Today's Data Files

- fuel.txt
- nematodes.txt

MATH 651: Regression Methods & Generalized Linear Models

Lecture 9: One-way Analysis of Variance (One way ANOVA)

Reading: KNNL Chapters 16-17

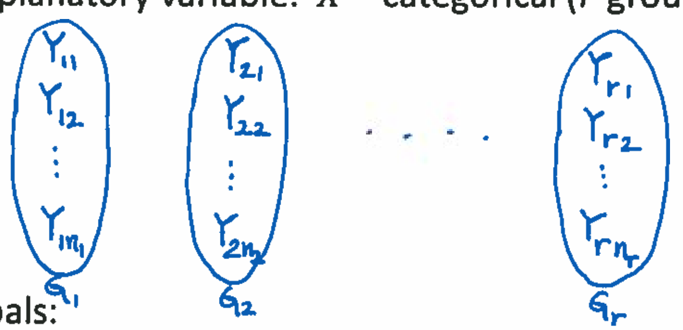
Kimberly F. Sellers
Department of Mathematics & Statistics

Regression and ANOVA

- ANOVA concept more general than our use in regression analysis suggests
- ANOVA concerns relationship between quantitative response variable and (usually, one or more) qualitative predictor variables

Single Factor ANOVA: Setup

- Response variable: Y – quantitative
- Explanatory variable: X – categorical (r groups)



- Goals:
  - Examine the relationship between X and Y
  - Compare means across > 2 two groups ($\mu_1, \mu_2, \ldots, \mu_r$)

Single Factor ANOVA: Notation

Sample 1: $Y_{11}, Y_{12}, \dots, Y_{1n_1} \Rightarrow \bar{Y}_{1\cdot}, s_1$

Sample 2: $Y_{21}, Y_{22}, \dots, Y_{2n_2} \Rightarrow \bar{Y}_{2\cdot}, s_2$

$\vdots$

Sample r: $Y_{r1}, Y_{r2}, \dots, Y_{rn_r} \Rightarrow \bar{Y}_{r\cdot}, s_r$

Total sample size: $n_T = n_1 + n_2 + \cdots n_r = \sum_{i=1}^{r} n_i$

Group mean: $\bar{Y}_{i\cdot} = \frac{1}{n_i}\sum_{j=1}^{n_i} Y_{ij}$

Group variance: $s_i^2 = \frac{1}{n_i - 1}\sum_{j=1}^{n_i}\left(Y_{ij} - \bar{Y}_{i\cdot}\right)^2$

Grand mean: $\bar{Y}_{\cdot\cdot} = \frac{\sum\sum Y_{ij}}{n_T}$

Motivating Example

The driver of a diesel-powered automobile decided to test the quality of three types of diesel fuel sold in the area based on mpg. Are the three average mpg equal, or does a difference exist based on brand?

Brand A	38.7	39.2	40.1	38.9	
Brand B	41.9	42.3	41.3		
Brand C	40.8	41.2	39.5	38.9	40.3

The data (**fuel.txt**) are provided on Blackboard.

Using R:

```
> fuel <- read.table("C:/MATH651data/fuel.txt", sep="\t", header=TRUE)
```

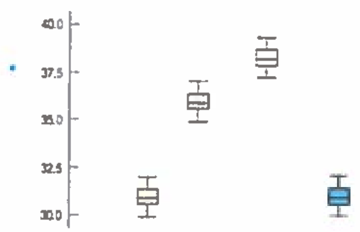
The Point

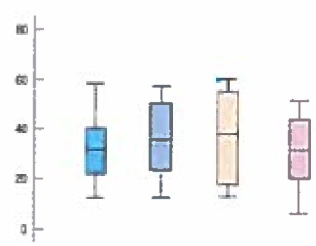
- Test equality of means (i.e., *Y* is not related to *X*) versus something otherwise (i.e., *Y* is somehow related to *X*)
- $H_0: \mu_1 = \mu_2 = \cdots = \mu_k$ vs. H_1: not all μ_i's equal
- Note: H_1 above $\not\Rightarrow H_1: \mu_1 \neq \mu_2 \neq \cdots \neq \mu_k$

Exercise: Make comparative boxplots of the **fuel.txt** data. What is your guess regarding the outcome of this hypothesis test?
Using R: > boxplot(mpg~brand, data=fuel)

Do Means Differ Across Groups?

