Statistics - Lecture Two

Charlotte Wickham
wickham@stat.berkeley.edu
http://www.stat.berkeley.edu/~wickham/

Outline

1. Large Sample Theory for MLE
2. Uncertainty
3. Hypothesis Testing

1 Large Sample Theory for Maximum Likelihood Estimates

Define the **Fisher Information** as,

$$I(\theta) = \mathrm{E}\left[\frac{\partial}{\partial\theta}\log f(X|\theta)\right]^2.$$

It can be shown under smoothness conditions that,

$$I(\theta) = -\mathrm{E}\left[\frac{\partial^2}{\partial\theta^2}\log f(X|\theta)\right].$$

The following is a very important result. It says that the mle is asymptotically unbiased, is normally distributed and gives us an estimate for its variance.

Asymptotic distribution of mle

Under certain regularity conditions,

$$\sqrt{nI(\theta_0)}(\hat{\theta} - \theta_0) \to \mathcal{N}(0, 1)$$

Proof

The following gives a outline of the proof. First we take a Taylor expansion of the first derivative of the likelihood around θ_0,

$$\begin{aligned}
0 = l'(\hat{\theta}) &\approx l'(\theta_0) + l''(\theta_0)(\hat{\theta} - \theta_0) \\
(\hat{\theta} - \theta_0) &\approx \frac{-l'(\theta_0)}{l''(\theta_0)} \\
\sqrt{n}(\hat{\theta} - \theta_0) &\approx \frac{-n^{-1/2}l'(\theta_0)}{n^{-1}l''(\theta_0)}
\end{aligned}$$

We need to show the numerator converges in distribution to a $\mathcal{N}(0, I(\theta_0))$ and the denominator converges in probability to $I(\theta_0)$. The result will then follow by application of Slutsky's Theorem. First consider the denominator,

$$n^{-1}l''(\theta_0) = \frac{1}{n}\sum_{i=1}^{n}\frac{\partial^2}{\partial\theta^2}\log f(x_i|\theta_0)$$

which by the law of large numbers converges in probability to,

$$\mathrm{E}\left[\frac{\partial^2}{\partial\theta^2}\log f(X|\theta_0)\right] = -I(\theta_0).$$

Now the numerator. Since $l'(\theta)$ is the sum of iid random variables we can apply the central limit theorem. Hence, the numerator will converge to a normal distribution. We just need to confirm the mean and variance of this normal.

$$\begin{aligned}
\mathrm{E}[n^{-1/2}l'(\theta_0)] &= n^{-1/2}\sum_{i=1}^{n}\mathrm{E}\left[\frac{\partial}{\partial\theta}\log f(X_i|\theta_0)\right] \\
&= 0 \\
\mathrm{Var}[n^{-1/2}l'(\theta_0)] &= \frac{1}{n}\sum_{i=1}^{n}\mathrm{E}\left[\frac{\partial}{\partial\theta}\log f(X_i|\theta_0)\right]^2 \\
&= I(\theta_0)
\end{aligned}$$

If follows then that

$$\begin{aligned}
\mathrm{E}[n^{1/2}(\hat{\theta} - \theta_0)] &\approx 0 \\
\mathrm{Var}[n^{1/2}(\hat{\theta} - \theta_0)] &\approx \frac{I(\theta_0)}{I(\theta_0)^2} \\
&= \frac{1}{I(\theta_0)}
\end{aligned}$$

We can use this result to give confidence intervals for the parameters and preform hypothesis tests. In practice we don't know the value of $I(\theta_0)$ since we don't know θ_0. There are two popular ways to estimate the value. The first simply substitutes the mle estimate for the true value,

$$\widehat{I(\theta_0)} = I(\hat{\theta}).$$

The second, known as the **observed Fisher Information** evaluates the negative of the second derivative at $\hat{\theta}$,

$$\widehat{I(\theta_0)} = -\frac{\partial^2}{\partial\theta^2} \log f(\boldsymbol{X}|\theta)\bigg|_{\hat{\theta}} \cdot \frac{1}{n}$$

This is often computationally the easiest as the Hessian (the matrix of second partial derivatives) is often calculated as part of numerical optimization.

