STAT 714 fa 2023 Lec 05

Inference in linear models, general linear hypothesis

Karl B. Gregory

University of South Carolina

These slides are an instructional aid; their sole purpose is to display, during the lecture, definitions, plots, results, etc. which take too much time to write by hand on the blackboard. They are not intended to explain or expound on any material.

Assume throughout that $\mathbf{y} = \mathbf{X}\mathbf{b} + \mathbf{e}$, $\mathbf{e} \sim \text{Normal}(\mathbf{0}, \sigma^2 \mathbf{I}_n)$.

We know that $\mathbf{c}^T\hat{\mathbf{b}}$ is the BLUE for $\mathbf{c}^T\mathbf{b}$, but is it the UMVUE?

Result (UMVUE for an estimable contrast under Normality)

Let $\mathbf{c}^T\mathbf{b}$ *be estimable in the model* $\mathbf{y} = \mathbf{X}\mathbf{b} + \mathbf{e}$, *where* $\mathbf{e} \sim \text{Normal}(\mathbf{0}, \sigma^2 \mathbf{I}_n)$, *and suppose* $\hat{\mathbf{b}}$ *satisfies* $\mathbf{X}^T\mathbf{X}\hat{\mathbf{b}} = \mathbf{X}^T\mathbf{y}$. *Then* $\mathbf{c}^T\mathbf{b}$ *is UMVUE for* $\mathbf{c}^T\mathbf{b}$.

See Res 6.2 and Cor 6.2 of Monahan (2008). Recall Lehmann-Scheffé.

Recall definition of UMVUE and prove the result.

Result (MLEs in Normal model)

Suppose $\mathbf{y} = \mathbf{X}\mathbf{b} + \mathbf{e}$, *where* $\mathbf{e} \sim \text{Normal}(\mathbf{0}, \sigma^2 \mathbf{I}_n)$, *and let* $\mathbf{c}^T\mathbf{b}$ *be estimable. The MLE of* $(\mathbf{c}^T\mathbf{b}, \sigma^2)$ *is given by* $(\mathbf{c}^T\hat{\mathbf{b}}, \|\hat{\mathbf{e}}\|^2/n)$, *where* $\hat{\mathbf{b}}$ *satisfies* $\mathbf{X}^T\mathbf{X}\hat{\mathbf{b}} = \mathbf{X}^T\mathbf{y}$.

See Res 6.3 and Cor 6.3 of Monahan (2008).

Derive the result.

The general linear hypothesis

In the context of the model $\mathbf{y} = \mathbf{X}\mathbf{b} + \mathbf{e}$, the *general linear hypothesis* is expressed

$$H_0\colon \mathbf{K}^T\mathbf{b} = \mathbf{m} \quad \text{versus} \quad H_1\colon \mathbf{K}^T\mathbf{b} \neq \mathbf{m},$$

where $\mathbf{K}$ is a $p \times s$ matrix and $\mathbf{m}$ is a $s \times 1$ vector.

The general linear hypothesis is *testable* provided that:

1. Each column of $\mathbf{K}$ defines an estimable contrast.
2. The columns of $\mathbf{K}$ are linearly independent (contrasts are non-redundant).

If one or the other condition does not hold, the hypothesis is *non-testable*.

See Def 6.1 of Monahan (2008).

Exercise: Consider the treatment effect model

$$Y_{ij} = \mu + \alpha_i + \varepsilon_{ij}, \quad \varepsilon_{ij} \overset{\text{ind}}{\sim} \text{Normal}(0, \sigma^2), \quad j = 1, \dots, n_i,\ i = 1, \dots, 3.$$

Formulate the general linear hypothesis for testing for a treatment effect.

Result (Helper result for building test statistic for gen. lin. hyp.)

If H_0: $\mathbf{K}^T\mathbf{b} = \mathbf{m}$ is testable, then the matrix $\mathbf{K}^T(\mathbf{X}^T\mathbf{X})^-\mathbf{K}$ is invertible.

