STATISTICS DEPARTMENT
M.S. EXAMINATION

PART I
CLOSED BOOK

Friday, May 17, 2002

9:00 a.m. - 1:00 p.m.

Biella Room (Library, First Floor)

Instructions: **Complete *all five* problems. Each problem counts 20 points. Unless otherwise noted, points are allocated approximately equally to lettered parts of a problem. Spend your time accordingly.**

Begin each problem on a new page. Write the problem number and the page number in the specified locations at the top of each page. Also write your chosen ID code number on every page. Please write only within the black borderlines, leaving at least 1" margins on both sides, top and bottom of each page. Write on one side of the page only.

At the end of this part of the exam you will turn in your answers sheets, but you will keep the question sheets and your scratch paper.

Tables of some distributions are provided. Use them as appropriate.

1. A biotech company is learning how to do large-scale production of a protein at two plants, A and B, according to a synthetic process. Below are summary statistics for batches taken from early production runs at each plant. Each observation is recorded as a percentage of the intended protein concentration for this process.

Plant	n	Sample Mean	Sample St Dev
A	30	98.47	8.65
B	30	92.57	11.84

(a) Perform an appropriate *t* test to judge whether the population means for A and B are equal. State the assumptions and the test criterion.

(b) Suppose that the population standard deviation is $\sigma = 10$ and that we want to be reasonably sure to detect a difference of 10 units on this measurement scale. Do you believe that $n = 30$ is an appropriate sample size at each plant? Why or why not? If not, recommend an appropriate value of *n*.

(c) For this part, focus on a sample from *one* of the two plants. We wish to make a 95% confidence interval for the population *variance* at that plant. Suppose this confidence interval will be based on a sample of size $n = 30$. If $\sigma^2 = 100$, give the expected length of this confidence interval.

(d) On the scale of these measurements, batches with values between 85 and 115 are considered to be acceptable. Based on the evidence in the table above for Plant A, about what percentage of the batches is acceptable?

(e) Here are the data used to make the table above. The observations from each plant were collected each morning and are recorded below *in the order in which they were observed* (reading across the rows). Based on what you see below, which of the assumptions you made in part (a) is most seriously in doubt? Explain briefly.

A:	91	92	92	114	96	97	108	105	112	86
	101	99	107	109	107	106	92	98	96	86
	96	91	88	103	85	107	85	101	94	110
B:	102	99	105	113	102	91	92	102	89	104
	95	90	98	110	106	85	97	74	77	77
	94	92	93	99	75	92	70	98	66	90

2. Martha works at her computer terminal. There are three possible states for her computer:

State	Meaning
0	Her computer is not online with the internet.
1	Her computer is online with the internet, but not logged onto her company's site.
2	Her computer is logged onto her company's site.

Let X_n be her computer's state at the end of the nth minute of her workday. Suppose that X_n, $n = 0,1,2,\cdots$ is a Markov chain with the transition probability matrix:

$$P = \begin{array}{c} \\ 0 \\ 1 \\ 2 \end{array} \begin{array}{c} \begin{array}{ccc} 0 & 1 & 2 \end{array} \\ \begin{bmatrix} .4 & .5 & .1 \\ .3 & .5 & .2 \\ 0 & 0 & 1 \end{bmatrix} \end{array}$$

4 (a) Indicate which state(s) are transient and which state(s) are absorbent.

4 (b) If her computer is in state 0 at the end of the 1st minute, then what is the probability that her computer is in state 1 at the end of the 3rd minute?

4 (c) Suppose that $\Pr(X_0 = 0) = 0.8 = 1 - \Pr(X_0 = 1)$. Find the probability that her computer is in state 2 at the end of the 3rd minute.

8 (d) If her computer starts in state 0, then what is the average time until her computer is in state 2?

3. The time until a newly purchased P-3 digital scanner will require repair is a random variable X having an exponential distribution with mean 10τ, where τ is the length of the warranty period.

(a) Find the probability that a newly purchased P-3 scanner will require repair during the warranty period.

The number N of P-3 scanners that a certain store sells on a given day has a Poisson distribution with mean 5. If $N = n$, let Y be the number of scanners, among the n scanners sold, that will require repair during the warranty period (if $N = 0$, then $Y = 0$). Assume independence where it is reasonable to do so.

(b) Find $E(Y)$ and $V(Y)$ by first finding $E(Y \mid N)$ and $V(Y \mid N)$.

(c) Find the probability generating function of Y by conditioning on N. Then identify the distribution of Y.

(d) Find $\mathrm{Cov}(Y, N)$.

(e) For $i = 1, 2, \ldots, 80$, let N_i be the number of P-3 scanners that the store will sell on day i. Suppose that $N_1, N_2, \ldots, N_{80}$ are independent random variables with a common Poisson ($\lambda = 5$) distribution. Find the moment generating function of the total number T of P-3 scanners that the store will sell over the eighty-day period. Then identify the distribution of T.

(f) Use the central limit theorem to approximate $P(T \geq 440)$, where T is defined in part (e).

4. One of the occupational hazards of being an airplane pilot is the hearing loss that results from being exposed to high noise levels. To document the magnitude of the problem, a team of researchers measured the cockpit noise levels of 18 commercial aircraft. The results (in decibels) are listed below.

Plane	Noise Level	Plane	Noise Level
1	74	10	72
2	77	11	90
3	80	12	87
4	82	13	73
5	82	14	83
6	85	15	86
7	80	16	83
8	75	17	83
9	75	18	80

Assume that the distribution of cockpit noise levels from plane to plane is normally distributed.

(a) Compute the m.l.e.'s for the parameters μ and σ^2 and give the estimated values of the m.l.e.'s.

(b) Compute the asymptotic variance for μ and σ^2 and give the estimated values.

(c) Compute the Cramer-Rao lower bound for μ and show that the m.l.e. $\hat{\mu}$ attains this bound. Explain what it means for an estimator to attain the CRLB.

(d) What is the exact distribution of $\hat{\mu}$?

(e) What is the exact distribution of $\hat{\sigma}^2$?

(f) Derive the exact $100(1-\alpha)\%$ confidence interval for the mean μ. Compute the CI.

(g) Derive the exact $100(1-\alpha)\%$ confidence interval for the variance σ^2. Compute the CI.

(h) Using the asymptotic variance, specify the large sample confidence interval for μ. Comment on any differences between the exact confidence interval and the asymptotic confidence interval. Repeat for σ^2.

5. The concrete beams manufactured by the ABC Company are more variable in strength than is considered desirable. In order to study sources of variability, 12 beams are made and tested for strength, with the results shown in the table below.

