

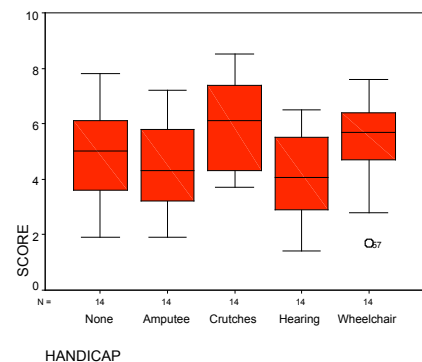
Resources: handicap.sav

The U.S. Vocational Rehabilitation Act of 1973 prohibited discrimination against people with physical disabilities. The act defined a handicapped person as any individual who had a physical or mental impairment that limits the person’s major life activities. In 1984, handicapped individuals in the labor force had an unemployment rate of 7% compared to 4.5% in the non-impaired labor force.

One study explored how physical handicaps affect people’s perception of employment qualifications (Cesare, Tannenbaum, Dalessio, “Interviewers’ Decisions Related to Applicant Handicap Type and Rater Empathy,” *Human Performance* 3(3) (1990): 157-71). The researchers prepared five videotaped job interviews using the same two male actors for each. A set script was designed to reflect an interview with an applicant of average qualifications. The tapes differed only in that the applicant appeared with a different handicap. In one, he appeared in a wheelchair; in a second, he appeared on crutches; in another, his hearing was impaired; in a fourth, he appeared to have one leg amputated; and in the final tape, he appeared to have no handicap.

Seventy undergraduate students from an American university were randomly assigned to view the tapes, fourteen to each tape. After viewing the tape, each subject rated the qualifications of the applicant on a 0-10 point scale. Here are the results of that study:

	No Handicap	Amputee	Crutches	Hearing	Wheelchair
1.90	1.90	3.70	1.40	1.70	
2.50	2.50	4.00	2.10	2.80	
3.00	2.60	4.30	2.40	3.50	
3.60	3.20	4.30	2.90	4.70	
4.10	3.60	5.10	3.40	4.80	
4.20	3.80	5.80	3.70	5.00	
4.90	4.00	6.00	3.90	5.30	
5.10	4.60	6.20	4.20	6.10	
5.40	5.30	6.30	4.30	6.10	
5.90	5.50	6.40	4.70	6.20	
6.10	5.80	7.40	5.50	6.40	
6.70	5.90	7.40	5.80	7.20	
7.40	6.10	7.50	5.90	7.40	
7.80	7.20	8.50	6.50	7.60	
Mean	4.9000	4.4286	5.9124	4.0500	5.3429
StDev	1.7936	1.5857	1.4818	1.5325	1.7483

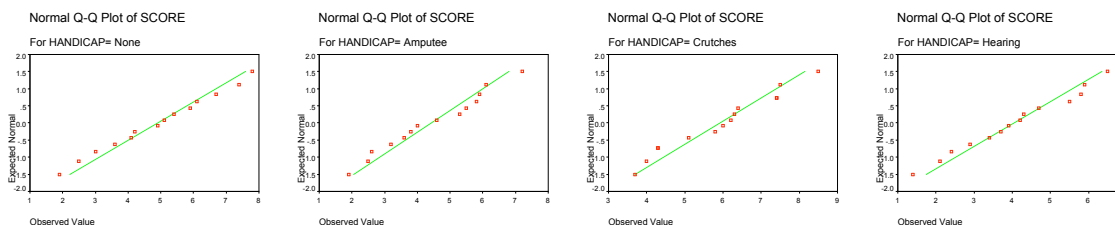


1) Conduct an appropriate analysis to determine if the treatments really did have an impact on the ratings. What assumptions must be met in order to conduct the analysis? Check these assumptions, select an appropriate significance level, create a summary table of your results (showing your calculations), and make a valid conclusion.

$H_0: \mu_N = \mu_A = \mu_C = \mu_H = \mu_W =$  All means are equal (or no treatment effect)  
 $H_A:$  Not all the population means are equal (at least one mean differs)

Type I: Incorrectly reject null hypothesis = conclude handicapped individuals get lower scores even though the differences are due to error  
 Type II: Incorrectly retain null hypothesis = conclude the ratings do not differ more than by chance even though the differences are real.  
 Neither one of these errors is too much to worry about in this simple study. We’ll stick with the standard alpha level of 0.05.

Assumptions: Independence, normality, homogeneity of variance. We use Q-Q plots or histograms to check for normality.



The normality assumption appears to be satisfied. Now we’ll check the homogeneity of variance assumption...

