model was fit that included the variables: time, swiss, infrac and viol_rec. (Note: In this model, the variable "newtime" divides the variable "time" by 12.)

```
. gen newtime= time/12
. logit violence newtime swiss infrac viol_rec

Iteration 0:   log likelihood = -182.56597
Iteration 1:   log likelihood = -144.96926
Iteration 2:   log likelihood = -144.07517
Iteration 3:   log likelihood =  -144.0726
Iteration 4:   log likelihood =  -144.0726
```

```
Logistic regression                               Number of obs   =        318
                                                  LR chi2(4)      =      76.99
                                                  Prob > chi2     =     0.0000
Log likelihood =  -144.0726                       Pseudo R2       =     0.2108

------------------------------------------------------------------------------
    violence |      Coef.   Std. Err.      z    P>|z|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
     newtime |   .1139039   .0486755     2.34   0.019     .0185016    .2093061
       swiss |  -1.029187   .3772512    -2.73   0.006    -1.768586   -.2897883
      infrac |   .2768106    .051555     5.37   0.000     .1757647    .3778564
    viol_rec |   .6653906   .3684903     1.81   0.071    -.0568372    1.387618
       _cons |  -1.948656   .2329258    -8.37   0.000    -2.405182    -1.49213
------------------------------------------------------------------------------
```

(c) [6 points] This model assumes that the variable "time" is linear in the logit. Carefully explain the implications of that assumption in this model by computing the odds ratio corresponding to a 12 month increase at two different starting points in time (e.g., 12 months and 36 months).

This means that the logit difference (i.e., $\ln(OR)$) is constant over the range of x.

e.g.,

Consider an individual with some arbitrary values of swiss, infrac and viol_rec. The logit for an individual who spends 12 months in prison (i.e., newtime=1) is:

$\text{logit}\ (12) = -1.948656 + .1139039 \times 1 - 1.029187 \times \text{swiss} + .2768106 \times \text{infrac} + .6653906 \times \text{viol\_rec}$ [1 point]

The logit for an individual who spends 24 months in prison (i.e., newtime=2) is:

$\text{logit}\ (24) = -1.948656 + .1139039 \times 2 - 1.029187 \times \text{swiss} + .2768106 \times \text{infrac} + .6653906 \times \text{viol\_rec}$ [1 point]

Hence, the logit difference (i.e., log odds ratio) is

$\text{logit}\ (24) - \text{logit}\ (12) = .1139039$ [1 point]

$e^{.1139} = 1.12$ [1 point]

Similarly, the difference in the logits of a subject who spent 48 months in prison (i.e., newtime=4) as compared to one who spent 36 months in prison (i.e., newtime=3) is:

$\text{logit}\ (48) - \text{logit}\ (36) = .1139039$ [1 point]

Hence the logit difference (and therefore the log odds ratio (and odds ratio)) is constant over the range of x. [1 point]

(d) A number of methods can be used to assess the validity of the assumption that "time" is linear in the logit. Please summarize which methods you might use for this purpose. [2 points]

Quartile Method
Lowess Plots
Fractional polynomials

Suppose you believe that the variable "time" is not linear in the logit. In fact, you believe that the relationship is quadratic rather than linear. As a result, the following model was fit. (Note: In this model, the variable "newtime" divides the variable "time" by 12 and "newtimesq" = "newtime"^{2}.

```
. gen newtimesq= newtime* newtime
. logit violence newtime newtimesq swiss infrac viol_rec

Iteration 0:   log likelihood = -182.56597
Iteration 1:   log likelihood = -140.48048
Iteration 2:   log likelihood = -139.14154
Iteration 3:   log likelihood = -139.13951
Iteration 4:   log likelihood = -139.13951

Logistic regression                               Number of obs   =        318
                                                  LR chi2(5)      =      86.85
                                                  Prob > chi2     =     0.0000
Log likelihood = -139.13951                       Pseudo R2       =     0.2379

------------------------------------------------------------------------------
    violence |      Coef.   Std. Err.      z    P>|z|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
     newtime |   .5424386   .1505048     3.60   0.000     .2474547    .8374226
   newtimesq |  -.0340503   .0115048    -2.96   0.003    -.0565994   -.0115013
       swiss |  -1.197425   .3885879    -3.08   0.002    -1.959043   -.4358066
      infrac |   .2515902   .0522881     4.81   0.000     .1491074    .3540731
    viol_rec |   .6185321   .3776603     1.64   0.101    -.1216685    1.358733
       _cons |  -2.413786   .2937536    -8.22   0.000    -2.989533    -1.83804
------------------------------------------------------------------------------
```

(e) [5 points] Provide a table estimating the odds ratio corresponding to a 12-month increase in time starting at 12 and at 60 months. Interpret these results, commenting on the appropriateness or inappropriateness of the assumption of linearity in the logit for time in this model.