See Res 6.4 of Monahan (2008).

Prove the result.

Exercise: Assume $\mathbf{y} = \mathbf{X}\mathbf{b} + \mathbf{e}$, $\mathbf{e} \sim \text{Normal}(\mathbf{0}, \sigma^2 \mathbf{I}_n)$ and consider the *ad-hoc rule*

$$F = \frac{(\mathbf{K}^T\hat{\mathbf{b}} - \mathbf{m})^T[\mathbf{K}^T(\mathbf{X}^T\mathbf{X})^-\mathbf{K}]^{-1}(\mathbf{K}^T\hat{\mathbf{b}} - \mathbf{m})/s}{\mathbf{y}^T(\mathbf{I}_n - \mathbf{P}_\mathbf{X})\mathbf{y}/(n-r)} > F_{s,n-r,\alpha},$$

for rejecting testable H_0: $\mathbf{K}^T\mathbf{b} = \mathbf{m}$, where $s = \operatorname{rank}\mathbf{K}$ and $r = \operatorname{rank}\mathbf{X}$.

1. Give the size of the test.
2. Give an expression for the power of the test.
3. Check if the test is unbiased (has greater power over H_1 than under H_0).

Exercise: Consider the multiple linear regression model

$$Y_i = \beta_0 + \beta_1 x_{i1} + \cdots + \beta_p x_{ip} + \varepsilon_i, \quad \varepsilon_i \overset{\text{ind}}{\sim} \text{Normal}(0, \sigma^2), \quad i = 1, \ldots, n.$$

Assume a full-rank design matrix.

1. Give the size α ad-hoc rule for testing H_0: $\beta_j = 0$ for any $j = 1, \ldots, p$.
2. Give an expression of the power of the test.

Likelihood ratio and LRT

Suppose $\theta \in \Theta$ is a parameter with likelihood $\mathcal{L}(\theta; \mathbf{y})$ for some data $\mathbf{y}$. Consider

$$H_0\colon \theta \in \Theta_0 \quad \text{vs} \quad H_1\colon \theta \in \Theta \setminus \Theta_0.$$

The *likelihood ratio* is defined as $\text{LR}(\mathbf{y}; \Theta_0) = \dfrac{\sup_{\theta \in \Theta_\mathbf{0}} \mathcal{L}(\theta; \mathbf{y})}{\sup_{\theta \in \Theta} \mathcal{L}(\theta; \mathbf{y})}$.

The *likelihood ratio test* is the test which rejects H_0 when $\text{LR}(\mathbf{y}; \Theta_0) < c$.

Choose c to give the test size α or find an equivalent test easier to calibrate.

Result (LRT is equivalent to ad-hoc test of general linear hyp)

Let $\mathbf{y} = \mathbf{X}\mathbf{b} + \mathbf{e}$, $\mathbf{e} \sim N(\mathbf{0}, \sigma^2 \mathbf{I}_n)$ *and* H_0: $\mathbf{K}^T\mathbf{b} = \mathbf{m}$ *be testable. Then the rule*

$$\text{Reject } H_0 \text{ if} \quad F := \frac{(\|\mathbf{y} - \mathbf{X}\hat{\mathbf{b}}_{\text{null}}\|^2 - \|\mathbf{y} - \mathbf{X}\hat{\mathbf{b}}\|^2)/s}{\|\mathbf{y} - \mathbf{X}\hat{\mathbf{b}}\|^2/(n-r)} > F_{s,n-r}$$

is the size-α *LRT, where* $\hat{\mathbf{b}}_{\text{null}}$ *minimizes* $\|\mathbf{y} - \mathbf{X}\mathbf{b}\|^2$ *subject to* $\mathbf{K}^T\mathbf{b} = \mathbf{m}$. *Also*

