HOS 6236 Molecular Marker Assisted Plant Breeding
Fall 2017

These slides attempt to answer the questions:

Why do we need multiple comparisons?

How do we perform a multiple comparison?

How do we interpret the results?

Hypothesis testing and P-value

ANOVA is a statistical method that allow us to test the null hypothesis of no difference in treatments means

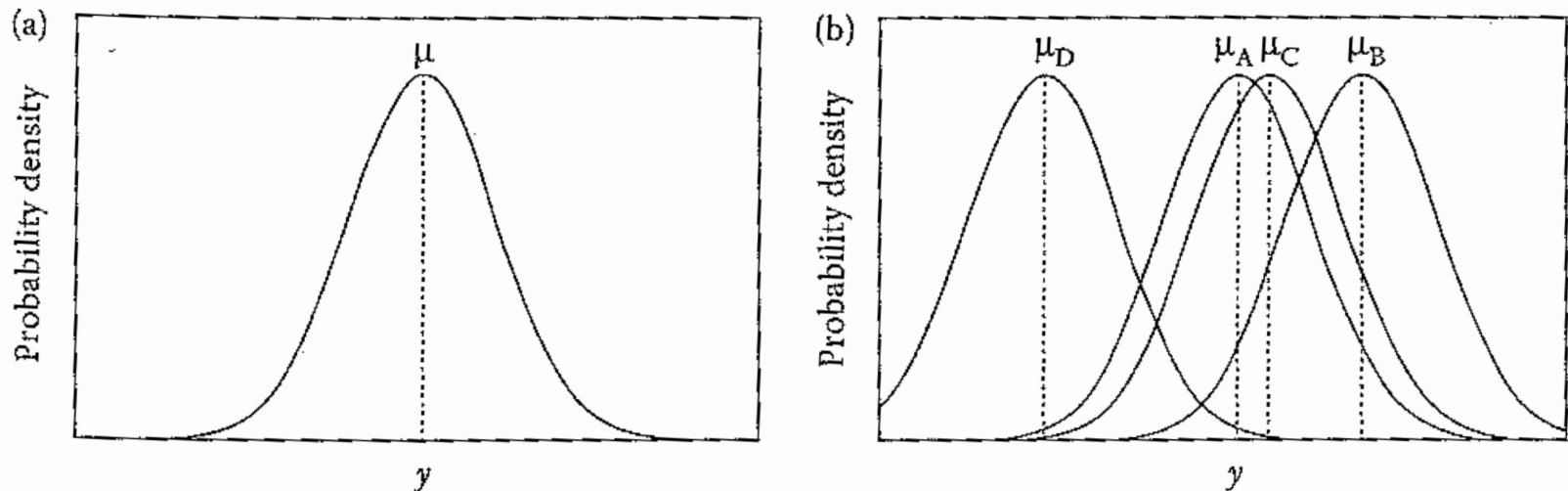$$H_0 : \mu_1 = \mu_2 = \mu_3 = \ldots = \mu_t$$

Figure 4.2 Welham et al 2014

P-value

The P-value indicate the probability that the values observed belong to the null hypothesis when the null hypothesis is assumed to be true.

Treatment Comparisons

If results from ANOVA show significance differences among treatments, then the next step is to find out what treatments are different.

Because, if significant, the F-test tells you there are differences, but it does not tell you where these differences lie.

There is many ways to do this, but usually misuse and/or misinterpreted.

Ideally, you should plan your comparisons at the design stage of your experiment. And no decide when data is being analyzed

Treatment Comparisons

These comparisons should:

- Have a biological meaning

- Hopefully be independent from each other

- Not more in number than the number of degree of freedom for the treatments

Treatment Comparisons

- You can either do contrast or planned comparisons, where number of comparisons are equal to the number of degree of freedom for the treatment, or (**These are not discussed here**)

- Use multiple comparisons (more comparisons that the degree of freedom). Here many methods to correct for the lack of degree of freedom

The most common and simple comparison is pairwise comparisons:

Treatment Comparisons

Most common and Simplest case: the aim is to test a null hypothesis of equality between a pair of treatment means:

$$H_0 : \mu_i = \mu_j$$

$$H_a : \mu_i \neq \mu_j$$

For what we could use
$$t = \frac{\hat{\mu}_i - \hat{\mu}_j}{SE(\hat{\mu}_i - \hat{\mu}_j)} = \frac{\hat{\mu}_i - \hat{\mu}_j}{SED}$$

With $SED = \sqrt{s^2 * (\frac{1}{r_1} + \frac{1}{r_2})}$ if the number of replicates (r_1 and r_2) for the two treatments is different, or $SED = \sqrt{2 \frac{s^2}{r}}$ if they are equal.

