

Jednoczynnikowa Analiza Wariancji (ANOVA)

Wykład 11

- Przypomnienie: wykłady i zadania kursu były zaczerpnięte z podręczników:
 - *Statystyka dla studentów kierunków technicznych i przyrodniczych*, J. Koronacki, J. Mielniczuk, WNT 2004, wyd. II
 - *Introduction to the Practice of Statistics*, D. Moore, G. McCabe, Freeman 2003, wyd. IV
 - *Statistics for the Life Sciences*, M. Samuels, J. Witmer, 2003, wyd. III
 - H. Jasiulewicz i W. Kordecki \Rachunek prawdopodobieństwa i statystyka matematyczna. Przykłady i zadania", wyd. II

Plan wykładu:

- Wnioskowanie statystyczne dla jednoczynnikowej Analizy wariancji:
 - Format danych
 - Hipotezy i model ANOVy
 - Tabela ANOVy i test Fishera
- Porównywanie średnich parami

Jednoczynnikowa Analiza wariancji

- Rozważamy zmienną losową o wartościach liczbowych, np. poziom cholesterolu we krwi.
- Obserwacje pochodzą z kilku grup/populacji.
- Rozstrzygniemy czy średnie w grupach różnią się istotnie przez porównanie zmienności **pomiędzy** grupami ze zmiennością indywidualnych obserwacji **wewnątrz** grup.
- To jest **AN**alysis **Of** **V**ariance=**ANOVA**
- Grupy etykietuje jedna dodatkowa zmienna jakościowa, np. nazwa zastosowanej terapii, stąd **jednoczynnikowa ANOVA**.

Typowe dane

- ✓ Czas do zagojenia się danej rany przy kilku metodach terapii.
- ✓ Zarobki kobiet i mężczyzn w danym sektorze gospodarki.
- ✓ Okres użytkowania opon pochodzących od kilku producentów.
- Jeżeli zmienna jakościowa ma tylko dwie wartości, ANOVA jest równoważna testowi Studenta dla dwóch niezależnych prób.
- ANOVA pozwala na 3 i więcej grup.

**Przypomnienie: Dwie populacje,
nieznane odchylenia standardowe**

- Cel: testuj $H_0: \mu_1 = \mu_2$ przeciw $H_a: \mu_1 \neq \mu_2$
- σ_1, σ_2 są nieznane. Statystyka Studenta t to:

$$t = \frac{\bar{x}_1 - \bar{x}_2}{\sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}}}$$

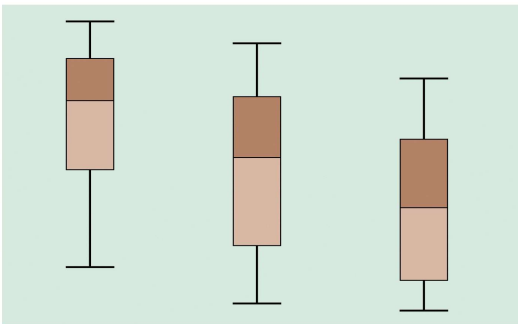
Wtedy $t^2 = \frac{n(\bar{x}_1 - \bar{x}_2)^2}{s_1^2 + s_2^2}$.

Statystyka Fishera F_S (poniżej) stanowi pewne uogólnienie.

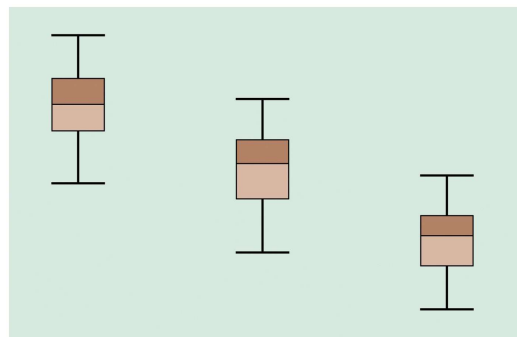
**Zasada: porównać zmienność *pomiędzy* i
wewnątrz grup**

- Badanie graficzne:
 - porównanie histogramów wszystkich grup
 - wykresy pudełkowe (box-plot) obok siebie
- *Statystyczna istotność* różnic pomiędzy grupami zależy od:
 - różnic średnich
 - odchyłeń standardowych
 - wielkości grup

**Wykresy pudełkowe dla terapii
antycholesterolowych. Czy terapie istotnie się
różnią?**



A teraz? (Mniejsza zmienność wewnątrz grup.)



