Suppose that m observations are available on each of t "treatments" or "groups," where $m \geq 2$ and $t \geq 4$. Assume that the j^{th} observation on the i^{th} treatment has the representation

$$y_{ij} = \mu + \alpha_i + \epsilon_{ij} \quad (i = 1, ..., t;\ j = 1, ..., m),$$

where μ, α_1, ..., α_t are unknown parameters, and the ϵ_{ij}'s are statistically independent, normally distributed random variables having zero means and common variance σ^2. Let $\underline{\alpha} = (\mu, \alpha_1, ..., \alpha_t)'$ and let $\overline{y}_{i\cdot} = \sum_{j=1}^{m} y_{ij}/m$ for $i = 1, ..., t$.

Note: You may use matrices in your derivations, but no matrix should appear in your final answers to the following questions except for part (e).

(a) Let $c\mu + \sum_{i=1}^{t} d_i\alpha_i$ denote an arbitrary linear function of the parameters. Determine necessary and sufficient conditions in terms of c, $d_1, \ldots, d_t$ for this linear function to be estimable.

(b) Characterize all possible estimable functions and then derive their best linear unbiased estimators (BLUE's).

(c) Obtain confidence intervals for all of the differences $\alpha_i - \alpha_j$ $(i, j = 1, \ldots, t, i \neq j)$ such that the probability of simultaneous coverage is exactly $1 - \gamma$.

(d) Derive a size-γ test of the null hypothesis H_0: $\alpha_2 = \alpha_3 = \cdots = \alpha_t$ versus the alternative hypothesis H_a: not H_0.

For parts (e)—(g) below, let $r \geq 2$ and $s = t - r \geq 2$ be two integers and let

$$u = \frac{1}{t}\sum_{i=1}^{t} \overline{y}_{i\cdot}, \quad v = \frac{1}{r}\sum_{i=1}^{r} \overline{y}_{i\cdot}, \quad w = \frac{1}{s}\sum_{i=r+1}^{t} \overline{y}_{i\cdot}.$$

Furthermore, define

$$SS = \sum_{i=1}^{t}(\overline{y}_{i\cdot} - u)^2, \quad SS_1 = \sum_{i=1}^{r}(\overline{y}_{i\cdot} - v)^2, \quad SS_2 = \sum_{i=r+1}^{t}(\overline{y}_{i\cdot} - w)^2, \quad SS_3 = \frac{rs}{t}(v - w)^2.$$

(e) Show that $SS = SS_1 + SS_2 + SS_3$.

(f) Show that SS_1, SS_2 and SS_3 are independently distributed and find their distributions.

(g) Obtain an F-statistic based on SS_1 and $SS_{\text{res}} = \sum_{i=1}^{t}\sum_{j=1}^{m}(y_{ij} - \overline{y}_{i\cdot})^2$. Then find the degrees of freedom and noncentrality parameter of this F-statistic. What null hypothesis can be tested using this F-statistic?

Ph.D. Prelim Exam
Spring 2001

Solutions
Linear Models

(a) In matrix notation, we have $\underline{y} = X\underline{\alpha} + \underline{\varepsilon}$, where

$$\underline{y} = \begin{bmatrix} y_{11} \\ \vdots \\ y_{1m} \\ \vdots \\ y_{t1} \\ \vdots \\ y_{tm} \end{bmatrix}, \quad X = \begin{bmatrix} \underline{1}_m & \underline{1}_m & \underline{0} & \cdots & \underline{0} \\ \underline{1}_m & \underline{0} & \underline{1}_m & \cdots & \underline{0} \\ \vdots & \vdots & & & \vdots \\ \underline{1}_m & \underline{0} & \cdots & & \underline{1}_m \end{bmatrix}, \quad \underline{\varepsilon} = \begin{bmatrix} \varepsilon_{11} \\ \vdots \\ \varepsilon_{1m} \\ \vdots \\ \varepsilon_{t1} \\ \vdots \\ \varepsilon_{tm} \end{bmatrix}.$$

and $\underline{1}_m$ is an $m \times 1$ vector of 1's.

Note that the row space of X, denoted by $R(X)$, is spanned by the $1 \times (t+1)$ row vectors $(1, 1, 0, \cdots, 0)$, $(1, 0, 1, 0, \cdots, 0)$, $\cdots$, and $(1, 0, \cdots, 0, 1)$. Thus any estimable function $\underline{\lambda}'\underline{\alpha}$ is of the form

$$\left[c_1(1, 1, 0, \cdots, 0) + c_2(1, 0, 1, 0, \cdots, 0) + \cdots + c_t(1, 0, \cdots, 0, 1)\right] \cdot \underline{\alpha}$$
$$= \sum_{i=1}^{t} c_i(\mu + \alpha_i) = \left(\sum_{i=1}^{t} c_i\right)\mu + \sum_{i=1}^{t} c_i \alpha_i,$$

for some constants $c_1, \cdots, c_t$. Therefore, a necessary and sufficient condition for $c\mu + \sum_{i=1}^{t} d_i \alpha_i$ to be estimable is that $c = \sum_{i=1}^{t} d_i$.

(b) By part (a), any estimable function can be written as $\left(\sum_{i=1}^{t} d_i\right)\mu + \sum_{i=1}^{t} d_i \alpha_i$ for some constants $d_1, \cdots, d_t$. Note that the normal equations

$X'X\hat{\underline{\alpha}} = X'\underline{y}$ are

$$\begin{bmatrix} mt & m & \cdots & m \\ m & m & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ m & 0 & \cdots & m \end{bmatrix} \begin{bmatrix} \hat{\mu} \\ \hat{\alpha}_1 \\ \vdots \\ \hat{\alpha}_t \end{bmatrix} = \begin{bmatrix} m\sum_{i=1}^{t} \bar{y}_{i\cdot} \\ m\bar{y}_{1\cdot} \\ \vdots \\ m\bar{y}_{t\cdot} \end{bmatrix}.$$ Thus a solution

$\hat{\underline{\alpha}}$ is $\hat{\mu} = 0$, $\hat{\alpha}_1 = \bar{y}_{1\cdot}, \cdots, \hat{\alpha}_t = \bar{y}_{t\cdot}$. By the Gauss-Markov theorem, the BLUE of any estimable function $(\sum_{i=1}^{t} d_i)\mu + \sum_{i=1}^{t} d_i \alpha_i$ is then $\sum_{i=1}^{t} d_i \bar{y}_{i\cdot}$.

(c) First note that the differences $\alpha_i - \alpha_j = (\mu+\alpha_i) - (\mu+\alpha_j)$ are estimable functions and their BLUE's are $\bar{y}_{i\cdot} - \bar{y}_{j\cdot}$. Since $\bar{y}_{1\cdot}, \cdots, \bar{y}_{t\cdot}$ are independently distributed random variables and $\bar{y}_{i\cdot} \sim N(0, \sigma^2/m)$ for $i=1,\cdots,t$, the Tukey method applies and it gives confidence intervals for all of the differences $\alpha_i - \alpha_j$ with the probability of simultaneous coverage being exactly $1-\gamma$. These intervals are

$$(\bar{y}_{i\cdot} - \bar{y}_{j\cdot}) \pm (q^*_{\gamma,\, t,\, tm-t}) \cdot \hat{\sigma} \cdot \sqrt{\frac{1}{m}},$$

for $i, j = 1, \cdots, t$ and $i \neq j$, where

$\hat{\sigma}^2 = \sum_{i=1}^{t} \sum_{j=1}^{m} (y_{ij} - \bar{y}_{i\cdot})^2 / (tm-t)$ and

$q^*_{\gamma,\, t,\, tm-t}$ is the upper γ point of the studentized range distribution with t and $tm-t$ degrees of freedom.

(d) For model $\underline{Y} = X\underline{\alpha} + \underline{\varepsilon}$ given in part (a), the regression sum of squares is $\sum_{i=1}^{t} m \cdot \bar{y}_{i\cdot}^2$ and the residual sum of squares is $\sum_{i=1}^{t}\sum_{j=1}^{m} (y_{ij} - \bar{y}_{i\cdot})^2$.

For the model $\underline{Y} = X\underline{\alpha} + \underline{\varepsilon}$ with $\alpha_2 = \alpha_3 = \cdots = \alpha_t$, minimizing $\sum_{i=1}^{t}\sum_{j=1}^{m} (y_{ij} - \mu - \alpha_i)^2$ subject to $\alpha_2 = \cdots = \alpha_t$ gives one solution $\hat{\mu} = 0$, $\hat{\alpha}_1 = \bar{y}_{1\cdot}$ and $\hat{\alpha}_2 = \cdots = \hat{\alpha}_t = \frac{1}{t-1}\sum_{i=2}^{t} \bar{y}_{i\cdot}$. Thus the regression sum of squares for this model is $m\bar{y}_{1\cdot}^2 + \frac{m}{t-1}\left(\sum_{i=2}^{t} \bar{y}_{i\cdot}\right)^2$.

