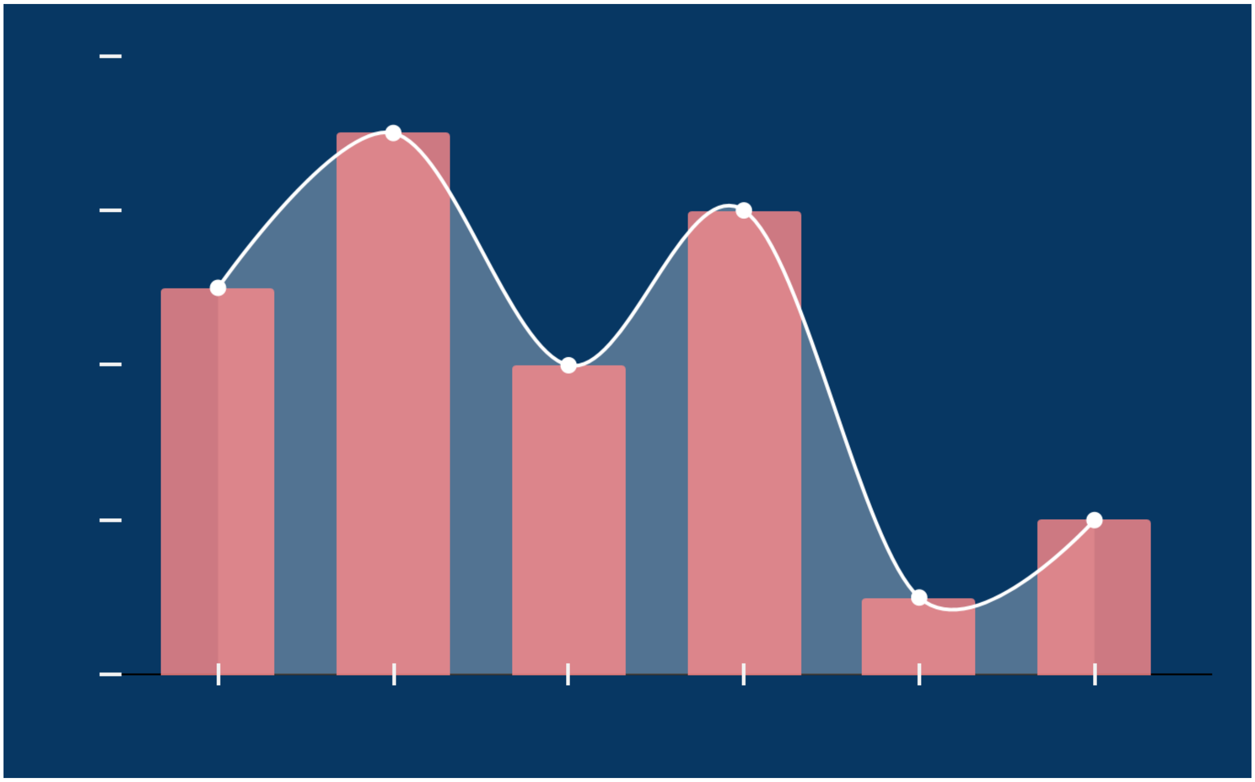
Post-Conflict Environmental Assessment - Section III (HW7)

Hazmat Suit Stress Analysis



Kya Allen

03.15.2021

301-919-7902

Arendelle Department of Environmental Safety

INTRODUCTION

In the Aftermath of the Arendellian-Atohallan Conflict, there have been reports of leftover radioactive hot spots across the border that separated Arendell and the formerly-Atohallan territory of the continent. The Radioactivity of these sites pose a serious risk to the health conducting environmental analysis in these zones. In this study we analyze stress test data for how much radioactivity our experimental Hazmat Suits can take at once without compromising the user. From the literature on stress tests of medical devices, it's clear that a log-normal distribution is common. For this study we'll be checking to see if our data is lognormally distributed, and we'll use log transformations to determine the Mean and Variance.

Methodology

To determine if our data is Lognormally distributed, we'll simply take the natural logarithm of our data set. Then from there we would expect the transformed data to be normally distributed, so to check if the transformed data is normally distributed, we regress the transformed-observed-quantiles on the Normal-theoretical quantiles and test for any significant nonlinearity (which would indicate non-normality of the transformed data)

To Determine the Mean and Variance, we'll use a set of equations to compute them from the Mean and Variance of the transformed data.

