One-way ANOVA

What is this used for?

- For comparison of more than three samples. (Recall that t-test is usually used for two sample comparison)
- Our concern here is whether there are the statistically significant difference in the mean among samples. In other words, the difference you observed is just because of sampling errors or not?
- Why “one-way”?: Just one independent variable.

Why "Analysis of Variance"?

- You will use three kinds of variance: Total variance, between variance, and within variance.
- H0: all of population means are the same. $\mu_1 = \mu_2 = \mu_3 = \ldots$
- H1: there is significant difference between "at least any two." E.g. $\mu_1 \neq \mu_2$ or $\mu_2 \neq \mu_3$ or $\mu_1 \neq \mu_2 \neq \mu_3 \ldots$
- The basic idea is if between variance is significantly bigger than within variance, you can reject H0.

Simple Example.
(Hours you worked last week)

White	Black	Other
10	3	4
6	4	1
5	5	6
6	2	5
7	6	8
4	5	7
5	8	9
9	7	8
8	9	12
9	9	10

What is H0 and H1?

- Some of you would ask there are three variables in the data. There is one independent variable (sometimes called “factor”), say race, but there are three “levels” for this variable. What is dependent variable? Your working hour.
- The sample mean is 6.9(White), 5.8(Black), and 7.0(Other). You are interested in whether the difference among three sample means is really significant at what level of confidence!
- H0: $\mu_{white}=\mu_{black}=\mu_{other}$
- H1: not H0. (there is the significant difference between at least any two groups)

Total Variance (SST)

- First of all, you have to calculate the mean for 30 cases (total cases!) I got 6.567.
- Secondly, you have to 1) subtract the grand mean from each score, 2) square each of these differences, and 3) add up squared scores.
- SST=189.367=$(10-6.567)^2+\ldots+(9-6.567)^2+$ $(3-6.567)^2+\ldots+(9-6.567)^2+(4-6.567)^2+\ldots+(10-6.567)^2$

  White Black other
- This equation is about this calculation process.

$$SST = \sum_{i=1}^{n} (Y_i - \bar{Y})^2$$

6.567

Between Variance (SSB)

- First of all, you have to calculate the mean for each group. 6.9(White), 5.8(Black), 7.0(Other).
- Secondly, you have to 1) subtract the grand mean from each sample mean, 2) square each of these differences, 3) multiply each squared score by corresponding each sample size, and 4) add up.
- SSB=$(6.9-6.567)^2(10)+(5.8-6.567)^2(10)+(7.0-6.567)^2(10)$

  =8.867.
- The following equation is about what you did.

$$SSB = \sum_{k=1}^{K} N_k(\bar{Y}_k - \bar{Y})^2$$

Within Variance (SSW)

- First, you already have got each sample mean: 6.9(White), 5.8(Black), 7.0(Other).
- Next, you have to 1) subtract each sample mean from each score within each group, 2) square the differences, 3) add up all of elements.
- SSW=180.5=$(10-6.9)^2+\ldots+(9-6.9)^2+$

  $(3-5.8)^2+\ldots+(9-5.8)^2+(4-7.0)^2+\ldots+(10-7.0)^2$

  White Black other

- This calculation process is expressed as the following equation. $SSW = \sum_K \sum_I (Y_{ik} - \bar{Y}_K)^2$

What if SSB is big enough?

- Wow! SST=SSB+SSW=189.367=8.867+180.5
- If SSB is big enough, what does this mean? The difference between the grand mean and each group mean is big. If null hypothesis is true, all of sample means should be equal, 6.567. However, your sample mean is 6.9, 5.8, 7.0.
- What if each sample mean is 12, 3.85, 3.85? (Note that the grand mean is the same as 6.567) Oh! There is big difference between the first sample mean and 3.85. So, we can say that as SSB become bigger, the chance of rejecting the null hypothesis becomes higher.

What if SSW is small enough?

- SSW is 180.5. In other words, SSW explains almost all of variance in SST. (180.5 is much bigger than 8.867)
- SSW indicates the difference between the sample mean for each group and each score in that group. Big SSW reflects a lot of variance in each group.
- For example, if 10 cases had 7.0 for the third group (other), then there is no variance.
- We might as well regard SSW as kind of standard error, that is, variability from one sample to another. As the fluctuation becomes smaller, the chance of rejecting the null hypothesis becomes higher.

F-test and F-distribution

- ANOVA is based on F-distribution. This distribution has two degree of freedoms, k-1 (df for between) and N-k (df for within). Here, k is the number of group, or the number of sample mean; N is total number of cases.
- Statistic F follows F-distribution (k-1, N-k)
- When MSB=SSB/dfb, MSW=SSW/dfw, then the statistic F is MSB/MSW.

So, What is the statistic F?

- SST=SSB+SSW=189.367=8.867+180.5
- What is df for between? K-1. What is the number of group? 3. Therefore, dfb is 2.
- What is df for within? N-k. What is the total case? 30. Therefore, dfw=30-3=27.
- What is MSB=SSB/dfb=8.867/2=4.434.
- What is MSW=SSW/dfw=180.5/27=6.685.
- What is F? F=MSB/MSW=4.434/6.685=0.663.
- Let's just suppose that your confidence level is 95%.