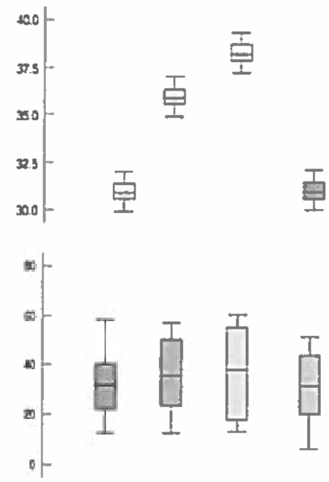
- Want to know if observed differences in sample means can occur by chance (because of random sampling) or not





Thought this was Analysis of VARIANCE, not Analysis of Means?!

- **First figure**: Can't imagine means that far apart just from natural sampling variability alone.
- **Second figure**: These observations *could* have occurred from treatments (populations) with the same means.
- **Moral**: Conclusions depend on both the **variation** between sample means, and how much **variation** exists within each sample.



F-test

- Compares differences/variation between group means with variation within groups
- When differences between means are large compared with variation within groups, we reject H_0 and conclude means are (probably) not equal
- F-statistic:

$$F = \frac{\text{variation among averages } \bar{Y}_{1\cdot}, \bar{Y}_{2\cdot}, \ldots, \bar{Y}_{r\cdot}}{\text{variation within samples}}$$

The One-Way ANOVA Model

$$Y_{ij} = \mu_i + \epsilon_{ij}; \quad i = 1, 2, \dots, r \quad j = 1, 2, \dots, n_i$$

- Y_{ij} = response variable value in jth trial for ith treatment
- μ_i = (fixed) parameters representing unknown population mean for ith treatment
- ϵ_{ij} = independent $N(0, \sigma^2)$ (random) errors

- Called the cell means model

Note: Formulation $\Rightarrow Y_{ij} \sim N(\mu_i, \sigma^2)$ indpt.

Parameter Estimation: Means

- Estimating $\mu_1, \mu_2, \dots, \mu_r$:

$$\hat{\mu}_i = \bar{Y}_{i\cdot} = \frac{\sum_{j=1}^{n_i} Y_{ij}}{n_i} = \frac{Y_{i\cdot}}{n_i}$$

$$\hat{\mu}_A = \bar{Y}_{A\cdot} = 39.225$$
$$\hat{\mu}_B = \bar{Y}_{B\cdot} = 41.833$$
$$\hat{\mu}_C = \bar{Y}_{C\cdot} = 40.140$$

Using R:

```
> summary(fuel$mpg[fuel$brand=="A"]) # similarly for B and C
> fuel.lm1 <- lm(mpg ~ brand-1,data=fuel)
> fuel.lm1 # alternative approach; see "coefficients"
```

Visual Depictions: Bar Graphs and Main Effects Plots

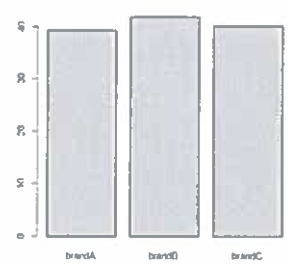
- Used to display estimated factor level means in two dimensions
- Used to compare magnitudes of different factor level means
- In main effects plots, trend lines most appropriate for quantitative factors; not particularly meaningful for qualitative factors

13

Using R: Plots of Estimated Factor Level Means

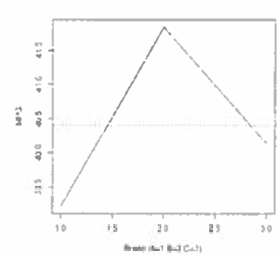
- Bar Graph:

```
> barplot(fuel.lm1$coeff)
```



- Main Effects Plot:

```
> plot(fuel.lm1$coeff,type="l",ylab="MPG",
xlab="Brand (A=1,B=2,C=3)")
> abline(h=mean(fuel.lm1$coeff),lty=2)
```



14

Parameter Estimation: Variance

- Pool information from all r samples to estimate σ
- Recall estimating σ for one sample: For data $y_1, y_2, \ldots, y_n$, we took their average ($\bar{y}$) and computed sample standard deviation as:

$$s = \sqrt{\frac{\sum_{i=1}^{n}(y_i - \bar{y})^2}{n-1}}$$

where deviations,

$$(y_1 - \bar{y}), (y_2 - \bar{y}), \ldots, (y_n - \bar{y})$$

sum to 0

- n terms but only $n-1$ are "free to vary"

15

Estimating σ From r Samples

- Based on residual estimates, $e_{ij} = Y_{ij} - \bar{Y}_i$
- Sample 1: $(Y_{11} - \bar{Y}_1), (Y_{12} - \bar{Y}_1), \ldots, (Y_{1n_1} - \bar{Y}_1) \Rightarrow n_1 - 1$ can vary
- Sample 2: $(Y_{21} - \bar{Y}_2), (Y_{22} - \bar{Y}_2), \ldots, (Y_{2n_2} - \bar{Y}_2) \Rightarrow n_2 - 1$ can vary