	Line	
Batch	A	B
1	461 450	428 428
2	346 342	578 584
3	683 693	552 561

(a) Mary discusses the design of this experiment with the engineers who performed it. Based on her discussion, she formulates the following description of the experimental design: Of the 12 beams, 6 were made in each of the two available production lines (A and B). The experiment used three batches of concrete, which were mixed in the company's standard way, but undoubtedly there were some differences among them. Each batch was used to make 4 beams, two in each production line.

(i) Write the model for this design, stating *explicitly* whether each effect is fixed or random, crossed or nested. Define the symbols you use, specify ranges of subscripts, state any restrictions, and give distributional information.

(ii) Complete the DF, MS, and F columns of the ANOVA table below (in which only the sums of squares are provided). State and test appropriate hypotheses at the 5% level. Interpret these results to describe the most important source(s) of variability?

```
Analysis of Variance for Strength

Source          DF          SS          MS          F
------------------------------------------------------
Line                      2028
Batch                    78041
Line*Batch               72189
Error                      177
------------------------------------------------------
Total                   152436
```

(b) John conducts an in-depth discussion with the engineers concerning the design of this experiment. Here is his understanding of the experimental design: Of the 12 beams, 6 were made in each of the two available production lines — which, it turns out, are located about 50 miles apart. The experiment used 6 batches of concrete, 3 of them mixed in the standard way at each location. Each batch was used to make 2 beams (strengths shown in each cell of the table).

(i) Write the model for this experiment, stating explicitly whether each effect is fixed or random, crossed or nested. Define the symbols you use, specify ranges of subscripts, state any restrictions, and give distributional information.

(ii) Modify Mary's ANOVA table to match John's understanding of the design. (Hint: "combine" two lines of the table.) State and test appropriate hypotheses at the 5% level. Interpret these results to describe the most important source(s) of variability.

Upper-tail Areas for the Normal Curve

z	.00	.01	.02	.03	.04	.05	.06	.07	.08	.09
0.00	.5000	.4960	.4920	.4880	.4840	.4801	.4761	.4721	.4681	.4641
0.10	.4602	.4562	.4522	.4483	.4443	.4404	.4364	.4325	.4286	.4247
0.20	.4207	.4168	.4129	.4090	.4052	.4013	.3974	.3936	.3897	.3859
0.30	.3821	.3783	.3745	.3707	.3669	.3632	.3594	.3557	.3520	.3483
0.40	.3446	.3409	.3372	.3336	.3300	.3264	.3228	.3192	.3156	.3121
0.50	.3085	.3050	.3015	.2981	.2946	.2912	.2877	.2843	.2810	.2776
0.60	.2743	.2709	.2676	.2643	.2611	.2578	.2546	.2514	.2483	.2451
0.70	.2420	.2389	.2358	.2327	.2296	.2266	.2236	.2206	.2177	.2148
0.80	.2119	.2090	.2061	.2033	.2005	.1977	.1949	.1922	.1894	.1867
0.90	.1841	.1814	.1788	.1762	.1736	.1711	.1685	.1660	.1635	.1611
1.00	.1587	.1562	.1539	.1515	.1492	.1469	.1446	.1423	.1401	.1379
1.10	.1357	.1335	.1314	.1292	.1271	.1251	.1230	.1210	.1190	.1170
1.20	.1151	.1131	.1112	.1093	.1075	.1056	.1038	.1020	.1003	.0985
1.30	.0968	.0951	.0934	.0918	.0901	.0885	.0869	.0853	.0838	.0823
1.40	.0808	.0793	.0778	.0764	.0749	.0735	.0721	.0708	.0694	.0681
1.50	.0668	.0655	.0643	.0630	.0618	.0606	.0594	.0582	.0571	.0559
1.60	.0548	.0537	.0526	.0516	.0505	.0495	.0485	.0475	.0465	.0455
1.70	.0446	.0436	.0427	.0418	.0409	.0401	.0392	.0384	.0375	.0367
1.80	.0359	.0351	.0344	.0336	.0329	.0322	.0314	.0307	.0301	.0294
1.90	.0287	.0281	.0274	.0268	.0262	.0256	.0250	.0244	.0239	.0233
2.00	.0228	.0222	.0217	.0212	.0207	.0202	.0197	.0192	.0188	.0183
2.10	.0179	.0174	.0170	.0166	.0162	.0158	.0154	.0150	.0146	.0143
2.20	.0139	.0136	.0132	.0129	.0125	.0122	.0119	.0116	.0113	.0110
2.30	.0107	.0104	.0102	.0099	.0096	.0094	.0091	.0089	.0087	.0084
2.40	.0082	.0080	.0078	.0075	.0073	.0071	.0069	.0068	.0066	.0064
2.50	.0062	.0060	.0059	.0057	.0055	.0054	.0052	.0051	.0049	.0048
2.60	.0047	.0045	.0044	.0043	.0041	.0040	.0039	.0038	.0037	.0036
2.70	.0035	.0034	.0033	.0032	.0031	.0030	.0029	.0028	.0027	.0026
2.80	.0026	.0025	.0024	.0023	.0023	.0022	.0021	.0021	.0020	.0019
2.90	.0019	.0018	.0018	.0017	.0016	.0016	.0015	.0015	.0014	.0014
3.00	.0013	.0013	.0013	.0012	.0012	.0011	.0011	.0011	.0010	.0010

