

Exam #2 (Practice Exam!!!!!!!)
Open Book, Open Notes, Individual Effort

NOTE: This practice exam is intentionally too long. I just kept adding questions so you can practice more. The real text will fit into the 1.5 hour slot. The points assigned to each question are about what I would assign in the real test, but the total test adds up to more than 100 points. Understand?

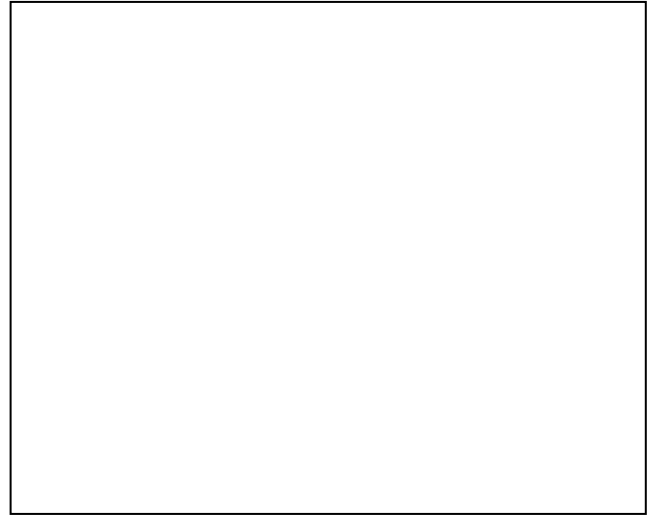
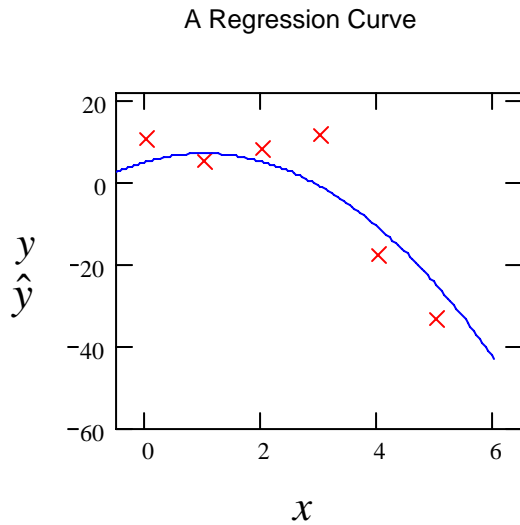
1. (5 pts) You are seeking to calibrate a thermometer. You wish to determine the regression line relating the output of the device (in Volts) to the input to be measured (temperature in degrees Celsius). What are the units of the slope of the regression line? What are the units of the sum of squares of the error? What are the units of R^2 ?

2. (5 pts) Find the mean, median, and mode of the sample 34, 29, 26, 37, 26.

3. (5 points) Provide an example of an engineering system and a measurable parameter thereof for which the sample median is a more appropriate way to communicate the central tendency than the sample average. Explain your reasoning briefly.

- 4) (2 pts) In order to resolve only the main effects in a system with 7 control factors each having 2 levels (with no replicates), how many experiments would one would need to conduct? You may assume that two-factor and higher order interactions are negligibly small.

5. (5 pts) The graph below resulted from a polynomial regression of six data points. Sketch a plot of the residuals versus the independent variable x .



You may make your sketch here. Please label your axes appropriately.

6. (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:
- R^2 decreases
 - The adjusted R^2 decreases
 - All three coefficients β_1 , β_2 , and β_{12} fail the t test for significance

7. (15 pts) Provide a persuasive and intellectually rigorous defense of your answer to question #5 above. The answer need not be primarily implemented with symbolic mathematics, but it could be.

8. Weibull describes a graph below as an example of how Weibull plots may reveal that two sample distributions have been mixed.

a) (5 pts) Briefly describe the procedure to make such a plot.

b) (5 pts) Under what conditions will you get two separate linear regions as depicted here? Under what conditions will you not get two different regions and therefore fail to detect there are two populations mixed?

