Stat/IE 531 Exam 1

February 24, 1997 Prof. Vardeman

There are 5 questions on this exam. Do Question #1 (40 points), Question #2 (20 points) and exactly 2 of Questions #3, #4 and #5 (20 points each). BEGIN EACH NEW QUESTION ON A NEW SHEET OF PAPER!

1. Consider the following hypothetical situation. A "variables" process monitoring scheme is to be set up for a production line, and two different measuring devices are available for data gathering purposes. Device A produces precise and expensive measurements and device B produces less precise and less expensive measurements. Let $\sigma_{\text{measurement}}$ for the two devices be respectively σ_{A} and σ_{B}, and suppose that the target for a particular critical diameter for widgets produced on the line is 200.0.

(a) A single widget produced on the line is measured $n = 10$ times with each device and $R_{\text{A}} = 2.0$ and $R_{\text{B}} = 5.0$. Give estimates of σ_{A} and σ_{B}.

(b) Explain why it would not be appropriate to use one of your estimates from (a) as a "σ" for setting up an $\overline{x}$ and R chart pair for monitoring the process based on measurements from one of the devices.

Using device A, 10 consecutive widgets produced on the line (under presumably stable conditions) have (single) measurements with $R = 8.0$.

(c) Set up reasonable control limits for *both* $\overline{x}$ and R for the future monitoring of the process based on samples of size $n = 10$ and measurements from device A. (No need to do arithmetic.)

(d) Combining the information above about the A measurements on 10 consecutive widgets with your answer to (a), under a model that says

$$observed\,diameter = real\,diameter + measurement\,error$$

where "$real\,diameter$" and "$measurement\,error$" are independent, give an estimate of the process standard deviation alone (the standard deviation of the real diameters).

(e) Based on your answers to parts (a) and (d), set up reasonable control limits for *both* $\overline{x}$ and R for the future monitoring of the process based on samples of size $n = 5$ and measurements from the cheaper device, device B. (Again, no need to do arithmetic.)

2. Miscellaneous *short* answers.

(a) Below are 4 hypothetical samples of size $n = 3$. A little calculation (that you do NOT need to do here) shows that ignoring the fact that there are 4 samples and simply computing "s" based on 12 observations will produce a "standard deviation" much larger than s_{pooled}. Why is this?

3,6,5   4,3,1   8,9,6   2,1,4

(b) What do multivariate control charts provide that is not available in separate monitoring of p quality dimensions?

(c) It is often helpful to state "standard errors" (estimated standard deviations) corresponding to point estimates of quantities of interest. In a context where a standard deviation, σ, is to be estimated by $\overline{R}/d_2(n)$ based on r samples of size n, what is a reasonable standard error to announce? (Be sure that your answer is computable from sample data, i.e. doesn't involve any unknown process parameters.)

Choose only 2 of Questions #3, #4 and #5.

3. In applying ANOVA methods to gage R&R studies, one often uses linear combinations of independent mean squares as estimators of their expected values. In fact, it is possible to also produce standard errors (estimated standard deviations) for these linear combinations.

Suppose that $MS_1, MS_2, ..., MS_k$ are independent random variables, $\frac{\nu_i MS_i}{\mathrm{E}MS_i} \sim \chi^2_{\nu_i}$. Consider the random variable

$$L = c_1 MS_1 + c_2 MS_2 + \cdots + c_k MS_k \quad .$$

(a) Find the standard deviation of L.

(b) Your expression from (a) should involve the means $\mathrm{E}MS_i$, that in applications will be unknown. Propose a sensible (data-based) estimator of the standard deviation of L that does not involve these quantities.

(c) Apply your result from (b) to give a sensible standard error for the ANOVA-based estimator of $\sigma^2_{\text{overall}}$.

4. In class, Vardeman considered a "two alarm rule monitoring scheme" due to Wetherill and showed how find the ARL for that scheme by solving two linear equations for quantities L and L^*. It is possible to extend the arguments presented in class and find the *variance* of the run length.

(a) For a generic random variable X, express both Var$X$ and $\mathrm{E}(X+1)^2$ in terms of $\mathrm{E}X$ and $\mathrm{E}X^2$.

(b) Let M be the expected square of the run length for the Wetherill scheme and let M^* be the expected square of the number of additional plotted points required to produce an out of control signal if there has been no signal to date and the current plotted point is in region B. Set up two equations for M and M^* that are linear in M, M^*, L and L^*.

(c) The equations from (b) can be solved simultaneously for M and M^*. Do not take time to solve therm here, but express the variance of the run length for the Wetherill scheme in terms of M, M^*, L and L^*.

