29 Lecture 29: April 7

Last time

- ANCOVA
- Linear contrasts of means

Today

- One last poll on alternative grading path
  - Consider the votes on the three polls (including today's), if we model them as binomial distributed random variables with probability for 'yes' as p_1, p_2 and p_3
  - What is the likelihood function?
  - How do we test the null hypothesis of $H_0 : p_1 = p_2 = p_3$?
  - What do you expect? Why?
- Final exam will be posted on April 30th (per requested by Grace), Due May 11th 11:59pm
- Sampling distribution of linear contrasts
- Multiple comparisons
- Sample size computations for one-way ANOVA
- Lack of fit test

Additional reference

Course notes by Dr. Jason Osborne.

Sampling distribution of linear contrast estimates

For a linear contrast

$$\theta = c_1\mu_1 + \cdots + c_t\mu_t$$

The *best* estimator for a contrast of interest can be obtained by substituting treatment group sample means $\bar{y}_{i+}$ for treatment population means μ_i in the contrast θ:

$$\hat{\theta} = c_1\bar{Y}_{1+} + c_2\bar{Y}_{2+} + \cdots + c_t\bar{Y}_{t+}$$

Example

Recall the binding fraction data that investigate binding fraction for several antibiotics using $n = 20$ bovine serum samples:

Antibiotic	Binding Percentage	Sample mean
Penicillin G	29.6 24.3 28.5 32.0	28.6
Tetracyclin	27.3 32.6 30.8 34.8	31.4
Streptomycin	5.8 6.2 11.0 8.3	7.8
Erythromycin	21.6 17.4 18.3 19	19.1
Chloramphenicol	29.2 32.8 25.0 24.2	27.8

Consider the pairwise contrast comparing penicillin (population) mean to Tetracyclin mean:

$$\theta = \mu_1 - \mu_2 = (1)\mu_1 + (-1)\mu_2 + (0)\mu_3 + (0)\mu_4 + (0)\mu_5$$

Obtain a point estimator of θ.
Answers:

Question: How good is this estimate? In other words, how much uncertainty associated with the estimate?

We want to characterize the sampling distribution of $\hat{\theta}$. According to our model setup, Y_{ij} follow normal distributions. $\hat{\theta}$ is a linear function of Y_{ij}, so that $\hat{\theta}$ follows a normal distribution. We want to derive the mean and variance (the two sufficient statistics) to characterize the normal distribution that $\hat{\theta}$ follows:

$$\hat{\theta} \sim \mathcal{N}(\theta, Var(\hat{\theta}))$$

Derive expressions for the mean and the variance:

Therefore, the standard error:

$$SE(\hat{\theta}) = \sqrt{Var(\hat{\theta})} = \sqrt{\sigma^2 \sum_{j=1}^{t} \frac{c_j^2}{n_j}}$$

which is estimated by

$$\widehat{SE}(\hat{\theta}) = \sqrt{MS[E] \sum_{j=1}^{t} \frac{c_j^2}{n_j}}$$

To test $H_0 : \theta = \theta_0$ (often 0) versus $H_1 : \theta \neq \theta_0$, use t-test:

$$t = \frac{\hat{\theta} - \theta_0}{\widehat{SE}(\hat{\theta})} \overset{H_0}{\sim} t_{N-t}$$

At level α, the critical value for this test is $t(N-t, \alpha/2)$ and $100(1-\alpha)\%$ confidence interval for a contrast $\theta = \sum c_j \mu_j$ is given by

$$\sum c_j \bar{Y}_{j+} \pm t(N-t, \alpha/2) \sqrt{MS[E] \sum \frac{c_j^2}{n_j}}$$

Multiple Comparisons

Let's first review type I and type II errors.

	H_0 is True	H_0 is False
Don't reject H_0	Probability $1-\alpha$	Probability β
Reject H_0	Probability α	Probability $1-\beta$

- Type I error: rejection of a true null hypothesis (false positive).
- Type II error: failure to reject a false null hypothesis (false negative).
- Type I error rate or significance level (α): the probability of rejecting the null hypothesis given the null hypothesis is true.
- Type II error rate (β): the probability of failure to reject the null hypothesis given the null hypothesis is false. $1-\beta$ gives the power of a test.

Now, let's consider all simple (pairwise) contrasts for the binding fraction data with $t = 5$ antibiotic treatments of the form $\theta = \mu_i - \mu_j$.

- We have $\begin{pmatrix} 5 \\ 2 \end{pmatrix} = 10$ tests for significance each at level $\alpha = 0.05$
- what is the probability of committing at least one type I error?

