11.3 Midterm Exam 2: Fall 2000

True/False: 2 points each

1. A histogram of the distribution of $\bar{x}$ always has the same shape as a histogram of the population being sampled.
2. The closer π is to 0 or 1, the larger n must be in order for the distribution of p to be approximately normal.
3. If $\pi = 0.8$, $n = 25$, and N is very large, then the standard deviation of the sampling distribution of p is 0.08.
4. $\bar{x}$, p, and s^2 are point estimators.
5. If the sampling distribution of a statistic has small variability, then the statistic is said to be unbiased.
6. The standard normal distribution is the limit of the t distribution as degrees of freedom go to infinity.
7. The standard error of $\bar{x}$ is s.
8. The confidence level of a confidence interval refers to a property of the method rather than a property of a particular interval.
9. The smallest sample size that would ensure that a population proportion π could be estimated to within $B = 0.1$ with 95% confidence, regardless of the value of π, is approximately 97.
10. The critical value for a 90% confidence interval for μ based on a sample of size $n = 20$ when σ^2 is unknown is 1.645.
11. The two possible decisions in a hypothesis test are accept H_0 and accept H_a.
12. A type II error is made by failing to reject a false H_0.
13. A type I error is made by rejecting a false H_0.
14. Suppose that the p-value for a test of $\mathrm{H}_0\colon \mu_1 - \mu_2 = 0$ against $\mathrm{H}_a\colon \mu_1 - \mu_2 \neq 0$ is 0.00001. It can be concluded that there is an important difference between μ_1 and μ_2.
15. H_0 is rejected if the observed results are unlikely to occur when H_0 is false.
16. A small p-value indicates that the observed data are contradictory to the null hypothesis.
17. If H_0 is not rejected by a statistical test, then there is strong evidence that H_0 is true.
18. The small sample t test for μ should be used only if the population being sampled is approximately normal.
19. If H_0 is false, then power can be increased by increasing sample size or by increasing α.

Multiple choice: 4 points each

20. The central limit theorem says that if you are take a simple random sample from a large population (n/N is small), then
    (a) $\mu_{\bar{x}} = \mu_x$
    (b) $\sigma_{\bar{x}} = \sigma_x/\sqrt{n}$
    (c) The sampling distribution of $\bar{x}$ is approximately normal if n is large.
    (d) p is unbiased for π.
21. Suppose that a 95% confidence interval is computed for μ resulting in the interval $(112.4, 121.6)$. Then
    (a) 95% of the time μ falls in the interval $(112.4, 121.6)$.
    (b) there is a 95% chance that μ falls within the interval $(112.4, 121.6)$.
    (c) 95% of all possible values of μ fall within the interval $(112.4, 121.6)$.
    (d) 95% of all possible samples produce intervals that capture μ.

22. If $\sigma = 10$, then the sample size required to estimate a population mean μ to within 0.5 with 95% confidence is
    (a) 40
    (b) 119
    (c) 1257
    (d) 1537

23. Which of the following does not influence the width of a large sample confidence interval for μ?
    (a) $\bar{x}$
    (b) σ
    (c) the confidence level
    (d) the sample size

24. If $z = 1.68$ for a test of $H_0\colon \pi = 0.6$ against $H_a\colon \pi \neq 0.6$, then the p-value is
    (a) 0.0930
    (b) 0.0465
    (c) 0.9170
    (d) 0.9535

25. If the goal of a study is to determine whether treatments cause a difference between two population means, then
    (a) the investigator should use very large sample sizes
    (b) the investigator should use a very small α
    (c) the investigator should randomly assign treatments to subjects
    (d) the investigator should perform a one-tailed test

26. An investigator wishes to construct a confidence interval for the difference between two population proportions. If $n_1 = 40$, $n_2 = 60$, $p_1 = 0.10$, and $p_2 = 0.80$, then the standard error of $p_1 - p_2$ is
    (a) it cannot be computed because s_1 and s_2 were not given.
    (b) it cannot be computed because $n_1 p_1$ is too small.
    (c) approximately 0.07012.
    (d) approximately 0.10198.

Short answer and computational questions

27. A study concerning the effects of stream-side logging on stream-bank stability is being performed on the Alsea River in Oregon. Measurements of stream-bank stability are taken before the area is logged and also after the area is logged at a random sample of 2 sites. Let μ_1 be the mean stream-bank stability measurement before logging and μ_2 the mean stream-bank stability measurement after logging. The data are listed below.

Site	Before	After
1	25.3	19.6
2	17.1	12.2

    (a) Construct a 95% confidence interval for $\mu_1 - \mu_2$. (7 pts)
    (b) Interpret your confidence interval. Is there evidence that stream-bank stability differed from before to after logging? If so, how did it differ? (4 pts)
    (c) This test requires that the sampling distribution of $\overline{X}_1 - \overline{X}_2$ be normal. Under what conditions will this be satisfied in this investigation? (3 pts)

