Cochran Theorem – from *The distribution of quadratic forms in a normal system, with applications to the analysis of covariance* published in 1934 – is probably the most import one in a regression course. It is an application of a nice result on quadratic forms of Gaussian vectors. More precisely, we can prove that if \boldsymbol{Y}\sim\mathcal{N}(\boldsymbol{0},\mathbb{I}_d) is a random vector with d \mathcal{N}(0,1) variable then (i) if A is a (squared) idempotent matrix \boldsymbol{Y}^\top A\boldsymbol{Y}\sim\chi^2_r where r is the rank of matrix A, and (ii) conversely, if \boldsymbol{Y}^\top A\boldsymbol{Y}\sim\chi^2_r then A is an idempotent matrix of rank r. And just in case, A is an idempotent matrix means that A^2=A, and a lot of results can be derived (for instance on the eigenvalues). The prof of that result (at least the (i) part) is nice: we diagonlize matrix A, so that A=P\Delta P^\top, with P orthonormal. Since A is an idempotent matrix observe thatA^2=P\Delta P^\top=P\Delta P^\top=P\Delta^2 P^\topwhere \Delta is some diagonal matrix such that \Delta^2=\Delta, so terms on the diagonal of \Delta are either 0 or 1‘s. And because the rank of A (and \Delta) is r then there should be r 1‘s and d-r 1‘s. Now write\boldsymbol{Y}^\top A\boldsymbol{Y}= \boldsymbol{Y}^\top P\Delta P^\top\boldsymbol{Y}=\boldsymbol{Z}^\top \Delta\boldsymbol{Z}where \boldsymbol{Z}=P^\top\boldsymbol{Y} that satisfies\boldsymbol{Z}\sim\mathcal{N}(\boldsymbol{0},PP^\top) i.e. \boldsymbol{Z}\sim\mathcal{N}(\boldsymbol{0},\mathbb{I}_d). Thus \boldsymbol{Z}^\top \Delta \boldsymbol{Z}=\sum_{i:\Delta_{i,i}-1}Z_i^2\sim\chi^2_rNice, isn't it. And there is more (that will be strongly connected actually to Cochran theorem). Let A=A_1+\dots+A_k, then the two following statements are equivalent (i) A is idempotent and \text{rank}(A)=\text{rank}(A_1)+\dots+\text{rank}(A_k) (ii) A_i‘s are idempotents, A_iA_j=0 for all i\neq j.

Now, let us talk about projections. Let \boldsymbol{y} be a vector in \mathbb{R}^n. Its projection on the space \mathcal V(\boldsymbol{v}_1,\dots,\boldsymbol{v}_p) (generated by those p vectors) is the vector \hat{\boldsymbol{y}}=\boldsymbol{V} \hat{\boldsymbol{a}} that minimizes \|\boldsymbol{y} -\boldsymbol{V} \boldsymbol{a}\| (in \boldsymbol{a}). The solution is\hat{\boldsymbol{a}}=(\boldsymbol{V}^\top \boldsymbol{V})^{-1} \boldsymbol{V}^\top \boldsymbol{y} \text{ and } \hat{\boldsymbol{y}} = \boldsymbol{V} \hat{\boldsymbol{a}}
Matrix P=\boldsymbol{V} (\boldsymbol{V}^\top \boldsymbol{V})^{-1} \boldsymbol{V}^\top is the orthogonal projection on \{\boldsymbol{v}_1,\dots,\boldsymbol{v}_p\} and \hat{\boldsymbol{y}} = P\boldsymbol{y}.

Now we can recall Cochran theorem. Let \boldsymbol{Y}\sim\mathcal{N}(\boldsymbol{\mu},\sigma^2\mathbb{I}_d) for some \sigma>0 and \boldsymbol{\mu}. Consider sub-vector orthogonal spaces F_1,\dots,F_m, with dimension d_i. Let P_{F_i} be the orthogonal projection matrix on F_i, then (i) vectors P_{F_1}\boldsymbol{X},\dots,P_{F_m}\boldsymbol{X} are independent, with respective distribution \mathcal{N}(P_{F_i}\boldsymbol{\mu},\sigma^2\mathbb{I}_{d_i}) and (ii) random variables \|P_{F_i}(\boldsymbol{X}-\boldsymbol{\mu})\|^2/\sigma^2 are independent and \chi^2_{d_i} distributed.