z	Area
3.500	.00023263
4.000	.00003167
4.500	.00000340
5.000	.00000029

Percentage Points of the t Distribution

df	a = .1	a = .05	a = .025	a = .01	a = .005	a = .001
1	3.078	6.314	12.706	31.821	63.657	318.309
2	1.886	2.920	4.303	6.965	9.925	22.327
3	1.638	2.353	3.182	4.541	5.841	10.215
4	1.533	2.132	2.776	3.747	4.604	7.173
5	1.476	2.015	2.571	3.365	4.032	5.893
6	1.440	1.943	2.447	3.143	3.707	5.208
7	1.415	1.895	2.365	2.998	3.499	4.785
8	1.397	1.860	2.306	2.896	3.355	4.501
9	1.383	1.833	2.262	2.821	3.250	4.297
10	1.372	1.812	2.228	2.764	3.169	4.144
11	1.363	1.796	2.201	2.718	3.106	4.025
12	1.356	1.782	2.179	2.681	3.055	3.930
13	1.350	1.771	2.160	2.650	3.012	3.852
14	1.345	1.761	2.145	2.624	2.977	3.787
15	1.341	1.753	2.131	2.602	2.947	3.733
16	1.337	1.746	2.120	2.583	2.921	3.686
17	1.333	1.740	2.110	2.567	2.898	3.646
18	1.330	1.734	2.101	2.552	2.878	3.610
19	1.328	1.729	2.093	2.539	2.861	3.579
20	1.325	1.725	2.086	2.528	2.845	3.552
21	1.323	1.721	2.080	2.518	2.831	3.527
22	1.321	1.717	2.074	2.508	2.819	3.505
23	1.319	1.714	2.069	2.500	2.807	3.485
24	1.318	1.711	2.064	2.492	2.797	3.467
25	1.316	1.708	2.060	2.485	2.787	3.450
26	1.315	1.706	2.056	2.479	2.779	3.435
27	1.314	1.703	2.052	2.473	2.771	3.421
28	1.313	1.701	2.048	2.467	2.763	3.408
29	1.311	1.699	2.045	2.462	2.756	3.396
30	1.310	1.697	2.042	2.457	2.750	3.385
40	1.303	1.684	2.021	2.423	2.704	3.307
60	1.296	1.671	2.000	2.390	2.660	3.232
120	1.289	1.658	1.980	2.358	2.617	3.160
240	1.285	1.651	1.970	2.342	2.596	3.125
inf.	1.282	1.645	1 960	2.326	2.576	3.090

CUMULATIVE DISTRIBUTION OF CHI-SQUARE*

Degrees of Freedom	Probability of a Greater Value												
	0.995	0.990	0.975	0.950	0.900	0.750	0.500	0.250	0.100	0.050	0.025	0.010	0.005
1	....	....	....	----	0.02	0.10	0.45	1.32	2.71	3.84	5.02	6.63	7.88
2	0.01	0.02	0.05	0.10	0.21	0.58	1.39	2.77	4.61	5.99	7.38	9.21	10.60
3	0.07	0.11	0.22	0.35	0.58	1.21	2.37	4.11	6.25	7.81	9.35	11.34	12.84
4	0.21	0.30	0.48	0.71	1.06	1.92	3.36	5.39	7.78	9.49	11.14	13.28	14.86
5	0.41	0.55	0.83	1.15	1.61	2.67	4.35	6.63	9.24	11.07	12.83	15.09	16.75
6	0.68	0.87	1.24	1.64	2.20	3.45	5.35	7.84	10.64	12.59	14.45	16.81	18.55
7	0.99	1.24	1.69	2.17	2.83	4.25	6.35	9.04	12.02	14.07	16.01	18.48	20.28
8	1.34	1.65	2.18	2.73	3.49	5.07	7.34	10.22	13.36	15.51	17.53	20.09	21.96
9	1.73	2.09	2.70	3.33	4.17	5.90	8.34	11.39	14.68	16.92	19.02	21.67	23.59
10	2.16	2.56	3.25	3.94	4.87	6.74	9.34	12.55	15.99	18.31	20.48	23.21	25.19
11	2.60	3.05	3.82	4.57	5.58	7.58	10.34	13.70	17.28	19.68	21.92	24.72	26.76
12	3.07	3.57	4.40	5.23	6.30	8.44	11.34	14.85	18.55	21.03	23.34	26.22	28.30
13	3.57	4.11	5.01	5.89	7.04	9.30	12.34	15.98	19.81	22.36	24.74	27.69	29.82
14	4.07	4.66	5.63	6.57	7.79	10.17	13.34	17.12	21.06	23.68	26.12	29.14	31.32
15	4.60	5.23	6.27	7.26	8.55	11.04	14.34	18.25	22.31	25.00	27.49	30.58	32.80
16	5.14	5.81	6.91	7.96	9.31	11.91	15.34	19.37	23.54	26.30	28.85	32.00	34.27
17	5.70	6.41	7.56	8.67	10.09	12.79	16.34	20.49	24.77	27.59	30.19	33.41	35.72
18	6.26	7.01	8.23	9.39	10.86	13.68	17.34	21.60	25.99	28.87	31.53	34.81	37.16
19	6.84	7.63	8.91	10.12	11.65	14.56	18.34	22.72	27.20	30.14	32.85	36.19	38.58
20	7.43	8.26	9.59	10.85	12.44	15.45	19.34	23.83	28.41	31.41	34.17	37.57	40.00
21	8.03	8.90	10.28	11.59	13.24	16.34	20.34	24.93	29.62	32.67	35.48	38.93	41.40
22	8.64	9.54	10.98	12.34	14.04	17.24	21.34	26.04	30.81	33.92	36.78	40.29	42.80
23	9.26	10.20	11.69	13.09	14.85	18.14	22.34	27.14	32.01	35.17	38.08	41.64	44.18
24	9.89	10.86	12.40	13.85	15.66	19.04	23.34	28.24	33.20	36.42	39.36	42.98	45.56
25	10.52	11.52	13.12	14.61	16.47	19.94	24.34	29.34	34.38	37.65	40.65	44.31	46.93
26	11.16	12.20	13.84	15.38	17.29	20.84	25.34	30.43	35.56	38.89	41.92	45.64	48.29
27	11.81	12.88	14.57	16.15	18.11	21.75	26.34	31.53	36.74	40.11	43.19	46.96	49.64
28	12.46	13.56	15.31	16.93	18.94	22.66	27.34	32.62	37.92	41.34	44.46	48.28	50.99
29	13.12	14.26	16.05	17.71	19.77	23.57	28.34	33.71	39.09	42.56	45.72	49.59	52.34
30	13.79	14.95	16.79	18.49	20.60	24.48	29.34	34.80	40.26	43.77	46.98	50.89	53.67
40	20.71	22.16	24.43	26.51	29.05	33.66	39.34	45.62	51.80	55.76	59.34	63.69	66.77
50	27.99	29.71	32.36	34.76	37.69	42.94	49.33	56.33	63.17	67.50	71.42	76.15	79.49
60	35.53	37.48	40.48	43.19	46.46	52.29	59.33	66.98	74.40	79.08	83.30	88.38	91.95
70	43.28	45.44	48.76	51.74	55.33	61.70	69.33	77.58	85.53	90.53	95.02	100.42	104.22
80	51.17	53.54	57.15	60.39	64.28	71.14	79.33	88.13	96.58	101.88	106.63	112.33	116.32
90	59.20	61.75	65.65	69.13	73.29	80.62	89.33	98.64	107.56	113.14	118.14	124.12	128.30
100	67.33	70.06	74.22	77.93	82.36	90.13	99.33	109.14	118.50	124.34	129.56	135.81	140.17