ANOVA: hipotezy

H_0 : Średnie w grupach (populacjach) są równe.

H_a : Nie wszystkie średnie są równe.

- W H_a nie specyfikujemy, które średnie się różnią.
- Można to następnie badać przez porównania parami.
- „Grupa” znaczy tutaj populacja.
- Dla *danych* „grupa” to próba z danej populacji.

ANOVA: notacja

n = liczba wszystkich obserwacji

I = liczna grup

$\bar{x}$ = (wielka) średnia dla wszystkich danych

- i -ta grupa ma:
 - n_i = # obserwacji
 - x_{ij} = wartość dla j -tej obserwacji
 - $\bar{x}_j$ = średnią
 - s_i = próbkowe odchylenie standardowe

ANOVA: model

$$x_{ij} = \mu_i + \varepsilon_{ij},$$

gdzie ε_{ij} niezal. $N(0, \sigma)$

- Tutaj μ_i oznacza (nieznana) średnią z i -tej populacji,
- σ jest (nieznany) odchyleniem standardowym, o którym zakładamy, że jest równe dla wszystkich populacji.

Estymatory parametrów:

- Reguła dla odchyłeń standardowych w ANOVie: Jeżeli iloraz różnych s_i nie przekracza 2, to możemy utrzymywać, że σ nie zależy od grupy i używać ANOVy.
- Estymatorem μ_i jest $\bar{x}_j$.
- Estymatorem σ jest

$$s_p = \sqrt{s_p^2}, \text{ where}$$

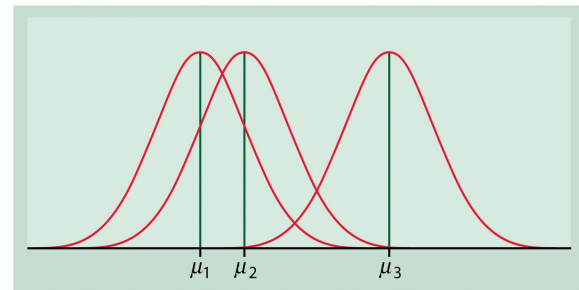
$$s_p^2 = \frac{(n_1 - 1)s_1^2 + \dots + (n_I - 1)s_I^2}{(n_1 - 1) + \dots + (n_I - 1)}$$

ANOVA: hipotezy (ponownie)

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

H_a : Nie wszystkie średnie z populacji są równe.

Typowa ilustracja: równe odchylenia standardowe, różne średnie w trzech populacjach:



Przykład: Zbadamy czytanie ze zrozumieniem dla trzech grup uczniów szkoły podstawowej **przed** i **po** zastosowania trzech różnych metod nauki czytania: Basal, DRTA, Strat.