Tested against a $t_{(rdf,\alpha/2)}$ = t-distribution with df equal to the residual df from the ANOVA table and $\alpha = 0.05$. s^2 comes from the square root of the Residual Mean Square (MSR) from the ANOVA.

If t-statistics is larger than $t_{(rdf,\alpha/2)}$ (Critical value) then reject H_0.

Treatment Comparisons

Most common and Simplest case: the aim is to test a null hypothesis of equality between a pair of treatment means: $H_0 : \mu_i = \mu_j$

For what we could use

$$t = \frac{\hat{\mu}_i - \hat{\mu}_j}{SE(\hat{\mu}_i - \hat{\mu}_j)} = \frac{\hat{\mu}_i - \hat{\mu}_j}{SED}$$

$H_a : \mu_i \neq \mu_j$

With $SED = \sqrt{s^2 * (\frac{1}{r_1} + \frac{1}{r_2})}$ if the number of replicates (r_1 and r_2) for the two treatments is different, or $SED = \sqrt{2\frac{s^2}{r}}$ if they are equal.

Tested against a $t_{(rdf, \alpha/2)}$ = t-distribution with df equal to the residual df from the ANOVA table and α_s=0.05.

 We test against a specific significance level, α_s=0.05 in this case.

Known as the "**comparison-wise significance level**"

Treatment Comparisons

α_s = is also **the Type I error** – probability of rejecting the null hypothesis when in reality is true

The **Type I error** applies to each individual hypothesis test done

The more test we do the higher the probability of a false-positive result – the problem of **multiple testing**

Treatment Comparisons

So if we make 15 independent test with α_s =0.05, then

$$\text{prob(at least one false positive)} = \alpha_f = 1 - (0.95)^{15} = 0.537$$

Meaning that there is 53.7% chance of one or more false-positive result.

The problem is that in the context of comparisons from a single experiment, multiple test are not independent as the SED (denominator in the t-statistics) are based on the same MSR= s^2 (residual mean square). So we use different approaches to deal with this:

Bonferroni Correction

Adjust the α_s (**the Type I error)** downwards.

Adjustment is based on number of comparisons (*m*) as:

$$\alpha_s{}^* = \alpha_s/m$$

When we use $\alpha_s{}^*$ instead of α_s the test statistic required to obtain a significant result for any individual comparison increases.

Bonferroni Correction

Main disadvantage is that as the number of comparisons (m) increases, the absolute treatment differences often have to become very large to exceed the Bonferroni critical value.

$$\alpha_s{}^* = \alpha_s / m$$

False Discovery Rate (FDR)

It does not attempt to control the experiment-wise error rate, but instead seeks to quantify the expected proportion of type I errors within the sets of rejected hypotheses.

So a FDR of 0.07 means that 7% of the treatment differences that have been found statistically significant are expected to be false-positive results.

Two forms of calculating FDR (we only show one here!). **Both methods are applied after the test results have been obtained.**

False Discovery Rate

Method 1: fixes the significant level α_s for individual test and then calculates the observed FDR

$$FDR = (\alpha_s * m)/s$$

Where s is the number of comparisons with statistical significant result.

Example: 200 comparisons, 24 with statistical significance difference, testing at an $0.05 = \alpha_s$. **FDR=0.417** ~41.7% of the 24 comparisons are false-positive (~10 comparisons).

Least significant Difference (LSD)

For a comparison of treatments is defined as the smallest absolute difference that would result in rejection of the null hypothesis $H_0 : \mu_i = \mu_j$ at a significance level α_s

$$\text{LSD} = t_{[\alpha_s/2],\text{rdf}} \times \text{SED}$$

The unprotected LSD approach to multiple comparisons reject the null hypothesis for any pair of treatments whose absolute difference exceed the LSD, regardless of the results of the F-test on the ANOVA. e.g.:

$$\left| \hat{\mu}_i - \hat{\mu}_j \right| \geq LSD$$

The protected version **required that the F-test (ANOVA) to be significant before calculating the LSD**

-

Least significant Difference (LSD)

For a comparison of treatments is defined as the smallest absolute difference that would result in rejection of the null hypothesis $H_0 : \mu_i = \mu_j$ at a significance level α_s

$$LSD = t_{\alpha_s/2, df_r} \times SED \qquad\qquad SED = \sqrt{s^2 * (\frac{1}{r_1} + \frac{1}{r_2})} \quad \text{or} \quad SED = \sqrt{2 \frac{s^2}{r}}$$

r is the number of replicates and s^2 is the MSE or MSR from the ANOVA table.

