17.1. Effect of Mis-specified Covariance Matrix

Assume we have specified $E[\mathbf{Y}] = \mathbf{X}\boldsymbol{\beta}$ correctly but suppose

$$\text{cov}(\boldsymbol{\varepsilon}) = \sigma^2\mathbf{V}$$

when we assume

$$\text{cov}(\boldsymbol{\varepsilon}) = \sigma^2\mathbf{I}.$$

Is $\hat{\boldsymbol{\beta}}$ unbiased?

Yes, we saw this when we discussed Generalized Least Squares. Full rank case: $E(\hat{\boldsymbol{\beta}}) = E((\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{Y}) = (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'E(\mathbf{Y}) = (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{X}\boldsymbol{\beta} = \boldsymbol{\beta}$.

When $\text{cov}(\boldsymbol{\varepsilon}) = \sigma^2\mathbf{V}$ but we proceed under the assumption that $\text{cov}(\boldsymbol{\varepsilon}) = \sigma^2\mathbf{I}$, we already know that we are (probably) using a less efficient estimator. In addition, the covariance matrix of $\hat{\boldsymbol{\beta}}$ will not necessarily be equal to $\sigma^2(\mathbf{X}'\mathbf{X})^{-1}$. We have instead:

$$\begin{aligned}
\text{cov}(\hat{\boldsymbol{\beta}}) &= \text{cov}[(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{Y}] \\
&= (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\sigma^2\mathbf{V}\mathbf{X}(\mathbf{X}'\mathbf{X})^{-1} \\
&= \sigma^2(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{V}\mathbf{X}(\mathbf{X}'\mathbf{X})^{-1}.
\end{aligned}$$

Also,

$$\begin{aligned}
E[S^2] &= \frac{1}{n-p}E[\mathbf{Y}'(\mathbf{I}-\mathbf{P})\mathbf{Y}] \\
&= \frac{1}{n-p}\text{tr}[\sigma^2\mathbf{V}(\mathbf{I}-\mathbf{P})] \\
&= \frac{\sigma^2}{n-p}\text{tr}[\mathbf{V}(\mathbf{I}-\mathbf{P})].
\end{aligned}$$

Therefore, in most cases, S^2 is biased, i.e., $E[S^2] \neq \sigma^2$.

17.2. Effect of Non-constant Variance on Two-sample t-test (Scheffé, 10.2)

Model:

$$Y_{ij} = \mu_i + \varepsilon_{ij}, \quad \text{var}(\varepsilon_{ij}) = \sigma_i^2, \quad i = 1, 2; j = 1, \ldots, n_i.$$

The t-statistic used for testing $\mu_1 - \mu_2 = 0$ is

$$T = \frac{\bar{Y}_1 - \bar{Y}_2}{S(n_1^{-1} + n_2^{-1})^{1/2}},$$

where

$$S^2 = \frac{1}{n-2} \sum_i \sum_j (Y_{ij} - \bar{Y}_i)^2 = \frac{(n_1 - 1)s_1^2 + (n_2 - 1)s_2^2}{n-2},$$

$n = n_1 + n_2$ and s_i^2 is the sample variance in the ith group. Now, if $\sigma_1^2 = \sigma_2^2$ and ε_{ij} is normally distributed, then

$$T \sim t_{n-2} \approx N(0, 1), \quad \text{for large } n.$$

We now derive the approximate distribution of T when $\sigma_1^2 \neq \sigma_2^2$.

For large n, $S^2 \approx \frac{1}{n}(n_1\sigma_1^2 + n_2\sigma_2^2)$
and T is approximately normally distributed with mean 0 and

$$\begin{aligned}
\text{var}(T) &\approx \frac{\text{var}(\bar{Y}_1 - \bar{Y}_2)}{\frac{1}{n}(n_1\sigma_1^2 + n_2\sigma_2^2)(n_1^{-1} + n_2^{-1})} \\
&= \frac{n_1^{-1}\sigma_1^2 + n_2^{-1}\sigma_2^2}{\frac{1}{n}(n_1\sigma_1^2 + n_2\sigma_2^2)(n_1^{-1} + n_2^{-1})} \times \frac{\frac{n_1}{\sigma_2^2}}{\frac{n_1}{\sigma_2^2}} \\
&= \frac{\frac{\sigma_1^2}{\sigma_2^2} + \frac{n_1}{n_2}}{\frac{n_1}{n_2}\frac{\sigma_1^2}{\sigma_2^2} + 1}.
\end{aligned}$$