TABLE 12.1 Pretest reading scores

Group	Subject	Score	Group	Subject	Score	Group	Subject	Score
Basal	1	4	DRTA	23	7	Strat	45	11
Basal	2	6	DRTA	24	7	Strat	46	7
Basal	3	9	DRTA	25	12	Strat	47	4
Basal	4	12	DRTA	26	10	Strat	48	7
Basal	5	16	DRTA	27	16	Strat	49	7
Basal	6	15	DRTA	28	15	Strat	50	6
Basal	7	14	DRTA	29	9	Strat	51	11
Basal	8	12	DRTA	30	8	Strat	52	14
Basal	9	12	DRTA	31	13	Strat	53	13
Basal	10	8	DRTA	32	12	Strat	54	9
Basal	11	13	DRTA	33	7	Strat	55	12
Basal	12	9	DRTA	34	6	Strat	56	13
Basal	13	12	DRTA	35	8	Strat	57	4
Basal	14	12	DRTA	36	9	Strat	58	13
Basal	15	12	DRTA	37	9	Strat	59	6
Basal	16	10	DRTA	38	8	Strat	60	12
Basal	17	8	DRTA	39	9	Strat	61	6
Basal	18	12	DRTA	40	13	Strat	62	11
Basal	19	11	DRTA	41	10	Strat	63	14
Basal	20	8	DRTA	42	8	Strat	64	8
Basal	21	7	DRTA	43	8	Strat	65	5
Basal	22	9	DRTA	44	10	Strat	66	8

Wstępne pytanie naukowe:

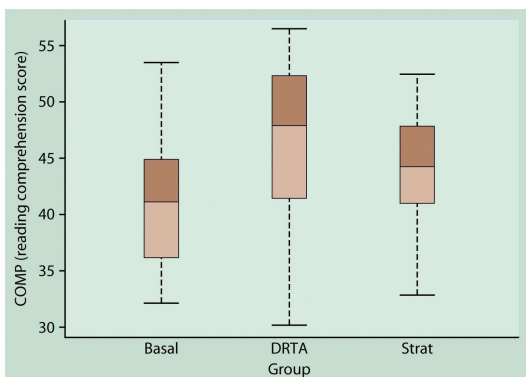
Czy te grupy różnią się poziomem przed rozpoczęciem kursu?

Symbolicznie:

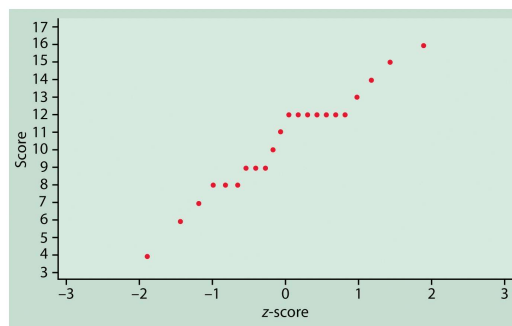
$$H_0: \mu_1 = \mu_2 = \mu_3$$

H_a : Nie wszystkie średnie są różne.

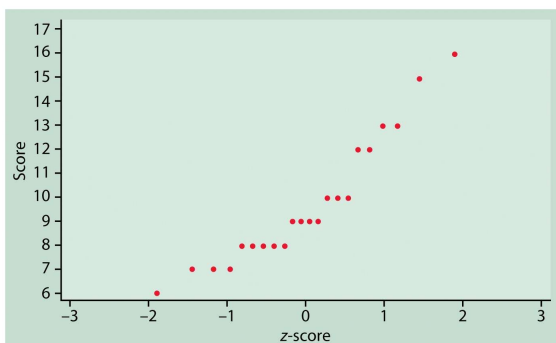
Wykresy pudełkowe dla grup **przed** kursem.



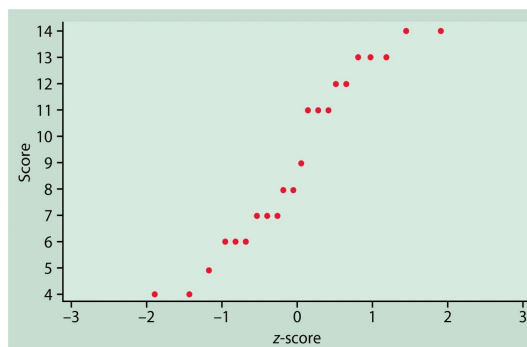
Ocena normalności (QQ-plot) dla grupy Basal przed kursem



Ocena normalności (QQ-plot) dla grupy DRTA przed kursem



Ocena normalności (QQ-plot) dla grupy Strat przed kursem



Statystyki dla każdej z grup. Przypomnienie:
pytamy, czy grupy różnią się.

