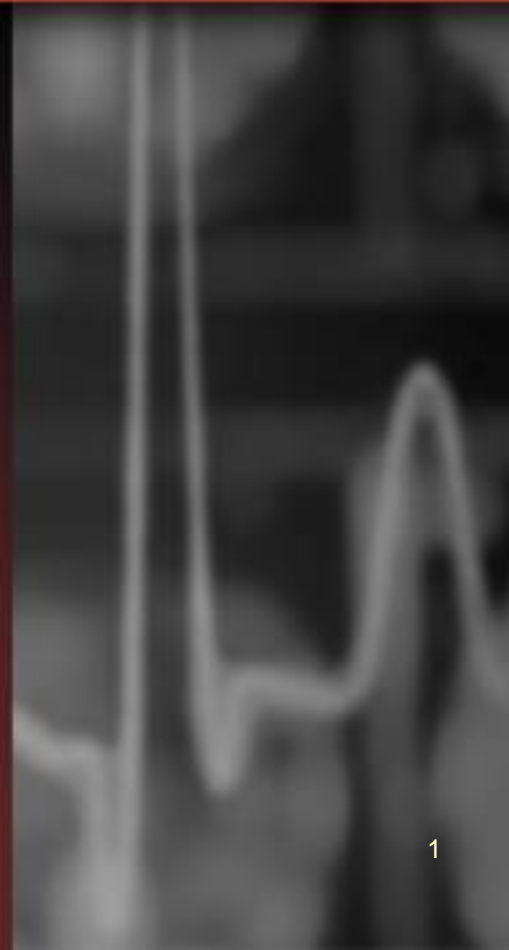


# Biostatistics

Week #14

6/02/2020



# Ch 12 – Analysis of Variance



# Chapter 11 review

- Paired samples – reducing to single random variable,  $t$ -test...
- Independent samples:
  - Equal variance  $S_p^2$
  - Unequal variance:

$$t = \frac{(\bar{x}_1 - \bar{x}_2) - (\mu_1 - \mu_2)}{\sqrt{s_p^2 [(1/n_1) + (1/n_2)]}}$$

$$t = \frac{(\bar{x}_1 - \bar{x}_2) - (\mu_1 - \mu_2)}{\sqrt{(s_1^2/n_1) + (s_2^2/n_2)}}.$$

$$DF = n_1 + n_2 - 2$$

$$v = \frac{[(s_1^2/n_1) + (s_2^2/n_2)]^2}{[(s_1^2/n_1)^2/(n_1 - 1) + (s_2^2/n_2)^2/(n_2 - 1)]};$$

Effective DF

# Introduction

- Analysis of variance compares two or more populations for their independent means.
- It is the extension of the two-sample  $t$ -test (in Chapter 11) to three or more samples.

		Group 1	Group 2	...	Group $k$
Population	Mean	$\mu_1$	$\mu_2$	...	$\mu_k$
	Standard Deviation	$\sigma_1$	$\sigma_2$	...	$\sigma_k$
Sample	Mean	$\bar{x}_1$	$\bar{x}_2$	...	$\bar{x}_k$
	Standard Deviation	$s_1$	$s_2$	...	$s_k$
	Sample Size	$n_1$	$n_2$	...	$n_k$

Similar to what we have stated in the last lecture (Chapter 11), we have statistics representing:

- Population (including “mean value  $\mu$ ” and “standard deviation  $\sigma$ ”) as well as
- Sample (including “mean  $\bar{x}$ ”, “standard deviation  $s$ ” and “sample size  $n$ ”).

# Example 1

- The individuals involved in a study were recruited from three medical centers.
- Before combining the subjects into one large group, we should first ensure that the patients from different centers are comparable.
- $FEV_1$  were recorded for each patients from three centers

$FEV_1$  : This is the amount of air that you can forcibly blow out in one second, measured in liters, and is considered one of the primary indicators of lung function.

**TABLE 12.1**

Forced expiratory volume in one second for patients with coronary artery disease sampled at three different medical centers

Johns Hopkins	Rancho Los Amigos	St. Louis
3.23	3.22	2.79

$$n_1 = 21$$

$$\bar{x}_1 = 2.63 \text{ liters}$$

$$n_2 = 16$$

$$\bar{x}_2 = 3.03 \text{ liters}$$

$$n_3 = 23$$

$$\bar{x}_3 = 2.88 \text{ liters}$$

2.10

2.81

3.28

3.36

2.61

2.91

1.98

2.57

2.08

2.47

2.47

2.74

2.88

2.63

2.53

3.39

3.86

2.64

2.71

2.71

3.41

2.87

2.61

3.39

3.17

2.95

3.56

2.88

2.63

3.38

3.07

2.81

3.17

2.23

2.19

4.06

1.98

2.81

2.85

2.43

3.20

3.53

$$n_1 = 21$$

$$\bar{x}_1 = 2.63 \text{ liters}$$

$$s_1 = 0.496 \text{ liters}$$

$$n_2 = 16$$

$$\bar{x}_2 = 3.03 \text{ liters}$$

$$s_2 = 0.523 \text{ liters}$$

$$n_3 = 23$$

$$\bar{x}_3 = 2.88 \text{ liters}$$

$$s_3 = 0.498 \text{ liters}$$

**Observation:** It looks like that the mean value from Johns Hopkins (2.63) may be a little lower than the means for the other two groups (3.03 and 2.88).



**TABLE 12.1**

Forced expiratory volume in one second for patients with coronary artery disease sampled at three different medical centers

Johns Hopkins	Rancho Los Amigos	St. Louis
3.23	3.22	2.79
3.47	2.88	3.22
1.86	1.71	2.25
2.47	2.89	2.98
3.01	3.77	2.47
1.69	3.29	2.77
2.10	3.39	2.95
2.81	3.86	3.56
3.28	2.64	2.88
3.36	2.71	2.63
2.61	2.71	3.38
2.91	3.41	3.07
1.98	2.87	2.81
2.57	2.61	3.17
2.08	3.39	2.23
2.47	3.17	2.19
2.47		4.06
2.74		1.98
2.88		2.81
2.63		2.85
2.53		2.43
		3.20
		3.53
$n_1 = 21$ $\bar{x}_1 = 2.63$ liters $s_1 = 0.496$ liters	$n_2 = 16$ $\bar{x}_2 = 3.03$ liters $s_2 = 0.523$ liters	$n_3 = 23$ $\bar{x}_3 = 2.88$ liters $s_3 = 0.498$ liters

**Question:** is it possible to join all these recruitments from 3 medical centers together into a bigger sample (so that  $n=21+16+23=60$ )?

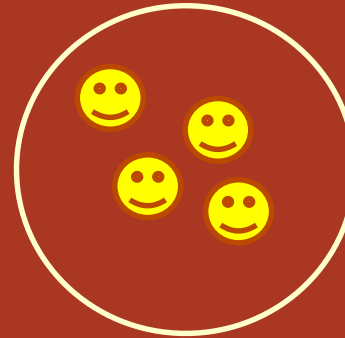


# Solution

- We might attempt to compare the three population means by evaluating all possible pairs of sample means using the 2-sample t-test (as did in last lecture).
- In this case, we'd compare group 1 vs group 2, group 1 vs group 3, and group 2 vs group 3.
- For the number of k cases getting large, we need to do  $C(k,2)$  such tests, which could be expensive.

# Cont'd

- To consider the overall instead of all paired means and variability, we need to estimate the following two types of variance:
  - $S_w^2$ : The pooled estimate of the “common variance **within**” each of these groups
  - $S_B^2$ : The variance **between** groups, that measures how mean of each group varies around the overall mean