centage Points of the F-Distribution

$f_{0.05,\nu_1,\nu_2}$

		Degrees of freedom for the numerator (ν_1)																		
	ν_2 \ ν_1	1	2	3	4	5	6	7	8	9	10	12	15	20	24	30	40	60	120	∞
Degrees of freedom for the denominator (ν_2)	1	161.4	199.5	215.7	224.6	230.2	234.0	236.8	238.9	240.5	241.9	243.9	245.9	248.0	249.1	250.1	251.1	252.2	253.3	254.3
	2	18.51	19.00	19.16	19.25	19.30	19.33	19.35	19.37	19.38	19.40	19.41	19.43	19.45	19.45	19.46	19.47	19.48	19.49	19.50
	3	10.13	9.55	9.28	9.12	9.01	8.94	8.89	8.85	8.81	8.79	8.74	8.70	8.66	8.64	8.62	8.59	8.57	8.55	8.53
	4	7.71	6.94	6.59	6.39	6.26	6.16	6.09	6.04	6.00	5.96	5.91	5.86	5.80	5.77	5.75	5.72	5.69	5.66	5.63
	5	6.61	5.79	5.41	5.19	5.05	4.95	4.88	4.82	4.77	4.74	4.68	4.62	4.56	4.53	4.50	4.46	4.43	4.40	4.36
	6	5.99	5.14	4.76	4.53	4.39	4.28	4.21	4.15	4.10	4.06	4.00	3.94	3.87	3.84	3.81	3.77	3.74	3.70	3.67
	7	5.59	4.74	4.35	4.12	3.97	3.87	3.79	3.73	3.68	3.64	3.57	3.51	3.44	3.41	3.38	3.34	3.30	3.27	3.23
	8	5.32	4.46	4.07	3.84	3.69	3.58	3.50	3.44	3.39	3.35	3.28	3.22	3.15	3.12	3.08	3.04	3.01	2.97	2.93
	9	5.12	4.26	3.86	3.63	3.48	3.37	3.29	3.23	3.18	3.14	3.07	3.01	2.94	2.90	2.86	2.83	2.79	2.75	2.71
	10	4.96	4.10	3.71	3.48	3.33	3.22	3.14	3.07	3.02	2.98	2.91	2.85	2.77	2.74	2.70	2.66	2.62	2.58	2.54
	11	4.84	3.98	3.59	3.36	3.20	3.09	3.01	2.95	2.90	2.85	2.79	2.72	2.65	2.61	2.57	2.53	2.49	2.45	2.40
	12	4.75	3.89	3.49	3.26	3.11	3.00	2.91	2.85	2.80	2.75	2.69	2.62	2.54	2.51	2.47	2.43	2.38	2.34	2.30
	13	4.67	3.81	3.41	3.18	3.03	2.92	2.83	2.77	2.71	2.67	2.60	2.53	2.46	2.42	2.38	2.34	2.30	2.25	2.21
	14	4.60	3.74	3.34	3.11	2.96	2.85	2.76	2.70	2.65	2.60	2.53	2.46	2.39	2.35	2.31	2.27	2.22	2.18	2.13
	15	4.54	3.68	3.29	3.06	2.90	2.79	2.71	2.64	2.59	2.54	2.48	2.40	2.33	2.29	2.25	2.20	2.16	2.11	2.07
	16	4.49	3.63	3.24	3.01	2.85	2.74	2.66	2.59	2.54	2.49	2.42	2.35	2.28	2.24	2.19	2.15	2.11	2.06	2.01
	17	4.45	3.59	3.20	2.96	2.81	2.70	2.61	2.55	2.49	2.45	2.38	2.31	2.23	2.19	2.15	2.10	2.06	2.01	1.96
	18	4.41	3.55	3.16	2.93	2.77	2.66	2.58	2.51	2.46	2.41	2.34	2.27	2.19	2.15	2.11	2.06	2.02	1.97	1.92
	19	4.38	3.52	3.13	2.90	2.74	2.63	2.54	2.48	2.42	2.38	2.31	2.23	2.16	2.11	2.07	2.03	1.98	1.93	1.88
	20	4.35	3.49	3.10	2.87	2.71	2.60	2.51	2.45	2.39	2.35	2.28	2.20	2.12	2.08	2.04	1.99	1.95	1.90	1.84
	21	4.32	3.47	3.07	2.84	2.68	2.57	2.49	2.42	2.37	2.32	2.25	2.18	2.10	2.05	2.01	1.96	1.92	1.87	1.81
	22	4.30	3.44	3.05	2.82	2.66	2.55	2.46	2.40	2.34	2.30	2.23	2.15	2.07	2.03	1.98	1.94	1.89	1.84	1.78
	23	4.28	3.42	3.03	2.80	2.64	2.53	2.44	2.37	2.32	2.27	2.20	2.13	2.05	2.01	1.96	1.91	1.86	1.81	1.76
	24	4.26	3.40	3.01	2.78	2.62	2.51	2.42	2.36	2.30	2.25	2.18	2.11	2.03	1.98	1.94	1.89	1.84	1.79	1.73
	25	4.24	3.39	2.99	2.76	2.60	2.49	2.40	2.34	2.28	2.24	2.16	2.09	2.01	1.96	1.92	1.87	1.82	1.77	1.71
	26	4.23	3.37	2.98	2.74	2.59	2.47	2.39	2.32	2.27	2.22	2.15	2.07	1.99	1.95	1.90	1.85	1.80	1.75	1.69
	27	4.21	3.35	2.96	2.73	2.57	2.46	2.37	2.31	2.25	2.20	2.13	2.06	1.97	1.93	1.88	1.84	1.79	1.73	1.67
	28	4.20	3.34	2.95	2.71	2.56	2.45	2.36	2.29	2.24	2.19	2.12	2.04	1.96	1.91	1.87	1.82	1.77	1.71	1.65
	29	4.18	3.33	2.93	2.70	2.55	2.43	2.35	2.28	2.22	2.18	2.10	2.03	1.94	1.90	1.85	1.81	1.75	1.70	1.64
	30	4.17	3.32	2.92	2.69	2.53	2.42	2.33	2.27	2.21	2.16	2.09	2.01	1.93	1.89	1.84	1.79	1.74	1.68	1.62
	40	4.08	3.23	2.84	2.61	2.45	2.34	2.25	2.18	2.12	2.08	2.00	1.92	1.84	1.79	1.74	1.69	1.64	1.58	1.51
	60	4.00	3.15	2.76	2.53	2.37	2.25	2.17	2.10	2.04	1.99	1.92	1.84	1.75	1.70	1.65	1.59	1.53	1.47	1.39
	120	3.92	3.07	2.68	2.45	2.29	2.17	2.09	2.02	1.96	1.91	1.83	1.75	1.66	1.61	1.55	1.55	1.43	1.35	1.25
	∞	3.84	3.00	2.60	2.37	2.21	2.10	2.01	1.94	1.88	1.83	1.75	1.67	1.57	1.52	1.46	1.39	1.32	1.22	1.00

