MULTIPLE PARAMETRIC / NON PARAMETRIC TESTS (DA_2022)

Andrea Giansanti
Dipartimento di Fisica, Sapienza Università di Roma

Andrea.Giansanti@roma1.infn.it

Lecture n. 20, Rome mon 16th of May 2022

DIPARTIMENTO DI FISICA

SAPIENZA
UNIVERSITÀ DI ROMA

LET US LOOK AT THIS SECTION OF THE PROGRAM

9. NON PARAMETRIC TESTS (R 9.1-9.2; 9.4 ,W&S chap.13)
The problem of non-normal distributed data
Lognormal distribution
Tests of normality
Ranks
Sign test
Wilcoxon rank-sum test (Mann-Whitney U test)

10. MULTIPLE PARAMETRIC/NON PARAMETRIC TESTS (R 12.1-12-4; 12.7)
Whitin group/between-group variability
One-way ANOVA
Bonferroni correction
Kruskal-Wallis tes

Outline L 20

- **Generalize tests to situations with multiple groups (beyond two)**
- **Data dredging, you make" the data eventually tell what you were looking for**" Motivation for the Bonferroni correction (see overleaf n. 8 in W&S) [see W&S chapter 15 for a compact nice discussion]
- https://en.wikipedia.org/wiki/Data_dredging

- Why multiple comparison tests induce type I errors?

- ONE WAY ANALYSIS OF VARIANCE (ANOVA)
- Between/within variability
- F test statistics (see also R chap. 8.6

(see Rosners' chapter 12)

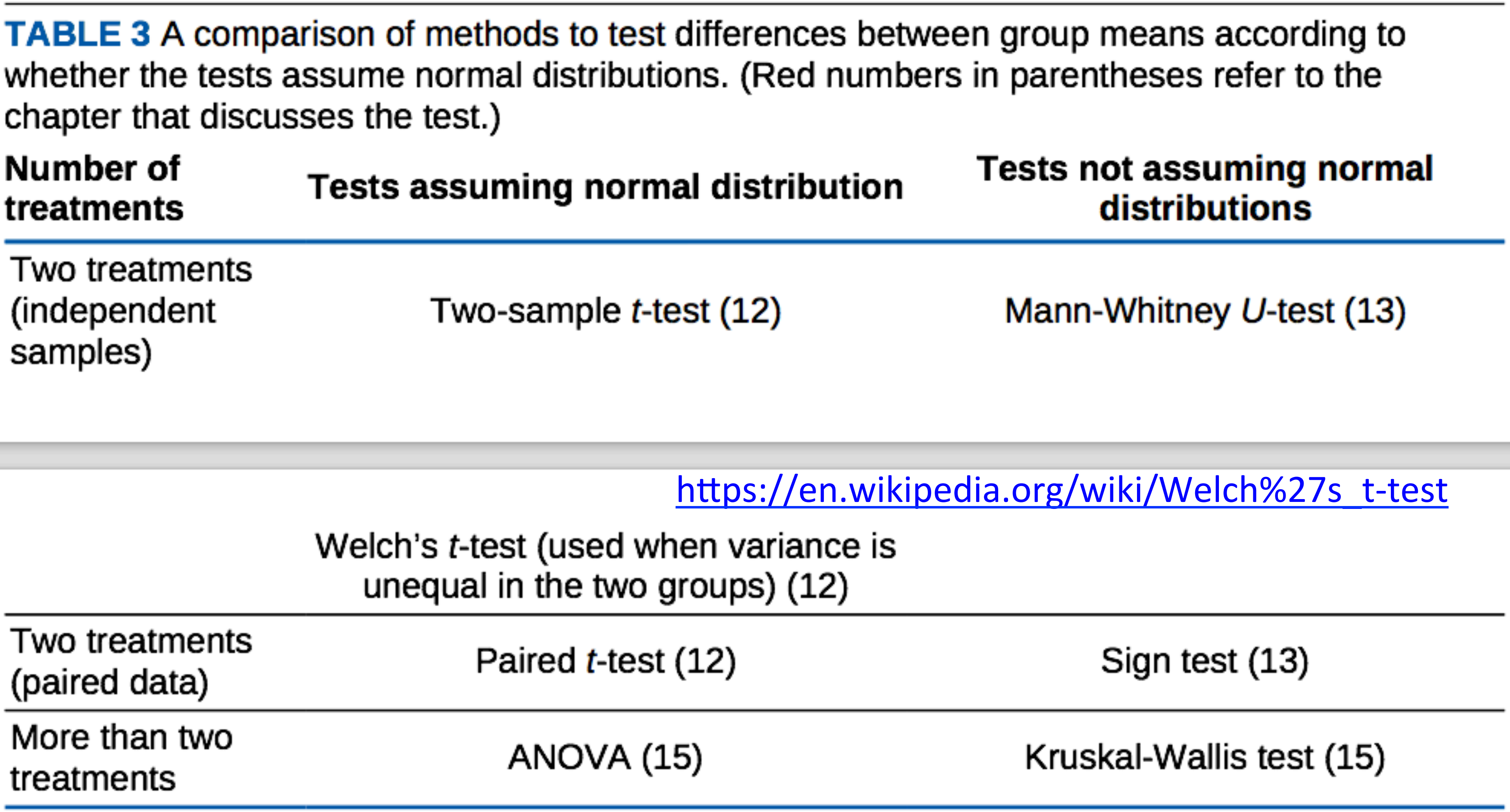


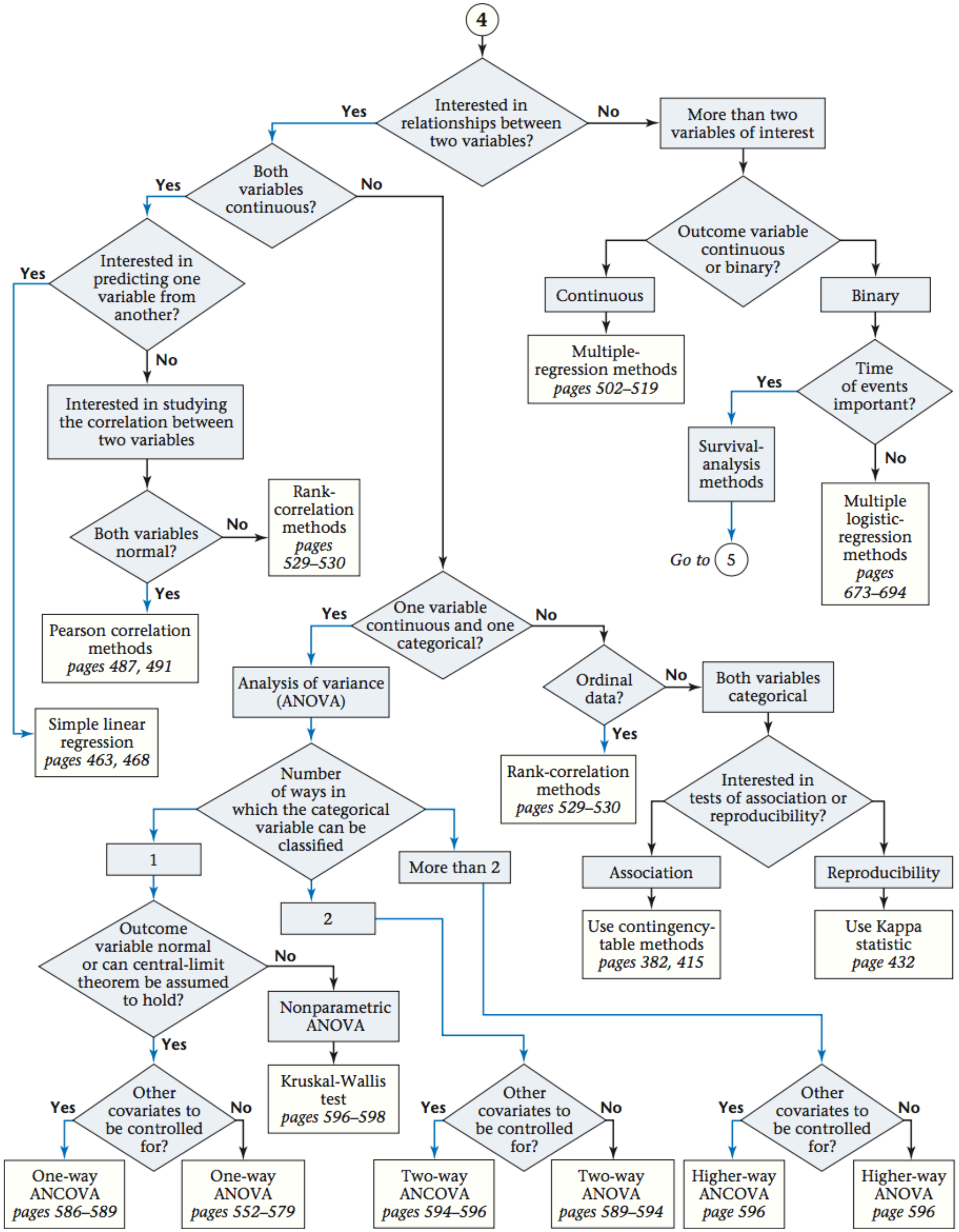
TABLE 3 A comparison of methods to test differences between group means according to whether the tests assume normal distributions. (Red numbers in parentheses refer to the chapter that discusses the test.)