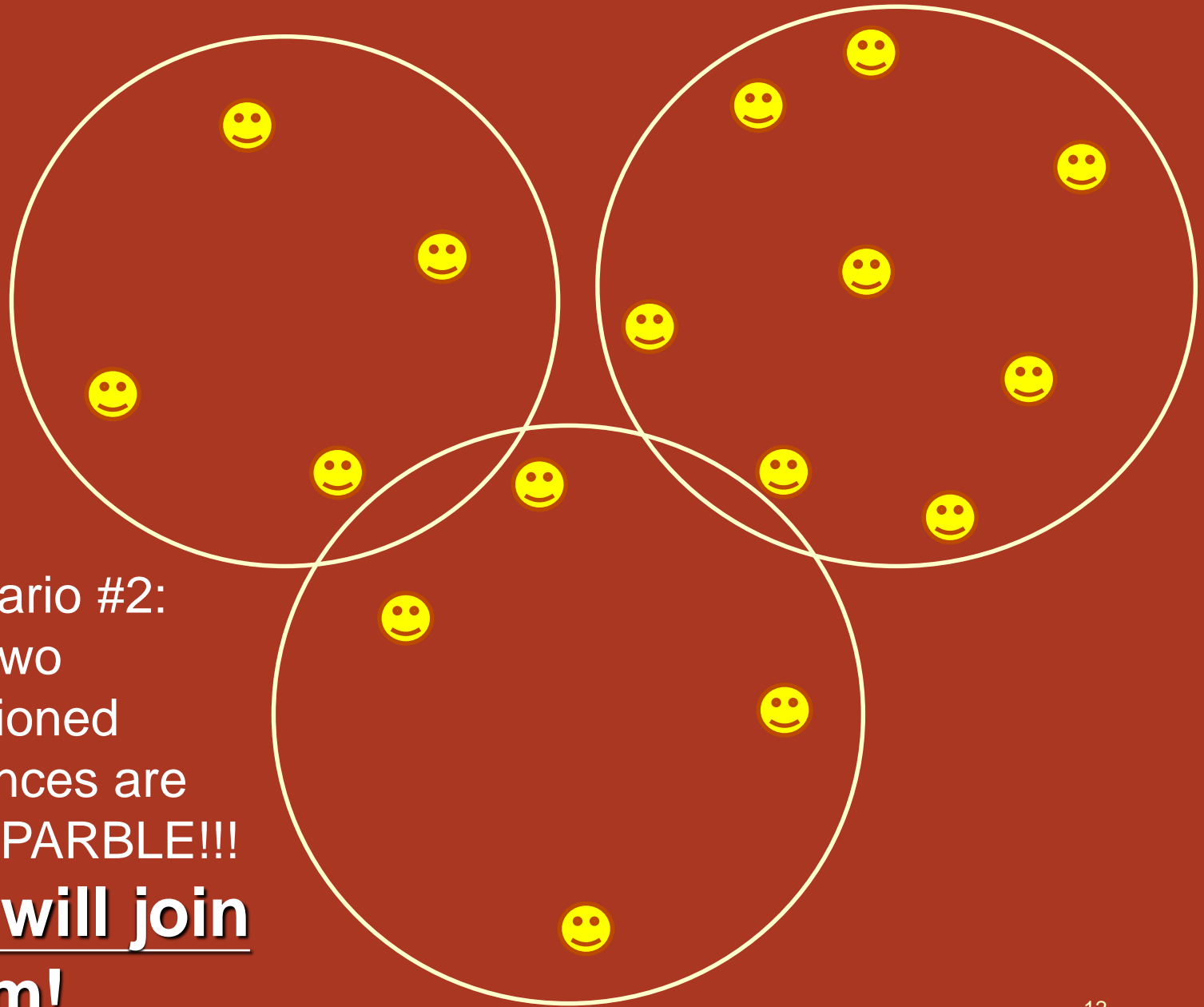
Scenario #1:

→  $S_w^2$ : The pooled estimate of the “common variance **within**”

each of these groups is **SMALL**.

→  $S_B^2$ : The variance **between** groups, that measures how mean of each group varies around the overall mean is **BIG**.

→ **We will NOT join them!**



Scenario #2:  
The two  
mentioned  
variances are  
COMPARBLE!!!  
**We will join**  
**them!**

# Defining the hypothesis

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

$H_A$ : **At least** two means differ  
(yet we don't know which  
two)

# Pool estimate of variance

- Recall that in the 2-sample test ( $H_0 : \mu_1 = \mu_2$ ), we first compute the **pooled estimate of the variance** using the formula shown below:

$$s_p^2 = \frac{(n_1 - 1)s_1^2 + (n_2 - 1)s_2^2}{n_1 + n_2 - 2}$$

# Defining the “difference” for 3 or more samples

- The following two types of variances will be defined:
  1.  $S_w^2$ : Generalized Pooled Estimate of the Common Variance (***Within*** Groups)
  2.  $S_B^2$ : Variance of the Means of each Group around the Grand Mean (***Between*** Group)



# 1. Generalized Pooled Estimate of the Common Variance (Within Groups)

**w** : within-groups variability

$$\begin{aligned}s_W^2 &= \frac{(n_1 - 1)s_1^2 + (n_2 - 1)s_2^2 + \cdots + (n_k - 1)s_k^2}{n_1 + n_2 + \cdots + n_k - k} \\ &= \frac{(n_1 - 1)s_1^2 + (n_2 - 1)s_2^2 + \cdots + (n_k - 1)s_k^2}{n - k}.\end{aligned}$$

given that  $n = n_1 + n_2 + \cdots + n_k$

Generalizing from 2 to k

*Recall this is the pooled variance we used in a 2-sample test before.*

$$s_p^2 = \frac{(n_1 - 1)s_1^2 + (n_2 - 1)s_2^2}{n_1 + n_2 - 2}$$

## 2. Variance of the Means of each Group around the Grand Mean (Between-Group)

**B** : between-group variability

$$s_B^2 = \frac{n_1(\bar{x}_1 - \bar{x})^2 + n_2(\bar{x}_2 - \bar{x})^2 + \cdots + n_k(\bar{x}_k - \bar{x})^2}{k - 1}$$

The terms  $(\bar{x}_i - \bar{x})^2$  are the squared deviations of the sample means  $\bar{x}_i$  from the grand mean  $\bar{x}$ . The grand mean is defined as the overall average of the  $n$  observations that make up the  $k$  different samples; consequently,

$$\begin{aligned}\bar{x} &= \frac{n_1\bar{x}_1 + n_2\bar{x}_2 + \cdots + n_k\bar{x}_k}{n_1 + n_2 + \cdots + n_k} \\ &= \frac{n_1\bar{x}_1 + n_2\bar{x}_2 + \cdots + n_k\bar{x}_k}{n}.\end{aligned}$$

*This is like considering each mean value as a random variable, and estimating the variance of these random variables.*

# The F-test

- Here we define a new statistics F that helps in answering the question:
  - Do the “sample means vary around the grand mean ( $S_B^2$ )” **more than** the “individual observations vary around their sample means ( $S_W^2$ )” (such as scenario #1 mentioned earlier)?
  - If they do, this implies that the corresponding population means are in fact different. (Thus we reject the null hypothesis and DO NOT join them together)

# The F-test

$$F = \frac{s_B^2}{s_W^2}.$$

- Under the null hypothesis  $H_0$  that all sample means are equal, both  $s_W^2$  and  $s_B^2$  estimate the common variance, and **F is close to 1.**
- If there is a difference among populations, the between-group variance exceeds the within-group variance and **F is greater than 1.**

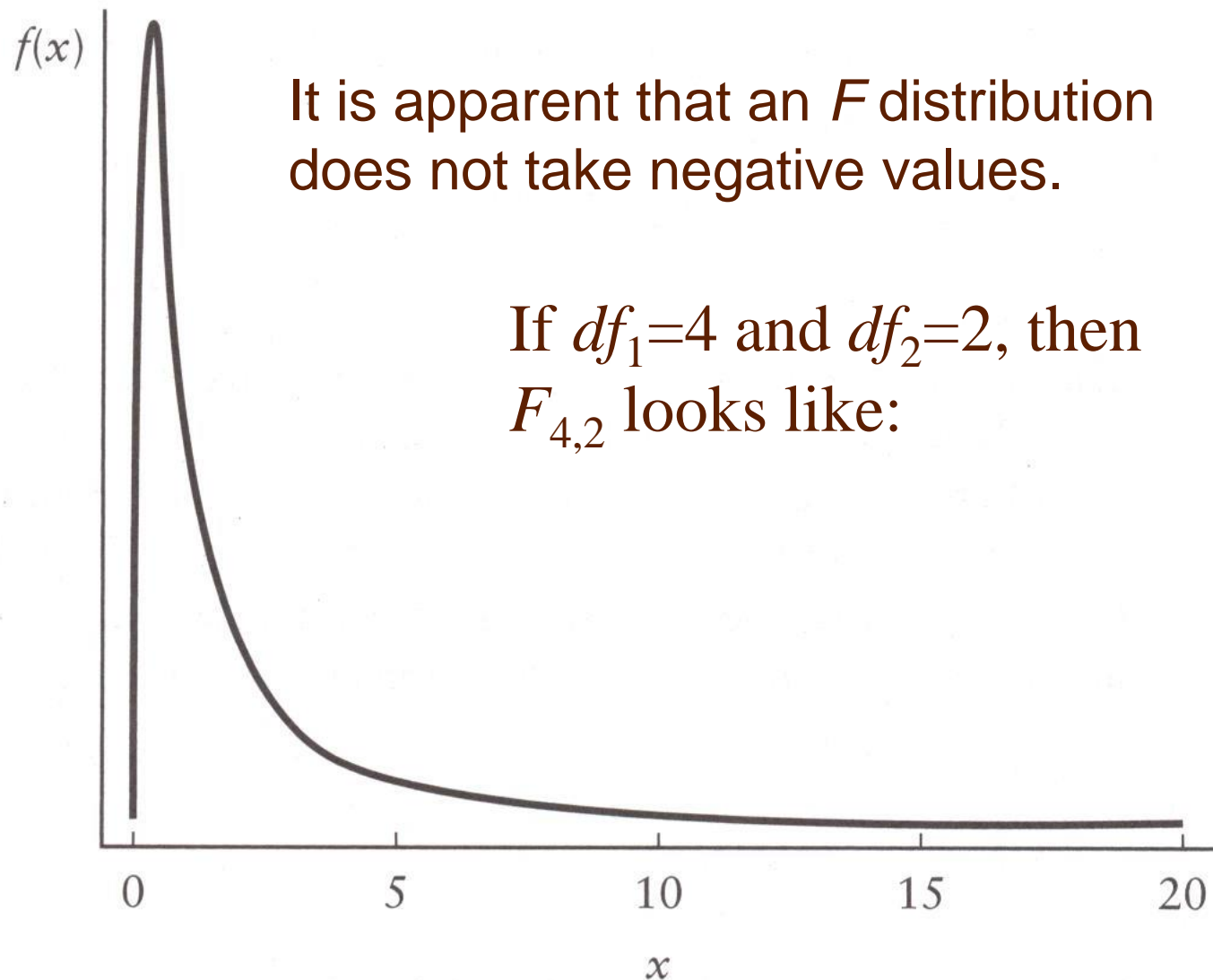
## Cont'd

$$F = \frac{s_B^2}{s_W^2}.$$

- Under  $H_0$ , the ratio  $F$  has an  $F$  distribution with  $k-1$  ( $df_1$ ) and  $n-k$  ( $df_2$ ) degrees of freedom; corresponding to the numerator (分子)  $s_B^2$  and the denominator (分母)  $s_W^2$ , respectively.
- We denote it  $F_{k-1, n-k}$ , or  $F_{df_1, df_2}$ .

