Chapter 9

Single–Factor Studies: Comparing Means and Determining Sample Sizes

We look a tests and confidence intervals of a multiple combination or contrast of means. From a "big picture" point of view, we continue to look at the details of the statistical inference of (large n) multiple–sample mean problems.

		mean μ	variance σ^2	proportion π
	one	large n, 3.7, 3.8, 3.9, 3.10, 4.6 small n, 4.3, 4.6	4.4	6.2
sample	two	large n, 3.11 small n, 4.3	4.4	6.3
	multiple	**chapters** 7, 8, **9**	not done	6.2, 6.3

9.1 Introduction

A *linear combination* of sample means

$$\hat{\theta} = c_1\overline{Y}_1 + \cdots + c_t\overline{Y}_t$$

is called a *contrast* of sample means if $c_1 + \cdots + c_t = 0$. Two linear contrasts of means (with equal sample sizes)

$$\hat{\theta}_1 = c_1\overline{Y}_1 + \cdots + c_t\overline{Y}_t, \quad \hat{\theta}_2 = d_1\overline{Y}_1 + \cdots + d_t\overline{Y}_t$$

are *mutually orthogonal* if

$$c_1d_1 + \cdots + c_td_t = 0$$

Exercise 9.1 (Linear Combinations, Contrasts and Orthogonal Contrasts)

1. *Examples of Linear Combinations: Drugs.* Fifteen different patients are subjected to three drugs.

drug 1	5.90	5.92	5.91	5.89	5.88	$Y_{1\cdot} \approx 5.90$
drug 2	5.51	5.50	5.50	5.49	5.50	$\bar{Y}_{2\cdot} \approx 5.50$
drug 3	5.01	5.00	4.99	4.98	5.02	$\bar{Y}_{3\cdot} \approx 5.00$

   (a) The difference between two factor level means, $\hat{\theta} = \overline{Y}_1 - \overline{Y}_2$, or,

$$\begin{aligned} \hat{\theta} &= c_1\bar{Y}_{1\cdot} + c_2\bar{Y}_{2\cdot} + c_3\bar{Y}_{3\cdot} \\ &= (1)\bar{Y}_{1\cdot} + (-1)\bar{Y}_{2\cdot} + 0\bar{Y}_{3\cdot} \\ &= (1)5.90 + (-1)5.50 + 0(5.00) \\ &= 0.40 \end{aligned}$$

   is a special case of a linear combination where
   $(c_1, c_2, c_3) =$ (choose one) **(0, 1, −1)** / **(1, −1, 0)** / **(1, 1, 1)**
   where because $c_1 + c_2 + c_3 = 1 - 1 + 0 = 0$, (circle one) **is** / **is not** also a contrast.

   (b) The single factor level mean, $\hat{\theta} = \overline{Y}_1$, or,

$$\begin{aligned} \hat{\theta} &= c_1\bar{Y}_{1\cdot} + c_2\bar{Y}_{2\cdot} + c_3\bar{Y}_{3\cdot} \\ &= (1)\bar{Y}_{1\cdot} + (0)\bar{Y}_{2\cdot} + 0\bar{Y}_{3\cdot} \\ &= (1)5.90 + (0)5.50 + 0(5.00) \\ &= 5.90 \end{aligned}$$

   is a special case of a linear combination where
   $(c_1, c_2, c_3) =$ (choose one) **(0, 1, −1)** / **(1, −1, 0)** / **(1, 0, 0)** where because $c_1 + c_2 + c_3 = 1 + 0 + 0 \neq 0$, (circle one) **is** / **is not** also a contrast.

   (c) If $(c_1, c_2, c_3) = (-1, 0, 1)$, then we are dealing with an example of a (choose none, one or more)

      i. single factor mean
      ii. difference between two factor level means
      iii. contrast among factor level means
      iv. linear combination among factor level means

   where $\hat{\theta} = -5.90 + 5.00 =$ (choose one) **−0.90** / **0** / **0.90**

   (d) If $(c_1, c_2, c_3) = (-1, -1, 1)$, then we are dealing with an example of a (choose none, one or more)

      i. single factor mean

ii. difference between two factor level means
iii. contrast among factor level means
iv. linear combination among factor level means

where $\hat{\theta} = -5.90 - 5.50 + 5.00 =$ (choose one) $\mathbf{-7.90}$ / $\mathbf{-6.40}$ / $\mathbf{-5.40}$

(e) If $(c_1, c_2, c_3) = \left(-\frac{1}{3}, -\frac{1}{3}, \frac{2}{3}\right)$, then we are dealing with an example of a (choose none, one or more)

i. single factor mean
ii. difference between two factor level means
iii. contrast among factor level means
iv. linear combination among factor level means

where $\hat{\theta} = -\frac{1}{3}(5.90) - \frac{1}{3}(5.50) + \frac{2}{3}(5.00) =$ (choose one) $-\mathbf{\frac{1.4}{3}}$ / $-\mathbf{\frac{0.4}{3}}$ / $\mathbf{\frac{1.4}{3}}$

(f) If $(c_1, c_2, c_3) = (-1, 0, 1)$ and $(d_1, d_2, d_3) = (-1, 2, -1)$ and so

$$c_1d_1 + c_2d_2 + c_3d_3 = (-1)(-1) + (0)(2) + (1)(-1) = 0$$

then these two contrasts (choose one) **are** / **are not** orthogonal.

