

Experimental Design
Examination 1 (Chapters 16-17)

Name _____

Show your work. Be neat. Each of the 25 problems is worth 4 points.

Part I For a single-factor fixed ANOVA model with normal errors

$$Y_{ij} = \mu_i + \epsilon_{ij} \quad i = 1, \dots, r; \quad j = 1, \dots, n_i; \quad n_T = n_1 + \dots + n_r,$$

answer the following questions about the model assumptions and the ANOVA method.

1. The i th factor level mean μ_i is (i) a random variable (ii) an unknown constant . (Circle)
2. The error term ϵ_{ij} has a _____ distribution with mean _____ and common variance σ^2 .
3. The response variable Y_{ij} has a _____ distribution with mean _____ and variance _____.
4. The total sum of squares, SSTO, has _____ degrees of freedom (df). SSTO can be partitioned into _____ with _____ df and _____ with _____ df.
5. Whether the null hypothesis ($H_0: \mu_1 = \dots = \mu_r$) holds or not, an unbiased estimator of σ^2 is _____.
6. If $\sum n_i (\bar{Y}_{i.} - \bar{Y}_{..})^2 / (r - 1)$ is very large relative to $\sum \sum (Y_{ij} - \bar{Y}_{i.})^2 / (n_T - r)$, we are likely to (i) reject H_0 (ii) not reject H_0 .
7. The fitted value \hat{Y}_{ij} is equal to _____.
8. The residual e_{ij} is equal to _____ - _____.

9. Which sum of squares is a measure of the (response) variation within treatments? _____

10. Which sum of squares is a measure of the (response) variation between treatments? _____

11. Which sum of squares is a measure of the total (response) variation? _____

12. Another term for “factor level” is _____.

Part II Use the attached Minitab output and any other sources to work the following problems.

13. The response value Y_{24} is (a) 264.3 (b) 248.6 (c) 261.06 (d) 249.42 (e) none of the prior .

14. The fitted value \hat{Y}_{24} is (a) 264.3 (b) 248.6 (c) 261.06 (d) 249.42 (e) none of the prior.

15. Using level of significance $\alpha = .05$ in testing whether the mean distances are the same for all the ball brands, our conclusion is that

- (a) there is not sufficient evidence to say that the means are the same.
- (b) there is sufficient evidence to say that the means are not all the same.
- (c) there is not sufficient evidence to say that the means are not all the same.
- (d) there is sufficient evidence to say that the means are the same.

16. An unbiased estimator of σ^2 , the variance of distance for all brands, is 903.1 yd².

True False

17. In terms of mean distance, the difference between Brand B and Brand C is statistically significant (using level of significance $\alpha = .05$).

True False

18. In terms of mean distance, the difference between Brand A and Brand D is statistically significant (using level of significance $\alpha = .05$).

True False

19. Using the Tukey multiple comparison procedure, the confidence intervals for pairwise comparisons of means all have the same width. Explain why this is so.
20. Using the Tukey procedure results, we are 95% confident that the mean distance for Brand C exceeds the mean distance for Brand B by at least _____ yards.
21. Using the Bonferroni procedure results, we are 95% confident that the mean distance for Brand C exceeds the mean distance for Brand B by at least _____ yards.
22. If the Scheffe multiple comparison procedure were used, we would be 95% confident that the mean distance for Brand C exceeds the mean distance for Brand B by at least _____ yards.
23. Find the Bonferroni critical value used for making all 6 pairwise comparisons.
24. The ANOVA table gives us $MSE = 25.2$. Show how we can derive the MSE using the estimated standard deviations for each brand.
25. Let $L = \frac{\mu_A + \mu_B}{2} - \frac{\mu_C + \mu_D}{2}$. Obtain a 95% confidence interval for L. Interpret your interval.

$$\leq L \leq$$

Interpretation:

(Graduate students.) What does “data snooping” mean? Describe a situation in which data snooping can lead to a misspecified significance level.

Golf Ball Problem and Minitab Output

Suppose the United States Golf Association (USGA) wants to compare the mean distances associated with four different brands of golf balls when struck with a driver. A completely randomized design is employed, with Iron Byron, the USGA's robotic golfer, using a driver to hit a random sample of 10 balls of each brand in a random sequence. The distance (in yards) is recorded for each hit, and the results are shown in the table below.

Distances hit by Byron

ROW	brandA	brandB	brandC	brandD
1	251.2	263.2	269.7	251.6
2	245.1	262.9	263.2	248.6
3	248.0	265.0	277.5	249.4
4	251.1	254.5	267.4	242.0
5	265.5	264.3	270.5	246.5
6	250.0	257.0	265.5	251.3
7	253.9	262.8	270.7	262.8
8	244.6	264.4	272.9	249.0
9	254.6	260.6	275.6	247.1
10	248.8	255.9	266.5	245.9

Decriptive Statistics by Brand

STAT> describe c1 c2 c3 c4

	N	MEAN	MEDIAN	TRMEAN	STDEV	SEMEAN
brandA	10	251.28	250.55	250.34	5.98	1.89
brandB	10	261.06	262.85	261.39	3.87	1.22
brandC	10	269.95	270.10	269.85	4.50	1.42
brandD	10	249.42	248.80	248.67	5.47	1.73

ANALYSIS OF VARIANCE

SOURCE	DF	SS	MS	F	p
TREATMENT	3	2709.2	903.1	35.81	0.000
ERROR	36	907.9	25.2		
TOTAL	39	3617.1			

INDIVIDUAL 95 PCT CI'S FOR MEAN

BASED ON POOLED STDEV

LEVEL	N	MEAN	STDEV	-----+-----+-----+-----+-----			
1	10	251.28	5.98	(---*---)			
2	10	261.06	3.87		(---*---)		
3	10	269.95	4.50			(---*---)	
4	10	249.42	5.47	(---*---)			
				-----+-----+-----+-----+-----			
POOLED STDEV =	5.02			248.0	256.0	264.0	272.0

MULTIPLE COMPARISONS

Tukey's multiple comparison procedure

Nominal level = 0.0500
 Family error rate = 0.0500
 Individual error rate = 0.0107

Critical value = 3.80

Intervals for (mean of column group) - (mean of row group)

	1	2	3
2	-15.821		
	-3.739		
3	-24.711	-14.931	
	-12.629	-2.849	
4	-4.181	5.599	14.489
	7.901	17.681	26.571

Bonferroni Confidence Intervals with Family Confidence Coefficient .95

Intervals for (mean of column group) - (mean of row group)

	1	2	3
2	-16.050		
	-3.510		
3	-24.940	-15.160	
	-12.400	-2.620	
4	-4.410	5.370	14.260
	8.130	17.910	26.800