Reject the null hypothesis (no difference between the pair of treatments) for any pair of treatments whose absolute difference exceed the LSD. e.g.:
-
$$\left| \hat{\mu}_i - \hat{\mu}_j \right| \geq LSD$$

Randomized Complete Block Design (RCBD)

Three new varieties of wheat (V2, V3 and V4) are compared to a standard commercial variety (V1) using a RCBD with three blocks. Is the yield of the new varieties higher than the commercial variety?

b-1 where b is number of reps (n)

Source of Variation	d.f.	Sum Squares	Mean Squares	F-Ratio	P-value
block (b)	2	9.78	4.89	12.22	0.0076
Treatment (t)	3	6.63	2.21	5.52	0.0367
Residual	6	2.40	0.40		
Total	11	18.81			

New varieties **yield** more than standard variety as p-value is **SMALLER** than 0.05

rdf= Residual degree of freedom S^2 = MSE = MSR

Least significant Difference (LSD)

Example

$LSD = t_{[\alpha_s/2], rdf} \times SED = t_{[0.05/2], rdf} \times SED$,   if $\alpha_s = 0.05$ is chosen

$t_{0.025, 6}$ = [in excel "t.inv.2T(0.05,6)"]=2.447

$$SED = \sqrt{2\frac{s^2}{r}} = \sqrt{2\frac{0.40}{3}} = 0.516$$

LSD=2.447 x 0.516 = 1.263

Any difference greater than 1.263 is significant

Means

V1=6.50;   V2=7.60;   V3=6.60;   V4=8.30

Tukey's Honestly Significant Difference test (HSD) or Tukey's test or Tukey's Studentized Range Test

One of the criticisms of LSD is that if the comparison-wise error is fixed, say 0.05, the experiment-wise error rate increase as the number of treatments increase.

So, Tukey's test controls the experiment-wise error rate.

Instead of the Least Significance Difference (LSD) a Minimum Significance Difference (MSD) is calculated instead as:

$$MSD = Q * \sqrt{\frac{s^2}{r}}$$

Tukey's Honestly Significant Difference test (HSD) or Tukey's test or Tukey's Studentized Range Test

$$MSD = Q * \sqrt{\frac{s^2}{r}}$$

MSD increases as the number of treatment increases, as there are more possible comparisons.

Where Q is the Studentize range (from tables or internet)

Table to calculate Q

The Studentized range upper quantiles q(k, df; 0.05)

df	k-> 2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20
1	17.969	26.976	32.819	37.082	40.408	43.119	45.397	47.357	49.071	50.592	51.957	53.194	54.323	55.361	56.320	57.212	58.044	58.824	59.558
2	6.085	8.331	9.798	10.881	11.734	12.435	13.027	13.539	13.988	14.389	14.749	15.076	15.375	15.650	15.905	16.143	16.365	16.573	16.769
3	4.501	5.910	6.825	7.502	8.037	8.478	8.852	9.177	9.462	9.717	9.946	10.155	10.346	10.522	10.686	10.838	10.980	11.114	11.240
4	3.926	5.040	5.757	6.287	6.706	7.053	7.347	7.602	7.826	8.027	8.208	8.373	8.524	8.664	8.793	8.914	9.027	9.133	9.233
5	3.635	4.602	5.218	5.673	6.033	6.330	6.582	6.801	6.995	7.167	7.323	7.466	7.596	7.716	7.828	7.932	8.030	8.122	8.208
6	3.460	4.339	4.896	5.305	5.628	5.895	6.122	6.319	6.493	6.649	6.789	6.917	7.034	7.143	7.244	7.338	7.426	7.508	7.586
7	3.344	4.165	4.681	5.060	5.359	5.606	5.815	5.997	6.158	6.302	6.431	6.550	6.658	6.759	6.852	6.939	7.020	7.097	7.169
8	3.261	4.041	4.529	4.886	5.167	5.399	5.596	5.767	5.918	6.053	6.175	6.287	6.389	6.483	6.571	6.653	6.729	6.801	6.869
9	3.199	3.948	4.415	4.755	5.024	5.244	5.432	5.595	5.738	5.867	5.983	6.089	6.186	6.276	6.359	6.437	6.510	6.579	6.643
10	3.151	3.877	4.327	4.654	4.912	5.124	5.304	5.460	5.598	5.722	5.833	5.935	6.028	6.114	6.194	6.269	6.339	6.405	6.467
11	3.113	3.820	4.256	4.574	4.823	5.028	5.202	5.353	5.486	5.605	5.713	5.811	5.901	5.984	6.062	6.134	6.202	6.265	6.325
12	3.081	3.773	4.199	4.508	4.750	4.950	5.119	5.265	5.395	5.510	5.615	5.710	5.797	5.878	5.953	6.023	6.089	6.151	6.209
13	3.055	3.734	4.151	4.453	4.690	4.884	5.049	5.192	5.318	5.431	5.533	5.625	5.711	5.789	5.862	5.931	5.995	6.055	6.112
14	3.033	3.701	4.111	4.407	4.639	4.829	4.990	5.130	5.253	5.364	5.463	5.554	5.637	5.714	5.785	5.852	5.915	5.973	6.029
15	3.014	3.673	4.076	4.367	4.595	4.782	4.940	5.077	5.198	5.306	5.403	5.492	5.574	5.649	5.719	5.785	5.846	5.904	5.958
16	2.998	3.649	4.046	4.333	4.557	4.741	4.896	5.031	5.150	5.256	5.352	5.439	5.519	5.593	5.662	5.726	5.786	5.843	5.896
17	2.984	3.628	4.020	4.303	4.524	4.705	4.858	4.991	5.108	5.212	5.306	5.392	5.471	5.544	5.612	5.675	5.734	5.790	5.842
18	2.971	3.609	3.997	4.276	4.494	4.673	4.824	4.955	5.071	5.173	5.266	5.351	5.429	5.501	5.567	5.629	5.688	5.743	5.794
19	2.960	3.593	3.977	4.253	4.468	4.645	4.794	4.924	5.037	5.139	5.231	5.314	5.391	5.462	5.528	5.589	5.647	5.701	5.752
20	2.950	3.578	3.958	4.232	4.445	4.620	4.768	4.895	5.008	5.108	5.199	5.282	5.357	5.427	5.492	5.553	5.610	5.663	5.714