Consider again an individual with some arbitrary values of swiss, infrac and viol_rec. The logit for an individual who spends 12 months in prison (i.e., newtime=1) is:

$$\text{logit}\ (12) = -2.413786 + .5424386 \times 1 - .0340503 \times 1 - 1.197425 \times \text{swiss} + .2515902 \times \text{infrac} + .6185321 \times \text{viol\_rec}$$

The logit for an individual who spends 24 months in prison (i.e., newtime=2) is:

$$\text{logit}\ (24) = -2.413786 + .5424386 \times 2 - .0340503 \times 4 - 1.197425 \times \text{swiss} + .2515902 \times \text{infrac} + .6185321 \times \text{viol\_rec}$$

Hence, the logit difference (i.e., log odds ratio) is

$\text{logit}(24) - \text{logit}(12) = 1\times(.5424386)+3\times(-.0340503) = .4402877$

$e^{.4402877} = 1.553$. [2 points]

Similarly, the difference in the logits of a subject who spent 60 months in prison (i.e., newtime=5) as compared to one who spent 72 months in prison (i.e., newtime=6) is:

$\text{logit}(72) - \text{logit}(60) = 1\times(.5424386)+11\times(-.0340503) = .1678853$

$e^{.1678853} = 1.183$. [2 points]

Hence, in the absence of linearity of the logit, the logit difference (and therefore the log odds ratio (and odds ratio)) is no longer constant over the range of x. [1 point]

After the model on page 2 was run, the variance-covariance matrix was generated as follows:

```
. vce
Covariance matrix of coefficients of logit model

             | violence
        e(V) |    newtime        swiss       infrac     viol_rec        _cons
-------------+-----------------------------------------------------------------
violence     |
     newtime |   .00236931
       swiss |  -.00919989    .14231844
      infrac |  -.00071935     .0001171    .00265791
    viol_rec |    .0009161   -.01637681    .00058747     .1357851
       _cons |  -.00311284   -.01062054   -.00501075   -.02734419    .05425445
```

(f) [4 points] Use this information to show how you would compute a 95% confidence interval for the probability that a non-swiss prisoner, in prison for 12 months, who had 3 previous nonviolent infractions and who was convicted for a violent offense prior to the index offense. Without actually carrying out the calculations, carefully show how you would set up the formulas you would need to solve.

The $\text{logit} = -1.948656+.1139039\times1-1.029187\times0+.2768106\times3+.6653906\times1$ and [1 point]

$$\pi(\mathbf{x}) = \frac{e^{-1.948656+.1139039\times1-1.029187\times0+.2768106\times3+.6653906\times1}}{1+e^{-1.948656+.1139039\times1-1.029187\times0+.2768106\times3+.6653906\times1}}$$ [1 point]

The variance of the logit, by algebraic manipulation, can be shown to be:

$$\begin{aligned}\widehat{\text{var}}(\text{logit}) = {} & \widehat{\text{var}}(\beta_0) + x_1^2\,\widehat{\text{var}}(\beta_1) + x_2^2\,\widehat{\text{var}}(\beta_2) + x_3^2\,\widehat{\text{var}}(\beta_3) + x_4^2\,\widehat{\text{var}}(\beta_4) + \\ & 2x_1\widehat{\text{cov}}(\beta_0,\beta_1) + 2x_2\widehat{\text{cov}}(\beta_0,\beta_2) + 2x_3\widehat{\text{cov}}(\beta_0,\beta_3) + 2x_4\widehat{\text{cov}}(\beta_0,\beta_4) + \\ & 2x_1x_2\widehat{\text{cov}}(\beta_1,\beta_2) + 2x_1x_3\widehat{\text{cov}}(\beta_1,\beta_3) + 2x_1x_4\widehat{\text{cov}}(\beta_1,\beta_4) + \\ & 2x_2x_3\widehat{\text{cov}}(\beta_2,\beta_3) + 2x_2x_4\widehat{\text{cov}}(\beta_2,\beta_4) + 2x_3x_4\widehat{\text{cov}}(\beta_3,\beta_4)\end{aligned}$$

|1 point|

All necessary variances and covariances are contained in the variance covariance matrix provided by Stata in the vce command.

