1. Statistics is about ______________.

2. Define each of the following terms:
   a) sampling distribution of a statistic
   b) expected value of a statistic
   c) standard error of a statistic

3. Give the shapes, expected values and standard errors of each of these sampling distributions:
   a) \( \bar{x} \), when the population is normally distributed with known mean and standard deviation
   b) \( \bar{x} \), when the population is normally distributed, the standard deviation is not known, and \( n < 30 \).
   c) \( \bar{x} \), when the shape of the population distribution is unknown or known to be non-normal but \( n \geq 30 \).
   d) \( p \), the sample proportion

4. The weights of grapefruit are normally distributed with \( \mu = 17 \) oz. and \( \sigma = 1 \) oz. Find the expected value of the sample mean and standard error of the mean for a sample of 25 of these grapefruit.

5. For these grapefruit,
   a) (N) what is the probability that a randomly selected grapefruit will have a weight of 18 oz or more?
   b) (N) a sample of 25 grapefruit will have a mean of 18 oz or more?
   c) draw a sketch illustrating how to find the probabilities in a) and b).

6. State the Central Limit Theorem and explain its significance in sampling problems.

7. Transit times of packages shipped via Kamikaze Couriers are skewed with \( \mu = 48 \) min. and \( \sigma = 12 \) minutes. Characterize the distribution for a sample of \( n = 18 \); for a sample of \( n = 64 \).

8. Twenty per-cent of the flashlight batteries in a military warehouse have been stored too long and are defective. For a sample of 400 of these batteries,
   a) find \( \sigma_p \)
   b) find \( E(p) \)
   c) (N) what’s the probability that the proportion of defectives in the sample will be less than 24%?
   d) draw a sketch of the distribution of \( p \) and use it to explain how to calculate the probability in c).

9. From a truckload of Watauga County cabbages we drew a sample of 36 and found their mean weight to be 2 lb. In this case \( \bar{x} \) is a __________ estimate of ________.

10. Why do statisticians generally prefer interval estimates to point estimates?

11. Asked whether they preferred blades or electric razors, 100 of a sample of 300 men said they preferred electric razors. The best point estimate we could give of the population proportion preferring electric razors is ________.

12. In each of the following cases, state whether you should use the z or t distribution in constructing confidence intervals. Be sure you know why in each case.
   a) We have a sample of 68 children, and we wish to establish the average time they will take to run 1 mile; we believe that these times will be skewed upwards, and \( \sigma \) is unknown.
b) We have a sample of 17 men who have been put on a weight-loss diet; we wish to estimate the average weight loss, and we believe that weight loss will be normally distributed but otherwise we know nothing about the population distribution.

c) In part b, the researchers got some additional funds and increased the sample size to 54.

d) A random sample of 1122 Americans were asked whether they thought we should impose import restrictions on Japanese products; 56% said we should.

e) Reading speed in the human population is known to be normally distributed with standard deviation = 85 words per minute. Using a sample of 22 people, we would like to estimate the average reading speed among last year’s graduates of Lizard Lick High School.

13. Suppose in 12.a. above, we had $\bar{x} = 11$ minutes, and on this basis we said, “We are 95% confident that the average American child can run a mile in somewhere between 10.2 and 11.8 minutes.” What exactly does “confident” mean?

14. Apple weights are normally distributed; a sample of 25 apples has $\bar{x} = 5$ oz with $s = 1.2$ oz. A 95% confidence interval for the weight of all apples, constructed using a t value of 2.069, is ______.

15. Starting salaries of college graduates are skewed with population standard deviation $\sigma = $4,000; in a sample of 100 recent graduates, $\bar{x} = $22,000. Since $z$ here = 2.33, a 98% confidence interval for the average starting salary of all recent college graduates will be __________.

16. A sample of 200 men hit by .357 magnum bullets was drawn from police files. Of these 200, 40 were either killed or downed and disabled by one shot. A 90% confidence interval for the proportion of the population who would be put out of action by one shot from a .357 is (use $z = \pm 1.64$) __.

