# Lecture 10
# ANalysis Of VAriance (ANOVA)

Fall 2013

Prof. Yao Xie, yao.xie@isye.gatech.edu
H. Milton Stewart School of Industrial Systems & Engineering
Georgia Tech

# Test difference in two means

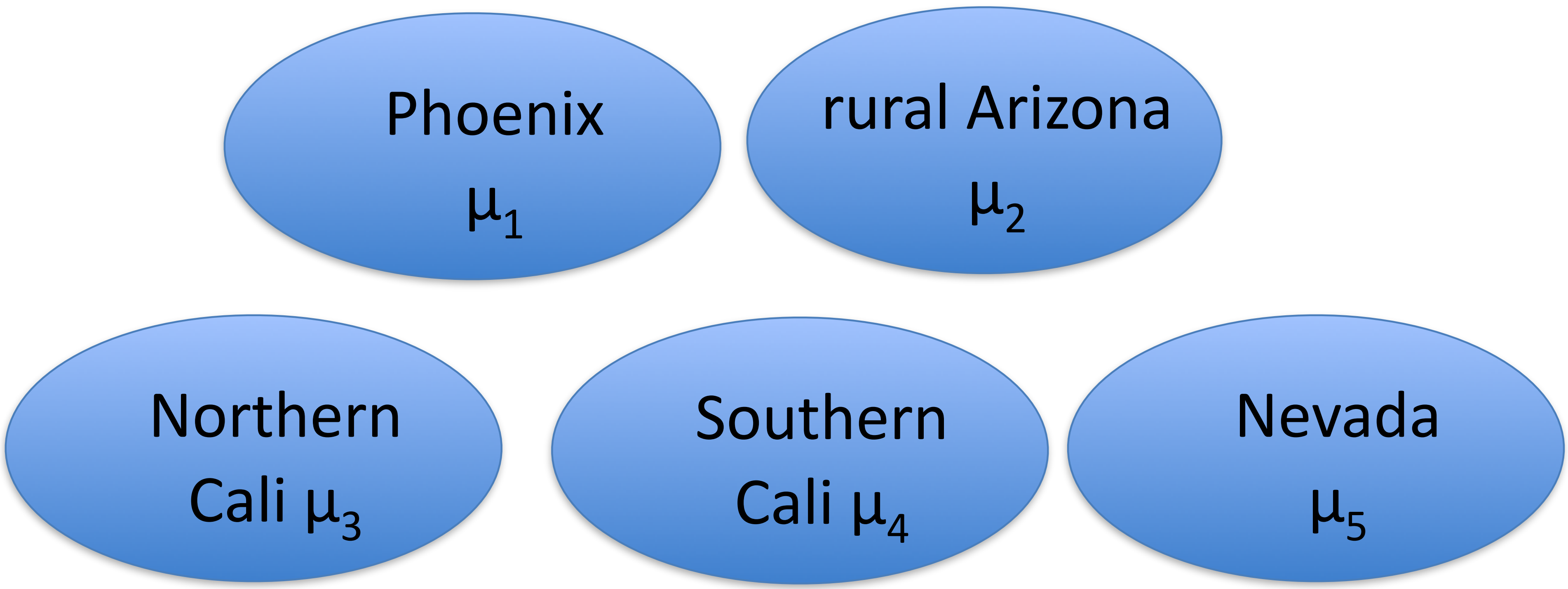
- Safety of drinking water (Arizona Republic, May 27, 2001)
- Water sampled from 10 communities in Pheonix
- And 10 communities from rural Arizona

Phoenix $\mu_1$

rural Arizona $\mu_2$

# Test difference in more than 2 mean

- In many cases we may want to compare means of more than two populations



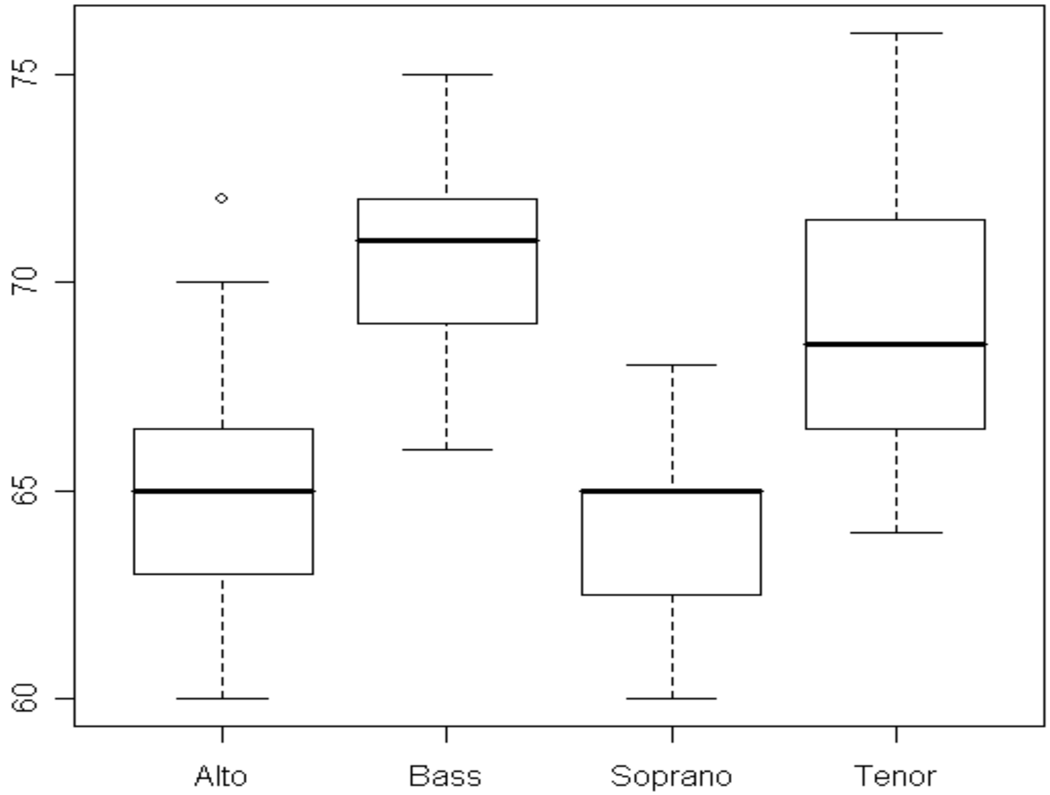

- In practice, many experiment requires comparing more than two levels
- Analysis of Variance (ANOVA)

