Math 362: Mathematical Statistics II

Le Chen
le.chen@emory.edu

Emory University
Atlanta, GA

Last updated on April 24, 2021

2021 Spring

Chapter 12. The Analysis of Variance

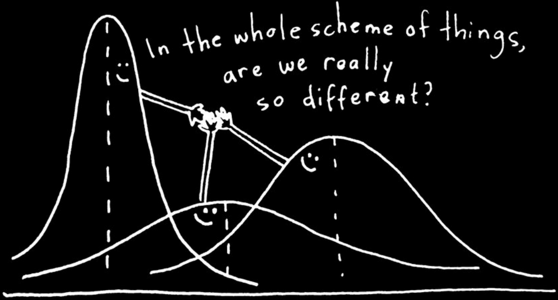
Plan

§ 12.1 Introduction

§ 12.2 The F Test

§ 12.3 Multiple Comparisons: Turkey's Method

§ 12.4 Testing Subhypotheses with Contrasts


In the whole scheme of things,
are we really
so different?

E.g. 1 Study the relation between smoking and heart rates.

Generations of athletes have been cautioned that cigarette smoking impedes performance. One measure of the truth of that warning is the effect of smoking on heart rate. In one study, six nonsmokers, six light smokers, six moderate smokers, and six heavy smokers each engaged in sustained physical exercise. Table 8.1.1 lists their heart rates after they had rested for three minutes.

Show whether smoking affects heart rates at $\alpha = 0.05$.

E.g. 1 Study the relation between smoking and heart rates.

Generations of athletes have been cautioned that cigarette smoking impedes performance. One measure of the truth of that warning is the effect of smoking on heart rate. In one study, six nonsmokers, six light smokers, six moderate smokers, and six heavy smokers each engaged in sustained physical exercise. Table 8.1.1 lists their heart rates after they had rested for three minutes.

Table 8.1.1 Heart Rates

	Nonsmokers	Light Smokers	Moderate Smokers	Heavy Smokers
	69	55	66	91
	52	60	81	72
	71	78	70	81
	58	58	77	67
	59	62	57	95
	65	66	79	84
Averages:	62.3	63.2	71.7	81.7

Show whether smoking affects heart rates at $\alpha = 0.05$.

E.g. 2 A certain fraction of antibiotics injected into the bloodstream are "bound" to serum proteins. This phenomenon bears directly on the effectiveness of the medication, because the binding decreases the systemic uptake of the drug. Table below lists the binding percentages in bovine serum measured for five widely prescribed antibiotics. Which antibiotics have similar binding properties, and which are different?

E.g. 2 A certain fraction of antibiotics injected into the bloodstream are "bound" to serum proteins. This phenomenon bears directly on the effectiveness of the medication, because the binding decreases the systemic uptake of the drug. Table below lists the binding percentages in bovine serum measured for five widely prescribed antibiotics. Which antibiotics have similar binding properties, and which are different?

Table 12.3.1

	Penicillin G	Tetra-cycline	Strepto-mycin	Erythro-mycin	Chloram-phenicol
	29.6	27.3	5.8	21.6	29.2
	24.3	32.6	6.2	17.4	32.8
	28.5	30.8	11.0	18.3	25.0
	32.0	34.8	8.3	19.0	24.2
$T_{.j}$	114.4	125.5	31.3	76.3	111.2
$\overline{Y}_{.j}$	28.6	31.4	7.8	19.1	27.8

Table 12.1.1

	Treatment Level			
	1	2	...	k
	Y_{11}	Y_{12}		Y_{1k}
	Y_{21}	Y_{22}		
	$\vdots$	$\vdots$	...	$\vdots$
	$Y_{n_1 1}$	$Y_{n_2 2}$		$Y_{n_k k}$
Sample sizes:	n_1	n_2	...	n_k
Sample totals:	$T_{.1}$	$T_{.2}$		$T_{.k}$
Sample means:	$\overline{Y}_{.1}$	$\overline{Y}_{.2}$		$\overline{Y}_{.k}$
True means:	μ_1	μ_2		μ_k