Number of treatments	Tests assuming normal distribution	Tests not assuming normal distributions
Two treatments (independent samples)	Two-sample *t*-test (12) Welch's *t*-test (used when variance is unequal in the two groups) (12)	Mann-Whitney *U*-test (13)
Two treatments (paired data)	Paired *t*-test (12)	Sign test (13)
More than two treatments	ANOVA (15)	Kruskal-Wallis test (15)

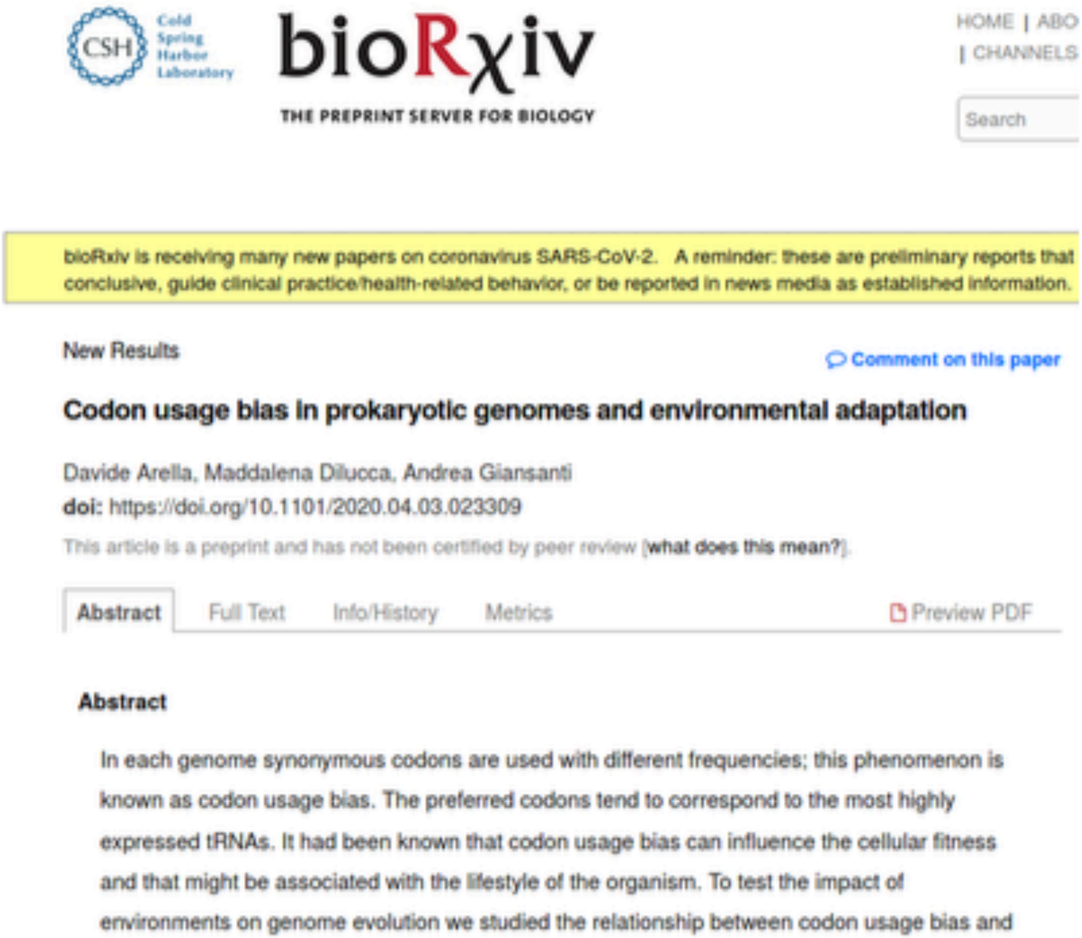
https://en.wikipedia.org/wiki/Welch%27s_t-test

…A GLIMPSE TO THE JUNGLE

FIGURE 11.33 Flowchart for appropriate methods of statistical inference



Paper already available on bioRxiv and submitted to *Molecular Genetics and Genomics*.



bioRxiv is receiving many new papers on coronavirus SARS-CoV-2. A reminder: these are preliminary reports that
conclusive, guide clinical practice/health-related behavior, or be reported in news media as established information.

New Results

Codon usage bias in prokaryotic genomes and environmental adaptation

Davide Arella, Maddalena Dilucca, Andrea Giansanti

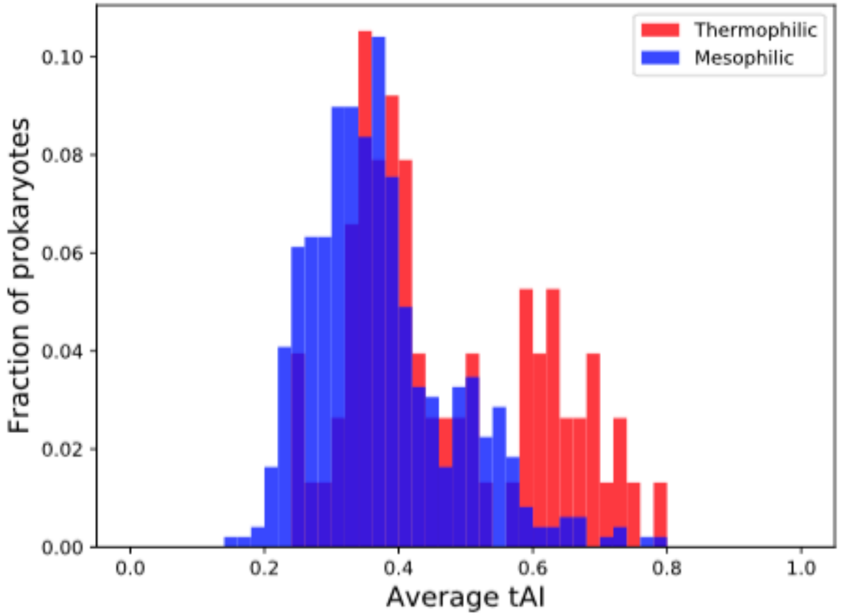
doi: https://doi.org/10.1101/2020.04.03.023309

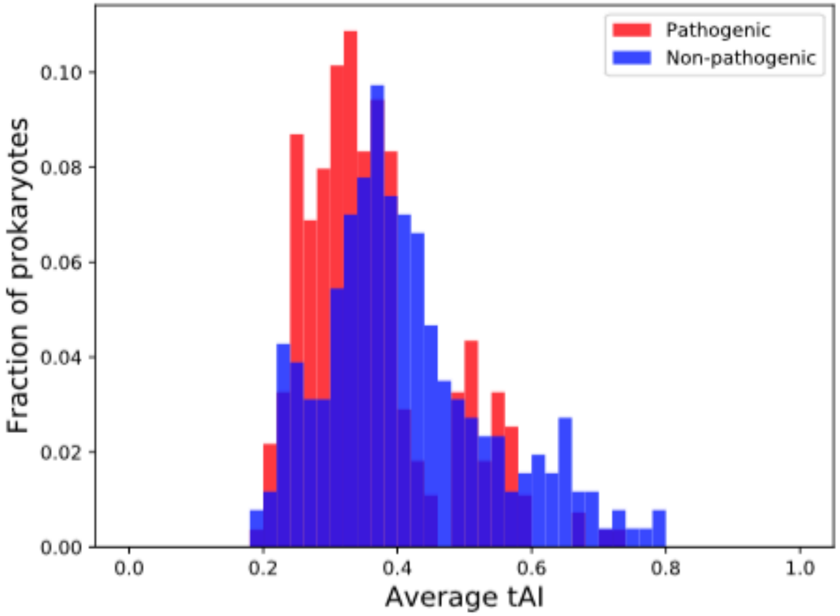
This article is a preprint and has not been certified by peer review [**what does this mean?**].