Prob #1, C.B

Answers:

(a) A two-sample *t* test is appropriate provided that the two populations are approximately normal and that the data are a random sample from their respective populations. Based on the similar sample standard deviations, seems safe to assume, for purposes of doing a *t* test, that the population standard deviations are equal. The computations are routine. Here is a Minitab printout of the results for the single-variance *t* test. Because of the large DF, it would be OK to assume that the *t* statistic is essentially normally distributed and base the P-value on normal tables (obtaining 0.028).

```
Two-sample T for Assay

Plant          N       Mean      StDev   SE Mean
1             30      98.47       8.65       1.6
2             30       92.6       11.8       2.2

Difference = mu (1) - mu (2)
Estimate for difference:  5.90
95% CI for difference: (0.54, 11.26)
T-Test of difference = 0 (vs not =): T-Value = 2.20  P-Value = 0.032  DF = 58
Both use Pooled StDev = 10.4
```

If you do the separate-variances test, results are very similar:

```
T-Value = 2.20  P-Value = 0.032  DF = 53
```

(b) It is OK to use the normal approximation and use standard formulas (e.g., see Ott/Longnecker, page 318) or direct computation using normal curves. Using $\Delta = \sigma = 10$ and $\alpha = \beta = 0.05$, one obtains $n = 2[(\sigma/\Delta)\,(z_{\alpha/2} + z_{\beta/2})]^2 \approx 31$. Other values of α and β might be used. For comparison, the exact computations below, using the noncentral *t* distribution are given by Minitab (menu: **STAT > power and sample size > 2-sample t**).

```
MTB > Power;
SUBC>   TTwo;
SUBC>     Difference 10;
SUBC>     Power .8 .9 .95;
SUBC>     Sigma 10.0.

Power and Sample Size

2-Sample t Test

Testing mean 1 = mean 2 (versus not =)
Calculating power for mean 1 = mean 2 + difference
Alpha = 0.05  Sigma = 10

             Sample  Target  Actual
Difference     Size   Power   Power
        10       17  0.8000  0.8070
        10       23  0.9000  0.9125
        10       27  0.9500  0.9501
```

If we choose $\alpha = \beta = 0.05$ and bear in mind that the values chosen for σ and Δ must be somewhat arbitrary in practice, there seems to be no reason to recommend a change from $n = 30$.

(c) The confidence interval for s^2 is based on the chi-squared distribution with DF = 29, for which the values 16.04 and 45.7 cut off 2.5% from the respective tails. So a 95% CI for s^2 is given by $(29s^2/45.72,\ 29s^2/16.04)$, where $E(29s^2) = 29(\sigma^2) = 2900$. Thus, the expected length of the CI is $2900(1/16.04 - 1/45.72) = 117.4$. [The expected length of the CI for σ would require finding $E(s) \neq \sigma$.]

(d) Assuming data distributed roughly *N*(100, 10), we are asking what percentage of the observations will lie within 1.5 standard deviations of the mean. Then the answer is about 87%. Similar elementary computations using $\mu = 98.5$ and $\sigma = 8.6$ give about 91%. Because μ and σ can only be estimated, the answer must be approximate. Any answer around 90% supported by a reasonable argument is acceptable. While it is true that all 30 of the batches at hand are within acceptable limits, it would be grossly overoptimistic to assume that batches outside the limits would *never* be produced.

(e) The data in the second sample appear to trend downward with time, so that the second sample is apparently not a random sample from a fixed distribution. Here is a Minitab time series plot of the data for Plant B.

```
  B       -    4
          -              4
          -
     105.0+   3      0    5
          - 1   5  8
          -  2          3          4   8
          -          1     7
          -      7             123  6
      90.0+     6  9  2                 0
          -
          -                6
          -
          -                   90
      75.0+                  8      5
          -
          -                           7
          -                             9
           +---------+---------+---------+
           0        10        20        30
```

Working by hand, you were not necessarily expected to have drawn this time series plot *(although the italics might have been a clue)*; we show it here because it is so revealing.

Here are some of the more likely ways in which you *were* expected to detect the downward trend: Just scanning across by eye, you should notice few low values in the first row for Plant B and many low values in the third row of the data (particularly striking: *none* 100 or above). By the criterion of part (d), three of the last 15 values are "unacceptably" low (below 75) and three more are near the lower acceptable limit. If you make a dot plot or stem plot for Plant B by hand, you should notice a pattern of low values while plotting the second half of the sample.

[In fact, the data were randomly generated: The first sample is N(100, 10). In the second sample, the first 15 observations are from N(100, 10) and the second 15 are from N(87, 10). This is a simulated reconstruction of proprietary data exhibiting similar behavior.]

#2 CB Answer

(a) States 0 & 1 are transient and State 2 is absorbent.

(b) $P_{01}(2) = (.4)(.5) + (.5)(.5) + (.1)0 = .45$

(c) $P[X_3 = 2] = P[X_0 \overset{0.8}{=} 0]\, P_{02}^{(3)} + P[X_0 \overset{0.2}{=} 1]\, P_{12}^{(3)} + P[X_0 \overset{0}{=} 2]\, P_{22}^{(3)}$

$= .8\, P_{02}(3) + .2\, P_{12}(3) = .8(.361) + .2(.437) = .3762$

$$P^2 = \begin{bmatrix} .31 & .45 & .24 \\ .27 & .40 & .33 \\ 0 & 0 & 1 \end{bmatrix} \qquad P^3 = \begin{bmatrix} .259 & .38 & .361 \\ .228 & .335 & .437 \\ 0 & 0 & 1 \end{bmatrix}$$

(d) Let ν_0 = average time from state 0 to state 2
ν_1 = " " " " 1 " " 2

$\nu_0 = 1 + .4\nu_0 + .5\nu_1$

$\nu_1 = 1 + .3\nu_0 + .5\nu_1 \Rightarrow \nu_1 = \dfrac{1 + .3\nu_0}{.5}$

$$\nu_0 = 1 + .4\nu_0 + .5\left(\frac{1 + .3\nu_0}{.5}\right)$$

$= 1 + .4\nu_0 + 1 + .3\nu_0$

$.3\nu_0 = 2$

$\nu_0 = \dfrac{2}{.3} = \dfrac{20}{3}$

Answer CB

③ a) $P(X<\tau) = \int_0^{\tau} \frac{1}{10\tau} e^{-\frac{x}{10\tau}} dx = -e^{-\frac{x}{10\tau}}\Big|_0^{\tau} = 1-e^{-\frac{1}{10}}$

$$\boxed{\begin{aligned}&= 1-e^{-.1}\\ &\text{or } .0952\end{aligned}}$$

b) Assuming independence among scanners,

$Y|N=n$ is Binomial n, $p = 1-e^{-.1}$

So $E(Y|N=n) = np$ and $E(Y|N) = Np$

and $V(Y|N=n) = npq$ and $V(Y|N) = Npq$.