(g) If $(c_1, c_2, c_3) = (-1, 0, 1)$ and $(d_1, d_2, d_3) = (1, -1, 0)$ and so

$$c_1d_1 + c_2d_2 + c_3d_3 = (-1)(1) + (0)(-1) + (1)(0) = -1$$

then these two contrasts (choose one) **are** / **are not** orthogonal.

2. *More Examples of Linear Combinations: Fiber Strength.* The tensile strength of fibers with varying amounts of cotton content is investigated.

5% cotton	7	7	15	11	9	10	$\bar{Y}_{1\cdot} \approx 9.8$
10% cotton	12	17	12	18	18	16	$\bar{Y}_{2\cdot} \approx 15.5$
15% cotton	14	18	18	19	19	17	$\bar{Y}_{3\cdot} \approx 17.5$
20% cotton	19	25	22	19	23	24	$\bar{Y}_{4\cdot} \approx 22$
25% cotton	7	10	11	15	11	14	$\bar{Y}_{5\cdot} \approx 11.3$

(a) The difference between two factor level means,

$$\begin{aligned} \hat{\theta} &= c_1\bar{Y}_{1\cdot} + c_2\bar{Y}_{2\cdot} + c_3\bar{Y}_{3\cdot} + c_4\bar{Y}_{4\cdot} + c_5\bar{Y}_{5\cdot} \\ &= (0)\bar{Y}_{1\cdot} + (0)\bar{Y}_{2\cdot} + (0)\bar{Y}_{3\cdot} + (-1)\bar{Y}_{4\cdot} + (1)\bar{Y}_{5\cdot} \end{aligned}$$

is a special case of a linear combination where
$(c_1, c_2, c_3, c_4, c_5) =$ (choose one)
$\mathbf{(0, 1, -1, 1, 1)}$ / $\mathbf{(1, -1, 0, 0, 0)}$ / $\mathbf{(0, 0, 0, -1, 1)}$
where, because $c_1 + c_2 + c_3 + c_4 + c_5 = 0 + 0 + 0 - 1 + 1 = 0$, (circle one)
is / **is not** also a contrast.

(b) If $(c_1, c_2, c_3, c_4, c_5) = \left(-\frac{1}{4}, -\frac{1}{4}, -\frac{1}{4}, -\frac{1}{4}, 1\right)$, then we are dealing with an example of a (choose none, one or more)

i. single factor mean
ii. difference between two factor level means
iii. contrast among factor level means
iv. linear combination among factor level means

where $\hat{\theta}$ = (choose one) **-3.9 / -4.9 / -5.9**

(c) If $(c_1, c_2, c_3, c_4, c_5) = (-1, 0, 1, 0, 0)$ and $(d_1, d_2, d_3, d_4, d_5) = (-1, 2, -1, 0, 0)$ and so

$$c_1d_1+c_2d_2+c_3d_3+c_4d_4+c_5c_5 = (-1)(-1)+(0)(2)+(1)(-1)+(0)(0)+(0)(0) = 0$$

then these two contrasts (choose one) **are** / **are not** orthogonal.

9.2 Comparisons Specified Before Performing the Experiment

We look a tests and confidence intervals of the contrast of means using two methods. One method uses the t distribution and allows one–sided and two–sided tests. The other method uses the F distribution which only allows only two–sided tests, but of multiple orthogonal contrasts.

Exercise 9.2 (Tests and Confidence Intervals For Contrasts)

1. *Test of Contrast, Using t Distribution.* Fifteen different patients are subjected to three drugs.

drug 1	5.90	5.92	5.91	5.89	5.88	$\bar{Y}_{1\cdot} \approx 5.90$
drug 2	5.51	5.50	5.50	5.49	5.50	$\bar{Y}_{2\cdot} \approx 5.50$
drug 3	5.01	5.00	4.99	4.98	5.02	$\bar{Y}_{3\cdot} \approx 5.00$

Test if the mean patient response to drug 1 is larger than the average of the mean patient response to drug 2 and drug 3 at $\alpha = 0.05$.

(a) *Statement.* Let $\theta = \mu_1 - \frac{1}{2}(\mu_2 + \mu_3)$; in other words,

$$\begin{aligned} \theta &= c_0 + c_1\mu_1 + c_2\mu_2 + c_3\mu_3 \\ &= 0 + (1)\mu_1 - \frac{1}{2}\mu_2 - \frac{1}{2}\mu_3 \end{aligned}$$

where $c_0 = 0$, $c_1 = 1$, $c_2 = -\frac{1}{2}$ and c_3 = (circle one) **-1** / **$-\frac{1}{2}$** / **0**. The statement of the test is (check none, one or more):

i. $H_0 : \theta = 0$ versus $H_1 : \theta < 0$.
ii. $H_0 : \theta = 0$ versus $H_1 : \theta > 0$.
iii. $H_0 : \theta = 0$ versus $H_1 : \theta \neq 0$.