df	k-> 2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20
21	2.941	3.565	3.942	4.213	4.424	4.597	4.743	4.870	4.981	5.081	5.170	5.252	5.327	5.396	5.460	5.520	5.576	5.629	5.679
22	2.933	3.553	3.927	4.196	4.405	4.577	4.722	4.847	4.957	5.056	5.144	5.225	5.299	5.368	5.431	5.491	5.546	5.599	5.648
23	2.926	3.542	3.914	4.180	4.388	4.558	4.702	4.826	4.935	5.033	5.121	5.201	5.274	5.342	5.405	5.464	5.519	5.571	5.620
24	2.919	3.532	3.901	4.166	4.373	4.541	4.684	4.807	4.915	5.012	5.099	5.179	5.251	5.319	5.381	5.439	5.494	5.545	5.594
25	2.913	3.523	3.890	4.153	4.358	4.526	4.667	4.789	4.897	4.993	5.079	5.158	5.230	5.297	5.359	5.417	5.471	5.522	5.570
26	2.907	3.514	3.880	4.141	4.345	4.511	4.652	4.773	4.880	4.975	5.061	5.139	5.211	5.277	5.339	5.396	5.450	5.500	5.548
27	2.902	3.506	3.870	4.130	4.333	4.498	4.638	4.758	4.864	4.959	5.044	5.122	5.193	5.259	5.320	5.377	5.430	5.480	5.528
28	2.897	3.499	3.861	4.120	4.322	4.486	4.625	4.745	4.850	4.944	5.029	5.106	5.177	5.242	5.302	5.359	5.412	5.462	5.509
29	2.892	3.493	3.853	4.111	4.311	4.475	4.613	4.732	4.837	4.930	5.014	5.091	5.161	5.226	5.286	5.342	5.395	5.445	5.491
30	2.888	3.486	3.845	4.102	4.301	4.464	4.601	4.720	4.824	4.917	5.001	5.077	5.147	5.211	5.271	5.327	5.379	5.429	5.475
31	2.884	3.481	3.838	4.094	4.292	4.454	4.591	4.709	4.812	4.905	4.988	5.064	5.134	5.198	5.257	5.313	5.365	5.414	5.460
32	2.881	3.475	3.832	4.086	4.284	4.445	4.581	4.698	4.802	4.894	4.976	5.052	5.121	5.185	5.244	5.299	5.351	5.400	5.445
33	2.877	3.470	3.825	4.079	4.276	4.436	4.572	4.689	4.791	4.883	4.965	5.040	5.109	5.173	5.232	5.287	5.338	5.386	5.432
34	2.874	3.465	3.820	4.072	4.268	4.428	4.563	4.680	4.782	4.873	4.955	5.030	5.098	5.161	5.220	5.275	5.326	5.374	5.420
35	2.871	3.461	3.814	4.066	4.261	4.421	4.555	4.671	4.773	4.863	4.945	5.020	5.088	5.151	5.209	5.264	5.315	5.362	5.408

K=number of treatments and df of the residual df (rdf).

Tukey's Honestly Significant Difference test (HSD) or Tukey's test or Tukey's Studentized Range Test

$$MSD = Q * \sqrt{\frac{s^2}{r}}$$

For the example with 6 =rdf and 4 treatments Q=4.89

S^2=0.40

r=3

So the MSD=1.79

Any difference greater than 1.79 is significant

Means

    V1=6.50;   V2=7.60;   V3=6.60;   V4=8.30