Strong Law of Large Numbers

Let $X_1, X_2, \ldots$ be a sequence of i.i.d. random variables. Assume that X_1 has finite mean μ and $E \mid X_1 \mid < \infty$

$$A = \left\{ \lim_{n \to \infty} \frac{X_1 + X_2 + \ldots + X_n}{n} = \mu \right\},$$

then

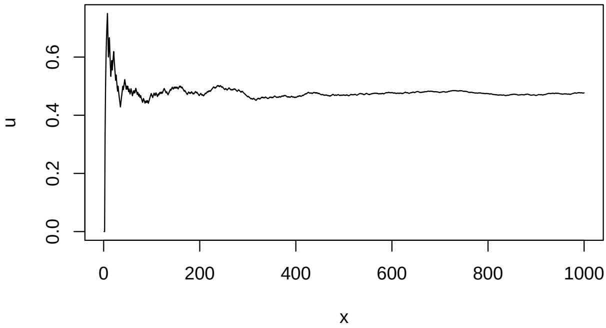
$$P(A) = 1.$$

Law of Large Numbers

```
> runningmean = function (x,N){
+ y = sample(x,N, replace=TRUE)
+ c = cumsum(y)
+ n = 1:N
+ c/n
+ }
> u = runningmean(c(0,1), 1000)
```

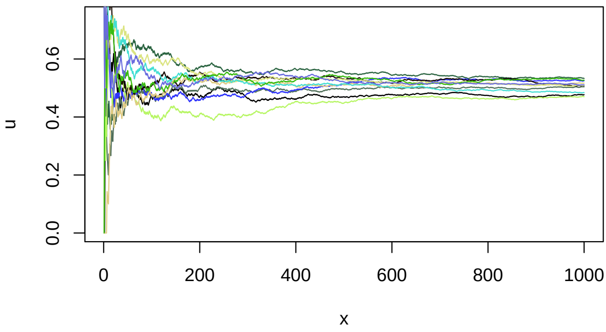
Law of Large Numbers

```
> x=1:1000; plot(u~x, type="l");
>
```

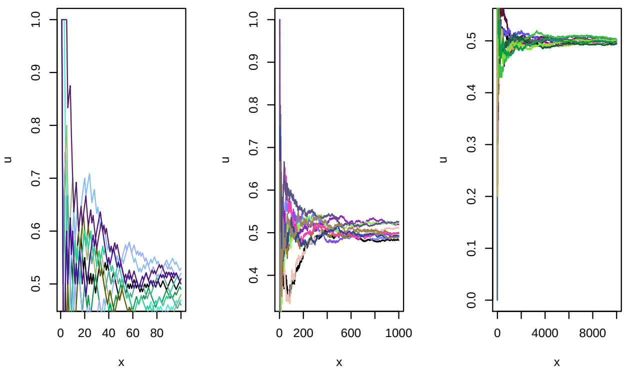


Law of Large Numbers

```
> x=1:1000; plot(u~x, type="l");
> replicate(10, lines(runningmean(c(0,1), 1000)~x, type="l",  col=rgb(runif(3),runif(3),runif(3))))
```



Law of Large Numbers



Simple Linear Regression: Relationship in Bivariate Data

- Key: conditional mean of response variable given the predictor cariable is a linear function.
- Model: For data points (x_i, y_i) with $1 \leq i \leq n$,

$$y_i = \beta_0 + \beta_1 x_i + \varepsilon_i,$$

  where ε_i assumed to be mean 0 and variance σ^2 Normal random variables.
- Observe only (x_i, y_i) for $1 \leq i \leq n$.