17. Length of time spent shopping for a color television set is normally distributed with $\sigma = 3$ hours; the mean time is unknown. In a sample of 28 shoppers, mean time before making a purchase decision was 9.4 hours. With $z = 2.58$, a 99% confidence interval for the mean time all shoppers would spend is 9.4 ± __.

18. Under what conditions may the z and the t distributions be used in calculating confidence intervals?

19. (N) The following sample values came from a normal population with $\sigma = 8$. Construct 90% and 99% confidence intervals for the population mean. Sample values: 54, 62, 48, 31, 51, 47, 53, 38.

20. (N) The following sample values came from a normally distributed population with unknown standard deviation. Construct a 95% and 98% confidence interval for the population mean. Sample values: 23, 18, 35, 41, 34, 26, 29, 19, 37, 26, 30, 31.

21. State Excel cell formulas that would give you each of the following:
   a) the probability of values between 20 and 30 on a normal distribution with $\mu = 35$ and $\sigma = 10$.
   b) a value $x^I$ such that only 2% of $x$’s exceed $x^I$ on a normal distribution with $\mu = 100$ and $\sigma = 15$.
   c) the appropriate z value to calculate a 92% confidence interval.
   d) the appropriate t value for a 96% confidence interval with 23 degrees of freedom.
   e) the standard error of the mean for sample data which are recorded in cells B6 through B28.

22. List the arguments needed for each of the following Excel functions and state what Excel displays in each of these cases:
   a) NORMDIST
   b) NORMINV
   c) NORMSDIST
   d) NORMSINV
   e) TDIST
   f) TINV
23. Explain the steps needed to use the Descriptive Statistics tool in the Excel Data Analysis Tool Pack. Explain each of the entries in the following output table:

<table>
<thead>
<tr>
<th>Errors Found Per Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>Mean</td>
</tr>
<tr>
<td>Standard Error</td>
</tr>
<tr>
<td>Median</td>
</tr>
<tr>
<td>Mode</td>
</tr>
<tr>
<td>Standard Deviation</td>
</tr>
<tr>
<td>Sample Variance</td>
</tr>
<tr>
<td>Kurtosis</td>
</tr>
<tr>
<td>Skewness</td>
</tr>
<tr>
<td>Range</td>
</tr>
<tr>
<td>Minimum</td>
</tr>
<tr>
<td>Maximum</td>
</tr>
<tr>
<td>Sum</td>
</tr>
<tr>
<td>Count</td>
</tr>
</tbody>
</table>

24. (N) The following data are a POPULATION: 142, 81, 65, 114, 121, 93, 87, 102, 127, 130. For this population, find
   a) the mean
   b) the variance and standard deviation
   c) the median
   d) the first and third quartiles
   e) the interquartile range

25. State in words what result each of the following formulas would give you if entered into an Excel spreadsheet. For example, NORMSINV(.05) returns $z_0$ such that only five per cent of the standard normal distribution is less than $z_0$.
   a) NORMDIST(23, 31, 2, TRUE)
   b) 1 – NORMDIST(33, 31, 2, TRUE)
   c) NORMINV(.95, 31, 2)
   d) NORMSDIST(-2.58)
   e) TDIST(2.33, 23, 2)
   f) TDIST(2.33, 23, 1)
   g) TINV(.05, 23)
SELF TEST TWO

1. In hypothesis testing, how are the values to be tested by the hypotheses chosen?

2. Fill in the blanks: The basic logic of hypothesis testing is that if we get a sample which is improbable if the _____ _____ is true, then we must either believe that we got an ________ sample or that the _____ _____ is ________.

3. What is meant by the significance level of an hypothesis test?

4. The probability of a Type I error = _______, but also α = the ________ _______ of the test.

5. What are the six formal steps in hypothesis testing?

6. What is the difference between a one-tailed and a two-tailed hypothesis test?

7. Explain the circumstances under which a z value can be used in hypothesis tests and those in which a t value must be used.