$$s_B^2 = \frac{n_1(\bar{x}_1 - \bar{x})^2 + n_2(\bar{x}_2 - \bar{x})^2 + \cdots + n_k(\bar{x}_k - \bar{x})^2}{k - 1}$$

$$s_W^2 = \frac{(n_1 - 1)s_1^2 + (n_2 - 1)s_2^2 + \cdots + (n_k - 1)s_k^2}{n - k}$$



**FIGURE 12.2**

The  $F$  distribution with 4 and 2 degrees of freedom

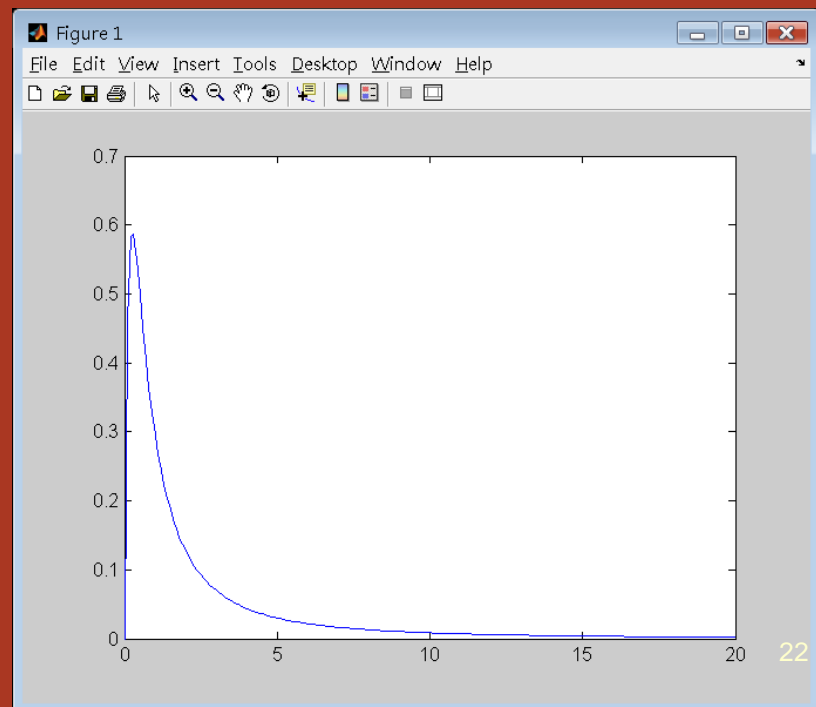
# MATLAB function fpdf

FPDF F probability density function.

$Y = \text{FPDF}(X, V1, V2)$  returns the F distribution probability density function with  $V1$  and  $V2$  degrees of freedom at the values in  $X$ .

For  $F_{4,2}$  shown in previous slide, we have

```
>> x=[0:0.1:20];  
>> y=fpdf(x,4,2);  
>> plot(x,y)
```



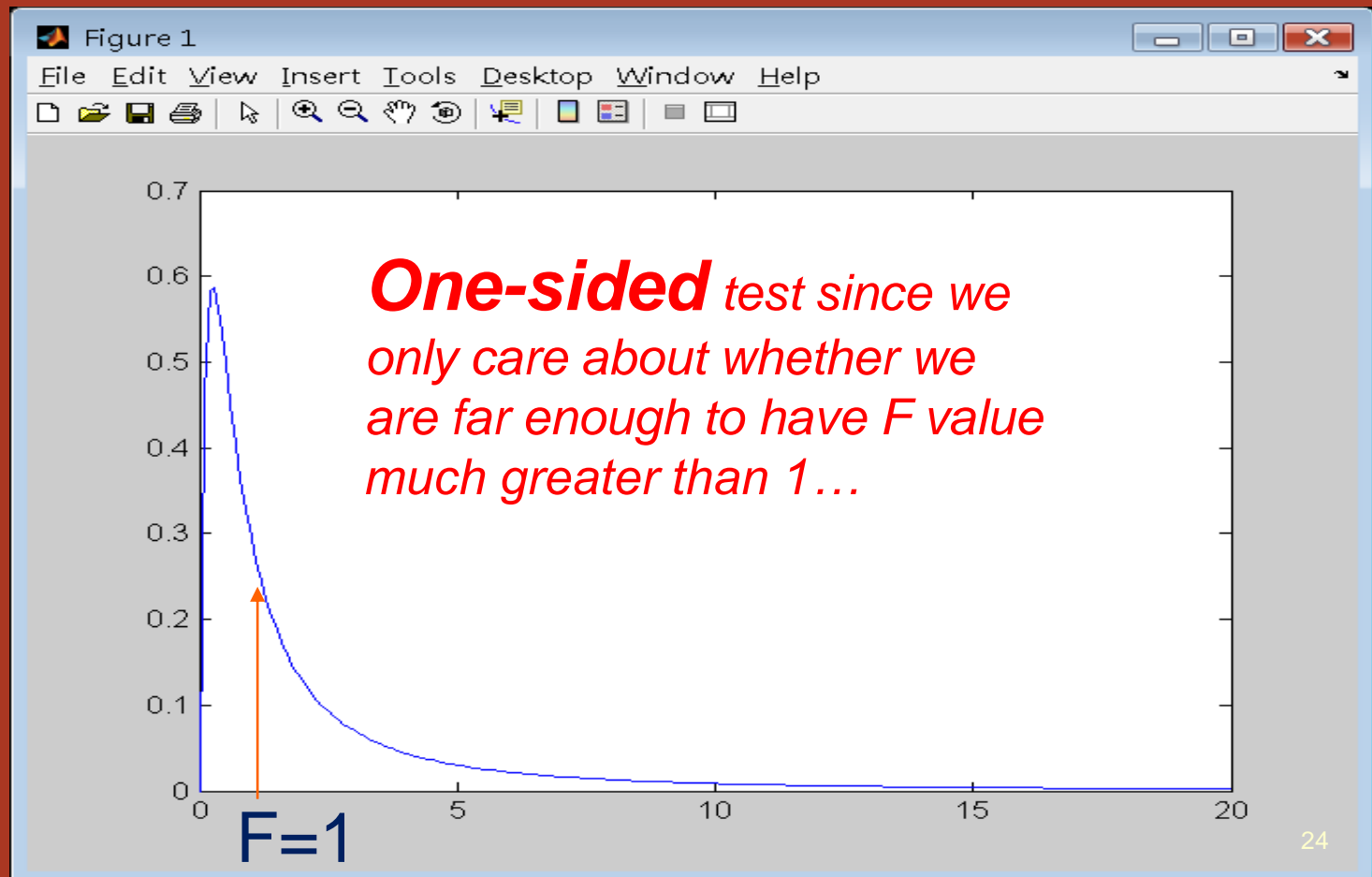


## More discussion about $F_{4,2}$

- We mentioned that  $df_1 = k - 1 = 4$ . Thus  $k = 5$ . This means we have 5 groups of data.
- We mentioned that  $df_2 = n - k = 2$ . Thus  $n = k + 2 = 7$ . This means we have total of 7 data points.
- This is an extreme example – 7 data points being divided into 5 groups.
- Later we will see a more realistic example...

- Is this a **one-sided** or two-sided test?
- What would be the boundary (or boundaries) for this 95% CI?

$$F = \frac{s_B^2}{s_W^2}.$$



```
>> finv(0.95,4,2)
```

```
ans =
```

```
19.2468
```

```
>>
```