Simple Linear Regression: Relationship in Bivariate Data

- Minimize residual sum of squares. Find β_0, β_1 such that

$$\sum_{i=1}^{n} (y_i - \beta_0 - \beta_1 x_i)^2$$

  is minimized.
- Can be solved: Calculus and Linear Algebra

Simple Linear Regression: Relationship in Bivariate Data

- Minimize residual sum of squares. Find β_0, β_1 such that

$$\sum_{i=1}^{n}(y_i - \beta_0 - \beta_1 x_i)^2$$

  is minimized.
- Can be solved using: Calculus and Linear Algebra

Simple Linear Regression: Relationship in Bivariate Data

$$\hat{\beta}_1 = \frac{\sum_{i=1}^{n}(x_i - \bar{x})(y_i - \bar{y})}{\sum_{i=1}^{n}(x_i - \bar{x})^2} = \text{correlation}(x, y)\frac{S_x^2}{S_y^2}$$

$$\hat{\beta}_0 = \bar{y} - \hat{\beta}_1\bar{x}$$

Simple Linear Regression: Relationship in Bivariate Data

$$\hat{\beta}_1 = \frac{\sum_{i=1}^{n}(x_i - \bar{x})(y_i - \bar{y})}{\sum_{i=1}^{n}(x_i - \bar{x})^2} = \text{correlation}(x, y)\frac{S_x^2}{S_y^2}$$

$$\hat{\beta}_0 = \bar{y} - \hat{\beta}_1\bar{x}$$

Observations:

- Slope of line is function of Correlation in standarised scale.
- Line passes through $(\bar{x}, \bar{y})$
- Roles of y and x are not interchangeable.

Simple Linear Regression: Relationship in Bivariate Data

History of Least Squares:- Please refer to the talk that was given in your Writing of Mathematics 2018-course.

- Shape of Earth: Method of Least Squares
- Has incorrect picture of Legendre: 

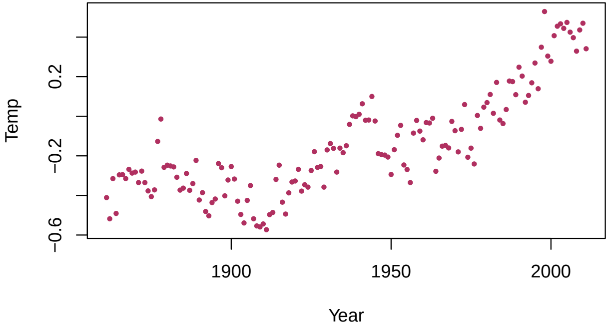
Simple Linear Regression

```
> y = read.csv("annual_temp.csv", header=TRUE)
```

```
> head(y)

  Year   Temp   CO2   CH4   NO2 Irradiance Nino_SST Volcano
1 1861 -0.411 286.5 838.2 288.9   1361.097 26.74233 0.00281
2 1862 -0.518 286.6 839.6 288.9   1360.987 26.39426 0.00859
3 1863 -0.315 286.8 840.9 289.0   1360.837 26.16013 0.01318
4 1864 -0.491 287.0 842.3 289.1   1360.753 26.28774 0.00707
5 1865 -0.296 287.2 843.8 289.1   1360.691 26.32374 0.00302
6 1866 -0.295 287.4 845.5 289.2   1360.600 26.31218 0.00128

> tail(y)

    Year  Temp   CO2    CH4   NO2 Irradiance Nino_SST Volcano
146 2006 0.425 381.9 1784.5 320.0   1361.005 27.25267 0.00342
147 2007 0.397 383.8 1790.4 320.8   1360.939 26.66768 0.00454
148 2008 0.329 385.6 1797.8 321.7   1360.849 26.43034 0.00374
149 2009 0.436 387.4 1802.7 322.4   1360.822 27.50094 0.00402
150 2010 0.470 389.8 1807.7 323.2   1360.841 26.80601 0.00449
151 2011 0.341 391.6 1813.1 324.2   1361.083 26.39182 0.00370
```