8. (N) Find the appropriate z value for each of the following tests:
   a) upper one-tailed test at 5% significance
   b) two-tailed test at 5% significance
   c) lower one-tailed test at 1% significance
   d) two-tailed test with α = 0.01
   e) lower one-tailed test with α = 0.025
   f) two-tailed test at 15% significance

9. (N) Find the appropriate t values for the tests in question 8. For a - c, assume n = 18; for d - f, assume n = 25.

   For questions 10 to 14, state the appropriate null and alternative hypotheses, state whether the test statistic is a z or a t and calculate the test statistic.

10. Engineers at ART, Inc. wish to test whether the mean length of a certain spring is at least 40 mm; it is known that the population standard deviation is 0.5 mm and that the population is normally distributed. In a sample of 16 springs, \( \bar{x} = 39.7 \) mm.

11. A disillusioned and discouraged professor claims that at least half his students have paying jobs which interfere with their studies. An enterprising student decides to test the claim by taking a sample of her classmates; in a sample of 120, she finds that 63 have jobs.

12. An auto company official claims that his company's cars on average have fewer than 8 defects when delivered to customers. A consumer testing company selects a sample of 23 of the company's cars and finds that \( \bar{x} = 7.5 \) with \( s = 3 \). The testing company feels pretty sure that the distribution of defects is normal.

13. An auditor intends to use a sampling procedure to count the number of spare parts on hand in a warehouse. According to the company's records, there should be an average of 47 parts in each of a large number of bins; nothing is known about how these parts are distributed. In a sample of 64 bins, the average number of parts is 44 with a sample standard deviation of 8.

14. A nutritionist claims that the average elementary school child eats 600 calories a day of junk food. Believing this distribution to be normal, a researcher attempts to check the nutritionist's claim by carefully recording the food intake of a sample of 25 children. In the sample the researcher finds \( \bar{x} = 640 \) with \( s = 100 \).
15. For question 10 – 14, parts a – e give the critical values of the appropriate test statistics. In each case, state whether the null should be rejected and what you are concluding about the population parameter.
   a) For $\alpha = 0.05$, $z_C = -1.645$.
   b) For significance level 0.01, $z = -2.33$.
   c) For significance level 0.05, $t$ with 22 degrees of freedom = $-1.717$.
   d) For $\alpha = 0.05$, $t = \pm 1.993$.
   e) For $\alpha = 0.05$, with 24 d.f., $t = \pm 2.064$.

16. Give a definition of the p-value of an hypothesis test and explain how the p values could be calculated for each of questions 10 – 14.

17. (N) Find the p value for each of the tests in questions 10 – 14.

18. In testing the hypothesis $H_0: \mu \geq 230$ vs. $H_A: \mu < 230$, engineers at ART, Inc. got a sample mean of 218, which they found (using NORMDIST) to have a probability of 0.03 if the null hypothesis were true. If the test is at $\alpha = 0.05$, what should be concluded?

19. Down the road at Mass Marketing, researchers tested $H_0: \mu = 3,000$ against $H_A: \mu \neq 3,000$. The result of the sample taken was $\bar{x} = 2840$; if the null were true, the probability of a sample mean that small or smaller is 0.03 (calculations using NORMDIST). If the test is conducted at 5% significance, what should the researchers conclude?

20. In an upper one-tailed hypothesis test using the z distribution, the calculated value of $z = 1.4325$. State an Excel cell formula that would give the p value of this test.

21. In a two-tailed test using the t distribution, the calculated $t = -2.8943$ with 17 degrees of freedom. State an Excel cell formula that would give the p value of this test.

22. What two kinds of errors can be made in hypothesis testing?

23. What is the probability of a Type I error? Why is it not possible in advance to say what the probability is of a Type II error?

24. (N) Safety engineers claim that the average speed on Interstate Highways is at least 72 mph; it is known that motorists' speeds are normally distributed with $\sigma = 9$ mph. For a test at 5% significance and $n = 100$, $H_0$ will be rejected if $\bar{x} < 70.524$ mph. Suppose that in fact $\mu = 69.8$ mph. Find $\beta$, the probability of a Type II error.