5. In class Vardeman presented "rounded data" likelihood methods for normal data with the 2 parameters μ and σ. The same kind of thing can be done for other families of distributions (which can have other numbers of parameters). For example, the exponential distributions with means θ^{-1} can be used. (Here there is the single parameter θ.) These exponential distributions have cdf's

$$F_\theta(x) = \begin{cases} 1 - \exp(-\theta x) & \text{for } x \geq 0 \\ 0 & \text{for } x < 0 \end{cases} .$$

Below is a frequency table for twenty exponential observations that have been rounded to the nearest integer.

rounded value	0	1	2	3	4
frequency	7	8	2	2	1

(a) Write out an expression for the appropriate "rounded data log likelihood function" for this problem,

$$\mathcal{L}(\theta) = \log L(data|\,\theta) \ .$$

(You should be slightly careful here. Exponential random variables only take values in the interval $(0,\infty)$.)

(b) Attached to this exam is a plot of $\mathcal{L}(\theta)$. Use it and identify the maximum likelihood estimate of θ based on the rounded data.

(c) Use the attached plot and make an approximate 90% confidence interval for θ. (The appropriate χ^2 value has 1 associated degree of freedom.)

Stat/IE 531 Exam 2

April 9, 1997 Prof. Vardeman

Do Both of Problems 1 and 2

1. Suppose that standard values of process parameters are $\mu = 17$ and $\sigma = 2.4$.

(a) **Using sample means $\bar{x}$ based on samples of size $n = 4$**, design either a combined high and low side CUSUM scheme (with 0 headstarts) or a EWMA scheme to have an "all ok" ARL of 370 and quickest possible detection of a shift in process of mean of size .6.

(b) If, in fact, the process mean is $\mu = 17.5$ and the process standard deviation is $\sigma = 3.0$, show how you would find the ARL associated with your scheme from (a). (You don't need to actually interpolate in any table, but DO compute the values you would need in order to enter a table, and say which table you must employ.)

2. A discrete variable X can take only values $1, 2, 3, 4$ and 5. Nevertheless, managers decide to "monitor process spread" using the range of samples of size $n = 2$. Suppose, for sake of argument, that under standard plant conditions observations are iid and uniform on the values 1 through 5 (i.e. $P[X = 1] = P[X = 2] = P[X = 3] = P[X = 4] = P[X = 5] = .2$).

(a) Find the distribution of R for this situation. (Note that R has possible values $0, 1, 2, 3$ and 4. You need to reason out the corresponding probabilities.)

(b) The correct answer to part (a) has $\mathrm{E}R = 1.6$. This implies that if many samples of size $n = 2$ are taken and $\bar{R}$ computed, one can expect a mean range near 1.6. Find *and criticize* corresponding normal theory control limits for R.

(c) Suppose that instead of using a normal-based Shewhart chart for R, one decides to use a high side Shewhart-CUSUM scheme with reference value $k_1 = 2$ and starting value 0, that signals the first time any range is 4 or that the CUSUM is 3 or more. Use your answer for (a) and show how to find the ARL for this scheme. (You need not actually carry through the calculations, but show explicitly how to set things up.)

(If you were unable to do part (a), complete part (c) using the (wrong) distribution for R, given by $f(0) = .15,\ f(1) = .2,\ f(2) = .3,\ f(3) = .25$ and $f(4) = .1$.)

Do Only 2 of Problems 3, 4 and 5
(if you choose 5, do only 5A or 5B, not both)

3. A process has a "good" state and a "bad" state. Suppose that when in the good state, the probability that an observation on the process plots outside of control limits is g, while the corresponding probability for the bad state is b. Assume further that if the process is in the good state at time $t-1$, there is a probability d of degradation to the bad state before an observation at time t is made. (Once the process moves into the bad state it stays there until that condition is detected via process monitoring and corrected.)

Find the "ARL"/mean time of alarm, if the process is in the good state at time $t=0$ and observation starts at time $t=1$.

4. Consider the following (nonstandard) process monitoring scheme for a variable X that has ideal value 0. Suppose $h(x)>0$ is a function with $h(x)=h(-x)$ that is decreasing in $|x|$. (h has its maximum at 0 and decreases symmetrically as one moves away from 0.) Then suppose that

i) control limits for X_1 are $\pm h(0)$, and

ii) for $t>1$ control limits for X_t are $\pm h(X_{t-1})$.

(Control limits vary. The larger that $|X_{t-1}|$ is, the tighter are the limits on X_t.)

Discuss how you would find an ARL for this scheme for iid X with marginal probability density f. (Write down an appropriate integral equation, briefly discuss how you would go about solving it and what you would do with the solution in order to find the desired ARL.)