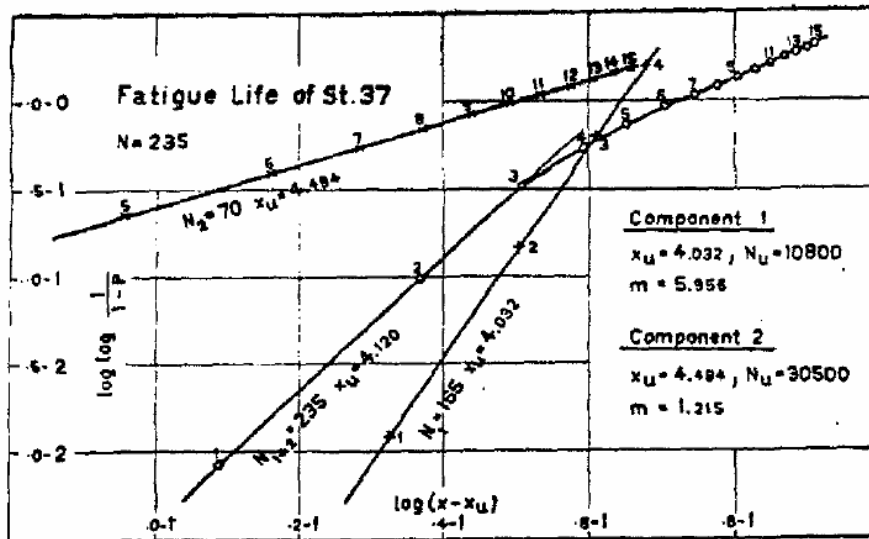


FIG. 6 FATIGUE LIFE OF ST-37 STEEL

Copyright © 1951 by ASME. Used with permission.

9. (15 pts) A key outcome of this class has been to show that beer has deeply influenced the development of statistics (and human affairs more generally). Below are some excerpts from a paper that illustrates the continuing influence. Select an idea expressed here and critique it. Suggest a better approach to either the analysis or the communication of the results.

From Koksalan, M., N. Erkip and H. Moskowitz. "Explaining beer demand: A residual modeling regression approach using statistical process control." *Int. J. Production Economics* 58 (1999): 265-276.

The regression model was run for a three-year period from 1989 to 1991, since the sales figures for the cities were most reliable for this period. There were 42 data points (cities) for each year, yielding a total of 126 data points. None of the pairwise correlation estimates between independent variables were very high, eliminating concerns of multicollinearity.

We ran the regression model by the backward elimination of insignificant independent variables.

The value of the adjusted R^2 turned out to be 0.60 indicating that about 60% of the variation in the dependent variable can be explained. We present a scatterplot of standardized residuals vs. predicted sales in Fig. 1, which shows an undesirable trend indicating that the variance may not be constant (the horizontal lines at \$1 standard deviation points and the highlighted data points are referred to in the next section). Our various efforts of transforming the data did not lead to an improvement. Thinking that increasing variance could be caused by missing variables, we developed and implemented an SPC-based approach discussed below to diagnose the existence of such a possibility.

10. In the spreadsheet below:

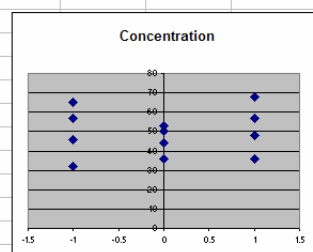
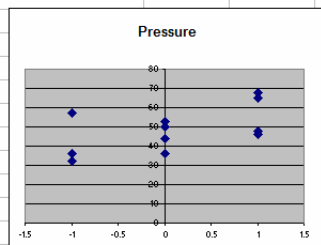
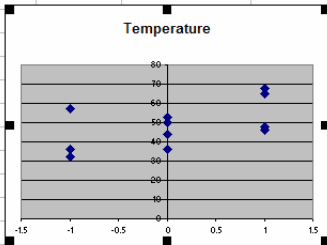
a) (5 pts) Why are there 4 DOF for the regression?

