Motivation: Projections, Quadratic Forms, χ^2 Distributions:

Suppose

$$\mathbf{Y}^{n\times 1} = \underbrace{\beta_0\mathbf{x}_0 + \beta_1\mathbf{x}_1 + \cdots + \beta_{p-1}\mathbf{x}_{p-1}}_{\equiv\boldsymbol{\mu}} + \boldsymbol{\varepsilon} = \mathbf{X}\boldsymbol{\beta} + \boldsymbol{\varepsilon},$$

where $\boldsymbol{\varepsilon}$ is $N_n(0, \sigma^2\mathbf{I}_n)$. That is, $\mathbf{Y} \sim N_n(\boldsymbol{\mu}, \sigma^2\mathbf{I}_n)$. Let $\hat{\mathbf{Y}}$ be the projection of $\mathbf{Y}$ onto Ω, the space spanned by $\{\mathbf{x}_0, \mathbf{x}_1, \ldots \mathbf{x}_{p-1}\}$. That is, $\hat{\mathbf{Y}} = \mathbf{PY}$ for some projection matrix $\mathbf{P}$. Also, since $\hat{\mathbf{Y}}$ is in the space spanned by $\{\mathbf{x}_0, \mathbf{x}_1, \ldots \mathbf{x}_{p-1}\}$, it follows that $\hat{\mathbf{Y}}$ is a linear combination of the columns of $\mathbf{X}$: $\hat{\mathbf{Y}} = \mathbf{X}\hat{\boldsymbol{\beta}}$.

This is our current definition of $\hat{\boldsymbol{\beta}}$.

Let $\hat{\boldsymbol{\varepsilon}} = \mathbf{Y} - \hat{\mathbf{Y}} = \mathbf{IY} - \mathbf{PY} = (\mathbf{I} - \mathbf{P})\mathbf{Y}$. Then $\hat{\boldsymbol{\varepsilon}}$ and $\hat{\mathbf{Y}}$ are orthogonal:

$$\hat{\boldsymbol{\varepsilon}}'\hat{\mathbf{Y}} = ((\mathbf{I} - \mathbf{P})\mathbf{Y})'\mathbf{PY} = \mathbf{Y}'(\mathbf{I} - \mathbf{P})\mathbf{PY} = \mathbf{Y}'\mathbf{PY} - \mathbf{Y}'\mathbf{P}^2\mathbf{Y} = 0$$

since $\mathbf{P}$ is idempotent.

To summarize,

$$\hat{\boldsymbol{\varepsilon}} \equiv (\mathbf{I} - \mathbf{P})\mathbf{Y} \perp \mathbf{PY} \equiv \hat{\mathbf{Y}},$$

with $\hat{\boldsymbol{\varepsilon}} \in \Omega^{\perp}$.

Since $\hat{\mathbf{Y}}$ and $\mathbf{Y} - \hat{\mathbf{Y}} = \hat{\boldsymbol{\varepsilon}}$ are orthogonal and sum to $\mathbf{Y}$,

$$\mathbf{Y}'\mathbf{Y} = \|\mathbf{Y}\|^2 = \|\hat{\mathbf{Y}} + \mathbf{Y} - \hat{\mathbf{Y}}\|^2 = \overbrace{\|\hat{\mathbf{Y}}\|^2 + \|\mathbf{Y} - \hat{\mathbf{Y}}\|^2}^{by\ Pythagorean\ Theorem}.$$

We can also write this as

$$\begin{aligned} \mathbf{Y}'\mathbf{Y} &= (\mathbf{P}\mathbf{Y})'(\mathbf{P}\mathbf{Y}) + ((\mathbf{I}_n - \mathbf{P})\mathbf{Y})'(\mathbf{I}_n - \mathbf{P})\mathbf{Y} \\ &= \underbrace{\mathbf{Y}'\mathbf{P}\mathbf{Y} + \overbrace{\mathbf{Y}'(\mathbf{I}_n - \mathbf{P})\mathbf{Y}}^{What\ is\ this?\ RSS}}_{What\ are\ these?\ quadratic\ forms} \end{aligned}$$

An important fact is that since $\mathbf{P}$ and $(\mathbf{I}_n - \mathbf{P})$ are projection matrices, $\mathbf{Y}'\mathbf{P}\mathbf{Y}$ and $\mathbf{Y}'(\mathbf{I}_n - \mathbf{P})\mathbf{Y}$ are χ^2–distributed (and vice-versa).

In classical linear model theory, test statistics arise from sums of squares (special cases of quadratic forms) with χ^2 distributions.