It follows from the general ANOVA table and Cochran's Theorem that

$$F = \frac{\left[\sum_{i=1}^{t} m \cdot \bar{y}_{i\cdot} - \left(m\bar{y}_{1\cdot}^2 + \frac{m}{t-1}\left(\sum_{i=2}^{t} \bar{y}_{i\cdot}\right)^2\right)\right]/(t-2)}{\sum_{i=1}^{t}\sum_{j=1}^{m} (y_{ij} - \bar{y}_{i\cdot})^2 / (mt - t)}$$

$$= \frac{m\sum_{i=2}^{t}\left[\bar{y}_{i\cdot} - \frac{1}{t-1}\sum_{j=2}^{t} \bar{y}_{j\cdot}\right]^2/(t-2)}{\sum_{i=1}^{t}\sum_{j=1}^{m} (y_{ij} - \bar{y}_{i\cdot})^2/(mt-t)}$$

has a noncentral F distribution with $t-2$ and $mt-t$ degrees of freedom and noncentrality parameter $\left(\frac{m}{2\sigma^2}\right)\sum_{i=2}^{t}\left[\alpha_i - \frac{1}{t-1}\sum_{j=2}^{t}\alpha_j\right]^2$, which equals 0 if and only if $\alpha_2 = \alpha_3 = \cdots = \alpha_t$. Thus a size-$\gamma$ test of $H_0: \alpha_2 = \alpha_3 = \cdots = \alpha_t$ versus H_a: not H_0 is obtained by rejecting H_0 if and only if $F \geq F_{\gamma:(t-2),\, mt-t}$.

(e) For simpler notation, let $z_i = \overline{y}_{i\cdot}$, $i=1,\cdots,t$, $\underline{z} = (z_1,\cdots,z_t)'$ and $\overline{z} = \frac{1}{t}\sum_{i=1}^{t} z_i$. Then $SS = \underline{z}'A\underline{z}$, where $A = I_t - \frac{1}{t}\underline{1}_t\underline{1}_t'$ and I_t is the $t\times t$ identity matrix. Similarly, we have $SS_i = \underline{z}'A_i\underline{z}$, $i=1,2,3$, where

$$A_1 = \begin{bmatrix} I_r - \frac{1}{r}\underline{1}_r\underline{1}_r' & 0 \\ 0 & 0 \end{bmatrix}, \quad A_2 = \begin{bmatrix} 0 & 0 \\ 0 & I_s - \frac{1}{s}\underline{1}_s\underline{1}_s' \end{bmatrix} \text{ and}$$

$$A_3 = \begin{bmatrix} \frac{s}{tr}\underline{1}_r\underline{1}_r' & -\frac{1}{t}\underline{1}_r\underline{1}_s' \\ -\frac{1}{t}\underline{1}_s\underline{1}_r' & \frac{r}{st}\underline{1}_s\underline{1}_s' \end{bmatrix}.$$

Since $r+s=t$, then $\frac{s}{tr} - \frac{1}{r} = -\frac{1}{t}$ and $\frac{r}{st} - \frac{1}{s} = -\frac{1}{t}$. It then follows that $A_1+A_2+A_3 = A$ and hence that $SS = SS_1 + SS_2 + SS_3$.

(f) Note that A_1, A_2, A_3 and A are symmetric and idempotent matrices, so is $I_t - A$. Then by Cochran's Theorem, SS_1, SS_2 and SS_3 are independently distributed, and their distributions are respectively $(\frac{\sigma^2}{m})\cdot\chi^2[r-1, (\frac{m}{2\sigma^2})\cdot\sum_{i=1}^{r}(\alpha_i - \frac{1}{r}\sum_{j=1}^{r}\alpha_j)^2]$, $(\frac{\sigma^2}{m})\cdot\chi^2[s-1, (\frac{m}{2\sigma^2})\cdot\sum_{i=r+1}^{t}(\alpha_i - \frac{1}{s}\sum_{j=r+1}^{t}\alpha_j)^2]$, and $(\frac{\sigma^2}{m})\cdot\chi^2[1, (\frac{m}{2\sigma^2})\cdot\frac{rs}{t}(\frac{1}{r}\sum_{j=1}^{r}\alpha_j - \frac{1}{s}\sum_{j=r+1}^{t}\alpha_j)^2]$.

(g) Note that $\underline{z} = \frac{1}{m}\begin{pmatrix} \underline{1}_m' & \underline{0} & \cdots & \underline{0} \\ \underline{0} & \underline{1}_m' & \cdots & \underline{0} \\ \vdots & \vdots & \ddots & \vdots \\ \underline{0} & \underline{0} & \cdots & \underline{1}_m' \end{pmatrix}_{t \times (mt)} \cdot \underline{y}$ and thus SS_1 can be written as $\underline{y}' H \underline{y}$, where $H = \begin{bmatrix} H_1 & 0 \\ 0 & 0 \end{bmatrix}$, with

$$H_1 = \frac{1}{m^2} \cdot \left\{ \begin{bmatrix} D & 0 & \cdots & 0 \\ 0 & D & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & D \end{bmatrix}_{(mr)\times(mr)} - \frac{1}{r} \begin{bmatrix} D & \cdots & D \\ \vdots & & \vdots \\ D & \cdots & D \end{bmatrix} \right\}, \text{ and}$$

$D = \underline{1}_m \underline{1}_m'$. Also, $SS_{res} = \underline{y}' P \underline{y}$, where

$$P = I - \frac{1}{m} \begin{bmatrix} D & 0 & \cdots & 0 \\ 0 & D & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & D \end{bmatrix}_{(mt)\times(mt)}.$$ Note that $H \cdot P = 0$. Thus SS_1 and SS_{res} are independent. By part (f), $\left(\frac{m}{\sigma^2}\right) SS_1 \sim \chi^2\left[r-1, \left(\frac{m}{2\sigma^2}\right) \cdot \sum_{i=1}^{r} (\alpha_i - \sum_{j=1}^{r} \alpha_j / r)^2\right]$. Furthermore, $\left(\frac{1}{\sigma^2}\right) SS_{res} \sim \chi^2(mt-t)$. Thus

$$F = \frac{m SS_1 / (r-1)}{SS_{res} / (mt-t)}$$ is an F statistic which has an F distribution with $r-1$ and $mt-t$ degrees of freedom and with noncentrality parameter $\left(\frac{m}{2\sigma^2}\right) \cdot \sum_{i=1}^{r} (\alpha_i - \sum_{j=1}^{r} \alpha_j / r)^2$. This parameter equals zero if and only if $\alpha_1 = \cdots = \alpha_r$. Thus the F statistic obtained here can be used to test $H_0: \alpha_1 = \cdots = \alpha_r$ versus H_a: not H_0.

1. A randomized double blind clinical trial was undertaken to evaluate two therapies for HIV-infected individuals. One outcome of interest was the percentage of subjects who had higher CD4 counts at 56 weeks than at baseline (i.e, the improvers). The 2×2 table below contains the data:

	Improvement at 56 weeks	
	Yes	No
Treatment A	114	92
Treatment B	84	80

a) Is the percentage of improvers on treatment A higher than on treatment B? Conduct a test and state your conclusions.

b) The above table contains only the observed data for the subjects who completed 56 weeks of treatment. Unfortunately, 34% of the CD4 counts at 56 weeks were missing on treatment A and 27% on treatment B. The reasons for missingness included loss to follow-up, skipped clinic visits, and discontinuing the assigned therapy. The investigators were concerned that subjects with missing CD4 counts at 56 weeks were different than those who provided CD4 counts at 56 weeks. Based on the investigators' concern, discuss the implications of ignoring the individuals with missing data in the analysis done in part (a).

c) Selection models are a common approach to dealing with missing data. In a selection model, an indicator of missingness (m) is used where

$$m = \begin{cases} 1 & \text{if the response is observed} \\ 0 & \text{if the response is missing} \end{cases} ,$$

Let $f(y,m)$ denote the pmf for the joint distribution of a binary response Y and the indicator of missingness m. In a selection model, $f(y,m)$ is factored as

$$f(y,m) = f(m|y)f(y) ,$$

where $f(m|y)$ represents the missing data (or selection) mechanism and $f(y)$ gives the marginal distribution of the response. Suppose the missing data mechanism is

$$m|y \sim Bernoulli(\exp(\alpha_0 + \alpha_1 y)/(1 + \exp(\alpha_0 + \alpha_1 y))) ,$$

where α_0 and α_1 are unknown parameters. What do the parameters α_0 and α_1 represent in this model?

