ESD.86
Models, Data, Inference for Socio-Technical Systems
MIT, Spring 2007

Exam #2
Open Book, Open Notes, Individual Effort

1. (5 pts) A systems engineering textbook (Blanchard and Fabrycky) says “When the system operating time is the Mean Time Between Failures (MTBF), the reliability is 37%.” Show that this is true for systems with constant failure rates.

2. (10 pts) The graph below is from Weibull's 1951 paper. Estimate the mean, median, and the mode of the fatigue life. You do not need to be very precise in your estimates, but you should explain clearly your approach and why it's valid. Discuss the relationships among these the mean, median, and mode and what the relationships among them imply about the distribution of the data.

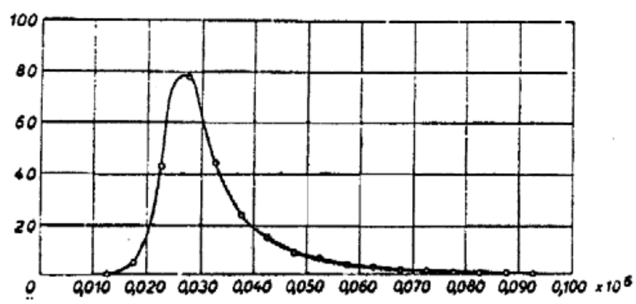

FIG. 5 FREQUENCY CURVE OF FATIGUE LIFE OF ST-37 STEEL
(Number of specimens versus number of stress cycles.)

3. Given that variables z_1, z_2, and z_3 are standardized, normally distributed, and independent, how are the following random variables distributed? (name the distribution and give the required parameters or else write the equation out in full):

a) (5 pts) $z_2 + z_3$

b) (5 pts) $z_2^{\,2} + z_3^{\,2}$

c) (5 pts) $\dfrac{z_1}{\sqrt{(z_2^{\,2} + z_3^{\,2})/k}}$

4. (10 pts) Given the following commands:
mu=[1 5 3];
sigma=[1 0 0; 0 2 0; 0 0 2];
n=15;
X = lhsnorm(mu,sigma,n)

0.1702	5.0223	2.3869
-0.6328	6.2121	3.4204
0.9821	4.662	4.0464
1.3432	4.5406	3.6269
-0.4473	3.8452	4.8308
2.1877	5.3442	3.257
1.9791	3.6864	4.2428
0.9132	5.658	1.174
1.4637	9.5435	2.8336
1.1773	4.3546	2.4748
0.4533	6.1233	-0.2493
1.6657	6.9879	3.1136
-0.0869	2.2616	1.5618
2.6525	3.3038	5.7646
0.7161	5.558	1.8655

These data are generated. The documentation says "X = lhsnorm(mu,sigma,n) generates a Latin Hypercube sample X of size n from the multivariate normal distribution with mean vector mu and covariance matrix sigma. What operations would you perform to check that the sample actually is what the documentation claims it to be? (You don't need to actually process the data, it's just provided in case you want to illustrate concretely what you'd do).

5. (5 pts) You perform a linear regression of 100 data points (n=100). There are two independent variables x_1 and x_2. The regression R^2 is 0.72. Both β_1 and β_2 pass a t test for significance. You decide to add the interaction x_1x_2 to the model. Select all the things that cannot happen:

a) R^2 decreases
c) The adjusted R^2 decreases
d) All three coefficients $\beta_{1,}$ $\beta_{2,}$ and β_{12} fail the t test for significance

6. (10 pts) Provide a persuasive and rigorous defense of your answer to question #5 above. The answer need not be primarily implemented with symbolic mathematics, but it could be.

7. The computer output below represents the results from a one way (single factor) ANOVA (α=0.05). Each row represents a different treatment condition.

a) (3 pts) What does the "F crit" value represent?

b) (3 pts) What does the "F" value represent?

c) (3 pts) What are the most critical assumptions of this analysis?

d) (3 pts) Do you think any of the assumptions were violated?

e) (3 pts) What conclusion do you draw about the effect of the treatment? Briefly justify your position.

Treatment	Replicates					
1	19	19	20	14	15	14
2	22	23	26	11	12	12
3	19	59	11	38	24	45
4	28	30	34	19	20	22

Anova: Single Factor

SUMMARY

Groups	*Count*	*Sum*	*Average*	*Variance*
1	6	101	16.83333	7.766667
2	6	106	17.66667	45.06667
3	6	196	32.66667	321.0667
4	6	153	25.5	36.7

ANOVA

Source of Variation	*SS*	*df*	*MS*	*F*	*P-value*	*F crit*
Between Groups	996.3333	3	332.1111	3.235374	0.043995	3.098393
Within Groups	2053	20	102.65			
Total	3049.333	23				

8. For the experimental plan and associated data in the Table below.

a) (5 pts) Sketch a factor effect plot.

b) (5 pts) Does this data strongly suggest that there are interactions among the factors?

Explain briefly.

	Control Factors				
Expt. No.	A	B	C	D	response
1	1	1	1	1	2.2
2	1	2	2	2	1.9
3	1	3	3	3	2.1
4	2	1	2	2	4
5	2	2	3	1	4.3
6	2	3	1	2	3.8
7	3	1	3	3	5.9
8	3	2	1	3	6.4
9	3	3	2	1	5.9

9. You run a 2^7 full factorial design. The response is the velocity of a paper helicopter when it hits the ground when dropped from 9 feet.   The standard deviation of the experimental error is estimated to be 0.3 seconds since it's known the greatest contribution to error is the response time in hitting the stopwatch.

a) (5 points) What is the standard deviation of a main effect estimate?

b) (5 points) Suggest three ways to reduce the uncertainty in the main effect estimate.

10. (10 points) You run a 2^7 full factorial design. You have the same situation as problem #9 above, except now it's also known that one of the helicopters was not set to the correct factor levels. Instead, it was set to the levels for some other helicopter in the matrix experiment. Can you still characterize the uncertainty in your main effect estimate?
If "yes", briefly describe a procedure would you suggest to estimate the uncertainty.
If "no", explain why not and describe what you'll do next.