# ANOVA Example 1: Voice Pitch and Height

Each singer in the NY Choral Society in 1979 self-reported his or her height to the nearest inch. Their voice parts in order from highest pitch to lowest pitch are **Soprano, Alto, Tenor, Bass**. The first two are typically sung by female voices and the last two by male voices.

One can examine how height varies across voice range, or make comparisons of sopranos and altos and separate comparisons of tenors and basses.

# Compare Singer Height by Voice Pitch




1. Is there a difference in the height by voice pitch?

2. Which singers are taller?

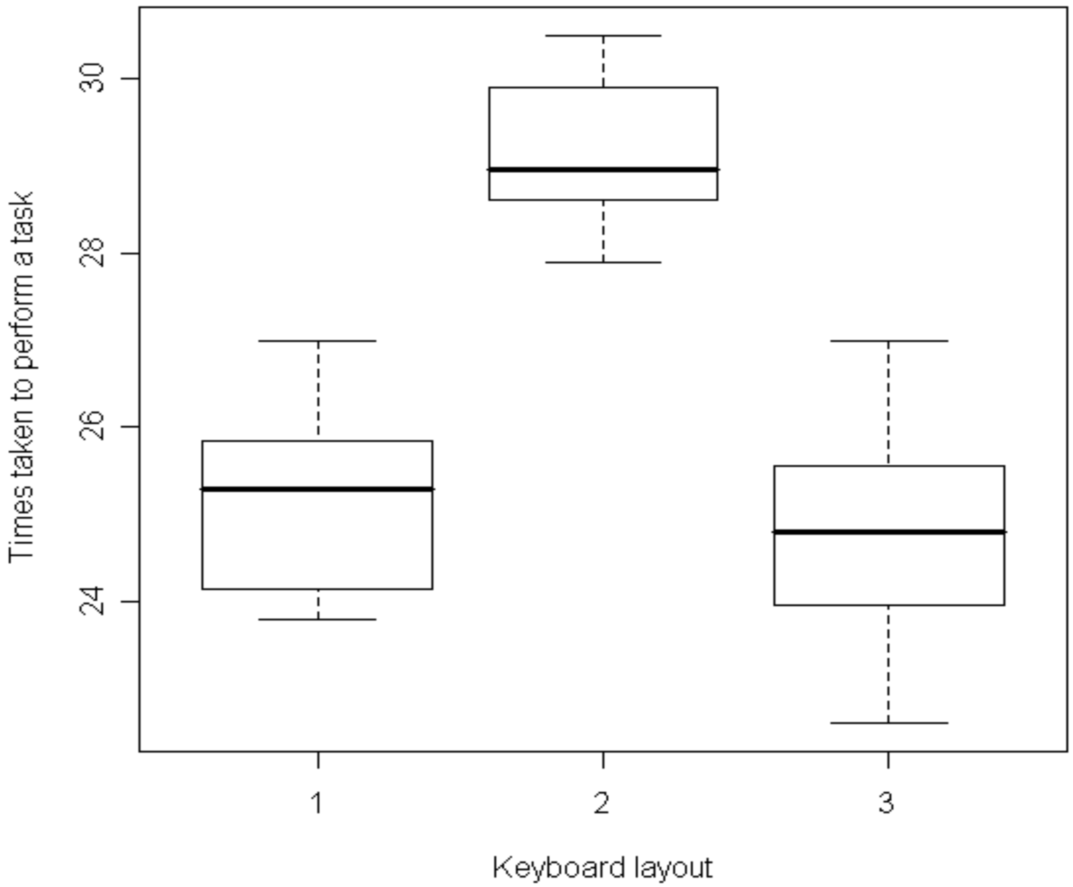
# ANOVA Example 2:  Keybord layout

Three different keyboard layouts are being compared in terms of typing speed.


TPSE WA 1971 P6. EA71A6

| Layout 1 | Layout 2 | Layout 3 |
|---|---|---|
| 23.8 | 30.2 | 27.0 |
| 25.6 | 29.9 | 25.4 |
| 24.0 | 29.1 | 25.6 |
| 25.1 | 28.8 | 24.2 |
| 25.5 | 29.1 | 24.8 |
| 26.1 | 28.6 | 24.0 |
| 23.8 | 28.3 | 25.5 |
| 25.7 | 28.7 | 23.9 |
| 24.3 | 27.9 | 22.6 |
| 26.0 | 30.5 | 26.0 |
| 24.6 | * | 23.4 |
| 27.0 | * | * |

# Operation Time by Keyboard Layout




1. Is there a difference in the time taken to perform a task?

2. Which layout is more effective?

# ANOVA Example 3: Carpet Wear

Six carpet fiber blends are tested for the amount of wear.

| Carpet Type | Number Sampled |
|---|---|
| 1 | 16 |
| 2 | 16 |
| 3 | 13 |
| 4 | 16 |
| 5 | 14 |
| 6 | 15 |

# Test means of multiple normals

Example:

A manufacturer of paper used for making grocery bags is interested in improving the tensile strength of the product. Product engineering thinks that tensile strength is a function of the hardwood concentration in the pulp and that the range of hardwood concentrations of practical interest is between 5 and 20%. A team of engineers responsible for the study decides to investigate four levels of hardwood concentration: 5%, 10%, 15%, and 20%.

