# Topic 22

# Analysis of Variance

## Contrasts

# Outline

## Example

Honey Bee Queen Development Time

## Contrasts

# Honey Bee Queen Development Time

- The development time for a European queen in a honey bee hive is suspected to depend on the temperature of the hive.
- To examine this, queens are reared in a low (31.1° C), a medium (32.8° C) and a high temperature hive (34.4° C).
- The hypothesis is that higher temperatures increase metabolism rate and thus reduce the time needed from the time the egg is laid until an adult queen honey bee emerges from the cell.

Figure: Emerging adult honey bee queen

# Honey Bee Queen Development Time

The hypothesis is

$$H_0 : \mu_{low} = \mu_{med} = \mu_{high} \quad \text{versus} \quad H_1 : \mu_{low}, \mu_{med}, \mu_{high} \text{ differ}$$

where $\mu_{low}$, $\mu_{med}$, and $\mu_{high}$ are, respectively, the mean development time in days for queen eggs reared in a low, a medium, and a high temperature hive.

Here are the data and a boxplot:

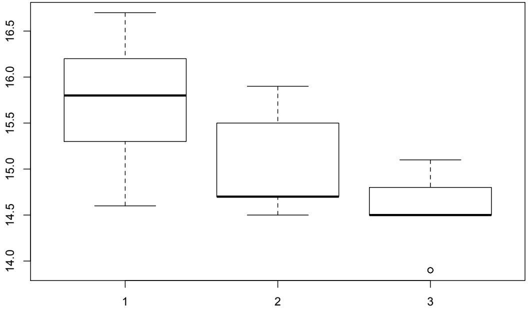
```
> ehblow<-c(16.2,14.6,15.8,15.8,15.8,15.8,16.2,16.7,15.8,16.7,15.3,14.6,
  15.3,15.8)
> ehbmed<-c(14.5,14.7,15.9,15.5,14.7,14.7,14.7,15.5,14.7,15.2,15.2,15.9,
  14.7,14.7)
> ehbhigh<-c(13.9,15.1,14.8,15.1,14.5,14.5,14.5,14.5,13.9,14.5,14.8,14.8,
  13.9,14.8,14.5,14.5,14.8,14.5,14.8)
> boxplot(ehblow,ehbmed,ehbhigh)
```

# Honey Bee Queen Development Time

```
> ehb<-c(ehblow,ehbmed,ehbhigh)
> temp<-c(rep(1,length(ehblow)),
  rep(2,length(ehbmed)),
  rep(3,length(ehbhigh)))
> ftemp<-factor(temp,c(1:3))
> anova(lm(ehb~ftemp))

Analysis of Variance Table

Response: ehb
          Df Sum Sq Mean Sq F value    Pr(>F)
ftemp      2 11.222  5.6111  23.307 1.252e-07 ***
Residuals 44 10.593  0.2407
---
Signif. codes:  0 *** 0.001 ** 0.01 * 0.05 . 0.1   1
```



# Contrasts

After completing a one way analysis of variance, resulting, as above, in rejecting the null hypotheses, a typical follow-up procedure is the use of contrasts. Contrasts use as a null hypothesis that some linear combination of the means equals to zero.

To see if the mean queen development time for medium hive temperature is midway between the time for the high and low temperature hives, we have the contrast,

$$H_0 : \frac{1}{2}(\mu_{low} + \mu_{high}) = \mu_{med} \quad \text{versus} \quad H_1 : \frac{1}{2}(\mu_{low} + \mu_{high}) \neq \mu_{med}$$

or

$$H_0 : \frac{1}{2}\mu_{low} - \mu_{med} + \frac{1}{2}\mu_{high} = 0 \quad \text{versus} \quad H_1 : \frac{1}{2}\mu_{low} - \mu_{med} + \frac{1}{2}\mu_{high} \neq 0.$$

# Contrasts

Notice that, under the null hypothesis, the mean

$$E\left[\frac{1}{2}\bar{Y}_{low} - \bar{Y}_{med} + \frac{1}{2}\bar{Y}_{high}\right] = \frac{1}{2}\mu_{low} - \mu_{med} + \frac{1}{2}\mu_{high} = 0$$

and the variance

$$\mathrm{Var}\left(\frac{1}{2}\bar{Y}_{low} - \bar{Y}_{med} + \frac{1}{2}\bar{Y}_{high}\right) = \frac{1}{4}\frac{\sigma^2}{n_{low}} + \frac{\sigma^2}{n_{med}} + \frac{1}{4}\frac{\sigma^2}{n_{high}}.$$

This leads to the test statistic

$$t = \frac{\frac{1}{2}\bar{y}_{low} - \bar{y}_{med} + \frac{1}{2}\bar{y}_{high}}{s_{residual}\sqrt{\frac{1}{4n_{low}} + \frac{1}{n_{med}} + \frac{1}{4n_{high}}}} = \frac{\frac{1}{2}15.743 - 15.043 + \frac{1}{2}14.563}{0.4906\sqrt{\frac{1}{4\cdot 14} + \frac{1}{14} + \frac{1}{4\cdot 19}}} = 0.7005.$$

The $p$-value,

```
> 2*(1-pt(0.7005,44))
[1] 0.487303
```

again, is considerably too high to reject the null hypothesis.

# Contrasts

If we want to see if the rain forest has seen a change in logged areas over the past 8 years in the mean number of trees. This can be written as

$$H_0 : \mu_2 = \mu_3 \quad \text{versus} \quad H_1 : \mu_2 \neq \mu_3$$

or

$$H_0 : \mu_2 - \mu_3 = 0 \quad \text{versus} \quad H_1 : \mu_2 - \mu_3 \neq 0$$

Under the null hypothesis, the test statistic has a $t$-distribution with $n - q = 33 - 3 = 30$ degrees of freedom. Here

$$t = \frac{\bar{y}_2 - \bar{y}_3}{s_{residual}\sqrt{\frac{1}{n_2} + \frac{1}{n_3}}} = \frac{14.083 - 15.778}{5.234\sqrt{\frac{1}{12} + \frac{1}{9}}} = -0.7344,$$

Exercise. Compute the $p$-value for this two-sided test and comment on the strength of the evidence against the null hypothesis.

# Contrasts

Exercise. Under the null hypothesis appropriate for one way analysis of variance, with $n_j$ observations in group $j = 1, \ldots, q$ and $\bar{Y}_j = \sum_{i=1}^{n_j} Y_{ij}/n_j$,

$$E[c_1\bar{Y}_1 + \cdots + c_q\bar{Y}_q] = c_1\mu_1 + \cdots + c_q\mu_q, \quad \mathrm{Var}(c_1\bar{Y}_1 + \cdots + c_q\bar{Y}_q) = \frac{c_1^2\sigma^2}{n_1} + \cdots + \frac{c_q^2\sigma^2}{n_q}.$$

In general, a contrast begins with a linear combination of the means

$$\psi = c_1\mu_1 + \cdots + c_q\mu_q.$$

The hypothesis is

$$H_0: \psi = 0 \quad \text{versus} \quad H_1: \psi \neq 0.$$

For sample means, $\bar{y}_1, \ldots, \bar{y}_q$, the test statistic is

$$t = \frac{c_1\bar{y}_1 + \cdots + c_q\bar{y}_q}{s_{residual}\sqrt{\frac{c_1^2}{n_1} + \cdots + \frac{c_q^2}{n_q}}}.$$

which, under the null hypothesis, has a $t$ distribution with $n - q$ degrees of freedom.