11.4 Midterm Exam 2: Spring 2001

True/False: 4 points each

1. The alternative hypothesis is strongly supported if the investigator obtains strong evidence against the null hypothesis.
2. The null hypothesis is strongly supported if the investigator fails to obtain strong evidence against the null.
3. A 95% confidence interval for μ is $(23, 46)$. If the study is repeated a large number of times, then the interval $(23, 46)$ will capture the population mean in 95% of the studies.
4. A one-sided test is always more powerful than a two-sided test.
5. If a population is strongly skewed, then sample means based on random samples from the population will not obey the central limit theorem.
6. If an estimator is applied to a sample, then the resulting value is called an estimate.
7. A parameter is a characteristic of a population and a statistic is a characteristic of a sample.
8. The standard error of $\overline{X}$ is S_x
9. The p-value is the probability that the null hypothesis is true.
10. If $\alpha = 0.01$, then the the probability of a type II error is 0.01.
11. I would like to determine if exam 1 is harder that exam 2 in statistics 401. If (a) the 10 students enrolled this term are considered to be a random sample of possible students; (b) the exam scores are approximately normal; and (c) the variance of exam 1 is equal to the variance of exam 2, then the two-sample pooled t test is appropriate.
12. A 95% confidence interval for $\mu_1 - \mu_2$ is $(11.5, 14.6)$. If the investigator tests $\mathrm{H}_0\colon \mu_1 = \mu_2$ against $\mathrm{H}_a\colon \mu_1 \neq \mu_2$ at level $\alpha = 0.05$, then the null hypothesis will be rejected.
13. If n is decreased, then the power of a test of $\mathrm{H}_0\colon \mu = \mu_0$ against $\mathrm{H}_a\colon \mu \neq \mu_0$ decreases.
14. An investigator has conducted a test of $\mathrm{H}_0\colon \pi_1 - \pi_2 = 0$ against $\mathrm{H}_a\colon \pi_1 - \pi_2 \neq 0$. If the p-value is 0.0001 and $\alpha = 0.05$, then the investigator should reject the null and conclude that a large difference between π_1 and π_2 has been detected.
15. Suppose that $\overline{X}_1$ and $\overline{X}_2$ are the sample means from independent samples from population 1 and population 2, respectively. Then the standard deviation of the sampling distribution of $\overline{X}_1 - \overline{X}_2$ is $\sigma_{\bar{x}_1 - \bar{x}_2} = \frac{\sigma_1}{\sqrt{n_1}} + \frac{\sigma_2}{\sqrt{n_2}}$.

Multiple choice: 5 points each

16. If the test statistic for testing $\mathrm{H}_0\colon \mu = 100$ against $\mathrm{H}_a\colon \mu < 100$ is $Z = 1.96$, then the p-value is
    - (a) 0.05
    - (b) 0.025
    - (c) 0.95
    - (d) 0.975
17. If samples of size $n = 2$ are taken from a binary population with $\pi = .4$, then the sampling distribution of p
    - (a) is normally distributed with mean equal to 0.4 and variance equal to 0.24.
    - (b) is not normally distributed but has mean equal to 0.4 and variance equal to 0.24.
    - (c) is normally distributed with mean equal to 0.4 and variance equal to 0.12.
    - (d) is not normally distributed but has mean equal to 0.4 and variance equal to 0.12.
    - (e) none of the above, because the sample size is too small.

18. Which of the following are valid statistical hypotheses? Choose all that are valid.
    (a) $H_0: \pi > 0.05$ versus $H_a: \pi < 0.05$
    (b) $H_0: \mu_1 - \mu_2 = 10$ versus $H_a: \mu_1 - \mu_2 > 10$
    (c) $H_0: S_x = 1$ versus $H_a: S_x \neq 1$
    (d) $H_0: \pi_1 - \pi_2 = 0$ versus $H_a: \pi_1 - \pi_2 < 0$

Short answer and computational questions

19. An investigator desires to estimate π to within 0.05 with confidence 0.95. The investigator knows that the value of π is no larger than 0.25. What sample size should the investigator use? (7 pts)

20. Consider a population of $N = 500$ numerical values. An investigator is interested in studying the sampling distribution of the sample median based on samples of size $n = 11$. Describe how the investigator could obtain the sampling distribution assuming that she has sufficient resources. (5 pts)

21. A consumer product testing group is interested in the lifetimes of name brand televisions and discount brand televisions. Independent random samples of each type of television were selected resulting in the summary table below.

Type	Sample Size	Sample Mean lifetime	Sample SD
Name Brand	12	7.2 years	1.2
Discount	8	5.9 years	1.5

    The consumer group wants to know if mean lifetimes differ between name brand and discount televisions.

    (a) State the appropriate null and alternative hypotheses. (4 pts)
    (b) Is the pooled two sample t test appropriate? Why or why not? (1 pts)
    (c) Compute the test statistic and the p-value. (4 pts)
    (d) Make a decision and draw conclusions. Use $\alpha = 0.05$. (4 pts)

11.5 Final Exam: Fall 2000

True/False: 3 points each

1. The value of r (correlation coefficient) depends on which variable is labeled x and which is labeled y.
2. The value of r is always between 0 and 1.
3. If $r = 0$, then you can conclude that there is no relationship between x and y.
4. The coefficient of determination, r^2, measures the proportion of variability in y that can be explained by a linear relationship between x and y.
5. In simple linear regression, s_e is a point estimate of the standard deviation of y when x has the value x^*.
6. In the simple linear regression model, b is an unbiased estimator of β.
7. The standard deviation of the sampling distribution of $a + bx^*$ increases as the difference between x^* and $\bar{x}$ increases.
8. In a simple linear regression, the mean of y when x has the value x^* is $\alpha + \beta x^* + \varepsilon$.
9. In a simple linear regression model, the residual for the i^{th} observation is $y_i - (\alpha + \beta x_i)$.
10. Residuals in simple linear regression do not have the same standard deviation because the standard deviation of a residual depends on x.
11. If a simple random sample of size n is obtained and each member of the sample is classified into one of k categories, then the X^2 statistic for testing goodness of fit has $n - 1$ degrees of freedom.