# MTP Report

*SPND Error Diagnosis and Correction*

IIT Bombay

Autumn -2016

# Notes

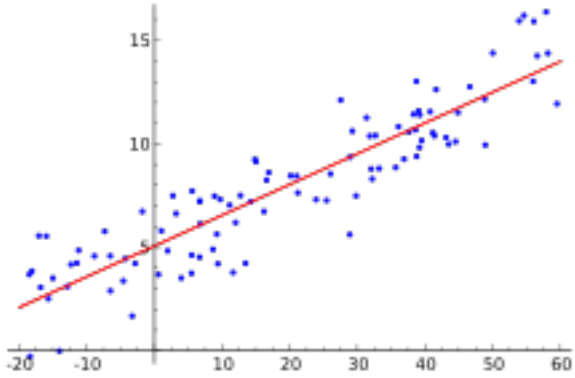
## PCA

**What is Variance?**

variance is the is the second moment of a frequency distribution taken about the arithmetic mean as the origin*(Dictionary of Statistical Terms - Sir Maurice G. Kendal)

**Regression analysis?(wiki)**

In statistical modeling, **regression analysis** is a statistical process for estimating the relationships among variables. It includes many techniques for modelling and analysing several variables, when the focus is on the relationship between a dependent variable and one or more independent variables (or 'predictors'). More specifically, regression analysis helps one understand how the typical value of the dependent variable (or 'criterion variable') changes when any one of the independent variables is varied, while the other independent variables are held fixed.




**Covariance?(wiki)**

In probability theory and statistics, covariance is a measure of how much two random variables change together. If the greater values of one variable mainly correspond with the greater values of the other variable, and the same holds for the lesser values, i.e., the variables tend to show similar behaviour, the covariance is positive.

The covariance between two jointly distributed real-valued random variables $X$ and $Y$ with finite second moments is defined as[2]
cov.

$$\operatorname{cov}(X,Y) = \operatorname{E}\left[(X - \operatorname{E}[X])(Y - \operatorname{E}[Y])\right],$$

where E[$X$] is the expected value of $X$, also known as the mean of $X$. By using the linearity property of expectations, this can be simplified to cov (X , Y )

$$\begin{aligned}
\operatorname{cov}(X,Y) &= \operatorname{E}\left[(X - \operatorname{E}[X])\,(Y - \operatorname{E}[Y])\right] \\
&= \operatorname{E}\left[XY - X\operatorname{E}[Y] - \operatorname{E}[X]\,Y + \operatorname{E}[X]\operatorname{E}[Y]\right] \\
&= \operatorname{E}[XY] - \operatorname{E}[X]\operatorname{E}[Y] - \operatorname{E}[X]\operatorname{E}[Y] + \operatorname{E}[X]\operatorname{E}[Y] \\
&= \operatorname{E}[XY] - \operatorname{E}[X]\operatorname{E}[Y].
\end{aligned}$$

**Covariance Matrix?**
For n variables x1,x2,….,xn for which the covariance of $x_i$ and $x_j$ is $c_{ij}$ is called the covariance matrix. The diagonal terms are carnage $x_i$.
When the variates are standardised with unit variance then this matrix is called correlation matrix.
every covariance matrix is positive semi-definite.

**Chi Squared Test?** $\chi^2$ **test**
is any statistical hypothesis test wherein the sampling distribution of the test statistic is a chi-square distribution when the null hypothesis is true.

**Chi-squared distribution** (also **chi-square** or **$\chi^2$-distribution**) with $k$ degrees of freedom is the distribution of a sum of the squares of $k$ independent standard normal random variables. It is a special case of the gamma distribution and is one of the most widely used probability distributions in inferential statistics, e.g., in **hypothesis testing** or in construction of confidence intervals.The chi-squared distribution is used in the common chi-squared tests for goodness of fit of an observed distribution to a theoretical one

**Khan academy:**

**chi square distribution:**
**for** random variable X~N(0,1) take another RV Q=X^2 , a chi squared distributed random variable, since here we have only one X we have Q with DOF 1.
For RV X1 and X2 which are standard normally distributed independent variables. we sample X1 and X2 from distribution square them and add to get Q2= X1^2 + X2^2 ,this is a chi squared distributed random variable with DOF 2.

FOR Q1 there will be aa high probability of getting Q1 to be close to zero for a sample close to the mean 0. To get a larger value say 4 we will have to sample 2 from X i.e. two standard deviation away from mean, needless to say probability of such a event will be very small.
For Q2 as it is some of two sample squares probability of getting a low value will be lower then that of Q1, but it still weighted towards lower value.
For Q3( DOF =3) the shape starts to shift towards right and slump may form. as DoF increase the lump shifts to right more and more.

**hypothesis test**

make a null hypothesis $H_0$ that the hypothesis is correct
$H_1$ that the hypothesis is not correct
a=0.05 is significance level

Assuming the distribution that we modelled is correct i.e. $H_0$ is correct. To calculate chi-squared statistic we take the diff b/w the the observed and the expected ,square it and normalise it by expected value. Add all these values this get the chi-squared statistic .

- Dof is all the independent Random Variables.
- Critical chi-square value is obtained from chi square char. Example for a=5% and DoF =5 we get a critical value 11.07 .

this tells us that the probability of getting a result atlas as extreme as 11.07 is 5%. If our result of chi-squared value as calculated above is lower then this critical value then we will reject such hypothesis, else we will accept the hypothesis.