Hence, $E(Y) = E(E(Y|N)) = E(Np) = pE(N) = p\lambda$

and $\qquad = (1-e^{-.1})(5) = \boxed{.476}$

$$V(Y) = E(V(Y|N)) + V(E(Y|N))$$
$$= E(Npq) + V(Np) = pq\,E(N) + p^2 V(N)$$
$$= pq\lambda + p^2\lambda = p(1-p)\lambda + p^2\lambda = p\lambda = \boxed{.476}$$

c) ~~$g_Y(t) = E(t^Y)$~~ $= E(E(t^Y|N)) = E((pt+q)^N)$

$$= \sum_{n=0}^{\infty} (pt+q)^n \frac{e^{-\lambda}\lambda^n}{n!} = e^{-\lambda} \sum_{n=0}^{\infty} \frac{[\lambda(pt+q)]^n}{n!}$$
$$= e^{-\lambda} \cdot e^{\lambda(pt+q)} = e^{\lambda(pt+q-1)} = e^{\lambda(pt-p)} = e^{\lambda p(t-1)}$$

$\therefore g_Y(t) = e^{(\lambda p)(t-1)}$. This is the p.g.f. of $Po(\lambda p)$

$\therefore$ Y must be $Po(\lambda p)$ or $Po(.476)$

(Note: Students cannot use this result to avoid doing the required work in part b.)

d) From (b), $E(Y) = \lambda p$. $\quad E(N) = \lambda$.

$$E(YN) = E(E(YN|N)) = E(N \cdot E(Y|N))$$
$$= E(N(Np)) = p\,E(N^2) = p(V(N) + E(N)^2)$$
$$= p(\lambda + \lambda^2).$$

So $\mathrm{Cov}(Y,N) = E(YN) - E(Y)E(N) = p(\lambda+\lambda^2) - (\lambda p)(\lambda)$

$$= p\lambda = \boxed{.476}$$

3(e) $T = \sum_{i=1}^{80} N_i$. $m_T(t) = E(e^{tT}) = E(e^{t\sum_{i=1}^{80} N_i})$

$$= E\left(\prod_{i=1}^{80} e^{tN_i}\right) = \prod_{i=1}^{80} m_{N_i}(t) = [m_N(t)]^{80}$$

$$= \left[e^{\lambda(e^t-1)}\right]^{80} = e^{(\lambda 80)(e^t-1)} = e^{400(e^t-1)}$$

$(\lambda = 5)$

$\therefore$ T is Po(400)

f) T is the sum of a large number of iid random variables and $E(T) = 400$ and $SD(T) = \sqrt{400} = 20$. So $P(T \geqslant 440) \cong P(Z \geqslant \frac{440-400}{20}) = P(Z \geqslant 2.0)$

$\cong \boxed{.0228}$

<u>solution.</u> #4 C6

a) $$f(x \mid \mu, \sigma^2) = \frac{1}{\sqrt{2\pi}\,\sigma} e^{-\frac{(x-\mu)^2}{2\sigma^2}} \qquad -\infty < x < \infty.$$

$$L(\mu, \sigma^2) = \prod_{i=1}^{n} f(x_i \mid \mu, \sigma^2) = \prod_{i=1}^{n} \frac{1}{\sqrt{2\pi}\,\sigma} e^{-\frac{(x_i-\mu)^2}{2\sigma^2}}$$

$$= (2\pi)^{-n/2} (\sigma^2)^{-n/2} e^{-\frac{1}{2\sigma^2}\sum_{i=1}^{n}(x_i-\mu)^2}$$

$$l(\mu, \sigma^2) = -\frac{n}{2}\log(2\pi) - \frac{n}{2}\log(\sigma^2) - \frac{1}{2\sigma^2}\sum_{i=1}^{n}(x_i-\mu)^2$$

$$\frac{\partial}{\partial \mu} l(\mu, \sigma^2) = -\frac{1}{\sigma^2}\sum_{i=1}^{n}(x_i-\mu)(-1)$$

$$= \frac{1}{\sigma^2}\sum^{n}(x_i-\mu) = 0$$

$$\Sigma x_i - n\mu = 0$$

$n = 18$

$$\Sigma x_i = n\mu$$

$\Sigma x = 1447$

$$\boxed{\hat{\mu} = \frac{\Sigma x_i}{n}. = \bar{x}}$$

[illegible] $\hat{\mu}. = 80.3889$

$$\frac{\partial}{\partial\sigma^2}\ \ell(\mu,\sigma^2) = -\frac{n}{2}\ \frac{1}{\sigma^2} + \frac{1}{2(\sigma^2)^2}\Sigma(x_i-\mu)^2$$

$$= -\frac{n}{2\sigma^2} + \frac{\Sigma(x_i-\mu)^2}{2(\sigma^2)^2}$$

$$\frac{\Sigma(x_i-\mu)^2}{2(\sigma^2)^2} = \frac{n}{2\sigma^2}$$

$$\frac{\Sigma(x_i-\hat{\mu})^2}{n} = \hat{\sigma}^2$$

$$\boxed{\frac{\Sigma(x_i-\bar{x})^2}{n} = \hat{\sigma}^2.}$$

$n = 18$

compute $\hat{\sigma}^2 = 25.0154$

$\Sigma x = 1447$

$\Sigma x^2 = 116,773$

b.) $I(\mu) = -E\left[\frac{\partial^2}{\partial \mu^2} \log f(x|\mu,\sigma^2) \right]$.