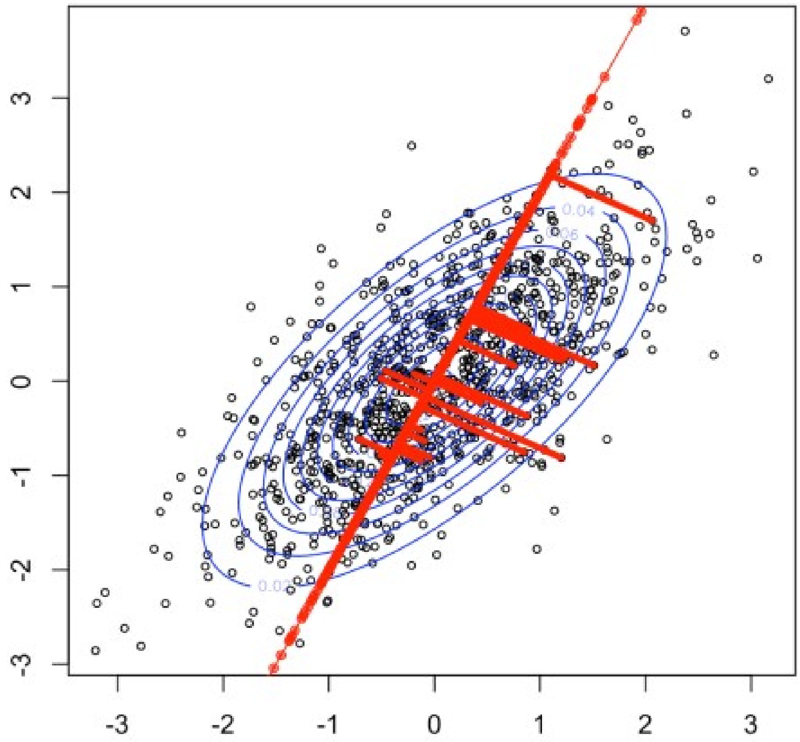
We can try to visualize those results. For instance, the orthogonal projection of a random vector has a Gaussian distribution. Consider a two-dimensional Gaussian vector

```
library(mnormt)
r = .7
s1 = 1
s2 = 1
Sig = matrix(c(s1^2,r*s1*s2,r*s1*s2,s2^2),2,2)
Sig
Y = rmnorm(n = 1000,mean=c(0,0),varcov = Sig)
plot(Y,cex=.6)
vu = seq(-4,4,length=101)
vz = outer(vu,vu,function (x,y) dmnorm(cbind(x,y),
mean=c(0,0), varcov = Sig))
contour(vu,vu,vz,add=TRUE,col='blue')
abline(a=0,b=2,col="red")
```

Consider now the projection of points \boldsymbol{y}=(y_1,y_2) on the straight linear with directional vector \overrightarrow{\boldsymbol{u}} with slope a (say a=2). To get the projected point \boldsymbol{x}=(x_1,x_2) recall that x_2=ay_1 and \overrightarrow{\boldsymbol{x},\boldsymbol{y}}\perp\overrightarrow{\boldsymbol{u}}

. Hence, the following code will give us the orthogonal projections

```
p = function(a){
x0=(Y[,1]+a*Y[,2])/(1+a^2)
y0=a*x0
cbind(x0,y0)
}
```

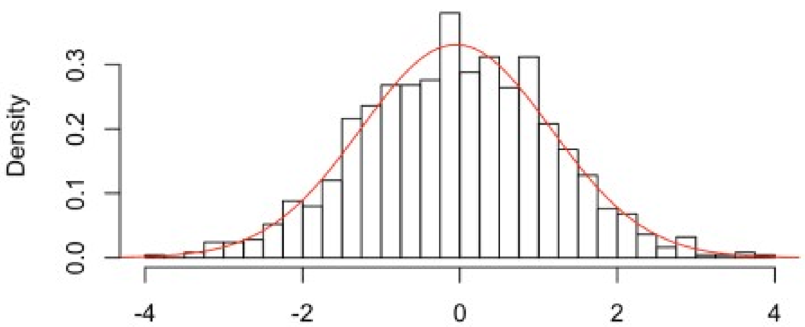


with

```
P = p(2)
for(i in 1:20) segments(Y[i,1],Y[i,2],P[i,1],P[i,2],lwd=4,col="red")
points(P[,1],P[,2],col="red",cex=.7)
```