(b) *Test.* From the ANOVA table, $MS[E] = 0.00018$ and so

$$\begin{aligned}\hat{\theta} &= c_0 + c_1\bar{y}_1 + c_2\bar{y}_2 + c_3\bar{y}_3 \\ &= 0 + (1)\bar{y}_1 - \frac{1}{2}\bar{y}_2 - \frac{1}{2}\bar{y}_3 \\ &= 0 + (1)5.90 - \frac{1}{2}(5.50) - \frac{1}{2}(5.00) =\end{aligned}$$

(circle one) **0.00** / **0.50** / **0.65**
and

$$\begin{aligned}\hat{\sigma}_{\hat{\theta}} &= \sqrt{\left\{\frac{c_1^2}{n_1} + \frac{c_2^2}{n_2} + \frac{c_3^2}{n_3}\right\} MS[E]} \\ &= \sqrt{\left\{\frac{(1)^2}{5} + \frac{(-1/2)^2}{5} + \frac{(-1/2)^2}{5}\right\} 0.00018}\end{aligned}$$

which is (circle one) **0.0042** / **0.00735** / **0.8485**
and so the test statistic is

$$t = \frac{\hat{\theta} - \theta_0}{\hat{\sigma}_{\hat{\theta}}} = \frac{0.65 - 0}{0.00735}$$

which is (circle one) **58.93** / **63.23** / **88.44**.
Since the test statistic is $t = 88.44$, the p–value, with $t - 1 = 3 - 1 = 2$ degrees of freedom, is given by

$$\text{p–value} = P(t \geq 88.44)$$

which equals (circle one) **0.00** / **0.35** / **0.43**.
The level of significance is 0.05.

(c) *Conclusion.* Since the p–value, 0.00, is smaller than the level of significance, 0.05, we (circle one) **accept** / **reject** the null hypothesis that mean patient response to drug 1 is equal to the average of the mean patient response to drug 2 and drug 3.

2. *Contrast: 95% Confidence Interval, Using t Distribution.* Fifteen different patients are subjected to three drugs.

drug 1	5.90	5.92	5.91	5.89	5.88	$\bar{Y}_{1\cdot} \approx 5.90$
drug 2	5.51	5.50	5.50	5.49	5.50	$\bar{Y}_{2\cdot} \approx 5.50$
drug 3	5.01	5.00	4.99	4.98	5.02	$\bar{Y}_{3\cdot} \approx 5.00$

A (two–sided) 95% CI for $\theta = \mu_1 - \frac{1}{2}(\mu_2 + \mu_3)$ is given by

$$\hat{\theta} \pm t(N-t, \alpha/2)\hat{\sigma}_{\hat{\theta}} = 0.65 \pm t(15-3, 0.05/2)(0.00735) =$$

(circle one) **(0.647, 0.659)** / **(0.657, 0.669)** / **(0.634, 0.666)**.
(To determine t(12, 0.025), type PRGM INVT 12 ENTER 0.975 ENTER.)

3. *Significance Test of Contrast $\theta = -\frac{1}{4}(\mu_1 + \mu_2 + \mu_3 + \mu_4) + \mu_5$, Using F Distribution.* The tensile strength of fibers with varying amounts of cotton content is investigated.

5% cotton	7	7	15	11	9	10	$\bar{y}_1 \approx 9.8$
10% cotton	12	17	12	18	18	16	$\bar{y}_2 \approx 15.5$
15% cotton	14	18	18	19	19	17	$\bar{y}_3 \approx 17.5$
20% cotton	19	25	22	19	23	24	$\bar{y}_4 \approx 22$
25% cotton	7	10	11	15	11	14	$\bar{y}_5 \approx 11.3$

Test if the mean tensile strength of the fiber with 25% cotton is different than the average of the tensile strengths of fibers with 5%, 10%, 15% and 20% cotton at $\alpha = 0.05$.

(a) *Statement.* Let $\theta = -\frac{1}{4}(\mu_1 + \mu_2 + \mu_3 + \mu_4) + \mu_5$; in other words,

$$\begin{aligned} \theta &= c_1\mu_1 + c_2\mu_2 + c_3\mu_3 + c_4\mu_4 + c_5\mu_5 \\ &= -\frac{1}{4}\mu_1 - \frac{1}{4}\mu_2 - \frac{1}{4}\mu_3 - \frac{1}{4}\mu_4 + \mu_5 \end{aligned}$$

The statement of the test is (check none, one or more):

i. $H_0 : \theta = 0$ versus $H_1 : \theta < 0$.
ii. $H_0 : \theta = 0$ versus $H_1 : \theta > 0$.
iii. $H_0 : \theta = 0$ versus $H_1 : \theta \neq 0$.