$$f(x|\mu,\sigma^2) = \frac{1}{\sqrt{2\pi}\,\sigma} e^{-\frac{(x-\mu)^2}{2\sigma^2}}$$

$$\log f(x|\mu,\sigma^2) = -\frac{1}{2}\log(2\pi) - \frac{1}{2}\log(\sigma^2) - \frac{1}{2\sigma^2}(x-\mu)^2$$

$$\frac{\partial}{\partial \mu} \log f(x|\mu,\sigma^2) = -\frac{1}{\sigma^2}(x-\mu)(-1) = \frac{1}{\sigma^2}(x-\mu)$$

$$\frac{\partial^2}{\partial \mu^2} \log f(x|\mu,\sigma^2) = -\frac{1}{\sigma^2}.$$

$$I(\mu) = -E\left[-\frac{1}{\sigma^2}\right] = \frac{1}{\sigma^2}$$

$$AV(\mu) = \frac{1}{n I(\mu)} = \frac{1}{n\frac{1}{\sigma^2}} = \frac{\sigma^2}{n}.$$

$$\hat{AV}(\mu) = \frac{\hat{\sigma}^2}{n} = \frac{25.015\checkmark}{18} = 1.3897\checkmark$$

$$I(\sigma^2) = -E\left[\frac{\partial^2}{\partial\sigma^2}\log f(x|\mu,\sigma^2)\right]$$

$$\frac{\partial}{\partial\sigma^2}\log f(x|\mu,\sigma^2) = -\frac{1}{2\sigma^2} + \frac{1}{2(\sigma^2)^2}(x-\mu)^2$$

$$\frac{\partial^2}{\partial(\sigma^2)^2}\log f(x|\mu,\sigma^2) = +\frac{1}{2(\sigma^2)^2} - \frac{1}{(\sigma^2)^3}(x-\mu)^2$$

$$I(\sigma^2) = -E\left[\frac{1}{2(\sigma^2)^2} - \frac{1}{(\sigma^2)^3}(x-\mu)^2\right]$$

$$= -\frac{1}{2(\sigma^2)^2} + \frac{1}{(\sigma^2)^3}E\left[(X-\mu)^2\right]$$

$$= -\frac{1}{2(\sigma^2)^2} + \frac{\sigma^2}{(\sigma^2)^3}$$

~~Compute the MLE's~~

$$-\frac{1}{2(\sigma^2)^2} + \frac{1}{(\sigma^2)^2} = \frac{1}{2(\sigma^2)^2} \Big) = \frac{1}{2\sigma^4}$$

$$AV(\sigma^2) = \frac{1}{nI(\sigma^2)} = \frac{1}{n \cdot \frac{1}{2\sigma^4}} = \frac{2\sigma^4}{n}.$$

$$\hat{AV}(\hat{\sigma}^2) = \frac{2(\hat{\sigma}^2)^2}{n} = \frac{2(25.0154)^2}{15}$$

$$= 69.5300$$

c.) $CRLB(\mu) = \frac{1}{nI(\mu)} = \frac{\sigma^2}{n}.$

$$Var(\hat{\mu}) = Var(\bar{X}) = \frac{1}{n^2} \sum^n Var(X_i)$$

$$= \frac{1}{n^2} \sum^n \sigma^2 = \frac{\sigma^2}{n}.$$

$\hat{\mu}$ attains the CRLB(μ).

$\hat{\mu}$ is UMVUE

$$\hat{\mu} = \bar{X}$$

$$\frac{(\bar{X} - \mu)}{\sigma/\sqrt{n}} \sim N(0, 1)$$

$$\therefore \bar{X} \sim N\left(\mu, \frac{\sigma^2}{n}\right).$$

$$\hat{\sigma}^2 = \frac{\Sigma(X_i - \mu)^2}{n}.$$

$$\frac{n\hat{\sigma}^2}{\sigma^2} \sim \chi^2_{n-1}$$

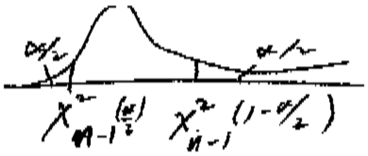
$$\therefore \hat{\sigma}^2 \sim \frac{\sigma^2}{n} \cdot \chi^2_{n-1}$$

f.) $P\left(\mu - z_{\alpha/2}\frac{\sigma}{\sqrt{n}} < \bar{X} < \mu + z_{\alpha/2}\frac{\sigma}{\sqrt{n}}\right) = 1-\alpha$

$P\left(\bar{X} - z_{\alpha/2}\frac{\sigma}{\sqrt{n}} < \mu < \bar{X} + z_{\alpha/2}\frac{\sigma}{\sqrt{n}}\right) = 1-\alpha$

$\therefore \bar{X} \pm z_{\alpha/2}\frac{\hat{\sigma}}{\sqrt{n}}$ is an approx

compute. $100(1-\alpha)\%$ CI f.o. μ.



3) $$1-\alpha = P\left(\frac{\sigma^2}{n}\chi^2_{(n-1)}(\alpha/2) < \hat{\sigma}^2 < \frac{\sigma^2}{n}\chi^2_{(n-1)}(1-\tfrac{\alpha}{2})\right)$$

$$1-\alpha = P\left(\frac{\chi^2_{(n-1)}(\frac{\alpha}{2})}{\hat{\sigma}^2 n} < \frac{1}{\sigma^2} < \frac{\chi^2_{(n-1)}(1-\frac{\alpha}{2})}{\hat{\sigma}^2 n}\right)$$

$$1-\alpha = P\left(\frac{n\hat{\sigma}^2}{\chi^2_{(n-1)}(1-\alpha/2)} < \sigma^2 < \frac{n\hat{\sigma}^2}{\chi^2_{(n-1)}(\frac{\alpha}{2})}\right)$$

$$\left[\frac{n\hat{\sigma}^2}{\chi^2_{(n-1)}(1-\frac{\alpha}{2})}, \frac{n\hat{\sigma}^2}{\chi^2_{(n-1)}(\frac{\alpha}{2})}\right]$$

compute.

4) $$\hat{\mu} \pm z_{\alpha/2}\sqrt{AV(\hat{\mu})}$$

$$\bar{x} \pm z_{\alpha/2}\frac{\hat{\sigma}}{\sqrt{n}}.$$

same compute

$$\hat{\sigma}^2 \pm z_{\alpha/2}\sqrt{AV(\hat{\sigma}^2)}$$

$$\hat{\sigma}^2 \pm z_{\alpha/2}\frac{\sqrt{2}\,\hat{\sigma}^2}{\sqrt{n}}$$

different compute.

#5 CB

Answers:

(a) ***Mary's Model.***

(i) Model: $Y_{ijk} = \mu + \alpha_i + B_j + (\alpha B)_{ij} + e_{ijk}$,

$i = 1, 2$ production lines (crossed, fixed effect), restriction: $\Sigma_i\, \alpha_i = 0$,

$j = 1, 2, 3$ batches (crossed, random effect), B_j iid $N(0, \sigma_B)$.