The Statistical Table for F-distribution

Within

Critical Values of the F Distribution for p = .05

between

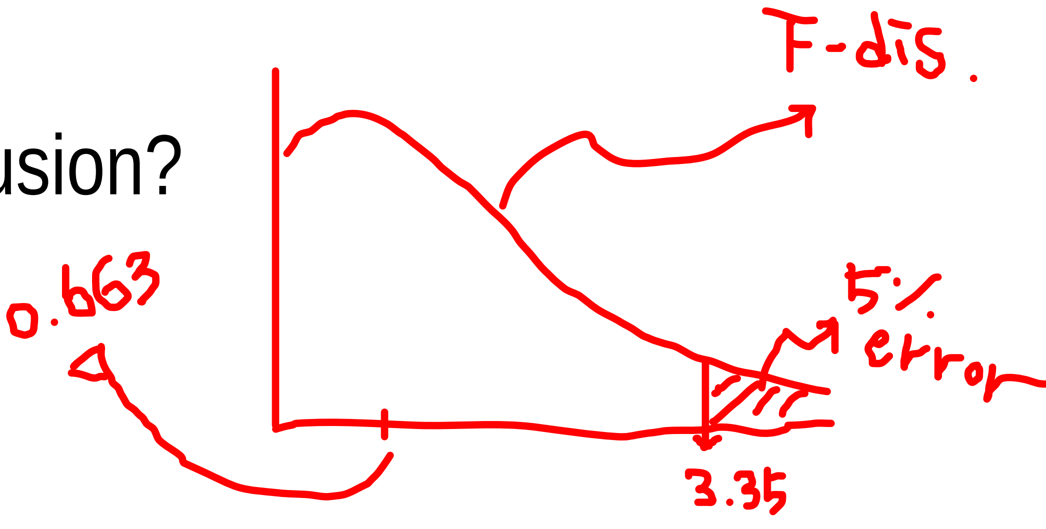
					df1					
df2	1	2	3	4	5	6	8	12	24	more
1	161.40	199.50	215.70	224.60	230.20	234.00	238.90	243.90	249.00	254.30
2	18.51	19.00	19.16	19.25	19.30	19.33	19.37	19.41	19.45	19.50
3	10.13	9.55	9.28	9.12	9.01	8.94	8.84	8.74	8.64	8.53
4	7.71	6.94	6.59	6.39	6.26	6.16	6.04	5.91	5.77	5.63
5	6.61	5.79	5.41	5.19	5.05	4.95	4.82	4.68	4.53	4.36
6	5.99	5.14	4.76	4.53	4.39	4.28	4.15	4.00	3.84	3.67
7	5.59	4.74	4.35	4.12	3.97	3.87	3.73	3.57	3.41	3.23
8	5.32	4.46	4.07	3.84	3.69	3.58	3.44	3.28	3.12	2.93
9	5.12	4.26	3.86	3.63	3.48	3.37	3.23	3.07	2.90	2.71
10	4.96	4.10	3.71	3.48	3.33	3.22	3.07	2.91	2.74	2.54
15	4.54	3.68	3.29	3.06	2.90	2.79	2.64	2.48	2.29	2.07
20	4.35	3.49	3.10	2.87	2.71	2.60	2.45	2.28	2.08	1.84
25	4.24	3.38	2.99	2.76	2.60	2.49	2.35	2.16	1.96	1.71
30	4.17	3.32	2.92	2.69	2.53	2.42	2.27	2.09	1.89	1.62
40	4.08	3.23	2.84	2.61	2.45	2.34	2.18	2.00	1.79	1.51
60	4.00	3.15	2.76	2.52	2.37	2.25	2.10	1.92	1.70	1.39
120	3.92	3.07	2.68	2.45	2.29	2.17	2.02	1.83	1.61	1.25
more	3.84	2.99	2.60	2.37	2.21	2.09	1.94	1.75	1.52	1.00

27

3.35 (appx.)

What's your judgment?

- Your statistic F is 0.663
- The critical value at the confidence level of 95% is 3.35 approximately.
- The statistic is less than the critical value. In other words, your statistic does not fall in the rejection region.
- So, what's your conclusion?



What is Eta?

- One question in your homework is about Eta squared. (Min is 0; Max is 1)
- To calculate this, use the following equation.

$$\eta^2 = \frac{SSB}{SST}$$

- What is Eta squared? 8.867/189.317=0.0468.
- Our result is very closer to 0. This implies that most of total variance could be explained by the variance within groups (rather than the variance between groups). In other words, the chance of rejecting the null hypothesis might be very small.

Keys

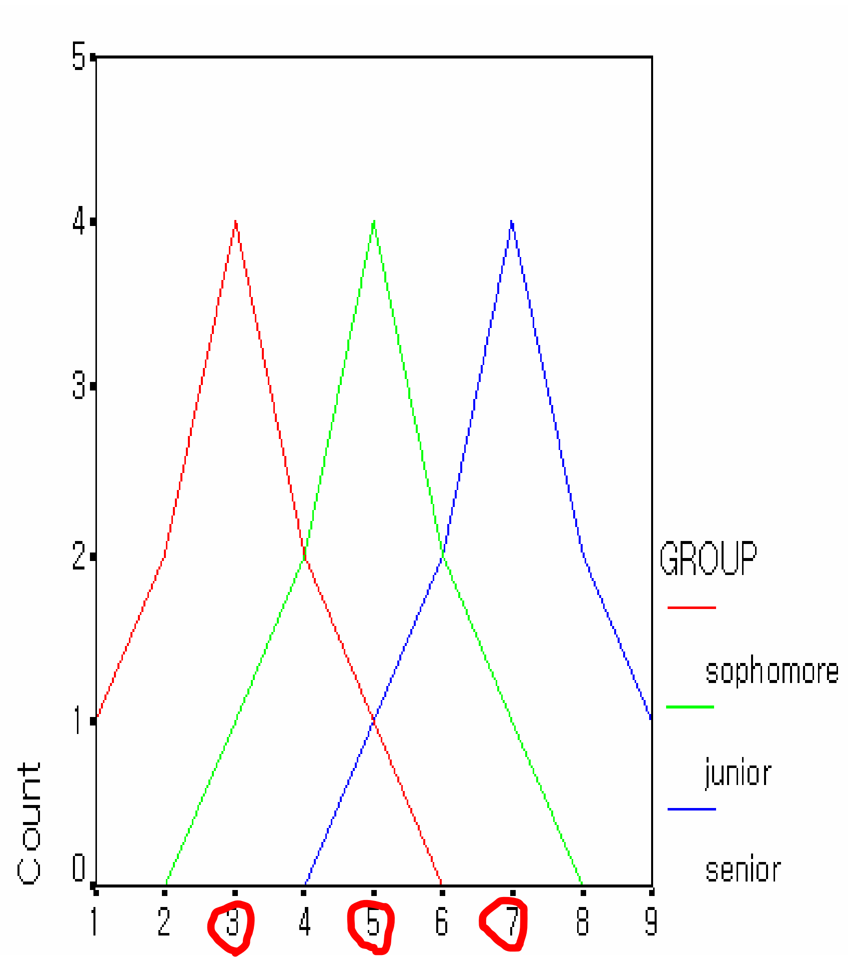
- Why do you prefer ANOVA (versus t-test)?
- What is the grand mean?
- What does alpha signify in ANOVA?
- What does J signify?
- SST=SSB+SSW (not SST=SSB/SSW or SST=SSB=SSW……)
- How to calculate each of them? (not computation, but conceptually)
- What are common sources of error?
- MSB=SSB/(J-1); MSW=SSW/(N-J)

- What is the definition of F-ratio?
- Why is this important?
- What is homoscedasticity?
- What is mean difference hypothesis test?
- What is eta-square (SSW/SST)? What is its range? Don’t confuse this with F-ratio (MSB/MSW)
- What is the purpose of post-hoc test?

