

multiple comparisons

- ▶ which means are different?
- ▶ controlling the Type I error rate

which means differ

- ▶ two sample t tests for differences in each possible pair of groups
- ▶ multiple tests → inflated Type I error rate
- ▶ solution: use modified significance level

multiple comparisons

- ▶ testing many pairs of groups is called multiple comparisons
- ▶ the Bonferroni correction suggests that a more stringent significance level is more appropriate for these tests
 - ▶ adjust α by the number of comparisons being considered

Bonferroni correction:

$$\alpha^* = \alpha / K \quad K: \text{number of comparisons}, K = \frac{k(k-1)}{2}$$

The social class variables has 4 levels. If $\alpha = 0.05$, what should the modified significance level be for two sample t tests for determining which pairs of groups have significantly different means?

$$K = 4$$

$$K = \frac{4 \times 3}{2} = 6$$

$$\alpha^* = 0.05 / 6 \approx 0.0083$$

pairwise comparisons

- ▶ for multiple comparisons after ANOVA, since the assumption of equal variability across groups must have been satisfied, re-think the standard error and the degrees of freedom:
 - ▶ use a consistent measure of standard error and consistent degrees of freedom for all tests
- ▶ compare the p-values from each test to the modified significance level

Standard error for multiple pairwise comparisons:

$$SE = \sqrt{\frac{MSE}{n_1} + \frac{MSE}{n_2}}$$

indep. groups test:

$$SE = \sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}}$$

Degrees of freedom for multiple pairwise comparisons:

$$df = df_E$$

$$df = \min(n_1 - 1, n_2 - 1)$$

Is there a difference between the average vocabulary scores between middle and lower class Americans?

	Df	Sum Sq	Mean Sq	F value	Pr(> F)
class	3	236.56	78.855	21.735	<0.0001
Residuals	791	2869.80	3.628		
Total	794	3106.36			

	n	mean
lower class	41	5.07
middle class	331	6.76

$H_0: \mu_{middle} - \mu_{lower} = 0$

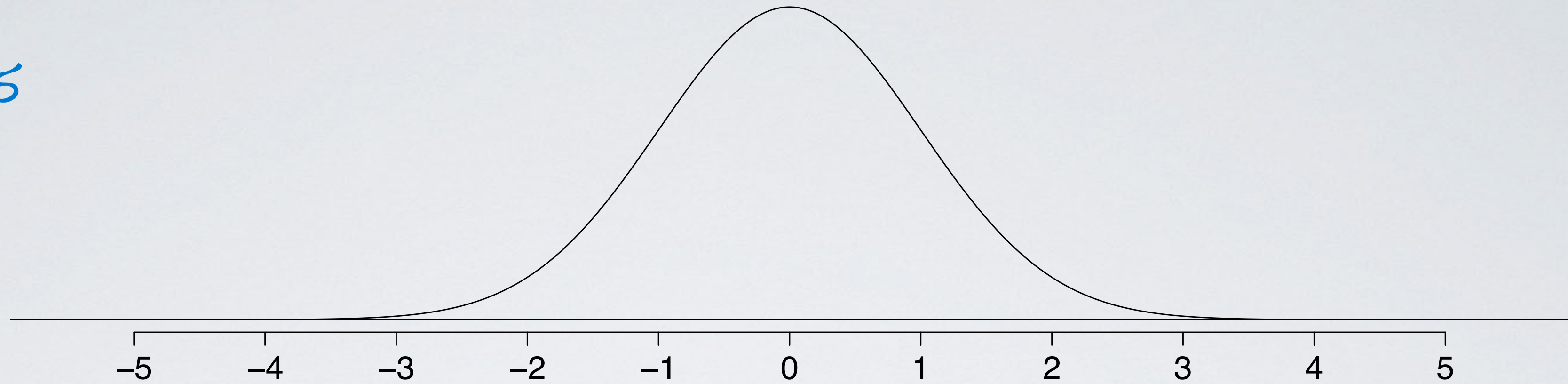
$H_A: \mu_{middle} - \mu_{lower} \neq 0$

$$T = \frac{(\bar{X}_{middle} - \bar{X}_{lower}) - 0}{\sqrt{\frac{MSE}{n_{middle}} + \frac{MSE}{n_{lower}}}} = \frac{(6.76 - 5.07)}{\sqrt{\frac{3.628}{331} + \frac{3.628}{41}}} = \frac{1.69}{0.315} = 5.365$$

 $df = 791$

$$T = 5.365$$

$$df = 791$$



R

```
> 2 * pt(5.365, df = 791, lower.tail = FALSE)
[1] 1.063895e-07
```

$$\alpha^* = 0.0083$$

$p\text{-value} < \alpha^* \rightarrow \text{Reject } H_0$