THE

QUALITY ENGINEER

SOLUTIONS TEXT

SECTION XI

ADVANCED STATISTICS -- TEST QUESTIONS

11.1. To state that a model in an experimental design is fixed indicates that:

a. The levels used for each factor are the only ones of interest
b. The levels were chosen from a fixed population
c. The equipment from which the data is collected must not be moved
d. The factors under consideration are qualitative

Solution: Answers **b, c**, and **d** are all filler answers. In fact, if you selected **c** as your answer, you should seriously consider taking the CQE exam at a later date. Experimental design levels are established (or fixed) based on the best advice of people knowledgeable of the process. A balanced design is then considered only at those levels. Based upon analysis, factors may then be adjusted to other fixed levels for subsequent experimentation. The objective is to achieve optimum performance.

Answer a is correct.

References: *CQE Primer*, Section XI - 79 and 87/88 (and logic). This question has been modified from a published 1984 CQE exam.

11.2. When finding a confidence interval for mean μ, based on a sample size of n:

a. Increasing n increases the interval
b. Having to use s_X instead of n decreases the interval
c. The larger the interval, the better the estimate of μ
d. Increasing n decreases the interval

Solution: To determine a confidence interval for μ (for a normal distribution, σ = 0.05).

$$\overline{X} \pm 1.96 \frac{s}{\sqrt{n}}$$

Where,

n = sample size and
σ = population standard deviation

Answer **b**, as presented, doesn't make sense. Perhaps there's a typo involved and n should be σ. Answer **c** is also incorrect. The larger the interval, the poorer the estimate of μ. From the above equation, as the sample size is increased, the confidence interval will become smaller.

Answer d is correct.

References: *CQE Primer*, Section XI - 4. This question has been modified from a published 1972 CQE exam and an old CQE Brochure.

ADVANCED STATISTICS -- TEST QUESTIONS

11.3. A two-way analysis of variance has r levels for one variable, c levels for the second variable, with 2 observations per cell. The degree of freedom for interaction is:

a. 2(r)(c)
b. (r - 1)(c - 1)
c. rc -1
d. 2(r -1)(c - 1)

Solution: The degrees of freedom for an interaction in a two-way analysis of variance, regardless of the number of repetitions, is always (r - 1)(c - 1). The number of repetitions is used in the determination of the sum of squares of the interaction, not in the degrees of freedom.

Answer b is correct.

References: *CQE Primer*, Section XI - 48 and 59. This question has been modified from a published 1984 CQE exam.

11.4. Determine whether the following two types of rockets have significantly different variances at the 5% level. Assume that the larger variance goes in the numerator.

Rocket A	Rocket B
61 readings	31 readings
1,347 $miles^2$	2,237 $miles^2$

a. Significant difference because $F_{calc} < F_{table}$
b. No significant difference because $F_{calc} < F_{table}$
c. Significant difference because $F_{calc} > F_{table}$
d. No significant difference because $F_{calc} > F_{table}$

Solution: This question requires an F test calculation. Note from the wording of the question that a two tailed test is required and that the variances are provided. The biggest variance is placed in the numerator.

$$F = \frac{s_1^2}{s_2^2}$$

$$H_0: s_A^2 = s_B^2 \qquad H_1: s_A^2 \neq s_B^2$$

$DF_1 = 30$, $DF_2 = 60$ The upper critical value of F, $\alpha = 0.025$, is 1.82.

$$F = \frac{s_1^2}{s_2^2} = \frac{2{,}237 \text{ miles}^2}{1{,}347 \text{ miles}^2} = 1.66$$

The question states that $(s_1)^2$ is the larger variance. This avoids the tricky problem of finding the critical F value for the left-tail test. Since F calculated is less than F critical, the null hypothesis cannot be rejected. Our conclusion is that the data only permits us to act as if the null hypothesis were true. One must say there is no significant difference. If the student wishes to determine the left-tail critical value, the degrees of freedom must be reversed in the F table and the reciprocal determined. That is, 1/1.94 = 0.515. In this case, F calculated is between the two critical values. The null hypothesis still cannot be rejected, but the answer choices would not apply.

Answer b is correct.

References: *CQE Primer*, Sections XI - 35/37 and XII - 12. This question has been modified from a published 1978 CQE exam.

ADVANCED STATISTICS -- TEST QUESTIONS

11.5. The advantage of using the "modern designed" method of experimentation, rather than 1 FAT, is that:

a. Holding all factors constant, except the factor under investigation, is less costly
b. Experimental error is recognized, but need not be stated in quantitative terms
c. Fewer terms and measurements are needed for valid and useful information
d. Measurement is often assumed to have no effect and provides greater test freedom

Solution: The term 1 FAT (one factor at a time) refers to varying one factor at a time, while holding all other factors constant. Although this approach may work for very simple problems, it causes havoc with moderately complex systems. The “fixed” factors do vary which can waste time, effort and money. This traditional approach can yield invalid or inconclusive results. Modern design experimentation (including fractional-factorials, improved three factor designs, and Latin squares) squeeze a large amount of valid information from a few trials.

Answer c is correct.

Reference: *CQE Primer*, Section XI - 74 and 86.

11.6. When constructing a power of test curve, one would not be surprised to discover that as alpha (α) increases:

a. The value of mu (μ) becomes greater
b. Mu (μ) approaches a value of 1
c. 1 - beta (β) increases
d. The sample size becomes larger

Solution: When constructing a power curve, 1 - beta is plotted against alternate values of mu. Answers **a** and **b** are fabricated distracters. In general, as alpha increases, beta decreases, and the power of 1 - beta increases. Increasing the sample size decreases both alpha risks and beta risks and increases the power of the (1 - beta) test.