One-tailed vs Two-tailed

- Two-tailed test is more conservative than one-tailed test. Why? You split the total error equally in two-tailed test! Recall that your test statistic will be positive or negative.
- Given the same amount of error you could tolerate, say 5%...
- In two-tailed test, you actually have 2.5% on one side (i.e., if your test statistic is positive, you apply right-side 2.5%; negative, left-side 2.5%)
- In one-tailed test, by contrast, you will apply generously right-side 5% if your test statistic is positive; left-side 5% if it is negative.
- Exactly saying, your generousness is double in one-tailed test than in two-tailed test!!!

Previous example (I)
(Another example of ANOVA.ppt)



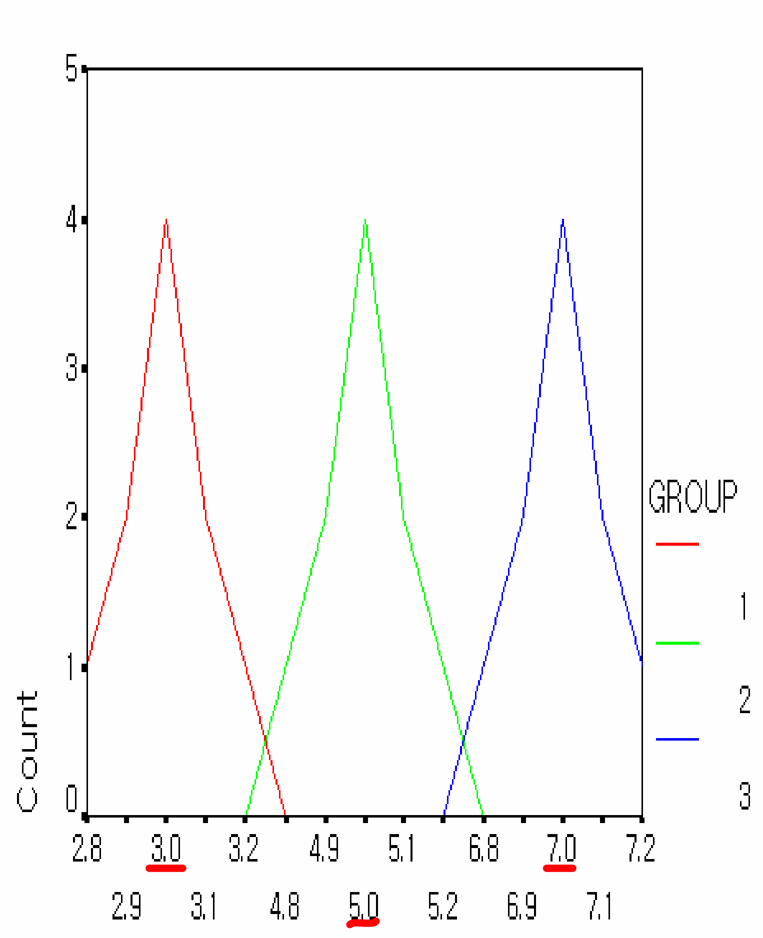
- Three line graphs overlap but not that much. This means that there is between variation.
- But, there is within variation as well because people have a variety of scores within a group. (Recall the concept of range. It is 4!)

One extreme case (II)

	score	group
1	2.8	1
2	2.9	1
3	2.9	1
4	3.0	1
5	3.0	1
6	3.0	1
7	3.0	1
8	3.1	1
9	3.1	1
10	3.2	1
11	4.8	2
12	4.9	2
13	4.9	2
14	5.0	2
15	5.0	2
16	5.0	2
17	5.0	2
18	5.1	2
19	5.1	2
20	5.2	2
21	6.8	3
22	6.9	3
23	6.9	3
24	7.0	3
25	7.0	3
26	7.0	3
27	7.0	3
28	7.1	3
29	7.1	3
30	7.2	3
31		

- Here is a situation where within variation is much smaller than before.
- Little variation around the average of each group: 2.8 to 3.2 for group 1; 4.8 to 5.2 for group 2; and 6.8 to 7.2 for group 3 (Range is 0.4!)