- k treatment levels; k independent random sample of size $n_1, \cdots, n_k$
- Total sample size: $n = \sum_{i=1}^{k} n_i$
- Y_{ij}: i-th observation for the j-th level.
- Sample total: $T_{.j} = \sum_{i=1}^{n_j} Y_{ij}$
- Sample mean: $\overline{Y}_{.j} = \frac{1}{n_j} \sum_{i=1}^{n_j} Y_{ij} = \frac{T_{.j}}{n_j}$
- Overall total: $T_{..} = \sum_{j=1}^{k} \sum_{i=1}^{n_j} Y_{ij} = \sum_{j=1}^{k} T_{.j}$
- Overall mean: $\overline{Y}_{..} = \frac{1}{n} \sum_{j=1}^{k} \sum_{i=1}^{n_j} Y_{ij} = \frac{1}{n} \sum_{j=1}^{k} n_j \overline{Y}_{.j} = \frac{1}{n} \sum_{j=1}^{k} T_{.j}$

Table 12.1.1

	Treatment Level			
	1	2	...	k
	Y_{11}	Y_{12}		Y_{1k}
	Y_{21}	Y_{22}		
	$\vdots$	$\vdots$	...	$\vdots$
	$Y_{n_1 1}$	$Y_{n_2 2}$		$Y_{n_k k}$
Sample sizes:	n_1	n_2	...	n_k
Sample totals:	$T_{.1}$	$T_{.2}$		$T_{.k}$
Sample means:	$\overline{Y}_{.1}$	$\overline{Y}_{.2}$		$\overline{Y}_{.k}$
True means:	μ_1	μ_2		μ_k

- k treatment levels; k independent random sample of size $n_1, \cdots, n_k$
- Total sample size: $n = \sum_{i=1}^{k} n_i$
- Y_{ij}: i-th observation for the j-th level.
- Sample total: $T_{.j} = \sum_{i=1}^{n_j} Y_{ij}$
- Sample mean: $\overline{Y}_{.j} = \frac{1}{n_j} \sum_{i=1}^{n_j} Y_{ij} = \frac{T_{.j}}{n_j}$
- Overall total: $T_{..} = \sum_{j=1}^{k} \sum_{i=1}^{n_j} Y_{ij} = \sum_{j=1}^{k} T_{.j}$
- Overall mean: $\overline{Y}_{..} = \frac{1}{n} \sum_{j=1}^{k} \sum_{i=1}^{n_j} Y_{ij} = \frac{1}{n} \sum_{j=1}^{k} n_j \overline{Y}_{.j} = \frac{1}{n} \sum_{j=1}^{k} T_{.j}$

Table 12.1.1

	Treatment Level			
	1	2	...	k
	Y_{11}	Y_{12}		Y_{1k}
	Y_{21}	Y_{22}		
	$\vdots$	$\vdots$	...	$\vdots$
	$Y_{n_1 1}$	$Y_{n_2 2}$		$Y_{n_k k}$
Sample sizes:	n_1	n_2	...	n_k
Sample totals:	$T_{.1}$	$T_{.2}$		$T_{.k}$
Sample means:	$\overline{Y}_{.1}$	$\overline{Y}_{.2}$		$\overline{Y}_{.k}$
True means:	μ_1	μ_2		μ_k

- k treatment levels; k independent random sample of size $n_1, \cdots, n_k$
- Total sample size: $n = \sum_{i=1}^{k} n_i$
- Y_{ij}: i-th observation for the j-th level.
- Sample total: $T_{.j} = \sum_{i=1}^{n_j} Y_{ij}$
- Sample mean: $\overline{Y}_{.j} = \frac{1}{n_j} \sum_{i=1}^{n_j} Y_{ij} = \frac{T_{.j}}{n_j}$
- Overall total: $T_{..} = \sum_{j=1}^{k} \sum_{i=1}^{n_j} Y_{ij} = \sum_{j=1}^{k} T_{.j}$
- Overall mean: $\overline{Y}_{..} = \frac{1}{n} \sum_{j=1}^{k} \sum_{i=1}^{n_j} Y_{ij} = \frac{1}{n} \sum_{j=1}^{k} n_j \overline{Y}_{.j} = \frac{1}{n} \sum_{j=1}^{k} T_{.j}$

Table 12.1.1

	Treatment Level			
	1	2	...	k
	Y_{11}	Y_{12}		Y_{1k}
	Y_{21}	Y_{22}		
	$\vdots$	$\vdots$	...	$\vdots$
	$Y_{n_1 1}$	$Y_{n_2 2}$		$Y_{n_k k}$
Sample sizes:	n_1	n_2	...	n_k
Sample totals:	$T_{.1}$	$T_{.2}$		$T_{.k}$
Sample means:	$\overline{Y}_{.1}$	$\overline{Y}_{.2}$		$\overline{Y}_{.k}$
True means:	μ_1	μ_2		μ_k