Question of interest: Is hardwood concentration an important factor in improving tensile strength?

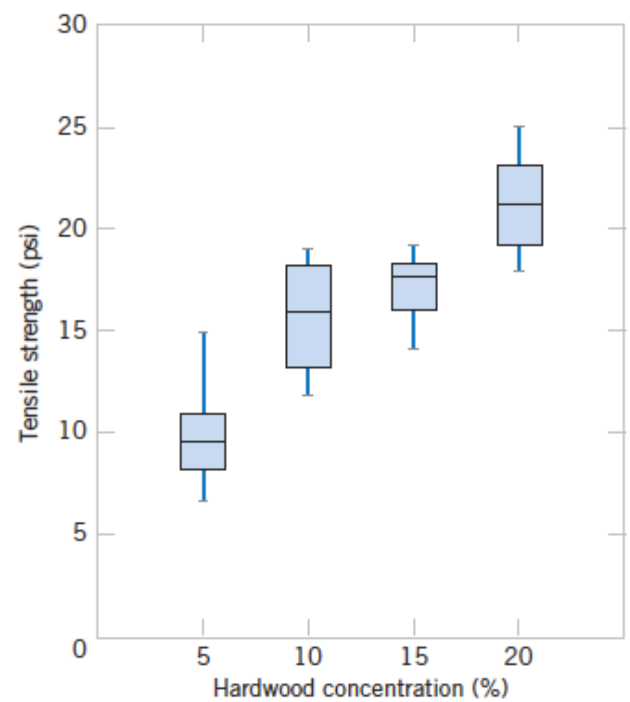
$$\begin{cases} H_0 : \mu_1 = \mu_2 = \mu_3 = \mu_4 \\ H_a : \text{at least one mean differs from others} \end{cases}$$

# Hypothesis Test on Means of Multiple Normal Distributions

Table 13-1 Tensile Strength of Paper (psi)

| Hardwood Concentration (%) | Observations 1 | 2 | 3 | 4 | 5 | 6 | Totals | Averages |
|---|---|---|---|---|---|---|---|---|
| 5 | 7 | 8 | 15 | 11 | 9 | 10 | 60 | 10.00 |
| 10 | 12 | 17 | 13 | 18 | 19 | 15 | 94 | 15.67 |
| 15 | 14 | 18 | 19 | 17 | 16 | 18 | 102 | 17.00 |
| 20 | 19 | 25 | 22 | 23 | 18 | 20 | 127 | 21.17 |
| | | | | | | | 383 | 15.96 |

$$\begin{cases} H_0 : \mu_1 = \mu_2 = \mu_3 = \mu_4 \\ H_a : \text{at least one mean differs from others} \end{cases}$$



(a)

# Hypothesis Test on Means of Multiple Normal Distributions

Table 13-2 Typical Data for a Single-Factor Experiment

| Treatment | Observations | | | | Totals | Averages |
|---|---|---|---|---|---|---|
| 1 | $y_{11}$ | $y_{12}$ | $\cdots$ | $y_{1n}$ | $y_{1\cdot}$ | $\bar{y}_{1\cdot}$ |
| 2 | $y_{21}$ | $y_{22}$ | $\cdots$ | $y_{2n}$ | $y_{2\cdot}$ | $\bar{y}_{2\cdot}$ |
| $\vdots$ | $\vdots$ | $\vdots$ | $\vdots\vdots\vdots$ | $\vdots$ | $\vdots$ | $\vdots$ |
| $a$ | $y_{a1}$ | $y_{a2}$ | $\cdots$ | $y_{an}$ | $y_{a\cdot}$ | $\bar{y}_{a\cdot}$ |
| | | | | | $y_{\cdot\cdot}$ | $\bar{y}_{\cdot\cdot}$ |