  $\vdots$
- Sample r: $(Y_{r1} - \bar{Y}_r), (Y_{r2} - \bar{Y}_r), \ldots, (Y_{rn_r} - \bar{Y}_r) \Rightarrow n_r - 1$ can vary

$$\overline{}$$
$$n_T - r$$

- Thus,

$$\widehat{\sigma^2} = s^2 = \frac{\sum_i \sum_j (Y_{ij} - \bar{Y}_i)^2}{n_T - r} \quad \text{and} \quad \hat{\sigma} = s$$

16

Estimating Variation

- We can reconsider the model as

$$Y_{ij} = \mu_i + \epsilon_{ij} = \mu + \tau_i + \epsilon_{ij}$$

  and ask $H_0: \tau_1 = \tau_2 = \cdots = \tau_r = 0$
- Data can be represented as

Data = Grand mean + Treatment effect + Residual

$$Y_{ij} = \bar{Y}_{..} + (\bar{Y}_i - \bar{Y}_{..}) + (Y_{ij} - \bar{Y}_i)$$
$$\Rightarrow Y_{ij} - \bar{Y}_{..} = (\bar{Y}_i - \bar{Y}_{..}) + (Y_{ij} - \bar{Y}_i)$$

- Total sum of squares (SSTO): $\sum_{i=1}^{r}\sum_{j=1}^{n_i}(Y_{ij} - \bar{Y}_{..})^2$

17

Estimating Variation (cont.)

- Doing the same on the right-hand side,

$$\sum_{i=1}^{r}\sum_{j=1}^{n_i}[(\bar{Y}_i - \bar{Y}_{..}) + (Y_{ij} - \bar{Y}_i)]^2 = \sum_{i=1}^{r}\sum_{j=1}^{n_i}(\bar{Y}_i - \bar{Y}_{..})^2 + \sum_{i=1}^{r}\sum_{j=1}^{n_i}(Y_{ij} - \bar{Y}_i)^2$$

  – Variation between groups:

$$\text{SSTR} = \sum_{i=1}^{r}\sum_{j=1}^{n_i}(\bar{Y}_i - \bar{Y}_{..})^2 = \sum_{i=1}^{r} n_i(\bar{Y}_i - \bar{Y}_{..})^2$$

  – Variation within groups:

$$\text{SSE} = \sum_{i=1}^{r}\sum_{j=1}^{n_i}(Y_{ij} - \bar{Y}_i)^2 = \sum_{i=1}^{r}\sum_{j=1}^{n_i} e_{ij}^2 = \sum_{i=1}^{r}(n_i - 1)s_i^2$$

- Total sum of squares: SSTO = SSTR + SSE

18

One-Way ANOVA Table

Source of Variation	df	Sum of Squares (SS)	Mean Square (MS)	F-ratio
Between treatments (groups)	$r-1$	$SSTR$	$MSTR = \frac{SSTR}{r-1}$	$F = \frac{MSTR}{MSE}$
Error (within groups)	$n_T - r$	SSE	$MSE = \frac{SSE}{n_T - r}$	
Total	$n_T - 1$	$SSTO$		

- $r =$ number of treatments/groups
- $n_T =$ total number of observations in the study
- *F*-statistic follows *F*-distribution with df $r-1, n_T - r$

19

Using R

- You already know how to do this (sort of)

```
> fuel.aov <- aov(mpg ~ brand, data=fuel)
> summary(fuel.aov)
> fuel.lm <- lm(mpg~brand, data=fuel)
> anova(fuel.lm)
Analysis of Variance Table

Response: mpg
          Df  Sum Sq Mean Sq  F value   Pr(>F)
Brand      2 11.7830  5.8915  10.224  0.004823 **
Residuals  9  5.1862  0.5762
```

- Clearly, you can use regression methods instead of ANOVA

20

Cell Means Linear Model

- To write as linear model:

$$Y_{ij} = \mu_i + \epsilon_{ij} \quad = \mu_1 X_{ij1} + \mu_2 X_{ij2} + \cdots + \mu_r X_{ijr} + \epsilon_{ij}$$
$$= \boldsymbol{X\beta} + \boldsymbol{\epsilon}$$

where $X_{ijk} = \begin{cases} 1, \text{if kth factor level used} \\ 0, \text{otherwise} \end{cases}$

- Example: for the **fuel.txt** dataset….