d) Since we are dealing with a binary response, $f(y)$ is determined by

$$\begin{aligned} \rho \equiv P(Y=1) &= P(Y=1|m=1)P(m=1) + P(Y=1|m=0)P(m=0) \\ &= P(Y=1|m=1)P(m=1) + P(Y=1|m=0)(1 - P(m=1)) \end{aligned}$$

In the HIV study, we would like to estimate ρ, the probability of improvement at 56 weeks, for each treatment. It is clear here that we can estimate $P(Y = 1|m = 1)$ and $P(m = 1)$ from the observed data for each treatment, but we cannot directly estimate $P(Y = 1|m = 0)$ for either treatment. As a first step in dealing with this problem, show that $P(Y = 1|m = 0)$, and consequently ρ, can be expressed as a function of only $P(Y = 1|m = 1)$, $P(m = 1)$, α_0, and α_1 [Hint: Bayes Rule might be useful here].

e) It can also be shown that α_0 is identified (α_0 can be estimated) when α_1 is known. (You are not asked to show this.) Consequently, $P(m = 1|y)$ can be estimated from the observed data when α_1 is known. Given this result, what might be a good strategy for presenting an analysis of data from studies such as the HIV study described above. In what types of situations could we draw a definitive conclusion?

2. The objective of a recent randomized clinical trial was to investigate the effect of a new treatment that combines a drug therapy with exercise (D+E) for building and maintaining muscle strength in the elderly, relative to a treatment that just uses exercise (E). Each subject's muscle strength was measured at baseline and after 12 months of treatment. The trial enrolled 78 subjects, with 40 subjects in the exercise (E) group and 38 subjects in the drug plus exercise group (D+E). Only 53 subjects completed the trial: 9 dropped out of the exercise group and 16 dropped out of the exercise plus drug group. The means and standard deviations for the completers in each treatment group at baseline and 12 months appear below. Standard deviations are given in parentheses.

	baseline	12 months	Number of completers
treatment E	64.8 (23.8)	72.5 (21.0)	31
treatment D+E	78.1 (23.7)	88.4 (32.1)	22

a) Consider only the data at 12 months. Perform a test to determine if exercise plus the new drug therapy (treatment D+E) was better than just exercise (treatment E) with respect to average muscle strength at 12 months? State your conclusions.

b) Suppose we want to compare the changes in mean muscle strength from baseline to 12 months for the two treatments. Do we have enough information in the above table to perform an appropriate test? Explain.

c) Twenty-five subjects did not complete the trial. Let's consider another way to model missing data, called pattern mixture models. In a pattern mixture model, $f(y, m)$ is factored as

$$f(y, m) = f(y|m)f(m),$$

where $f(y|m)$ gives the distribution of the response conditional on the observations being missing or not missing, and $f(m)$ gives the marginal distribution of m, the variable that indicates the missingness pattern.

Let Y be a 2×1 random vector in which the first component is muscle strength at baseline and the second component is muscle strength at 12 months of treatment, and let x denote a corresponding vector of subject-specific covariates. Consider the following pattern mixture model for each treatment:

$$\begin{aligned} Y|(m=k) &\sim N(\mu^{(k)}, \Sigma^{(k)}) \quad \text{for } k = 0, 1 \\ m|x &\sim Bernoulli(\exp(x'\beta)/(1+\exp(x'\beta))) \ , \end{aligned}$$

where β is a vector of unknown parameters. For each completer ($m = 1$), we observe $(y_0,\ y_1)$ where y_0 is the observed baseline muscle strength and y_1 is the observed muscle strength at 12 months. For each dropout ($m = 0$), we observe $(y_0, -)$ because the observation at 12 months is missing. Let

$$\mu^{(k)} = \begin{pmatrix} \mu_0^{(k)} \\ \mu_1^{(k)} \end{pmatrix} \quad \text{and} \quad \Sigma^{(k)} = \begin{bmatrix} \Sigma_{00}^{(k)} & \Sigma_{01}^{(k)} \\ \Sigma_{10}^{(k)} & \Sigma_{11}^{(k)} \end{bmatrix}.$$

Obviously, the following parameters are not identified and cannot be estimated: $\mu_1^{(0)}$ and $\Sigma_{11}^{(0)}$, the mean and variance at 12 months for the dropouts, and $\Sigma_{01}^{(0)}$, the covariance between the observations at baseline and 12 months for the dropouts. One way to deal with this situation is to incorporate additional restrictions into the model.

Consider the restriction that $f(y_1|y_0, x)$ is the same for both completers and dropouts (i.e., the conditional distribution of the 12 month observation given the baseline observation and the covariates for the subject is the same among completers and dropouts). Is this assumption verifiable from the data? Explain.

d) Show that under this restriction, $P(m = 1|y_0, y_1, x) = h(y_0; x)$ (i.e., the dropout mechanism is ignorable). Also, give the form of h.

e) What happens if we use the restriction that $f(y_0|y_1, x)$ is the same for completers and dropouts?

3. Keeping in mind the scenario in part 2, consider the following parametric selection model.

$$\begin{aligned} m|y &\sim Bernoulli(\exp(\gamma_0 + \gamma_1 y)/(1+\exp(\gamma_0 + \gamma_1 y))) \\ y &\sim N(\mu, \Sigma) \end{aligned}$$

In most cases, no restrictions are necessary for all of the parameters to be identified (unlike in part 2). Give some intuition on how this is possible. (Think about the assumptions we are making in this model that tell us what the data should look like if we could observe both the baseline and 12 month strength measurements for every subject in the study).

1) a) X^2 test (or difference between proportions)

$X^2 = .62 \qquad p > .05$

No significant difference

b) No, cannot just ignore the missing data. The missingness here depends on the values of the unobserved data so we will get biased estimates of treatment percentages and their differences.

c) α_1 log odds ratio for unit change in response y

α_0 log odds of missingness for response $y_1 = 0$

d) Given α, show can identify e

$$e = P(Y=1) = \underbrace{P(Y=1 \mid m=0)}_{e_0} \underbrace{P(m=0)}_{1-\gamma} + \underbrace{P(Y=1 \mid m=1)}_{e_1} \underbrace{P(m=1)}_{\gamma} \qquad (1)$$

apply Bayes rule

$$P(Y=1 \mid m=0) = \frac{P(m=0 \mid Y=1) P(Y=1)}{P(m=0)} = \frac{\left[e^{\alpha_0+\alpha_1}/(1+e^{\alpha_0+\alpha_1})\right] e}{1-\gamma}$$

know plug this back into (1)

$$e = \frac{\left[e^{\alpha_0+\alpha_1}/(1+e^{\alpha_0+\alpha_1})\right] e}{1-\gamma} (1-\gamma) + e_1 \gamma$$

solve for e

$$e = e_1 (1+e^{\alpha_0+\alpha_1}) \gamma$$

e) A good strategy for presenting the results would be to do a sensitivity analysis. For reasonable values of α_1, examine and present, e.g., the difference between treatments. If over this range, the inferences do not change, could draw a definitive conclusion of the superiority of 1 treatment over the other.

2) a) Two sample t-test $T = \dfrac{\bar{Y}_E - \bar{Y}_{DE}}{s_p\sqrt{\frac{1}{n_E} + \frac{1}{n_{DE}}}} \overset{\text{assume const}^t \text{ variance}}{=} \dfrac{72.5 - 88.4}{26.1\sqrt{\frac{1}{31} + \frac{1}{22}}} = -2.2$ $p < .05$

D+E is better

b) No. Would use paired t-test and then need variance of differences which is unavailable from the table

c)

assumption not verifiable

have no information about conditional distribution of $y_1 | y_0$ in dropout pattern $(k=0)$

use Bayes rule

$$P(m=1 | y_0, y_1, x) = \frac{f(y_0, y_1 | m=1, x)\, P(m=1|x)}{f(y_0, y_1 | m=1, x)\, P(m=1|x) + f(y_0, y_1 | m=0, x) P(m=0|x)}$$

$$= \frac{f(y_1 | y_0, m=1, x)\, f(y_0 | m=1, x)\, P(m=1|x)}{f(y_1 | y_0, m=1, x) P(m=1|x) + f(y_1 | y_0, m=0, x) f(y_0 | m=0, x) P(m=0)}$$

$f(y_0 | m=1, x)$

under restriction can factor $f(y_1 | y_0, x)$ out of denominator and cancel with term in numerator

recall: $f(y_1 | y_0, m=1, x) = f(y_1 | y_0, m=0, x)$

$$= \frac{f(y_0 | m=1, x) P(m=1|x)}{f(y_0 | m=1, x) P(m=1|x) + f(y_0 | m=0, x) P(m=0|x)} = h(y_0, x)$$

not depend on y_1 ✓

d) under restriction $f(y_0 | y_1, m=1, x) = f(y_0 | y_1, m=0, x)$

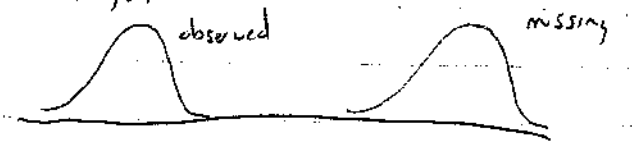
form similar to above with y_0 replaced by y_1 ✓

3)

~~under~~ given the restriction that the data are multivariate normal and given the observed data, there should often be a unique value of α which 'supports' the data being multivariate normal.