(b) *Test.* Since

$$\begin{aligned} \hat{\theta} &= c_1\bar{y}_1 + c_2\bar{y}_2 + c_3\bar{y}_3 + c_4\bar{y}_4 + c_5\bar{y}_5 \\ &= -\frac{1}{4}(9.8) - \frac{1}{4}(15.5) - \frac{1}{4}(17.5) - \frac{1}{4}(22) + (1)(11.3) = \end{aligned}$$

(circle one) **−5.6** / **−4.9** / **−3.3**
and

$$\begin{aligned} MS[\hat{\theta}] &= SS[\hat{\theta}] \\ &= \frac{\hat{\theta}^2}{\left(\frac{c_1^2}{n_1} + \frac{c_2^2}{n_2} + \frac{c_3^2}{n_3} + \frac{c_4^2}{n_4} + \frac{c_5^2}{n_5}\right)} \\ &= \frac{(-4.9)^2}{\left(\frac{(-0.25)^2}{6} + \frac{(-0.25)^2}{6} + \frac{(-0.25)^2}{6} + \frac{(-0.25)^2}{6} + \frac{(1)^2}{6}\right)} \end{aligned}$$

which equals (circle one) **110.25** / **115.25** / **120.25**
and since the ANOVA table is given by,

Source	Degrees of Freedom	Sum Of Squares	Mean Squares	F
Treatment (% Cotton)	4	572.2	143.05	20.42
Error	25	175.17	7.007	
Total	29	747.37		

where $MS[E]$ = (circle one) **7.007** / **143.05** / **175.17**.
(Type data into L_1, L_2, L_3, L_4 and L_5 and then type STAT TESTS F:ANOVA(L_1,L_2,L_3,L_4,L_5) ENTER.)
then the test statistic is

$$F = \frac{MS[\hat{\theta}]}{MS[E]} = \frac{115.25}{7.007}$$

which is (circle one) **15.45** / **16.45** / **17.45**.
Since the test statistic is $F = 16.45$, the p–value, with $(1, N-t) = (1, 30-5) = (1, 25)$ degrees of freedom, is given by

$$\text{p–value} = P(F \geq 16.45)$$

which equals (circle one) **0.00043** / **0.0043** / **0.043**.
(Use 2nd DISTR 8:Fcdf(16.45,E99,1,25).)
The level of significance is 0.05.

(c) *Conclusion.* Since the p–value, 0.00043, is smaller than the level of significance, 0.05, we (circle one) **accept** / **reject** the null hypothesis that the mean tensile strength of the fiber with 25% cotton is equal to the average of the tensile strengths of fibers with 5%, 10%, 15% and 20% cotton.

4. *Another Significance Test of Contrast $\theta = \mu_2 - \mu_3$, Using F Distribution.* The tensile strength of fibers with varying amounts of cotton content is investigated.

5% cotton	7	7	15	11	9	10	$\bar{y}_1 \approx 9.8$
10% cotton	12	17	12	18	18	16	$\bar{y}_2 \approx 15.5$
15% cotton	14	18	18	19	19	17	$\bar{y}_3 \approx 17.5$
20% cotton	19	25	22	19	23	24	$\bar{y}_4 \approx 22$
25% cotton	7	10	11	15	11	14	$\bar{y}_5 \approx 11.3$

Test if the mean tensile strength of the fiber with 10% cotton is different than the mean tensile strength of the fiber with 15% cotton at $\alpha = 0.05$.

(a) *Statement.* Let $\theta = \mu_2 - \mu_3$; in other words,

$$\begin{aligned} \theta &= c_1\mu_1 + c_2\mu_2 + c_3\mu_3 + c_4\mu_4 + c_5\mu_5 \\ &= (0)\mu_1 + (1)\mu_2 - (1)\mu_3 + (0)\mu_4 + (0)\mu_5 \end{aligned}$$

The statement of the test is (check none, one or more):

i. $H_0 : \theta = 0$ versus $H_1 : \theta < 0$.
ii. $H_0 : \theta = 0$ versus $H_1 : \theta > 0$.
iii. $H_0 : \theta = 0$ versus $H_1 : \theta \neq 0$.

(b) *Test.* Since

$$\begin{aligned} \hat{\theta} &= c_1\bar{y}_1 + c_2\bar{y}_2 + c_3\bar{y}_3 + c_4\bar{y}_4 + c_5\bar{y}_5 \\ &= (0)(9.8) + (1)(15.5) + (-1)(17.5) + (0)(22) + (0)(11.3) = \end{aligned}$$