Analysis Variable : SCORE			
----- GROUP=Basal -----			
N	Mean	Std Dev	
22	10.5000000	2.9720924	
----- GROUP=DRTA -----			
N	Mean	Std Dev	
22	9.7272727	2.6935871	
----- GROUP=Strat -----			
N	Mean	Std Dev	
22	9.1363636	3.3423039	

Czy można użyć wielu testów Studenta dla porównań parami?

Tak, ale..

- **Wielokrotne porównania** (tutaj 3) pomiędzy wszystkimi parami zwiększają prawdopodobieństwo fałszywego odrzucenia hipotezy zerowej.
- **Oszacowanie odchylenia standardowego σ** będzie dokładniejsze przy wykorzystaniu wszystkich obserwacji, jak w ANOVie.
- Użyjemy **ANOVy** i testu Fishera **F**.

Tabela ANOVy

- Trzy kategorie: „pomiędzy”, „wewnątrz” i „całkowite” (between, within, total).
- Trzy wielkości do obliczenia: SS, df, MS.
- Będziemy często przetwarzać dane komputerem.

	DF stopnie swobod y	SS suma kwadratów odchylen	MS Średnie kwadraty odchylen
Between Groups	I-1	z danych	SSG/DFG
Within groups (called Error)	n-I	z danych	SSE/DFE
Total	n-1	z danych	SST/DFT

Wzory:

$$SSG = \sum_{groups} n_i (\bar{x}_i - \bar{x})^2$$

$$SSE = \sum_{groups} (n_i - 1) s_i^2$$

$$SST = \sum_{obs} (x_{ij} - \bar{x})^2$$

Note: SST=SSG+SSE, DFT=DFG+DFE.

Wyniki ANOVy w SASie

General Linear Models Procedure					
Dependent Variable: SCORE					
Source	DF	Sum of Squares	Mean Square	F Value	Pr > F
Model	2	20.575758	10.287879	1.13	0.3288
Error	63	572.454545	9.086580		
Corrected Total	65	593.030303			
	R-Square	C.V.	Root MSE	SCORE Mean	
	0.034696	30.79723	3.0144	9.7879	

F test Fishera

- Statystyka testowa:

$$F_s = \text{MSG}/\text{MSE}.$$

- Przy H_0 , F_s ma tzw. rozkład Fishera-Snedecora z parametrami DFG, DFE.
- W tablicy F na WWW podane są wartości krytyczne/progi odrzucenia dla danych DFG, DFE.
 - „Degrees of freedom in the numerator” = DFG.
 - „Degrees of freedom in the denominator” = DFE.
- Duże wartości statystyki F_s odrzucają H_0 .