For example, consider a single observation of each subject and suppose the observed data look as below

for α large and positive, the missing observations would lie far to the right of the above distribution, and $f(y)$ would be bimodal (inconsistent with normality).



So α should be 'identified' (though weakly) by the restriction that $f(y)$ is normal

A study was performed at the Iowa State University Horticulture Station to examine properties of a certain bioinsecticide that is used to control insects that damage grass. A patch of grass was grown at each of four different locations in Iowa. The same variety of grass was grown in each patch. Each patch was divided into a 3x3 grid of 1.0m x 1.0m plots. Treatments consisted of the application of one of three levels of a liquid bioinsecticide (60, 120 or 240 gallons per acre (GPA) followed by one of three post-application levels of irrigation (0, 1, or 2 cm of water per day). The nine combinations of these two factors were randomly assigned to the nine plots within each patch with a different randomization for each patch.

The bioinsecticide is used to control certain insects that damage grass. It consists of spores of a certain fungus that are suspended in a solution. After they are introduced into the soil, the spores develop into fungal colonies that infect and destroy the insects. The objective of this study was to determine how different methods of application affect the viability of the spores in the soil. Each of the three bioinsecticide application levels (60, 120, or 240 gallons per acre) deposited the same number spores per unit area (5.1×10^8 spores per 10 cm^2). The 240 gallons per acre application, for example, deposited the same number of spores in a plot as the 60 gallons per acre application, but the 240 gallon per acre application used fours times more water than the 60 gallon per acre application.

Soil samples were taken from each plot at 1, 5, 10, and 50 days after the application of the bioinsecticide. Each plot was divided into four sub-plots and the soil sample was taken from a different subplot on each sampling date. Sampling dates were randomly assigned to the four subplots. A soil sample was obtained by inserting a probe into the soil. Then, soil sections were taken from each soil sample at depths of 1cm, 3cm, and 5cm below the surface. The soil sections were frozen at 4°C until measurements were taken. Serial dilutions of each soil section were suspended in phosphate buffered saline and plated on two plates per dilution of nutrient agar containing rifamycin B (15 ug/ml), Nystatin (50 ug/ml) and cyclohexamide (100 ug/ml). The average of the results from the two plated dilutions was used to calculate the number of colony forming units (CFU) per gram of soil for each soil section. The researchers decided to use $Y=\log_{10}(CFU+1)$ as the response in the analysis of these data.

The ANOVA corresponding to the table outlined on the top of the next page was used to analyze the data from this study. You are first asked to write down a corresponding linear model and then go back to this table to fill in degrees of freedom and expectations of some mean squares.

Source of Variation	Degrees of freedom	Expected Mean Squares
Patches of grass (locations)		
Application levels		
Post irrigation levels		
Application × Irrigation interaction		
Error a (Whole plot error)		
Sampling dates (sub-plots)		
Dates × application interaction		
Dates × irrigation interaction		
Dates × application × irrigation		
Error b (sub-plot error)		
Depths		
Depths × application		
Depths × irrigation		
Depths × application × irrigation		
Depths × dates		
Depths × application × dates		
Depths × irrigation × dates		
Depths × application × irrigation × dates		
Error c		
Corrected total		

A. Write out a linear model corresponding to the ANOVA table shown above. Identify the fixed effects and the random effects in your model and describe the distributional properties of the random effects. What does your model imply about the variances of the $\log_{10}(CFU+1)$ observations and correlations among the $\log_{10}(CFU+1)$ observations?

B. Your answer to part A will depend on whether you consider the location effects as fixed effects or as random effects.

(i) Give a general criterion, or a set of general criteria, that one could use to decide if block effects should be considered as fixed effects or random effects.

(ii) For this particular bioinsecticide study, describe the types of inferences that could be affected by the decision to consider location effects as fixed or random effects. What types of inferences would be unaffected by this decision?

C. Fill in the values of the degrees of freedom in the ANOVA table shown above. Using the notation you established in your answer to Part A, give formulas for the expectations of the three error mean squares.

D. Describe how you would evaluate the decision to use $Y = \log_{10}(CFU+1)$ as the response in the analysis of these data. Describe the data displays you would examine and other procedures you would use in this evaluation.

E. The following table shows sample means for the $\log_{10}(CFU+1)$ values at 50 days after application of the bioinsecticide. These sample means were computed by averaging across the four locations and the three post-application levels of irrigation.

Bioinsecticide Application Level

Soil Depth	*60 gal/acre*	*120 gal/acre*	*240 gal/acre*
1 cm	6.1	6.4	7.0
3 cm	5.3	5.4	6.1
5 cm	5.1	5.3	5.5

Evaluate the sum of squares for the (linear)x(linear) interaction contrast for these two factors. Show how to construct an F-test of the null hypothesis that the expected value of this contrast is zero.

F. Using the means squares from the ANOVA table you outlined in part (A), give formulas for the standard errors of the following differences between sample means at 50 days shown in part (D):

(i) Difference between the sample means for the 240 gal/acre and 60 gal/acre applications at soil depth 1 cm.

(ii) Difference between the sample means at soil depths of 5 cm and 1 cm for the 60 gal/acre application.

(iii) Difference between the sample mean at 1 cm for the 60 gal/acre application and the sample mean at 5 cm for the 240 gal/acre application.

G. The description of the study indicates that randomization was used in performing the experiment. What are the potential benefits of the use of randomization in this study?

H. While the method used in this study to determine the CFU value for each soil sample (call it method C) is relatively accurate, it is also expensive and time consuming. There are two faster and cheaper methods that the researchers could have used. Call these methods A and B. A second study was done to estimate the correlations between the results for these three methods. In this study, all three methods were applied to each of forty soil samples and sample correlations between the results from these three methods, r_{AC}, r_{AB}, and r_{BC} were computed.

Describe the procedure you would use to determine if the true correlation between the results from methods A and C is stronger or weaker than the true correlation between the results from methods B and C. Give some justification for the procedure you selected.

A. For the $\log_{10}$(CFU+1) measurement taken at the m-*th* depth on the ℓ -*th* date on the plot at the i-*th* location that received the j-*th* application level with the k-*th* level of post application irrigation, we have

$$\begin{aligned} Y_{ijk\ell m} = {} & \mu + \beta_i + A_j + C_k + (AC)_{jk} + \omega_{ijk} + T_\ell + (AT)_{j\ell} + (CT)_{k\ell} + (ACT)_{jk\ell} + \eta_{ijk\ell} \\ & + D_m + (AD)_{jm} + (CD)_{km} + (ACD)_{jkm} \\ & + (TD)_{\ell m} + (ATD)_{j\ell m} + (CTD)_{k\ell m} + (ACTD)_{jk\ell m} + \varepsilon_{ijk\ell m} \end{aligned}$$

where

$\beta_i \sim NID(0, \sigma_g^2)$ are random location (block) effects

$\omega_{ijk} \sim NID(0, \sigma_a^2)$ are random whole plot effects

$\eta_{ijk\ell} \sim NID(0, \sigma_b^2)$ are random sub-plot effects

$\varepsilon_{ijk\ell m} \sim NID(0, \sigma_c^2)$ are random errors

and any random effect is independent of any other random effect. This model implies that each observation has the same variance $(\sigma_c^2 + \sigma_b^2 + \sigma_a^2 + \sigma_g^2)$ and there are three levels of correlation:

$(\sigma_b^2 + \sigma_a^2 + \sigma_g^2)/(\sigma_c^2 + \sigma_b^2 + \sigma_a^2 + \sigma_g^2)$	between observations at different depths from the same soil sample
$(\sigma_a^2 + \sigma_g^2)/(\sigma_c^2 + \sigma_b^2 + \sigma_a^2 + \sigma_g^2)$	between observations taken on different dates within the same whole plot
$\sigma_g^2/(\sigma_c^2 + \sigma_b^2 + \sigma_a^2 + \sigma_g^2)$	between observations from different whole plots at the same location (block)

Random location (block) effects were incorporated into this solution. A solution using fixed block effects is not shown here. This model could be modified by including additional random effects to account for more complex patterns of correlations among responses. The normality assumption is not used in the construction of the ANOVA table or the derivation of formulas for variances and covariances. We could have simply assumed that random components are sets of i.i.d. random variables without making the normality assumption.