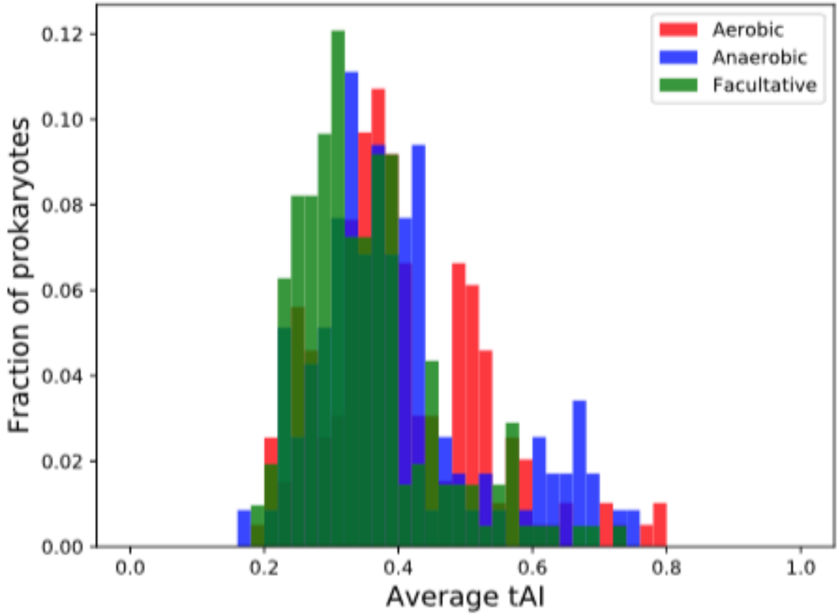
Abstract

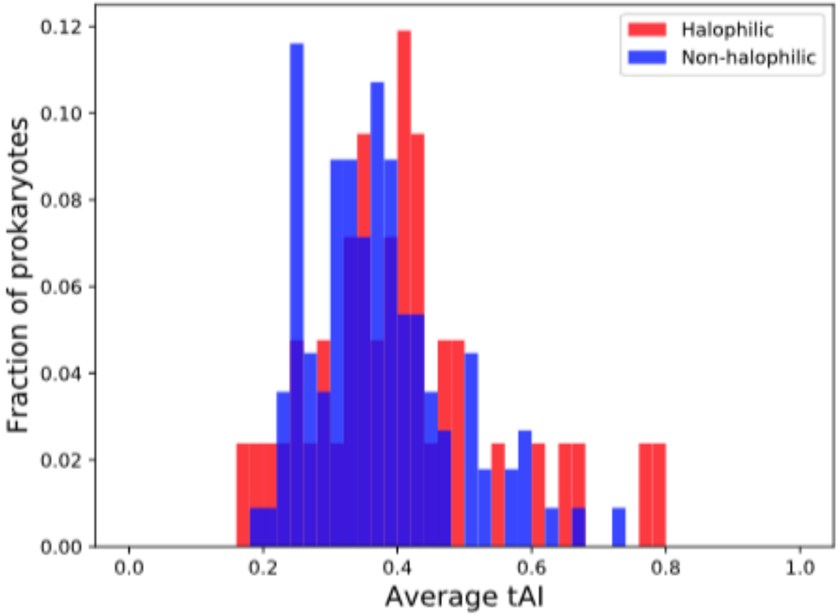
In each genome synonymous codons are used with different frequencies; this phenomenon is known as codon usage bias. The preferred codons tend to correspond to the most highly expressed tRNAs. It had been known that codon usage bias can influence the cellular fitness and that might be associated with the lifestyle of the organism. To test the impact of environments on genome evolution we studied the relationship between codon usage bias and

- **temperature range**: $p = 1.62 \times 10^{-8}$ (Mann-Whitney test)
- **pathogenicity**: $p = 8.67 \times 10^{-8}$ (Mann-Whitney test)
- **oxygen requirement**: $p = 7.09 \times 10^{-8}$ (Kruskal-Wallis test)
  $p = 4.53 \times 10^{-8}$, facultative vs aerobic (Mann-Whitney test)
  $p = 1.35 \times 10^{-4}$, facultative vs anaerobic (Mann-Whitney test)
  $p = 0.301^*$, aerobic vs anaerobic (Mann-Whitney test)
- **salinity**: $p = 0.161^*$ (Mann-Whitney test)
- **habitat**: $p = 2.66 \times 10^{-6}$, specialized vs multiple (Mann-Whitney test)


Thermophilic
Mesophilic
Fraction of prokaryotes
0.00
0.02
0.04
0.06
0.08
0.10
0.0
0.2
0.4
0.6
0.8
1.0
Average tAI


Pathogenic
Non-pathogenic
Fraction of prokaryotes
0.00
0.02
0.04
0.06
0.08
0.10
0.0
0.2
0.4
0.6
0.8
1.0
Average tAI


Aerobic
Anaerobic
Facultative
Fraction of prokaryotes
0.00
0.02
0.04
0.06
0.08
0.10
0.12
0.0
0.2
0.4
0.6
0.8
1.0
Average tAI


Halophilic
Non-halophilic
Fraction of prokaryotes
0.00
0.02
0.04
0.06
0.08
0.10
0.12
0.0
0.2
0.4
0.6
0.8
1.0
Average tAI

Pulmonary Disease A topic of public-health interest is whether *passive smoking* (exposure among nonsmokers to cigarette smoke in the atmosphere) has a measurable effect on pulmonary health. White and Froeb studied this question by measuring pulmonary function in several ways in the following six groups [1]:

(1) *Nonsmokers (NS)*: People who themselves did not smoke and were not exposed to cigarette smoke either at home or on the job.

(2) *Passive smokers (PS)*: People who themselves did not smoke and were not exposed to cigarette smoke in the home but were employed for 20 or more years in an enclosed working area that routinely contained tobacco smoke.

(3) *Noninhaling smokers (NI)*: People who smoked pipes, cigars, or cigarettes but who did not inhale.

(4) *Light smokers (LS)*: People who smoked and inhaled 1–10 cigarettes per day for 20 or more years. (*Note:* There are 20 cigarettes in a pack.)