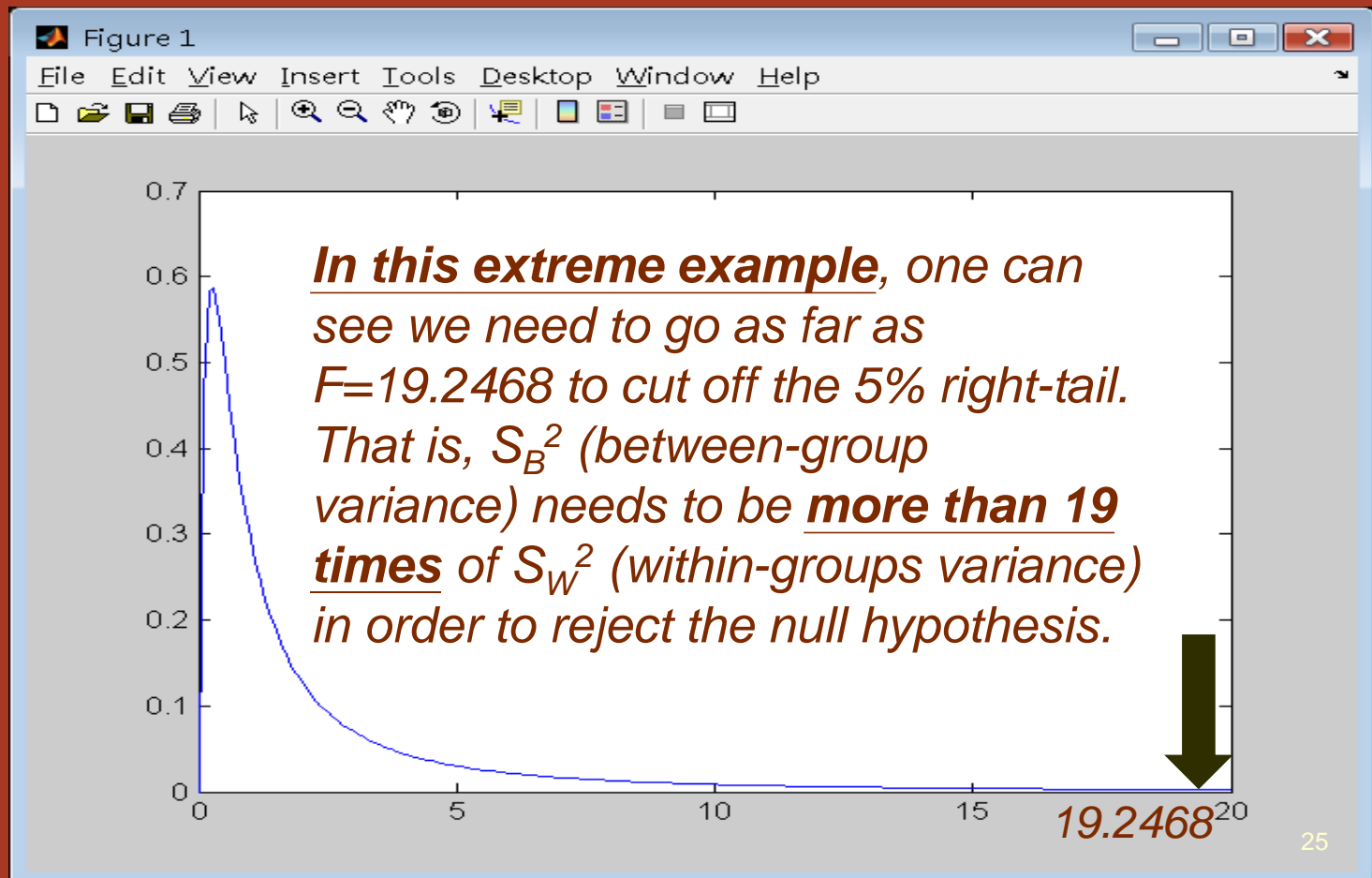
```
>> fcdf(ans,4,2)
```

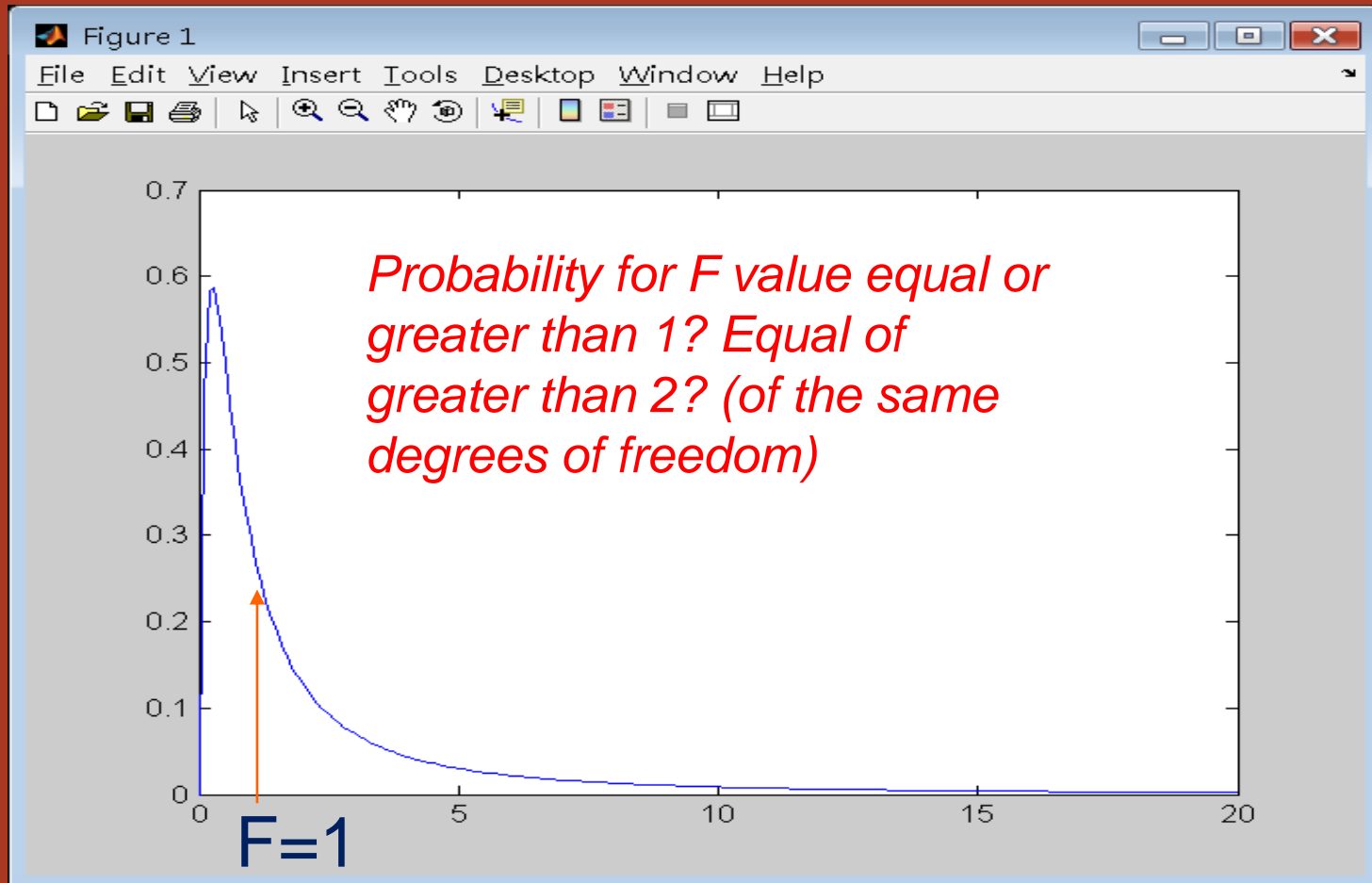
```
ans =
```

```
0.9500
```

```
>>
```