B. Discussion.

C.

Source of Variation	Degrees of freedom	Expected Mean Squares
Locations (blocks)	3	$\sigma_c^2 + 3\sigma_b^2 + 12\sigma_a^2 + 108\sigma_g^2$
Application levels	2	
Post irrigation levels	2	
Application × Irrigation interaction	4	
Error a (Whole plot error)	24	$\sigma_c^2 + 3\sigma_b^2 + 12\sigma_a^2$
Sampling dates (sub-plots)	3	
Dates × application interaction	6	
Dates × irrigation interaction	6	
Dates × application × irrigation	12	
Error b (sub-plot error)	81	$\sigma_c^2 + 3\sigma_b^2$
Depths	2	
Depths × application	4	
Depths × irrigation	4	
Depths × application × irrigation	8	
Depths × dates	6	
Depths × application × dates	12	
Depths × irrigation × dates	12	
Depths × application × irrigation × dates	24	
Error d	216	σ_c^2
Corrected total	431	

D. To check if the $Y=\log_{10}(CFU+1)$ transformation helps to approximate the normality assumptions for the random errors included in the model in part (B), one can look at a series of Q-Q plots of residuals and corresponding Shapiro-Wilk or Anderson-Darling tests for normality. The first set of residuals is obtained by using ordinary least squares (OLS) estimation to fit the model described in part (B) and compute residuals $e_{ijk\ell m} = Y_{ijk\ell m} - \hat{Y}_{ijk\ell m}$. A second set of residuals is obtained by averaging observations across the three depths for each soil sample and using OLS estimation to fit the model

$$\bar{Y}_{ijk\ell\bullet} = \mu + \beta_i + A_j + C_k + (AC)_{jk} + \omega_{ijk} + T_\ell + (AT)_{j\ell} + (CT)_{k\ell} + (ACT)_{jk\ell} + \eta_{ijk\ell}.$$

A third set of residuals is obtained by computing the average $\log_{10}(CFU+1)$ value for each whole plot (averaging across dates and soil depths at each date) and using OLS estimation to fit the model

$$\bar{Y}_{ijk\bullet\bullet} = \mu + \beta_i + A_j + C_k + (AC)_{jk} + \omega_{ijk}.$$

Homogeneity of variances can be examined by plotting residuals against estimated means for these three regression models. You could examine residual plots for other transformations of the CFU values to determine if some other transformation better promotes normality or homogeneity of variances for the conditional error distributions.

E. Note that the levels of the soil depth factor are equally spaced and coefficients for the linear contrast are –1, 0, 1. The levels of the application factor are not equally spaced, and coefficients for the linear contrast are (60-140)/20=-4, (120-140)/20=-1, and (240-140)/20=5. The estimates of the interaction contrast is

$$\hat{C} = -4(-1\bar{Y}_{\bullet 1\bullet 41} + 0\bar{Y}_{\bullet 1\bullet 42} + \bar{Y}_{\bullet 1\bullet 43}) - (-1\bar{Y}_{\bullet 2\bullet 41} + 0\bar{Y}_{\bullet 2\bullet 42} + \bar{Y}_{\bullet 2\bullet 43})$$

$$+ 5(-1\bar{Y}_{\bullet 3\bullet 41} + 0\bar{Y}_{\bullet 3\bullet 42} + \bar{Y}_{\bullet 3\bullet 43}) \quad = \quad -2.4$$

and $SS_C = 12\hat{C}^2/84 = \hat{C}^2/7$. Reject the null hypothesis that $E(\hat{C}) = 0$ if $F = SS_C / MS_{error\,c} > F_{(1,216),\alpha}$.

F. (i) $\mathrm{Var}(\bar{Y}_{\bullet 3\bullet 41} - \bar{Y}_{\bullet 1\bullet 41}) = \dfrac{\sigma_a^2 + \sigma_b^2 + \sigma_c^2}{6}$ and a method of moments estimator for the standard error of $(\bar{Y}_{\bullet 3\bullet 41} - \bar{Y}_{\bullet 1\bullet 41})$ is

$$S_{\bar{Y}_{\bullet 3\bullet 41} - \bar{Y}_{\bullet 1\bullet 41}} = \left(\frac{MS_{\text{error c}} + 3MS_{\text{error b}} + 8MS_{\text{error c}}}{72} \right)^{0.5}.$$

(ii) $\mathrm{Var}(\bar{Y}_{\bullet 1\bullet 43} - \bar{Y}_{\bullet 1\bullet 41}) = \dfrac{\sigma_c^2}{6}$ and a method of moments estimator for the standard error of $(\bar{Y}_{\bullet 1\bullet 43} - \bar{Y}_{\bullet 1\bullet 41})$ is $S_{\bar{Y}_{\bullet 1\bullet 43} - \bar{Y}_{\bullet 1\bullet 41}} = \left(\dfrac{MS_{\text{error c}}}{6} \right)^{0.5}$.

(iii) $\mathrm{Var}(\bar{Y}_{\bullet 1\bullet 41} - \bar{Y}_{\bullet 3\bullet 43}) = \dfrac{\sigma_a^2 + \sigma_b^2 + \sigma_c^2}{6}$ and a method of moments estimator for the standard error of $(\bar{Y}_{\bullet 1\bullet 41} - \bar{Y}_{\bullet 3\bullet 43})$ is

$$S_{\bar{Y}_{\bullet 1\bullet 41} - \bar{Y}_{\bullet 3\bullet 43}} = \left(\frac{MS_{\text{error c}} + 3MS_{\text{error b}} + 8MS_{\text{error c}}}{72} \right)^{0.5}.$$

G. Discussion

H. Methods based on the assumption that the sample correlations are independent random variables would not be appropriate in this situation. Why?

Here is one idea. A bootstrap procedure could be applied. Use simple random sampling with replacement to select B bootstrap samples from the original sample. Each of the bootstrap samples would select 40 soil samples from the original set of 40 soil samples using simple random sampling with replacement. Estimates of the correlation coefficients, $r_{AC,j}$ and $r_{BC,j}$, would be computed for each of the j=1,2, …., B bootstrap samples. Then, one could construct a bootstrapped confidence interval for an appropriate function of the population correlations, e.g. $\rho_{AC,j} - \rho_{BC,j}$ or $\rho_{AC,j} / \rho_{BC,j}$.

Statistics Ph.D. Preliminary Examination Methods Question, Spring 2001

An experiment was conducted to study the life-time distribution of of an electronic device to be used in an undersea communications system. In order to get information more quickly, most units were tested at higher than usual levels of temperature (the nominal temperature at the bottom of the Atlantic Ocean is generally taken to be 10°C).

The life time of the devices being tested could not be observed directly, but could be inferred from a complicated, expensive computer-based test that could only be run periodically. In this test, inspections were performed every 1000 hours. The last inspections were at 5000 hours. The following table gives the data from the experiment, showing the number of failures discovered at each inspection, at each temperature.

	Temperature			
Hours	10°C	40°C	60°C	80°C
0-1000	0	0	2	5
1000-2000	0	2	2	8
2000-3000	0	0	3	1
3000-4000	0	4	0	0
4000-5000	0	4	2	0
> 5000	30	90	11	1
Total	30	100	20	15

1. Without making any assumptions about the shape of the life-time distribution, use the data at 60°C to compute an estimate and an approximate confidence interval for F(5000) at 60°C, where $F(t)$ is the probability of failure before t hours of operation.

2. Using the data at 10°C alone, what can be said about F(5000) at 10°C?

3. Make a meaningful plot or plots of the data and comment on what the plot tells you. You may use the graph paper attached to this question.

The cause of failure was known to be a chemical reaction and the rate of the chemical reaction increases with temperature according to the Arrhenius law from physical chemistry. In particular, the reaction rate is

$$k = \gamma \exp\left(\frac{-E_a}{k_B \times (\mathtt{temp} + 273.15)}\right) = \gamma \exp\left(\frac{-E_a \times 11605}{\mathtt{temp} + 273.15}\right) \tag{1}$$

where temp is temperature in °C and $11605 = 1/k_B$ is the reciprocal of Boltzmann's constant and E_a is known as the "activation energy" of the reaction.