(circle one) **−2.0** / **−1.5** / **−0.5**
and

$$\begin{aligned} MS[\hat{\theta}] &= SS[\hat{\theta}] \\ &= \frac{\hat{\theta}^2}{\left(\frac{c_1^2}{n_1} + \frac{c_2^2}{n_2} + \frac{c_3^2}{n_3} + \frac{c_4^2}{n_4} + \frac{c_5^2}{n_5}\right)} \\ &= \frac{(-2.0)^2}{\left(\frac{(0)^2}{6} + \frac{(1)^2}{6} + \frac{(-1)^2}{6} + \frac{(0)^2}{6} + \frac{(0)^2}{6}\right)} \end{aligned}$$

which equals (circle one) **12** / **14** / **16**
and since, from the ANOVA table, $MS[E] = 7.007$, the test statistic is

$$F = \frac{MS[\hat{\theta}]}{MS[E]} = \frac{12}{7.007}$$

which is (circle one) **1.713** / **1.813** / **1.913**.
Since the test statistic is $F = 1.713$, the p–value, with $(1, N-t) = (1, 30-5) = (1, 25)$ degrees of freedom, is given by

$$\text{p–value} = P(F \geq 1.713)$$

which equals (circle one) **0.203** / **0.303** / **0.4.03**.
(Use 2nd DISTR Fcdf(1.713,E99,1,25).)
The level of significance is 0.05.

(c) *Conclusion.* Since the p–value, 0.203, is larger than the level of significance, 0.05, we (circle one) **accept** / **reject** the null hypothesis that the mean tensile strength of the fiber with 10% cotton is equal to the mean tensile strength of the fiber with 15% cotton.

5. *Test Multiple Orthogonal Contrasts, Using F Distribution.* The tensile strength of fibers with varying amounts of cotton content is investigated.

5% cotton	7	7	15	11	9	10	$\bar{y}_1 \approx 9.83$
10% cotton	12	17	12	18	18	16	$\bar{y}_2 \approx 15.5$
15% cotton	14	18	18	19	19	17	$\bar{y}_3 \approx 17.5$
20% cotton	19	25	22	19	23	24	$\bar{y}_4 \approx 22$
25% cotton	7	10	11	15	11	14	$\bar{y}_5 \approx 11.33$

(a) Four orthogonal contrasts and their associated sum of squares are given below.

$(c_1, c_2, c_3, c_4, c_5)$	$\hat{\theta}$	$SS[\hat{\theta}]$
$(-\frac{1}{4}, -\frac{1}{4}, -\frac{1}{4}, -\frac{1}{4}, 1)$	$\hat{\theta}_1 = -4.9$	$SS[\hat{\theta}_1] = 115.248$
$(0, 1, -1, 0, 0)$	$\hat{\theta}_2 = -2.0$	$SS[\hat{\theta}_2] = 12$
$(-1, 0, 0, 1, 0)$	$\hat{\theta}_3 = 12.2$	$SS[\hat{\theta}_3] = 446.52$
$(-\frac{1}{2}, \frac{1}{2}, \frac{1}{2}, -\frac{1}{2}, 0)$	$\hat{\theta}_4 = 0.6$	$SS[\hat{\theta}_4] = 2.16$

$SS[\hat{\theta}_1] + SS[\hat{\theta}_2] + SS[\hat{\theta}_3] + SS[\hat{\theta}_4] =$ (circle one) **572.2** / **575.9** / **583.4**.

(b) The ANOVA table is given by,

Source	Degrees of Freedom	Sum Of Squares	Mean Squares	F
Treatment (% Cotton)	4	572.2	143.05	20.42
Error	25	175.17	7.007	
Total	29	747.37		

where $SS[T] =$ (circle one) **572.2** / **575.9** / **583.4**.

(c) The total of the sum of squares for the *mutually orthogonal* contrasts, 575.9, *should* add up to $SS[T] = 572.2$, but, because of round–off error, they do not. In other words, the ANOVA table with the four contrasts is given by

Source	Degrees of Freedom	Sum Of Squares	Mean Squares	F
Treatment (% Cotton)	4	572.2	143.05	20.42
$(-\frac{1}{4}, -\frac{1}{4}, -\frac{1}{4}, -\frac{1}{4}, 1)\ \hat{\theta}_1$	1	115.3	115.3	16.5
$(0, 1, -1, 0, 0)\ \hat{\theta}_2$	1	12	12	1.7
$(-1, 0, 0, 1, 0)\ \hat{\theta}_3$	1	446.5	446.5	63.7
$(-\frac{1}{2}, \frac{1}{2}, \frac{1}{2}, -\frac{1}{2}, 0)\ \hat{\theta}_4$	1	2.2	2.2	0.32
Error	25	175.17	7.007	
Total	29	747.37		

and so, the p–value for $\hat{\theta}_3$, with $(1, N - t) = (1, 30 - 5) = (1, 25)$ degrees of freedom, is given by

$$\text{p–value} = P(F \geq 63.7)$$

which equals (circle one) **0.00** / **0.303** / **0.4.03**.
(Use 2nd DISTR 8:Fcdf(63.7,E99,1,25).)