Continued…



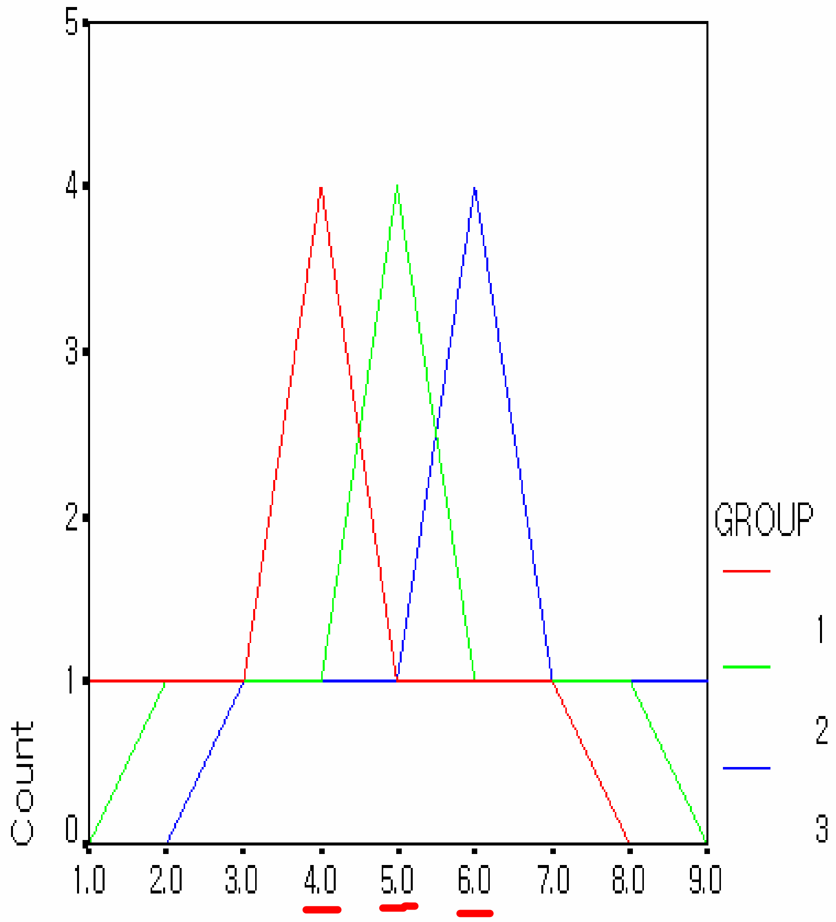
- Less overlap, which means that, although the average of each group is still the same (3, 5, 7), between variation becomes bigger compared to within variation.
- As is shown in the narrow width of graphs, within variation is much smaller.

Another extreme case (III)

	score	group
1	1.0	1
2	2.0	1
3	3.0	1
4	4.0	1
5	4.0	1
6	4.0	1
7	4.0	1
8	5.0	1
9	6.0	1
10	7.0	1
11	2.0	2
12	3.0	2
13	4.0	2
14	5.0	2
15	5.0	2
16	5.0	2
17	5.0	2
18	6.0	2
19	7.0	2
20	8.0	2
21	3.0	3
22	4.0	3
23	5.0	3
24	6.0	3
25	6.0	3
26	6.0	3
27	6.0	3
28	7.0	3
29	8.0	3
30	9.0	3
31		

- Here is another situation where between variation is much smaller. Why? The average is 4, 5, and 6 (Recall the interval is 2 in example I and II).
- But, there is still much of within variation: 1 to 7 (group 1); 2 to 8 (group 2); 3 to 9 (group 3). Range is 6!

Continued…



- As you expected, the degree of overlap is much bigger, which means that little between variation
- There is still much of within variation for each group, as it is shown in the broad width of graphs

In sum…

- The first case – we examined shortly today – is somewhere between the two extreme cases. Much of between variation is found, but this is the case with within variation.
- Again, the main goal of ANOVA is which one can explain more about the total variation (or total sum of squares, SST).
- Hence, it would be interesting to check the SPSS outcome.

Outcome I

N=30
J=3

within variance

Range

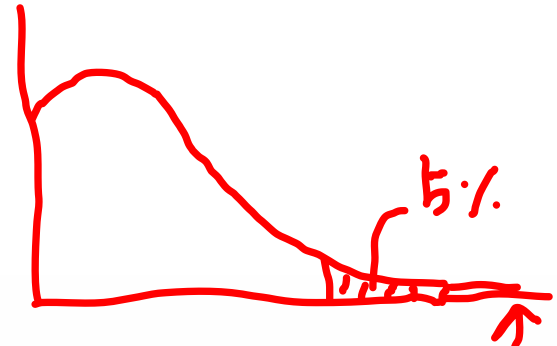
Descriptives

SCORE

	N	Mean	Std. Deviation	Std. Error	95% Confidence Interval for Mean		Minimum	Maximum
					Lower Bound	Upper Bound		
sophomore	10	3,00	1,155	,365	2,17	3,83	1	5
junior	10	5,00	1,155	,365	4,17	5,83	3	7
senior	10	7,00	1,155	,365	6,17	7,83	5	9
Total	30	5,00	2,000	,365	4,25	5,75	1	9

- Average is 3, 5, 7, and standard deviation is 1.155…
- So, this difference from one group to another (e.g. 2 for [sophomore vs junior] or [junior vs senior] or 4 for [sophomore vs senior]) is really statistically significant?
- In other words, is there any significant variation of the outcome (i.e., score) across groups? (i.e., between variation is bigger compared to within variation)

Outcome I



ANOVA

SCORE

	Sum of Squares	df	Mean Square	F	Sig.
Between Groups	80.000	2	40.000	30.000	.000
Within Groups	36.000	27	1.333		
Total	116.000	29			

- Yes! We have to reject the null hypothesis! P-value under “Sig.” is 0.000 which means that the amount of error you will have if you reject the null is actually 0%. Hence, you could reject it really really safely.
- What was the null? No difference in the average across groups. In other words, no between variation.
- Your conclusion is that between variation (rather than within variation) can explain more about the total variation. The difference of the average across groups is significant.

Outcome II

Descriptives

SCORE

	N	Mean	Std. Deviation	Std. Error	95% Confidence Interval for Mean		Minimum	Maximum
					Lower Bound	Upper Bound		
1	10	3,000	,1155	,0365	2,917	3,083	2,8	3,2
2	10	5,000	,1155	,0365	4,917	5,083	4,8	5,2
3	10	7,000	,1155	,0365	6,917	7,083	6,8	7,2
Total	30	5,000	1,6646	,3039	4,378	5,622	2,8	7,2

- Each average is the same (3, 5, 7), but standard deviation is 0.1155 which means that within variation becomes much smaller.
- Your question again… this difference from one group to another (e.g. 2 for [sophomore vs junior] or [junior vs senior] or 4 for [sophomore vs senior]) is really statistically significant but given the big change of within variation?
- In other words, is there any significant variation of the outcome (i.e., score) across groups? (i.e., between variation is bigger compared to within variation)

Outcome II

ANOVA

SCORE

	Sum of Squares	df	Mean Square	F	Sig.
Between Groups	80,000	2	40,000	3000,000	,000
Within Groups	,360	27	,013		
Total	80,360	29			

- Yes! You have to reject the null hypothesis again. Look at F-ratio! 3000. Much much bigger than 30 (outcome 1). Why? No wonder! Within variation is much much smaller here. F-ratio is 40/0.013. (FYI: F-ratio is 40/1.333)
- If you have big test-statistic (whatever it is), the chance that you have to reject the null hypothesis become higher. Why? The chance that your test-statistic falls into the rejection region becomes higher.