Let T denote the failure time random variable for the electronic devices. Physical/chemical theory also suggests that the life-time distribution at any fixed level of temperature should follow a lognormal distribution with cdf

$$F(t;x) = \Pr[T \le t; x] = \Phi\left[\frac{\log(t) - \mu(x)}{\sigma}\right]$$

where μ and σ are the mean and standard deviation, respectively, of the logarithm of the failure time random variable. Because the life time of a device should be inversely proportional to the rate of the failure-causing chemical reaction, the Arrhenius rate-reaction model implies that $\mu(x) = \beta_0 + \beta_1 x$ where $x = 11605/(\mathtt{temp} + 273.15)$, and $\beta_1 = E_a$ is the activation energy. Because the Arrhenius rate-reaction model is effectively a time-scaling model, the Arrhenius model also implies that σ does not depend on temperature.

4. Provide an expression for the log likelihood for the Arrhenius-lognormal regression model.

5. The ML estimates of the parameters $\theta = (\beta_0, \beta_1, \sigma)$ will be correlated, raising the potential for having a log likelihood that has a shape that might be difficult to maximize. Explain what steps you might take to insure more stable optimization of the likelihood.

The Arrhenius-lognormal regression model was fit to the data using maximum likelihood. The following table provides a summary of the results.

Parameter	ML Estimate	Standard Error
β_0	−13.9	3.37
β_1	.64	.097
σ	.97	.16

An estimate of the variance-covariance matrix for the ML estimates $\widehat{\theta} = (\widehat{\beta}_0, \widehat{\beta}_1, \widehat{\sigma})$ is

$$\widehat{\Sigma}_{\widehat{\theta}} = \begin{bmatrix} 11.37 & -0.327 & -0.348 \\ -0.327 & 0.0094 & 0.0104 \\ -0.348 & 0.0104 & 0.0259 \end{bmatrix}. \quad (2)$$

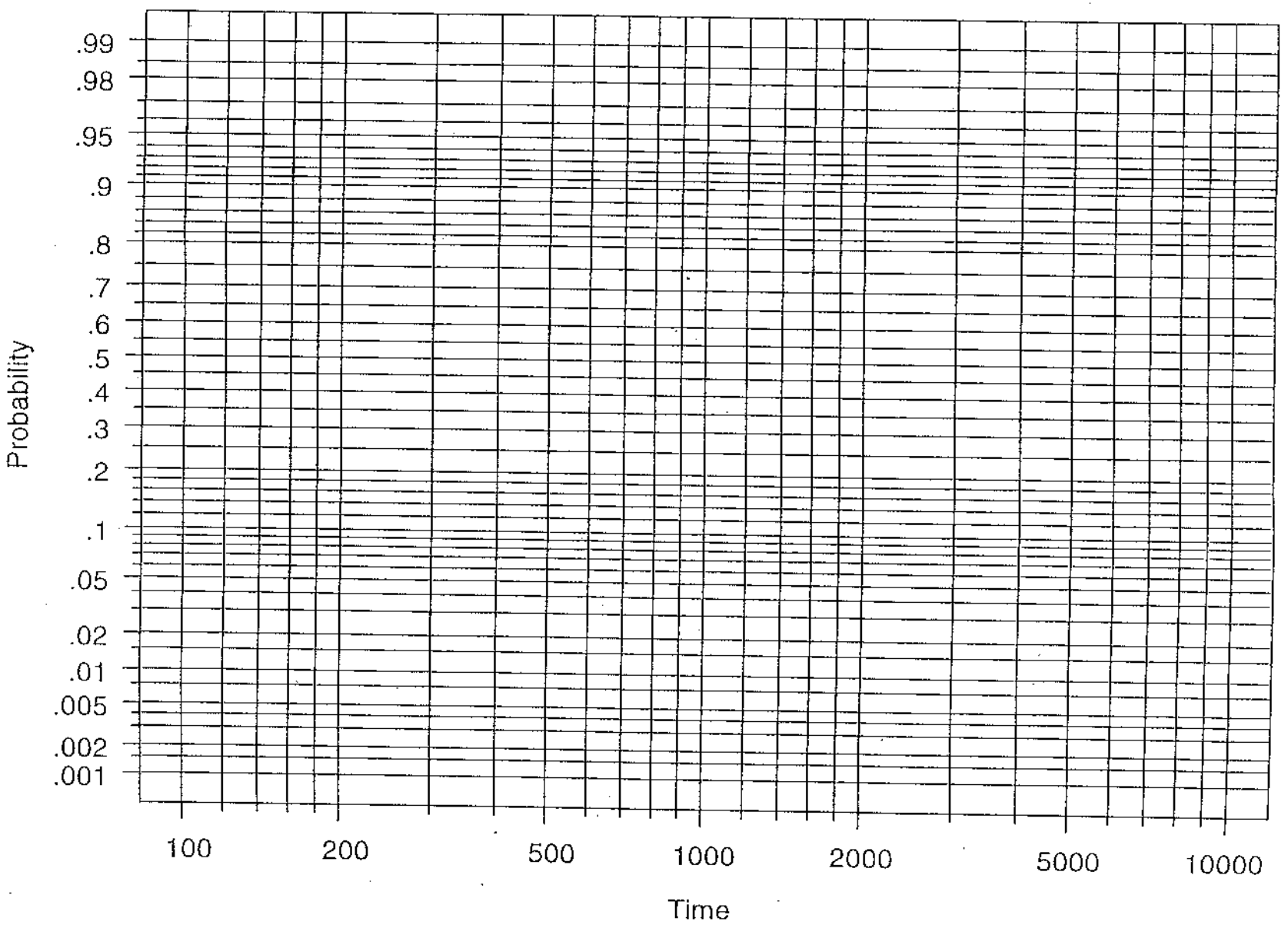
The engineers responsible for the system in which the device was to be used were interested in determining the time at which 10% of the devices in a large population of such devices would fail, assuming operation at 10°C. This time is also known as the .1 quantile or the 10th percentile of the life time distribution. The p quantile of the distribution of T at temperature temp is denoted by $t_p(\texttt{temp})$.

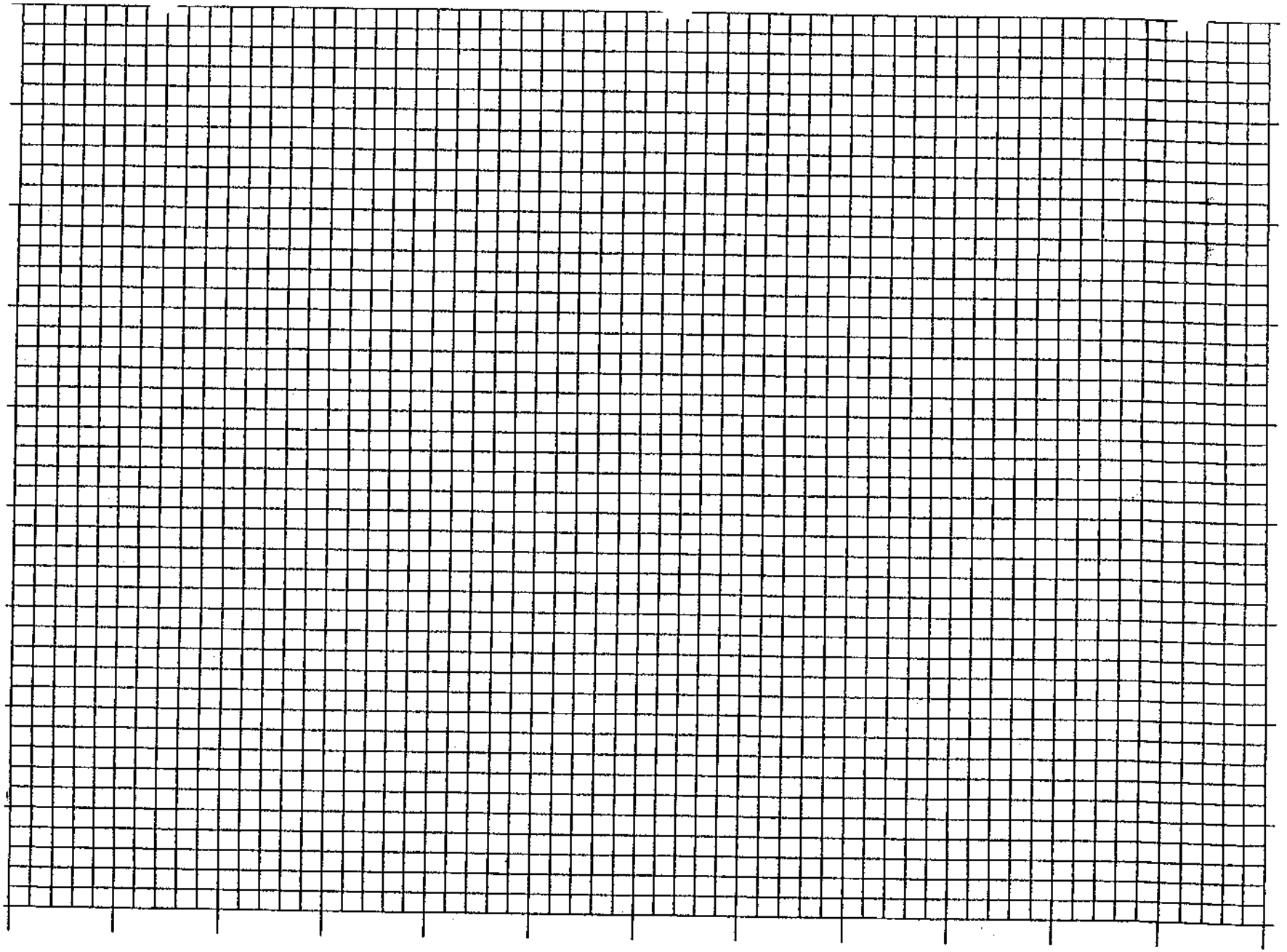
6. With reference to the plot made in part 3, comment on the ability of the assumed model and the available data to provide a useful estimate of the .1 quantile of the life-time distribution at 10°C.