$i = 1,2,\cdots,a \longrightarrow$ Number of populations

$j = 1,2,\cdots,n \longrightarrow$ Sample size

$y_{ij}$ Observation *j* from population *i*

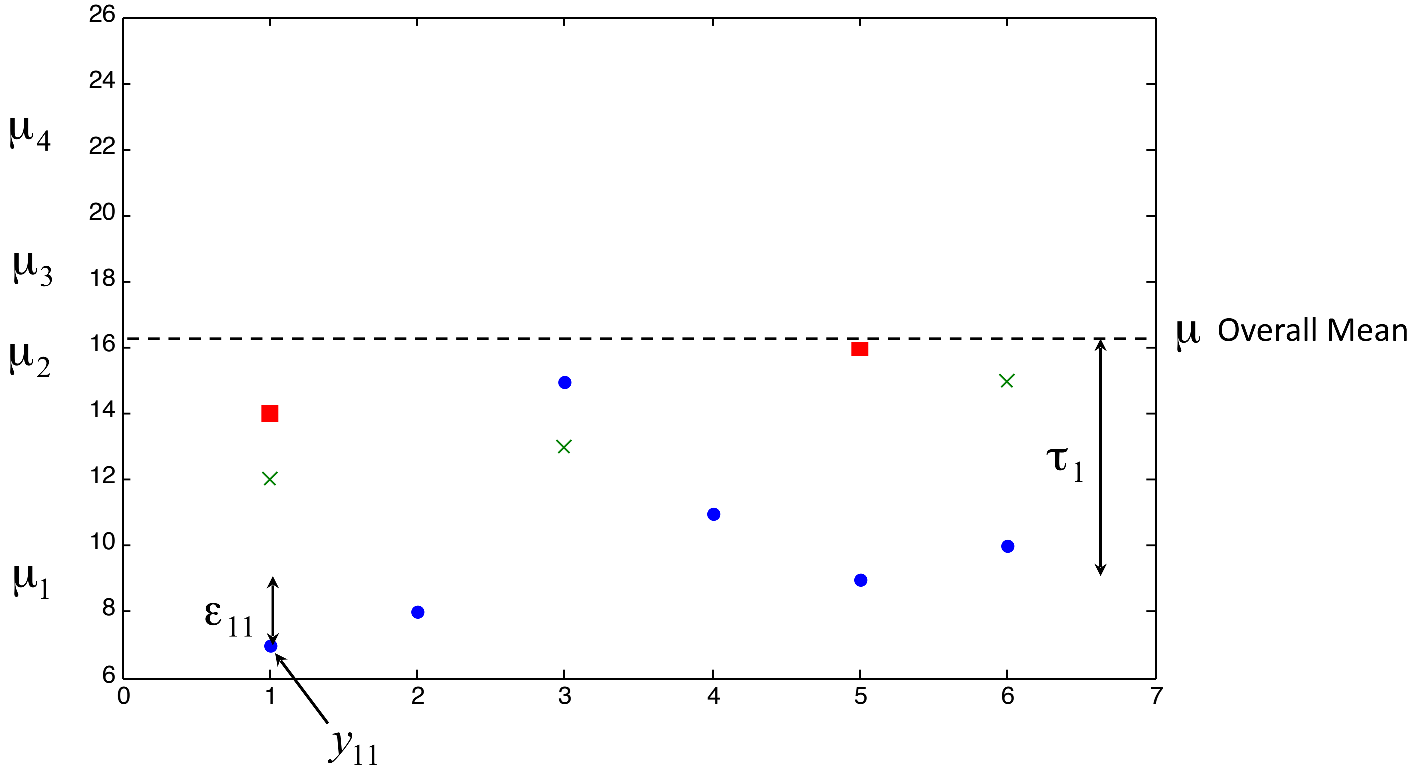
Population *i* Average

Population *i* Total $\quad y_{i\cdot} = \sum_{j=1}^{n} y_{ij} \qquad \bar{y}_{i\cdot} = y_{i\cdot}/n \qquad i = 1, 2, \ldots, a$

Grand Total $\quad y_{\cdot\cdot} = \sum_{i=1}^{a} \sum_{j=1}^{n} y_{ij} \qquad \bar{y}_{\cdot\cdot} = y_{\cdot\cdot}/N$ Grand Average

- The levels of the factor are sometimes called treatments.
- Each treatment has six observations or replicates.
- The runs are run in random order.
- This setting is known as Completely Randomized Single-Factor Experiment.

# Hypothesis Test on Means of Multiple Normal Distributions




$$Y_{ij} = \mu_i + \epsilon_{ij} \begin{cases} i = 1, 2, \ldots, a \\ j = 1, 2, \ldots, n \end{cases}$$

Linear Statistical Model (1)

$$Y_{ij} = \mu + \tau_i + \epsilon_{ij} \begin{cases} i = 1, 2, \ldots, a \\ j = 1, 2, \ldots, n \end{cases}$$

Linear Statistical Model (2)

$$\sum_{i=1}^{a} \tau_i = 0$$

# Hypothesis Test on Means of Multiple Normal Distributions

We wish to test the hypotheses:

$$\begin{cases} H_0 : \mu_1 = \mu_2 = \mu_3 = \mu_4 \\ H_a : \text{at least one mean differs from others} \end{cases}$$

We know that $\mu_i = \mu + \tau_i$

Therefore, the hypothesis test can be written as

$$H_0\colon \tau_1 = \tau_2 = \cdots = \tau_a = 0$$
$$H_1\colon \tau_i \neq 0 \quad \text{for at least one } i$$

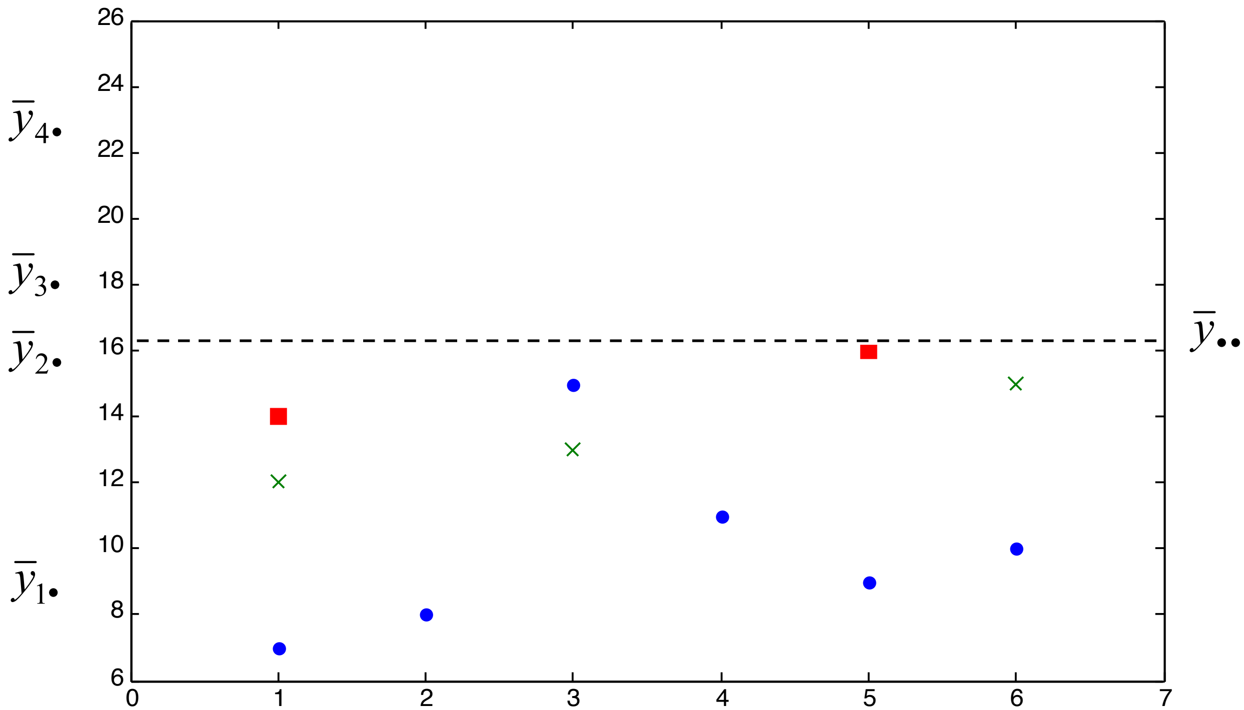
# Analysis of Variance (ANOVA)