(d) Looking at the observed F values in the ANOVA table, it appears as though (circle *two*!)

i. $\theta_1 = -\frac{1}{4}(\mu_1 + \mu_2 + \mu_3 + \mu_4) + \mu_5 \neq 0$ and $\theta_3 = -\mu_1 + \mu_4 \neq 0$
ii. $\theta_2 = \mu_2 - \mu_3 \neq 0$ and $\theta_4 = \frac{1}{2}(-\mu_1 + \mu_2 + \mu_3 - \mu_4) \neq 0$
iii. $\theta_2 = \mu_2 - \mu_3 = 0$ and $\theta_4 = \frac{1}{2}(-\mu_1 + \mu_2 + \mu_3 - \mu_4) = 0$

(e) The four mutually orthogonal contrasts are *not* unique. For example, another set of four mutually orthogonal contrasts are

$(c_1, c_2, c_3, c_4, c_5)$	$\hat{\theta}$	$SS[\hat{\theta}]$
$(1, -\frac{1}{4}, -\frac{1}{4}, -\frac{1}{4}, -\frac{1}{4})$	$\hat{\theta}_1 = -6.8$	$SS[\hat{\theta}_1] = 221.952$
$(0, 1, -1, 0, 0)$	$\hat{\theta}_2 = -2.0$	$SS[\hat{\theta}_2] = 12$
$(0, 0, 0, 1, -1)$	$\hat{\theta}_3 = 10.7$	$SS[\hat{\theta}_3] = 343.47$
$(0, \frac{1}{2}, \frac{1}{2}, -\frac{1}{2}, -\frac{1}{2})$	$\hat{\theta}_4 = -0.17$	$SS[\hat{\theta}_4] = 0.1734$

Source	Degrees of Freedom	Sum Of Squares	Mean Squares	F
Treatment (% Cotton)	4	572.2	143.05	20.42
$(1, -\frac{1}{4}, -\frac{1}{4}, -\frac{1}{4}, -\frac{1}{4})\ \hat{\theta}_1$	1	222.0	222.0	31.7
$(0, 1, -1, 0, 0)\ \hat{\theta}_2$	1	12	12	1.7
$(0, 0, 0, 1, -1)\ \hat{\theta}_3$	1	343.5	343.5	49.0
$(0, \frac{1}{2}, \frac{1}{2}, -\frac{1}{2}, -\frac{1}{2})\ \hat{\theta}_4$	1	0.2	0.2	0.03
Error	25	175.17	7.007	
Total	29	747.37		

and so, the p–value for $\hat{\theta}_1$, say, with $(1, N-t) = (1, 30-5) = (1, 25)$ degrees of freedom, is given by

$$\text{p–value} = P(F \geq 31.7)$$

which equals (circle one) **0.00** / **0.303** / **0.403**.
(Use 2nd DISTR 8:Fcdf(31.7,E99,1,25).)

(f) Looking at the observed F values in the ANOVA table, it appears as though (circle one)

i. $\theta_1 = \mu_1 - \frac{1}{4}(\mu_2 + \mu_3 + \mu_4 + \mu_5) \neq 0$ and $\theta_3 = \mu_4 - \mu_5 \neq 0$

ii. $\theta_2 = \mu_2 - \mu_3 \neq 0$ and $\theta_4 = \frac{1}{2}(\mu_2 + \mu_3 - \mu_4 - \mu_5) \neq 0$

9.3 Testing Contrasts Suggested By Data

We continue to look at how to test contrasts, this time, using three different types of *critical contrast values* (CCVs). For $\hat{\theta}_1, \ldots, \hat{\theta}_k$ contrasts,

1. *Fisher's CCV*: if $|\hat{\theta}_i| > t(N-t, \alpha/2)\hat{\sigma}_{\hat{\theta}}$ then $\hat{\theta}_i$ is significant
2. *Bonferroni's CCV*: if $|\hat{\theta}_i| > t(N-t, \alpha/(2 \times k))\hat{\sigma}_{\hat{\theta}}$ then $\hat{\theta}_i$ is significant
3. *Scheffe's CCV*: if $|\hat{\theta}_i| > \sqrt{(t-1)F(t-1, N-t, \alpha)}\hat{\sigma}_{\hat{\theta}}$ then $\hat{\theta}_i$ is significant

Generally speaking, it is "easier" to get significance using Fisher's CCV, than it is using either Bonferroni's CCV or Scheffe's CCV.

Exercise 9.3 (Critical Contrast Values: Multiple Comparison of Means) The tensile strength of fibers with varying amounts of cotton content is investigated.

5% cotton	7	7	15	11	9	10	$\bar{y}_1 \approx 9.8$
10% cotton	12	17	12	18	18	16	$\bar{y}_2 \approx 15.5$
15% cotton	14	18	18	19	19	17	$\bar{y}_3 \approx 17.5$
20% cotton	19	25	22	19	23	24	$\bar{y}_4 \approx 22$
25% cotton	7	10	11	15	11	14	$\bar{y}_5 \approx 11.3$

See Lab 11: Critical Contrast Values

1. *CCVs, For Contrast* $\theta = -\frac{1}{4}(\mu_1 + \mu_2 + \mu_3 + \mu_4) + \mu_5$. Use Fisher's CCV to decide if the mean tensile strength of the fiber with 25% cotton is different than the average of the tensile strengths of fibers with 5%, 10%, 15% and 20% cotton at $\alpha = 0.05$. Use all versions of the CCV. Assume that this contrast is one of $k = 4$ tested.