$$F = \frac{(\mathbf{K}^T\hat{\mathbf{b}} - \mathbf{m})^T[\mathbf{K}^T(\mathbf{X}^T\mathbf{X})^-\mathbf{K}]^{-1}(\mathbf{K}^T\hat{\mathbf{b}} - \mathbf{m})/s}{\mathbf{y}^T(\mathbf{I}_n - \mathbf{P}_\mathbf{X})\mathbf{y}/(n-r)},$$

which has distribution $F_{s,n-r}\left(\phi = \frac{1}{\sigma^2}(\mathbf{K}^T\mathbf{b} - \mathbf{m})^T[\mathbf{K}^T(\mathbf{X}^T\mathbf{X})^-\mathbf{K}]^{-1}(\mathbf{K}^T\mathbf{b} - \mathbf{m})\right)$.

See Thm 6.1 of Monahan (2008).

Prove the result.

Full- reduced-model F test

Let $\mathbf{y} = \mathbf{X}\mathbf{b} + \mathbf{e}$, $\mathbf{e} \sim \text{Normal}(\mathbf{0}, \sigma^2 \mathbf{I}_n)$, and let $\mathbf{X} = [\mathbf{X}_0 \ \mathbf{X}_1]$, where $\mathbf{X}_0$ has rank r_0 and $\mathbf{X}$ has rank r. Then we can test

$$H_0\text{: ``the columns in } \mathbf{X}_1 \text{ make no contribution to } \mathbf{y}\text{''}$$

at significance level α with the test

$$\frac{(\text{SSE}_{\text{Red}} - \text{SSE}_{\text{Full}})/(r - r_0)}{\text{SSE}_{\text{Full}}/(n - r)} > F_{r_1, n-r, \alpha},$$

where $\text{SSE}_{\text{Red}} = \mathbf{y}^T(\mathbf{I} - \mathbf{P}_{\mathbf{X}_0})\mathbf{y}$ and $\text{SSE}_{\text{Full}} = \mathbf{y}^T(\mathbf{I}_n - \mathbf{P}_{\mathbf{X}})\mathbf{y}$.

Why is this formulated in a funny way? We don't know if $\mathbf{b}_1$ is estimable.

Exercise: To justify the above test, reparameterize as $\mathbf{y} = \mathbf{W}\mathbf{d} + \mathbf{e}$:

- Let $\mathbf{W} = [\mathbf{W}_0 \ \mathbf{W}_1]$ be full-rank with $\text{Col}\,\mathbf{W}_0 = \text{Col}\,\mathbf{X}_0$ and $\text{Col}\,\mathbf{W}_0 = \text{Col}\,\mathbf{X}_1$.
- Show that LRT for H_0: $\mathbf{d}_1 = \mathbf{0}$ is equivalent to the full- reduced model F test.

Exercise: Consider the treatment effect model

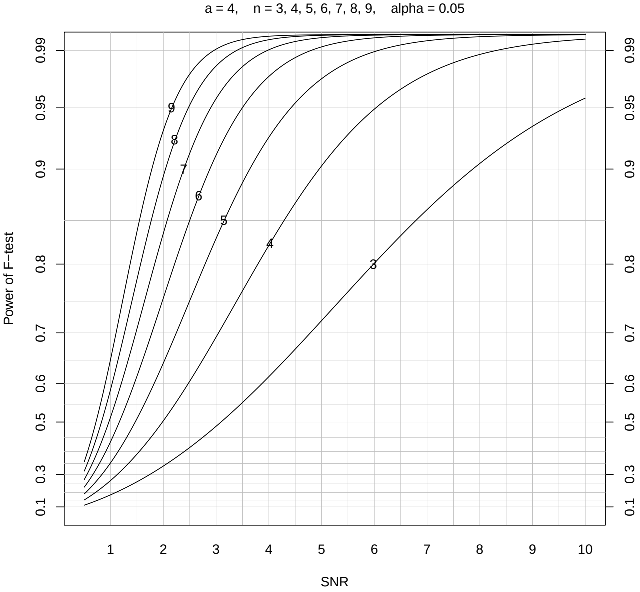
$$Y_{ij} = \mu + \alpha_i + \varepsilon_{ij}, \quad \varepsilon_{ij} \overset{\text{ind}}{\sim} \text{Normal}(0, \sigma^2), \quad j = 1, \dots, n, \ i = 1, \dots, a.$$

