Sample Problems
Lecture 6: Linear Regression

Problems prepared by: Prof. F. Jalayer and Dr. H. Ebrahimian

1. The Cloud Method (linear logarithmic regression)

Suppose that have calculated the normalized roof displacement demand (D) for a structure subjected to a set of n = 9 ground motion records:

D = {0.001, 0.002, 0.003, 0.025, 0.01, 0.009, 0.0065, 0.015, 0.009}

Moreover, suppose that the peak ground acceleration values corresponding to each of the records in question are equal to (units g): PGA = {0.09, 0.16, 0.22, 0.80, 0.55, 0.27, 0.20, 0.40, 0.33}

- **Calculate the sample mean and standard deviation of roof displacement**

$$\mu_{\ln D} = \frac{\sum_{i=1}^{9} \ln D_i}{9} = -5.098$$

$$\sigma_{\ln D} = \sqrt{\frac{\sum_{i=1}^{9}\left(\ln D_i - \mu_{\ln X}\right)^2}{9-1}} = 1.0227$$

- **Calculate the mean and standard deviation by ordered statistics of roof displacement**

$$\mu_{\ln X} = \ln\left(x_{50}\right) = -4.7105 = \ln\left(\eta_X\right) = \ln\left(0.009\right)$$

$$\sigma_{\ln X} = \frac{\ln x_{84} - \ln x_{16}}{2} = \frac{1}{2}\ln\left(\frac{x_{84}}{x_{16}}\right) = 1.0075$$

- **We have done a linear regression of the form: ln(*D*) = ln(*a*) + *b* ln(*PGA*) and have calculated the constants *a* and *b* via linear least squares method to be equal to *a* = 0.0385 and *b* = 1.4377. Calculate the standard error of regression.**

$$\sigma_{\ln D|PGA} = \beta_{D|PGA} = \sqrt{\frac{\sum_{i=1}^{9}\left(\ln D_i - \ln a - b \cdot \ln PGA_i\right)^2}{9-2}} = \sqrt{\frac{\sum_{i=1}^{n}\left(\ln \frac{D_i}{a \cdot PGA_i^{b}}\right)^2}{n-2}} = 0.4074$$

- **Calculate the correlation between two data set**

$$\sigma_{\ln D,\ln PGA} = \frac{\sum_{i=1}^{n}\left(\ln D_i - \ln\eta_D\right)\left(\ln PGA_i - \ln\eta_{PGA}\right)}{n-1}$$

$$\rho_{\ln D,\ln PGA} = \frac{\sigma_{\ln D,\ln PGA}}{\sigma_{\ln D}\cdot\sigma_{\ln PGA}} = \frac{\sum_{i=1}^{n}\left(\ln D_i - \ln\eta_D\right)\left(\ln PGA_i - \ln\eta_{PGA}\right)}{\sqrt{\sum_{i=1}^{n}\left(\ln D_i - \ln\eta_D\right)^2 \sum_{i=1}^{n}\left(\ln PGA_i - \ln\eta_{PGA}\right)^2}} = 0.928$$

2. Building a landslide trigger predictive model

A predictive model for rainfall-induced landslide trigger (*innesco*) on a susceptible slope can be created based on information available about previous rainfall events that have led to triggering of a landslide. In particular, a simplified predictive model can be built based on information about cumulative rainfall intensity (in mm/hour) H and rainfall duration (in hours) D. Imagine that the following small database of previous critical rainfall scenarios (that have led to a landslide trigger) is available:
$H = \{200, 0.6, 5, 40, 10\}$
$D = \{0.5, 2000, 30, 4, 100\}$
Let us build a simple predictive model for landslide triggering by employing the linear regression in the logarithmic space in the form of:

$$\ln H = \ln a + b \cdot \ln D$$

1- Calculate a and b and the standard error of regression $\sigma_{\ln H|D}$ (i.e., the standard deviation of $\ln H$ given D).

See the lecture notes for the expressions of a and b:

$$x_i = \ln D_i,\ \ y_i = \ln H_i$$

$$b = \frac{\sum_{i=1}^{n} x_i y_i - n\bar{x}\,\bar{y}}{\sum_{i=1}^{n} x_i^2 - n\,\bar{x}^2} = -0.6719$$

$$\ln a = \bar{y} - b\,\bar{x} \Rightarrow a = 106.5136$$

$$\sigma_{\ln H|D} = \sqrt{\frac{\sum_{i=1}^{5}\left(\ln H_i - \ln a - b \cdot \ln D_i\right)^2}{5-2}} = \sqrt{\frac{\sum_{i=1}^{5}\left(\ln \dfrac{H_i}{a \cdot D_i^{\,b}}\right)^2}{5-2}} = \sqrt{\frac{\mathbb{RSS}}{5-2}} = 0.6228$$

$$\mathbb{RSS} = 1.1637$$

The deviator n-2 (the degrees of freedom) is used rather than n or n-1. Since the n data points are already used in order to estimate the two regression coefficients, i.e. the slope and the intercept of the regression line, n–2 degree of freedom remains.