$$\begin{pmatrix} Y_{11} \\ \vdots \\ Y_{14} \\ \hline Y_{21} \\ \vdots \\ Y_{23} \\ \hline Y_{31} \\ \vdots \\ Y_{35} \end{pmatrix} = \left(\begin{array}{c|c|c} 1 & & \\ \vdots & & 0 \\ 1 & & \\ \hline 0 & 1 & \\ \vdots & \vdots & 0 \\ 0 & 1 & \\ \hline & & 1 \\ 0 & & \vdots \\ & & 1 \end{array}\right) \begin{pmatrix} \beta_1 = \mu_1 \\ \beta_2 = \mu_2 \\ \beta_3 = \mu_3 \end{pmatrix} + \begin{pmatrix} \epsilon_{11} \\ \vdots \\ \epsilon_{14} \\ \hline \epsilon_{21} \\ \vdots \\ \epsilon_{23} \\ \hline \epsilon_{31} \\ \vdots \\ \epsilon_{35} \end{pmatrix}$$

$\underset{\sim}{Y} = \beta_1 \underset{\sim A}{I} + \beta_2 \underset{\sim B}{I} + \beta_3 \underset{\sim C}{I} + \underset{\sim}{\epsilon}$

where $\beta_i = \mu_i$

- Note: model has no intercept term

Linear Model (cont.)

- Can use least squares for estimation
- $\boldsymbol{b} = (\hat{\mu}_1, \dots, \hat{\mu}_r)' = (\boldsymbol{X'X})^{-1}\boldsymbol{X'Y}$ is BLUE for $\boldsymbol{\beta} = (\mu_1, \dots, \mu_r)'$

where $\boldsymbol{X'X} = \begin{pmatrix} n_1 & 0 \dots & 0 \\ 0 & n_2 & 0 \\ 0 \dots & 0 & \ddots\, n_r \end{pmatrix}, \boldsymbol{X'Y} = \begin{pmatrix} Y_{1\cdot} \\ \vdots \\ Y_{r\cdot} \end{pmatrix}$

- $E(\boldsymbol{b}) = \boldsymbol{\beta}$

- $V(\boldsymbol{b}) = \sigma^2(\boldsymbol{X'X})^{-1} = \sigma^2 \begin{pmatrix} 1/n_1 & 0 \dots & 0 \\ 0 & 1/n_2 & 0 \\ 0 \dots & 0 & \ddots\, 1/n_r \end{pmatrix}$

Linear Model: Background for F-Statistic

- $Var(b_i) = Var(\hat{\mu}_i) = \frac{\sigma^2}{n_i}$
- $\boldsymbol{b} \sim N(\boldsymbol{\beta}, \sigma^2(\boldsymbol{X}'\boldsymbol{X})^{-1})$
- $MSE = \frac{1}{n_T - r}\sum_i \sum_j \left(Y_{ij} - \bar{Y}_{i\cdot}\right)^2 = \frac{1}{n_T - r}\sum_i (n_i - 1)s_i^2$
- $E\left(s_i^2\right) = \sigma^2 \Rightarrow E(MSE) = \frac{1}{n_T - r}\sum_i (n_i - 1)\sigma^2 = \sigma^2$

  $\Rightarrow$ MSE unbiased estimator of σ^2
- Similarly,

  $E(MSTR) = \sigma^2 + \frac{\sum_{i=1}^{r} n_i(\mu_i - \mu_\cdot)^2}{r-1}$ where $\mu_\cdot = \frac{\sum_{i=1}^{r} n_i \mu_i}{n_T}$

  – All treatment means equal $\Rightarrow E(MSTR) = \sigma^2$

23

F-Statistic (cont.)

- Consider $H_0: \mu_1 = \cdots = \mu_r$ vs. H_1: not all μ_i equal
- Test statistic: $F = \frac{MSTR}{MSE}$
  – Large F support H_1; F near 1 supports H_0
- Background: Cochran's theorem implies that, if H_0 holds, then $\frac{SSE}{\sigma^2}, \frac{SSTR}{\sigma^2}$ independent χ^2 variables with $n_T - r$ and $r - 1$ df, respectively

24

Alternative Parametrization

- Reconsider model as $Y_{ij} = \mu + \tau_i + \epsilon_{ij}$, where
  - μ = constant component, common to all observations
  - τ_i = ith treatment effect
  - ϵ_{ij} = independent $N(0, \sigma^2)$ random errors
  - – Called treatment effects formulation; also linear model
- Exercise: rewrite $Y_{ij} = \mu + \tau_i + \epsilon_{ij}$ in matrix form