Example - Poisson

In a previous lecture we found that the mle for a sample of size n from a Poisson distribution was

$$\hat{\lambda} = \overline{X}.$$

We can now find the approximate large sample distribution for this estimate. We know,

$$\frac{\partial}{\partial\lambda} \log f(x|\lambda) = \frac{X}{\lambda} - 1$$

We have two ways to proceed from here depending on which formula for $I(\theta)$ is easier to use. Here we will present both routes. The first:

$$\begin{aligned}
I(\lambda) &= -\mathrm{E}\left(\frac{\partial}{\partial\lambda} \log f(x|\lambda)\right)^2 \\
&= -\mathrm{E}\left(\frac{X}{\lambda} - 1\right)^2 \\
&= -\mathrm{E}\left(\frac{X^2}{\lambda^2} - 2\frac{X}{\lambda} + 1\right) \\
&= -\left(\frac{\lambda(\lambda+1)}{\lambda^2} - 2 + 1\right) \\
&= \frac{1}{\lambda}.
\end{aligned}$$

The second:

$$\begin{aligned} I(\lambda) &= -\mathrm{E}\left(\frac{\partial^2}{\partial\lambda^2}\log f(x|\lambda)\right) \\ &= -\mathrm{E}\left(\frac{\partial}{\partial\lambda}\frac{X}{\lambda} - 1\right) \\ &= -\mathrm{E}\left(-\frac{X}{\lambda^2}\right) \\ &= \frac{1}{\lambda}. \end{aligned}$$

So, asymptotically $\hat{\lambda}$ is distributed $\mathcal{N}(\lambda, \frac{\lambda}{n})$.

Example - Confidence intervals for mle

Suppose we wish to find a 95% confidence interval for a population parameter, θ_0, that has been estimated by maximum likelihood. We rely on the large sample properties of the mle that,

$$\sqrt{nI(\theta_0)}(\hat{\theta} - \theta_0) \dot{\sim} \mathcal{N}(0, 1).$$

We find,

$$P\left(-z(\alpha/2) \leq \sqrt{nI(\theta_0)}(\hat{\theta} - \theta_0) \leq z(\alpha/2)\right) = 1 - \alpha.$$

Then rearrange to find an interval of the form,

$$\hat{\theta} \pm z(\alpha/2)\sqrt{\frac{1}{nI(\theta_0)}}.$$

In practice an estimate of $I(\theta_0)$, such as those discussed above, is used.

So, for example, using the poisson example above we find a 95% confidence interval for λ is,

$$\overline{X} \pm z(\alpha/2)\sqrt{\frac{\overline{X}}{n}}$$

Multivariate case

The results of the previous section extend to the case where θ_0 is a vector. Except now the asymptotic distribution of $\hat{\theta}$ is multivariate normal,

$$\hat{\theta} \dot{\sim} \mathcal{N}_p\left(\theta_0, \frac{1}{n} I(\theta_0)^{-1}\right),$$

where $I(\theta)$ is a matrix with the ij component,

$$\mathrm{E}\left(\frac{\partial}{\partial\theta_i}\log f(X|\theta)\frac{\partial}{\partial\theta_j}\log f(X|\theta)\right) = -\mathrm{E}\left(\frac{\partial^2}{\partial\theta_i\partial\theta_j}\log f(X|\theta)\right)$$

These results don't hold if the true parameter is on the boundary of parameter space.

Functions of parameters

The delta method can be directly applied to maximum likelihood estimates. Since we know,

$$\sqrt{n}(\hat{\theta} - \theta_0) \to \mathcal{N}(0, \frac{1}{I(\theta_0)}).$$

We can use the delta theorem to find that,

$$\sqrt{n}(g(\hat{\theta}) - g(\theta_0)) \to \mathcal{N}(0, g'(\theta)^2 I(\theta_0)^{-1}).$$

This extends easily to the multivariate case as well.

2 Uncertainty

Cramer Rao

The Cramer Rao Inequality gives us a lower bound on the variance of an unbiased estimator.

