

1. If children are given more choices within a class of products, will they tend to prefer that product to a competing product that offers fewer choices? Marketers want to know. An experiment prepared three sets of beverages. Set 1 contained two milk drinks and two fruit drinks. Set 2 had two fruit drinks and four milk drinks. Set 3 contained four fruit drinks but only two milk drinks. The researchers divided 210 children aged 4 to 12 years into three groups at random. They offered each group one of the sets. As each child chose a beverage to drink from the set presented, the researchers noted whether the choice was a milk drink or a fruit drink.
 - a) What are the experimental subjects?
 - b) What is the factor and what are its levels? What is the response variable?
 - c) Use a diagram to outline a completely randomized design for the study
 - d) Would it be reasonable to use a matched pair for this study? Why or why not?
 - e) Use a diagram to outline a blocked design for the study
2. A random sample of 26 offshore oil workers took part in a simulated escape exercise, and their times (sec) to complete the escape are recorded. The sample mean is 370.69 sec and the sample standard deviation is 24.36 sec. Construct a 95% lower confidence bound on the true average escape time. Interpret your interval.
3. The life in hours of a battery is known to be approximately normally distributed. The manufacture claims that the average battery life exceeds 40 hours. A random sample of 10 batteries has a mean life of 40.5 hours and sample standard deviation $s=1.25$ hours. Carry out a hypothesis test for $H_0: \mu = 40 \text{ hrs}$ vs $H_a: \mu > 40 \text{ hrs}$. $\alpha = 0.05$
4. The overall distance traveled by a golf ball is tested by hitting the ball with Iron Byron, a mechanical golfer with a swing that is said to emulate the legendary champion, Byron Nelson. Ten randomly selected balls of two different brands are tested and the overall distance measured. The data follow:

Brand 1: 275, 286, 287, 271, 283, 271, 279, 275, 263, 267
Brand 2: 258, 244, 260, 265, 273, 281, 271, 270, 263, 268

 - a) Which procedure is the most appropriate, 2-sample independent or 2-sample pairs? Explain
 - b) Find a 95 % confidence interval for the difference of the mean.
 - c) Use the four-step procedure to carry out a hypothesis test to determine whether the mean overall distance for brand 1 and brand 2 are different.
5. The Indiana State Police wish to estimate the average mph being traveled on the Interstate Highways, which cross the state. If the estimate is to be within ± 5 mph of the true mean with 95% confidence and the estimated population standard deviation is 25 mph, how large a sample size must be taken?
6. A laboratory is testing the concentration level in mg/ml for the active ingredient found in a pharmaceutical product. In a random sample of 10 vials of the product, the mean and the sample standard deviation of the concentrations are 2.58 mg/ml and 0.09 mg/ml. Find a 95% confidence interval for the mean concentration level in mg/ml for the active ingredient found in this product.
7. An investigator wishes to estimate the difference between two population mean lifetimes of two different brands of batteries under specified conditions. If the population standard deviations are both roughly 2 hr and the sample size from the first brand will be twice the sample size from the second brand, what values of the sample sizes will be necessary to estimate the difference to within 0.5 hours with 99% confidence?

8. The following summary data on proportional stress limits for two different type of woods, Red oak and Douglas fir.

Type of Wood	Sample Size	Sample Mean	Sample Standard Deviation
Red oak	50	8.51	1.52
Douglas fir	62	7.69	3.25

- Find a 90% confidence interval for the difference between true average proportional stress limits for the Red oak and that for the Douglas fir. Interpret your result.
- A test of hypotheses is conducted at $\alpha=0.10$ to determine if the stress limits are the same for the two type of woods.
- Explain how you can use the confidence interval in part (a) to draw a conclusion in the test of hypotheses.

9. The accompanying summary data on the ratio of strength to cross-sectional area for knee extensors is from the article "Knee Extensor and Knee Flexor Strength: Cross Sectional Area Ratios in Young and Elderly Men":

Group	Sample Size	Sample Mean	Sample Standard Deviation
Young Men	50	7.47	0.44
Elderly Men	45	6.71	0.56

Does the data suggest that the true average ratio for young men exceeds that for elderly men? Carry out a test of significance using $\alpha = 0.01$.

10. Coronary heart disease (CHD) begins in young adulthood and is the fifth leading cause of death among adults aged 20 to 24 years. Studies of serum cholesterol levels among college students, however, are very limited. A 1999 study looked at a large sample of students from a large southeastern university and reported that the mean serum cholesterol level among women is 168 mg/dl with a standard deviation of 27 mg/dl. A more recent study at a southern university investigated the lipid levels in a cohort of sedentary university students. The mean total cholesterol level among $n = 71$ females was $\bar{x} = 173.7$. Is there evidence that the mean cholesterol level among sedentary students differs from this average over all students? Use the four-step procedure to carry out a test of significance. Use $\alpha = 0.05$.