Outcome III

Descriptives

SCORE

	N	Mean	Std. Deviation	Std. Error	95% Confidence Interval for Mean		Minimum	Maximum
					Lower Bound	Upper Bound		
1	10	4.000	1.7638	.5578	2.738	5.262	1.0	7.0
2	10	5.000	1.7638	.5578	3.738	6.262	2.0	8.0
3	10	6.000	1.7638	.5578	4.738	7.262	3.0	9.0
Total	30	5.000	1.8937	.3457	4.293	5.707	1.0	9.0

- Average changed: 4, 5, 6. Standard deviation is 1.7638, which is the biggest. Much more overlap, which means we have less between variation here.
- The difference from one group to another (e.g. 1 for [sophomore vs junior] or [junior vs senior] or 2 for [sophomore vs senior]) is still statistically significant given the increase in within variation?
- In other words, is there any significant variation of the outcome (i.e., score) across groups? (i.e., between variation is bigger compared to within variation)

Outcome III

ANOVA

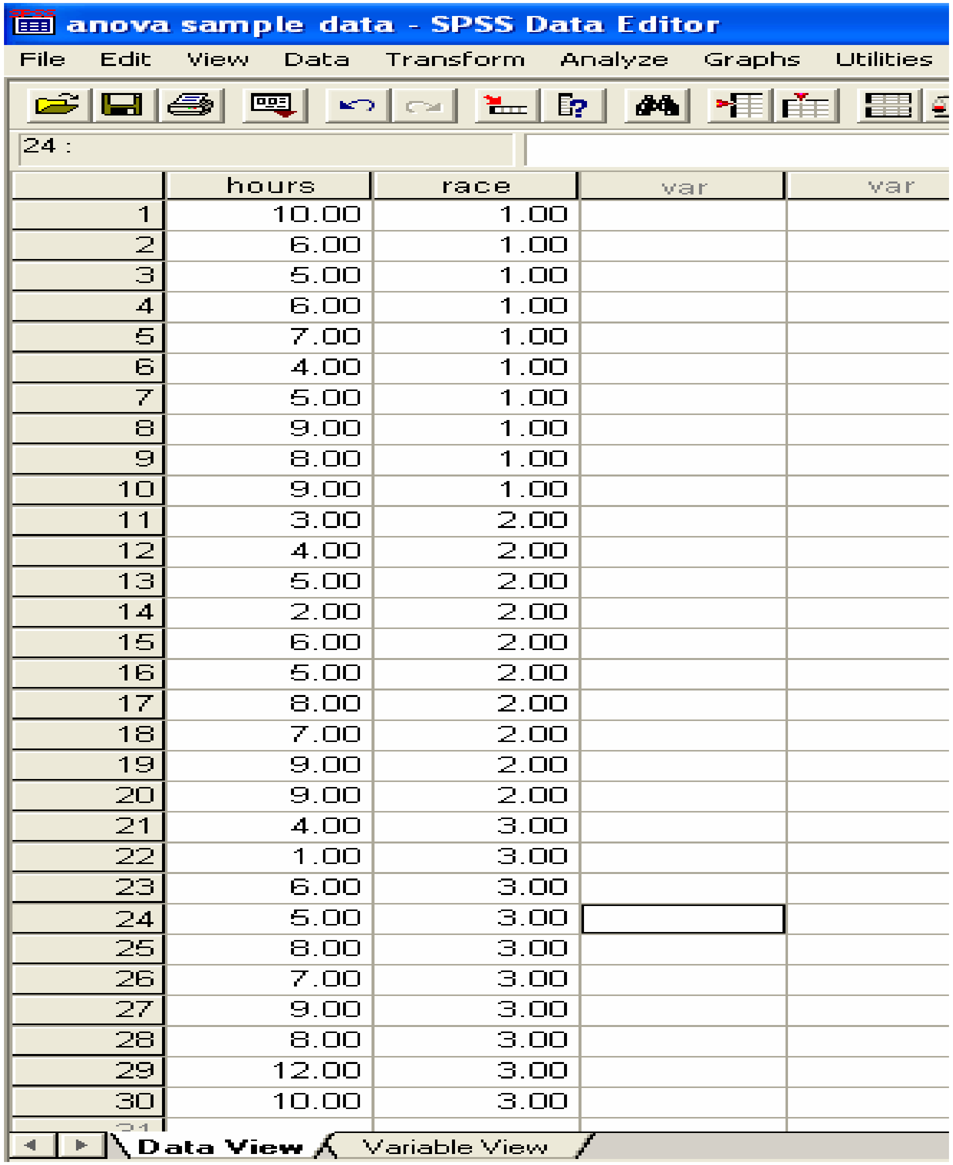
SCORE

	Sum of Squares	df	Mean Square	F	Sig.
Between Groups	20.000	2	10.000	3.214	.056
Within Groups	84.000	27	3.111		
Total	104.000	29			

- Unfortunately, you cannot reject the null hypothesis. Why? If you reject it, you will have 5.6% of error (Look at 0.056 under “Sig.” again), which is a little bit above the max amount of error you can tolerate (i.e., 5%).
- Within variation increased (i.e., big range) but between variation decreased (i.e., more overlap). Given this, it is not surprising that the ratio of MSB to MSW (i.e., F-ratio) decreased drastically.
- Again, you do not have a big test statistic enough to falls into the rejection region. It falls a little bit outside of that region.