Now, if we look at the distribution of points on that line, we get… a Gaussian distribution, as expected,

```
z = sqrt(P[,1]^2+P[,2]^2)*c(-1,+1)[(P[,1]>0)*1+1]
vu = seq(-6,6,length=601)
vv = dnorm(vu,mean(z),sd(z))
hist(z,probability = TRUE,breaks = seq(-4,4,by=.25))
lines(vu,vv,col="red")
```



Or course, we can use the matrix representation to get the projection on \overrightarrow{\boldsymbol{u}}, or a normalized version of that vector actually

```
a=2
U = c(1,a)/sqrt(a^2+1)
U
[1] 0.4472136 0.8944272
matP = U %*% solve(t(U) %*% U) %*% t(U)
matP %*% Y[1,]
[,1]
[1,] -0.1120555
[2,] -0.2241110
P[1,]
x0 y0
-0.1120555 -0.2241110
```

(which is consistent with our manual computation). Now, in Cochran theorem, we start with independent random variables,

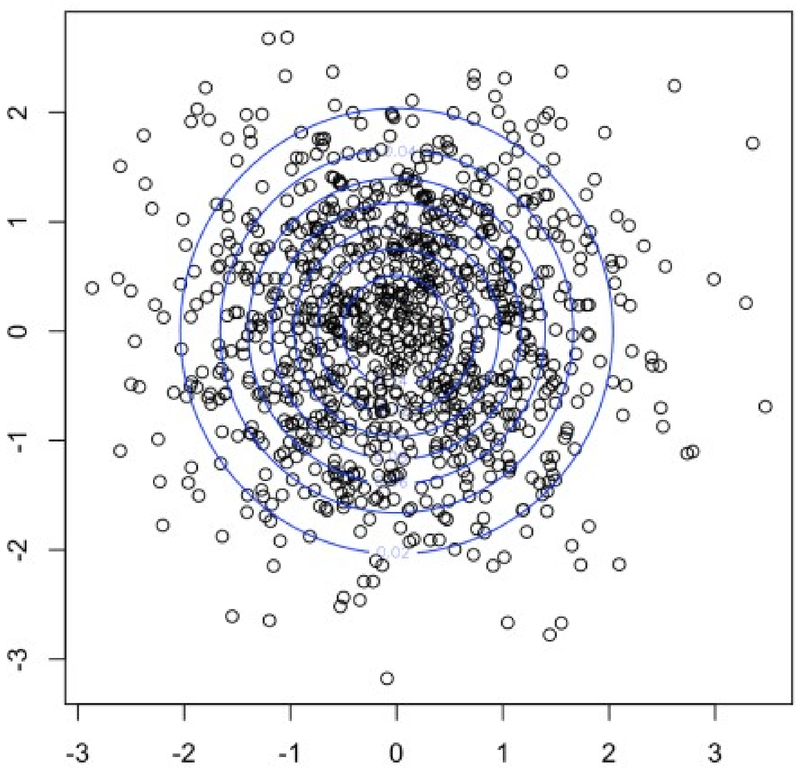
```
Y = rmnorm(n = 1000,mean=c(0,0),varcov = diag(c(1,1)))
```

Then we consider the projection on \overrightarrow{\boldsymbol{u}} and \overrightarrow{\boldsymbol{v}
}=\overrightarrow{\boldsymbol{u}}^\perp

```
U = c(1,a)/sqrt(a^2+1)
matP1 = U %*% solve(t(U) %*% U) %*% t(U)
P1 = Y %*% matP1
z1 = sqrt(P1[,1]^2+P1[,2]^2)*c(-1,+1)[(P1[,1]>0)*1+1]
V = c(a,-1)/sqrt(a^2+1)
matP2 = V %*% solve(t(V) %*% V) %*% t(V)
P2 = Y %*% matP2
z2 = sqrt(P2[,1]^2+P2[,2]^2)*c(-1,+1)[(P2[,1]>0)*1+1]
```

We can plot those two projections

```
plot(z1,z2)
```



and observe that the two are indeed, independent Gaussian variables. And (of course) there squared norms are χ^2_{1} distributed.