2- Calculate the coefficient of determination R^2.

$$R^2 = 1 - \frac{\mathbb{RSS}}{SS_{yy}} = 1 - \frac{1.1637}{19.1387} = 0.9392$$

$$SS_{yy} = \sum_{i=1}^{5}\left(\ln H_i - \overline{\ln H}\right)^2 = 19.1387$$

3- What is the 98% confidence interval on the slope of the regression b? (hint: $\sigma_b = \frac{\sigma_{\ln H|D}}{\sqrt{SS_{xx}}}$ where $\sigma_{\ln H|D}$ is the standard error of regression and $SS_{xx} = \sum_i (X_i - \bar{X})^2$, where X is the independent variable of the linear regression).

$$SS_{xx} = \sum_{i=1}^{5}\left(\ln D_i - \overline{\ln D}\right)^2 = 39.8110$$

$$\sigma_b = \frac{\sigma_{\ln H|D}}{\sqrt{SS_{xx}}} = \frac{0.6228}{\sqrt{39.8110}} = 0.0987$$

$$\hat{b} - \Phi^{-1}\left(\frac{1+0.98}{2}\right) \times \frac{\sigma_{\ln H|D}}{\sqrt{S_{xx}}} \le \tilde{b} \le \hat{b} + \Phi^{-1}\left(\frac{1+0.98}{2}\right) \times \frac{\sigma_{\ln H|D}}{\sqrt{S_{xx}}}$$

$$\Rightarrow -0.6719 - 2.3263 \times 0.0987 \le \tilde{b} \le -0.6719 + 2.3263 \times 0.0987$$

$$\Rightarrow -0.9016 \le \tilde{b} \le -0.4423$$

4- If we do a test of hypothesis on whether the slope of the regression is statistically significant with confidence 98%, would such hypothesis be accepted (in other words, is b=0 out of the 98% confidence interval on the slope of the regression?)

Since the slope b=0 is out of the 98% confidence interval, the hypothesis will be accepted.

Landslide trigger probability:

Now we use the above regression model in order to make landslide trigger predictions. Figure 1 below demonstrates the probabilistic model that is created based on the regression results for d=48h. **In such model, the landslide trigger probability is calculated based on the percentage of critical rainfall events (for a given duration) that have had a rainfall intensity less than the intensity of the rainfall event in question** (for example, the regression prediction indicates 50% probability of trigger):

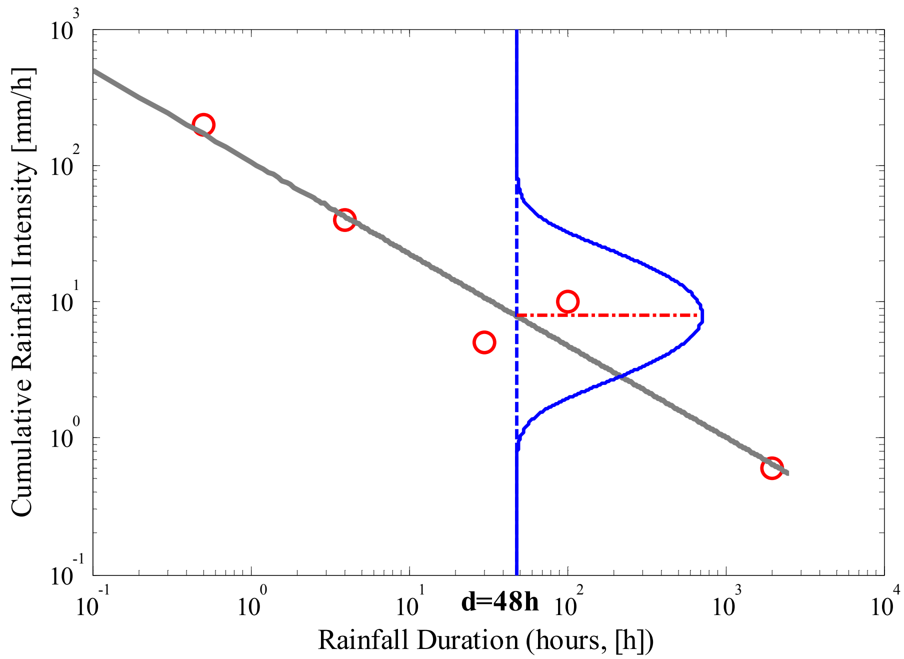


Figure 1- Regression probability model for landslide trigger d=48h

1- What is the rainfall intensity that corresponds to a rainfall trigger probability of 50% for d=48h?