Analysis procedure can be seen in the python notebook attached to this submission

Analysis I - Testing for Log-Normality

So first we take the data vector x and take the natural log of x.

$$ln(x)$$

Then With that we order the vector ln(x) for the Y-axis of our Quantile-Quantile Plot. Then to obtain the Normal Quantiles, we use the presented Normal order statistics from the paper presented in class. Then we regress, with Ordinary Least Squares

$$H_0\!: \widehat{Y} = \beta_0 + \beta_1 X + \epsilon$$

$$H_1\!: \widehat{Y} = \beta_0 + \beta_1 X + \beta_2 X^2 + \epsilon$$

To test this Hypothesis by performing a quadratic regression (the second equation) on the data with ordinary least squares. This can be done with calculus by taking the partial derivatives of the loss function with respect to beta-1 and beta-2, setting the system of equations to zero, and then solving for the Beta's. The loss function is the Sum of Squared Errors (SSE), and this method minimizes this quantity.

$$SSE = \sum_{i=0}^{n} (Y_i - \hat{Y}_i)^2$$

Alternatively, the Beta's can be found algebraically with Matrix Algebra

$$\beta = (X^T X)^{-1} X^T Y$$

Where each term here is a Vector or Matrix of the data it represents. For the purpose of this Analysis, We computed the regression using the OLS function in the Statsmodels api for python. The package automatically computes the following T-Test on each predictor variable.

$$T = \frac{\hat{\beta}_i - 0}{SE(\hat{\beta}_i)}$$

Once again, the results of this will be reported in the Results Section

Analysis II - Mean and Variance

In class we were presented with a set of equations for finding the Mean and Variance of the data from the mean and variance of the transformed data. They are called Moment Generating Functions (MGF) of the population parameters. Here they are:

$$Moment\ of\ a\ Lognormal:\ MGF of\ Y = E(e^{ty}) = e^{t\mu - \frac{t^2\sigma^2}{2}}$$

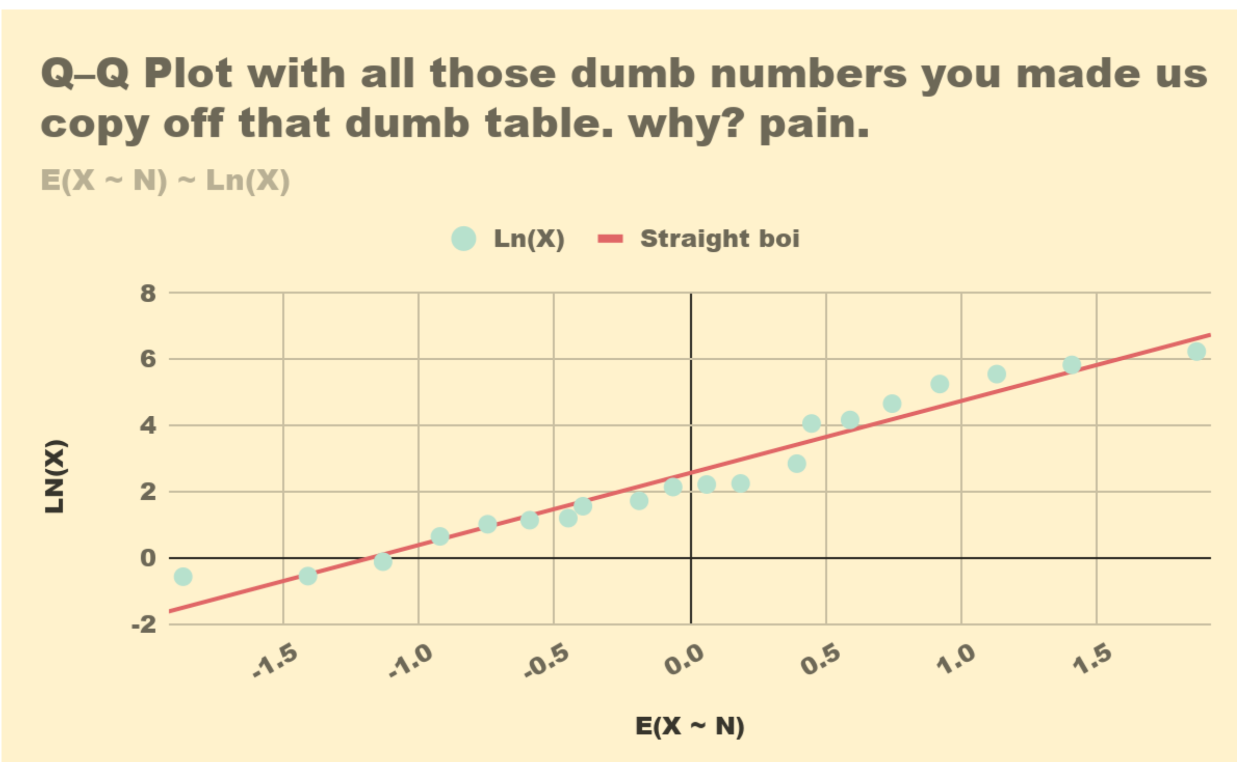
$$\hat{\eta} = e^{\bar{\mu} + \frac{\hat{\sigma}^2}{2}}$$