- k treatment levels; k independent random sample of size $n_1, \cdots, n_k$
- Total sample size: $n = \sum_{i=1}^{k} n_i$
- Y_{ij}: i-th observation for the j-th level.
- Sample total: $T_{.j} = \sum_{i=1}^{n_j} Y_{ij}$
- Sample mean: $\overline{Y}_{.j} = \frac{1}{n_j} \sum_{i=1}^{n_j} Y_{ij} = \frac{T_{.j}}{n_j}$
- Overall total: $T_{..} = \sum_{j=1}^{k} \sum_{i=1}^{n_j} Y_{ij} = \sum_{j=1}^{k} T_{.j}$
- Overall mean: $\overline{Y}_{..} = \frac{1}{n} \sum_{j=1}^{k} \sum_{i=1}^{n_j} Y_{ij} = \frac{1}{n} \sum_{j=1}^{k} n_j \overline{Y}_{.j} = \frac{1}{n} \sum_{j=1}^{k} T_{.j}$

Table 12.1.1

	Treatment Level			
	1	2	...	k
	Y_{11}	Y_{12}		Y_{1k}
	Y_{21}	Y_{22}		
	$\vdots$	$\vdots$	...	$\vdots$
	$Y_{n_1 1}$	$Y_{n_2 2}$		$Y_{n_k k}$
Sample sizes:	n_1	n_2	...	n_k
Sample totals:	$T_{.1}$	$T_{.2}$		$T_{.k}$
Sample means:	$\overline{Y}_{.1}$	$\overline{Y}_{.2}$		$\overline{Y}_{.k}$
True means:	μ_1	μ_2		μ_k

- k treatment levels; k independent random sample of size $n_1, \cdots, n_k$
- Total sample size: $n = \sum_{i=1}^{k} n_i$
- Y_{ij}: i-th observation for the j-th level.
- Sample total: $T_{.j} = \sum_{i=1}^{n_j} Y_{ij}$
- Sample mean: $\overline{Y}_{.j} = \frac{1}{n_j} \sum_{i=1}^{n_j} Y_{ij} = \frac{T_{.j}}{n_j}$
- Overall total: $T_{..} = \sum_{j=1}^{k} \sum_{i=1}^{n_j} Y_{ij} = \sum_{j=1}^{k} T_{.j}$
- Overall mean: $\overline{Y}_{..} = \frac{1}{n} \sum_{j=1}^{k} \sum_{i=1}^{n_j} Y_{ij} = \frac{1}{n} \sum_{j=1}^{k} n_j \overline{Y}_{.j} = \frac{1}{n} \sum_{j=1}^{k} T_{.j}$

Table 12.1.1

	Treatment Level			
	1	2	…	k
	Y_{11}	Y_{12}		Y_{1k}
	Y_{21}	Y_{22}		
	$\vdots$	$\vdots$	…	$\vdots$
	$Y_{n_1 1}$	$Y_{n_2 2}$		$Y_{n_k k}$
Sample sizes:	n_1	n_2	…	n_k
Sample totals:	$T_{.1}$	$T_{.2}$		$T_{.k}$
Sample means:	$\overline{Y}_{.1}$	$\overline{Y}_{.2}$		$\overline{Y}_{.k}$
True means:	μ_1	μ_2		μ_k

- k treatment levels; k independent random sample of size $n_1, \cdots, n_k$
- Total sample size: $n = \sum_{i=1}^{k} n_i$
- Y_{ij}: i-th observation for the j-th level.
- Sample total: $T_{.j} = \sum_{i=1}^{n_j} Y_{ij}$
- Sample mean: $\overline{Y}_{.j} = \frac{1}{n_j} \sum_{i=1}^{n_j} Y_{ij} = \frac{T_{.j}}{n_j}$
- Overall total: $T_{..} = \sum_{j=1}^{k} \sum_{i=1}^{n_j} Y_{ij} = \sum_{j=1}^{k} T_{.j}$
- Overall mean: $\overline{Y}_{..} = \frac{1}{n} \sum_{j=1}^{k} \sum_{i=1}^{n_j} Y_{ij} = \frac{1}{n} \sum_{j=1}^{k} n_j \overline{Y}_{.j} = \frac{1}{n} \sum_{j=1}^{k} T_{.j}$