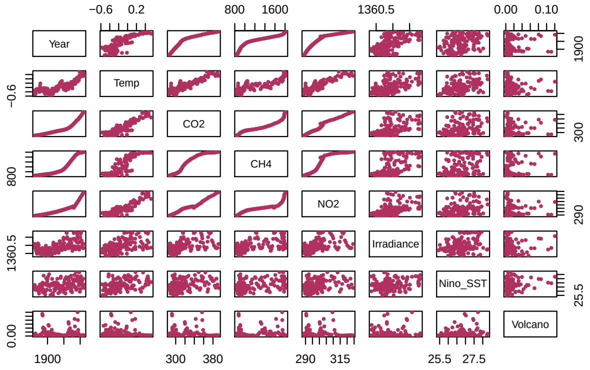
Simple Linear Regression

```
> plot(Temp ~ Year, data = y, pch=19,cex=.5, col="maroon")
```



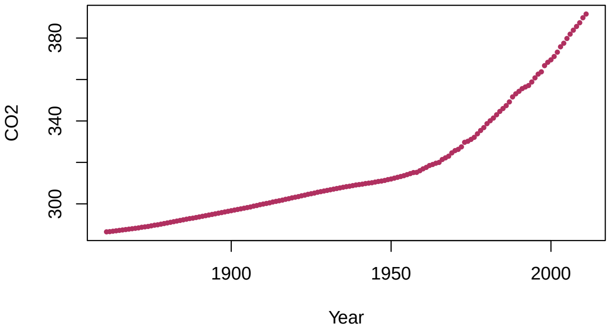
Simple Linear Regression

```
> plot(y, pch=19,cex=.5, col="maroon")
```



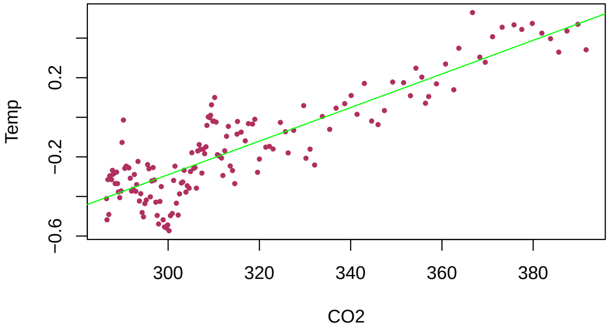
Simple Linear Regression

```
> plot(CO2 ~ Year, data=y, pch=19,cex=.5, col="maroon")
```



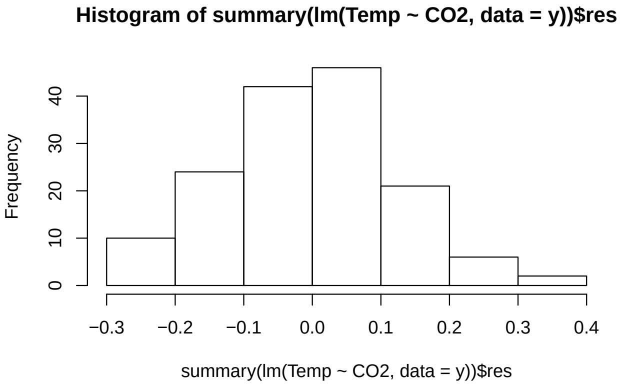
Simple Linear Regression

```
> plot(Temp ~ CO2,  data=y, pch=19,cex=.5, col="maroon")
> abline(lm(Temp ~ CO2, data=y), col="green")
```



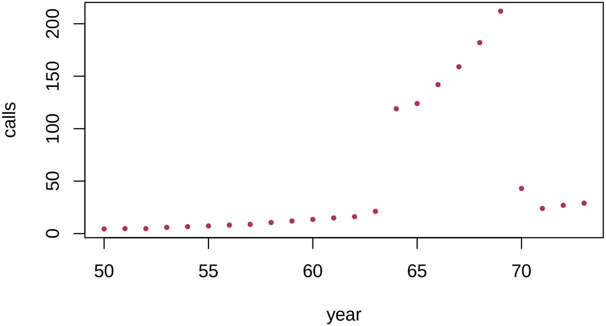
Simple Linear Regression

```
> hist(summary(lm(Temp ~ CO2, data=y))$res)
```