Density curve of an F distribution

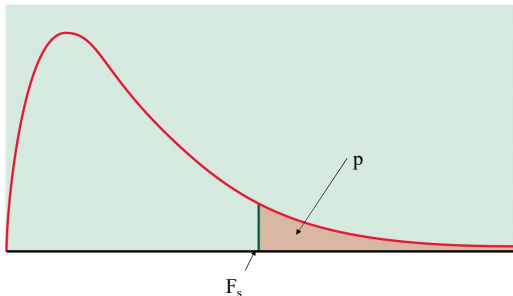


TABLE F F critical values									
		Degrees of freedom in the numerator							
		1	2	3	4	5	6	7	8
Degrees of freedom in the denominator	1	.100	39.86	49.50	53.59	55.83	57.24	58.20	58.91
	.050	161.45	199.50	215.71	224.58	230.16	233.99	236.77	238.88
	.025	647.79	799.50	864.16	899.58	921.85	937.11	948.22	956.66
	.010	4052.2	4999.5	5403.4	5624.6	5763.6	5859.0	5928.4	5981.1
	.001	405284	500000	540375	562500	576405	585937	592873	598144
	.100	8.53	9.00	9.16	9.24	9.29	9.33	9.35	9.37
	.050	18.51	19.00	19.16	19.25	19.30	19.33	19.35	19.37
	.025	38.51	39.00	39.17	39.25	39.30	39.33	39.36	39.37
	.010	98.50	99.00	99.17	99.25	99.30	99.33	99.36	99.37
	.001	998.50	999.00	999.17	999.25	999.30	999.33	999.36	999.37
2	.100	5.54	5.46	5.39	5.34	5.31	5.28	5.27	5.25
	.050	10.13	9.55	9.28	9.12	9.01	8.94	8.89	8.85
	.025	17.44	16.04	15.44	15.10	14.88	14.73	14.62	14.54
	.010	34.12	30.82	29.46	28.71	28.24	27.91	27.67	27.49
	.001	167.03	148.50	141.11	137.10	134.58	132.85	131.58	130.62
3	.100	4.54	4.32	4.19	4.11	4.05	4.01	3.98	3.95
	.050	7.71	6.94	6.59	6.39	6.26	6.16	6.09	6.04
	.025	12.22	10.65	9.98	9.60	9.36	9.20	9.07	8.98
	.010	21.20	18.00	16.69	15.98	15.52	15.21	14.98	14.80
	.001	74.14	61.25	56.18	53.44	51.71	50.53	49.66	49.00
4	.100	4.06	3.78	3.62	3.52	3.45	3.40	3.37	3.34
	.050	6.61	5.79	5.41	5.19	5.05	4.95	4.88	4.82
	.025	10.01	8.43	7.76	7.39	7.15	6.98	6.85	6.76
	.010	16.26	13.27	12.06	11.39	10.97	10.67	10.46	10.29
	.001	47.18	37.12	33.20	31.09	29.75	28.83	28.16	27.65
5	.100	3.78	3.46	3.29	3.18	3.11	3.05	3.01	2.98
	.050	5.99	5.14	4.76	4.53	4.39	4.28	4.21	4.15
	.025	8.81	7.26	6.60	6.23	5.99	5.82	5.70	5.60
	.010	13.75	10.92	9.78	9.15	8.75	8.47	8.26	8.10
	.001	35.51	27.00	23.70	21.92	20.80	20.03	19.46	19.03
6	.100	3.59	3.26	3.07	2.96	2.88	2.83	2.78	2.75
	.050	5.59	4.74	4.35	4.12	3.97	3.87	3.79	3.73
	.025	8.07	6.54	5.89	5.52	5.29	5.12	4.99	4.90
	.010	12.25	9.55	8.45	7.85	7.46	7.19	6.99	6.84
	.001	29.25	21.69	18.77	17.20	16.21	15.52	15.02	14.63
7	.100	3.46	3.12	2.92	2.81	2.73	2.68	2.63	2.60
	.050	5.39	4.54	4.15	3.92	3.77	3.67	3.59	3.53
	.025	7.91	6.38	5.73	5.36	5.13	4.96	4.83	4.74
	.010	12.19	9.49	8.39	7.79	7.40	7.13	6.93	6.78
	.001	28.99	21.54	18.62	17.05	16.06	15.37	14.87	14.48
8	.100	3.37	3.03	2.83	2.72	2.64	2.59	2.54	2.51
	.050	5.30	4.45	4.06	3.83	3.68	3.58	3.50	3.44
	.025	7.83	6.30	5.65	5.28	5.05	4.88	4.75	4.66
	.010	12.16	9.46	8.36	7.76	7.37	7.10	6.90	6.75
	.001	28.96	21.51	18.59	17.02	16.03	15.34	14.84	14.45
9	.100	3.30	2.96	2.76	2.65	2.57	2.52	2.47	2.44
	.050	5.23	4.38	3.99	3.76	3.61	3.51	3.43	3.37
	.025	7.76	6.23	5.58	5.21	4.98	4.81	4.68	4.59
	.010	12.13	9.43	8.33	7.73	7.34	7.07	6.87	6.72
	.001	28.93	21.48	18.56	17.00	16.01	15.32	14.82	14.43
10	.100	3.24	2.90	2.70	2.59	2.51	2.46	2.41	2.38
	.050	5.17	4.32	3.93	3.70	3.55	3.45	3.37	3.31
	.025	7.70	6.17	5.52	5.15	4.92	4.75	4.62	4.53
	.010	12.10	9.40	8.30	7.70	7.31	7.04	6.84	6.69
	.001	28.90	21.45	18.53	16.97	15.98	15.29	14.79	14.40

Ćwiczenie: Zweryfikuj tabele ANOVy i P-wartość testu F.