When is $\text{var}(T) \approx 1$ *for large* n*?*

$$\begin{aligned}
\frac{\sigma_1^2}{\sigma_2^2} + \frac{n_1}{n_2} &\approx \frac{n_1}{n_2}\frac{\sigma_1^2}{\sigma_2^2} + 1 \\
\frac{\sigma_1^2}{\sigma_2^2} - 1 &\approx \frac{n_1}{n_2}\frac{\sigma_1^2}{\sigma_2^2} - \frac{n_1}{n_2} \\
&= \frac{n_1}{n_2}(\frac{\sigma_1^2}{\sigma_2^2} - 1)
\end{aligned}$$

Answer: If either

1. $\sigma_1^2 = \sigma_2^2$, i.e., equal variance assumption holds, OR
2. $n_1 = n_2$, i.e., sample sizes are equal, regardless of equality of variances

Effect on error rate of 95% CI for $\mu_1 - \mu_2$.

A the 95% CI for $\mu_1 - \mu_2$:

$$[\bar{Y}_1 - \bar{Y}_2 - t^{.025}_{n-2} S(n_1^{-1} + n_2^{-1})^{1/2}, \bar{Y}_1 - \bar{Y}_2 + t^{.025}_{n-2} S(n_1^{-1} + n_2^{-1})^{1/2}].$$

The error rate of this CI is

$$\begin{aligned} P(\mu_1 - \mu_2 \notin CI) &= P(|T| > t^{.025}_{n-2}) \\ &\approx P(|N(0, v)| > t^{.025}_{n-2}), \end{aligned}$$

where $v = (\frac{\sigma_1^2}{\sigma_2^2} + \frac{n_1}{n_2})/(\frac{n_1}{n_2}\frac{\sigma_1^2}{\sigma_2^2} + 1)$.

Some values of the error rate based on the above normal approximation are given in the table below. The error rate does not deviate too far from the nominal value of 0.05 unless both the sample sizes and the variances differ substantially between groups.

	σ_1^2/σ_2^2				
n_1/n_2	$\frac{1}{5}$	$\frac{1}{2}$	1	2	5
1	.05	.05	.05	.05	.05
2	.12	.08	.05	.029	.014
5	.22	.12	.05	.014	.002

17.3. Effect of Serial Correlation on CI for a Mean

Assume we have a set of normally distributed observations from the following model:

$$Y_i = \mu + \varepsilon_i, \quad \text{var}(\varepsilon_i) = \sigma^2, \quad i = 1, \ldots, n.$$

Assuming the observations are independent, we use the following CI for μ: $[\bar{Y} - t_{n-1}^{\alpha/2} s/\sqrt{n}, \bar{Y} + t_{n-1}^{\alpha/2} s/\sqrt{n}]$, with nominal coverage probability $1 - \alpha$, where s is the sample variance. This CI is based on the t statistic $T \equiv \frac{\bar{Y}-\mu}{s/\sqrt{n}} \sim t_{n-1}$.

Now suppose that adjacent observations have correlation ρ and all other pairs are uncorrelated:

$$\text{corr}(\varepsilon_i, \varepsilon_{i-1}) = \rho, \quad i = 2, \ldots, n,$$

and

$$\text{corr}(\varepsilon_i, \varepsilon_j) = 0, \quad |i - j| > 1.$$

Then the variance of $\bar{Y}$ is

$$\text{var}(\bar{Y}) = \frac{\sigma^2}{n}\left[1 + 2\rho\left(1 - \frac{1}{n}\right)\right],$$

and

$$E[s^2] = \sigma^2\left(1 - \frac{2\rho}{n}\right).$$

Note that since $\left(1 - \frac{2\rho}{n}\right) < 1$ when $\rho > 0$, this indicates a tendency to underestimate the variance - anti-conservative inference.