We need to consider the familywise error rate (fwe) when testing k contrasts:

$$fwe = \Pr(\text{at least one type I error})$$

Methods for simultaneous inference for multiple contrasts include

- Bonferroni
- Scheffé
- Tukey

When the number of comparisons is in the hundreds or thousands (e.g. genome-wide association studies), and FWE control is hopeless, more manageable type I error rate is the False Discovery Rate (FDR):

$$FDR = E(\frac{\text{Falsely rejected null hypotheses}}{\text{Number of rejected null hypotheses}})$$

Bonferroni correction

Suppose interest lies in exactly k contrasts. The Bonferroni adjustment to α controls fwe is

$$\alpha_{bonferroni} = \frac{\alpha}{k}$$

and simultaneous 95% confidence intervals for the k contrasts are given by

$$a_1\bar{Y}_{1+} + \cdots + a_t\bar{Y}_{t+} \pm t(\frac{\alpha_{bonferroni}}{2}, \nu)\sqrt{MS[E]\sum\frac{a_j^2}{n_j}}$$

$$b_1\bar{Y}_{1+} + \cdots + b_t\bar{Y}_{t+} \pm t(\frac{\alpha_{bonferroni}}{2}, \nu)\sqrt{MS[E]\sum\frac{b_j^2}{n_j}}$$

$$\cdots$$

$$k_1\bar{Y}_{1+} + \cdots + k_t\bar{Y}_{t+} \pm t(\frac{\alpha_{bonferroni}}{2}, \nu)\sqrt{MS[E]\sum\frac{k_j^2}{n_j}}$$

where ν denotes *df* for error.

Example: for the binding fraction example, consider only pairwise comparisons with Penicillin:

$$\theta_1 = \mu_1 - \mu_2, \theta_2 = \mu_1 - \mu_3, \theta_3 = \mu_1 - \mu_4, \theta_4 = \mu_1 - \mu_5$$

We have $k = 4, \alpha_{bonferroni} = 0.05/k = 0.0125$ and $t(\frac{\alpha_{bonferroni}}{2}, 15) = 2.84$. Substitution leads to

$$t(\frac{\alpha_{bonferroni}}{2}, 15)\sqrt{MS[E]\left(\frac{1^2}{4} + \frac{(-1)^2}{4} + \frac{0^2}{4} + \cdots + \frac{0^2}{4}\right)}$$

$$=2.84\sqrt{(9.05)\frac{2}{4}} = 6.0$$

so that **simultaneous** 95% confidence intervals for θ_1, θ_2, θ_3 and θ_4 take the form

$$\bar{y}_{1+} - \bar{y}_{i+} \pm 6.0$$

Scheffé

Another method to construct **simultaneous** 95% confidence intervals for **ALL** contrasts, use

$$\sum_{j=1}^{t} c_j\bar{y}_{j+} \pm \sqrt{(t-1)(F^*)MS[E]\sum_{j=1}^{t}\frac{c_j^2}{n_j}}$$

where $F^* = F(\alpha, t-1, N-t)$. For a pairwise comparisons of means, μ_j and μ_k, this yields

$$\bar{y}_{j+} - \bar{y}_{k+} \pm \sqrt{(t-1)(F^*)MS[E](1/n_j + 1/n_k)}$$

Using $\alpha = 0.05$, need to specify

- t (from the design)
- F^* (same critical value as for $H_0 : \alpha_i \equiv 0$).
- $MS[E]$ (from the data)
- $\bar{y}_{j+}$, $\bar{y}_{k+}$

- n_j, n_k (from the data)

For binding fraction data,

$$\sqrt{(t-1)(F^*)MS[E](\frac{1}{n_j}+\frac{1}{n_k})} = \sqrt{(5-1)(3.06)9.05(\frac{1}{4}+\frac{1}{4})} = 7.44$$

If any two sample means differ by more than 7.44, they differ significantly.

Tukey

Tukey's method is better than Scheffé's method when making **all pairwise** comparisons in balanced designs ($n = n_1 = n_2 = \cdots = n_t$). It is conservative, controlling the experimentwise error rate, and has a lower type II error rate in these cases than Scheffé. (It is more powerful.)

For simple contrasts of the form

$$\theta = \mu_j - \mu_k$$

to test

$$H_0 : \theta = 0 vs H_1 : \theta \neq 0$$

reject H_0 at level α if

$$|\hat{\theta}| > q(t, N-t, \alpha)\sqrt{\frac{MS[E]}{n}}$$

where $q(t, N-t, \alpha)$ denotes α level studentized range for t means and $N-t$ degrees of freedom, the quantity $q(t, N-t, \alpha)\sqrt{\frac{MS[E]}{n}}$ is referred to as Tukey's honestly significant difference (HSD). The studentized ranges can be calculated using R function `qtukey(1 − α, t, N − t)`.