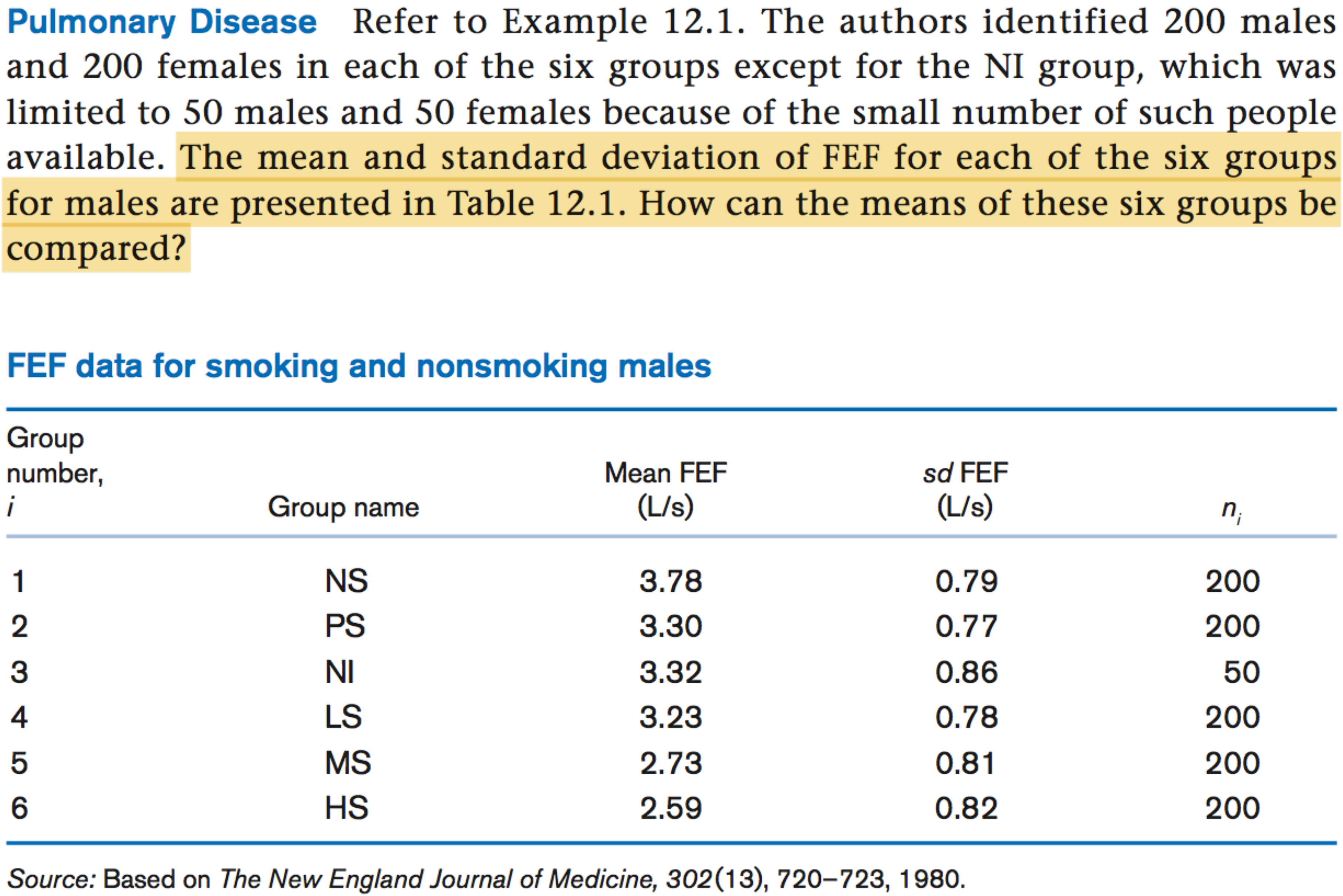
(5) *Moderate smokers (MS)*: People who smoked and inhaled 11–39 cigarettes per day for 20 or more years.

(6) *Heavy smokers (HS)*: People who smoked and inhaled 40 or more cigarettes per day for 20 or more years.

A principal measure used by White and Froeb to assess pulmonary function was forced mid-expiratory flow (FEF). They were interested in comparing mean FEF among the six groups.

The *t* test methodology generalizes nicely in this case to a procedure called the *one-way analysis of variance* (ANOVA).

THE EXPERIMET: PULMONARY FUNCTION & SMOKING HABITS



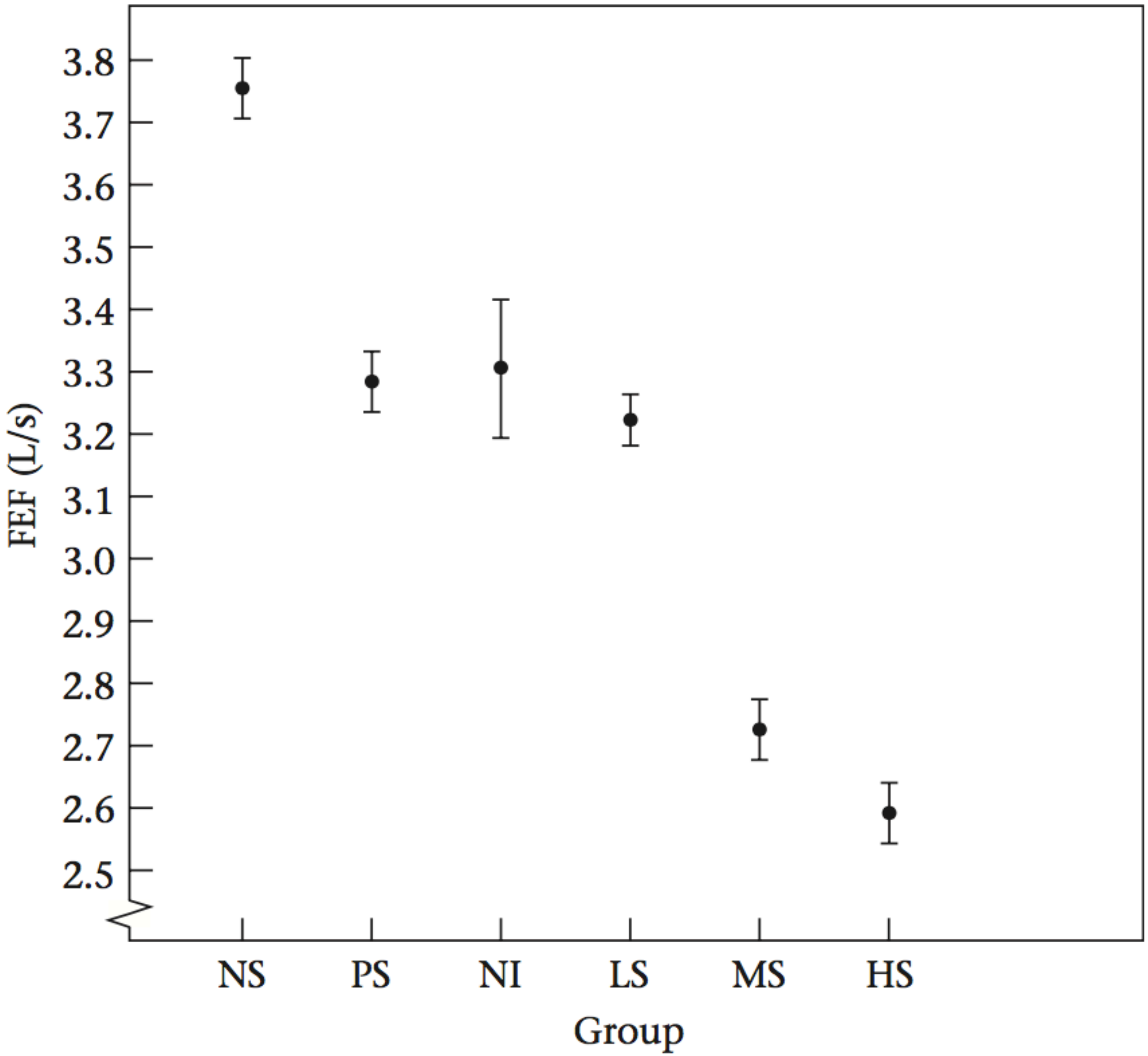
Pulmonary Disease Refer to Example 12.1. The authors identified 200 males and 200 females in each of the six groups except for the NI group, which was limited to 50 males and 50 females because of the small number of such people available. The mean and standard deviation of FEF for each of the six groups for males are presented in Table 12.1. How can the means of these six groups be compared?

FEF data for smoking and nonsmoking males

Group number, *i*	Group name	Mean FEF (L/s)	*sd* FEF (L/s)	n_i
1	NS	3.78	0.79	200
2	PS	3.30	0.77	200
3	NI	3.32	0.86	50
4	LS	3.23	0.78	200
5	MS	2.73	0.81	200
6	HS	2.59	0.82	200

Source: Based on *The New England Journal of Medicine, 302*(13), 720–723, 1980.

Mean ± *se* for FEF for each of six smoking groups



Source: Based on *The New England Journal of Medicine*, *302* (13), 720–723, 1980.