$$F = \frac{S_B^2}{S_W^2}.$$





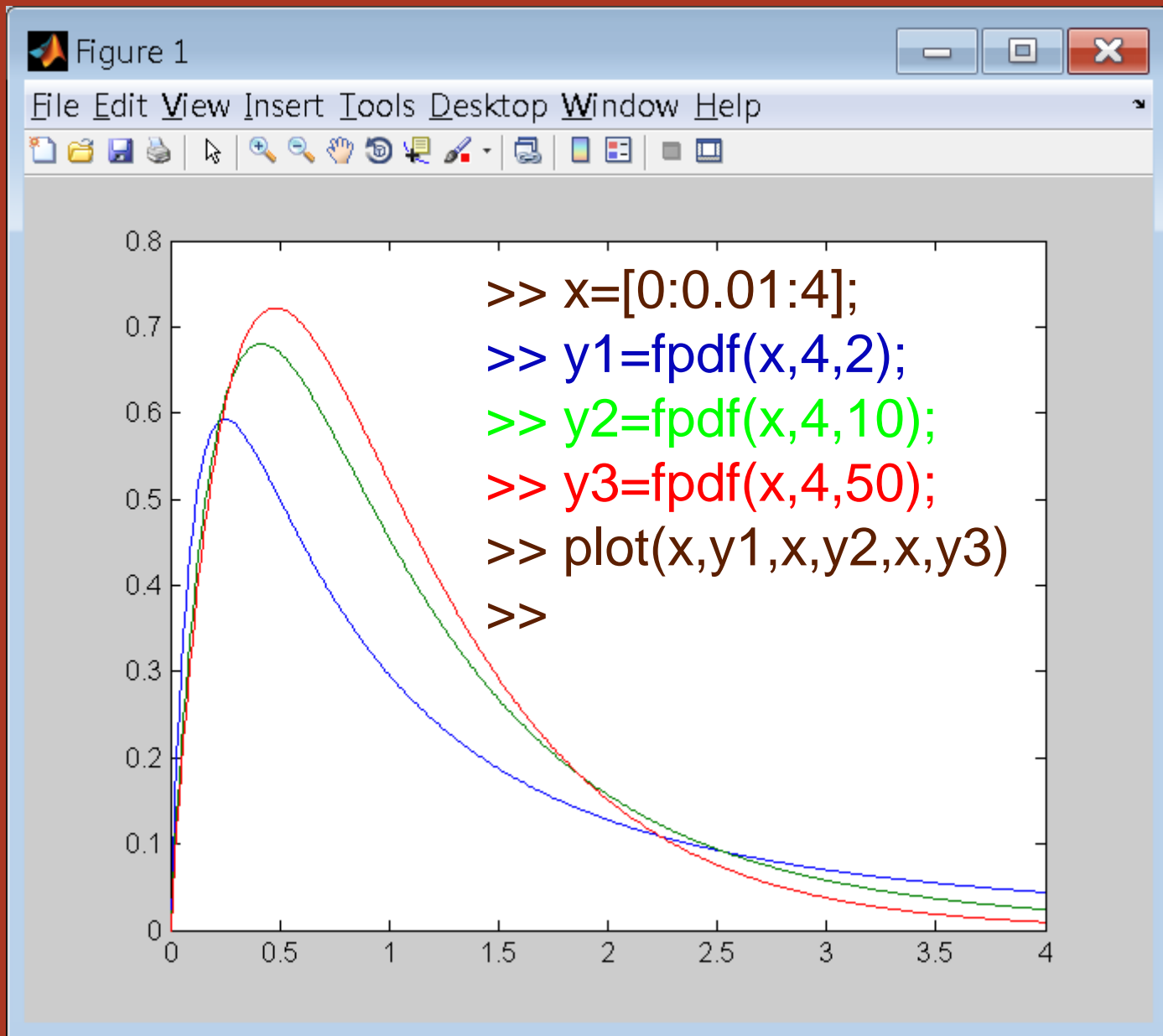
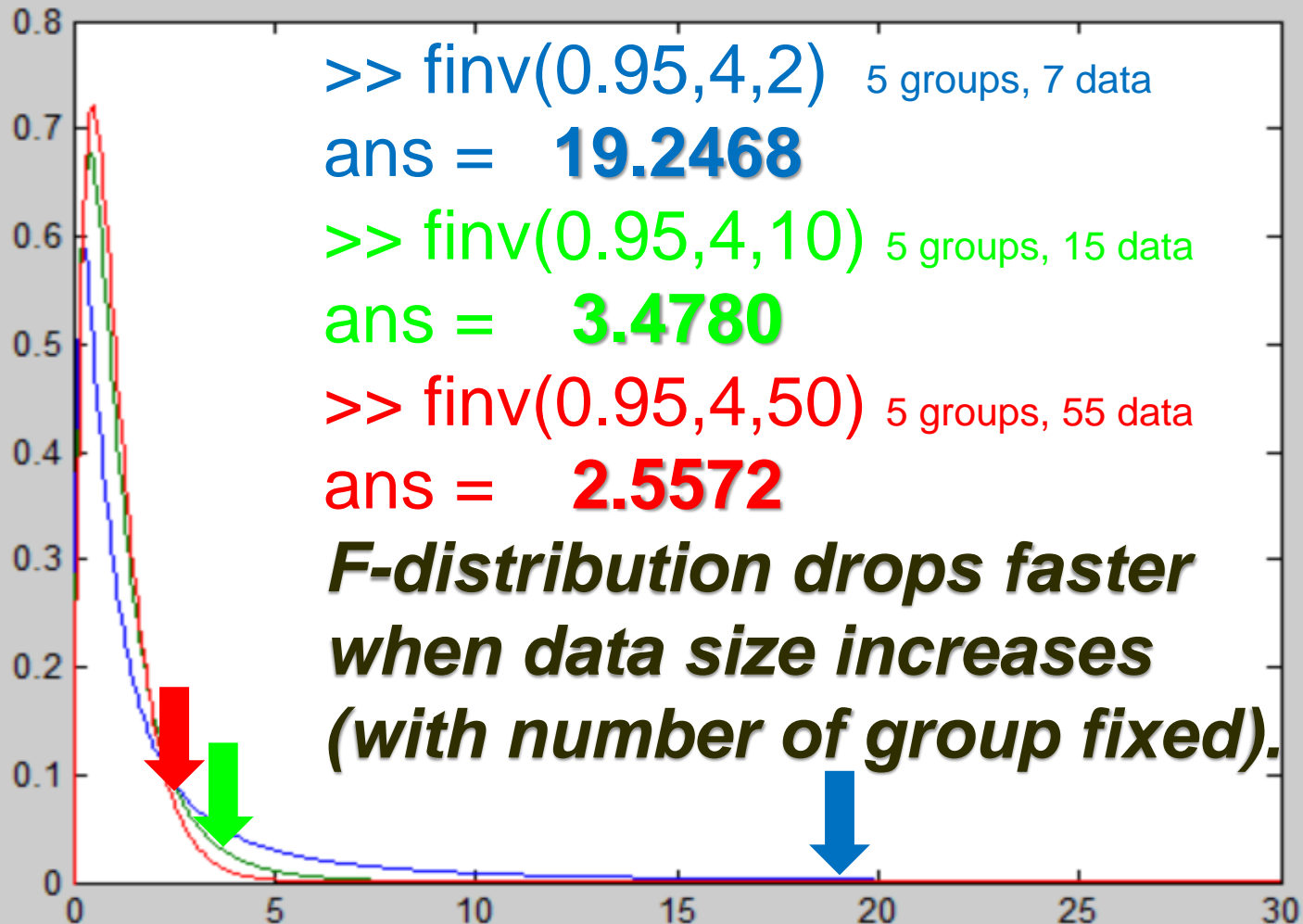


Figure 1

File Edit View Insert Tools Desktop Window Help



$F_{k-1, n-k}$



*F-distribution drops faster  
when data size increases  
(with number of group fixed)*

$$F = \frac{s_B^2}{s_W^2}.$$

- Drop faster means smaller F value (e.g., 2.5) will be able to go over the threshold and reject the null hypothesis. 【Cannot join them~】
- This suggests ***merging these (large amount of) data into one big dataset is challenging.*** 【Slightly bigger  $s_B^2$  or smaller  $s_W^2$  will fail the join~】



Figure 1

File Edit View Insert Tools Desktop Window Help

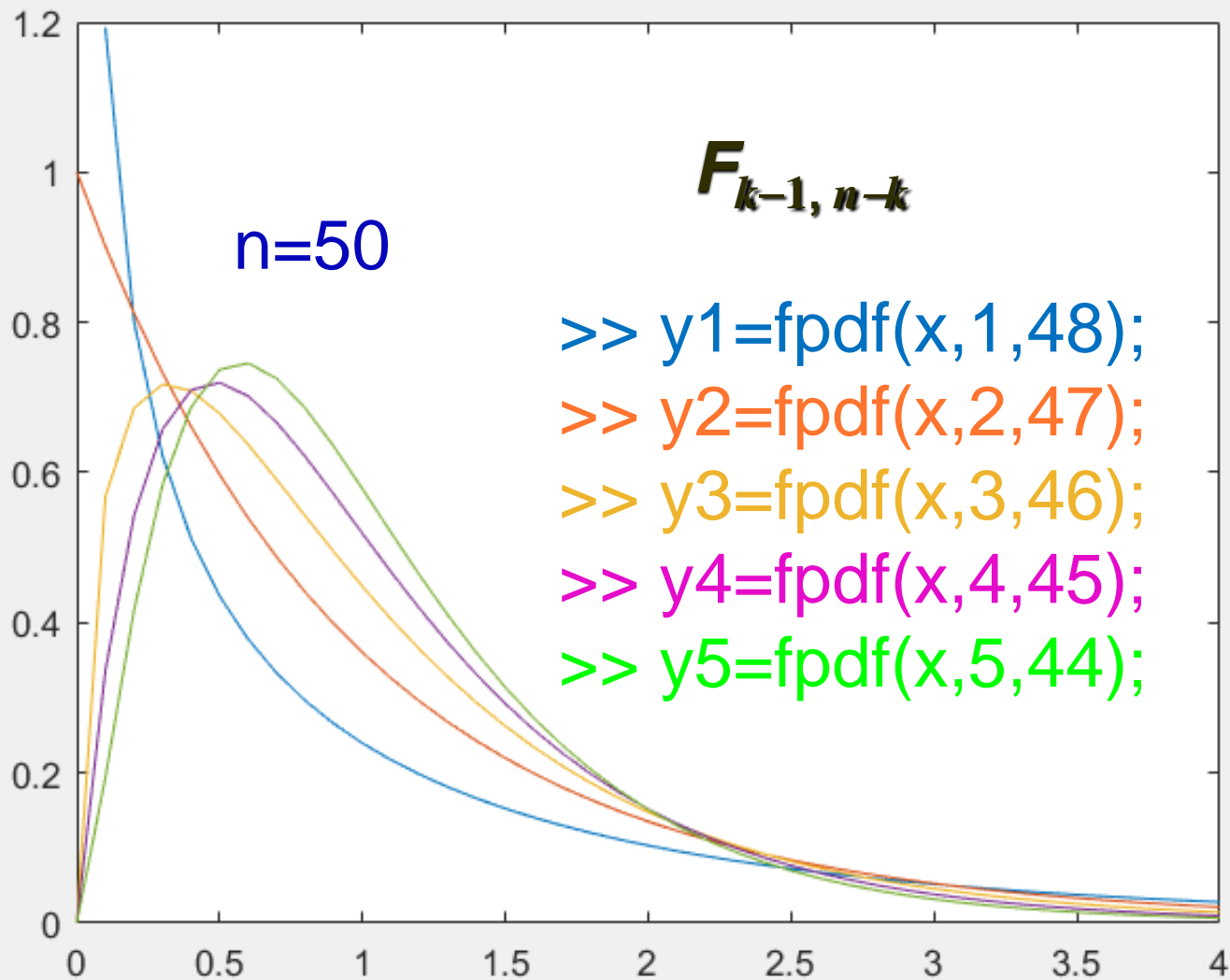
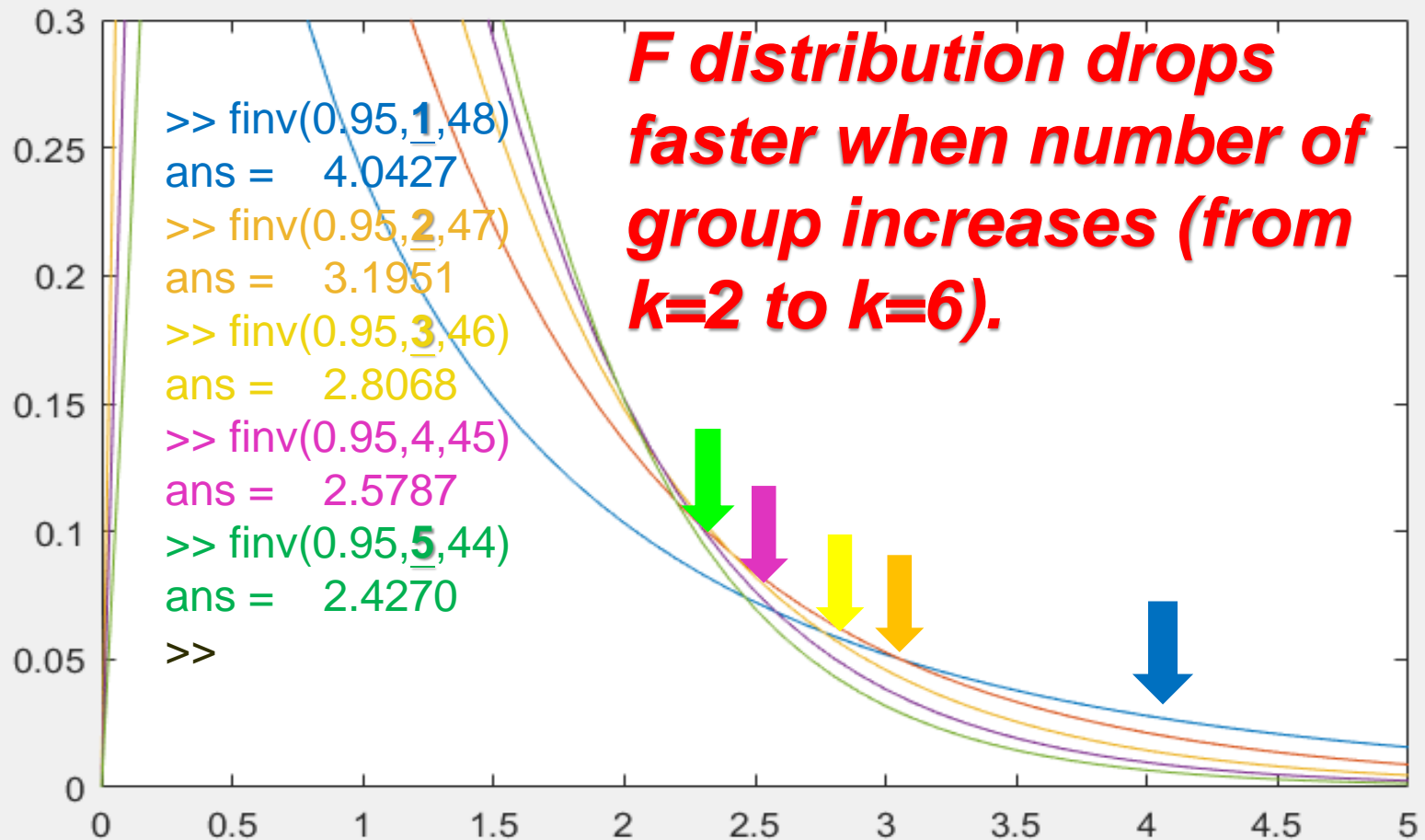