b) (5 pts) The standard error is listed as 5.978... Name something you can do to inspect the graphs and check that this is not substantially off.

| Regression Statistics | | | | | | | |
|-----------------------|----------|--|--|--|--|--|--|
| Multiple R | 0.907831 | | | | | | |
| R Square | 0.824157 | | | | | | |
| Adjusted R Square | 0.723675 | | | | | | |
| Standard Error | 5.978135 | | | | | | |
| Observations | 12 | | | | | | |

| ANOVA | | | | | |
|------------|----|-------------|---------|----------|----------------|
| | df | SS | MS | F | Significance F |
| Regression | 4 | 1172.5 | 293.125 | 8.202032 | 0.008857 |
| Residual | 7 | 250.1666667 | 35.7381 | | |
| Total | 11 | 1422.666667 | | | |

| | Coefficients | Standard Error | t Stat | P-value | Lower 95% | Upper 95% | Lower 95.0% | Upper 95.0% |
|--------------|--------------|----------------|----------|----------|-----------|-----------|-------------|-------------|
| Intercept | 49.33333 | 1.725738857 | 28.58679 | 1.65E-08 | 45.25261 | 53.41405 | 45.25261 | 53.41405 |
| X Variable 1 | 5.625 | 2.113589815 | 2.661349 | 0.032404 | 0.627158 | 10.62284 | 0.627158 | 10.62284 |
| X Variable 2 | 10.625 | 2.113589815 | 5.026992 | 0.001518 | 5.627158 | 15.62284 | 5.627158 | 15.62284 |
| X Variable 3 | 1.125 | 2.113589815 | 0.53227 | 0.611009 | -3.87284 | 6.122842 | -3.87284 | 6.122842 |
| X Variable 4 | -0.875 | 2.113589815 | -0.41399 | 0.691273 | -5.87284 | 4.122842 | -5.87284 | 4.122842 |



11. The passage below is from a poem by Oliver Wendell Holmes:

HAVE you heard of the wonderful one-hoss-shay, that was built in such a logical way it ran a hundred years to a day...

Now in building of chaises, I tell you what, there is always somewhere a weakest spot,-- in hub, tire, fellow, in spring or thill, in panel, or crossbar, or floor, or sill...

But the Deacon swore .. he would build one shay to beat the taown 'n' the keounty 'n' all the kentry raoun'; It should be so built that it couldn' break daown! --"Fur," said the Deacon, "t 's mighty plain that the weakes' place mus' stan' the strain; 'n' the way t' fix it, uz I maintain, is only jest t' make that place uz strong uz the rest"...

So the Deacon inquired of the village folk where he could find the strongest oak, That could n't be split nor bent nor broke,--

Eighteen hundred and twenty came;-- Running as usual; much the same. Thirty and forty at last arrive, And then come fifty, and fifty-five...

There are traces of age in the one-hoss-shay-- A general flavor of mild decay, But nothing local, as one may say. There couldn't be,--for the Deacon's art had made it so like in every part that there wasn't a chance for one to start...

And yet, as a whole, it is past a doubt in another hour it will be worn out!

First a shiver, and then a thrill, Then something decidedly like a spill,-- ...What do you think the parson found, when he got up and stared around? The poor old chaise in a heap or mound, as if it had been to the mill and ground! You see, of course, if you're not a dunce, how it went to pieces all at once,-- all at once, and nothing first,-- just as bubbles do when they burst.

End of the wonderful one-hoss-shay. Logic is logic. That's all I say.

a) (5 pts) What would the Weibull distributions of the individual components of the one hoss shay have to look like for the events in the poem to be somewhat likely.

b) (10 pts) Discuss the merits and demerits of actually designing engineering systems in the way the Deacon's Masterpiece is arranged.

12. (10 pts) I manufacture two rectangles of material with a known density and thickness. The sides of the two rectangles are normally distributed with means 1 and 2 meters for each rectangle and standard deviations of 1% of the nominal. What is the distribution of the ratio of the weights of the two sheets?