Wniosek:

Rozważane grupy uczniów nie różnią się istotnie przed przeprowadzeniem kursów ($P=0.33$).

Ewentualne późniejsze różnice będą wskazywały na różnice w efektywności metod nauczania: Basal, DRTA i Strat.

Ważne: odchylenia standardowe grup są podobne.

Współczynnik determinacji:

$$R^2 = \text{SSG} / \text{SST}$$

- Pokazuje jaka część całkowitej zmienności wynika z różnic pomiędzy grupami.
- Tutaj tylko 3%.
- Na wydruku SAS-a: R-square
- Pełni podobną rolę jak r^2 w regresji liniowej.

Pytanie naukowe:

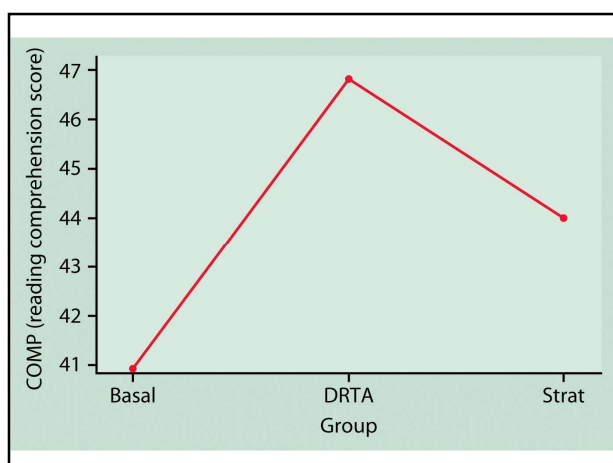
Po kilku miesiącach nauki metodami Basal, DRTA i Strat, został przeprowadzony kolejny egzamin czytania ze zrozumieniem. Wynik egzaminu będzie oznaczany symbolem COMP.

Czy te metody różnią się efektywnością?

Sformułuj hipotezy:...

Statystyki dla każdej z grup „po”:

Analysis Variable : COMP			
----- GROUP=Basal -----			
N	Mean	Std Dev	
22	41.0454545	5.6355781	
----- GROUP=DRTA -----			
N	Mean	Std Dev	
22	46.7272727	7.3884196	
----- GROUP=Strat -----			
N	Mean	Std Dev	
22	44.2727273	5.7667505	



Wyniki z SASa dla zmiennej COMP:

General Linear Models Procedure					
Dependent Variable: COMP					
Source	DF	Sum of Squares	Mean Square	F Value	Pr > F
Model	2	357.30303	178.65152	4.48	0.0152
Error	63	2511.68182	39.86797		
Corrected Total	65	2868.98485			
		R-Square	C.V.	Root MSE	COMP Mean
		0.124540	14.34531	6.3141	44.015
Contrast	DF	Contrast SS	Mean Square	F Value	Pr > F
B vs D and S	1	291.03030	291.03030	7.30	0.0088
D vs S	1	66.27273	66.27273	1.65	0.2020

Wniosek:

Wyjaśnienia:

- DRTA oraz Strat to innowacyjne metody oparte na podobnej filozofii dydaktyki.
- Basal jest metodą standardową.