$$\widehat{\delta^2} = e^{2\bar{\mu}+\sigma^2}[e^{\sigma^2}-1]$$

So I just wrote the equations up in Python to get the Mean and Variance. I'll post the results.

RESULTS

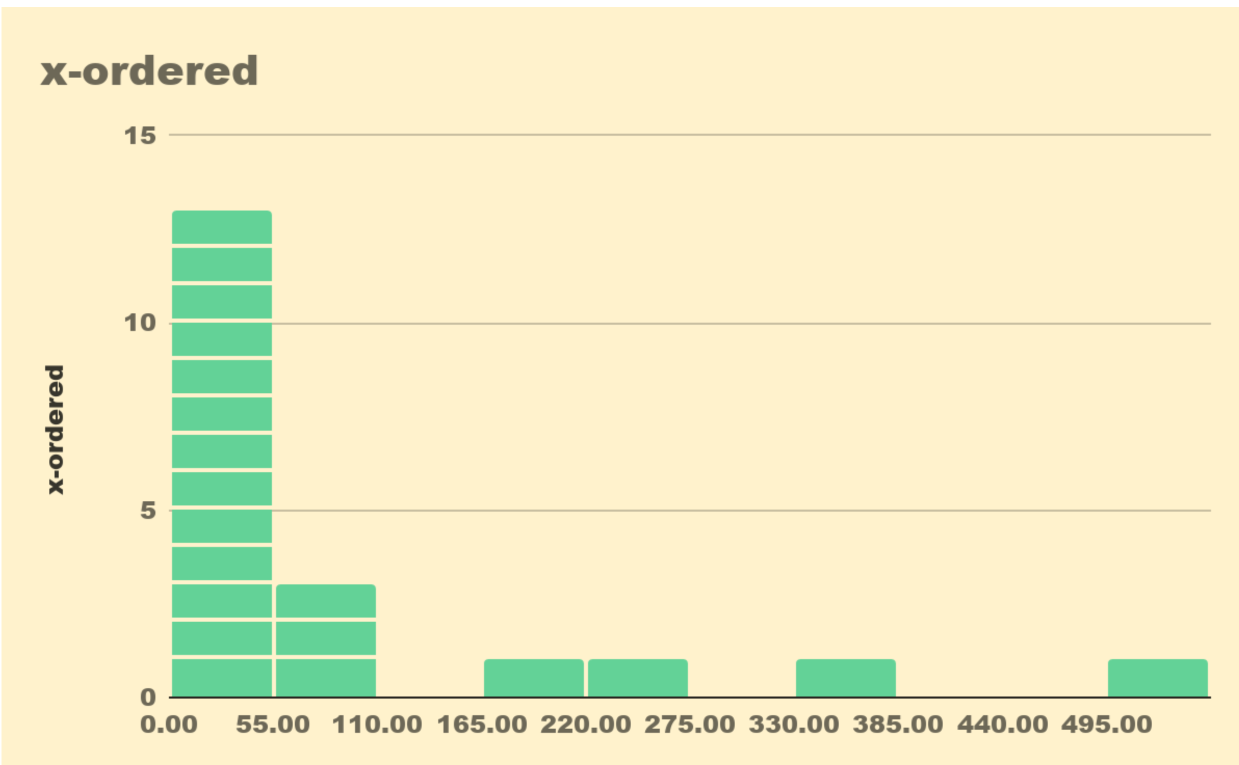
Table 1				
Regression Analysis Summary for Hazmat Max Stress data ~ Theoretical Normal Quantile				
Predictor Variable	Coefficient	Standard Error	T-Statistic	P-value
Intercept	2.4289**	0.132	18.351	< 0.001
Normal Quantile	2.1767**	0.107	20.327	< 0.001
Normal Quantile ^ 2	0.1578	0.096	1.64	0.119