$$\mu_{\ln H|D=d} = \ln \eta_{H|D=d} = \ln a + b \ln D \Rightarrow \eta_{H|D=48} = a \cdot 48^b = 7.9014$$

2- What is the rainfall intensity that corresponds to a rainfall trigger probability of 95% for d=48h? (hint: assume that the conditional distribution of ln$H$ given D is Normal).

Denoting by h^P the rainfall intensity that corresponds to P=0.95 probability of not being exceeded:

$$u = \Phi^{-1}\left(P = 0.95\right) = \frac{\ln h^P - \mu_{\ln H|D=d}}{\sigma_{\ln H|D}} = \frac{\ln h^P - \ln \eta_{H|D=48}}{\sigma_{\ln H|D}}$$

$$\Rightarrow h^P = \eta_{H|D=48} \times \exp\left(\sigma_{\ln H|D} \Phi^{-1}(0.95)\right) = 22.0092 \text{ mm}$$

$$\Phi^{-1}\left(0.95\right) = 1.6449$$

$$\Phi^{-1}\left(0.84\right) = 1,\ \Phi^{-1}\left(0.16\right) = -1$$

3- What is the rainfall trigger probability for a rainfall event with d=24h and h=15mm/hour? (hint: use Normal distribution and see Figure 2).

$$\eta_{H|D=24} = a \cdot 24^b = 12.5887$$

$$u = \frac{\ln h - \mu_{\ln H|D=24}}{\sigma_{\ln H|D}} = \frac{\ln 15 - \ln \eta_{H|D=24}}{\sigma_{\ln H|D}} = 0.2814$$

$$\Rightarrow P = \Phi(0.2814) = 0.6108$$

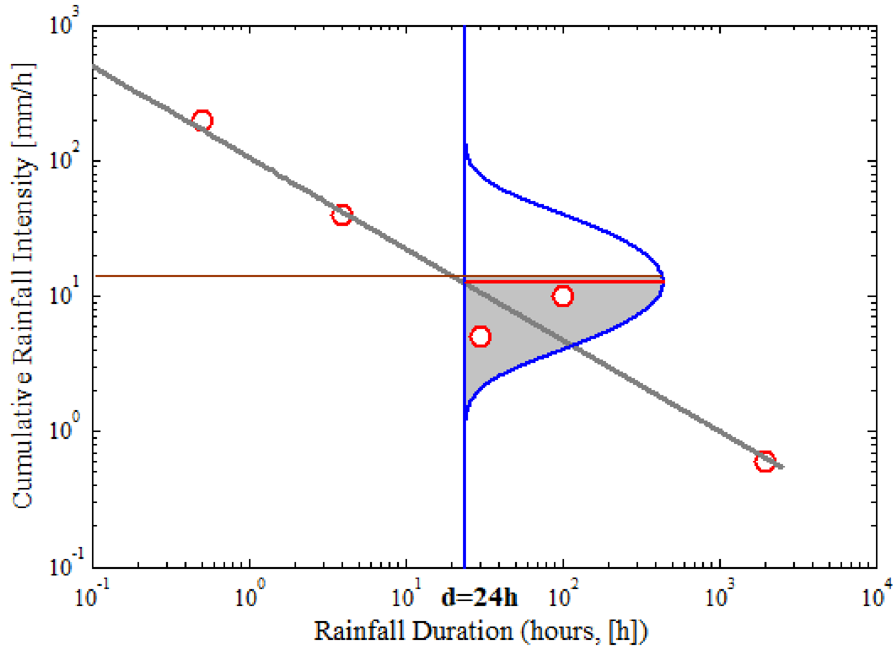


Figure 2- Regression probability model for landslide trigger d=24h

Landslide return period

The return period for various rainfall events is usually calculated from IDF curves known also as the rainfall intensity/duration/frequency curves. Let us assume that the below equation shows the IDF curves for a certain location:

$$h(d, T_R) = K_{T_R} \cdot (18.34\, d^{0.25})$$

where $h(d, T_R)$ is the rainfall intensity (in mm/h) for a given duration d and return period T_R. The term K_{T_R} is known as the *growing factor* for return period T_R: K_2=0.92, K_{10}=1.42, K_{30}=1.70, K_{50}=1.83, K_{100}=2.01, K_{300}=2.29.