Simple Linear Regression

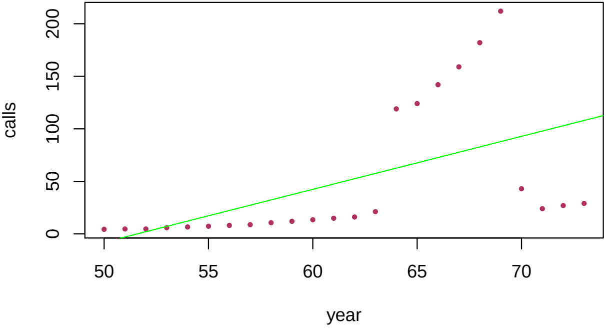
```
> require(MASS)
> plot(calls ~ year,  data=phones, pch=19,cex=.5, col="maroon")
```



Simple Linear Regression

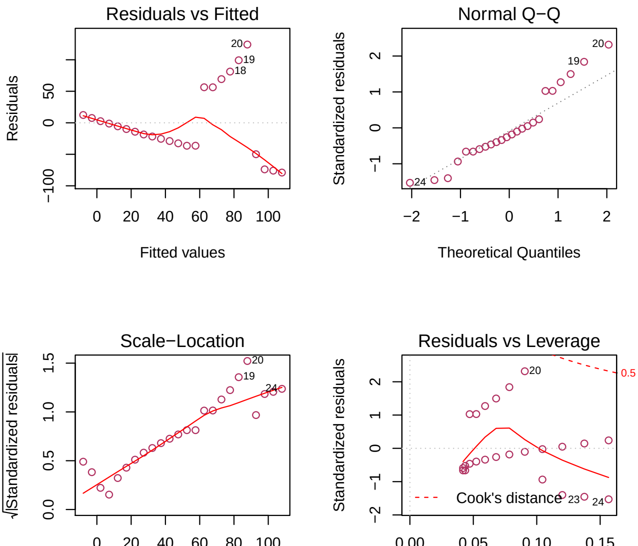
are there better fits ?

```
> plot(calls ~ year,  data=phones, pch=19,cex=.5, col="maroon")
> abline(lm(calls ~ year, data=phones), col="green")
```



Simple Linear Regression

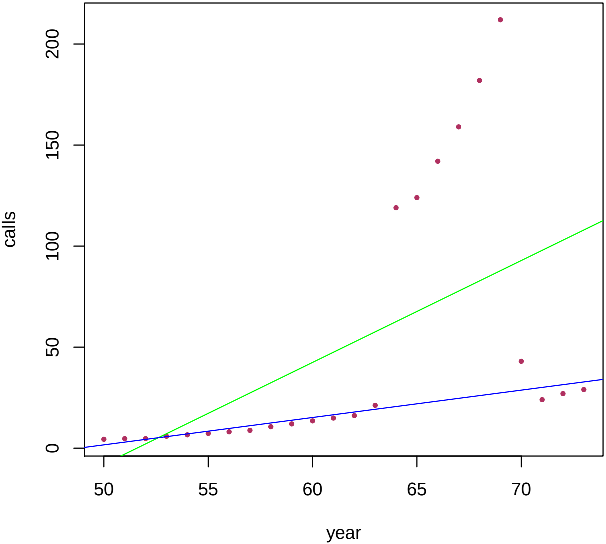
are there better fits ?



Simple Linear Regression

```
> ABSMINLINE = function(x)
+ { with (phones, sum(abs(calls- x[1] -x[2]*year)))
+ }
> OPTIMAL = optim(c(0,0), fn = ABSMINLINE)
> OPTIMAL
$par
[1] -66.053297   1.353735

$value
[1] 844

$counts
function gradient
     117       NA
```

Simple Linear Regression



Simple Linear Regression

```
> plot(calls ~ year,  data=phones, pch=19,cex=.5, col="maro
> abline(lm(calls ~ year, data=phones), col="green")
> abline(rlm(calls ~ year, data=phones), col="green")
```

Simple Linear Regression