A 95% confidence interval for the logit is computed as follows:

$$\text{lower bound for the logit} = \text{logit} - 1.96\sqrt{\widehat{\text{var}}(\text{logit})}$$
$$\text{upper bound for the logit} = \text{logit} + 1.96\sqrt{\widehat{\text{var}}(\text{logit})}$$

The confidence interval for the probability is computed as:

$$\pi(\mathbf{x})_L = \frac{e^{\text{lower bound for the logit}}}{1+e^{\text{lower bound for the logit}}},\quad \pi(\mathbf{x})_U = \frac{e^{\text{upper bound for the logit}}}{1+e^{\text{upper bound for the logit}}}$$

|1 point|

A.) $$h_{pop}(t) = \frac{f_{pop}(t)}{S_{pop}(t)}$$

$$f_{pop}(t) = \frac{\partial}{\partial t}\left[1 - S_{pop}(t)\right]$$

$$= \theta\lambda \exp\{-\theta(1-e^{-\lambda t})\} e^{-\lambda t}$$

$$h_{pop}(t) = \frac{\theta\lambda \exp\{-\theta(1-e^{-\lambda t})\} e^{-\lambda t}}{\exp\{-\theta(1-e^{-\lambda t})\}}$$

$$= \theta\lambda e^{-\lambda t}$$

B.) $$\lim_{t\to\infty} S_{pop}(t) = \lim_{t\to\infty} \exp\{-\theta(1-e^{-\lambda t})\}$$

$e^{-\lambda t} \to 0$ as $t \to \infty$

So:

$$\lim_{t\to\infty} \exp\{-\theta(1-e^{-\lambda t})\} = \exp\{-\theta(1-0)\}$$

$$= e^{-\theta}$$

c.) $S_{pop}(t) = P(N=0) + P(Y_1 > t, \ldots, Y_N > t, N \geq 1)$

$$= e^{-\theta} + \sum_{k=1}^{\infty} P(Y_1 > t, \ldots, Y_k > t \mid N=k) P(N=k)$$

$$= e^{-\theta} + \sum_{k=1}^{\infty} (e^{-\lambda t})^k \frac{\theta^k e^{-\theta}}{k!}$$

$$= e^{-\theta} + \exp\{-\theta + \theta^*\} \sum_{k=1}^{\infty} \frac{\theta^{*k} e^{-\theta^*}}{k!}$$

$$\theta^* = \theta e^{-\lambda t}$$

$$= e^{-\theta} + \exp\{-\theta + \theta^*\} P(N^* \geq 1)$$

$$N^* \sim \text{Poisson}(\theta^*)$$

$$= e^{-\theta} + \exp\{-\theta + \theta^*\}(1 - e^{-\theta^*})$$

$$= e^{-\theta} + \exp\{-\theta + \theta e^{-\lambda t}\} - \exp\{-\theta + \theta^* - \theta^*\}$$

$$= \exp\{-\theta(1 - e^{-\lambda t})\}$$

= Equation 1

D.) $S^*(t) = P(T > t \mid N \geq 1)$

$$= \frac{P(Y_1 > t, \ldots, Y_N > t, N \geq 1)}{P(N \geq 1)}$$

$$= \frac{S_{pop}(t) - P(N=0)}{P(N \geq 1)}$$

$$= \frac{\exp\{-\theta[1 - e^{-\lambda t}]\} - e^{-\theta}}{1 - e^{-\theta}}$$

Q2 2010 Answer Key (Q3) ①

(Q3)

(a) $W = \sum_{i=1}^{T} Y_i x_i - \hat{p} \sum_{i=1}^{T} m_i x_i$ $\qquad \hat{p} = \frac{\sum_{i=1}^{T} Y_i}{\sum_{i=1}^{T} m_i} = \frac{\sum_{i=1}^{T} Y_i}{m_+}$

$= \sum_{i=1}^{T} Y_i x_i - \left(\frac{1}{m_+}\sum_{i=1}^{T} Y_i\right)\left(\sum_{j=1}^{T} m_j x_j\right)$

$= \sum_{i=1}^{T} Y_i \left(x_i - \frac{\sum_{j=1}^{T} m_j x_j}{m_+} \right)$

$\mathrm{Var}(W) = \sum_{i=1}^{T} \mathrm{Var}(Y_i) \left(x_i - \frac{\sum_{j=1}^{T} m_j x_j}{m_+} \right)^2$ since Y_i's are independent