One can show that the size α LRT for whether there is any treatment effect is:

$$\frac{n \sum_{i=1}^{a} (\bar{y}_{1.} - \bar{y}_{..})^2 / (a-1)}{\sum_{i=1}^{a} \sum_{j=1}^{n} (y_{ij} - \bar{y}_{i.}) / (na - a)} > F_{a-1, n-a, \alpha}.$$

1. Give the noncentrality parameter as $\phi = n \cdot \text{SNR}$.
2. Let $a = 4$, $\sigma = 1/2$, and suppose $\sum_{i=1}^{a} (\alpha_i - \bar{\alpha})^2 = 1$. Find n such that the power of the F-test is at least 0.90 when using $\alpha = 0.05$.

```
alpha <- 0.05
a <- 4
nn <- 3:9
snr <- seq(1/2,10,length = 500)
powF <- matrix(NA,length(nn),length(snr))

for(i in 1:length(nn)) powF[i,] <- 1-pf(qf(1-alpha,a-1,a*(nn[i]-1)),a-1,a*(nn[i]-1),nn[i]*snr)

plot(NA, xlim = range(snr), ylim = exp(exp(c(.1,.99))),
     yaxt = "n", xaxt = "n", ylab = "Power of F-test",xlab = "SNR")

at <- c(.1,.3,.5,.6,.7,.8,.9,.95,.99)
axis(side = 2, at = exp(exp(at)), labels = at)
axis(side = 4, at = exp(exp(at)), labels = at)
abline(h = exp(exp(c(seq(.1,.95, by = .05),.99))), lwd = .5, col = "gray")

axis(side = 1, at = 1:10, tick = FALSE)
abline(v = seq(1,10, by = 0.5), lwd = .5, col = "gray")

pow_at <- seq(.8,.95,length = length(nn))
for(i in 1:length(nn)){
  lines(exp(exp(powF[i,])) ~ snr)
  snr_pow <- sum(exp(exp(powF[i,])) < exp(exp(pow_at[i])))
  text(x = snr[snr_pow], y = exp(exp(pow_at[i])), label = nn[i])
}

mtext(side = 3, text = paste("a = ",a,",    n = ",paste(nn,collapse=", "),
                             ",    alpha = ",alpha,sep = ""), line = 1)
```



a = 4,   n = 3, 4, 5, 6, 7, 8, 9,   alpha = 0.05
Power of F−test
SNR
0.1
0.3
0.5
0.6
0.7
0.8
0.9
0.95
0.99
1
2
3
4
5
6
7
8
9
10

Simultaneous confidence interval setup

Let $\mathbf{y} = \mathbf{X}\mathbf{b} + \mathbf{e}$, $\mathbf{e} \sim \text{Normal}(\mathbf{0}, \sigma^2 \mathbf{I}_n)$ with $\text{rank}\,\mathbf{X} = r$ and let $\mathbf{c}_1^T\mathbf{b}, \dots, \mathbf{c}_K^T\mathbf{b}$ be estimable contrasts. We seek L_k and U_k, $k = 1, \dots, K$ such that

$$P\left(\bigcap_{k=1}^{K} \{L_k \leq \mathbf{c}_k^T\mathbf{b} \leq U_k\}\right) \geq 1 - \alpha.$$

The intersection event above is called *simultaneous coverage*.

Result (Simul. coverage guarantee for Bonferroni-adj. t- intervals)

Under the simultaneous confidence interval setup, the intervals

$$[L_k, U_k] = \left[\mathbf{c}_k^T\hat{\mathbf{b}} \pm t_{n-r,\alpha/(2K)}\hat{\sigma}\sqrt{\mathbf{c}_k^T(\mathbf{X}^T\mathbf{X})^-\mathbf{c}_k}\right], \quad k = 1, \ldots, K$$

ensure simultaneous coverage with probability at least $1-\alpha$.