Theorem: If $\mathbf{Y} \sim N_n(\boldsymbol{\theta}, \boldsymbol{\Sigma})$ and $\boldsymbol{\Sigma}$ is p.d., then

$$(\mathbf{Y} - \boldsymbol{\theta})'\boldsymbol{\Sigma}^{-1}(\mathbf{Y} - \boldsymbol{\theta}) \sim \chi^2_n.$$

Proof: Since $\boldsymbol{\Sigma}$ is p.d., $\boldsymbol{\Sigma} = \mathbf{A}\mathbf{A}'$ for *non-singular* $\mathbf{A}_{n\times n}$ (Lecture 2, Seber & Lee, A4.2).

By definition 3 of the multivariate normal distribution (Lecture 4), $\mathbf{Y} = \mathbf{A}\mathbf{Z} + \boldsymbol{\theta}$ where $\mathbf{Z}$ is a random sample from $N(0, 1)$. Therefore,

$$\begin{aligned}
(\mathbf{Y} - \boldsymbol{\theta})'\boldsymbol{\Sigma}^{-1}(\mathbf{Y} - \boldsymbol{\theta}) &= (\mathbf{Y} - \boldsymbol{\theta})'(\mathbf{A}\mathbf{A}')^{-1}(\mathbf{Y} - \boldsymbol{\theta}) \\
&= (\mathbf{Y} - \boldsymbol{\theta})'(\mathbf{A}')^{-1}\mathbf{A}^{-1}(\mathbf{Y} - \boldsymbol{\theta}) \\
&= [\mathbf{A}^{-1}(\mathbf{Y} - \boldsymbol{\theta})]'[\mathbf{A}^{-1}(\mathbf{Y} - \boldsymbol{\theta})] \\
&= \mathbf{Z}'\mathbf{Z} \\
&\sim \chi^2_n.
\end{aligned}$$

The next result concerns the spherically symmetric case ($\mathbf{\Sigma} = \sigma^2 \mathbf{I}_n$).

Theorem: (Seber & Lee, Thm 2.7). Let $\mathbf{Y} \sim N_n(\boldsymbol{\theta}, \sigma^2 \mathbf{I})$ and $\mathbf{P}_{n\times n}$ be symmetric of rank r. Then $Q \equiv (\mathbf{Y} - \boldsymbol{\theta})'\mathbf{P}(\mathbf{Y} - \boldsymbol{\theta})/\sigma^2 \sim \chi^2_r$, if and only if $\mathbf{P}$ is idempotent (i.e., $\mathbf{P}^2 = \mathbf{P}$) and hence a projection.

Proof of $\mathbf{P}^2 = \mathbf{P} \Longrightarrow Q \sim \chi^2_r$:

If $\mathbf{P}^2 = \mathbf{P}$, then $\mathbf{P}$ has r eigenvalues $= 1$ and $n - r$ eigenvalues $= 0$. Therefore, there is an orthogonal matrix $\mathbf{T}$ s.t. $\mathbf{T}'\mathbf{P}\mathbf{T} = \mathbf{\Lambda}$, where

$$\mathbf{\Lambda} = diag[\mathbf{I}_r, \mathbf{0}_{n-r}].$$

Now let

$$\mathbf{Z} = \mathbf{T}'(\mathbf{Y} - \boldsymbol{\theta}).$$

Recall that orthogonal transformations of independent normal random vectors are independent with the same variance. This implies that $\mathbf{Z} \sim N_n(\mathbf{0}, \sigma^2 \mathbf{I})$. Therefore,

$$\begin{aligned} Q &\equiv (\mathbf{Y} - \boldsymbol{\theta})'\mathbf{P}(\mathbf{Y} - \boldsymbol{\theta})/\sigma^2 \\ &= (\mathbf{Y} - \boldsymbol{\theta})'\mathbf{T}\mathbf{\Lambda}\mathbf{T}'(\mathbf{Y} - \boldsymbol{\theta})/\sigma^2 \\ &= \frac{1}{\sigma}\mathbf{Z}'\mathbf{\Lambda}\frac{1}{\sigma}\mathbf{Z} \\ &= \sum_{i=1}^{r} (\frac{Z_i}{\sigma})^2 \\ &\sim \chi^2_r. \end{aligned}$$

Proof of $Q \sim \chi^2_r \Longrightarrow \mathbf{P}^2 = \mathbf{P}$: (See Seber, 1977, Thm 2.8, p. 37).

Interpretation: in the spherically symmetric case ($\mathbf{\Sigma} = \sigma^2 \mathbf{I}$), the only quadratic forms with χ^2 distributions are sums of squares, i.e., squared lengths of projections ($\mathbf{x}'\mathbf{P}\mathbf{x} = ||\mathbf{P}\mathbf{x}||^2$).