How to use SPSS

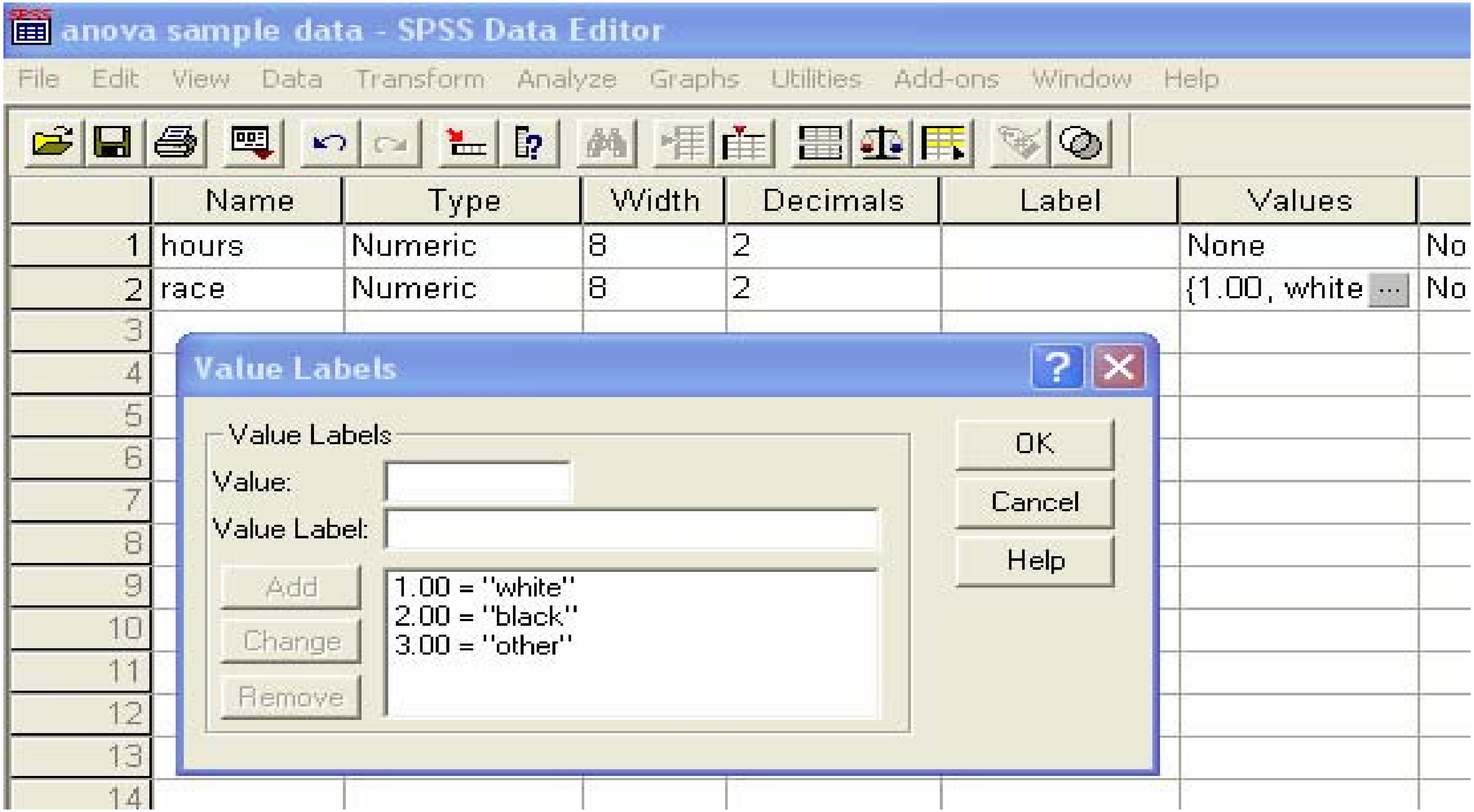
- Unless I am crazy, I am not going to do this kind of calculation again.
- Let’s use SPSS.



	hours	race	var	var
1	10.00	1.00		
2	6.00	1.00		
3	5.00	1.00		
4	6.00	1.00		
5	7.00	1.00		
6	4.00	1.00		
7	5.00	1.00		
8	9.00	1.00		
9	8.00	1.00		
10	9.00	1.00		
11	3.00	2.00		
12	4.00	2.00		
13	5.00	2.00		
14	2.00	2.00		
15	6.00	2.00		
16	5.00	2.00		
17	8.00	2.00		
18	7.00	2.00		
19	9.00	2.00		
20	9.00	2.00		
21	4.00	3.00		
22	1.00	3.00		
23	6.00	3.00		
24	5.00	3.00		
25	8.00	3.00		
26	7.00	3.00		
27	9.00	3.00		
28	8.00	3.00		
29	12.00	3.00		
30	10.00	3.00		