Cramer Rao Inequality

Let $X_1, \ldots, X_n$ be iid with density function $f(x|\theta)$. Let $T = t(X_1, \ldots, X_n)$ be an unbiased estimate of θ. Then under smoothness conditions on $f(x|\theta)$,

$$\text{Var}(T) \geq \frac{1}{nI(\theta)}$$

Proof

The following inequality forms the basis for this proof and follows from the fact that the correlation between two random variables must be less than 1 in absolute value,

$$\text{Cov}(T, W)^2 \leq \text{Var}(T)\text{Var}(W).$$

We let,

$$\begin{aligned} W &= \sum_{i=1}^{n} \frac{\partial}{\partial\theta} \log f(X_i|\theta) \\ &= \sum_{i=1}^{n} \frac{\frac{\partial}{\partial\theta} f(X_i|\theta)}{f(X_i|\theta)}. \end{aligned}$$

We will now find the expected value and variance for W. We will need the following fact a couple of times in this proof,

$$\sum_{i=1}^{n} \frac{\frac{\partial}{\partial\theta} f(x_i|\theta)}{f(x_i|\theta)} \prod_{j=1}^{n} f(x_j|\theta) = \frac{\partial}{\partial\theta} \prod_{i=1}^{n} f(x_i|\theta)$$

which follows by application of the chain rule of differentiation. First, we show $\mathrm{E}(W) = 0$,

$$\begin{aligned}
\mathrm{E}(W) &= \int \dots \int \sum_{i=1}^{n} \frac{\frac{\partial}{\partial\theta} f(x_i|\theta)}{f(x_i|\theta)} \prod_{j=1}^{n} f(x_j|\theta) dx_j \\
&= \int \dots \int \frac{\partial}{\partial\theta} \prod_{i=1}^{n} f(x_i|\theta) dx_i \\
&= \frac{\partial}{\partial\theta} \int \dots \int \prod_{i=1}^{n} f(x_i|\theta) dx_i \\
&= \frac{\partial}{\partial\theta} 1 \\
&= 0
\end{aligned}$$

Now we find the variance of W,

$$\begin{aligned}
\mathrm{Var}(W) &= E(W^2) - (E(W))^2 \\
&= E(W^2) \\
&= E\left[\left(\sum_{i=1}^{n} \frac{\partial}{\partial\theta} \log f(X_i|\theta)\right)^2\right] \\
&= \sum_{i=1}^{n} E\left[\frac{\partial}{\partial\theta} \log f(X_i|\theta)\right]^2 \quad \text{since } X_i \text{ are independent} \\
&= \sum_{i=1}^{n} I(\theta) \\
&= nI(\theta)
\end{aligned}$$

To complete the proof it remains only to show $\mathrm{Cov}(T, W) = 1$.

$$\begin{aligned}
\mathrm{Cov}(W, T) &= \mathrm{E}(WT) \quad \text{since } \mathrm{E}(W) = 0 \\
&= \int \dots \int t(x_1, ..., x_n) \sum_{i=1}^{n} \frac{\frac{\partial}{\partial\theta} f(x_i|\theta)}{f(x_i|\theta)} \prod_{j=1}^{n} f(x_j|\theta) dx_j \\
&= \int \dots \int t(x_1, ..., x_n) \frac{\partial}{\partial\theta} \prod_{i=1}^{n} f(x_i|\theta) dx_i \\
&= \frac{\partial}{\partial\theta} \int \dots \int t(x_1, ..., x_n) \prod_{i=1}^{n} f(x_i|\theta) dx_i \\
&= \frac{\partial}{\partial\theta} \mathrm{E}(T) \\
&= \frac{\partial}{\partial\theta} \theta \quad \text{since } T \text{ is unbiased} \\
&= 1
\end{aligned}$$

Note the mle has asymptotically minimum variance so we say it is asymptotically efficient. Also note the theorem only postulates about **unbiased** estimators.

3 Hypothesis Testing

The basic idea is that we want to make a decision about the underlying process generating our data. We start by setting up two mutually exclusive hypotheses. One we call the null hypothesis. This is generally the simpler of the two and often gives the simplest explanation of the observed data. The other hypothesis is known as the alternative hypothesis.

Simple Hypotheses

We describe the hypotheses as simple when both null and alternative completely specify the probability distribution.

Example of simple hypotheses

We have two coins and know one has probability 0.5 of being a head the other has 0.7 of being a head. We choose one coin and toss it ten times. Let X = the number of heads.

H_0 : Null: The coin we chose has probability 0.5 of being a head. $X \sim$ Bin(10,0.5)

H_1 : Alternative: The coin we chose has probability 0.7 of being a head, $X \sim$ Bin(10,0.7)

Example of non simple hypotheses

We have a coin and toss it ten times. Is it fair? Let X = the number of heads.