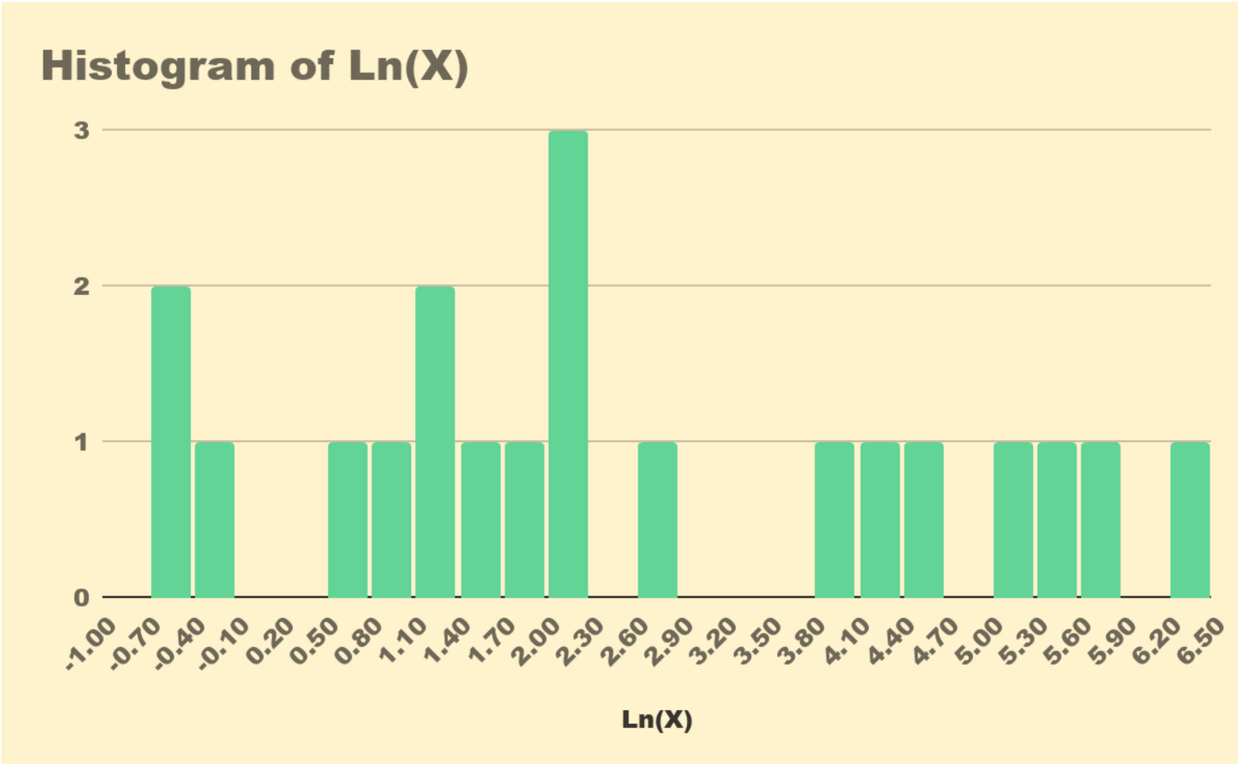


The T-test for the quadratic error was not significant at the 0.05 alpha level. We fail to reject the null hypothesis that the transformed data quantiles have a linear relationship with the Normal quantiles, and thus we reject the hypothesis that the transformed data is non-normal and thus we reject the hypothesis that the raw data is anything other than lognormal.

Table 2		
Parameters/moments		
	Ln(x)	x
Mean	2.569	38.367
Variance	4.644	11239.53

There they are. More graphs now





CONCLUSION

Our results on the test for distributional assumptions tells us that the data on our hazmat suits Time-To-Failure is indeed Lognormal like we expected based on existing research with stress testing medical devices. Our estimations indicate that the "average" suit can withstand about 38 [units] of radioactivity before failure. However, the data has a high variance, with a significant skew towards the right. So many of our hazmat suits could perform above expectation, but at the same time, most of our suits are performing very poorly, some failing after less than 1 [unit] of radioactivity.