$= \sum_{i=1}^{T} m_i p_i (1-p_i) \left(x_i - \underbrace{\frac{\sum_{j=1}^{T} m_j x_j}{m_+}}_{\bar{x}_w} \right)^2$

(b) If $p_1 = \cdots = p_T = p$,

$\mathrm{Var}(W) = p(1-p) \sum_{i=1}^{T} m_i \left(x_i - \underbrace{\frac{\sum_{j=1}^{T} m_j x_j}{m_+}}_{\bar{x}_w} \right)^2$ $\qquad \bar{x}_w = \frac{\sum_{i=1}^{T} m_i x_i}{\sum_{i=1}^{T} m_i}$

$= p(1-p) \sum_{i=1}^{T} m_i (x_i - \bar{x}_w)^2$ //

(c) $\mathrm{logit}(p_i) = \beta_0 + \beta_1 x_i \qquad H_0: \beta_1 = 0$ vs $H_1: \beta_1 \neq 0$

$\eta_i = f(p_i) = \log\left(\frac{p_i}{1-p_i}\right) = \log(p_i) - \log(1-p_i)$

$\frac{\partial \eta_i}{\partial p_i} = \frac{1}{p_i} - \frac{1}{1-p_i} = \frac{1}{p_i(1-p_i)}$

$\frac{\partial p_i}{\partial \beta_0} = \frac{\partial p_i}{\partial \eta_i} \cdot \frac{\partial \eta_i}{\partial \beta_0} = p_i(1-p_i) \cdot 1 = p_i(1-p_i)$

$\frac{\partial p_i}{\partial \beta_1} = \frac{\partial p_i}{\partial \eta_i} \cdot \frac{\partial \eta_i}{\partial \beta_1} = p_i(1-p_i) \cdot x_i = x_i p_i (1-p_i)$

$L = \prod_{i=1}^{T} \binom{m_i}{Y_i} p_i^{Y_i} (1-p_i)^{m_i - Y_i}$

$\log L = \log\left[\prod_{i=1}^{T} \binom{m_i}{Y_i}\right] + \sum_{i=1}^{T} y_i \log(p_i) + \sum_{i=1}^{T} (m_i - Y_i) \log(1-p_i)$

$= \log\left[\prod_{i=1}^{T} \binom{m_i}{Y_i}\right] + \sum_{i=1}^{T} \left[y_i \log(p_i) + (m_i - y_i) \log(1-p_i) \right]$

$\frac{\partial \log L}{\partial \beta_0} = \sum_{i=1}^{T} \left[\frac{y_i}{p_i} \cdot \frac{\partial p_i}{\partial \beta_0} - \frac{(m_i - y_i)}{1-p_i} \cdot \frac{\partial p_i}{\partial \beta_0} \right] = \sum_{i=1}^{T} \left[\left(\frac{y_i}{p_i} - \frac{m_i - y_i}{1-p_i} \right) p_i(1-p_i) \right]$

$= \sum_{i=1}^{T} \left[\frac{y_i - m_i p_i}{p_i(1-p_i)} \cdot p_i(1-p_i) \right] = \sum_{i=1}^{T} (y_i - m_i p_i) = U_0(\beta_0, \beta_1)$

2

$$\frac{\partial \log L}{\partial \beta_0} = \sum_{i=1}^{T}\left[\frac{y_i}{p_i}\cdot\frac{\partial p_i}{\partial \beta_1} - \frac{(m_i-y_i)}{1-p_i}\cdot\frac{\partial p_i}{\partial \beta_1}\right] = \sum_{i=1}^{T}\left[\left(\frac{y_i}{p_i}-\frac{m_i-y_i}{1-p_i}\right)x_i p_i(1-p_i)\right]$$

$$= \sum_{i=1}^{T}\left(\frac{y_i - m_i p_i}{p_i(1-p_i)}\cdot x_i p_i(1-p_i)\right) = \sum_{i=1}^{T} x_i(y_i - m_i p_i) = U_1(\beta_0,\beta_1)$$

Under H_0, $\beta_1 = 0$:

$$U_0 = \sum_{i=1}^{T}(y_i - m_i p_i) = 0 \rightarrow \hat{p} = \frac{\sum_{i=1}^{T} y_i}{\sum_{i=1}^{T} m_i}$$

