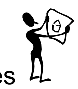
Chapter 24

One-Way Analysis of Variance: Comparing Several Means

Comparing Means

- Chapter 18: compared the means of <u>two</u> populations or the mean responses to <u>two</u> treatments in an experiment
  - two-sample *t* tests
- This chapter: compare <u>any number</u> of means
  - Analysis of Variance
    - Remember: we are comparing <u>**means**</u> even though the procedure is Analysis of <u>Variance</u>

Case Study

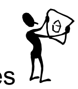


Gas Mileage for Classes of Vehicles

Data from the Environmental Protection Agency's *Model Year 2003 Fuel Economy Guide*, www.fueleconomy.gov.

Do SUVs and trucks have lower gas mileage than midsize cars?

Case Study



Gas Mileage for Classes of Vehicles
Data collection

- Response variable: gas mileage (mpg)
- Groups: vehicle classification
  - 31 midsize cars
  - 31 SUVs
  - 14 standard-size pickup trucks

Case Study

Gas Mileage for Classes of Vehicles

Data

TABLE 22.1 Highway gas mileage for 2003 model vehicles

Midsize Cars		Sport-Utility Vehicles		Standard Pickup Trucks	
Model	MPG	Model	MPG	Model	MPG
Acura 3.5TL	29	Cadillac Escalade	18	Chevrolet Silverado	20
Audi A6	27	Chevrolet Avalanche	18	Dodge Dakota	19
Audi A8	25	Chevrolet Blazer	23	Dodge Ram	20
Buick Century	29	Chevrolet Suburban	18	Ford Explorer	20
Buick Regal	27	Chevrolet Tahoe	18	Ford F150	20
Cadillac CTS	26	Chevrolet Tracker	26	Ford Ranger	26
Cadillac Seville	27	Chevrolet Trailblazer	22	GMC Sierra	20
Chevrolet Monte Carlo	32	Chrysler PT Cruiser	25	Lincoln Blackwood	17
Chrysler Sebring	30	Dodge Durango	19	Mazda B2300	26
Honda Accord	33	Ford Escape	28	Mazda B3000	22
Hyundai Sonata	30	Ford Expedition	19	Mazda B4000	21
Hyundai XG350	26	Ford Explorer	19	Nissan Frontier	23
Infiniti I35	26	Honda CR-V	28	Toyota Tacoma	25
Jaguar S-Type	26	Hyundai Santa Fe	27	Toyota Tundra	19
Jaguar Vanden Plas	24	Infiniti QX4	21		
Kia Optima	30	Isuzu Axiom	21		
Lexus ES300	29	Isuzu Rodeo	23		
Lexus GS300	25	Jeep Grand Cherokee	21		
Lincoln LS	26	Jeep Liberty	24		
Mercedes-Benz E320	27	Kia Sorento	20		
Mercedes-Benz E500	23	Lincoln Navigator	17		
Mitsubishi Diamante	25	Mazda Tribute	28		
Mitsubishi Galant	27	Mitsubishi Montero	22		
Nissan Altima	29	Nissan Pathfinder	21		
Nissan Maxima	26	Nissan Xterra	24		
Saab 9-5	29	Saturn Vue	28		
Saturn L200	32	Suzuki Vitara	25		
Saturn L300	29	Toyota 4Runner	20		
Toyota Camry	32	Toyota Highlander	27		
Volkswagen Passat	31	Toyota Rav4	29		
Volvo S80	28	Volvo XC 90	24		

Case Study

Gas Mileage for Classes of Vehicles

Data

<u>Means ($\bar{x}$ s)</u>:
Midsize: 27.903
SUV: 22.677
Pickup: 21.286

Midsize		SUV		Pickup	
17		17	0	17	0
18		18	0000	18	
19		19	000	19	00
20		20	00	20	00000
21		21	0000	21	0
22		22	00	22	0
23	0	23	00	23	0
24	0	24	000	24	
25	000	25	00	25	0
26	000000	26	0	26	00
27	00000	27	00	27	
28	0	28	0000	28	
29	000000	29	0	29	
30	000	30		30	
31	0	31		31	
32	000	32		32	
33	0	33		33	

Case Study

Gas Mileage for Classes of Vehicles

Data analysis

Means ($\bar{x}$ s):
Midsize: 27.903
SUV: 22.677
Pickup: 21.286

- Mean gas mileage for SUVs and pickups appears less than for midsize cars
- Are these differences statistically significant?

