18. **(B)** The number of runners whose times were less than or equal to 20 minutes was $80(0.2625) = 21$ and the number of runners whose times were less than or equal to 22 minutes was $80(0.5125) = 41$. Therefore, the number of runners whose times were more than 20 minutes and at most 22 minutes was $41 - 21 = 20$.

19. **(B)** The first (lower) quartile is the time below which 25% of the runners achieved. Thus the first quartile is between 18 and 20 minutes. The third (upper) quartile is the time below which 75% of the runners achieved. Thus the third quartile is between 24 and 26 minutes. According to this information, the smallest possible value of the interquartile range is $24 - 20 = 4$ minutes, and the largest possible value of the interquartile range is $26 - 18 = 8$ minutes. The only one of the five options that is between 4 and 8 minutes is option (B).

20. **(E)** The statement in (A) is correct, since the *p*-value for the test is given to be 0.127, which is greater than 0.05. The statement in (B) is correct, since a two-sided *t*-test with 90 degrees of freedom is being performed; when the value of the *t*-statistic is positive, failure to reject H_0 is equivalent to this value being less than the positive critical value for a single-tail probability of 0.025. The statement in (C) is correct, since zero being in the confidence interval is equivalent to non-rejection of H_0 at the 0.05 level. The statement in (D) is a correct interpretation of the *p*-value: the probability of getting a test statistic at least as extreme as the one obtained, given that H_0 is true. The attempted reversal in (E) of this statement, however, is incorrect: no probability can be attached to the population means.

21. **(B)** Statistic 1 and Statistic 3 have similar amounts of variability. However, the center of the distribution of Statistic 1 seems to be at, or close to, 5, the true value of the parameter, and this cannot be said for Statistic 3. So Statistic 1 is preferable to Statistic 3. The center of the distribution of Statistic 2 also seems to be at, or close to, 5, but the variability of Statistic 2 seems to be less than that of Statistic 1. Therefore, Statistic 2 is preferable to Statistic 1, and the required list is 2, 1, 3.

22. **(D)** The power of a hypothesis test is the probability of correctly rejecting H_0 given that H_0 is false, for a specific alternative value of the parameter. The relevant parameter in this question is the population mean, μ, and we are told to assume that the true value of μ is greater than the hypothesized value, 8. Thus, for this value of μ, the power of the test is the probability that H_0 will be (correctly) rejected.

The test statistic here will be either z or t. Changing the significance level to 0.1 will increase the size of the rejection region at the right end of the distribution of the test statistic, thus increasing the probability that H_0 will be rejected; so the power of the test is increased. (Another way to think of this is that changing the significance level has opposite effects on the probabilities of Type I and Type II errors, and the probability of a Type I error is exactly the significance level of the test. Thus, changing the significance level from 0.05 to 0.1 increases the probability of a Type I error and decreases the probability of a Type II error. Decreasing the probability of a Type II error is increasing the power of the test.) This change has resulted in a test that has greater power.

If the test is changed to the two-tailed version, the 5% critical region will be apportioned equally between the left and right extremes of the distribution of the test statistic. Thus, the size of the critical region on the right will be reduced, making it less likely that H_0 will be rejected. Therefore, this change has resulted in a test that has smaller power.

Increasing the sample size from 50 to 100 means that more information is being provided, and therefore it is more likely that the hypothesis test will reach the correct conclusion (that μ is greater than 8). Thus the power of the test has increased.

(A more precise way of explaining why an increase in the sample size will increase the power of the test is as follows. We'll assume that a t-test is being performed. (A very similar argument will apply in the case of a z-test.) The test statistic is given by

$$t = \frac{\overline{X} - 8}{s/\sqrt{n}}.$$

For the sake of argument, let's assume that the true population mean is 8.5. For either sample size the value of $\overline{X}$ is then likely to be around 8.5. However, the value of $s/\sqrt{n}$ is almost certain to be smaller when $n = 100$ than when $n = 50$. Consequently, the value of t is almost certain to be larger for the larger sample size. Therefore, for the larger sample size there is a greater probability of H_0 being rejected, and thus a greater power.)

23. **(D)** In systematic sampling, the members of the population are numbered, and then, for example, every 100th item in the population will be included in the sample. This method can often provide a sample that represents the population well. Meanwhile, virtually any sampling procedure that involves randomness, including simple random sampling, has some chance of producing a sample that does *not* represent the population well. Therefore it is not true to say that a simple random sample will always represent the population better than a systematic sample.

24. **(E)** This is a chi-square test for goodness of fit. The observed counts are 192, 133, 118, and 57. The corresponding expected counts are $500(0.4) = 200$, $500(0.3) = 150$, $500(0.18) = 90$, and $500(0.12) = 60$. So the value of the test statistic is

$$\chi^2 = \sum \frac{(\text{observed} - \text{expected})^2}{\text{expected}}$$

$$= \frac{(192-200)^2}{200} + \frac{(133-150)^2}{150} + \frac{(118-90)^2}{90} + \frac{(57-60)^2}{60} = 11.108.$$

25. **(C)** This process consists of repeated trials, where the outcome of each trial is either "success" (commercial vehicle) or "failure" (non-commercial vehicle). The probability of success on each trial is the same $(1/5)$, and the outcomes of the trials are independent of each other. The trials are repeated until a success is obtained, and the random variable we're interested in is defined to be the number trials up to and including the first success. Thus, the random variable has a geometric distribution.