METHODS QUALIFYING EXAM

JANUARY 2006

INSTRUCTIONS:

1. DO NOT put your NAME on the exam. Place the NUMBER assigned to you on the UPPER LEFT HAND CORNER of EACH PAGE of your exam.
2. Please start your answer to EACH QUESTION on a SEPARATE sheet of paper.
3. Answer all the questions.
4. Be sure to attempt all parts of every question. It may be possible to answer a later part of a question without having solved the earlier parts.
5. Be sure to hand in all of your exam. No additional material will be accepted once the exam has ended and you have left the exam room.

Problem 1

The National Institute for Standards and Technology conducted a study to develop standards for asbestos concentration. Asbestos dissolved in water was spread on a filter, and a section 3 *mm* in diameter was taken from the filter and mounted on a transmission electron microscope. An operator counted the number of asbestos fibers on the section. This procedure was repeated for 200 such samples. The 200 sections yielded the following counts: (the researcher no longer had the original counts – just the following grouped data and the mean of the 200 counts, $\bar{Y} = (1/200)\sum_{i=1}^{200} Y_i = 4940/200 = 27.7$)

	Grouped Counts							
	$0-10$	$11-15$	$16-20$	$21-24$	$25-27$	$28-30$	≥ 31	Total
O_i	2	1	36	52	50	39	20	200
E_i	0.12	2.43	34.62	57.51	45.36	32.62	27.34	200
$\frac{(O_i - E_i)^2}{E_i}$	30.25	0.85	0.06	0.53	0.48	1.25	1.97	35.39

a) The consulting statistician computed the chi-square goodness of fit and obtained a $p-\text{value} < 0.001$, and then stated that the Poisson model provided an inadequate fit to the data. Do you agree with his results? If not, correct his computations and reassess the fit of the Poisson model.

b) Assuming that the Poisson model provided a reasonable fit to the counts, construct a 95% confidence interval for the average number of asbestos fibers per 3 *mm* diameter area. (Hint: If $Y_1, \ldots, Y_n$ are iid Poisson(λ) random variables, then by the central limit theorem, the distribution of

$$\frac{\sqrt{n}\left(\bar{Y} - \lambda\right)}{\sqrt{\lambda}}$$

is approximately $N(0,1)$ for large n.

c) Under the assumption that the Poisson model provides an adequate representation of the distribution of asbestos fibers, is there sufficient evidence ($\alpha = 0.05$) that the average number of asbestos fibers per 3 *mm* diameter area is greater than 27?

(continued)

d) What is the power of the test developed in c) if the true average number of asbestos fibers per 3 *mm* diameter area equals 28?

e) Determine the sample size such that the test developed in b) will have a power of at least 0.90 when $\lambda = 28$.

Problem 2

This question tests your knowledge of fractional factorial designs.

a) For a resolution III design, main effects are confounded only with 3rd and higher order interactions. True or False?

b) For a resolution IV design, second order interactions are confounded with other 2nd order interactions. True or False?

c) A 2^{n-p} design that has resolution III requires how many generators?

d) What resolution are Plackett-Burman designs, and under what class of designs do they fall?

e) For an experiment that has eight (8) binary factors, what is the highest resolution you can get from a $2^{(8-4)}$ design and how many generators do you need?

Problem 3

Assume the model:

$$y_i = \beta_0 + \beta_1 x_i + \varepsilon_i, \; i = 1, \ldots, n$$

with

$$\varepsilon_i\text{'s iid } N\left(0, \sigma^2\right).$$

Recall that:

1) If w is $N(\mu, \sigma^2)$ and q is $\chi^2(p)$ and w and q are independent, then $t = \dfrac{(w/\sigma)}{\sqrt{q/p}}$ is distributed as $t(p, \delta)$, the non-central t with p degrees of freedom and non-centrality parameter $\delta = \mu/\sigma$.
2) $\hat{\beta}_1$ is $N\left[\beta_1, \sigma^2 / \sum_{i=1}^{n} (x_i - \bar{x})^2\right]$.
3) $(n-2)s^2 / \sigma^2$ is $\chi^2(n-2)$.
4) $\hat{\beta}_1$ and s^2 are independent.

Use the above to develop a test for:

$$\begin{aligned} H_0 &: \beta_1 = c \\ H_A &: \beta_1 \neq c \end{aligned}$$

Give details and the critical region.

Problem 4

Let $a_1 + b_1(x - \bar{x}_1)$ and $a_2 + b_2(x - \bar{x}_2)$ be two regression lines estimated from independent samples of sizes n_1 and n_2. Further let $S_{yy}^{(1)}$, $S_{xy}^{(1)}$ and $S_{xx}^{(1)}$ be the corrected sum of squares and products for the first sample, with superscript (2) replacing (1) to represent the analogous quantities from the second sample. We wish to find a confidence interval for the value ξ of x at which the true regression functions intersect. Assume that the usual regression assumptions apply and define the quantity z to be:

$$z = a_1 + b_1(\xi - \bar{x}_1) - a_2 - b_2(\xi - \bar{x}_2)$$

a) Show that $E(z) = 0$.

b) Show that $Var(z) = c\sigma^2$, where

$$c = \frac{1}{n_1} + \frac{(\xi - \bar{x}_1)^2}{S_{xx}^{(1)}} + \frac{1}{n_2} + \frac{(\xi - \bar{x}_2)^2}{S_{xx}^{(2)}}.$$

c) Show that $\hat{\sigma}^2$ is an unbiased estimator of σ^2, where

$$\hat{\sigma}^2(n_1 + n_2 - 4) = S_{yy}^{(1)} - \frac{\left[S_{xy}^{(1)}\right]^2}{S_{xx}^{(1)}} + S_{yy}^{(2)} - \frac{\left[S_{xy}^{(2)}\right]^2}{S_{xx}^{(2)}}.$$

d) Using the quantities z and $c\hat{\sigma}^2$ defined above, form a quadratic (in ξ) equation whose solution yields a $100(1-\alpha)$ percent confidence region for ξ. (You do not need to solve this equation, rather just set up the equation that can be solved to form the confidence interval.) Carefully define any notation you introduce and justify any additional distributional results that you use to construct your confidence interval.