BPS - 5th Ed. Chapter 24 7

Case Study

Gas Mileage for Classes of Vehicles

Data analysis

Means ($\bar{x}$ s):
Midsize: 27.903
SUV: 22.677
Pickup: 21.286

Null hypothesis:
The true means (for gas mileage) are the same for all groups (the three vehicle classifications)

For example, could look at separate *t* tests to compare each pair of means to see if they are different:

27.903 vs. 22.677, 27.903 vs. 21.286, & 22.677 vs. 21.286

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

Problem of *multiple comparisons*!

BPS - 5th Ed. Chapter 24 8

Multiple Comparisons

- Problem of how to do many comparisons at the same time with some overall measure of confidence in all the conclusions
- Two steps:
  - overall test to test for any differences
  - follow-up analysis to decide which groups differ and how large the differences are
- Follow-up analyses can be quite complex; we will look at only the overall test for a difference in several means, and examine the data to make follow-up conclusions

BPS - 5th Ed. Chapter 24 9

Analysis of Variance *F* Test

- $\mathbf{H_0}$: $\mu_1 = \mu_2 = \mu_3$
- $\mathbf{H_a}$: not all of the means are the same
- To test H_0, compare how much variation exists among the sample means (how much the $\bar{x}$ s differ) with how much variation exists within the samples from each group
  - is called the **analysis of variance *F* test**
  - test statistic is an *F* statistic
    - use *F* distribution (*F* table) to find *P*-value
  - analysis of variance is abbreviated ***ANOVA***

BPS - 5th Ed. Chapter 24 10

Case Study

Gas Mileage for Classes of Vehicles

Using Technology

One-way ANOVA: Midsize, SUV, Pickup

```
Analysis of Variance
Source     DF        SS        MS         F        P
Factor      2    606.45    303.22     31.61    0.000
Error      73    700.34      9.59
Total      75   1306.79

                                  Individual 95% CIs For Mean
                                  Based on Pooled StDev
Level       N      Mean     StDev ---+---------+---------+---------+----
Midsize    31    27.903     2.561                          (---*---)
SUV        31    22.677     3.673       (---*---)
Pickup     14    21.286     2.758 (---*---)
                                  ---+---------+---------+---------+----
Pooled StDev =    3.097             21.0      24.0      27.0      30.0
```

P-value<.05 significant differences

Follow-up analysis

BPS - 5th Ed. Chapter 24 11

Case Study

Gas Mileage for Classes of Vehicles

Data analysis

- *F* = 31.61
- *P*-value = 0.000 (rounded) (is <0.001)
  - there is significant evidence that the three types of vehicle do not all have the same gas mileage
  - from the confidence intervals (and looking at the original data), we see that SUVs and pickups have similar fuel economy and both are distinctly poorer than midsize cars

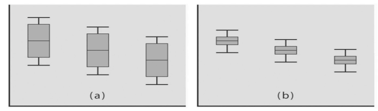
BPS - 5th Ed. Chapter 24 12

ANOVA Idea

- *ANOVA tests whether several populations have the same mean by comparing how much variation exists <u>among the sample means</u> (how much the $\bar{x}$ s differ) with how much variation exists <u>within the samples</u> from each group*
  - the decision is not based only on how far apart the sample means are, but instead on how far apart they are *<u>relative to the variability of the individual observations within each group</u>*

BPS - 5th Ed. Chapter 24 13

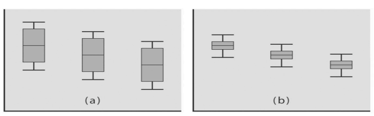
ANOVA Idea



- Sample means for the three samples are the same for each set (a) and (b) of boxplots (shown by the center of the boxplots)
  - variation among sample means for (a) is identical to (b)
- Less spread in the boxplots for (b)
  - variation among the individuals within the three samples is much less for (b)

BPS - 5th Ed. Chapter 24 14

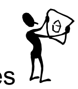
ANOVA Idea



- CONCLUSION: the samples in (b) contain a larger amount of variation among the sample means *<u>relative to</u>* the amount of variation within the samples, so ANOVA will find *<u>more significant differences among the means in (b)</u>*
  - assuming equal sample sizes here for (a) and (b)
  - larger samples will find more significant differences

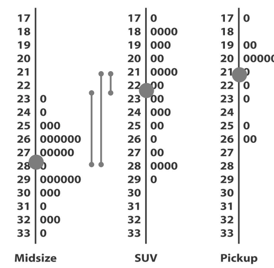
BPS - 5th Ed. Chapter 24 15

Case Study



Gas Mileage for Classes of Vehicles

Variation among sample means (how much the $\bar{x}$ s differ from each other)