Non-Central χ^2 Distribution

Definition: The *non-central chi-squared distribution with n degrees of freedom and non-centrality parameter λ*, denoted $\chi^2_n(\lambda)$, is defined as the distribution of $\sum_{i=1}^n Z_i^2$, where $Z_1, \ldots, Z_n$ are independent $N(\mu_i, 1)$ r.v.'s and $\lambda = \sum_{i=1}^n \mu_i^2/2$.

Theorem: If $\mathbf{Y} \sim N_n(\boldsymbol{\mu}, \mathbf{I})$ $\boldsymbol{\mu} \in \Re^n$, then $\mathbf{Y}'\mathbf{Y}$ has moment generating function

$$M_{\mathbf{Y}'\mathbf{Y}}(t) = (1-2t)^{-\frac{n}{2}} exp\{\frac{\boldsymbol{\mu}'\boldsymbol{\mu}}{2}[\frac{1}{1-2t} - 1]\}, \quad t < 1/2.$$

Proof: Suppose first that $n = 1$. For any $t < 1/2$,

$$M_{Y^2}(t) = E[e^{Y^2 t}] = \int_{-\infty}^{\infty} e^{y^2 t} \frac{e^{-(y-\mu)^2/2}}{\sqrt{2\pi}}\, dy = \int_{-\infty}^{\infty} \frac{exp\{y^2 t - (y-\mu)^2/2\}}{\sqrt{2\pi}}\, dy$$

Expand the exponential and collect terms:

$$= \int_{-\infty}^{\infty} \frac{1}{\sqrt{2\pi}} exp\{-\frac{(1-2t)}{2} y^2 + y\mu - \frac{\mu^2}{2}\}\, dy$$

Substitute $x = y\sqrt{1-2t}$:

$$= \int_{-\infty}^{\infty} \frac{1}{\sqrt{2\pi}} exp\{-\frac{x^2}{2} + \frac{x\mu}{\sqrt{1-2t}} - \frac{\mu^2}{2}\}\, \frac{dx}{\sqrt{1-2t}}$$

Complete the square:

$$= (1-2t)^{-\frac{1}{2}} \int_{-\infty}^{\infty} \frac{1}{\sqrt{2\pi}} exp\{-\frac{1}{2}(x - \frac{\mu}{\sqrt{1-2t}})^2 + \frac{\mu^2}{2}[\frac{1}{1-2t} - 1]\}\, dx$$

$$= (1-2t)^{-\frac{1}{2}} exp\{\frac{\mu^2}{2}[\frac{1}{1-2t} - 1]\}.$$

Now suppose $n \geq 2$. For any $t < 1/2$, we have

$$M_{\mathbf{Y}'\mathbf{Y}}(t) = E[exp\{t(\mathbf{Y}'\mathbf{Y})\}] = E[exp\{\sum_{i=1}^{n} tY_i^2\}] = E[\prod_{i=1}^{n} exp\{tY_i^2\}]$$

$$\stackrel{\text{ind}}{=} \prod_{i=1}^{n} E[exp\{tY_i^2\}] = \prod_{i=1}^{n} M_{Y_i^2}(t) = \prod_{i=1}^{n}[(1-2t)^{-\frac{1}{2}} exp\{\frac{\mu_i^2}{2}[\frac{1}{1-2t} - 1]\}]$$

$$= (1-2t)^{-\frac{n}{2}} exp\{\frac{\sum_{i=1}^{n} \mu_i^2}{2}[\frac{1}{1-2t} - 1]\} = (1-2t)^{-\frac{n}{2}} exp\{\frac{\boldsymbol{\mu}'\boldsymbol{\mu}}{2}[\frac{1}{1-2t} - 1]\}.$$

This is the m.g.f. of the non-central chi-square distribution, $\chi_n^2(\boldsymbol{\mu}'\boldsymbol{\mu}/2)$, with noncentrality parameter $\lambda = \boldsymbol{\mu}'\boldsymbol{\mu}/2$.

We refer to $\chi_n^2(0) \equiv \chi_n^2$ as the *central chi-square distribution.*

Basic facts:

1. If $Y \sim \chi^2(n, \lambda)$, then $E[Y] = n + 2\lambda$
2. If $Y \sim \chi_n^2$ with $n > 2$, then $E[\frac{1}{Y}] = \frac{1}{n-2}$

   Theorem: Let $\mathbf{Y} \sim N_n(\boldsymbol{\mu}, \sigma^2\mathbf{I}_n)$ and $\mathbf{P} = \mathbf{P}'$. $\mathbf{P} = \mathbf{P}^2$ and $rank[\mathbf{P}] = r$ if and only if

   $$\mathbf{Y}'\mathbf{P}\mathbf{Y}/\sigma^2 \sim \chi_r^2(\boldsymbol{\mu}'\mathbf{P}\boldsymbol{\mu}/2\sigma^2).$$

3. $\chi_n^2(\boldsymbol{\mu}'\boldsymbol{\mu}/2)$ depends upon $\boldsymbol{\mu}$ only through $\boldsymbol{\mu}'\boldsymbol{\mu} = \sum_{i=1}^{n} \mu_i^2$

Our next goal is to understand the conditions under which the difference of two χ^2-distributed quadratic forms is χ^2 (to be applied to the ANOVA decomposition of the sum of squares). To get there, we will need to know when the difference of two projection matrices is a projection matrix.