25. For the situation in question 24, draw a sketch illustrating the areas on the hypothesized and actual distributions that represent $\alpha$ and $\beta$ respectively.

26. It is alleged that in households with children, mean monthly expenditures for food = $800. If $\sigma = 75$, and we wish to test this claim at 2% significance with a sample of 25 households.
   a) what values of $\bar{x}$ will cause us to reject the null hypothesis? (The critical z value in this case = $\pm 2.33$.)
   b) (N) Suppose that $\mu = 850$; find $\beta$.
   c) (N) Suppose that $\mu = 780$; find $\beta$.

27. Other things being equal, as $\alpha$ increases, $\beta$ __________.

28. If both $\alpha$ and $\beta$ are too high in a test, the only thing we can sensibly do is to ________.

29. What is meant by the power of an hypothesis test?
30. As the actual mean approaches the hypothesized mean, we expect the power of a test to ____________.

31. Generally speaking, a test will be most powerful when the actual mean is __________ to (from) the hypothesized mean.

32. If we are dissatisfied with the power of a particular test, what can we do to change it?

33. In a particular test, when $\mu = 40$, $\beta = 0.32$. Accordingly, when $\mu = 40$, the power of the test = __________.

34. Without being too specific, explain how we would go about determining the appropriate significance level and sample size for an hypothesis test.

35. (N) We are manufacturing graphite rods for use in a nuclear reactor. From experience, we know that these rods have standard deviation $\sigma = 0.5$ mm, and that diameters are normally distributed. We require that our rods have a mean diameter of no more than 30 mm. We choose a sample from the last production run and find diameters of 30.4, 31, 29.8, 30.6, 30.2, 29.5, 31.1, 29.9, and 30.7 mm. Frame the appropriate hypotheses and use these data to test them at 5% significance. Also state the p value of the test.

36. (N) A consumer group has charged that Belchfire’s inspectors miss an average of at least 10 defects in every car that comes out of the Belchfire auto plant. In an effort to defend themselves, the Belchfire inspection staff select a sample of cars to “post-inspect”; for each of these, the number of previously undetected flaws is recorded. The results are: 8, 6, 10, 8, 11, 8, 10, 9, 10, 11. It appears that the number of undetected flaws is normally distributed. Frame the appropriate hypotheses to reflect the consumer group’s charge and use these data to test them at $\alpha = 0.01$. Also state the p value of the test.

37. (N) A candidate for the US House of Representatives claims that at least 70% of college students are in favor of reviving the military draft. In a sample of 600 students, 390 favor a new military draft. State the appropriate null and alternative hypothesis to use to test the candidate’s claim and use this sample result to test those hypotheses at 1% significance. State the p value of the test and explain its meaning.

38. What is ANOVA used for? Why is this technique preferred to a series of pair-wise hypothesis tests?

39. Why is it appropriate to use the F distribution as the test statistic in an analysis of variance test?

40. What are the two variances used in ANOVA?

41. What are the null and alternative hypotheses being tested with an analysis of variance?

42. The variation between sample means represents the variation in sample values due to __________, while the variation within each treatment sample represents variation due to __________.

The following data are for questions 43 to 46: Participants in a weight loss program at Ramapithecus County Community College were each placed on one of four different diets; their weight loss after six weeks was measured and the results are presented in the table:

<table>
<thead>
<tr>
<th>Diet #</th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
</tr>
</thead>
<tbody>
<tr>
<td>Observation:</td>
<td>1</td>
<td>4</td>
<td>6</td>
<td>2</td>
</tr>
<tr>
<td></td>
<td>2</td>
<td>8</td>
<td>3</td>
<td>1</td>
</tr>
<tr>
<td></td>
<td>3</td>
<td>3</td>
<td>2</td>
<td>12</td>
</tr>
<tr>
<td></td>
<td>4</td>
<td>0</td>
<td>6</td>
<td></td>
</tr>
</tbody>
</table>
43. These data would produce an ANOVA table with an F value which has _______ numerator and _______ denominator degrees of freedom.