11. Fifteen adult males between the ages 35 and 45 participated in a study to evaluate the effect of diet and exercise on blood cholesterol levels. The total cholesterol was measured in each subject initially, and then three months after participating in an aerobic exercise program and switching to a low-fat diet. The data are shown in the accompanying table.

Table I: Blood Cholesterol Levels for 15 Adult Males

Subject	1	2	3	4	5		6	7	8	9	10	11	12	13	14	15
Before	265	240	258	296	251		245	287	314	260	279	283	240	238	225	247
After	229	231	227	240	238		241	234	256	247	239	246	218	219	226	233

	N	Mean	StDev	SE Mean
Before	15	261.80	24.96	6.45
After	15	234.93	10.48	2.71
Diff (Before - After)	15	26.87	19.04	4.92

- Find a 90% lower confidence bound for the true mean reduction of the cholesterol reduction.
- Carry out a test of hypotheses to determine if the data support the claim that the low-fat diet and aerobic exercise are of value in producing a mean reduction in blood cholesterol levels? Use $\alpha=0.05$.

12. True or False Questions (explain why):

- ANOVA tests the null hypothesis that the sample means are all equal.
- A strong case for causation is best made in an observational study.
- You use ANOVA to compare the variances of the populations.
- In rejecting the null hypothesis, one can conclude that all the means are different from one another.
- One-way ANOVA can be used only when there are two means to be compared.
- The ANOVA F statistic will be large when the within-group variation is much larger than the between-group variation.

13. For each of the following situations, identify the response variable and the populations to be compared, and give I , N and (a) Degrees of freedom for group, for error, and for the total (b) Null and alternative hypotheses (c) Numerator and denominator degrees of freedom for the F statistic

- A poultry farmer is interested in reducing the cholesterol level in his marketable eggs. He wants to compare two different cholesterol-lowering drugs added to the hens' standard diet as well as an all-vegetarian diet. He assigns 25 of his hens to each of the three treatments.
- A researcher is interested in students' opinions regarding an additional annual fee to support non-income-producing varsity sports. Students were asked to rate their acceptance of this fee on a seven-point scale. She received 94 responses, of which 31 were from students who attend varsity football or basketball games only, 18 were from students who also attend other varsity competitions, and 45 were from students who did not attend any varsity games
- A professor wants to evaluate the effectiveness of his teaching assistants. In one class period, the 42 students were randomly divided into three equal-sized groups, and each group was taught power calculations from one of the assistants. At the beginning of the next class, each student took a quiz on power calculations, and these scores were compared.

14. Various studies have shown the benefits of massage to manage pain. In one study, 125 adults suffering from osteoarthritis of the knees were randomly assigned to one of five 8-week regimens. The primary outcome was the change in the Western Ontario and McMaster Universities Arthritis Index (WOMAC-Global). This index is used extensively to assess pain and functioning in those suffering from arthritis. Negative values indicate improvement. The following table summarizes the results of those completing the study.

Regimen	n	\bar{x}	s
30 min massage 1 \times /wk	22	-17.4	17.9
30 min massage 2 \times /wk	24	-18.4	20.7
60 min massage 1 \times /wk	24	-24.0	18.4
60 min massage 2 \times /wk	25	-24.0	19.8
Usual care, no massage	24	-6.3	14.6

- What proportion of adults dropped out of the study before completion?
 - Is it reasonable to use the assumption of equal standard deviations when we analyze these data? Give a reason for your answer.
 - The $SSA = 5060.346$. $MSE = 339.32$ Test the null hypothesis that the mean change in WOMAC-Global score is the same for all regimens at a 5% significance level. Be sure to include the degrees of freedom for the test statistic. I would suggest that you create the ANOVA table before you start.
 - There are 10 pairs of means to compare. For this part, assume that the number of observations for each case is 24. Determine the critical value for the Tukey multiple-comparisons method at a 5% significance level. Which pairs of means are found to be significantly different? Be sure to include a visual assessment. Write a short summary of your analysis including which of the methods is best. Note: on the exam, if this question is asked, there will be at most 3 different pairs.
15. Data show that married, divorced, and widowed men earn quite a bit more than men the same age who have never been married. This does not mean that a man can raise his income by getting married, because men who have never been married are different from married men in many ways other than marital status. Suggest several lurking variables that might help explain the association between marital status and income.