THE STATISTICAL MODEL

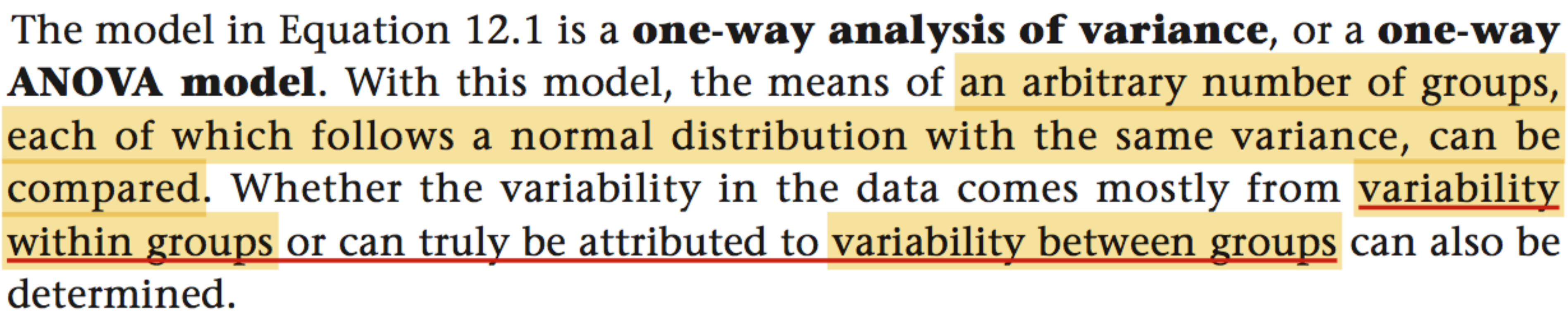
Suppose there are k groups with n_i observations in the ith group. The jth observation in the ith group will be denoted by y_{ij}. Let's assume the following model.

$$y_{ij} = \mu + \alpha_i + e_{ij}$$

where μ is a constant, α_i is a constant specific to the ith group, and e_{ij} is an error term, which is normally distributed with mean 0 and variance σ^2. Thus, a typical observation from the ith group is normally distributed with mean $\mu + \alpha_i$ and variance σ^2.

It is not possible to estimate both the overall constant μ as well as the k constants α_i, which are specific to each group. The reason is that we only have k observed mean values for the k groups, which are used to estimate $k + 1$ parameters. As a result, we need to constrain the parameters so that only k parameters will be estimated. Some typical constraints are (1) the sum of the α_i's is set to 0, or (2) the α_i for the last group (α_k) is set to 0. We use the former approach in this text. However, SAS uses the latter approach.

THE CORE IDEA BEHIND: DO THE STATISTICS OF THE DIFFERENT GROUP
COME FROM THE SAME DISTRIBUTION ?



The model in Equation 12.1 is a **one-way analysis of variance**, or a **one-way ANOVA model**. With this model, the means of an arbitrary number of groups, each of which follows a normal distribution with the same variance, can be compared. Whether the variability in the data comes mostly from variability within groups or can truly be attributed to variability between groups can also be determined.

CLEARLY, IF THE DIFFERENT GROUPS ARE SIGNIFICANTLY SEPARATED
WE HAVE TO REJECT THE NULL HYPOTHESIS OF JUST THE SAME DISTRIBUTION

Interpretation of the Parameters of a One-Way ANOVA Fixed-Effects Model

(1) μ represents the underlying mean of all groups taken together.

(2) α_i represents the difference between the mean of the *i*th group and the overall mean.

(3) e_{ij} represents random error about the mean $\mu + \alpha_i$ for an individual observation from the *i*th group.

Intuitively, in Table 12.1 an observed FEF is represented as a sum of an overall mean FEF plus an effect of each smoking group plus random variability within each smoking group. Group means are compared within the context of this model.

HYPOTHESIS TESTING IN ONE-WAY ANOVA—FIXED-EFFECTS MODEL

The null hypothesis (H_0) in this case is that the underlying mean FEF of each of the six groups is the same. This hypothesis is equivalent to stating that each $\alpha_i = 0$ because the α_i sum up to 0. The alternative hypothesis (H_1) is that at least two of the group means are not the same. This hypothesis is equivalent to stating that at least one $\alpha_i \neq 0$. Thus, we wish to test the hypothesis H_0: all $\alpha_i = 0$ vs. H_1: at least one $\alpha_i \neq 0$.

If H_1 should be accepted we cannot say which is the group that is at variance with the others

F Test for Overall Comparison of Group Means

The mean FEF for the *i*th group will be denoted by $\bar{y}_i$, and the mean FEF over all groups by $\bar{\bar{y}}$. The deviation of an individual observation from the overall mean can be represented as

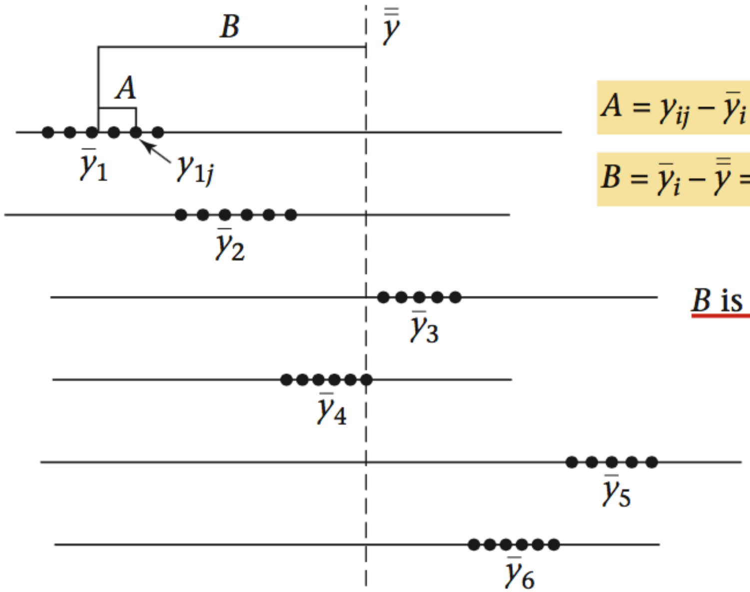
HOW THINGS WORK

$$y_{ij} - \bar{\bar{y}} = (y_{ij} - \bar{y}_i) + (\bar{y}_i - \bar{\bar{y}})$$

The first term on the right-hand side $(y_{ij} - \bar{y}_i)$ represents the deviation of an individual observation from the group mean for that observation and is an indication of *within-group variability*. The second term on the right-hand side $(\bar{y}_i - \bar{\bar{y}})$ represents the deviation of a group mean from the overall mean and is an indication of *between-group variability*. These terms are depicted in Figure 12.1.

Generally speaking, if the between-group variability is large and the within-group variability is small, as in Figure 12.1a, then H_0 is rejected and the underlying group means are declared significantly different. Conversely, if the between-group variability is small and the within-group variability is large, as in Figure 12.1b, then H_0, the hypothesis that the underlying group means are the same, is accepted.

Comparison of between-group and within-group variability

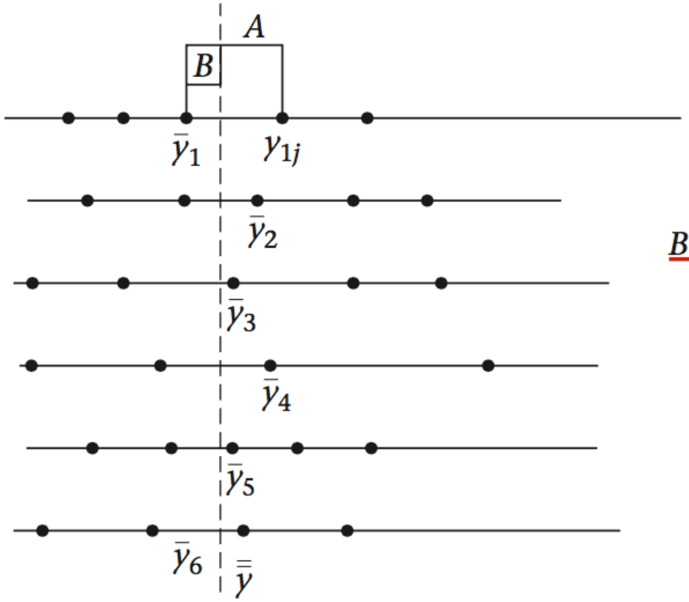


$A = y_{ij} - \bar{y}_i$ = within-group variability

$B = \bar{y}_i - \bar{\bar{y}}$ = between-group variability

B is large relative to A.

(a)



B is small relative to A.