Answer c is correct.

Reference: *CQE Primer*, Section XI - 8 and 27/29.

11.7. Given the data below is normally distributed, and the population standard deviation is 3.1, what is the 90% confidence interval for the mean?

22, 23, 19, 17, 29, 25

a. 20.88 - 24.12
b. 20.42 - 24.59
c. 21.65 - 23.35
d. 17.4 - 27.60

Solution: The expression for computing a confidence interval for the mean, given the population is normal, and the standard deviation is known, is:

$$\overline{X} - Z_{\frac{\alpha}{2}} \frac{\sigma}{\sqrt{n}} < \mu < \overline{X} + Z_{\frac{\alpha}{2}} \frac{\sigma}{\sqrt{n}}$$

σ is the population standard deviation, n is the sample size, and *z* is the standard normal deviate. For a 90% confidence interval, 5% of the area under the standard normal curve should be to the left of -*z*, and 5% of the area under the standard normal curve should be to the right of *z*, thus, *z* = 1.645. Entering the equation above with X-bar = 22.5, σ = 3.1, and n = 6 gives a confidence interval of 20.42 to 24.59.

Answer b is correct.

Reference: *CQE Primer,* Section XI - 4.

11.8. A study was conducted on the relationship between the speed of different cars and their gasoline mileage. The correlation coefficient was found to be 0.35. Later, it was discovered that there was a defect in the speedometers and they had all been set 5 miles per hour too fast. The correlation coefficient was computed using the corrected scores. Its new value will be:

a. 0.30
b. 0.35
c. 0.40
d. -0.35

Solution: A positive value for r indicates a correlation line that slopes upward. The input factor must be the speed and the output factor will be gasoline mileage. The correlation chart (scatter diagram) might look like Figure A.. If all mileage readings were found to be 5 miles per hour too fast, the new chart would look like Figure B.

Figure A

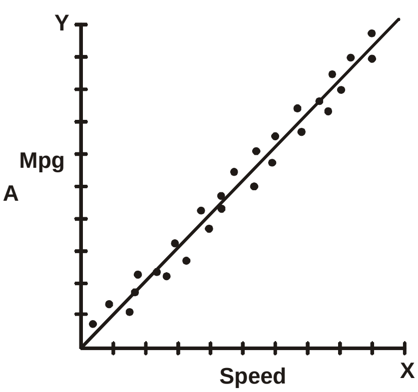
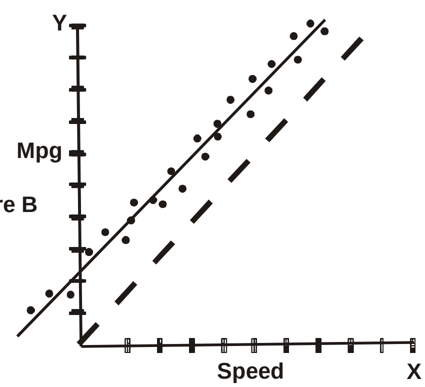


Figure B



The r value would remain the same. Since the sign of r tracks the slope of the line, the real r value, in this case, would be negative (speed has a negative effect on gas mileage), but that is not what we are to assume from the question.

Answer b is correct.

References: *CQE Primer*, Section XI - 70 (and logic). This question has been modified from a published 1984 CQE exam.

11.9. Assume that data for a goodness of fit test has been structured into effective cells for a chi-square calculation. Which of the following distributions would have the fewest degrees of freedom?

a. Normal
b. Poisson
c. Binomial
d. Uniform

Solution: Assuming the same data is being used for a determination of appropriate distribution using a chi-square test, the appropriate degree of freedom are:

Distribution	**DOF**
Normal	No. of cells -3 sigma
Poisson	No. of cells -2 sigma
Binomial	No. of cells -2 sigma
Uniform	No. of cells -1 sigma

The normal distribution would have the smallest DOF. The uniform distributions would have the largest DOF.

Answer a is correct.

Reference: *CQE Primer*, Section XI - 39.

11.10. In performing an analysis of variance for a single factor experiment, a fundamental assumption is that the factor:

a. Means are equal
b. Means are unequal
c. Variances are equal
d. Variances are unequal

Solution: Whether testing the values of a single or multiple population means the underlying assumption is that variation within each factor(s) is the same. This is referred to as homogeneity of variance.

Answer c is correct.

References: *CQE Primer*, Section XI - 50. This question has been modified from a published 1984 CQE exam.

11.11. The critical value of the t distribution is:

a. Always greater than the critical value of the normal distribution
b. Always less than the critical value of the normal distribution
c. Not related to the critical value of the normal distribution
d. Approaches the critical value of the normal distribution as the degrees of freedom increase

Solution: The t distribution compensates for not knowing the variance of the population. As the sample size increases, a more accurate estimate of the variance is obtained, and the t distribution approaches normal.

Answer d is correct.

Reference: *CQE Primer,* Section XI - 13/17.

11.12. A designed experiment has been conducted at three levels (A, B, and C) yielding the following "coded" data:

A	B	C
6	5	3
3	9	4
5	1	2
2		

As a major step in the analysis, the degrees of freedom for the "error" sum of squares is determined to be:

a. 7
b. 9
c. 6
d. 3

Solution: This question requires a calculation of the DF for the error sum of squares.

$$MSE = \frac{SSE}{N - t}$$

Where,
MSE = mean sum of squares (error term)
N = number of tests
SSE = sum of squares (error term)
t = number of treatments

From the above equation, the DF for the error sum equals:

$$DF_{ERROR} = N - t = 10 - 3 = 7$$

Answer a is correct.

References: *CQE Primer*, Section XI - 51. This question has been modified from a published 1974 CQE exam.