H_0 : Null: The coin has probability 0.5 of being a head. $X \sim$ Bin(10,0.5)

H_1 : Alternative: The does not have probability 0.5 of being a head. $X \sim$ Bin(10,?)

Error Types

Type I

A type I error occurs when we reject the null hypothesis when it is in fact true. This error rate is generally denoted α.

Type II

A type II error occurs when we fail to reject the null hypothesis when it is in fact false. This error rate is generally denoted β.

The probability of correctly rejecting the null when it is indeed false is called the **power** of the test and is $(1 - \beta)$.

Neyman-Pearson Approach

The Neyman-Pearson approach to hypothesis tests fixes α and then tries to minimize β. This means we are looking for the most power for a specified type I error rate.

Test statistic

We generally start by formulating a test statistic $T(x)$ (that is a function of the data). We compare this to some critical value, c, that defines a rejection and acceptance region. We choose the critical value to control the Type I error (α) under the null hypothesis. So, in general c is chosen to satisfy

$$P(T(x) > c | H_0 \text{ is true}) = \alpha$$

	Do not reject H_0	Reject H_0
H_0 is true		Type I error
H_0 is false	Type II error	Power

Table 1: Types of Error

But how do we choose $T(x)$?

The Likelihood Ratio Test

The likelihood ratio test rejects the null hypothesis if the likelihood ratio is small. We reject H_0 if,

$$\frac{f_0(X)}{f_1(X)} < c$$

Intuitively if the null is true we would expect the likelihood of the observed data under the null to be larger than under the alternative so that the ratio > 1. Alternatively, if the ratio is small the likelihood of the observed data under the alternative is larger than the likelihood under the null. This would give some evidence the null should be rejected. The Neyman Pearson Lemma gives us a good reason to use the likelihood ratio test.

Neyman Pearson Lemma

Suppose that H_0 and H_1 are simple hypotheses and that the test rejects H_0 whenever the likelihood ratio is less than c and significance level α. Then any other test for which the significance level is less than or equal to α has power less than or equal to that of the likelihood ratio test.

Proof

Let $f(x)$ be the probability density function (or frequency function) of the observations. The hypotheses can be stated as: $H_0 : f(x) = f_0(x)$ versus $H_A : f(x) = f_A(x)$. The test of these hypothesis is equivalent to using a decision function $d(x)$, where $d(x) = 0$ means we accept H_0 and $d(x) = 1$ means we reject H_0. $d(X)$ is a Bernoulli random variable with $P_0(d(X) = 1) = \mathrm{E}_0(d(X)) = \alpha$, the significance level, and $P_A(d(X) = 0) = \mathrm{E}_A(d(X)) = 1 - \beta$, the power. Here E_0 (P_0) means the expectation (probability) under the hypothesis H_0.

Let $d(X)$ correspond to the likelihood ratio test. Then $d(X) = 1$ if $f_0(X) \leq cf_A(X)$. Let $d^*(x)$ be another decision function of another test satisfying $\mathrm{E}_0(d^*(x)) \leq \mathrm{E}_0(d(X)) = \alpha$. We will show that $\mathrm{E}_A(d^*(X)) \leq \mathrm{E}_A(d(X))$.

Consider the following inequality

$$d^*(x)[cf_A(x) - f_0(x)] \leq d(x)[cf_A(x) - f_0(x)]$$

If $cf_A(x) - f_0(x) \geq 0$ then by the definition of the test $d(x) = 1$. Note that $d^*(x)$ can only be zero or one so that the inequality holds. Similarly if $cf_A(x) - f_0(x) \leq 0$, $d(x) = 0$ and the inequality still

holds. Integrating (or summing) both sides with respect to x gives,

$$c\mathrm{E}_A(d^*(X)) - \mathrm{E}_0(d^*(X)) \leq c\mathrm{E}_A(d(X)) - \mathrm{E}_0(d(X)).$$

Rearranging gives,

$$\mathrm{E}_0(d(X)) - \mathrm{E}_0(d^*(X)) \leq c[\mathrm{E}_A(d^*(X)) - \mathrm{E}_A(d(X))].$$

The left hand side is nonnegative by assumption and the result follows directly.