**Test of Homogeneity of Variances**

SCORE			
Levene Statistic	df1	df2	Sig.
.196	4	65	.940

$$F_{\max} = \frac{1.7936^2}{1.4818^2} = 1.465$$

## ANOVA

SCORE					
	Sum of Squares	df	Mean Square	F	Sig.
Between Groups	30.521	4	7.630	2.862	.030
Within Groups	173.321	65	2.666		
Total	203.843	69			

$$SS_A = 14[(4.9 - 4.92678)^2 + (4.4286 - 4.92678)^2 + \dots] = 14(2.16) = 30.24$$

$$df_A = a - 1 = 5 - 1 = 4$$

$$SS_W = 13(1.7936^2 + 1.5857^2 + \dots) = 13(13.33) = 173.319$$

$$df_W = N - a = 70 - 5 = 65$$

$$SS_T = 30.24 + 173.319 = 203.559$$

$$df_T = N - 1 = 70 - 1 = 69$$

$$\eta^2 = \frac{30.521}{203.843} = 0.15$$

**Conclusion:** We have significance, so we reject the null hypothesis. If differences were simply due to sampling error, we would expect to get results as extreme as ours 3% of the time. We conclude that the group means do differ.

- 2) Can we conclude that handicapped individuals received significantly lower ratings than non-impaired individuals? Can we conclude that the wheelchair group received significantly lower ratings than the crutches group?
- 3) If we want to compare specific treatment means, we'll need to conduct follow-up tests. These tests are often called *post hoc* tests (literally, *after the fact* tests). In this course we'll learn how to conduct 3 types of post hoc tests. The first methods we'll use is called the **Bonferroni Method**. This method allows us to compare pairs of group means (*pairwise comparisons*). Before we begin, list all of the pairwise differences we could examine in this study. How many pairwise comparisons would we have in a study with  $n$  treatments?

- 4) If we used independent samples t-tests on our 10 pairwise comparisons, our alpha level (the probability of making a Type I error) will inflate.

$$P(\text{retain null hypothesis all 10 times}) = (.95)(.95)\dots(.95)(.95) = .5987$$

$$P(\text{reject null hypothesis at least once}) = 1 - .5987 = .4013$$

Instead of having a .05 significance level, we have a .40 probability of making a Type I error.

Since we want  $\alpha=0.05$  for the entire study (all 10 pairwise comparisons), we'll have to lower alpha for each pairwise comparison. Let's arbitrarily lower alpha to 0.025. What would be the *familywise alpha* for our follow-up tests?

This time, let's divide alpha by the number of pairwise comparisons we want to make. What is the familywise error rate?

This is the key concept of *Bonferroni adjustments* – in order to have a family-wise error rate equal to  $\alpha$ , we need to divide our overall  $\alpha$ -level by the number of tests you will conduct. You then conduct independent samples t-tests of the pairwise differences

Bonferroni Method for pairwise differences: (Compare observed t-value to critical value)

$$t_{\text{observed}} = \frac{\bar{X}_1 - \bar{X}_2}{\sqrt{MS_w \left( \frac{1}{n_1} + \frac{1}{n_2} \right)}}$$

$$t_{\text{critical}} = t_{\frac{\alpha}{2k}, (N-a)} \quad \text{where} \quad k = \frac{a(a-1)}{2} \quad \binom{a}{2} = \frac{a!}{(a-2)!2!} = \frac{(a)(a-1)(a-2)!}{(a-2)!2} = \frac{(a)(a-1)}{2}$$

and a = number of groups in the overall study

- 5) Let's try the Bonferroni method on our study. Assuming we want a family-wise error rate of  $\alpha=0.05$ , what significance level should we choose for each pairwise test we will conduct? What significance level should be chosen if we want a family-wise error rate of  $\alpha=0.10$ .

6) Assuming we want a family-wise error rate of  $\alpha=0.05$ , conduct a post hoc test to compare the “No Handicapped” treatment to the “Amputee” treatment. What is the sample mean difference? Is this difference significant? State your conclusion.

7) Compare the “Crutches” group to the “Hearing” group. Is this difference significant? State your conclusion.

8) Instead of repeating the same steps over and over, we can speed up the process of comparing pairs of means by creating the following table:

Comparison	Difference	Standard Error	t-critical	Confidence Interval	Significant?
$\bar{X}_1$ vs $\bar{X}_2$	$\bar{X}_1 - \bar{X}_2$	$\sqrt{MS_W \left( \frac{1}{n_1} + \frac{1}{n_2} \right)}$	$t_{critical} = t_{\frac{\alpha}{2k}, (N-a)}$	$Difference \pm t(SE)$	Yes, if zero isn't in the CI
None vs. Amputee	0.4174	0.61714	2.91	(-1.378 , 2.267)	No
None vs. Crutches	-1.0214	0.61714	2.91	( -2.82 , 0.774 )	No
None vs. Hearing	0.85	0.61714	2.91	( -0.95 , 2.65 )	No
None vs. Wheelchair	-0.4429	0.61714	2.91	( -2.24 , 1.35 )	No

9) If you conduct the ANOVA and Bonferroni follow-ups on SPSS, you will obtain the following output. Make sure you can interpret each piece of the output.