$$\begin{pmatrix} Y_{11} \\ \vdots \\ Y_{1n_1} \\ \hline Y_{21} \\ \vdots \\ Y_{2n_2} \\ \hline \cdots \\ \hline Y_{r1} \\ \vdots \\ Y_{rn_r} \end{pmatrix} = \begin{pmatrix} 1 & 1 & & 0 \\ \vdots & \vdots & & \\ 1 & 1 & & \\ \hline 1 & 0 & 1 & 0 \\ \vdots & \vdots & \vdots & \\ 1 & 0 & 1 & \\ \hline & \cdots & & \\ \hline 1 & & 0 & 1 \\ \vdots & & & \vdots \\ 1 & & & 1 \end{pmatrix} \begin{pmatrix} \mu \\ \tau_1 \\ \tau_2 \\ \vdots \\ \tau_r \end{pmatrix} + \begin{pmatrix} \epsilon_{11} \\ \vdots \\ \epsilon_{1n_1} \\ \hline \epsilon_{21} \\ \vdots \\ \epsilon_{2n_2} \\ \hline \cdots \\ \hline \epsilon_{r1} \\ \vdots \\ \epsilon_{rn_r} \end{pmatrix}$$

25

Alternative Parametrization: Linear Model

- Normal equations: $\boldsymbol{b} = (\boldsymbol{X}'\boldsymbol{X})^{-1}\boldsymbol{X}'\boldsymbol{Y}$
- Problem: $\boldsymbol{X}'\boldsymbol{X} = \begin{pmatrix} n_T & n_1 & \cdots & n_r \\ n_1 & n_1 & 0 & 0 \\ \vdots & 0 & \ddots & 0 \\ n_r & 0 & 0 & n_r \end{pmatrix}$ singular
- Resolve by imposing restrictions on parameters to obtain full rank $\boldsymbol{X}$ matrix
  - – Here, assume $\sum_{i=1}^{r} \tau_i = 0$

26

Factor Effects Linear Model

- μ: grand (or overall) mean
- $\tau_i = (\mu + \tau_i) - \mu = \mu_i - \mu =$ Treatment Mean - Grand Mean = Treatment Effect
- $\sum_{i=1}^{r} \tau_i = 0 \Rightarrow \tau_r = -\tau_1 - \tau_2 - \cdots - \tau_{r-1}$ (e.g.)

  $\Rightarrow \mu_r = \mu + \tau_r = \mu - \tau_1 - \tau_2 - \cdots - \tau_{r-1}$

  $\Rightarrow$ model only needs $\mu, \tau_1, \tau_2, \ldots, \tau_{r-1}$

SEE ATTACHED

- Different reparametrization details produce different $\boldsymbol{X}$ and $\boldsymbol{\beta}$; in all cases, columns of $\boldsymbol{X}$ span same space

Multiple Comparisons

- Failing to reject $H_0 \Rightarrow$ Done.
- Rejecting $H_0 \Rightarrow$ Asking which means are different?
- Can't do multiple simple *t*-tests
  - Performing multiple tests increases the risk of making Type I error
  - Performing enough tests can result in rejecting a H_0 by mistake but won't know which one.
- Methods of multiple comparisons avoid inflation of experiment-wise error (overall Type I error)
- Can also calculate simultaneous confidence intervals

$$Y_{ij} = \mu + \tau_i + \epsilon_{ij}$$

Slide 25

$$\underbrace{\begin{pmatrix} Y_{11} \\ \vdots \\ Y_{1n_1} \\ \hline Y_{21} \\ \vdots \\ Y_{2n_2} \\ \hline \cdots \\ \hline Y_{r1} \\ \vdots \\ Y_{rn_r} \end{pmatrix}}_{(n_T \times 1)} = \underbrace{\begin{pmatrix} 1 & 1 & & \\ \vdots & \vdots & 0 & \\ 1 & 1 & & \\ \hline 1 & 0 & 1 & 0 \\ \vdots & \vdots & \vdots & \\ 1 & 0 & 1 & \\ \hline & & \cdots & \\ \hline 1 & & & 1 \\ \vdots & & 0 & \vdots \\ 1 & & & 1 \end{pmatrix}}_{n_T \times (r+1)} \underbrace{\begin{pmatrix} \mu \\ \tau_1 \\ \tau_2 \\ \vdots \\ \tau_r \end{pmatrix}}_{(r+1) \times 1} + \underbrace{\begin{pmatrix} \epsilon_{11} \\ \vdots \\ \epsilon_{1n_1} \\ \hline \epsilon_{21} \\ \vdots \\ \epsilon_{2n_2} \\ \hline \cdots \\ \hline \epsilon_{r1} \\ \vdots \\ \epsilon_{rn_r} \end{pmatrix}}_{n_T \times 1}$$