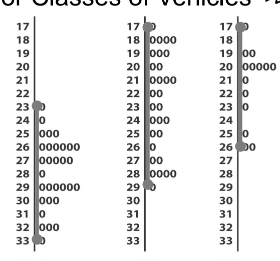


BPS - 5th Ed. Chapter 24 16

Case Study

Gas Mileage for Classes of Vehicles

Variation within the individual samples



BPS - 5th Ed. Chapter 24 17

ANOVA *F* Statistic

- To determine statistical significance, we need a test statistic that we can calculate
  - ANOVA *F* Statistic:

$$F = \frac{\text{variation among the sample means}}{\text{variation among individuals in the same sample}}$$

  - must be zero or positive
    - only zero when all sample means are identical
    - gets larger as means move further apart
  - large values of *F* are evidence against H_0: *equal means*
  - the *F* test is upper one-sided (like the chi-square test)

BPS - 5th Ed. Chapter 24 18

ANOVA *F* Test

- Calculate value of *F* statistic
  - by hand (cumbersome)
  - using technology (computer software, etc.)
- Find *P*-value in order to reject or fail to reject H_0
  - ***F* table** (not provided in book. Will provide on website)
  - from computer output
- If significant relationship exists (small *P*-value):
  - follow-up analysis
    - observe differences in sample means in original data
    - formal multiple comparison procedures (not covered here)

ANOVA *F* Test

- *F* test for comparing I populations, with an SRS of size n_i from the i^{th} population (thus giving $N = n_1+n_2+\cdots+n_I$ total observations) uses critical values from an *F* distribution with the following *numerator* and *denominator degrees of freedom*:
  - ***numerator df = I - 1***
  - ***denominator df = N - I***
- *P*-value is the area to the right of *F* under the density curve of the *F* distribution

Case Study

Gas Mileage for Classes of Vehicles
Using Technology

One-way ANOVA: Midsize, SUV, Pickup

```
Analysis of Variance
Source    DF      SS      MS       F      P
Factor     2   606.45  303.22  31.61  0.000
Error     73   700.34    9.59
Total     75  1306.79

                               Individual 95% CIs For Mean
                               Based on Pooled StDev
Level     N    Mean   StDev  ----+---------+---------+---------+----
Midsize  31  27.903   2.561                          (---*---)
SUV      31  22.677   3.673        (---*---)
Pickup   14  21.286   2.758  (---*---)
                             ----+---------+---------+---------+----
Pooled StDev =  3.097            21.0      24.0      27.0      30.0
```


Case Study

Gas Mileage for Classes of Vehicles

***F* = 31.61** $\quad$ ***I* = 3** classes of vehicle

n_1 = 31 midsize, n_2 = 31 SUVs, n_3 = 14 trucks

***N* = 31 + 31 + 14 = 76**

df_{num} = (I-1) = (3-1) = **2**

df_{den} = (N-I) = (76-3) = **73**

***P*-value** from technology output is **0.000**. This probability is not 0, but is very close to 0 and is smaller than 0.001, the smallest value the technology can record.

** *P*-value < .05, so we conclude significant differences **

ANOVA Model, Assumptions

- Conditions required for using ANOVA *F* test to compare population means
  1) have I *independent SRSs*, one from each population.
  2) the i^{th} population has a *Normal distribution* with unknown mean μ_i (means may be different).
  3) all of the populations have the *same standard deviation* σ, whose value is unknown.

Robustness

- ANOVA *F* test is not very sensitive to lack of Normality (is **robust**)
  - what matters is Normality of the sample means
  - ANOVA becomes safer as the sample sizes get larger, due to the Central Limit Theorem
  - if there are no outliers and the distributions are roughly symmetric, can safely use ANOVA for sample sizes as small as 4 or 5

Robustness

- ANOVA *F* test is not too sensitive to violations of the assumption of equal standard deviations
  - especially when all samples have the same or similar sizes and no sample is very small
  - statistical tests for equal standard deviations are very sensitive to lack of Normality (not practical)
  - check that sample standard deviations are similar to each other (next slide)

Checking Standard Deviations

- The results of ANOVA *F* tests are approximately correct when the largest sample standard deviation (*s*) is no more than twice as large as the smallest sample standard deviation

Case Study

Gas Mileage for Classes of Vehicles

One-way ANOVA: Midsize, SUV, Pickup

```
Analysis of Variance
Source     DF        SS        MS
Factor      2    606.45    303.22
Error      73    700.34      9.59
Total      75   1306.79



Level       N      Mean    StDev
Midsize    31    27.903    2.561
SUV        31    22.677    3.673
Pickup     14    21.286    2.758


Pooled StDev =    3.097
```

$s_1 = 2.561$
$s_2 = 3.673$
$s_3 = 2.758$

$$\frac{\text{largest } s}{\text{smallest } s} = \frac{3.673}{2.561} = 1.434$$

$\Rightarrow$ safe to use ANOVA *F* test

ANOVA Details

- ANOVA *F* statistic:

$$F = \frac{\textit{variation among the sample means}}{\textit{variation among individuals in the same sample}}$$