Figure 1

$$F_{k-1, n-k}$$



***F distribution drops faster when number of group increases (from  $k=2$  to  $k=6$ )***

$$F = \frac{s_B^2}{s_W^2}.$$

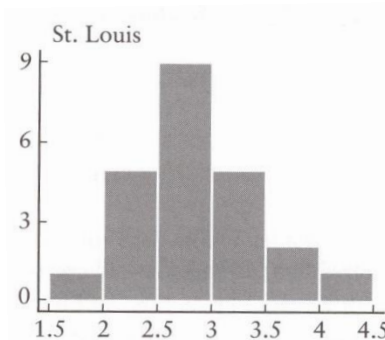
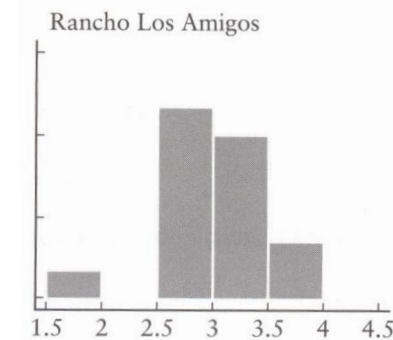
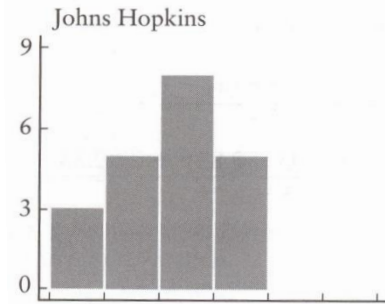
- Drop faster means smaller F value (e.g., 2.5) will be able to go over the threshold and reject the null hypothesis. 【Cannot join them~】
- This suggests ***merging many smaller groups into one is challenging.*** 【Slightly bigger  $s_B^2$  or smaller  $s_W^2$  will fail the join~】

# Back to the FEV Example

**TABLE 12.1**

Forced expiratory volume in one second for patients with coronary artery disease sampled at three different medical centers

Johns Hopkins	Rancho Los Amigos	St. Louis
3.23	3.22	2.79
3.47	2.88	3.22
1.86	1.71	2.25
2.47	2.89	2.98
3.01	3.77	2.47
1.69	3.29	2.77
2.10	3.39	2.95
2.81	3.86	3.56
3.28	2.64	2.88
3.36	2.71	2.63
2.61	2.71	3.38
2.91	3.41	3.07
1.98	2.87	2.81
2.57	2.61	3.17
2.08	3.39	2.23
2.47	3.17	2.19
2.47		4.06
2.74		1.98
2.88		2.81
2.63		2.85
2.53		2.43
		3.20
		3.53
$n_1 = 21$ $\bar{x}_1 = 2.63$ liters $s_1 = 0.496$ liters	$n_2 = 16$ $\bar{x}_2 = 3.03$ liters $s_2 = 0.523$ liters	$n_3 = 23$ $\bar{x}_3 = 2.88$ liters $s_3 = 0.498$ liters



1. An initial test upon their histogram show that they are approximately normal distributions.
2.  $n=21+16+23=60$  and  $k=3$ .
3. So  $df_1=k-1=2$ ,  $df_2=n-k=57$

4.  $F_{2,57}$

# Cont'd

- We first compute the grand mean needed in computing the between-group variance  $S_W^2$ :

$$\begin{aligned}\bar{x} &= \frac{n_1\bar{x}_1 + n_2\bar{x}_2 + n_3\bar{x}_3}{n_1 + n_2 + n_3} \\ &= \frac{21(2.63) + 16(3.03) + 23(2.88)}{21 + 16 + 23} \\ &= 2.83 \text{ liters,}\end{aligned}$$

$$\begin{aligned}n_1 &= 21 \\ \bar{x}_1 &= 2.63 \text{ liters} \\ s_1 &= 0.496 \text{ liters}\end{aligned}$$

$$\begin{aligned}n_2 &= 16 \\ \bar{x}_2 &= 3.03 \text{ liters} \\ s_2 &= 0.523 \text{ liters}\end{aligned}$$

$$\begin{aligned}n_3 &= 23 \\ \bar{x}_3 &= 2.88 \text{ liters} \\ s_3 &= 0.498 \text{ liters}\end{aligned}$$

- We then compute the two variances needed to get to F:

$$\begin{aligned}
 s_W^2 &= \frac{(n_1 - 1)s_1^2 + (n_2 - 1)s_2^2 + (n_3 - 1)s_3^2}{n_1 + n_2 + n_3 - 3} \\
 &= \frac{(21 - 1)(0.496)^2 + (16 - 1)(0.523)^2 + (23 - 1)(0.498)^2}{21 + 16 + 23 - 3} \\
 &= 0.254 \text{ liters}^2.
 \end{aligned}$$

$$\begin{aligned}
 s_B^2 &= \frac{n_1(\bar{x}_1 - \bar{x})^2 + n_2(\bar{x}_2 - \bar{x})^2 + n_3(\bar{x}_3 - \bar{x})^2}{3 - 1} \\
 &= \frac{21(2.63 - 2.83)^2 + 16(3.03 - 2.83)^2 + 23(2.88 - 2.83)^2}{3 - 1}
 \end{aligned}$$

$$= 0.769 \text{ liters}^2.$$

$n_1 = 21$ $\bar{x}_1 = 2.63$ liters $s_1 = 0.496$ liters	$n_2 = 16$ $\bar{x}_2 = 3.03$ liters $s_2 = 0.523$ liters	$n_3 = 23$ $\bar{x}_3 = 2.88$ liters $s_3 = 0.498$ liters
---	---	---