(b)

EQUATION 12.3

$$y_{ij} - \bar{\bar{y}} = (y_{ij} - \bar{y}_i) + (\bar{y}_i - \bar{\bar{y}})$$

Square both sides of 12.3 to get

$$\sum_{i=1}^{k}\sum_{j=1}^{n_i}\left(y_{ij} - \bar{\bar{y}}\right)^2 = \sum_{i=1}^{k}\sum_{j=1}^{n_i}\left(y_{ij} - \bar{y}_i\right)^2 + \sum_{i=1}^{k}\sum_{j=1}^{n_i}\left(\bar{y}_i - \bar{\bar{y}}\right)^2$$

TOTAL SS = whitin groups SS + between groups SS
THE CROSS TERM IN THE SQUARE OF 12.3 IS ZERO
[…HW execise]

The term

$$\sum_{i=1}^{k}\sum_{j=1}^{n_i}\left(y_{ij}-\bar{\bar{y}}\right)^2$$

is called the **Total Sum of Squares (Total SS).**

The term

$$\sum_{i=1}^{k}\sum_{j=1}^{n_i}\left(y_{ij}-\bar{y}_i\right)^2$$

is called the **Within Sum of Squares (Within SS).**

The term

$$\sum_{i=1}^{k}\sum_{j=1}^{n_i}\left(\bar{y}_i-\bar{\bar{y}}\right)^2$$

is called the **Between Sum of Squares (Between SS).**

Short Computational Form for the Between SS and Within SS

$$\text{Between SS} = \sum_{i=1}^{k} n_i \bar{y}_i^2 - \frac{\left(\sum_{i=1}^{k} n_i \bar{y}_i\right)^2}{n} = \sum_{i=1}^{k} n_i \bar{y}_i^2 - \frac{y_{..}^2}{n}$$

$$\text{Within SS} = \sum_{i=1}^{k} (n_i - 1) s_i^2$$

Between Mean Square = **Between MS** = Between SS/(k – 1)

Within Mean Square = **Within MS** = Within SS/(n – k)

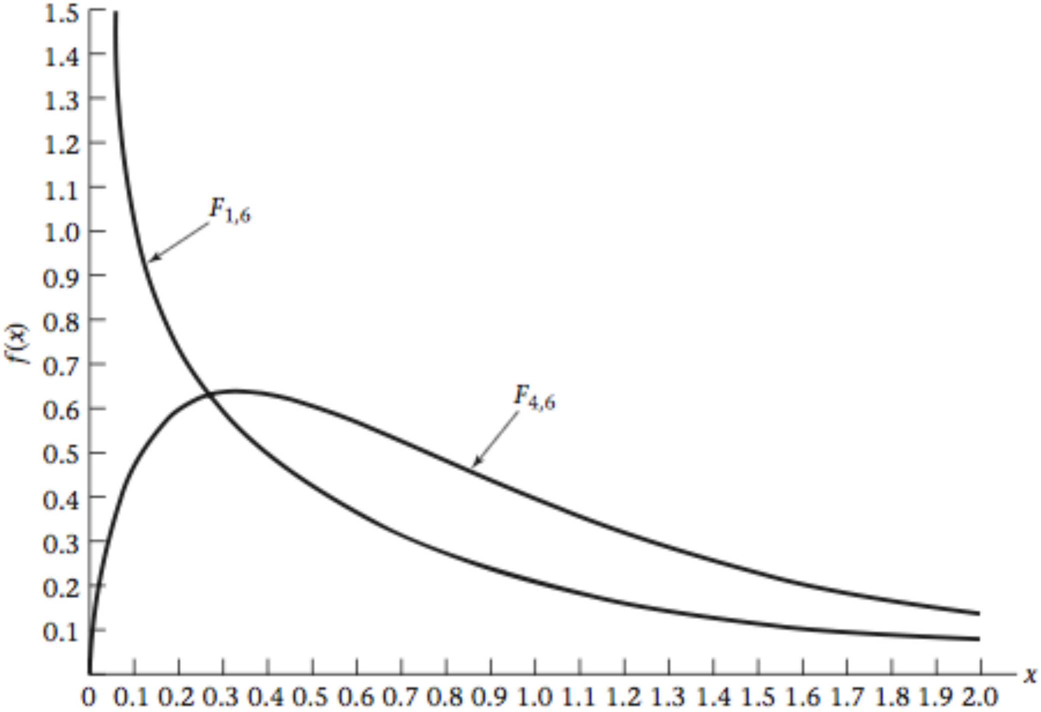
The significance test will be based on the ratio of the **Between MS** to the **Within MS.** If this ratio is large, then we reject H_0; if it is small, we accept (or fail to reject) H_0. Furthermore, under H_0, the ratio of Between MS to Within MS follows an F distribution with $k - 1$ and $n - k$ degrees of freedom. Thus, the following test procedure for a level α test is used.

The *F* Distribution

The distribution of the variance ratio $\left(S_1^2/S_2^2\right)$ was studied by statisticians R. A. Fisher and G. Snedecor. It can be shown that the variance ratio follows an ***F* distribution** under the null hypothesis that $\sigma_1^2 = \sigma_2^2$. There is no unique *F* distribution but instead a family of *F* distributions. This family is indexed by two parameters termed the *numerator* and *denominator degrees of freedom*, respectively. If the sizes of the first and second samples are n_1 and n_2 respectively, then the variance ratio follows an *F* distribution with $n_1 - 1$ (numerator *df*) and $n_2 - 1$ (denominator *df*), which is called an F_{n_1-1,n_2-1} distribution.

The *F* distribution is generally positively skewed, with the skewness dependent on the relative magnitudes of the two degrees of freedom. If the numerator *df* is 1 or 2, then the distribution has a mode at 0; otherwise, it has a mode greater than 0. The distribution is illustrated in Figure 8.5. Table 8 in the Appendix gives the percentiles of the *F* distribution for selected values of the numerator and denominator *df*.

Probability density for the *F* distribution



The **100 × *p*th percentile of an *F* distribution** with d_1 and d_2 degrees of freedom is denoted by $F_{d_1,d_2,p}$. Thus,

$$Pr\left(F_{d_1,d_2} \le F_{d_1,d_2,p}\right) = p$$

The *F* table is organized such that the numerator *df* (d_1) is shown in the first row, the denominator *df* (d_2) is shown in the first column, and the various percentiles (p) are shown in the second column.

Find the upper first percentile of an *F* distribution with 5 and 9 *df*.

Solution: $F_{5,9,.99}$ must be found. Look in the 5 column, the 9 row, and the subrow marked .99 to obtain

$$F_{5,9,.99} = 6.06$$

TABLE 8 Percentage points of the *F* distribution ($F_{d_1,d_2,p}$)