  - the measures of variation in the numerator and denominator are ***mean squares***
    - general form of a sample variance
    - ordinary s^2 is "an average (or mean) of the squared deviations of observations from their mean"

ANOVA Details

- Numerator: Mean Square for Groups (**MSG**)
  - an average of the I squared deviations of the means of the samples from the overall mean $\bar{X}$

$$MSG = \frac{n_1(\bar{x}_1 - \bar{x})^2 + n_2(\bar{x}_2 - \bar{x})^2 + \cdots + n_I(\bar{x}_I - \bar{x})^2}{I - 1}$$

- n_i is the number of observations in the i^{th} group
- $$\bar{x} = \frac{n_1\bar{x}_1 + n_2\bar{x}_2 + \cdots + n_I\bar{x}_I}{N}$$

ANOVA Details

- Denominator: Mean Square for Error (**MSE**)
  - an average of the individual sample variances (s_i^2) within each of the I groups

$$MSE = \frac{(n_1 - 1)s_1^2 + (n_2 - 1)s_2^2 + \cdots + (n_I - 1)s_I^2}{N - I}$$

- MSE is also called the *pooled sample variance*, written as s_p^2 (s_p is the *pooled standard deviation*)
- **s_p^2 estimates the common variance σ^2**

ANOVA Details

- the numerators of the mean squares are called the ***sums of squares*** (*SSG* and *SSE*)
- the denominators of the mean squares are the two degrees of freedom for the *F* test, (I-1) and (N-I)
- usually results of ANOVA are presented in an ***ANOVA table***, which gives the *source of variation*, *df*, *SS*, *MS*, and *F* statistic
  - ANOVA *F* statistic: $F = \dfrac{MSG}{MSE} = \dfrac{SSG/dfG}{SSE/dfE}$

Case Study

Gas Mileage for Classes of Vehicles Using Technology

One-way ANOVA: Midsize, SUV, Pickup

```
Analysis of Variance
Source     DF        SS        MS        F        P
Factor      2    606.45    303.22    31.61    0.000
Error      73    700.34      9.59
Total      75   1306.79
```

For detailed calculations, see Examples 24.7 and 24.8 on pages 652-654 of the textbook.

THE ANOVA *F* TEST **Summary**

Draw an independent SRS from each of I populations. The ith population has the $N(\mu_i, \sigma)$ distribution, where σ is the common standard deviation in all the populations. The ith sample has size n_i, sample mean $\overline{x}_i$, and sample standard deviation s_i.

The **ANOVA *F* statistic** tests the null hypothesis that all I populations have the same mean:

$$H_0: \mu_1 = \mu_2 = \cdots = \mu_I$$
$$H_a: \text{not all of the } \mu_i \text{ are equal}$$

The statistic is

$$F = \frac{\text{MSG}}{\text{MSE}}$$

The numerator of F is the **mean square for groups**

$$\text{MSG} = \frac{n_1(\overline{x}_1 - \overline{x})^2 + n_2(\overline{x}_2 - \overline{x})^2 + \cdots + n_I(\overline{x}_I - \overline{x})^2}{I - 1}$$

The denominator of F is the **mean square for error**

$$\text{MSE} = \frac{(n_1 - 1)s_1^2 + (n_2 - 1)s_2^2 + \cdots + (n_I - 1)s_I^2}{N - I}$$

When H_0 is true, F has the **F distribution** with $I - 1$ and $N - I$ degrees of freedom.

ANOVA Confidence Intervals

- **Confidence interval** for the mean μ_i of any group:

$$\overline{x}_i \pm t^* \frac{s_p}{\sqrt{n_i}}$$

  - t^* is the critical value from the t distribution with N-I degrees of freedom
  - s_p (pooled standard deviation) is used to estimate σ because it is better than any individual s_i

Case Study

Gas Mileage for Classes of Vehicles Using Technology

One-way ANOVA: Midsize, SUV, Pickup

```
Analysis of Variance
Source     DF        SS        MS        F        P
Factor      2    606.45    303.22    31.61    0.000
Error      73    700.34      9.59
Total      75   1306.79
                                  Individual 95% CIs For Mean
                                  Based on Pooled StDev
Level      N      Mean     StDev  ---+---------+---------+---------+----
Midsize   31    27.903     2.561                         (---*---)
SUV       31    22.677     3.673       (---*---)
Pickup    14    21.286     2.758  (---*---)
                                  ---+---------+---------+---------+----
Pooled StDev =    3.097              21.0      24.0      27.0      30.0
```