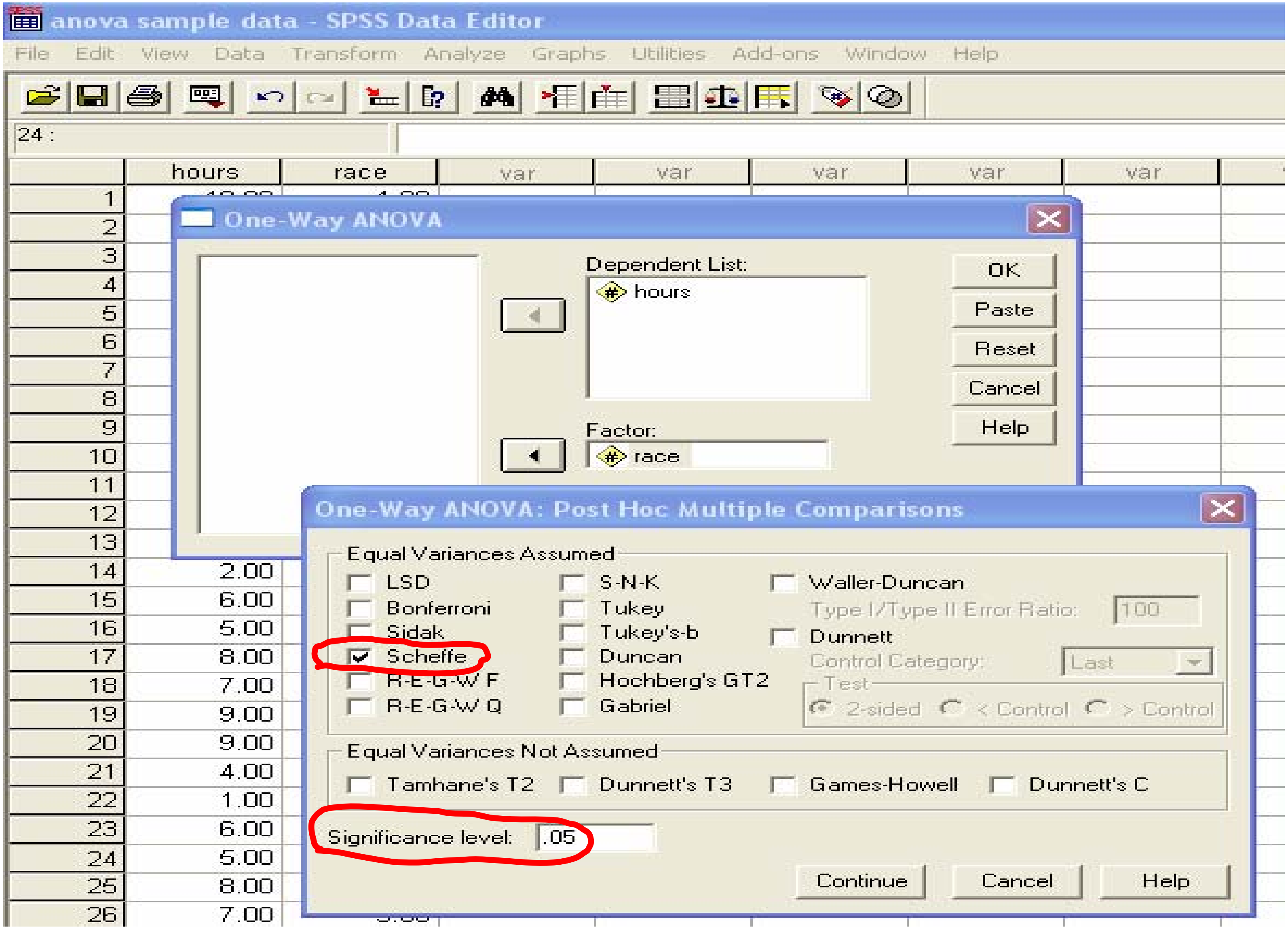
- “Hours” is dependent variable.
- “Race” is independent. (Factor) This has three levels (White, black, other)
- The total cases is 10+10+10=30.
- White is 1; black is 2; other is 3.

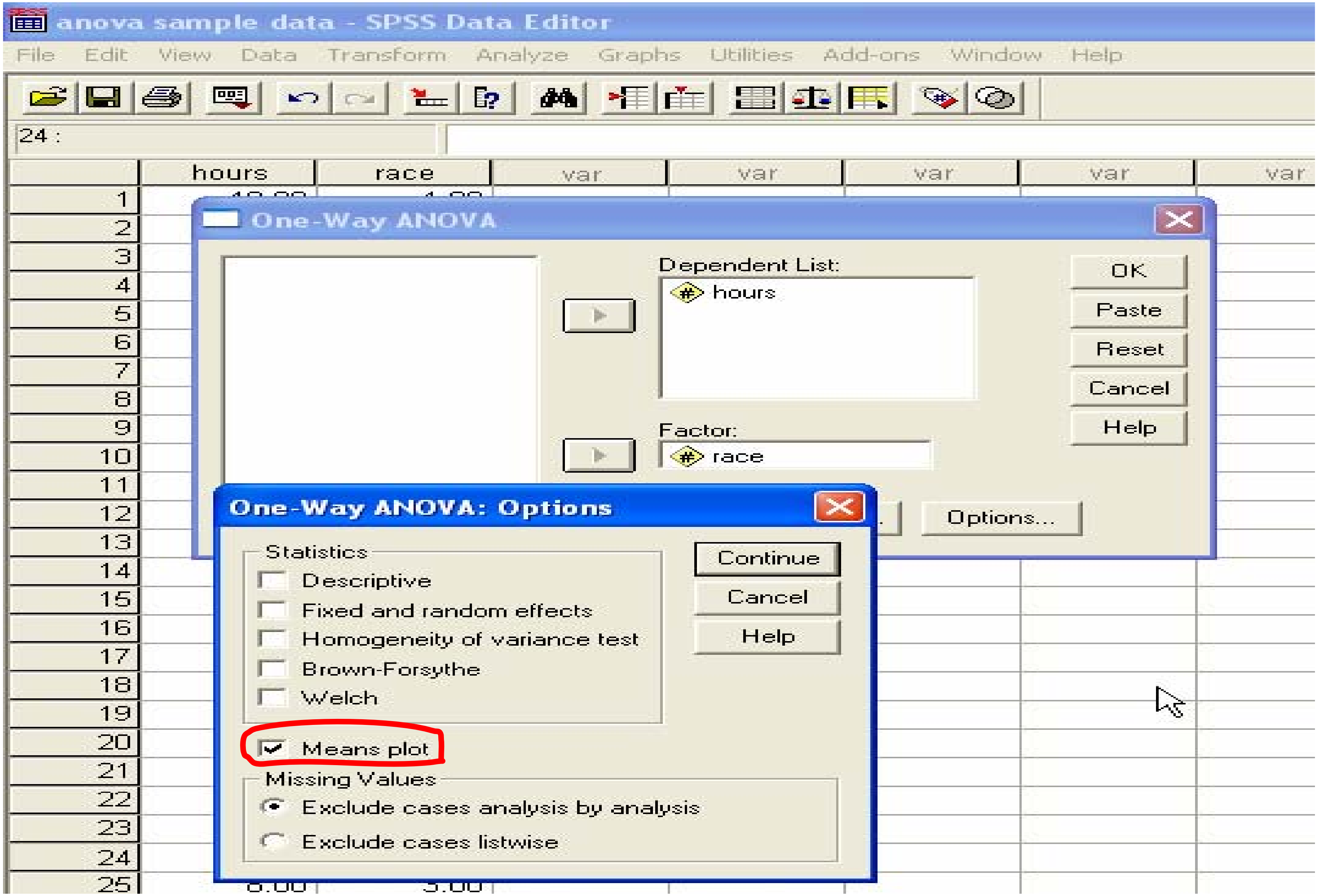
Variable View Capture



Analyze>Compare Means> One-way ANOVA

- Click the second tab about “Post Hoc”, and then click ‘Scheffe’ in the window about “equal variance assumed.” (You might as well choose “Bonferroni,” or “LSD” approach)
- Significance level is 0.05.
- Click the third tab about “option”, and then check the box about “Means plot.”
- Again, race is factor (independent variable), while working hour is dependent variable.