Slide 27

$$\underbrace{\begin{pmatrix} Y_{11} \\ \vdots \\ Y_{1n_1} \\ \hline Y_{21} \\ \vdots \\ Y_{2n_2} \\ \hline \cdots \\ \hline Y_{r-1,1} \\ \vdots \\ Y_{r-1,n_{r-1}} \\ \hline Y_{r1} \\ \vdots \\ Y_{rn_r} \end{pmatrix}}_{n_T \times 1} = \underbrace{\begin{pmatrix} 1 & 1 & & & \\ \vdots & \vdots & & 0 & \\ 1 & 1 & & & \\ \hline 1 & 0 & 1 & 0 & \\ \vdots & \vdots & \vdots & & \\ 1 & 0 & 1 & & \\ \hline & & \cdots & & \\ \hline 1 & & & & 1 \\ \vdots & & 0 & & \vdots \\ 1 & & & & 1 \\ \hline 1 & -1 & -1 & \cdots & -1 \\ \vdots & \vdots & \vdots & & \vdots \\ 1 & -1 & -1 & \cdots & -1 \end{pmatrix}}_{n_T \times r} \underbrace{\begin{pmatrix} \mu \\ \tau_1 \\ \tau_2 \\ \vdots \\ \tau_{r-1} \end{pmatrix}}_{r \times 1} + \underbrace{\begin{pmatrix} \epsilon_{11} \\ \vdots \\ \epsilon_{1n_1} \\ \hline \epsilon_{21} \\ \vdots \\ \epsilon_{2n_2} \\ \hline \cdots \\ \hline \epsilon_{r-1,1} \\ \vdots \\ \epsilon_{r-1,n_{r-1}} \\ \hline \epsilon_{r1} \\ \vdots \\ \epsilon_{rn_r} \end{pmatrix}}_{n_T \times 1}$$

Estimation and Testing of Treatment Effects

- Inferences for treatment effects usually concerned with:
  - Single treatment mean, μ_i
  - Difference between two treatment means
  - Contrast among treatment means
  - Linear combination of treatment means

29

Inference for Single Treatment Mean

- Estimate $\hat{\mu}_i = \bar{Y}_{i\cdot}$ has $E(\bar{Y}_{i\cdot}) = \mu_i$ and $V(\bar{Y}_{i\cdot}) = \frac{\sigma^2}{n_i}$
- Confidence limits for μ_i:

$$\bar{Y}_{i\cdot} \pm t_{n_T-r}\left(1 - \frac{\alpha}{2}\right) \cdot s(\bar{Y}_{i\cdot})$$

- Consider hypotheses, $H_0: \mu_i = c$ vs. $H_1: \mu_i \neq c$
  - Test statistic: $T = \frac{\bar{Y}_{i\cdot} - c}{s(\bar{Y}_{i\cdot})} \sim t_{n_T-r}$
  - Rejection region: reject H_0 when $|T| > t_{n_T-r}\left(1 - \frac{\alpha}{2}\right)$

30

Inference for Differences Between Treatment Means

- Pairwise comparisons: consider difference
$$D = \mu_i - \mu_{i'}$$
- Has unbiased estimator, $\widehat{D} = \bar{Y}_{i.} - \bar{Y}_{i'.}$
- $\bar{Y}_{i.}$ and $\bar{Y}_{i'.}$ indpt
$$\Rightarrow V(\widehat{D}) = V(\bar{Y}_{i.}) + V(\bar{Y}_{i'.}) = \sigma^2\left(\frac{1}{n_i} + \frac{1}{n_{i'}}\right)$$
- $V(\widehat{D})$ estimated by $s^2(\widehat{D}) = MSE\left(\frac{1}{n_i} + \frac{1}{n_{i'}}\right)$

31

Inference for Differences Between Treatment Means (cont.)

- Resulting inference:
$$\frac{\widehat{D} - D}{s(\widehat{D})} \sim t_{n_T - r}$$
- Confidence interval for D:
$$\widehat{D} \pm t_{n_T - r}\left(1 - \frac{\alpha}{2}\right) \cdot s(\widehat{D})$$
- Hypothesis test: $H_0: \mu_i = \mu_{i'}$ vs. $H_1: \mu_i \neq \mu_{i'}$
  - Test statistic: $T = \frac{\widehat{D}}{s(\widehat{D})}$
  - Rejection region: reject H_0 when $|T| > t_{n_T - r}\left(1 - \frac{\alpha}{2}\right)$

Inferences for Contrasts of Treatment Means

- Contrasts: linear combination of treatment means

$$L = \sum_{i=1}^{r} c_i \, \mu_i$$

where c_i non-random constants summing to zero, i.e. $\sum_{i=1}^{r} c_i = 0$

33

Contrast Estimator

- Unbiased contrast estimator: $\hat{L} = \sum_{i=1}^{r} c_i \, \bar{Y}_{i.}$
- $Var(\hat{L}) = \sigma^2 \sum_{i=1}^{r} \frac{c_i^2}{n_i}$ is estimated by

$$s^2(\hat{L}) = MSE \sum_{i=1}^{r} \frac{c_i^2}{n_i}$$

- Resulting inference: $\frac{\hat{L}-L}{s(\hat{L})} \sim t_{n_T - r}$

34

Inferences for Contrasts of Treatment Means (cont.)