44. The Grand Mean is ________.

45. These data will produce the following ANOVA table:

<table>
<thead>
<tr>
<th>Source of Variation:</th>
<th>Sum of Squares</th>
<th>Degrees of Freedom</th>
<th>Mean Squares</th>
</tr>
</thead>
<tbody>
<tr>
<td>Treatment</td>
<td>125.25</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Within Groups</td>
<td>36.75</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Total</td>
<td>162</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Fill in the blanks in the table.

46. If the critical F value for a test at 1% significance is 6.99, interpret the result in terms of the hypothesis test being conducted.

Answers for ANOVA questions:
43. 3, 9
44. 5
45. F=10.22
46. reject H₀; the weight-loss programs are NOT equally effective
Answers to Self Test One:

1. variation
2. see notes/text
3. 
   a. z
   b. t
   c. t
   d. z
4. 17, 0.2
5. 
   a. 0.1587
   b. 0
6. see notes/text
7. unknown; normal
8. 
   a. 0.02165
   b. 0.25
   c. 0.3227
9. point estimate of \( \mu \)
10. see notes/text
11. 0.33
12. 
   a. t
   b. t
   c. t
   d. z
   e. z
13. see notes/text
14. ±0.5
15. 22000±932

Answers to Self Test Two:

1 – 7. see notes/text
8. 
   a. +1.645
   b. ±1.96
   c. −2.33
   d. ±2.58
   e. −1.96
   f. ±1.44
9. 
   a. +1.7396
   b. ±2.1098
   c. −2.5669
   d. ±2.79695
   e. −2.0639
   f. ±1.4871
10. \( H_0: \mu \geq 40; z = -2.4 \)
11. \( H_0: \pi \geq 0.5; z = 0.5477 \)
12. \( H_0: \mu \geq 8 \) vs. \( H_1: \mu < 8; t = 0.7993 \)
13. \( H_0: \mu = 47; t = -3 \)
14. \( H_0: \mu = 600; t = 2 \)
15. 
   a. reject \( H_0: \mu < 40 \)

16. 20%±4.6%
17. 1.46
18. see notes/text
19. 
   a. 48±5.61; 99%:
   b. 48±8.83
   c. 90%: 29.08±4.49; 98%:
   d. 29.08±5.54
20. see notes/text
21. 
   a. nordmist(30,35,10,true)-normdist(20,35,10,true)
   b. norminv(98,100,15)
   c. normsinv(.04) or normsinv(.96)
   d. tinv(.04,23)
   e. stdev(b6:b28)/sqrt(count(b6:b28))
22. see notes/handout
23. see notes/handout; ignore Kurtosis
24. 
   a. 106.2
   b. 545.36, 23.35
   c. 108
   d. 88.5, 125.5
   e. 37
25. 
   a. on a normal distribution, with mean 31 and st.dev.
   b. as a, except probability values greater than 33
   c. on a normal distribution with mean 31 and st. dev.
   d. probability of \( z \) value less than – 2.58
   e. area in two tails of \( t \) distribution with 23 df beyond ±2.33
   f. area in one tail of \( t \) distribution with 23 df beyond 2.33
   g. a \( t \) value such that in a distribution with 23 df .05 is in the tails, or .95 is between ± \( t \)
26. 
   a. 0.840662
decreases
27. increase sample size
28. probability of rejecting a false null
29. decrease
30. far from
31. see notes/text
32. undefined since \( H_0 \) is true
33. see notes/text
34. see notes/text
35. \( H_0: \mu \leq 30; z=2.133>1.645; reject \ H_0; p-value = 0.016 \)
36. \( H_0: \mu \geq 10; t = -1.78> -1.8331; fail to reject; p-value = 0.054 \)
37. \( H_0: \pi \geq 0.7; z= -2.67<-2.33; reject; p-value=0.0038 \)