Ott and many other basic texts include the additional restriction $\Sigma i\, (\alpha B)_{ij} = 0$, $(\alpha B)_{ij} \sim N(0, \sigma_{\alpha B})$, but not independent on account of the restriction. Some texts (and SAS software) prefer to omit this restriction, in which case $(\alpha B)_{ij}$ are iid. As shown in Minitab printouts below, whether the additional restriction is made makes a difference in the EMSs, hence in the F-ratios. We expect to see answers using the restricted model, but will count answers based on the unrestricted model as correct.

(ii) The degrees of freedom are: DF(Line) = 2 – 1 = 1, DF(Batch) = 3 – 1 = 2, DF(Interaction) = DF(Line)DF(Batch) = (1)(2) = 1, DF(Total) = 12 –1, and DF(Error) by subtraction. In each line of the table MS = SS/DF.

According to the restricted model:

F(Interaction) = MS(Interaction)/MS(Error) = 1224 (clearly significant)

F(Line) = MS(Line)/MS(Int) = 0.06 (clearly not significant),

F(Batch) = MS(Batch)/MS(Error) = 1323 (clearly significant, but interpretation must take into account the highly significant interaction). All computations are straightforward once the SSs are known. The Minitab printout is shown below.

```
MTB > anova Strength = Line | Batch;
SUBC> random Batch;
SUBC> restrict;
SUBC> ems.
```

ANOVA: Strength versus Line, Batch

```
Factor     Type Levels Values
Line      fixed      2      1      2
Batch    random      3      1      2      3


Analysis of Variance for Strength

Source          DF          SS          MS        F       P
Line             1        2028        2028     0.06   0.835
Batch            2       78041       39021  1322.73   0.000
Line*Batch       2       72189       36095  1223.55   0.000
Error            6         177          30
Total           11      152436


Source          Variance Error Expected Mean Square for Each Term
               component term  (using restricted model)
 1 Line                     3   (4) + 2(3) + 6Q[1]
 2 Batch          9747.8    4   (4) + 4(2)
 3 Line*Batch    18032.6    4   (4) + 2(3)
 4 Error            29.5        (4)
```

The interpretation is that there is no significant difference between the two lines, but that variability involving how batches are mixed is very large compared with the beam-to-beam variability from a given batch in a given line. (Because interaction of this magnitude is rare in practice, one must wonder whether Mary has the right model.)

For completeness, we show the analysis for the unrestricted model. The difference is that the Batch effect is tested against Interaction rather than against Error, and hence is not significant. However, the interpretation remains fundamentally the same, namely that highly significant variability attaches to what batch is used at a site.

```
MTB > anova Strength = Line | Batch;
SUBC> random Batch;
SUBC> ems.
```

ANOVA: Strength versus Line, Batch

```
Factor      Type Levels Values
Line       fixed      2      1      2
Batch     random      3      1      2      3


Analysis of Variance for Strength

Source          DF          SS          MS         F       P
Line             1        2028        2028      0.06   0.835
Batch            2       78041       39021      1.08   0.481
Line*Batch       2       72189       36095   1223.55   0.000
Error            6         177          30
Total           11      152436


Source         Variance Error Expected Mean Square for Each Term
              component  term (using unrestricted model)
  1 Line                    3    (4) + 2(3) + Q[1]
  2 Batch         731.5     3    (4) + 2(3) + 4(2)
  3 Line*Batch  18032.6     4    (4) + 2(3)
  4 Error          29.5          (4)
```

Notes on scoring: Testing the Line effect against Error is wrong in either the restricted or unrestricted model and would give a seriously bogus significant Line effect. The six cell means in the table are so erratic that it should be *intuitively* obvious that a significant Line effect would make no sense. Furthermore, a key feature of mixed two-way models is that the fixed effect is never to be tested against Error. Thus, there is a nontrivial point penalty for calling the Line effect significant. (Confusion as to whether to test the Batch effect against Interaction or Error is, perhaps, excusable on a closed-book question.)

(b) ***John's Model.***

(i) Model: $Y_{ijk} = \mu + \alpha_i + B(\alpha)_{j(i)} + e_{ijk}$,

$i = 1, 2$ production lines (crossed, fixed effect), restriction: $\Sigma_i \alpha_i = 0$,

$j = 1, 2, 3$ batches (nested within lines, random effect), $B(\alpha)_{j(i)}$ iid $N(0, \sigma_{B(\alpha)})$.

Interaction is not supported by this model. First, because it makes no practical sense to talk about interaction of a factor with one nested within its levels. Second, if we tried to introduce an interaction term it would involve i and j in its subscripts, indicating confounding with the nested batch effect which involves these same subscripts.

(ii) The ANOVA table for this nested model can be obtained by adding DFs for Batch and Interaction in Mary's table (2 + 2 = 4), and by adding SSs for Batch and Interaction in Mary's table (78041 + 72189 = 150230) to obtain the DF and SS, respectively, for the nested Batch effect. Then MS and F for the nested batch effect are computed in the obvious way to obtain F = 1273, indicating a highly significant Batch effect. In practical importance, variation among batches completely swamps variation among beams within a batch. Again here, the Line effect is nowhere near significant. The Minitab printout is shown below.

```
MTB > anova Strength = Line Batch(Line);
SUBC> random Batch;
SUBC> ems.
```

ANOVA: Strength versus Line, Batch

```
Factor         Type Levels Values
Line          fixed      2      1      2
Batch(Line)  random      3      1      2      3


Analysis of Variance for Strength

   Source          DF          SS         MS        F       P
α  Line             1        2028       2028     0.05   0.828
β(α)Batch(Line)     4      150231      37558  1273.14   0.000
   Error            6         177         30
   Total           11      152436


Source          Variance Error Expected Mean Square for Each Term
               component term (using unrestricted model)
 1 Line                      2   (3) + 2(2) + Q[1]
 2 Batch(Line)   18764.1     3   (3) + 2(2)
 3 Error            29.5         (3)
```

Notes:

Data were randomly generated according to John's model with $\sigma_{B(\alpha)} = 75$ and $\sigma = 5$ (equivalently, but unrealistically in practice, according to the unrestricted version of Mary's model with $\sigma_B = 0$ and $\sigma_{\alpha B} = 75$). The respective variances are $75^2 = 5625$ (estimated here as 18764, illustrating that estimation of a variance when DF is small is risky business) and $5^2 = 25$ (estimated here as 29.5, by chance, a lot better). A difference of 20 in means between production lines was included in the simulation, but is much too small a signal to be detected against the overwhelming noise of the nested Batch variance. (While the data are fake, the part about overlooking that a design is nested based on too hurried a discussion with a client is real.)

For a coherent elementary discussion of the practical differences between restricted and unrestricted mixed models see Oehlert: *A First Course in Design and Analysis of Experiments*, Freeman, 2000.