5A. Suppose that a model $\mathcal{F}$ implies that

$$\mathrm{E}_{\mathcal{F}}[Z(t+1)|Z^t]=\frac{1}{2}Z(t)$$

while

$$\mathrm{E}_{\mathcal{F}}[Z(s)|Z^t]=0 \text{ for } s\geq t+2 \ .$$

(I'm not sure that there *is* any such model, but ingore this worry for the purposes of this problem.) If $T(t)\doteq 0$ find optimal (MV) control strategies for two different situtations involving numerical process adjustments a.

(a) First suppose that $A(a,s)=a$ for all $s\geq 1$

(b) Then suppose the impact of a control action is similar to that in (a), except there is one period of delay, i.e.

$$A(a,s)=\begin{cases} a & \text{for } s\geq 2 \\ 0 & \text{for } s=1 \end{cases}$$

5B. Attached are two tables giving some values of an uncontrolled process $Z(t)$ that has target $T(t) \doteq 0$. Suppose that a manipulated variable X is available and that the simple (integral only) control algorithm

$$\Delta X(t) = E(t)$$

will be employed, based on an observed process $Y(t)$ that is the sum of $Z(t)$ and the effects of all relevant changes in X.

Consider two different scenarios:

(a) a change of ΔX in the manipulated variable impacts all subsequent values of $Y(t)$ by the addition of an amount ΔX, and

(b) there is one period of dead time, after which a change of ΔX in the manipulated variable impacts all subsequent values of $Y(t)$ by the addition of an amount ΔX.

Fill in the two tables according to these two scenarios and **then comment** on the lesson they seem to suggest about the impact of dead time on the effectiveness of PID control.

Problem 5B Scenario (a) (no dead time)

t	$Z(t)$	$T(t)$	$Y(t)$	$E(t) = \Delta X(t)$
0	-1	0	-1	
1	-1	0		
2	-1	0		
3	-1	0		
4	-1	0		
5	-1	0		
6	-1	0		
7	-1	0		
8	-1	0		
9	-1	0		

Problem 5B Scenario (b) (one period of dead time)

t	$Z(t)$	$T(t)$	$Y(t)$	$E(t) = \Delta X(t)$
0	-1	0	-1	
1	-1	0		
2	-1	0		
3	-1	0		
4	-1	0		
5	-1	0		
6	-1	0		
7	-1	0		
8	-1	0		
9	-1	0		

Stat/IE 531 Final Exam

May 6, 1997 Prof. Vardeman

1. In their article "Robust Design Through Optimization Techniques," that appeared in *Quality Engineering* in 1994, Lawson and Madrigal modeled an impedance (Z) in a thin film redistribution layer as

$$Z = 41.0 \times \ln\left(\frac{5.98A}{.80B + C}\right)$$

where A is an insulator thickness, B is a line width and C is the line height. Suppose that in manufactured circuits, means and standard deviations of the variables A, B and C are respectively $\mu_A = 25$, $\sigma_A = .3333$, $\mu_B = 15$, $\sigma_B = .2222$, $\mu_C = 5$ and $\sigma_C = .1111$. (The units of A, B and C are all 10^{-6} m, but the article neglects to state the units of Z.)

(a) **Find** an approximate mean and standard deviation to associate with Z in the design of these devices.
(b) Manufacturing variation on **which** of the variables A, B or C **is predicted** to be the largest contributor to variation in impedance Z? **Explain**.

2. Suppose that on $I = 2$ different days (A), $J = 4$ different heats (B) of cast iron are studied, with $m = 3$ tests (C) being made on each. Suppose further that the resulting percent carbon measurements produce $SSA = .0355$, $SSB(A) = .0081$ and $SSC(B(A)) = SSE = .4088$.

(a) If one completely ignores the hierarchical structure of the data set, **what "sample variance"** is produced? **Does** this quantity estimate the variance that would be produced if on many different days a single heat was selected and a single test made? **Explain** carefully! (**Find** the expected value of the grand sample variance under the hierarchical random effects model and **compare** it to this variance of single measurements made on a single day.)
(b) **Give** point estimates of the variance components σ_α^2, σ_β^2 and σ^2.
(c) Your estimate of σ_α^2 should involve a linear combination of mean squares. **Give** the variance of that linear combination in terms of the model parameters and I, J and m. Use that expression and **propose** a sensible estimated standard deviation (a standard error) for this linear combination.

3. Individual items produced on a manufacturer's line may be graded as "Good" (G), "Marginal" (M) or "Defective" (D). Under stable process conditions, each successive item is (independently) G with probability p_G, M with probability p_M and D with probability p_D, where $p_G + p_M + p_D = 1$. Suppose that ultimately, defective items cause three times as much extra expense as marginal ones.