   (a) *Statement.* Let $\theta = -\frac{1}{4}(\mu_1 + \mu_2 + \mu_3 + \mu_4) + \mu_5$; in other words,

   $$\begin{aligned} \theta &= c_1\mu_1 + c_2\mu_2 + c_3\mu_3 + c_4\mu_4 + c_5\mu_5 \\ &= -\frac{1}{4}\mu_1 - \frac{1}{4}\mu_2 - \frac{1}{4}\mu_3 - \frac{1}{4}\mu_4 + \mu_5 \end{aligned}$$

   The statement of the test is (check none, one or more):

   i. $H_0 : \theta = 0$ versus $H_1 : \theta < 0$.
   ii. $H_0 : \theta = 0$ versus $H_1 : \theta > 0$.
   iii. $H_0 : \theta = 0$ versus $H_1 : \theta \neq 0$.

   (b) *Test.* Since the observed value of the contrast is

   $$\begin{aligned} \hat{\theta} &= c_1\bar{y}_1 + c_2\bar{y}_2 + c_3\bar{y}_3 + c_4\bar{y}_4 + c_5\bar{y}_5 \\ &= -\frac{1}{4}(9.8) - \frac{1}{4}(15.5) - \frac{1}{4}(17.5) - \frac{1}{4}(22) + (1)(11.3) = \end{aligned}$$

   (choose one) **−3.3** / **−4.9** / **−5.6**
   and the three CCV values are given by

contrast	$\lvert\hat{\theta}\rvert$	Fisher *CCV*	Bonferroni *CCV*	Scheffe *CCV*
$\left(-\frac{1}{4}, -\frac{1}{4}, -\frac{1}{4}, -\frac{1}{4}, 1\right)$	$\lvert -4.9\rvert$	2.49	_____	_____

   (First type (or UNGROUP) cotton tensile strength data into $L_1, \ldots, L_5$. Then PRGM CCV ENTER 5 ENTER 4 ENTER 0.05 ENTER 2nd{ -0.25, -0.25, -0.25, -0.25, 1 2nd} ENTER and reading the three CCV values.)

   (c) *Conclusion.* Since the observed value of the contrast, $|\hat{\theta}_1| = |-4.9| = 4.9$, is greater than all three CCVs, (2.49, 3.25, 4.01), we (circle one) **accept** / **reject** the null hypothesis that the mean tensile strength of the fiber with 25% cotton is equal to the average of the tensile strengths of fibers with 5%, 10%, 15% and 20% cotton according to all three CCVs.

2. *CCVs For Contrast* $\theta = \frac{1}{2}(\mu_2 + \mu_3) - \frac{1}{2}(\mu_4 + \mu_5)$. Decide if the average tensile strength of the fiber with 10% cotton and 15% cotton is different than the average of the tensile strengths of fibers with 20% and 25% cotton at $\alpha = 0.10$. Use all versions of the CCV. Assume that this contrast is one of $k = 2$ tested.

   (a) *Statement.* Let $\theta = \frac{1}{2}(\mu_2 + \mu_3) - \frac{1}{2}(\mu_4 + \mu_5)$; in other words,

$$\begin{aligned} \theta &= c_1\mu_1 + c_2\mu_2 + c_3\mu_3 + c_4\mu_4 + c_5\mu_5 \\ &= 0\mu_1 + \frac{1}{2}\mu_2 + \frac{1}{2}\mu_3 - \frac{1}{2}\mu_4 - \frac{1}{2}\mu_5 \end{aligned}$$

   The statement of the test is (check none, one or more):

   i. $H_0 : \theta = 0$ versus $H_1 : \theta < 0$.
   ii. $H_0 : \theta = 0$ versus $H_1 : \theta > 0$.
   iii. $H_0 : \theta = 0$ versus $H_1 : \theta \neq 0$.

   (b) *Test.* The observed value of the contrast is

$$\begin{aligned} \hat{\theta} &= c_1\bar{y}_1 + c_2\bar{y}_2 + c_3\bar{y}_3 + c_4\bar{y}_4 + c_5\bar{y}_5 \\ &= 0(9.8) + \frac{1}{2}(15.5) + \frac{1}{2}(17.5) - \frac{1}{2}(22) - \frac{1}{2}(11.3) = \end{aligned}$$

   (circle one) **−0.15** / **−0.30** / **−0.45**
   and the three CCV values are given by

contrast	$\lvert\hat{\theta}\rvert$	Fisher *CCV*	Bonferroni *CCV*	Scheffe *CCV*
$\left(0, \frac{1}{2}, \frac{1}{2}, -\frac{1}{2}, -\frac{1}{2}\right)$	−0.15	1.85	_____	3.19

   (c) *Conclusion.* Since the observed value of the contrast, $|\hat{\theta}_1| = |-0.15| = 0.15$, is less than all three critical contrast values (CCVs), (1.85, 2.23, 3.19), we (circle one) **accept** / **reject** the null hypothesis that average tensile strength of the fiber with 10% cotton and 15% cotton is the same as the average of the tensile strengths of fibers with 20% and 25% cotton

9.4 Multiple Comparison of Means

We continue to look at the Fisher, Bonferroni and Scheffe CCVs. We find out that Fisher has a *comparison error rate* of α, whereas Bonferroni and Scheffe both have *experimentwise error rates* of less than or equal to α. Bonferroni and Scheffe, better than Fisher, account for multiple tests of contrasts (which have been specified after observing the outcome of the experiment).