We typically use the Neyman-Pearson Lemma in the following way:

- Write down the likelihood ratio
- Observe that extreme values of a test statistic correspond to small values of the likelihood ratio
- Use distributional theory on the test statistic to find c to obtain the required significance level

Example - Normal means

We have a sample $X_1, ..., X_n$ from Normal distribution with mean μ and known variance σ^2. Our hypotheses are:

$$H_0 : \mu = \mu_0$$
$$H_A : \mu = \mu_1$$

The likelihood ratio is of the form,

$$\frac{f_0(\boldsymbol{X})}{f_1(\boldsymbol{X})} = \frac{\exp\left[\sum_{i=1}^n -\frac{1}{2\sigma^2}(X_i - \mu_0)^2\right]}{\exp\left[\sum_{i=1}^n -\frac{1}{2\sigma^2}(X_i - \mu_1)^2\right]}$$

where H_0 is rejected for small values. A small value of this ratio corresponds to a small value of $\sum_{i=1}^n (X_i - \mu_1)^2 - \sum_{i=1}^n (X_i - \mu_0)$. Expanding this gives,

$$2n\overline{X}(\mu_0 - \mu_1) + n\mu_1^2 - n\mu_0^2.$$

We see that if $\mu_1 > \mu_0$ then small values will correspond to large values of $\overline{X}$ and conversely if $\mu_1 < \mu_0$ small values will correspond to small values of $\overline{X}$. We will assume $\mu_1 > \mu_0$. So we will reject the null if $\overline{X} > c$ where c satisfies $P(\overline{X} > c|H_0) = \alpha$. Under the null hypothesis $\sqrt{n}(\overline{X} - \mu_0)/\sigma \sim \mathcal{N}(0, 1)$.

$$P(\overline{X} > c) = P\left(\frac{\overline{X} - \mu_0}{\sigma/\sqrt{n}} > \frac{c - \mu_0}{\sigma/\sqrt{n}}\right).$$

So we can find c by solving,

$$\frac{c - \mu_0}{\sigma/\sqrt{n}} = z(\alpha)$$

in order to find the rejection region for significance level α.

Example - Normal means again

Again consider a sample $X_1, ..., X_n$ from Normal distribution with mean μ and known variance σ^2. It is very common to want to test if $\mu = 0$. Our hypotheses are

$$H_0 : \mu = 0$$
$$H_A : \mu \neq 0.$$

Here our hypotheses are not simple so the Neyman Pearson Lemma does not help us. We can get halfway with the idea of uniformly most powerful.

Uniformly most powerful tests

A test is considered **uniformly most powerful** if for every simple alternative the test is most powerful.

Example - back to Normal means

If our hypotheses in the previous example were,

$$H_0 : \mu = 0$$
$$H_A : \mu > 0,$$

then the test devised would be uniformly most powerful. The test did not depend on μ_1 so that it is most powerful for every alternative and the test is the same for every alternative.

There is no uniformly most powerful test for the hypotheses, $H_0 : \mu = 0$, $H_A : \mu \neq 0$ since the test in the previous example only works for $\mu_1 > \mu_0$.

It turns out the tests based on the likelihood ratio generally work pretty well even when the hypotheses aren't simple. They are often not optimal but in most of these situations no optimal test exists. The generalized likelihood tests extends the likelihood test to situations with non simple hypotheses.

Generalized Likelihood Ratio Test

Assume we have a sample $\boldsymbol{X} = (X_1, \ldots, X_n)$ with joint density function $f(\boldsymbol{X}|\theta)$. Let H_0 specify that $\theta \in \omega_0$ and H_1 specify $\theta \in \omega_1$ where ω_1 is disjoint from ω_0 and let $\Omega = \omega_0 \cup \omega_1$. The generalized likelihood ratio is defined as,

$$\Lambda* = \frac{\max_{\theta \in \omega_0}[lik(\theta)]}{\max_{\theta \in \Omega}[lik(\theta)]}.$$

Example - Normal means

Back to the sample $X_1, ..., X_n$ from Normal with mean μ and known variance σ^2. For the hypotheses,

$$H_0 : \mu = 0$$
$$H_A : \mu \neq 0,$$

the generalized likelihood ratio is of the form,

$$\begin{aligned}\Lambda* &= \frac{lik(0)}{\max_{\mu \in (-\infty, \infty)}[lik(\mu)]} \\ &= \frac{\exp\left[\sum_{i=1}^n -\frac{1}{2\sigma^2} X_i^2\right]}{\exp\left[\sum_{i=1}^n -\frac{1}{2\sigma^2}(X_i - \overline{X})^2\right]}.\end{aligned}$$