df for denominator, d_2	p	*df* for numerator, d_1 1	2	3	4	5	6	7	8	12	24	∞
1	.90	39.86	49.50	53.59	55.83	57.24	58.20	58.91	59.44	60.71	62.00	63.33
	.95	161.4	199.5	215.7	224.6	230.2	234.0	236.8	238.9	243.9	249.1	254.3
	.975	647.8	799.5	864.2	899.6	921.8	937.1	948.2	956.7	976.7	997.2	1018.
	.99	4052.	5000.	5403.	5625.	5764.	5859.	5928.	5981.	6106.	6235.	6366.
	.995	16211.	20000.	21615.	22500.	23056.	23437.	23715.	23925.	24426.	24940.	25464.
	.999	405280.	500000.	540380.	562500.	576400.	585940.	592870.	598140.	610670.	623500.	636620.
2	.90	8.53	9.00	9.16	9.24	9.29	9.33	9.35	9.37	9.41	9.45	9.49
	.95	18.51	19.00	19.16	19.25	19.30	19.33	19.35	19.37	19.41	19.45	19.50
	.975	38.51	39.00	39.17	39.25	39.30	39.33	39.36	39.37	39.42	39.46	39.50
	.99	98.50	99.00	99.17	99.25	99.30	99.33	99.36	99.37	99.42	99.46	99.50
	.995	198.5	199.0	199.2	199.2	199.3	199.3	199.4	199.4	199.4	199.5	199.5
	.999	998.5	999.0	999.2	999.2	999.3	999.3	999.4	999.4	999.4	999.5	999.5
3	.90	5.54	5.46	5.39	5.34	5.31	5.28	5.27	5.25	5.22	5.18	5.13
	.95	10.13	9.55	9.28	9.12	9.01	8.94	8.89	8.85	8.74	8.64	8.53
	.975	17.44	16.04	15.44	15.10	14.88	14.74	14.62	14.54	14.34	14.12	13.90
	.99	34.12	30.82	29.46	28.71	28.24	27.91	27.67	27.49	27.05	26.60	26.13
	.995	55.55	49.80	47.47	46.20	45.39	44.84	44.43	44.13	43.39	42.62	41.83
	.999	167.00	148.5	141.1	137.1	134.6	132.8	131.6	130.6	128.3	125.9	123.5
4	.90	4.54	4.32	4.19	4.11	4.05	4.01	3.98	3.95	3.90	3.83	3.76
	.95	7.71	6.94	6.59	6.39	6.26	6.16	6.09	6.04	5.91	5.77	5.63
	.975	12.22	10.65	9.98	9.60	9.36	9.20	9.07	8.98	8.75	8.51	8.26
	.99	21.20	18.00	16.69	15.98	15.52	15.21	14.98	14.80	14.37	13.93	13.46
	.995	31.33	26.28	24.26	23.16	22.46	21.98	21.62	21.35	20.70	20.03	19.32
	.999	74.14	61.25	56.18	53.44	51.71	50.53	49.66	49.00	47.41	45.77	44.05
5	.90	4.06	3.78	3.62	3.52	3.45	3.40	3.37	3.34	3.27	3.19	3.10
	.95	6.61	5.79	5.41	5.19	5.05	4.95	4.88	4.82	4.68	4.53	4.36
	.975	10.01	8.43	7.76	7.39	7.15	6.98	6.85	6.76	6.52	6.28	6.02
	.99	16.26	13.27	12.06	11.39	10.97	10.67	10.46	10.29	9.89	9.47	9.02
	.995	22.78	18.31	16.53	15.56	14.94	14.51	14.20	13.96	13.38	12.78	12.14
	.999	47.18	37.12	33.20	31.09	29.75	28.83	28.16	27.65	26.42	25.13	23.79
6	.90	3.78	3.46	3.29	3.18	3.11	3.05	3.01	2.98	2.90	2.82	2.72
	.95	5.99	5.14	4.76	4.53	4.39	4.28	4.21	4.15	4.00	3.84	3.67
	.975	8.81	7.26	6.60	6.23	5.99	5.82	5.70	5.60	5.37	5.12	4.85
	.99	13.75	10.92	9.78	9.15	8.75	8.47	8.26	8.10	7.72	7.31	6.88
	.995	18.64	14.54	12.92	12.03	11.46	11.07	10.79	10.57	10.03	9.47	8.88
	.999	35.51	27.00	23.70	21.92	20.80	20.03	19.46	19.03	17.99	16.90	15.75
7	.90	3.59	3.26	3.07	2.96	2.88	2.83	2.78	2.75	2.67	2.58	2.47
	.95	5.59	4.74	4.35	4.12	3.97	3.87	3.79	3.73	3.57	3.41	3.23
	.975	8.07	6.54	5.89	5.52	5.29	5.12	4.99	4.90	4.67	4.42	4.14
	.99	12.25	9.55	8.45	7.85	7.46	7.19	6.99	6.84	6.47	6.07	5.65
	.995	16.24	12.40	10.88	10.05	9.52	9.16	8.89	8.68	8.18	7.65	7.08
	.999	29.25	21.69	18.77	17.20	16.21	15.52	15.02	14.63	13.71	12.73	11.70
8	.90	3.46	3.11	2.92	2.81	2.73	2.67	2.62	2.59	2.50	2.40	2.29
	.95	5.32	4.46	4.07	3.84	3.69	3.58	3.50	3.44	3.28	3.12	2.93
	.975	7.57	6.06	5.42	5.05	4.82	4.65	4.53	4.43	4.20	3.95	3.67
	.99	11.26	8.65	7.59	7.01	6.63	6.37	6.18	6.03	5.67	5.28	4.86
	.995	14.69	11.04	9.60	8.81	8.30	7.95	7.69	7.50	7.01	6.50	5.95
	.999	25.42	18.49	15.83	14.39	13.49	12.86	12.40	12.04	11.19	10.30	9.33
9	.90	3.36	3.01	2.81	2.69	2.61	2.55	2.51	2.47	2.38	2.28	2.16
	.95	5.12	4.26	3.86	3.63	3.48	3.37	3.29	3.23	3.07	2.90	2.71
	.975	7.21	5.71	5.08	4.72	4.48	4.32	4.20	4.10	3.87	3.61	3.33
	.99	10.56	8.02	6.99	6.42	6.06	5.80	5.61	5.47	5.11	4.73	4.31
	.995	13.61	10.11	8.72	7.96	7.47	7.13	6.88	6.69	6.23	5.73	5.19
	.999	22.86	16.39	13.90	12.56	11.71	11.13	10.70	10.37	9.57	8.72	7.81

Generally, *F* distribution tables give only upper percentage points because the symmetry properties of the *F* distribution make it possible to derive the lower percentage points of any *F* distribution from the corresponding upper percentage points of an *F* distribution with the degrees of freedom reversed. Specifically, note that under H_0, S_2^2/S_1^2 follows an F_{d_2,d_1} distribution. Therefore,

$$Pr\left(S_2^2/S_1^2 \ge F_{d_2,d_1,1-p}\right) = p$$

By taking the inverse of each side and reversing the direction of the inequality, we get

$$Pr\left(\frac{S_1^2}{S_2^2} \le \frac{1}{F_{d_2,d_1,1-p}}\right) = p$$

Under H_0, however, S_1^2/S_2^2 follows an F_{d_1,d_2} distribution. Therefore,

$$Pr\left(\frac{S_1^2}{S_2^2} \le F_{d_1,d_2,p}\right) = p$$

It follows from the last two inequalities that

$$F_{d_1,d_2,p} = \frac{1}{F_{d_2,d_1,1-p}}$$

This principle is summarized as follows.

Computation of the Lower Percentiles of an *F* Distribution

The **lower *p*th percentile** of an *F* distribution with d_1 and d_2 *df* is the reciprocal of the **upper *p*th percentile** of an *F* distribution with d_2 and d_1 *df*. In symbols,

$$F_{d_1,d_2,p} = 1/F_{d_2,d_1,1-p}$$

Thus, from Equation 8.14 we see that the lower *p*th percentile of an *F* distribution is the same as the inverse of the upper *p*th percentile of an *F* distribution with the degrees of freedom reversed.

Overall F Test for One-Way ANOVA

To test the hypothesis H_0: $\alpha_i = 0$ for all i vs. H_1: at least one $\alpha_i \neq 0$, use the following procedure:

(1) Compute the Between SS, Between MS, Within SS, and Within MS using Equation 12.5 and Definitions 12.5 and 12.6.

(2) Compute the test statistic F = Between MS/Within MS, which follows an F distribution with $k - 1$ and $n - k$ df under H_0.

(3) If $F > F_{k-1,n-k,1-\alpha}$ then reject H_0
If $F \leq F_{k-1,n-k,1-\alpha}$ then accept H_0

(4) The exact p-value is given by the area to the right of F under an $F_{k-1,n-k}$ distribution $= Pr(F_{k-1,n-k} > F)$.

The acceptance and rejection regions for this test are shown in Figure 12.2. Computation of the exact p-value is illustrated in Figure 12.3. The results from the ANOVA are typically displayed in an ANOVA table, as in Table 12.2.