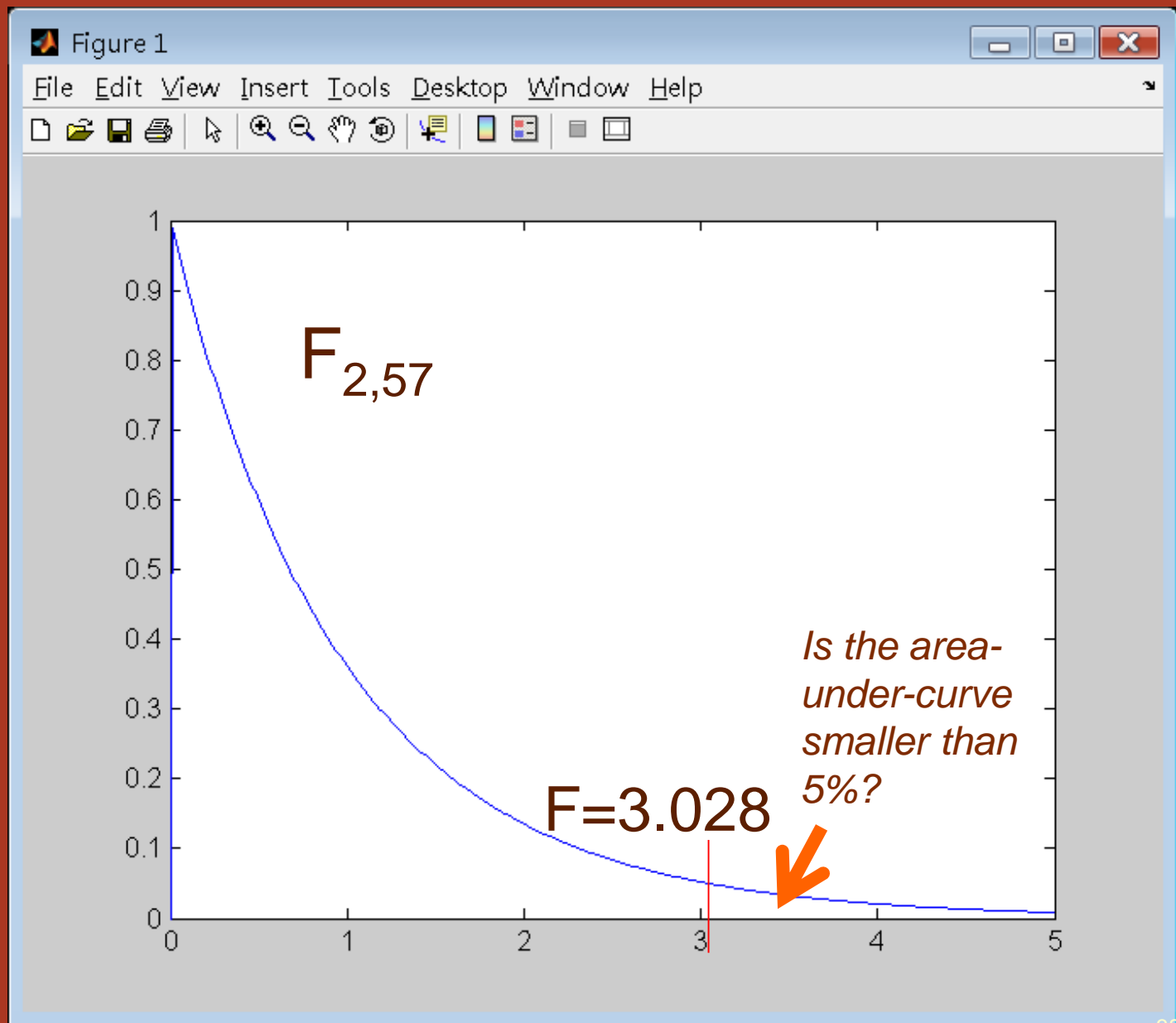
$$\begin{aligned} F &= \frac{s_B^2}{s_W^2} \\ &= \frac{0.769}{0.254} \\ &= 3.028. \end{aligned}$$

- We now see that “between-group variance” is more than 3 times the “within-groups variance”.
- Is this large enough to reject the null hypothesis (that the between-group variance is statistically large to prevent them from being joined together)?



$$\begin{aligned} F &= \frac{s_B^2}{s_W^2} \\ &= \frac{0.769}{0.254} \\ &= 3.028. \end{aligned}$$

- The two degrees of freedom needed are  $df_1 = k - 1 = 2$ , and  $df_2 = n - k = (21 + 16 + 23) - 3 = 57$ .
- So we will look for the distribution for  $F_{2,57}$ .



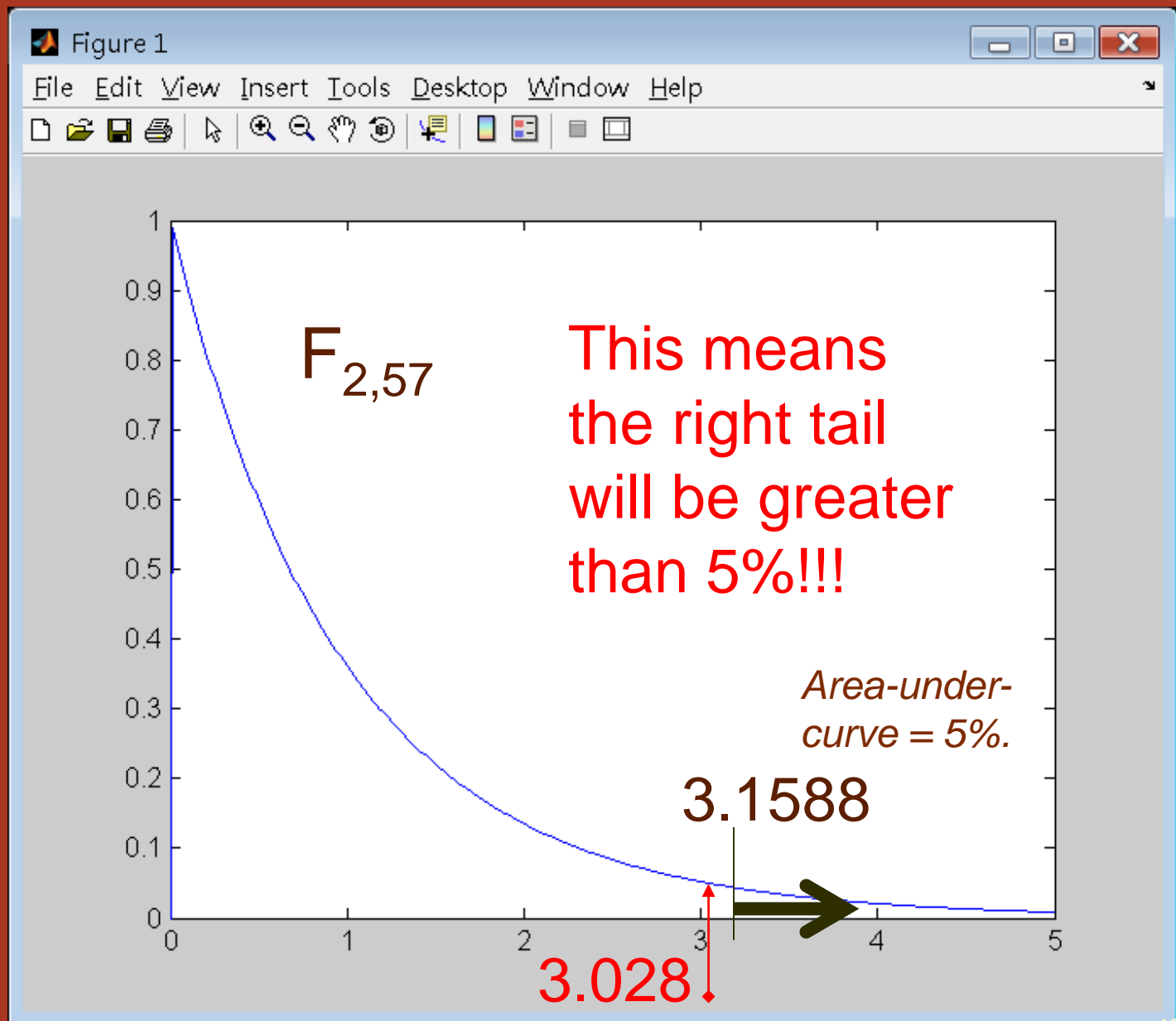
# MATLAB function FINV

FINV Inverse of the F cumulative distribution function.

$X = \text{FINV}(P, V1, V2)$  returns the inverse of the F distribution function with  $V1$  and  $V2$  degrees of freedom, at the values in  $P$ .

```
>> finv(0.95,2,57)  
ans = 3.1588  
>>
```

$x = 3.1588$  cuts  $F_{2,57}$  off a 5% right-tail. This means  $x = 3.028$  will cut more than 5%, or p-value > 0.05.



# MATLAB function FCDF

FCDF F cumulative distribution function.

$P = \text{FCDF}(X, V1, V2)$  returns the F cumulative distribution function with  $V1$  and  $V2$  degrees of freedom at the values in  $X$ .

```
>> fcdf(3.028,2,57)
ans = 0.9437
>>
```

$x=0$  to 3.028  
accumulates  $F_{2,57}$  to 0.9437. It leaves a right-tail of  $1-0.9437 = \underline{0.0563}$ . This is the p-value of the test.