Lemma: (Seber & Lee, A6.5). Assume $\mathbf{P}_1$ and $\mathbf{P}_2$ are projection matrices and $\mathbf{P}_1 - \mathbf{P}_2$ is p.s.d. Then (i) $\mathbf{P}_1\mathbf{P}_2 = \mathbf{P}_2\mathbf{P}_1 = \mathbf{P}_2$, and (ii) $\mathbf{P}_1 - \mathbf{P}_2$ is a projection matrix.

Property (i) above says that applying both projections (in either order) is equivalent to just applying $\mathbf{P}_2$.

Interpretation:

1. $\mathbf{P}_1$ is a projection onto a linear space Ω.
2. $\mathbf{P}_2$ is a projection onto a *subspace* ω of Ω.
3. $\mathbf{P}_1 - \mathbf{P}_2$ is a projection onto the orthogonal complement of ω within Ω.

Example: Projections of $(x, y, z)'$.

$$\mathbf{P}_1 = \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 0 \end{pmatrix} \text{ projection onto } (x, y) \text{ plane}$$

$$\mathbf{P}_2 = \begin{pmatrix} 1 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{pmatrix} \text{ projection onto } x \text{ axis}$$

$$\mathbf{P}_1 - \mathbf{P}_2 = \begin{pmatrix} 0 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 0 \end{pmatrix} \text{ projection onto } y \text{ axis}$$

Theorem: Let $\mathbf{Y} \sim N_n(\boldsymbol{\theta}, \sigma^2\mathbf{I})$ and let $Q_1 = (\mathbf{Y}-\boldsymbol{\theta})'\mathbf{P}_1(\mathbf{Y}-\boldsymbol{\theta})/\sigma^2$ and $Q_2 = (\mathbf{Y}-\boldsymbol{\theta})'\mathbf{P}_2(\mathbf{Y}-\boldsymbol{\theta})/\sigma^2$ where $\mathbf{P}_1$ and $\mathbf{P}_2$ are symmetric $n \times n$ matrices. If $Q_i \sim \chi^2_{r_i}$ and $Q_1 - Q_2 \geq 0$, then $Q_1 - Q_2$ and Q_2 are independent, and $Q_1 - Q_2 \sim \chi^2_{r_1 - r_2}$.

Proof:

Because $Q_i \sim \chi^2_{r_i}$, we know $\mathbf{P}_i$ is idempotent by the previous theorem (Seber & Lee, Thm 2.7), and so a projection matrix. Because

$$Q_1 - Q_2 = (\mathbf{Y}-\boldsymbol{\theta})'(\mathbf{P}_1 - \mathbf{P}_2)(\mathbf{Y}-\boldsymbol{\theta})/\sigma^2 \geq 0,$$

we know $\mathbf{P}_1 - \mathbf{P}_2$ is p.s.d. Therefore, by the lemma, $\mathbf{P}_1 - \mathbf{P}_2$ is a projection matrix. Now the previous theorem implies

$$Q_1 - Q_2 = (\mathbf{Y}-\boldsymbol{\theta})'(\mathbf{P}_1 - \mathbf{P}_2)(\mathbf{Y}-\boldsymbol{\theta})/\sigma^2 \sim \chi^2_r,$$

where $r = \text{rank}(\mathbf{P}_1 - \mathbf{P}_2)$. But

$$\text{rank}(\mathbf{P}_1 - \mathbf{P}_2) = \text{rank}(\mathbf{P}_1) - \text{rank}(\mathbf{P}_2),$$

because rank=trace for projection matrices.

The independence of Q_2 and $Q_1 - Q_2$ follows from $\mathbf{P}_2(\mathbf{P}_1 - \mathbf{P}_2) = \mathbf{0}$. (Check this.) (Seber & Lee, Theorem 2.5)

$\mathbf{P}_2(\mathbf{P}_1 - \mathbf{P}_2) = \mathbf{P}_2\mathbf{P}_1 - \mathbf{P}_2^2 = \mathbf{P}_2\mathbf{P}_1 - \mathbf{P}_2 = 0$ by lemma p.7