$$\begin{cases} H_0: \mu_1 = \mu_2 = \mu_3 = \mu_4 \\ H_a: \text{at least one mean differs from others} \end{cases}$$

ANOVA partitions the total variability into two parts

Total Variations = Between-group Variations + Within-group Variations

SST = SSB + SSW or ($SST = SS_{treatments} + SS_{Error}$)




$$\sum_{i=1}^{a} \sum_{j=1}^{n} (y_{ij} - \bar{y}_{..})^2 = n \sum_{i=1}^{a} (\bar{y}_{i.} - \bar{y}_{..})^2 + \sum_{i=1}^{a} \sum_{j=1}^{n} (y_{ij} - \bar{y}_{i.})^2 \qquad (13\text{-}5)$$

# Calculation in ANOVA

Table 13-2 Typical Data for a Single-Factor Experiment

| Treatment | Observations | | | | Totals | Averages |
|---|---|---|---|---|---|---|
| 1 | $y_{11}$ | $y_{12}$ | $\cdots$ | $y_{1n}$ | $y_{1\cdot}$ | $\bar{y}_{1\cdot}$ |
| 2 | $y_{21}$ | $y_{22}$ | $\cdots$ | $y_{2n}$ | $y_{2\cdot}$ | $\bar{y}_{2\cdot}$ |
| $\vdots$ | $\vdots$ | $\vdots$ | $\vdots\vdots\vdots$ | $\vdots$ | $\vdots$ | $\vdots$ |
| $a$ | $y_{a1}$ | $y_{a2}$ | $\cdots$ | $y_{an}$ | $y_{a\cdot}$ | $\bar{y}_{a\cdot}$ |
| | | | | | $y_{\cdot\cdot}$ | $\bar{y}_{\cdot\cdot}$ |

$$SS_T = \sum_{i=1}^{a} \sum_{j=1}^{n} y_{ij}^2 - \frac{y_{\cdot\cdot}^2}{N}$$

$$= \sum_{i=1}^{a} \sum_{j=1}^{n} (y_{ij} - \bar{y}_{\cdot\cdot})^2$$

$$SS_{\text{Treatments}} = \sum_{i=1}^{a} \frac{y_{i\cdot}^2}{n} - \frac{y_{\cdot\cdot}^2}{N}$$

$$= n \sum_{i=1}^{a} (\bar{y}_{i\cdot} - \bar{y}_{\cdot\cdot})^2$$

$$SS_E = SS_T - SS_{\text{Treatments}}$$

# Analysis of Variance (ANOVA)

$$\begin{cases} H_0 : \mu_1 = \mu_2 = \cdots = \mu_a \\ H_a : \text{at least one mean differs from others} \end{cases}$$

ANOVA partitions the total variability into two parts

$$SST = SS_{treatments} + SS_{Error}$$

The appropriate test statistic is

$$F_0 = \frac{SS_{\text{Treatments}}/(a-1)}{SS_E/[a(n-1)]} = \frac{MS_{\text{Treatments}}}{MS_E} \tag{13-7}$$

We would reject $H_0$ if

$$F_0 > F_{\alpha,a-1,a(n-1)} \quad \text{or} \quad F_0 > F_{\alpha,a-1,N-a}$$

# ANOVA test

- Under $H_0$

$$F_0 = \frac{SS_{\text{Treatments}}/(a-1)}{SS_E/[a(n-1)]} = \frac{MS_{\text{Treatments}}}{MS_E} \tag{13-7}$$

is a F distribution with degree-of-freedom (a-1, a(n-1))