Table 12.1.1				
	Treatment Level			
	1	2	...	k
	Y_{11}	Y_{12}		Y_{1k}
	Y_{21}	Y_{22}		
	$\vdots$	$\vdots$	...	$\vdots$
	$Y_{n_1 1}$	$Y_{n_2 2}$		$Y_{n_k k}$
Sample sizes:	n_1	n_2	...	n_k
Sample totals:	$T_{.1}$	$T_{.2}$		$T_{.k}$
Sample means:	$\overline{Y}_{.1}$	$\overline{Y}_{.2}$		$\overline{Y}_{.k}$
True means:	μ_1	μ_2		μ_k

- k treatment levels; k independent random sample of size $n_1, \cdots, n_k$
- Total sample size: $n = \sum_{i=1}^{k} n_i$
- Y_{ij}: i-th observation for the j-th level.
- Sample total: $T_{.j} = \sum_{i=1}^{n_j} Y_{ij}$
- Sample mean: $\overline{Y}_{.j} = \frac{1}{n_j} \sum_{i=1}^{n_j} Y_{ij} = \frac{T_{.j}}{n_j}$
- Overall total: $T_{..} = \sum_{j=1}^{k} \sum_{i=1}^{n_j} Y_{ij} = \sum_{j=1}^{k} T_{.j}$
- Overall mean: $\overline{Y}_{..} = \frac{1}{n} \sum_{j=1}^{k} \sum_{i=1}^{n_j} Y_{ij} = \frac{1}{n} \sum_{j=1}^{k} n_j \overline{Y}_{.j} = \frac{1}{n} \sum_{j=1}^{k} T_{.j}$

Assume For $j = 1, \cdots, k$, $Y_{ij} \sim N(\mu_j, \sigma_j^2)$ and $\sigma_1^2 = \cdots = \sigma_k^2 = \sigma^2$ (unknown).

Problem Testing

$$H_0 : \mu_1 = \mu_2 = \cdots = \mu_k$$

versus

$$H_1 : \text{not all the } \mu_j\text{'s are equal}$$

Or testing *subhypotheses* such as

$$H_0 : \mu_i = \mu_j \quad \text{or} \quad H_0 : \mu_3 = (\mu_1 + \mu_2)/2$$

Assume For $j = 1, \cdots, k$, $Y_{ij} \sim N(\mu_j, \sigma_j^2)$ and $\sigma_1^2 = \cdots = \sigma_k^2 = \sigma^2$ (unknown).

Problem Testing

$$H_0 : \mu_1 = \mu_2 = \cdots = \mu_k$$

versus

$$H_1 : \text{not all the } \mu_j\text{'s are equal}$$

Or testing *subhypotheses* such as

$$H_0 : \mu_i = \mu_j \quad \text{or} \quad H_0 : \mu_3 = (\mu_1 + \mu_2)/2$$

Assume For $j = 1, \cdots, k$, $Y_{ij} \sim N(\mu_j, \sigma_j^2)$ and $\sigma_1^2 = \cdots = \sigma_k^2 = \sigma^2$ (unknown).

Problem Testing

$$H_0 : \mu_1 = \mu_2 = \cdots = \mu_k$$

versus

$$H_1 : \text{not all the } \mu_j\text{'s are equal}$$

Or testing *subhypotheses* such as

$$H_0 : \mu_i = \mu_j \quad \text{or} \quad H_0 : \mu_3 = (\mu_1 + \mu_2)/2$$

ANOVA was developed by statistician and evolutionary biologist —

Ronald Fisher

Statistician

Sir Ronald Aylmer Fisher FRS was a British statistician and geneticist. For his work in statistics, he has been described as "a genius who almost single-handedly created the foundations for modern statistical science" and "the single most important figure in 20th century statistics". Wikipedia

Born: February 17, 1890, East Finchley, London, United Kingdom

Died: July 29, 1962, Adelaide, Australia

Known for: Fisher's principle, Fisher information

Residence: United Kingdom, Australia

Education: Gonville & Caius College, University of Cambridge, Harrow School https://www.youtube.com/watch?v=0XsovsSnRuw