See Lab 11: Critical Contrast Values

Exercise 9.4 (Critical Contrast Values: Multiple Comparison of Means)

1. *Comparison of the Fisher, Bonferroni and Scheffe CCVs.* Consider the following table below which compares the Fisher, Bonferroni and Scheffe critical contrast values (CCVs) at $\alpha = 0.05$ for three contrasts.

	contrast	$\lvert\hat{\theta}\rvert$	Fisher CCV	Bonferroni CCV	Scheffe CCV
1	$(-\frac{1}{4}, -\frac{1}{4}, -\frac{1}{4}, -\frac{1}{4}, 1)$	3.9	2.19	3.25	4.01
2	$(0, 1, -1, 0, 0)$	2.0	3.25	4.01	5.18
3	$(-\frac{1}{2}, \frac{1}{2}, \frac{1}{2}, -\frac{1}{2}, 0)$	4.1	2.73	3.25	3.39

Circle true or false.

(a) **True** / **False** The first contrast, contrast 1, is significant with respect to both the Fisher CCV and Bonferroni CCV, but not the Scheffe CCV.

(b) **True** / **False** The second contrast, contrast 2, is significant with respect to all three CCVs.

(c) **True** / **False** The third contrast, contrast 3, is significant with respect to all three CCVs.

(d) **True** / **False** So say the second contrast, contrast 2, is significant implies $\mu_2 \neq \mu_3$.

2. *Understanding the Fisher, Bonferroni and Scheffe CCVs: Tensile Strength of Cotton.* The tensile strength of fibers with varying amounts of cotton content is investigated.

5% cotton	7	7	15	11	9	10	$\bar{y}_1 \approx 9.8$
10% cotton	12	17	12	18	18	16	$\bar{y}_2 \approx 15.5$
15% cotton	14	18	18	19	19	17	$\bar{y}_3 \approx 17.5$
20% cotton	19	25	22	19	23	24	$\bar{y}_4 \approx 22$
25% cotton	7	10	11	15	11	14	$\bar{y}_5 \approx 11.3$

(a) Complete the following table below which compares the Fisher, Bonferroni and Scheffe CCVs at $\alpha = 0.05$.

contrast	$\lvert\hat{\theta}\rvert$	Fisher CCV	Bonferroni CCV	Scheffe CCV
$(-\frac{1}{4}, -\frac{1}{4}, -\frac{1}{4}, -\frac{1}{4}, 1)$	4.9	2.49	3.25	4.01
$(0, 1, -1, 0, 0)$	2.0	3.15	4.11	5.08
$(-1, 0, 0, 1, 0)$	12.2	3.14	4.11	5.08
$(-\frac{1}{2}, \frac{1}{2}, \frac{1}{2}, -\frac{1}{2}, 0)$	0.6	2.23	_____	3.59

(b) In this example, the Scheffe CCV is (circle one) **larger than / the same as / smaller than** either the Bonferroni CCV or Fisher CCV. This means the observed contrast $\hat{\theta}$ has to be larger to achieve significance when using the Scheffe CCV as opposed to using either the Bonferroni CCV or Fisher CCV.

(c) The Fisher CCV does *not* take into account the number of contrasts, k, used in the experiment. This can be "bad" since, the larger k, (the larger the number of contrasts tested) the (circle one) **lesser / same as / greater** the chance of mistakenly rejecting the null (of falsely saying a contrast is significant). The Fisher CCV is said to have a *comparison error rate* of α.

(d) The Scheffe CCV does take into account the number of contrasts, k, used in the experiment. However, it assumes "all" possible mutually orthogonal contrasts are used and does *not* account for the case when less than all contrasts are tested. This can be "bad" because if k is small, (a small number of contrasts is tested) then there can be (circle one) **a lesser / the same as / a greater** chance of mistakenly rejecting the null (of falsely saying a contrast is significant). The Scheffe CCV is said to have an *experimentwise error rate* of less than or equal to α.

(e) The Bonferroni CCV not only takes into account the number of contrasts, k, used in the experiment, but also accounts for the case when less than all contrasts are tested. The Bonferroni CCV, however, is not always smaller than the Scheffe CCV and so, often, both are calculated and the (circle one) **larger / smaller** of the two is chosen for testing purposes. The Bonferroni CCV is said to have an *experimentwise error rate* of less than or equal to α.