STA 442/2101F 2009 Homework 3. *due December 1 before 4 pm*

When answering questions requiring numerical work, the results are to be reported in a narrative summary, in your own words. Tables and Figures may be included, but must be formatted along with the text. DO NOT include in this summary printouts of computer code with the relevant selections highlighted.

All computer code used to obtain the results summarized in the response should be provided as an appendix. In this appendix you may highlight the relevant results.

1. The attached paper from *The Lancet* was published online on October 29, 2009. A news item about the paper from *The Daily Express* is also attached. Read both the paper and the news item before answering the following questions.

   (a) What is the difference between the Diabetes Prevention Program (DPP) and the Diabetes Prevention Program Outcomes Study (DPPOS)? Which of the two is this paper mainly concerned with?

   (b) What is the primary outcome variable (response variable) for the study reported in *The Lancet* article? What are the secondary outcome variables?

   (c) The authors refer to an "intention-to-treat" analysis. What is this, and what does it mean in the context of this study?

   (d) Figure 3 shows cumulative incidence plots for diabetes for two time periods. What time periods are these? What are the main qualitative differences between the two sets of plots? What is the connection between the information presented in Figures 3 and 4? What explanations do the authors suggest in the Discussion for the differences in the primary outcome between DPP and DPPOS?

   (e) Summarize in one or two non-technical sentences the main findings about weight loss.

   (f) In the Discussion, the authors note (p.8, col.2) that "Cumulative data are completely contained within ... should be interpreted cautiously". Explain.

   (g) The *Daily Express* headline is "Cheap Pill to Keep Diabetes Under Control". Can you find information in the news article or the *Lancet* paper on the cost of the drug treatment? There are (at least) two further factual errors in the news article. Describe these and correct them. Find a news report on this study from a different news agency, give the details of your news article, and discuss whether or not the coverage was more accurate in this second news report.

Table 1: Numbers of faults in rolls of textile fabric

Roll No.	Roll length (metres)	No. of faults	Roll No.	Roll length (metres)	No. of faults
1	551	6	17	543	8
2	651	4	18	842	9
3	832	17	19	905	23
4	375	9	20	542	9
5	715	14	21	522	6
6	868	8	22	122	1
7	271	5	23	657	9
8	630	7	24	170	4
9	491	7	25	738	9
10	372	7	26	371	14
11	645	6	27	735	17
12	441	8	28	749	10
13	895	28	29	495	7
14	458	4	30	716	3
15	642	10	31	952	9
16	492	4	32	417	2

2. Set 2 of Cox & Snell, p.169: The data in Table 1 give the number of faults in rolls of textile fabric. This data is available in `R` as `cloth` in the library `boot`.

   (a) Assume that the number of faults in roll i, say y_i, follows a Poisson distribution with rate λx_i, where x_i is the length of roll i. Show that the maximum likelihood estimate of λ is given by $\hat{\lambda} = \bar{y}/\bar{x}$ and find an expression for the variance of $\hat{\lambda}$.

   (b) Fit this Poisson model to the data, and summarize the results. Does the Poisson model appear to fit the data?

   (c) **STA 2101**: Show that if it is assumed that λ follows at Gamma distribution with shape and scale parameters α and β respectively, that the distribution of y_i is of the form

   $$f(y_i \mid \alpha, \beta) = \frac{\Gamma(y+\alpha)}{\Gamma(y+1)\Gamma(\alpha)}(1-\pi_i)^y_i\pi_i^\alpha, \quad y_i = 0, 1, 2, \ldots,$$

   where $\pi_i = \beta/(\beta + x_i)$. Fit this model and discuss whether it fits the data better than the Poisson model of part (b).

3. **STA 442**:

   (a) Suppose $Y_1, \ldots, Y_n$ are independently and identically distributed, with $E(Y_i) = \mu$ and $\text{var}(Y_i) = \sigma_Y^2$. Let $\bar{Y} = n^{-1}\sum Y_i$. Show by Taylor series expansion that for

$X = g(\bar{Y})$:

$$\begin{aligned} E(X) &\doteq g(\mu), \\ \mathrm{var}(X) &\doteq \frac{\sigma_Y^2}{n}\{g'(\mu)\}^2. \end{aligned}$$

(b) Suppose Y follows a Poisson distribution with mean μ. What are the approximate mean and variance of $X = \bar{Y}^{1/2}$?

(c) Suppose $Y_1, \ldots, Y_n$ are independent identically distributed random variables from a $N(\mu, \sigma^2)$ distribution. Let $s^2 = \sum(Y_i - \bar{Y})^2/(n-1)$. Show that $\mathrm{var}(\log s^2) \doteq 2/(n-1)$.

(d) Bonus: Show in (a) that $E(X) = g(\mu) + O(1/n)$ and $\mathrm{var}(X) = g'(\mu)^2/n + O(1/n^2)$.

4. **STA 1201**: Consider the model

$$y_{ij} = \mu + \beta_i + \epsilon_{ij}, \quad j = 1, \ldots n; \quad i = 1, \ldots k. \tag{1}$$

Assume that ϵ_{ij} are independent, identically distributed as $N(0, \sigma^2)$.

(a) Show that

$$MSE = \sum_{i=1}^{k}\sum_{j=1}^{n}(y_{ij} - \bar{y}_{i.})^2/k(n-1)$$

is an unbiased estimator of σ^2.

(b) Suppose before we collect the data we decide that standard normal-theory 95% confidence intervals for $\beta_i - \beta_{i'}$ should have a width of no more than 10 units. Show that we should choose n to be (bigger than) approximately $0.32\sigma^2$.

(c) Assume now that instead of having a fixed factor with k levels, that the k groups of observations are randomly sampled from some population. For example the groups could be classrooms, medical centres, batches of paint, etc. The model is

$$y_{ij} = \mu + b_i + \epsilon_{ij}, \quad j = 1, \ldots n; \quad i = 1, \ldots k, \tag{2}$$

where b_i are independently and identically distributed as $N(0, \sigma_B^2)$, and the ϵ_{ij} are i.i.d. $N(0, \sigma^2)$ as before. Show that

$$MSB = \sum_{i=1}^{k}\sum_{j=1}^{n}(\bar{y}_{i.} - \bar{y}_{..})^2/(k-1)$$

is an unbiased estimate of $\sigma^2 + n\sigma_B^2$.

(d) Use the result that

$$\frac{MSB/(\sigma^2 + n\sigma_B^2)}{MSE/\sigma^2} \sim F_{(k-1),k(n-1)}$$

to construct a $100(1-\alpha)\%$ confidence interval for the variance ratio σ_B^2/σ^2.

(e) Construct 90%, 95% and 99% confidence intervals for this ratio using the blood data from Davison (Table 9.22).