1- What is the mean annual rate of occurrence for a rainfall event with d=10h and h=30mm/h?

$$K_{T_R} = \frac{30}{18.34 \times 10^{0.25}} \cong 0.92 \Rightarrow T_R = 2 \text{ years}$$

$$\Rightarrow \lambda_r = \frac{1}{T_R} = \frac{1}{2} = 0.5/\text{years}$$

2- Assuming that the triggered landslide can be modeled with a filtered Poisson distribution, what is the mean annual rate of occurrence for the triggered landslide with *d*=10h and *h*=30mm/h?

$$\lambda_{ls} = \lambda_r \times P\left[H \le h \middle| D = d\right]$$

$$P\left[H \le 30 \middle| D = 10\right] = \Phi\left(\frac{\ln 30 - \log\left(a \cdot 10^b\right)}{\sigma_{\ln H|D}}\right) = \Phi\left(0.4498\right) = 0.6736$$

$$\Rightarrow \lambda_{ls} = 0.50 \times 0.6736 = 0.3368$$

3- What is the probability that at least one landslide event is triggered in one year from a rainfall event with *d*=10h and *h*=30mm/h?

$$P\left(r \ge 1 \middle| \lambda_{ls}, t = 1\right) = 1 - e^{-0.3368 \times 1} = 0.2859$$

Normal probability information needed

$\Phi^{-1}(0.99) = 2.3263$, $\Phi^{-1}(0.95) = 1.6449$, $\Phi(0.28) = 0.6103$, $\Phi(0.45) = 0.6736$

3. Estimating material mechanical properties for based on in-situ tests for an existing building

For an existing reinforced concrete building, a series of destructive and non-destructive tests have been carried out to determine the concrete compressive resistance.

a) The drilled core (destructive) tests of diameter are reported in Table 1. Assuming that concrete quality is homogenous over the entire building, what is the median and the standard deviation (of the logarithm) of the concrete core resistance?

$$\mu_{\ln f_{DT}} \cong \overline{\ln f_{DT}} = \frac{\sum_{i=1}^{15} \ln f_{core,i}}{15} = 2.6137 \;\therefore\; \ln\left(\eta_{f_{DT}}\right) = \mu_{\ln f_{DT}} \Rightarrow \eta_{f_{DT}} = e^{\mu_{\ln f_{DT}}} = 13.6491$$

$$\sigma_{\ln f_{DT}} \cong \sqrt{\frac{\sum_{i=1}^{15}\left(\ln f_{core,i} - \overline{\ln f_{DT}}\right)^2}{15-1}} = 0.3595$$

b) What is the estimated error for the (logarithmic) average core resistance of concrete? (hint: use the results of **Problem 1** applied to the average of the logarithm of core resistance data)

$$\sigma\left(\overline{\ln f_{DT}}\,\middle|\,\mathbf{I}\right) = \frac{\sigma_{\ln f_{DT}}}{\sqrt{15}} = 0.0928$$

Table 1: Drilled core (destructive) tests

No.	**Location**	f_{core} **(MPa)**	**symbol**
1	beam- third floor	11.45	c1
2	beam- third floor	23.16	c6
1	column-third floor	13.85	c2
2	column third floor	12.61	c5
1	beam-second floor	16.85	c3
2	beam-second floor	19.52	c8
1	column-second floor	9.65	c4
2	column-second floor	6.65	c7
3	column-second floor	11.93	c9
1	beam-first floor	20.8	c10
2	beam-first floor	11.18	c14
3	beam-first floor	23.05	c15
1	column-first floor	15.72	c11
2	column-first floor	9.88	c12
3	column-first floor	10.95	c13

c) Now assume that we have the following ultrasonic (non-destructive) test results available (reported in Table 2). As it can be seen from the table, quite a few of the ultrasonic tests results are executed at the same position where the core tests were carried out. These core tests are going to be used for calibrating the ultrasonic tests. Calculate the ultrasonic resistance by calibrating the velocity results to the drilled core tests. Hint. Find the ultrasonic tests for which the drilled core test results are available in Table 2 and then fit a regression in the form:

$$\ln f_{core} = \ln a + b \ln V \tag{3}$$

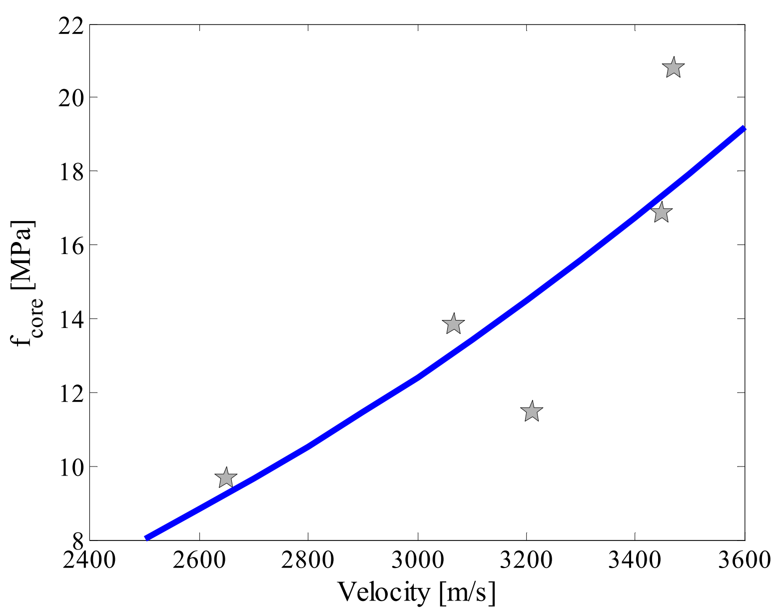
What are the coefficients a and b? What is the standard error of regression?

See the Lecture Note 6 for the expressions of a and b:

$$\ln a = -16.5911 \Rightarrow a = 6.2311\times 10^{-8}$$

$$b = 2.3868$$

$$\sigma_{\ln f_{UT}|V} = \sqrt{\frac{\sum_{i=1}^{5}\left(\ln(f_{core,i}) - \ln a - b\ln v_i\right)^2}{5-2}} = 0.1767$$



d) Given the error of measurement in the ultrasonic tests, we use a more conservative estimate of resistance compared to the mean estimate. Use the 16th percentile of the fitted relation using regression (hint: mean minus one standard deviation in the logarithmic space) to calculate the cylindrical resistance for all the other ultrasonic tests based on the velocity values reported in Table 2.

$$\ln f_{core}^{\;16th} = \ln a + b\ln V - \sigma_{\ln f_{UT}|V} = \ln\left(a\cdot V^b\right) - \sigma_{\ln f_{UT}|V}$$

$$\Rightarrow f_{core}^{\;16th} = \left(a\cdot V^b\right)\cdot \exp\left(-\sigma_{\ln f_{UT}|V}\right) = \eta_{f_{UT}|V}\cdot \exp\left(-\sigma_{\ln f_{UT}|V}\right)$$

The value are added to the last column of Table 2.

Table 2: Non-destructive tests

No.	**Location**	**Velocity V(m/s)**	**symbol**	**f_{core} (MPa)**	**f_{core}^{16th} (MPa)**
1	column- third floor	2733.6			8.3378
2	column- third floor	3272.1			12.8068
3	column- third floor	3275			12.8339
4	column- third floor	2760			8.5312
5	column- third floor	3025.9			10.6257
6	column- third floor	3067.5	c2	13.85	10.9777
1	beams-third floor	3212	c1	11.45	12.2525
2	beams-third floor	3272			12.8059
1	column- second floor	3378.4			13.8223
2	column-second floor	2613.2			7.48791
3	column-second floor	2883.4			9.4701
4	column-second floor	2650	c4	9.65	7.7421
5	column-second floor	2822.2			8.9974
1	beams -second floor	3448.3	c3	16.85	14.5187
2	beams-second floor	3385.7			13.8937
3	beams-second floor	3359.2			13.6356
1	columns-first floor	3566.9			15.7348
2	columns-first floor	3441.5			14.4465
3	columns-first floor	3359.8			13.6414
4	columns-first floor	3274.6			12.8302
5	columns-first floor	3258.8			12.6829
6	columns-first floor	3808.5			18.3991
1	beams-first floor	3470.3	c10	20.8	14.7368
2	beams-first floor	3508.8			15.1300

e) What is the median and standard deviation of the logarithm for the concrete resistance considering only the ultrasonic tests (based on the 16th percentiles of resistance calculated in the previous part)?

$$\mu_{\ln f_{UT}} \cong \overline{\ln f_{UT}} = \frac{\sum_{i=1}^{24} \ln f_{core,i}^{16th}}{24} = 2.4868 \;\therefore\; \ln\left(\eta_{f_{UT}}\right) = \mu_{\ln f_{UT}} \Rightarrow \eta_{f_{UT}} = e^{\mu_{\ln f_{UT}}} = 12.0226$$

$$\sigma_{\ln f_{UT}} \cong \sqrt{\frac{\sum_{i=1}^{24}\left(\ln f_{core,i}^{16th} - \overline{\ln f_{UT}}\right)^2}{24-1}} = 0.2416$$