- Confidence interval for L:

$$\hat{L} \pm t_{n_T - r}\left(1 - \frac{\alpha}{2}\right) \cdot s(\hat{L})$$

- Hypothesis test: $H_0: L = 0$ vs. $H_1: L \neq 0$
  - Test statistic: $T = \frac{\hat{L}}{s(\hat{L})}$
  - Rejection region: reject H_0 when $|T| > t_{n_T - r}\left(1 - \frac{\alpha}{2}\right)$

35

Inferences for a Linear Combination of Treatment Means

- Supposed interested in linear combination of factor levels that is not a contrast, i.e. consider no restrictions on c_i coefficients with $L = \sum_{i=1}^{r} c_i \mu_i$
  - Confidence limits and test statistics for a linear combination L obtained like those for contrast
- $\Rightarrow$ Tests of a single treatment mean, two treatment means, and contrasts all special cases of:

$$H_0: \sum_{i=1}^{r} c_i \mu_i = c \quad \text{vs.} \quad H_1: \sum_{i=1}^{r} c_i \mu_i \neq c$$

36

Inferences for a Linear Combination of Treatment Means (cont.)

- Corresponding t-statistics can be considered equivalently as F-tests:

$$F = T^2$$

where F follows F_{1,n_T-r} distribution when H_0 true

  - Numerator df always 1 in these cases, these tests called single degree-of-freedom tests

37

Multiple Comparison

- Suppose we want to test simultaneously $k \geq 2$ contrasts
- Two limitations for comparing multiple contrasts:
  - Confidence coefficient $1 - \alpha$ only applies to particular estimate, not series of estimates; similarly, Type I error rate (α) applies to a particular test, not a series of tests
    - Example: three "indpt" t-tests at $\alpha = 0.05 \Rightarrow$ confidence $(0.95)^3 = 0.857$ (Thus $\alpha = 0.143$, not .05)
  - The confidence coefficient $1 - \alpha$ and significance level α appropriate only if estimate not suggested by the data
    - Often, experiment's results suggest important (potentially significant) relationships

38

Tukey Multiple Comparison Procedure

- Applies for considering all pairwise comparisons of factor level means, i.e. consider
$$H_0: \mu_i - \mu_{i'} = 0 \text{ vs. } H_1: \mu_i - \mu_{i'} \neq 0$$
  - Equal group sample sizes $\Rightarrow$ Tukey method family confidence coefficient $\equiv 1-\alpha$; significance level $\equiv \alpha$
  - Unequal group sample sizes $\Rightarrow$ Tukey method family confidence coefficient $> 1-\alpha$; significance level $< \alpha$ (i.e. conservative test)

39

Tukey Multiple Comparison Procedure (cont.)

- Considers studentized range distribution:
  - Suppose have $Y_1, \ldots, Y_r$ observations from $N(\mu, \sigma^2)$ distribution.
  - Let $w = \max(Y_i) - \min(Y_i) =$ observation range
  - s^2 estimates σ^2 with ν df
  - Studentized range: $q(r, \nu) = \frac{w}{s}$
  - Distribution of q can be determined (see Table B.9 in KLLN)

40

Tukey Multiple Comparison Procedure (cont.)

- Tukey multiple comparison CIs for all pairwise comparisons ($D = \mu_i - \mu_{i'}$) with family confidence interval $1 - \alpha$ are:

$$\widehat{D} \pm T \cdot s(\widehat{D})$$

where

$$\widehat{D} = \bar{Y}_{i\cdot} - \bar{Y}_{i'\cdot}$$

$$s(\widehat{D}) = MSE\left(\frac{1}{n_i} + \frac{1}{n_{i'}}\right)$$

$$T = \left(\frac{1}{\sqrt{2}}\right) q_{r,n_T - r}(1 - \alpha)$$

41

Simultaneous Testing: Tukey

- Consider $H_0: \mu_i - \mu_{i'} = 0$ vs. $H_1: \mu_i - \mu_{i'} \neq 0$
  - Test statistic: $q^* = \frac{\sqrt{2}\widehat{D}}{s(\widehat{D})}$
  - Reject H_0 if $|q^*| > q_{r,n_T - r}(1 - \alpha)$
- Notes:
  - When not all pairwise comparisons of interest, family confidence coefficient for family of comparisons considered $> 1 - \alpha$ (with the significance level $< \alpha$)
  - Tukey procedure can be used for "data snooping" as long as effects to be studied on basis of preliminary data analysis are pairwise comparisons