- Under $H_1$

the mean of the numerator is much bigger than the mean of the denominator

# F distribution

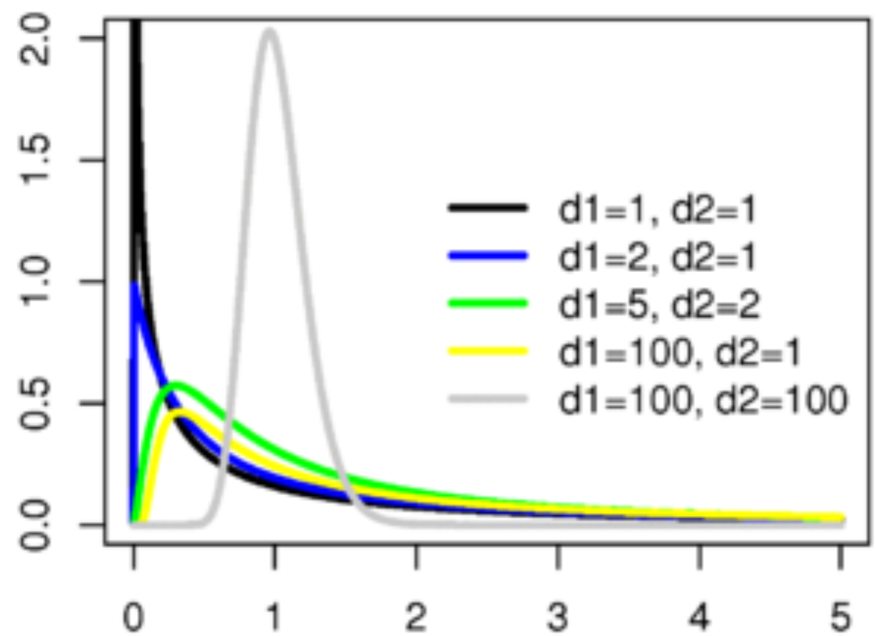
- A continuous distribution

$$f(x; d_1, d_2) = \frac{\sqrt{\frac{(d_1\,x)^{d_1}\,\,d_2^{d_2}}{(d_1\,x+d_2)^{d_1+d_2}}}}{x\,\mathrm{B}\!\left(\frac{d_1}{2}, \frac{d_2}{2}\right)}$$

$$= \frac{1}{\mathrm{B}\!\left(\frac{d_1}{2}, \frac{d_2}{2}\right)} \left(\frac{d_1}{d_2}\right)^{\frac{d_1}{2}} x^{\frac{d_1}{2}-1} \left(1 + \frac{d_1}{d_2}\,x\right)^{-\frac{d_1+d_2}{2}}$$

- mean = $\frac{d_2}{d_2 - 2}$

- we should reject $H_0$ when the statistic is large



# Analysis of Variance (ANOVA)

$$\begin{cases} H_0 : \mu_1 = \mu_2 = \cdots = \mu_a \\ H_a : \text{at least one mean differs from others} \end{cases}$$

$SST = SS_{treatments} + SS_{Error}$

The sums of squares computing formulas for the ANOVA with equal sample sizes in each treatment are

$$SS_T = \sum_{i=1}^{a} \sum_{j=1}^{n} y_{ij}^2 - \frac{y_{..}^2}{N} \tag{13-8}$$

and

$$SS_{\text{Treatments}} = \sum_{i=1}^{a} \frac{y_{i.}^2}{n} - \frac{y_{..}^2}{N} \tag{13-9}$$

The error sum of squares is obtained by subtraction as

$$SS_E = SS_T - SS_{\text{Treatments}} \tag{13-10}$$

# ANOVA Table

$$\begin{cases} H_0 : \mu_1 = \mu_2 = \cdots = \mu_a \\ H_a : \text{at least one mean differs from others} \end{cases}$$

Table 13-3 The Analysis of Variance for a Single-Factor Experiment, Fixed-Effects Model

| Source of Variation | Sum of Squares | Degrees of Freedom | Mean Square | $F_0$ |
|---|---|---|---|---|
| Treatments | $SS_{\text{Treatments}}$ | $a - 1$ | $MS_{\text{Treatments}}$ | $\dfrac{MS_{\text{Treatments}}}{MS_E}$ |
| Error | $SS_E$ | $a(n - 1)$ | $MS_E$ | |
| Total | $SS_T$ | $an - 1$ | | |

We would reject $H_0$ if

$$F_0 > F_{\alpha, a-1, a(n-1)}$$

# ANOVA

Example:

A manufacturer of paper used for making grocery bags is interested in improving the tensile strength of the product. Product engineering thinks that tensile strength is a function of the hardwood concentration in the pulp and that the range of hardwood concentrations of practical interest is between 5 and 20%. A team of engineers responsible for the study decides to investigate four levels of hardwood concentration: 5%, 10%, 15%, and 20%.

| Hardwood Concentration (%) | Observations 1 | 2 | 3 | 4 | 5 | 6 | Totals | Averages |
|---|---|---|---|---|---|---|---|---|
| 5 | 7 | 8 | 15 | 11 | 9 | 10 | 60 | 10.00 |
| 10 | 12 | 17 | 13 | 18 | 19 | 15 | 94 | 15.67 |
| 15 | 14 | 18 | 19 | 17 | 16 | 18 | 102 | 17.00 |
| 20 | 19 | 25 | 22 | 23 | 18 | 20 | 127 | 21.17 |
| | | | | | | | 383 | 15.96 |

$$SS_T = \sum_{i=1}^{a} \sum_{j=1}^{n} y_{ij}^2 - \frac{y_{..}^2}{N}$$

$$SS_{\text{Treatments}} = \sum_{i=1}^{a} \frac{y_{i.}^2}{n} - \frac{y_{..}^2}{N}$$

$$SS_E = SS_T - SS_{\text{Treatments}}$$

$a = 4$

$n = 6$

$N = 24$

$a = 4$

$n = 6$

$N = 24$

$$SS_T = \sum_{i=1}^{4} \sum_{j=1}^{6} y_{ij}^2 - \frac{y_{..}^2}{N}$$

$$= (7)^2 + (8)^2 + \cdots + (20)^2 - \frac{(383)^2}{24} = 512.96$$

$$SS_{\text{Treatments}} = \sum_{i=1}^{4} \frac{y_{i.}^2}{n} - \frac{y_{..}^2}{N}$$

$$= \frac{(60)^2 + (94)^2 + (102)^2 + (127)^2}{6} - \frac{(383)^2}{24}$$

$$= 382.79$$

$$SS_E = SS_T - SS_{\text{Treatments}}$$

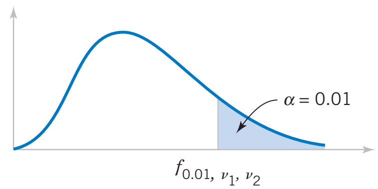
$$= 512.96 - 382.79 = 130.17$$

$a - 1 = 3$

$a(n-1) = 4 \times 5 = 20$

under $\mathrm{H}_0$ : statistic distributed as $F_{3,20}$

# F table




I Percentage Points $f_{\alpha,\nu_1,\nu_2}$ of the $F$ Distribution (*continued*)