Thus under H_0,
$$U = \begin{bmatrix} U_0(\tilde{\beta}_0, 0) \\ U_1(\tilde{\beta}_0, 0)\end{bmatrix} = \begin{bmatrix} 0 \\ \sum_{i=1}^{T} x_i(y_i - m_i\hat{p}_0)\end{bmatrix}$$

Information matrix:
$$I(\beta_0,\beta_1) = \begin{bmatrix} -\frac{\partial^2 \log L}{\partial \beta_0^2} & -\frac{\partial^2 \log L}{\partial \beta_0 \partial\beta_1} \\ -\frac{\partial^2 \log L}{\partial \beta_0\partial\beta_1} & -\frac{\partial^2 \log L}{\partial \beta_1^2}\end{bmatrix}$$

$$-\frac{\partial^2 \log L}{\partial \beta_0^2} = \frac{-\partial U_0}{\partial \beta_0} = \frac{-\partial U_0}{\partial p_i}\cdot\frac{\partial p_i}{\partial \beta_0} = \frac{-\partial}{\partial p_i}\left(\sum_{i=1}^{T}(y_i - m_i p_i)\right)\cdot p_i(1-p_i) = \sum_{i=1}^{T}(m_i p_i(1-p_i))$$

$$-\frac{\partial^2 \log L}{\partial \beta_0\partial\beta_1} = \frac{-\partial U_0}{\partial \beta_1} = \frac{-\partial U_0}{\partial p_i}\,\frac{\partial p_i}{\partial \beta_1} = \sum_{i=1}^{T}(x_i m_i p_i(1-p_i))$$

$$-\frac{\partial^2 \log L}{\partial \beta_1^2} = \frac{-\partial U_1}{\partial \beta_1} = \frac{-\partial U_1}{\partial p_i}\,\frac{\partial p_i}{\partial \beta_1} = \frac{-\partial}{\partial p_i}\left(\sum_{i=1}^{T} x_i(y_i - m_i p_i)\right)\cdot x_i p_i(1-p_i) = \sum_{i=1}^{T}(x_i^2 m_i p_i(1-p_i))$$

To find $I^{-1}(\beta_0,\beta_1)$ let $I(\beta_0,\beta_1) = \begin{bmatrix} a & b \\ c & d\end{bmatrix} \rightarrow I^{-1}(\beta_0,\beta_1) = \begin{bmatrix} \frac{d}{ad-bc} & \frac{-b}{ad-bc} \\ \frac{-c}{ad-bc} & \frac{a}{ad-bc}\end{bmatrix}$

Note that when you find $I^{-1}(\hat{\beta}_0, 0)$ you only need the lower right corner of I^{-1}, evaluated at $p_i = \hat{p}$:

Lower right of $I^{-1} = \frac{a}{ad-bc} = \frac{1}{d - \frac{bc}{a}}$

$$d - \frac{bc}{a} = \sum_{i=1}^{T} x_i^2 m_i p_i(1-p_i) - \frac{\left(\sum_{i=1}^{T} x_i m_i p_i(1-p_i)\right)^2}{\sum_{i=1}^{T} m_i p_i(1-p_i)} \quad \text{eval at } p_i = \hat{p}$$

$$= \hat{p}(1-\hat{p})\left[\sum_{i=1}^{T} x_i^2 m_i - \frac{\left(\sum_{i=1}^{T} x_i m_i\right)^2}{\sum_{i=1}^{T} m_i}\right]$$

Con't next page

(3)

$$= \hat{p}(1-\hat{p})\left[\sum_{i=1}^{T} x_i^2 m_i - 2\left(\sum_{i=1}^{T} m_i x_i\right)\underbrace{\left(\frac{\sum_{i=1}^{T} m_i x_i}{\sum_{i=1}^{T} m_i}\right)}_{\bar{x}_w} + \left(\sum_{i=1}^{T} m_i\right)\underbrace{\left(\frac{\sum_{i=1}^{T} m_i x_i}{\sum_{i=1}^{T} m_i}\right)^2}_{\bar{x}_w^2}\right]$$

$$= \hat{p}(1-\hat{p})\left[\sum_{i=1}^{T} m_i x_i^2 - 2\sum_{i=1}^{T} m_i x_i \bar{x}_w + \sum_{i=1}^{T} m_i \bar{x}_w^2\right]$$

$$= \hat{p}(1-\hat{p})\sum_{i=1}^{T} m_i (x_i - \bar{x}_w)^2$$

$\Rightarrow$ lower right of $I^{-1} = \dfrac{1}{\hat{p}(1-\hat{p})\sum_{i=1}^{T} m_i (x_i-\bar{x}_w)^2}$ $(*)$

