Statistics

Week 11: Single Factor Experiments (Chapters 12)

ESD, SUTD

Term 5, 2017

SINGAPORE UNIVERSITY OF TECHNOLOGY AND DESIGN

Established in collaboration with MIT

A summary of ANOVA

Regression

$H_0 : \beta_1 = \beta_2 = \cdots = \beta_k = 0.$

If $k = 1$ (only 1 predictor): simple linear regression, easy to test H_0.

If $k > 1$, use ANOVA.

If H_0 is rejected, need to find a good subset of predictors: r^2_{adj}, standardized data, AIC, ...

Single factor experiment

$H_0 : \mu_1 = \mu_2 = \cdots = \mu_k.$

If $k = 2$ (only 2 groups): independent samples design, easy to test H_0.

If $k > 2$, use ANOVA.

If H_0 is rejected, need to find which groups are different: Bonferroni method, Tukey method, ...

ANOVA table

Obtained in *Excel* under `Data Analysis` $\rightarrow$ `Anova: Single Factor`.

	SS	**df**	**MS**	F	p-value of F
Between	SSA	$k-1$	MSA	MSA/MSE	1−F.DIST(…)
Within	SSE	$N-k$	MSE		
Total	SST	$N-1$			

This is very similar to ANOVA for regression. In regression, SSA and MSA are replaced by SSR and MSR; the df's are replaced by k, $n-k-1$ and $n-1$.

An informal argument

Why does MSA/MSE follow an F distribution?

We argue informally that regardless to whether H_0 is true,

$$\frac{\mathsf{MSE}}{\sigma^2} = \frac{1}{N-k} \sum_{i,j} \left(\frac{y_{ij} - \bar{y}_i}{\sigma} \right)^2 \sim \frac{\chi^2_{N-k}}{N-k}.$$

When H_0 is true, another informal argument gives

$$\frac{\mathsf{MSA}}{\sigma^2} = \frac{1}{k-1} \sum_{i} \left(\frac{\bar{y}_i - \bar{\bar{y}}}{\sigma/\sqrt{n_i}} \right)^2 \sim \frac{\chi^2_{k-1}}{k-1}.$$

Their ratio follows a $F_{k-1,N-k}$ distribution; their values should be comparable, and a smaller ratio supports H_0 while a larger one supports H_1.

Once H_0 is rejected...

If $H_0 : \mu_1 = \mu_2 = \cdots = \mu_k$ is rejected, we wish to find which treatments have different means. Suppose we naïvely tested each pair: for all $i \neq j$, test a new $H_0 : \mu_i = \mu_j$ by checking if

$$|\bar{y}_i - \bar{y}_j| > t_{N-k,\,1-\alpha/2} \sqrt{\frac{s^2}{n_i} + \frac{s^2}{n_j}},$$

where $s^2 =$ MSE (this is based on independent samples design).

The results would be *incorrect*, because there can be many different pairs, amplifying the probability of observing rare events.

This is another example of the *multiple testing problem*.

Multiple testing problem

If we test across many parameters, then by random chance at least one parameter might show a difference; we are likely to observe coincidences not specified in advance.

Although testing for many things at once is fine as an exploratory method, one must use follow-up studies to confirm or refute any patterns that emerge.

Many poorly designed studies fall victim to multiple testing.

Bonferroni method

Because there are $m := \binom{k}{2}$ pairs involved in multiple testing , if the type I error for each test is α (and α is small enough), then the overall error is about $m\alpha$.

To fix this, we insist that the error for *each* test has to be α/m, so that the overall error is about α.

This approach is known as the *Bonferroni method*: we reject $H_0 : \mu_i = \mu_j$ if

$$|\bar{y}_i - \bar{y}_j| > t_{N-k,\, 1-\alpha/(2m)} \sqrt{\frac{s^2}{n_i} + \frac{s^2}{n_j}}.$$

There are other methods to check whether two treatments are different (textbook, Section 12.2); no method is perfect. Bonferroni works well for small k but can be too conservative.

ANOVA – further information

ANOVA is very widely used, but also widely abused.

ANOVA is reasonably robust against violations of its assumptions, namely,

$$Y_{ij} = \mu_i + \epsilon_{ij},$$

where ϵ_{ij} are iid $N(0, \sigma^2)$ random variables.

The *residuals* $e_{ij} := y_{ij} - \bar{y}_i$ can be used to test for normality, via a Q-Q plot.

There are tests to check if the variances are equal, and methods to transform the data if they are not (Section 12.1.3).

Exercise

Check the normality assumption for the '*anorexia*' example.

Use the Bonferroni method to determine which treatment is better.