Uwagi:

- Dla dwóch prób o tym samym rozmiarze n , mamy $F_s = \frac{n(\bar{x}_1 - \bar{x})^2 + n(\bar{x}_2 - \bar{x})^2}{[(n-1)s_1^2 + (n-1)s_2^2] : (2n-2)} \dots$
- Wartości krytyczne spełniają $F(1, df) = (t_{df})^2$.
- Zatem test Fishera F dla ANOVy rzeczywiście jest uogólnieniem testu Studenta t dla dwóch niezależnych prób.

Porównania parami:

- Jeżeli H_0 jest odrzucona, to można zbadać, które średnie faktycznie się różnią.

$$t_{ij} = \frac{\bar{x}_i - \bar{x}_j}{s_p \sqrt{\frac{1}{n_i} + \frac{1}{n_j}}}, \text{ if } |t_{ij}| > t^{**}, \text{ declare means } \mu_i, \mu_j \text{ different}$$

- Stosujemy np. poprawkę Bonferroniego dla k porównań i danego $\alpha=0.05$, $t^{**}=t^*(DFE, \alpha/k)$.
- Tu: $k = \binom{I}{2}$, jeżeli porównywać wszystkie pary.

Jednoczesne przedziały ufności dla różnic średnich:

$$CI \text{ for } \mu_i - \mu_j \text{ is } \bar{x}_i - \bar{x}_j \pm t^{**} s_p \sqrt{\frac{1}{n_i} + \frac{1}{n_j}}$$

General Linear Models Procedure

Bonferroni (Dunn) T tests for variable: COMP

NOTE: This test controls the type I experimentwise error rate, but generally has a higher type II error rate than REGWQ.

Alpha= 0.05 df= 63 MSE= 39.86797

Critical Value of T= 2.46

Minimum Significant Difference= 4.6825

Means with the same letter are not significantly different.

Bon Grouping	Mean	N	GROUP
A	46.727	22	DRTA
A			
B	44.273	22	Strat
B			
B	41.045	22	Basal

Bonferroni (Dunn) T tests for variable: COMP

NOTE: This test controls the type I experimentwise error rate but generally has a higher type II error rate than Tukey's for all pairwise comparisons.

Alpha= 0.05 Confidence= 0.95 df= 63 MSE= 39.86797

Critical Value of T= 2.45958

Minimum Significant Difference= 4.6825

Comparisons significant at the 0.05 level are indicated by '***'.

GROUP Comparison	Simultaneous Lower Confidence Limit	Difference Between Means	Simultaneous Upper Confidence Limit	
DRTA - Strat	-2.228	2.455	7.137	
DRTA - Basal	0.999	5.682	10.364	***
Strat - DRTA	-7.137	-2.455	2.228	
Strat - Basal	-1.455	3.227	7.910	
Basal - DRTA	-10.364	-5.682	-0.999	***
Basal - Strat	-7.910	-3.227	1.455	

Excel's output (for a different problem):

Excel

	A	B	C	D	E	F	G
1	Anova Single Factor						
2							
3	SUMMARY						
4	Groups	Count	Sum	Average	Variance		
5	Column 1	55	248	4.509091	0.476768		
6	Column 2	36	162	5.055556	0.68254		
7	Column 3	36	195	5.416667	0.764286		
8							
9							
10	ANOVA						
11	Source of Variation	SS	df	MS	F	P-value	F crit
12	Between Groups	18.82625	2	9.414127	15.28261	1.17E-06	3.069269
13	Within Groups	76.38434	124	0.616003			
14							
15	Total	95.2126	126				

One-Way Analysis of Variance

Level	N	Mean	StDev	
Undermin	55	4.5091	0.6905	(-----+-----+-----+-----+)
Affirm	36	5.0556	0.8262	(-----+-----+-----+-----+)
Control	36	5.4167	0.8742	(-----+-----+-----+-----+)
Pooled StDev =				4.40 4.80 5.20 5.60

One-way ANOVA

```
F=15.28260579
p=1.1673896E-6
Factor
  df=2
  SS=18.828255
  MS=9.4141275
Error
  df=124
  SS=76.3843434
  MS=.61600277
Sxp=.784858439
```