Sample size computations for one-way ANOVA

Now consider the null hypothesis in a balanced experiment using one-way ANOVA to compare t treatment means and $\alpha = 0.05$:

$$H_0 : \mu_1 = \mu_2 = \cdots = \mu_t = \mu$$

versus the alternative

$$H_a : \mu_i \neq \mu_j \text{ for some } i \neq j$$

Suppose that we intend to use a balanced design. How big does our sample size $n_1 = n_2 = \cdots = n_t = n$ need to be?

The answer depends on lots of things, namely, σ^2 and how many treatment groups t and how much of a difference among the means we hope to be able to detect, and with how big a probability.

Given α, $\mu_1, \ldots, \mu_t$ and σ^2, we can choose n to ensure a power of at least β (i.e. type II error rate) using the noncentral F distribution.

Recall that the critical region for the statistic $F = MS[Trt]/MS[E]$ is everything bigger than $F(\alpha, t-1, N-t) = F^*$.

The power of the F-test conducted using $\alpha = 0.05$ to reject H_0 under this alternative is given by

$$1 - \beta = \Pr(MS[Trt]/MS[E] > F^*; H_1 \text{ is true}). \tag{1}$$

Let $\tau_i = \mu_i - \mu$ for each treatment i so that

$$H_0 : \tau_1 = \tau_2 = \cdots = \tau_t = 0$$

When some H_1 is true and the sample size n is used in each group, it can be shown that the F ratio has the noncentral F distribution with noncentrality parameter

$$\gamma = \sum_{j=1}^{t} n_j \left(\frac{\tau_j}{\sigma}\right)^2 = n \sum_{j=1}^{t} \left(\frac{\tau_j}{\sigma}\right)^2$$

This is the parameterization for the F distribution used in both SAS and R.

One way to obtain an adequate sample size is trial and error. Software packages can be used to get probabilities of the form 1 for various values of n.

Example

Suppose we want to test equal mean binding fractions among antibiotics against the alternative

$$H_1 : \mu_P = \mu + 3, \mu_T = \mu + 3, \mu_S = \mu - 6, \mu_E = \mu, \mu_C = \mu$$

so that

$$\tau_1 = \tau_2 = 3, \tau_3 = -6, \tau_4 = \tau_5 = 0.$$

Assume $\sigma = 3$ (is it arbitrary? any idea of how to guess?) and we need to use $\alpha = \beta = 0.05$. The noncentrality parameter is given by

$$\gamma = n[(\frac{3}{3})^2 + (\frac{3}{3})^2 + (\frac{-6}{3})^2]$$

The $\alpha = 0.05$ critical value for H_0 is given by

$$F^* = F(5-1, 5n-5, 0.05).$$

We need the area to the right of F^* for the noncentral F distribution with degrees of freedom 4 and $5(n-1)$ and noncentrality parameter $\gamma = 6n$ to be greater or equal to the desired power level of $1 - \beta = 0.8$.

We will revisit this example in the lab session on Friday.

Lack-of-fit test

Hiking example: completely randomized experiment involving alpine meadows in the White Mountains of New Hampshire. $N = 20$ lanes of dimension $0.5m \times 1.5m$ randomized to 5 trampling treatments:

i: trt group	x: Number of passes	y_{ij}: Height (cm)			
1	0	20.7	15.9	17.8	17.6
2	25	12.9	13.4	12.7	9.0
3	75	11.8	12.6	11.4	12.1
4	200	7.6	9.5	9.9	9.0
5	500	7.8	9.0	8.5	6.7

Two models for mean plant height:

$$\text{SLR model: } \mu(x) = \beta_0 + \beta_1 x$$
$$\text{one-factor ANOVA model: } \mu_{ij} = \mu + \alpha_i$$

When the t treatments have an interval scale, the SLR model, and all polynomials of degree $p \leqslant t - 2$ (why?), are nested in one-factor ANOVA model with t treatment means.

Answer:

F-ratio for lack-of-fit test

To test for lack-of-fit of a polynomial (reduced) model of degree p, use extra sum-of-squares F-ratio on $t - 1 - p$ and $N - t$ df:

$$F = \frac{SS[\text{lack of fit}]/(t-1-p)}{MS[\text{pure error}]}$$

where

$$MS[\text{pure error}] = MS[E]_{full}$$

and

$$\begin{aligned} SS[\text{lack-of-fit}] &= SS[Trt] - SS[Reg]_{poly} \\ &= SS[E]_{poly} - SS[E]_{full} \end{aligned}$$

What is the $SS[\text{lack of fit}]$ for the meadows data?

Next step: either go with the one-factor ANOVA model or specify some other model, such as quadratic.