Heuristically (substituting $s^2 \approx \sigma^2$), when n is large we have that T is approximately normal with mean 0 and variance

$$\text{var}(T) \approx \frac{\text{var}(\bar{Y})}{\sigma^2/n} = \frac{(\sigma^2/n)[1 + 2\rho(1 - 1/n)]}{\sigma^2/n} \approx 1 + 2\rho.$$

17.4. Effect of Non-normality

Suppose we have correctly specified the model

$$\mathbf{Y} = \mathbf{X}\boldsymbol{\beta} + \boldsymbol{\varepsilon}, \quad E[\boldsymbol{\varepsilon}] = \mathbf{0}, \quad \text{cov}(\boldsymbol{\varepsilon}) = \sigma^2\mathbf{I},$$

but suppose that $\boldsymbol{\varepsilon}$ is not necessarily MVN.

We have seen previously that, in the full rank case, $\hat{\boldsymbol{\beta}}$ is unbiased, and $\text{cov}(\hat{\boldsymbol{\beta}}) = \sigma^2(\mathbf{X}'\mathbf{X})^{-1}$, without any distributional assumptions. In fact, we proved $\hat{\boldsymbol{\beta}}$ that is the BLUE without any distributional assumptions on $\boldsymbol{\varepsilon}$ except $E[\boldsymbol{\varepsilon}] = \mathbf{0}$ and $\text{cov}(\boldsymbol{\varepsilon}) = \sigma^2\mathbf{I}$.

Under some regularity conditions, the usual distributional properties of $\hat{\boldsymbol{\beta}}$ and the F test statistic still hold approximately for large n. For example,

$$\hat{\boldsymbol{\beta}} \approx N_p(\boldsymbol{\beta}, \sigma^2(\mathbf{X}'\mathbf{X})^{-1}).$$

In addition, inferences based on F-tests using an F distribution will be approximately correct.

17.5. Effect of Non-normality on Inference for a Mean

Assume we have an independent sample from a distribution with mean μ and variance σ^2, i.e.,

$$Y_i = \mu + \varepsilon_i, \quad \text{var}(\varepsilon_i) = \sigma^2, \quad i = 1, \ldots, n.$$

If ε_i is normally distributed, then the one-sample t statistic satisfies

$$T \equiv \frac{\bar{Y} - \mu}{s/\sqrt{n}} \sim t_{n-1} \approx N(0,1), \quad \text{for large } n.$$

For large n, $s^2 \approx \sigma^2$ and

$$T \approx \frac{\bar{Y} - \mu}{\sigma/\sqrt{n}} \approx N(0,1),$$

by the Central Limit Theorem, even if the errors are not normal.

The usual 95% CI for μ (assuming ε_i is normal) is

$$[\bar{Y} - t^{.025}_{n-1} s/\sqrt{n}, \bar{Y} + t^{.025}_{n-1} s/\sqrt{n}],$$

and the error rate of this CI is

$$P(\mu \notin CI) = P(|T| > t^{.025}_{n-1}).$$

As an extreme example we consider the case of a Bernoulli population with success probability p, i.e., $P(Y_i = 1) = p$, $P(Y_i = 0) = 1 - p$. Note p is the mean of the distribution, which is decidedly non-normal. The error rate for the CI is given in the following table. It does not deviate greatly from the nominal value of 0.05 unless n is fairly small and/or p is fairly extreme.

Error rate for the CI for μ for a Bernoulli population with success probability p:

	p				
n	0.1	0.2	0.3	0.4	0.5
10	0.350	0.114	0.039	0.059	0.021
20	0.124	0.079	0.053	0.072	0.041
50	0.121	0.062	0.053	0.044	0.065
100	0.068	0.058	0.050	0.052	0.057
200	0.073	0.059	0.056	0.051	0.056

General comments on the effect of non-normality:

1. The effect of non-normality on the type I error rate of F-tests depends more critically on the *kurtosis* of the distribution (heaviness of the tails) rather than the *skewness*. (On the other hand, beware of one-sided t-tests with skewed data.)
2. In ANOVA, the effect of non-normality is less severe in balanced designs (see Seber & Lee §9.5.2).