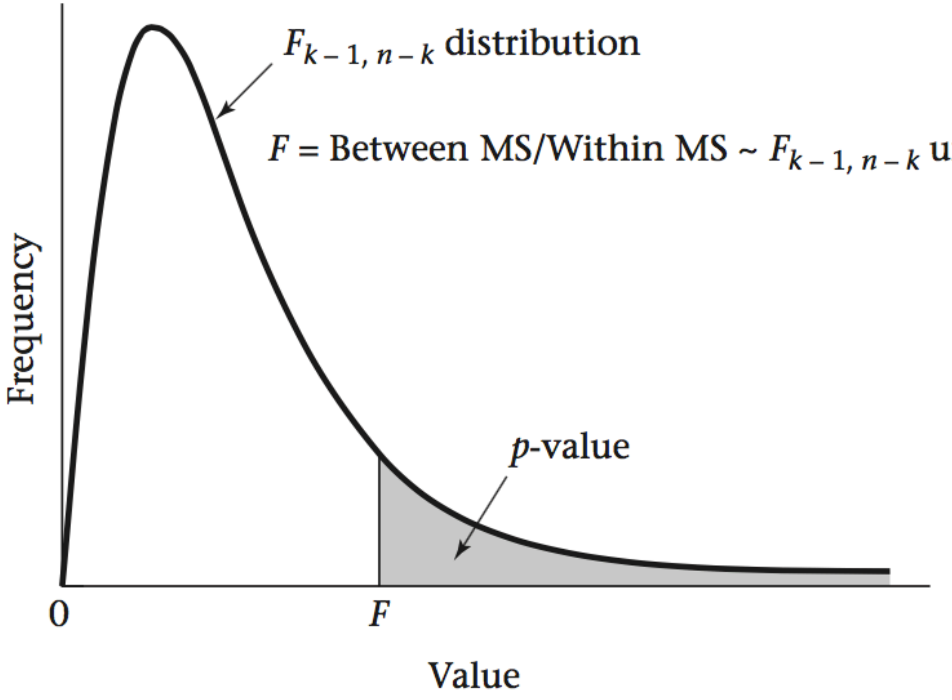
Display of one-way ANOVA results

Source of variation	SS	*df*	MS	*F* statistic	*p*-value
Between	$\sum_{i=1}^{k} n_i \bar{y}_i^2 - \frac{y_{..}^2}{n} = B$	$k - 1$	$\frac{B}{k-1}$	$\frac{B/(k-1)}{A/(n-k)} = F$	$Pr(F_{k-1,n-k} > F)$
Within	$\sum_{i=1}^{k} (n_i - 1)s_i^2 = A$	$n - k$	$\frac{A}{n-k}$		
Total	Between SS + Within SS				

ANOVA table for FEF data in Table 12.1

	SS	*df*	MS	*F* statistic	*p*-value
Between	184.38	5	36.875	58.0	$p < .001$
Within	663.87	1044	0.636		
Total	848.25				

Computation of the exact *p*-value for the overall *F* test for one-way ANOVA



Refer to Table 8 in the Appendix and find that

CONCLUSION

$$F_{5,120,.999} = 4.42$$

Because $F_{5,1044,.999} < F_{5,120,.999} = 4.42 < 58.0 = F$

it follows that $p < .001$. The exact *p*-value obtained from Stata = *F*tail(5,1044,58) = 2.5×10^{-53}. Therefore, we can reject H_0, that all the means are equal, and can conclude that at least two of the means are significantly different. These results are displayed in an ANOVA table (Table 12.3).

THE KRUSKAL-WALLIS TEST

In some instances we want to compare means among more than two samples, but either the underlying distribution is far from being normal or we have ordinal data. In these situations, a nonparametric alternative to the one-way ANOVA described in Sections 12.1–12.4 of this chapter must be used.

THE KW TEST IS BASED ON THE CHI_SQUARE STATISTICS

$$H = \frac{12}{N(N+1)} \sum_{i=1}^{k} n_i \left(\overline{R}_i - \overline{\overline{R}}\right)^2$$

where $\overline{R}_i$ = average rank in the ith sample and $\overline{\overline{R}}$ = average rank over all samples combined. Thus, if the average rank is about the same in all samples, then $\left|\overline{R}_i - \overline{\overline{R}}\right|$ will tend to be small and H_0 will be accepted. On the contrary, if the average rank is very different across samples, then $\left|\overline{R}_i - \overline{\overline{R}}\right|$ will tend to be large and H_0 will be rejected.

The Kruskal-Wallis Test

To compare the means of k samples ($k > 2$) using nonparametric methods, use the following procedure:

(1) Pool the observations over all samples, thus constructing a combined sample of size $N = \Sigma n_i$

(2) Assign ranks to the individual observations, using the average rank in the case of tied observations.

(3) Compute the rank sum R_i for each of the k samples.

(4) If there are no ties, compute the test statistic

$$H = H^* = \frac{12}{N(N+1)} \times \sum_{i=1}^{k} \frac{R_i^2}{n_i} - 3(N+1)$$

If there are ties, compute the test statistic

$$H = \frac{H^*}{1 - \frac{\sum_{j=1}^{g}\left(t_j^3 - t_j\right)}{N^3 - N}}$$

where t_j refers to the number of observations (i.e., the frequency) with the same value in the jth cluster of tied observations and g is the number of tied groups.

(5) For a level α test,

if $H > \chi^2_{k-1,1-\alpha}$ then reject H_0

if $H \leq \chi^2_{k-1,1-\alpha}$ then accept H_0

(6) To assess statistical significance, the p-value is given by

$$p = Pr\left(\chi^2_{k-1} > H\right)$$

(7) This test procedure should be used only if minimum $n_i \geq 5$ (i.e., if the smallest sample size for an individual group is at least 5).

The acceptance and rejection regions for this test are shown in Figure 12.12. Computation of the exact p-value is given in Figure 12.13.

Multiple Comparisons—Bonferroni Approach

In many studies, comparisons of interest are specified before looking at the actual data, in which case the *t* test procedure in Equation 12.12 and the linear-contrast procedure in Equation 12.13 are appropriate. In other instances, comparisons of interest are only specified after looking at the data. In this case a large number of potential comparisons are often possible. Specifically, if there are a large number of groups and every pair of groups is compared using the *t* test procedure in Equation 12.12, then some significant differences are likely to be found just by chance.

Suppose there are 10 groups. Thus, there are $\binom{10}{2} = 45$ possible pairs of groups to be compared. Using a 5% level of significance would imply that .05(45), or about two comparisons, are likely to be significant by chance alone. How can we protect ourselves against the detection of falsely significant differences resulting from making too many comparisons?

Several procedures, referred to as **multiple-comparisons procedures,** ensure that too many falsely significant differences are not declared. The basic idea of these procedures is to ensure that the *overall probability of declaring any significant differences between all possible pairs of groups* is maintained at some fixed significance level (say α). One of the simplest and most widely used such procedures is the method of *Bonferroni adjustment*. This method is summarized as follows.

Comparison of Pairs of Groups in One-Way ANOVA—Bonferroni Multiple-Comparisons Procedure

Suppose we wish to compare two specific groups, arbitrarily labeled as group 1 and group 2, among k groups. To test the hypothesis $H_0{:}\,\alpha_1 = \alpha_2$ vs. $H_1{:}\,\alpha_1 \neq \alpha_2$, use the following procedure:

(1) Compute the pooled estimate of the variance s^2 = Within MS from the one-way ANOVA.

(2) Compute the test statistic

$$t = \frac{\bar{y}_1 - \bar{y}_2}{\sqrt{s^2\left(\frac{1}{n_1} + \frac{1}{n_2}\right)}}$$

(3) For a two-sided level α test, let $\alpha^* = \alpha \Big/ \binom{k}{2}$

If $t > t_{n-k,1-\alpha^*/2}$ or $t < t_{n-k,\alpha^*/2}$ then reject H_0

If $t_{n-k,\alpha^*/2} \le t \le t_{n-k,1-\alpha^*/2}$ then accept H_0

The acceptance and rejection regions for this test are given in Figure 12.7.

(4) The Bonferroni corrected p-value $= \min\left[2\binom{k}{2}\Pr(t_{n-k} > |t|),\, 1\right]$

$= \min[k(k-1)\Pr(t_{n-k} > |t|, 1)] = \binom{k}{2}$ *LSD* p-value.

This test is called the *Bonferroni multiple-comparisons procedure*.

The rationale behind this procedure is that in a study with k groups, there are $\binom{k}{2}$ possible two-group comparisons. Suppose each two-group comparison is conducted at the α^* level of significance. Let E be the event that at least one of the two-group comparisons is statistically significant. $Pr(E)$ is sometimes referred to as the "experiment-wise type I error." We wish to determine the value α^* such that $Pr(E) = \alpha$. To find α^*, we note that