$f_{0.01,\nu_1,\nu_2}$

| Degrees of freedom for the numerator ($\nu_1$) | | | | | | | | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 | 12 | 15 | 20 | 24 | 30 | 40 | 60 | 120 | |
| 4052 | 4999.5 | 5403 | 5625 | 5764 | 5859 | 5928 | 5982 | 6022 | 6056 | 6106 | 6157 | 6209 | 6235 | 6261 | 6287 | 6313 | 6339 | 6 |
| 98.50 | 99.00 | 99.17 | 99.25 | 99.30 | 99.33 | 99.36 | 99.37 | 99.39 | 99.40 | 99.42 | 99.43 | 99.45 | 99.46 | 99.47 | 99.47 | 99.48 | 99.49 | |
| 34.12 | 30.82 | 29.46 | 28.71 | 28.24 | 27.91 | 27.67 | 27.49 | 27.35 | 27.23 | 27.05 | 26.87 | 26.69 | 26.00 | 26.50 | 26.41 | 26.32 | 26.22 | |
| 21.20 | 18.00 | 16.69 | 15.98 | 15.52 | 15.21 | 14.98 | 14.80 | 14.66 | 14.55 | 14.37 | 14.20 | 14.02 | 13.93 | 13.84 | 13.75 | 13.65 | 13.56 | |
| 16.26 | 13.27 | 12.06 | 11.39 | 10.97 | 10.67 | 10.46 | 10.29 | 10.16 | 10.05 | 9.89 | 9.72 | 9.55 | 9.47 | 9.38 | 9.29 | 9.20 | 9.11 | |
| 13.75 | 10.92 | 9.78 | 9.15 | 8.75 | 8.47 | 8.26 | 8.10 | 7.98 | 7.87 | 7.72 | 7.56 | 7.40 | 7.31 | 7.23 | 7.14 | 7.06 | 6.97 | |
| 12.25 | 9.55 | 8.45 | 7.85 | 7.46 | 7.19 | 6.99 | 6.84 | 6.72 | 6.62 | 6.47 | 6.31 | 6.16 | 6.07 | 5.99 | 5.91 | 5.82 | 5.74 | |
| 11.26 | 8.65 | 7.59 | 7.01 | 6.63 | 6.37 | 6.18 | 6.03 | 5.91 | 5.81 | 5.67 | 5.52 | 5.36 | 5.28 | 5.20 | 5.12 | 5.03 | 4.95 | |
| 10.56 | 8.02 | 6.99 | 6.42 | 6.06 | 5.80 | 5.61 | 5.47 | 5.35 | 5.26 | 5.11 | 4.96 | 4.81 | 4.73 | 4.65 | 4.57 | 4.48 | 4.40 | |
| 10.04 | 7.56 | 6.55 | 5.99 | 5.64 | 5.39 | 5.20 | 5.06 | 4.94 | 4.85 | 4.71 | 4.56 | 4.41 | 4.33 | 4.25 | 4.17 | 4.08 | 4.00 | |
| 9.65 | 7.21 | 6.22 | 5.67 | 5.32 | 5.07 | 4.89 | 4.74 | 4.63 | 4.54 | 4.40 | 4.25 | 4.10 | 4.02 | 3.94 | 3.86 | 3.78 | 3.69 | |
| 9.33 | 6.93 | 5.95 | 5.41 | 5.06 | 4.82 | 4.64 | 4.50 | 4.39 | 4.30 | 4.16 | 4.01 | 3.86 | 3.78 | 3.70 | 3.62 | 3.54 | 3.45 | |
| 9.07 | 6.70 | 5.74 | 5.21 | 4.86 | 4.62 | 4.44 | 4.30 | 4.19 | 4.10 | 3.96 | 3.82 | 3.66 | 3.59 | 3.51 | 3.43 | 3.34 | 3.25 | |
| 8.86 | 6.51 | 5.56 | 5.04 | 4.69 | 4.46 | 4.28 | 4.14 | 4.03 | 3.94 | 3.80 | 3.66 | 3.51 | 3.43 | 3.35 | 3.27 | 3.18 | 3.09 | |
| 8.68 | 6.36 | 5.42 | 4.89 | 4.36 | 4.32 | 4.14 | 4.00 | 3.89 | 3.80 | 3.67 | 3.52 | 3.37 | 3.29 | 3.21 | 3.13 | 3.05 | 2.96 | |
| 8.53 | 6.23 | 5.29 | 4.77 | 4.44 | 4.20 | 4.03 | 3.89 | 3.78 | 3.69 | 3.55 | 3.41 | 3.26 | 3.18 | 3.10 | 3.02 | 2.93 | 2.84 | |
| 8.40 | 6.11 | 5.18 | 4.67 | 4.34 | 4.10 | 3.93 | 3.79 | 3.68 | 3.59 | 3.46 | 3.31 | 3.16 | 3.08 | 3.00 | 2.92 | 2.83 | 2.75 | |
| 8.29 | 6.01 | 5.09 | 4.58 | 4.25 | 4.01 | 3.84 | 3.71 | 3.60 | 3.51 | 3.37 | 3.23 | 3.08 | 3.00 | 2.92 | 2.84 | 2.75 | 2.66 | |
| 8.18 | 5.93 | 5.01 | 4.50 | 4.17 | 3.94 | 3.77 | 3.63 | 3.52 | 3.43 | 3.30 | 3.15 | 3.00 | 2.92 | 2.84 | 2.76 | 2.67 | 2.58 | |
| 8.10 | 5.85 | 4.94 | 4.43 | 4.10 | 3.87 | 3.70 | 3.56 | 3.46 | 3.37 | 3.23 | 3.09 | 2.94 | 2.86 | 2.78 | 2.69 | 2.61 | 2.52 | |
| 8.02 | 5.78 | 4.87 | 4.37 | 4.04 | 3.81 | 3.64 | 3.51 | 3.40 | 3.31 | 3.17 | 3.03 | 2.88 | 2.80 | 2.72 | 2.64 | 2.55 | 2.46 | |
| 7.95 | 5.72 | 4.82 | 4.31 | 3.99 | 3.76 | 3.59 | 3.45 | 3.35 | 3.26 | 3.12 | 2.98 | 2.83 | 2.75 | 2.67 | 2.58 | 2.50 | 2.40 | |
| 7.88 | 5.66 | 4.76 | 4.26 | 3.94 | 3.71 | 3.54 | 3.41 | 3.30 | 3.21 | 3.07 | 2.93 | 2.78 | 2.70 | 2.62 | 2.54 | 2.45 | 2.35 | |
| 7.82 | 5.61 | 4.72 | 4.22 | 3.90 | 3.67 | 3.50 | 3.36 | 3.26 | 3.17 | 3.03 | 2.89 | 2.74 | 2.66 | 2.58 | 2.49 | 2.40 | 2.31 | |
| 7.77 | 5.57 | 4.68 | 4.18 | 3.85 | 3.63 | 3.46 | 3.32 | 3.22 | 3.13 | 2.99 | 2.85 | 2.70 | 2.62 | 2.54 | 2.45 | 2.36 | 2.27 | |
| 7.72 | 5.53 | 4.64 | 4.14 | 3.82 | 3.59 | 3.42 | 3.29 | 3.18 | 3.09 | 2.96 | 2.81 | 2.66 | 2.58 | 2.50 | 2.42 | 2.33 | 2.23 | |
| 7.68 | 5.49 | 4.60 | 4.11 | 3.78 | 3.56 | 3.39 | 3.26 | 3.15 | 3.06 | 2.93 | 2.78 | 2.63 | 2.55 | 2.47 | 2.38 | 2.29 | 2.20 | |
| 7.64 | 5.45 | 4.57 | 4.07 | 3.75 | 3.53 | 3.36 | 3.23 | 3.12 | 3.03 | 2.90 | 2.75 | 2.60 | 2.52 | 2.44 | 2.35 | 2.26 | 2.17 | |
| 7.60 | 5.42 | 4.54 | 4.04 | 3.73 | 3.50 | 3.33 | 3.20 | 3.09 | 3.00 | 2.87 | 2.73 | 2.57 | 2.49 | 2.41 | 2.33 | 2.23 | 2.14 | |
| 7.56 | 5.39 | 4.51 | 4.02 | 3.70 | 3.47 | 3.30 | 3.17 | 3.07 | 2.98 | 2.84 | 2.70 | 2.55 | 2.47 | 2.39 | 2.30 | 2.21 | 2.11 | |
| 7.31 | 5.18 | 4.31 | 3.83 | 3.51 | 3.29 | 3.12 | 2.99 | 2.89 | 2.80 | 2.66 | 2.52 | 2.37 | 2.29 | 2.20 | 2.11 | 2.02 | 1.92 | |
| 7.08 | 4.98 | 4.13 | 3.65 | 3.34 | 3.12 | 2.95 | 2.82 | 2.72 | 2.63 | 2.50 | 2.35 | 2.20 | 2.12 | 2.03 | 1.94 | 1.84 | 1.73 | |
| 6.85 | 4.79 | 3.95 | 3.48 | 3.17 | 2.96 | 2.79 | 2.66 | 2.56 | 2.47 | 2.34 | 2.19 | 2.03 | 1.95 | 1.86 | 1.76 | 1.66 | 1.53 | |
| 6.63 | 4.61 | 3.78 | 3.32 | 3.02 | 2.80 | 2.64 | 2.51 | 2.41 | 2.32 | 2.18 | 2.04 | 1.88 | 1.79 | 1.70 | 1.59 | 1.47 | 1.32 | |