Small values of λ_* correspond to small values of $\sum_{i=1}^n (X_i - \overline{X})^2 - \sum_{i=1}^n X_i^2$. This is equivalent to small values of $-\overline{X}^2$ or $|\overline{X}|$ being large. Under the null hypothesis $\sqrt{n}(\overline{X})/\sigma \sim \mathcal{N}(0, 1)$. So,

$$P(|\overline{X}| > c) = P\left(\left|\frac{\overline{X}}{\sigma/\sqrt{n}}\right| > \left|\frac{c}{\sigma/\sqrt{n}}\right|\right).$$

So we can find c by solving,

$$\frac{c}{\sigma/\sqrt{n}} = z(\alpha/2).$$

Distribution of the generalized likelihood

Under smoothness conditions on the probability density (or frequency function) involved, the null distribution of $-2\log\Lambda*$ tends to a chi-square distribution with degrees of freedom $\dim\Omega - \dim\omega_0$ as the sample size tends to infinity.

Example - Normal means

The above result says for our previous example we expect,

$$\begin{aligned}
-2\log\Lambda* &= -2\log\left(\frac{\exp\left[\sum_{i=1}^{n} -\frac{1}{2\sigma^2}X_i^2\right]}{\exp\left[\sum_{i=1}^{n} -\frac{1}{2\sigma^2}(X_i-\overline{X})^2\right]}\right) \\
&= -2\left(\left[\sum_{i=1}^{n} -\frac{1}{2\sigma^2}X_i^2\right] - \left[\sum_{i=1}^{n} -\frac{1}{2\sigma^2}(X_i-\overline{X})^2\right]\right) \\
&= \frac{n\overline{X}^2}{\sigma^2},
\end{aligned}$$

to be asymptotically distributed χ_1^2. Here, this is actually an exact result since $\sqrt{n}\overline{X}/\sigma$ is distributed $\mathcal{N}(0,1)$.

p-value

A ***p*-value** is a convenient way of summarizing the evidence against the null hypothesis. There two ways of looking at a p-value. This first is that it is the smallest significance level at which the null hypothesis would be rejected. The second is that the p-value is the probability under the null hypothesis of a result as or more extreme than that observed. The smaller the p-value the stronger the evidence against the null hypothesis.

Duality of hypothesis tests and confidence intervals

There is a duality between hypothesis tests and confidence intervals. We can use confidence intervals to define an acceptance region for a test and equally invert a test to create a confidence interval. The following two statements define this formally.

Let θ be a parameter of a probability distribution and Θ be all possible values of θ. Let $\boldsymbol{X}$ be the random variables constituting the data.

Hypothesis test to confidence interval

Suppose that for every value θ_0 in Θ there is a test at level α of the hypothesis $H_0 : \theta = \theta_0$. Denote the acceptance region of the test by $A(\theta_0)$. Then the set

$$C(\boldsymbol{X}) = \{\theta : \boldsymbol{X} \in A(\theta)\}$$

is a $100(1-\alpha)\%$ confidence region for θ.

Confidence interval to hypothesis test

Suppose $C(\boldsymbol{X})$ is a $100(1-\alpha)\%$ confidence region for θ Then an acceptance region for a test at level α of the hypothesis $H_0 : \theta = \theta_0$ is

$$A(\theta_0) = \{\boldsymbol{X} | \theta_0 \in C(\boldsymbol{X})\}$$

4 Multiple Comparisons

Here we just introduce the problem of multiple comparisons. In 215B you will be exposed to some possible solutions. The problem arises when we are making many hypothesis tests (or equivalently constructing many confidence intervals). Individually they will have a test wise type I error rate of α but if we do many tests we will expect our experiment wise error rate to be larger. Imagine we preform J tests. Then,

$$\begin{aligned} P(\text{ Reject at least one hypothesis } | H_0 \text{ is true }) &= 1 - P(\text{ Accept all hypotheses } | H_0 \text{ is true}) \\ &= 1 - (1-\alpha)^J. \end{aligned}$$

Some possible solutions are Scheffé's method, Bonferroni's method and Tukey's honest method. False Discovery Rate is alternative way of approaching the problem.