Score test: $u^T I^{-1} u = \begin{bmatrix} 0 & u_1(\tilde{\beta}_0, 0)\end{bmatrix}\begin{bmatrix} ? & ? \\ ? & (*)\end{bmatrix}\begin{bmatrix} 0 \\ u_1(\tilde{\beta}_0,0)\end{bmatrix}$

$$= \frac{\left(\sum_{i=1}^{T} x_i (y_i - m_i \hat{p})\right)^2}{\hat{p}(1-\hat{p})\sum_{i=1}^{T} m_i (x_i - \bar{x}_w)^2} \quad //$$

(d) Describe intuitively, by examining the quantity being squared in the numerator of the of (1), why this test statistic will be large if $\text{logit}(p_i)$ is positively associated with x_i.

Solution:

The quantity being squared can be rewritten in several ways that provide intuition. One way is

$$N = \sum_{i=1}^{T} Y_i x_i - \widehat{p} \sum_{i=1}^{T} m_i x_i = \sum_{i=1}^{T} \left(\widehat{p}_i m_i x_i - \widehat{p} m_i x_i\right) = \sum_{i=1}^{T} m_i x_i \left(\widehat{p}_i - \widehat{p}\right).$$

The final expression can be thought of in several ways that suggest it will be "large" (positive or negative) if the p_i systematically increase or decrease in the x_i. Of course, in this model, the increase is not linear in the x_i.

One viewpoint is that N is (nearly) the numerator of the correlation between the pairs $(p_i - \widehat{p}, m_i x_i)$ and hence will be large if these quantities are monotonically associated. A second viewpoint is geometric; recall that $\sum_{i=1^T} w_i z_i$ is equal to the cosine of the angle between the vectors $(w_1, \ldots, w_T)$ and $(z_1, \ldots, z_T)$ times the length of each of vector, and that the cosine of an angle is large when and only when the angle is close to zero degrees or close to 180 degrees. Thus N will be large when either the $p_i - \widehat{p}$ are positively or negative associated with the values of $x_i m_i$. Other interpretations are possible.

(4)

Q2 2010
Answer Key
Q4

(a)(i) yes - if x & z are correlated (multicollinearity)

(ii) yes - leaving out a variable may cause increase in SSE that causes effects to be non-significant

(iii) yes - if x & z negatively correlated, for example:

$V(y) = V(x) = V(z) = 1 \quad Corr(y,x) = 0.3 \quad Corr(y,z) = 0.2$ as given

Let $\rho = Corr(x,z)$

$V(x+z) = 1 + 1 + 2\rho = 2(1+\rho)$

$Cov(y, x+z) = 0.3 + 0.2 = 0.5$

$Corr(y, x+z) = \frac{0.5}{\sqrt{2(1+\rho)}} = 0.7 \rightarrow \rho = -0.7449$

Check that $\begin{bmatrix} 1 & 0.3 & 0.2 \\ 0.3 & 1 & -0.74 \\ 0.2 & -0.74 & 1 \end{bmatrix}$ is positive semi-definite (yes) to ensure a valid covar matrix

(b) (i) $R^2 = [Corr(\hat{y}, y)]^2$ $\hat{y}$ = predicted / fitted from $y|x,z,w$
y = observed

(ii) overall F-test

(iii) $Corr(y, x | z, w)$ = correlation of residuals from $y|z,w$ and $x|z,w$ models

(iv) $y = \beta_0 + \beta_x x + \beta_z z + \beta_w w + e$ test $H_0: \beta_x = 0$ to test partial corr.

(c) Use partial regressions:

Step 1: Regress y on x $\rightarrow$ get residuals $y - \hat{y}$

Step 2: Regress z on x $\rightarrow$ get residuals $z - \hat{z}$

Step 3: Regress $(y-\hat{y})$ on $(z-\hat{z})$ with NO INTERCEPT

i.e. $(y-\hat{y}) = \beta(z-\hat{z}) + e$

Resulting β will be the coef. for z in the model

$y = \beta_0 + \beta_1 x + \beta_2 z + e$

Repeat with y on z, x on z, $(y-\hat{y})$ on $(x-\hat{x})$ to get coef for x.