$$F_{0.01,3,20} = 4.94$$

# ANOVA table we come up with

**Table 13-4** ANOVA for the Tensile Strength Data

| Source of Variation | Sum of Squares | Degrees of Freedom | Mean Square | $f_0$ | $P$-value |
|---|---|---|---|---|---|
| Hardwood concentration | 382.79 | 3 | 127.60 | $\boxed{19.60}$ | 3.59 E-6 |
| Error | 130.17 | 20 | 6.51 | | |
| Total | 512.96 | 23 | | | |

$F_{0.01,3,20} = 4.94$ $\Rightarrow$ Reject H0

p-value: $P = P(F_{3,20} > 19.60) \simeq 3.59 \times 10^{-6}$

R command

```
> p <- 1-pf(19.6,3,20)
> p
[1] 3.599599e-06
```

# Multiple comparisons following the ANOVA

- ANOVA only tells **whether or not** the means are the same
- to determine which means are different: multiple comparison methods
- Fisher's least significant difference (LSD) method
  - compute pairwise group sample means, and claim the groups to be different if their sample mean difference satisfies: $|\bar{y}_{i\cdot} - \bar{y}_{j\cdot}| > \text{LSD}$

$$MS_E = SS_E/[a(n-1)]$$

$$\text{LSD} = t_{\alpha/2,a(n-1)}\sqrt{\frac{2MS_E}{n}}$$