Prove the result.

Result (Exact simul. coverage for max-abs intervals)

Under the simultaneous confidence interval setup, the intervals

$$[L_k, U_k] = \left[\mathbf{c}_k^T\hat{\mathbf{b}} \pm |t|^{\vee}_{n-r,\alpha}\hat{\sigma}\sqrt{\mathbf{c}_k^T(\mathbf{X}^T\mathbf{X})^{-}\mathbf{c}_k}\right], \quad k = 1,\dots,K$$

give simultaneous coverage probability $1-\alpha$, *where* $|t|^{\vee}_{n-r}$ *is the dist. such that*

$$\max_{1\leq k\leq K}\left|\frac{\mathbf{c}_k^T\hat{\mathbf{b}} - \mathbf{c}_k^T\mathbf{b}}{\hat{\sigma}\sqrt{\mathbf{c}_k^T(\mathbf{X}^T\mathbf{X})^{-}\mathbf{c}_k}}\right| \sim |t|^{\vee}_{n-r}.$$

The upper α quantile $|t|^{\vee}_{n-r,\alpha}$ of $|t|^{\vee}_{n-r}$ can be found by Monte Carlo simulation.

Tukey's pairwise comparisons of means in one-way ANOVA is exactly this.

Prove the result.

Monte Carlo simulation for $|t|^{\vee}_{n-r}$

1. Note
$$\frac{\mathbf{c}_k^T\hat{\mathbf{b}} - \mathbf{c}_k^T\mathbf{b}}{\hat{\sigma}\sqrt{\mathbf{c}_k^T(\mathbf{X}^T\mathbf{X})^{-}\mathbf{c}_k}} = \frac{Z_k}{\sqrt{W/(n-r)}}, \quad k = 1,\dots,K,$$
where $W = \frac{(n-r)\hat{\sigma}^2}{\sigma^2} \sim \chi^2_{n-r}$ and $(Z_1,\dots,Z_K)^T = \frac{1}{\sigma}\mathbf{D}(\mathbf{C}^T\hat{\mathbf{b}} - \mathbf{C}^T\mathbf{b})$ with
$$\mathbf{C} = [\mathbf{c}_1 \ \dots \ \mathbf{c}_K] \quad \text{and} \quad \mathbf{D}^{-2} = \text{diag}\left(\mathbf{c}_1^T(\mathbf{X}^T\mathbf{X})^{-}\mathbf{c}_1,\dots,\mathbf{c}_K^T(\mathbf{X}^T\mathbf{X})^{-}\mathbf{c}_K\right).$$
2. Since $\mathbf{C}^T\hat{\mathbf{b}} - \mathbf{C}^T\mathbf{b} \sim \text{Normal}(0,\sigma^2\mathbf{C}^T(\mathbf{X}^T\mathbf{X})^{-}\mathbf{C})$, we have
$$(Z_1,\dots,Z_K)^T \sim \text{Normal}(\mathbf{0},\mathbf{D}\mathbf{C}^T(\mathbf{X}^T\mathbf{X})^{-}\mathbf{C}\mathbf{D}).$$
3. To obtain an MC approx. to $|t|^{\vee}_{n-r,\alpha}$, we generate many realizations of
$$\max\{|Z_1|,\dots,|Z_K|\}/\sqrt{W/(n-r)}, \quad W \text{ independent of } (Z_1,\dots,Z_K)^T,$$
and take the upper α quantile.

The covariance matrix $\mathbf{D}\mathbf{C}^T(\mathbf{X}^T\mathbf{X})^{-}\mathbf{C}\mathbf{D}$ may not be positive definite dwbh.

Monahan, J. F. (2008). *A primer on linear models*. CRC Press.