7. Give an expression for $t_p(10)$ and compute the ML estimate of this quantity.

8. Give an expression (numerical computations not needed) for an approximate 95% confidence interval for $t_p(10)$ based on a normal distribution approximation. Provide a technical justification for the approximate interval.

9. Explain how to obtain an approximate 95% confidence interval for $t_p(10)$ based inverting a likelihood ratio test. Describe the method. You do not need to present explicit formulas.


Probability
.99
.98
.95
.9
.8
.7
.6
.5
.4
.3
.2
.1
.05
.02
.01
.005
.002
.001
100
200
500
1000
2000
5000
10000
Time



Method III Solution

Solution
Statistics Ph.D. Preliminary Examination Methods Question, Spring 2001

1. Let x denote the observed number of failures before 5000 hours out of n tested at 60°C. The binomial distribution is appropriate. A simple approximate 95% confidence interval for $F(5000)$ at 60°C can be obtained from

$$x/n \pm z_{.975}\sqrt{\frac{(x/n)(1-x/n)}{n}} = 9/20 \pm 1.96\sqrt{\frac{(9/20)(11/20)}{20}} = [0.232, \quad 0.668].$$

2. Let p denote $F(5000)$ at 10°C. It is easy to show that the ML estimate of p is 0. A conservative $100(1-\alpha)\%$ upper confidence bound for p can be obtained by using the standard method for getting a confidence interval on the parameter of a discrete distribution (e.g., page 420 of Cassela and Berger). For the binomial distribution, given x observed failures, solve

$$\Pr(X \le x) = \sum_{i=0}^{x} \binom{n}{i} (\tilde{p})^i (1-\tilde{p})^{n-i} = \alpha$$

for $\tilde{p}$. With $x = 0$ failures, $P(X = 0) = (1-\tilde{p})^n = \alpha$ and solving for $\tilde{p}$ gives

$$\tilde{p} \;=\; 1 - \alpha^{(1/n)}. \tag{1}$$

For the electronic devices, the 95% conservative upper confidence bound for p is

$$\tilde{p} \;=\; 1 - (.05)^{(1/20)} = 0.1391.$$

Another alternative is to plot the relative likelihood $R(p) = L(p)/L(\widehat{p})$ versus the binomial proportion p. An upper confidence bound for p could be defined by the set values of $p > \widehat{p}$ for which $R(p) > R_c$, calibrating according to the distribution of $R(p)$. With $x = 0$ failures, $R(p) = (1-p)^n$ decreases from 1 to 0 as p goes from 0 to 1. It is not clear how well the usual Chisquare approximation for $-2\log[R(p)]$ would work for this problem. Using the exact distribution of $R(p)$, from the binomial distribution, leads to the conservative upper confidence bound in (1).

3. See the attached graphs.

4. The log likelihood for the Arrhenius-lognormal regression model can be expressed in a number of different ways. For example:

$$\mathcal{L}(\beta_0, \beta_1, \sigma) \;=\; \sum_{i=1}^{n} w_{ij} \log\left[\Phi\left(\frac{\log(t_i^U) - (\beta_0 + \beta_1 x_j)}{\sigma}\right) - \Phi\left(\frac{\log(t_i^L) - (\beta_0 + \beta_1 x_j)}{\sigma}\right)\right]$$

where w_{ij} is the number of units in interval i at temperature j and t_i^L and t_i^U are, respectively, the lower and upper limits of the inspection time intervals and $\Phi[\log(t_i^L) - (\beta_0 + \beta_1 x)/\sigma] \equiv 0$ when $t_i^L = 0$ and $\Phi[\log(t_i^U) - (\beta_0 + \beta_1 x)/\sigma] \equiv 1$ when the upper endpoint of the time interval is unbounded.

5. There are many different ways to do this. One could, for example, obtain the asymptotic variance-covariance matrix of the parameter estimates for the given experimental plan and find linear function of the parameters that would diagonalize the matrix. Another particularly simple fix is to replace x with $x - \bar{x}$ and fit the model

$$\mu = \beta_0^* + \beta_1(x - \bar{x}) = (\beta_0^* - \beta_1\bar{x}) + \beta_1 x$$

effectively bringing the constant term into the middle of the data so that $\widehat{\beta}_0^*$ and $\widehat{\beta}_1$ will be approximately uncorrelated. Then $\beta_0 = \beta_0^* - \beta_1\bar{x}$.

Method III Solution

6. The inspection data, relative to having exact failure times decreases precision in estimation, but not substantially. The lack of resolution in time makes it somewhat more difficult to assess distributional fit. Stopping the test at 5000 hours makes it difficult to assess the adequacy of the assumed relationship between life time and temperature. A major concern is the extrapolation to 10°C. In spite of the fact that units were tested at 10°C, there were no failures at that condition. Physical/chemical theory has been used to suggest both the lognormal distribution and the Arrhenius relationship and to the extent that the model implied by this theory is adequate, it might be reasonable to use it to make the desired predictions at 10°C.

7. Solving

$$p = F(t) = \Pr[T \le t; \texttt{temp}] = \Phi\left[\frac{\log(t_p) - (\beta_0 + \beta_1 x)}{\sigma}\right]$$

for t_p gives

$$t_p = \exp(\beta_0 + \beta_1 x + \Phi^{-1}(p)\sigma) = \exp(-13.9 + .64 \times 40.985 + (-1.2816) \times .97) = 65336$$

where $x = 11605/(\texttt{temp} + 273.15) = 11605/(10 + 273.15) = 40.985$.

8. An approximate $100(1-\alpha)\%$ confidence interval for $\log(t_p)$ can, for example, be based on the standard normal limiting distribution of

$$Z_{\log(\widehat{t}_p)} = \frac{\log(\widehat{t}_p) - \log(t_p)}{\widehat{\text{se}}_{\log(\widehat{t}_p)}}$$

where

$$\widehat{\text{se}}_{\log(\widehat{t}_p)} = \sqrt{\xi' \widehat{\Sigma} \xi},$$

$\xi = (1, x_{10}, \Phi^{-1}(p))$ and $x_{10} = 11605/(10 + 273.15)$. From the definition of quantiles of the standard normal distribution

$$\Pr(-z_{(1-\alpha/2)} \le Z_{\log(\widehat{t}_p)} < z_{(1-\alpha/2)}) \approx 1 - \alpha$$

where $z_{(1-\alpha/2)}$ is the $1-\alpha/2$ quantile of the standard normal distribution. From this it follows that