42

Using R: Tukey's Honest Significant Difference (HSD)

```
> TukeyHSD(fuel.aov)
  Tukey multiple comparisons of means
    95% family-wise confidence level

Fit: aov(formula = mpg ~ brand, data = fuel)

$brand
         diff        lwr        upr     p adj
B-A  2.608333  0.9895950  4.2270717 0.0038199
C-A  0.915000 -0.5067542  2.3367542 0.2248005
C-B -1.693333 -3.2411433 -0.1455234 0.0332056
```

Scheffé Multiple Comparison Procedure

- Method applies when family of interest is set of possible contrasts among treatment means:

$$L = \sum_{i=1}^{r} c_i \mu_i \text{ where } \sum_{i=1}^{r} c_i = 0$$

- Consider H_0: $L = 0$ vs. H_1: $L \neq 0$
  - Infinite number of such tests
  - Family confidence level = $1 - \alpha$ (i.e., significance level equals α) whether sample sizes equal or not

Scheffé Multiple Comparison Procedure (cont.)

- Scheffé CIs for family of contrasts L:

$$\hat{L} \pm S \cdot s(\hat{L})$$

where

$$\hat{L} = \sum_{i=1}^{r} c_i \bar{Y}_{i.}$$

$$s^2(\hat{L}) = MSE \sum_{i=1}^{r} \frac{c_i^2}{n_i}$$

$$S^2 = (r-1)F_{r-1,n_T-r}(1-\alpha)$$

45

Simultaneous Testing: Scheffé

- Consider $H_0: L = 0$ vs. $H_1: L \neq 0$
  - Test statistic:

$$F^* = \frac{\hat{L}^2}{(r-1) \cdot s^2(\hat{L})}$$

  - Reject H_0 if

$$F^* > F_{r,n_T-r}(1-\alpha)$$

46

Scheffé Multiple Comparison Procedure: Notes

- Since applications of Scheffé procedure never involve all conceivable contrasts, finite family confidence coefficient will be larger than $1-\alpha$, so $1-\alpha$ is lower bound. Thus, people often consider larger α (e.g., 90% confidence interval)
- Scheffé procedure can be used for wide variety of "data snooping" since family of statements contains all possible contrasts
- If only pairwise comparisons considered, Tukey procedure gives narrower confidence limits

47

Bonferroni Multiple Comparison Procedure

- Bonferroni approach can be used for ANOVA when family of interest is a particular set of pairwise comparisons, contrasts, or linear combinations - specified in advance
- Applicable whether sample sizes are equal or unequal

48

Bonferroni Multiple Comparison Procedure (cont.)

- CIs for g linear combinations L:

$$\hat{L} \pm B \cdot s(\hat{L})$$

where

$$B = t_{n_T - r}\left(1 - \frac{\alpha}{2g}\right)$$

- $g =$ number of comparisons in family

Simultaneous Testing: Bonferroni

- Consider $H_0: L = 0$ vs. $H_1: L \neq 0$
- Test statistic:

$$T = \frac{\hat{L}}{s(\hat{L})}$$

- Rejection region:

Reject H_0 if $|T| > t_{n_T - r}\left(1 - \frac{\alpha}{2g}\right)$

Comparing Multiple Comparison Procedures

- If all pairwise comparisons of interest, Tukey procedure is superior (narrower confidence intervals). If not all pairwise comparisons are of interest, Bonferroni may be better.
- Bonferroni better than Scheffé when number of contrasts about the same as treatment levels (or less)
- Best approach: compute Bonferroni, Tukey, Scheffé intervals and pick the smallest
  - This is a valid approach!
- Bonferroni can't be used for "snooping" unless one decides in advance the family of interest

51

Exercise: Nemotode Levels

Do nematodes affect plant growth? A botanist prepares 16 identical planting pots and adds different numbers of nematodes into the pots. Seedling growth (in mm) is recorded two weeks later; see **nematodes.txt** on Blackboard. Does a difference in seedling growth exist among the four levels?

- State the null and alternative hypotheses.
- Perform an ANOVA on these data. What do you conclude?
- Check the assumptions and conditions for an ANOVA. Are they satisfied?
- Perform a multiple comparisons test to determine which levels differ in terms of mean seedling growth.

52