# Conclusion

- Using  $\alpha=0.05$  level, we **will not reject** the null hypothesis. That is, there exists **negligible difference** in terms of mean values among these three samples.
- However, since this computed p-value is very close to 0.05, there might possibly be some difference in these samples to consider.
- We'd reject the hypothesis if  $\alpha=0.1$ , for example.

# 12.2 Multiple Comparisons Procedures

- Previously we have performed a “so-called” one-way ANalysis Of Variance (one-way ANOVA) that can test if  $k$  population means are identical or not.

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

$H_A$ : At least two means differ

# Cont'd

- If the computed p-value is smaller than the prescribed threshold, for example,  $\alpha=0.05$ , we would reject the hypothesis.
- That means there exists differences among those populations, so that the population means are not equal.
- *We, however, do not know whether all the means are different from one another, or only some of them are different.*



# Cont'd

- Once we reject the null hypothesis, therefore, we often want to conduct additional tests to find out **where the differences exist.**
- Many of these techniques are so-called **“multiple comparisons”**.
- They typically involve testing **each pair of means** individually.

# The Problems are...

- We mentioned that, to compare all pairs of 2-sample tests for  $k$  populations, we need to test  $C(k,2)$  pairs. When  $k$  gets large, the pairs to test would become expensive.
- **Even more important, another problem that arises when all possible 2-sample tests are conducted is that such procedure is likely to lead to an incorrect conclusion.**

# Falsely rejecting $H_0$

- Taking  $k = 3$ , for example. We assume that three population means are in fact equal, and we conduct all three pairwise tests (1 vs 2, 1 vs 3 and 2 vs 3).
- We set  $\alpha = 0.05$  (the significance level) for each of these 3 tests.
- Since each pair presumably contains “equal” population mean, the probability to (falsely) reject such a pairwise test would be 0.05 or less. (Recall that we call this type 1 error or rejecting error.)

## Cont'd

- The probability for ***failing to reject one single test*** here is at most  $1 - 0.05 = 0.95$ .
- By multiplicative rule, the probability of failing to reject this particular null hypothesis (of no difference in **all THREE** instances) would be

$$\begin{aligned} &P(\text{fail to reject in all 3 tests}) \\ &= (1 - 0.05)^3 = (0.95)^3 = \\ &0.857. \end{aligned}$$

# Cont'd

- Consequently, the probability of rejecting the null hypothesis in at least one of the 3 tests would be

$$\begin{aligned} &P(\text{reject in } \underline{\text{at least one test}}) \\ &= 1 - 0.857 = 0.143. \end{aligned}$$

- Note that this is much larger than our ideal  $\alpha = 0.05$  for a single 2-sample test.
- This is called a ***type I error*** ( $P(\text{reject } H_0 \mid H_0 \text{ is true})$ ), or so-called false negative. [We should not reject (since the p-value is greater than 0.05) but we rejected.]

# The solution...

- When evaluating pairwise 2-sample tests from multiple populations, we must ensure the overall probability for making a type I error shown earlier at  $\alpha=0.05$  too.
- **So a new significance level  $\alpha^*$  should be used:**

$$\alpha^* = \frac{0.05}{\binom{k}{2}}$$

This is called a Bonferrino correction

## Cont'd

- For example, for  $k = 3$ , we have  $C(3,2) = 3$  pairs to consider.
- The new significance level for each pair would be  $\alpha^* = 0.05/3 = 0.0167$ .
- This makes the overall type I error (as did before) still within 0.05.

$$\begin{aligned} P(\text{reject in at least one test}) &= \\ 1 - (1 - \underline{0.05/3})^3 &= 1 - (0.9833)^3 = \\ 1 - 0.9507 &= \underline{0.0493} \end{aligned}$$

More accurately, of course, you may compute  $\alpha^* = 1 - 0.95^{(1/3)}$   
 $= 1 - 0.9830 = 0.0170$

# Cont'd

- Use the same example for  $k = 3$ . Assume that we now change the predetermined  $\alpha = 0.1$ .
- The new significance level for each pair would be  $\alpha^* = 0.1/3 = 0.0333$ .
- This makes the overall type I error (as did before) still within 0.1.

$$\begin{aligned} P(\text{reject in at least one test}) &= \\ 1 - (1 - 0.1/3)^3 &= 1 - (0.9667)^3 = \\ 1 - 0.9034 &= \underline{\underline{0.0966}} \end{aligned}$$

More accurately, of course, you may compute  $\alpha^* = 1 - 0.90^{(1/3)}$   
 $= 1 - 0.9655 = 0.0345$



# The FEV example (again)

- This is the same example containing FEV measurements from 3 medical centers.
- We know that, at  $\alpha = 0.1$ , we'd reject the null hypothesis that all three means are equal. [Recall that we had the p-value 0.0563]
- ***The question now is – how is each pair differing? Could it be possible that two groups might contain comparable means so they can be merged?***

## Cont'd

- The new significance level for conducting each of these individual 2-sample tests would be

$$\alpha^* = \frac{0.10}{3} \\ = 0.033$$

More accurately, of course, you may compute  $\alpha^* = 1 - 0.90^{(1/3)}$   
 $= 1 - 0.9655 = 0.0345$

# Cont'd

- To conduct the 2-sample t test, we establish the null hypothesis below and use the formula to compute the t-statistics.
- Based the computed t, and the associated t-distribution, we compute for p-value to decide whether to reject the hypothesis or not. [**Subject to  $\alpha^* = 0.033$ , not  $\alpha = 0.1$** ] **1**

$$H_0: \mu_i = \mu_j,$$

$$s_p^2 = \frac{(n_1 - 1)s_1^2 + (n_2 - 1)s_2^2}{n_1 + n_2 - 2}$$

$$t_{ij} = \frac{\bar{x}_i - \bar{x}_j}{\sqrt{s_w^2[(1/n_i) + (1/n_j)]}},$$

We previously had, for a paired independent samples, a pooled variance  $s_p^2$ .



## Cont'd

- Note that we take advantage on using the  
**2** previously computed within-groups  $s_w^2$  for the estimated common variance, instead of simply using data from group i and j.
- The degree of freedom for the t-distribution, based on this variation  $s_w^2$ , **would be  $n-k$** , or  $21+16+23-3 = 57$  here. **3**

$$\begin{aligned}s_w^2 &= \frac{(n_1 - 1)s_1^2 + (n_2 - 1)s_2^2 + (n_3 - 1)s_3^2}{n_1 + n_2 + n_3 - 3} \\&= \frac{(21 - 1)(0.496)^2 + (16 - 1)(0.523)^2 + (23 - 1)(0.498)^2}{21 + 16 + 23 - 3} \\&= 0.254 \text{ liters}^2. \quad \text{We have already computed this in the F-test.}\end{aligned}$$

# Group 1 vs Group 2

$n_1 = 21$ $\bar{x}_1 = 2.63$ liters $s_1 = 0.496$ liters	$n_2 = 16$ $\bar{x}_2 = 3.03$ liters $s_2 = 0.523$ liters	$n_3 = 23$ $\bar{x}_3 = 2.88$ liters $s_3 = 0.498$ liters
---	---	---

$$\begin{aligned}t_{12} &= \frac{\bar{x}_1 - \bar{x}_2}{\sqrt{s_w^2[(1/n_1) + (1/n_2)]}} \\&= \frac{2.63 - 3.03}{\sqrt{0.254[(1/21) + (1/16)]}} \\&= -2.39.\end{aligned}$$

Note we took advantage on using the same variability (0.254) and associated DF (57) here that we previously computed in the F test.

For  $df=57$ , the t-distribution is nearly standard. So  $\pm 1.96$  cuts off 5% of the probability density distribution. Here we have  $\pm 2.39$ , which surely cuts off less than 5%. The problem is – does it cut off less than  $\alpha^*=0.033$ ?

$$\begin{aligned}
 t_{12} &= \frac{\bar{x}_1 - \bar{x}_2}{\sqrt{s_w^2[(1/n_1) + (1/n_2)]}} \\
 &= \frac{2.63 - 3.03}{\sqrt{0.254[(1/21) + (1/16)]}} \\
 &= -2.39.
 \end{aligned}$$

```
>> tcdf(-2.39,57)
ans = 0.0101
```

And p-value =  $2 \times 0.0101 = 0.0202$  (this is a 2-tailed test), which is smaller than 0.033. We thus conclude the **two means are NOT comparable.**

If using the formula  $S_p^2$  in Chapter 11, which is computed as 0.258. The new  $t_{12} = -2.37$ , and  $df = 21 + 16 - 2 = 35$ .

$$s_p^2 = \frac{(n_1 - 1)s_1^2 + (n_2 - 1)s_2^2}{n_1 + n_2 - 2}$$

```
>> tcdf(-2.37,35)
ans =
    0.0117
>>
```

This p-value is comparable with the one shown above using  $S_w^2$ .

# Comments

- Know your question clearly.
- Scenario #1 – Are you testing whether there exists difference between two samples (with no prior condition)? [As in Chap 11]
- Scenario #2 – Are you testing whether there exists difference between two samples, following a test that rejected the null hypothesis that 3 (or more) such samples are not comparable? [As in Chap 12]
- You can take advantage of using  $S_w^2$  in scenario #2, or using  $S_p^2$  as did in scenario #1. 【Note the DFs in two scenarios are different!】