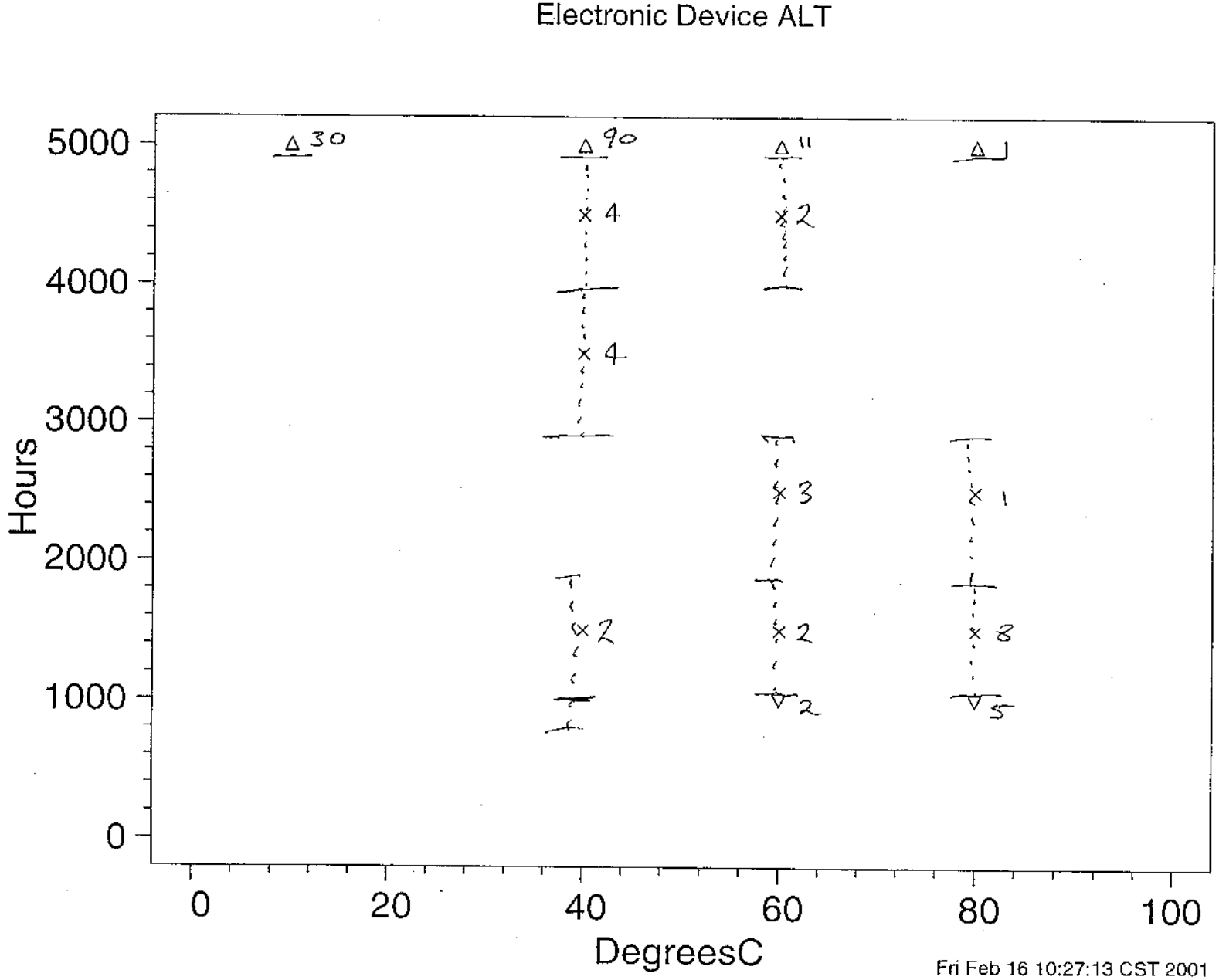
$$\Pr(\log(\widehat{t}_p) - z_{(1-\alpha/2)}\widehat{\text{se}}_{\log(\widehat{t}_p)} \le \log(t_p) < \log(\widehat{t}_p) + z_{(1-\alpha/2)}\widehat{\text{se}}_{\log(\widehat{t}_p)}) \approx 1 - \alpha.$$

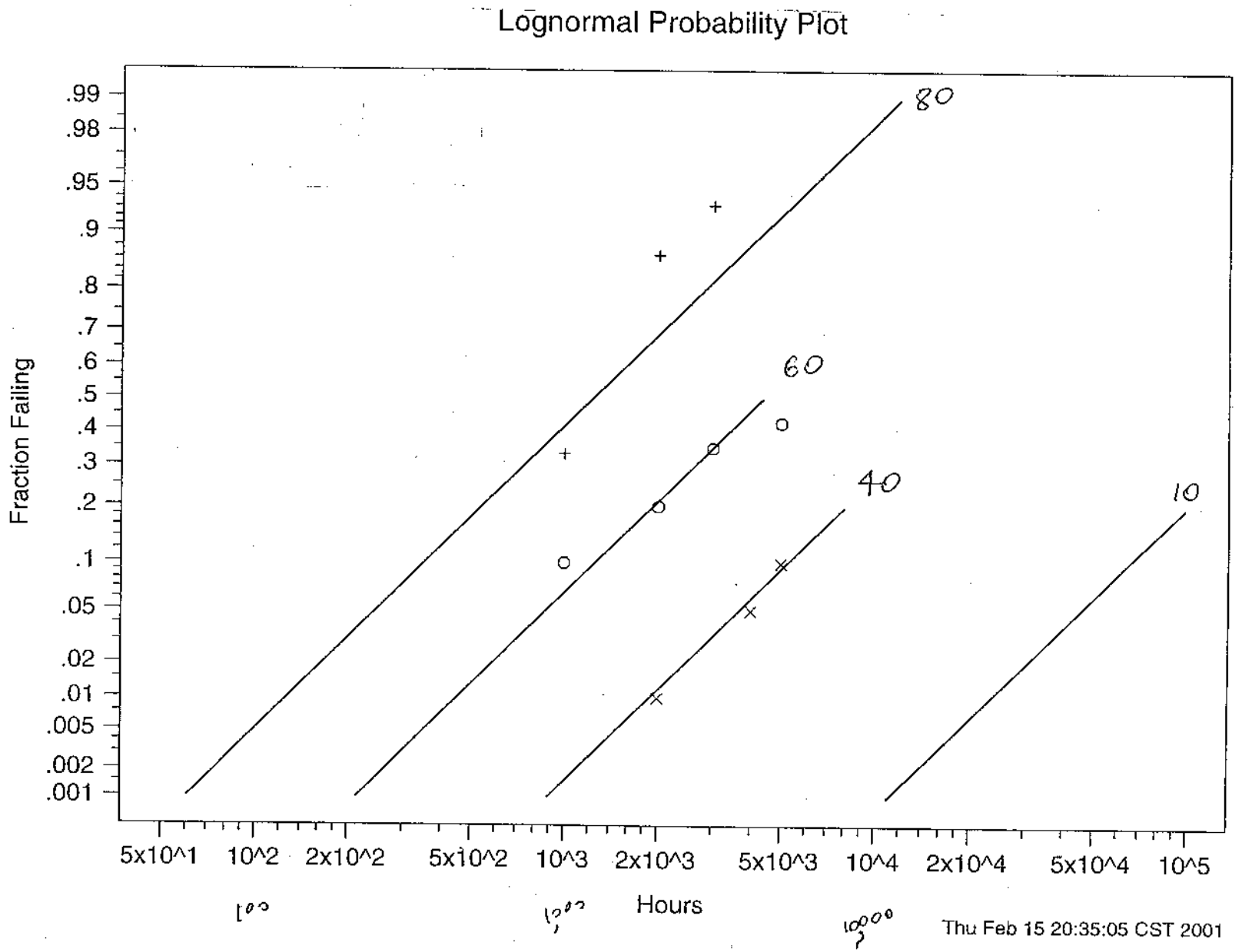
Thus an approximate $100(1-\alpha)\%$ confidence interval for $\log(t_p)$ is

$$[\underline{\log(t_p)}, \quad \overline{\log(t_p)}] = \log(\widehat{t}_p) \pm z_{(1-\alpha/2)}\widehat{\text{se}}_{\log(\widehat{t}_p)}$$

and the corresponding interval for t_p is obtained by taking antilogs of the endpoints.

9. There are several possible ways to do this. The simplest conceptually is to define a likelihood ratio test statistic for testing a null hypothesis about $t_p(10)$ and taking the likelihood-based confidence interval as the set of all values of $t_p(10)$ that would not be rejected at the .05 level of significance.


Electronic Device ALT
Hours
5000
4000
3000
2000
1000
0
0
20
40
60
80
100
DegreesC
Δ 30
Δ 90
Δ 11
Δ 1
x 4
x 4
x 2
x 3
x 1
x 2
x 2
x 8
▽ 2
▽ 5
Fri Feb 16 10:27:13 CST 2001


Lognormal Probability Plot
Fraction Failing
.99
.98
.95
.9
.8
.7
.6
.5
.4
.3
.2
.1
.05
.02
.01
.005
.002
.001
80
60
40
10
5x10^1
10^2
2x10^2
5x10^2
10^3
2x10^3
5x10^3
10^4
2x10^4
5x10^4
10^5
Hours
Thu Feb 15 20:35:05 CST 2001