anova sample data - SPSS Data Editor
File Edit View Data Transform Analyze Graphs Utilities Add-ons Window Help
24 :
hours
race
var
One-Way ANOVA
Dependent List:
hours
Factor:
race
OK
Paste
Reset
Cancel
Help
One-Way ANOVA: Post Hoc Multiple Comparisons
Equal Variances Assumed
LSD
Bonferroni
Sidak
Scheffe
R-E-G-W F
R-E-G-W Q
S-N-K
Tukey
Tukey's-b
Duncan
Hochberg's GT2
Gabriel
Waller-Duncan
Type I/Type II Error Ratio: 100
Dunnett
Control Category: Last
Test
2-sided
< Control
> Control
Equal Variances Not Assumed
Tamhane's T2
Dunnett's T3
Games-Howell
Dunnett's C
Significance level: .05
Continue
Cancel
Help



anova sample data - SPSS Data Editor
File Edit View Data Transform Analyze Graphs Utilities Add-ons Window Help
24 :
hours
race
var
One-Way ANOVA
Dependent List:
hours
Factor:
race
OK
Paste
Reset
Cancel
Help
Options...
One-Way ANOVA: Options
Statistics
Descriptive
Fixed and random effects
Homogeneity of variance test
Brown-Forsythe
Welch
Means plot
Missing Values
Exclude cases analysis by analysis
Exclude cases listwise
Continue
Cancel
Help

➔ Oneway

ANOVA

hours

	Sum of Squares	df	Mean Square	F	Sig.
Between Groups	8.867	2	4.433	.663	.523
Within Groups	180.500	27	6.685		
Total	189.367	29			

- The result is the same as what I calculated before. You cannot reject the null hypothesis. (p-value is 0.523)

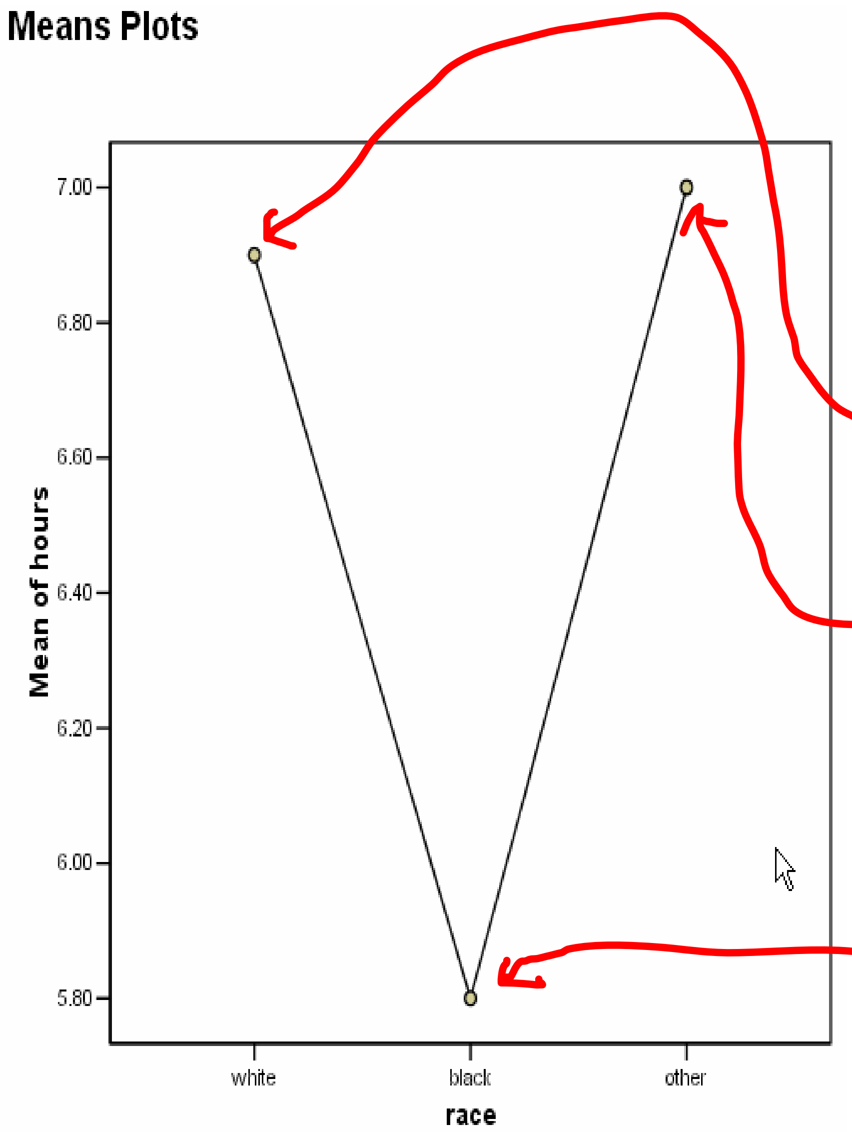
Post Hoc Tests

Multiple Comparisons

Dependent Variable: hours
Scheffe

(I) race	(J) race	Mean Difference (I-J)	Std. Error	Sig.	95% Confidence Interval	
					Lower Bound	Upper Bound
white	black	1.10000	1.15630	.641	-1.8949	4.0949
	other	-.10000	1.15630	.996	-3.0949	2.8949
black	white	-1.10000	1.15630	.641	-4.0949	1.8949
	other	-1.20000	1.15630	.590	-4.1949	1.7949
other	white	.10000	1.15630	.996	-2.8949	3.0949
	black	1.20000	1.15630	.590	-1.7949	4.1949

- What is so-called Scheffe test? H1 is that there is significant difference between at least any two. Do you remember? In a word, this test tells that there is difference between (white, black) or (white, other) or (black, other). The result is that all differences between any two are not significant.
- Not surprisingly, the difference between white and other is the most insignificant (p-value is 0.996) because two sample means are most similar!



- This plot is just about the position of each sample mean.
- The sample mean is 6.9(White), 5.8(Black), and 7.0(Other).