Based on the kind of cost information alluded to above, one might give each inspected item a "score" s according to

$$s = \begin{cases} 3 & \text{if the item is D} \\ 1 & \text{if the item is M} \\ 0 & \text{if the item is G} \end{cases} \quad .$$

It is possible to argue (don't bother to do so here) that $\mathrm{E}(s) = 3p_{\mathrm{D}} + p_{\mathrm{M}}$ and $\mathrm{Var}(s) = 9p_{\mathrm{D}}(1 - p_{\mathrm{D}}) + p_{\mathrm{M}}(1 - p_{\mathrm{M}}) - 3p_{\mathrm{D}}p_{\mathrm{M}}$.

Process Monitoring

(a) **Give** formulas for standards given Shewhart control limits for average scores $\overline{s}$ based on samples of size n. **Describe** how you would obtain the information necessary to calculate limits for future control of $\overline{s}$.
(b) Ultimately, suppose that "standard" values are set at $p_{\mathrm{G}} = .90$, $p_{\mathrm{M}} = .07$ and $p_{\mathrm{D}} = .03$ and $n = 100$ is used for samples of a high volume product. Use a normal approximation to the distribution of $\overline{s}$ and **find** an approximate ARL for your scheme from part (a) if in fact the mix of items shifts to where $p_{\mathrm{G}} = .85$, $p_{\mathrm{M}} = .10$ and $p_{\mathrm{D}} = .05$.
(c) Suppose that one decides to use a high side CUSUM scheme to monitor *individual* scores as they come in one at a time. Consider a scheme with $k_1 = 1$ and no head-start that signals the first time that a CUSUM of scores of at least $h_1 = 6$ is reached. **Set up** an appropriate transition matrix and **say** how you would use that matrix to find an ARL for this scheme for an arbitrary set of probabilities $(p_{\mathrm{G}}, p_{\mathrm{M}}, p_{\mathrm{D}})$.

Acceptance Sampling

(d) Suppose that inspecting an item costs 1/5th of the extra expense caused by an undetected marginal item. A plausible (single sampling) acceptance sampling plan for lots of N=10,000 of these items then accepts the lot if

$$\overline{s} \leq .20 \quad .$$

If rejection of the lot will result in 100% inspection of the remainder, consider the ("perspective B") economic choice of sample size for plans of this form, in particular the comparison of $n = 100$ and $n = 400$ plans. The following table gives some approximate acceptance probabilities for these plans under two sets of probabilities $\boldsymbol{p} = (p_{\mathrm{G}}, p_{\mathrm{M}}, p_{\mathrm{D}})$.

	$n = 100$	$n = 400$
$\boldsymbol{p} = (.9, .07, .03)$	$Pa \approx .76$	$Pa \approx .92$
$\boldsymbol{p} = (.85, .10, .05)$	$Pa \approx .24$	$Pa \approx .08$

Find expected costs for these two plans ($n = 100$ and $n = 400$) if costs are accrued on a per item and per inspection basis and "prior" probabilities of these two sets of process conditions are respectively .8 for $\boldsymbol{p} = (.9, .07, .03)$ and .2 for $\boldsymbol{p} = (.85, .10, .05)$.

4. Consider the classical problem of acceptance sampling plan design. Suppose that one wants plans whose OC "drops" near $p = .03$ (wants $Pa \approx .5$ for $p = .03$) also wants $p = .04$ to have $Pa \approx .05$.

(a) Design an attributes single sampling plan approximately meeting the above criteria.

Suppose that in fact "nonconforming" is defined in terms of a measured variable, X, being less than a lower specification $L = 13$, and that it is sensible to use a normal model for X.

(b) Design a "known σ" variables plan for the above criteria if $\sigma = 1$.
(c) Design an "unknown σ" variables plan for the above criteria.

5. Miscellaneous Short Answer

(a) A single operator measures a single widget diameter 15 times and obtains a range of $R = 3 \times 10^{-4}$ inches. Then this person measures the diameters of 12 different widgets once each and obtains a range of $R = 8 \times 10^{-4}$ inches. **Give** an estimated standard deviation of widget diameters (*not* including measurement error).
(b) Confidence intervals for μ, prediction intervals and tolerance intervals do different jobs. **What** are these different jobs?
(c) SQC novices faced with the task of analyzing a sequence of (say) m individual observations collected over time often do the following: Compute "$\overline{x}$" and "s" from the m data values and apply "control limits" $\overline{x} \pm 3s$ to the m individuals. **Say** why this method of operation is essentially useless.
(d) Attached is a photocopy of a corporate "capability analysis" form. Reading from this form, what do you estimate to be the mean and standard deviation of thread lengths of this type (in .001 inch over nominal)?