### ANOVA

SCORE					
	Sum of Squares	df	Mean Square	F	Sig.
Between Groups	30.521	4	7.630	2.862	.030
Within Groups	173.321	65	2.666		
Total	203.843	69			

### Multiple Comparisons

Dependent Variable: SCORE

Bonferroni

(I) HANDICAP	(J) HANDICAP	Mean Difference (I-J)	Std. Error	Sig.	95% Confidence Interval	
					Lower Bound	Upper Bound
None	Amputee	.4714	.61719	1.000	-1.3221	2.2650
	Crutches	-1.0214	.61719	1.000	-2.8150	.7721
	Hearing	.8500	.61719	1.000	-.9436	2.6436
	Wheelchair	-.4429	.61719	1.000	-2.2364	1.3507
Amputee	None	-.4714	.61719	1.000	-2.2650	1.3221
	Crutches	-1.4929	.61719	.184	-3.2864	.3007
	Hearing	.3786	.61719	1.000	-1.4150	2.1721
	Wheelchair	-.9143	.61719	1.000	-2.7079	.8793
Crutches	None	1.0214	.61719	1.000	-.7721	2.8150
	Amputee	1.4929	.61719	.184	-.3007	3.2864
	Hearing	1.8714*	.61719	.035	.0779	3.6650
	Wheelchair	.5786	.61719	1.000	-1.2150	2.3721
Hearing	None	-.8500	.61719	1.000	-2.6436	.9436
	Amputee	-.3786	.61719	1.000	-2.1721	1.4150
	Crutches	-1.8714*	.61719	.035	-3.6650	-.0779
	Wheelchair	-1.2929	.61719	.401	-3.0864	.5007
Wheelchair	None	.4429	.61719	1.000	-1.3507	2.2364
	Amputee	.9143	.61719	1.000	-.8793	2.7079
	Crutches	-.5786	.61719	1.000	-2.3721	1.2150
	Hearing	1.2929	.61719	.401	-.5007	3.0864

\*. The mean difference is significant at the .05 level.

10) Write out your conclusions from this study. Does a handicapped individual receive lower job qualification ratings than a non-handicapped individual?

11) Recall our drug-binding study (from the previous activity). We conducted an ANOVA to see if drugs had different binding percentages. Here is a summary of our analysis

						Sample Mean	Sample Size
(P) Penicillin G	29.6	24.3	28.5	32.0		28.600	4
(T) Tetracycline	27.3	32.6	30.8	34.8	32.0	31.500	5
(S) Streptomycin	05.8	06.2	11.0	08.3		7.825	4
(E) Erythromycin	21.6	17.4	18.3			19.100	3
(C) Chloramphenicol	29.2	32.8	25.0	24.2		27.800	4
					Total: M =	23.585	20

ANOVA

BINDING

	Sum of Squares	df	Mean Square	F	Sig.
Between Groups	1538.758	4	384.690	42.389	.000
Within Groups	136.127	15	9.075		
Total	1674.886	19			

We concluded that not all of the means are equal.

12) Conduct pairwise comparisons using the *Bonferroni method*. Highlight significant comparisons.

Comparison	Difference	Standard Error	t-critical	Confidence Interval	Significant?
P vs. T	-2.900	$\sqrt{9.075\left(\frac{1}{4} + \frac{1}{5}\right)} = 2.0208$	alpha = .0025 df = 20-5 = 15 = 3.3	-9.57, 3.77	No
P vs. S	20.775	$\sqrt{9.075\left(\frac{1}{4} + \frac{1}{4}\right)} = 2.13$	3.3	13.746, 27.804	Yes
P vs. E	9.5	$\sqrt{9.075\left(\frac{1}{4} + \frac{1}{3}\right)} = 2.301$	3.3	1.91, 17.09	Yes
P vs. C	0.8	$\sqrt{9.075\left(\frac{1}{4} + \frac{1}{4}\right)} = 2.13$	3.3	-6.229, 7.829	No

### Multiple Comparisons

Dependent Variable: BINDING

Bonferroni

(I) DRUG	(J) DRUG	Mean Difference (I-J)	Std. Error	Sig.	95% Confidence Interval	
					Lower Bound	Upper Bound
P	T	-2.900	2.0208	1.000	-9.541	3.741
	S	20.775*	2.1302	.000	13.775	27.775
	E	9.500*	2.3008	.009	1.939	17.061
	C	.800	2.1302	1.000	-6.200	7.800
T	P	2.900	2.0208	1.000	-3.741	9.541
	S	23.675*	2.0208	.000	17.034	30.316
	E	12.400*	2.2000	.000	5.171	19.629
	C	3.700	2.0208	.870	-2.941	10.341
S	P	-20.775*	2.1302	.000	-27.775	-13.775
	T	-23.675*	2.0208	.000	-30.316	-17.034
	E	-11.275*	2.3008	.002	-18.836	-3.714
	C	-19.975*	2.1302	.000	-26.975	-12.975
E	P	-9.500*	2.3008	.009	-17.061	-1.939
	T	-12.400*	2.2000	.000	-19.629	-5.171
	S	11.275*	2.3008	.002	3.714	18.836
	C	-8.700*	2.3008	.018	-16.261	-1.139
C	P	-.800	2.1302	1.000	-7.800	6.200
	T	-3.700	2.0208	.870	-10.341	2.941
	S	19.975*	2.1302	.000	12.975	26.975
	E	8.700*	2.3008	.018	1.139	16.261

\*. The mean difference is significant at the .05 level.

Question: If we run an ANOVA and reject the null hypothesis, will we always find at least one significant pairwise difference?