

Final Exam Review Questions

You are responsible for everything that we have covered in the course including lecture, lecture notes, and text readings. The exam is two hours so there will be somewhere around 30-35 multiple choice questions, comprised of a good mix of calculation and concept questions.

As I've mentioned in class: we need information like the sample mean and standard deviation to calculate our t stats; we need to know about sampling distributions to understand hypothesis testing and confidence intervals; we need to know about probabilities to calculate p -values and give proper interpretations. Everything we've done ties together in one way or another. So, you are responsible for everything we have covered in the course. However, there will be more emphasis on material covered since Midterm 2 because you have not yet been tested on that material in an exam setting: confidence intervals, Z -tests, t -tests, etc.

With respect to studying, you may wish to start with the most recent material in that Chapters 6, and 7 build off of Chapters 1-5. So if you are having trouble with the recent material, it may be an indication of some of the important topics covered in previous chapters that you may have had problems with (e.g. standardizing data, looking up probabilities or values from a probability table, computing means or standard deviations, finding the probability of an event for a random variable, distribution of linear sums of random variables, etc.).

What follows are extra questions to help you review as you prepare for the final exam, with more emphasis on material in Chapters 6 and 7 since you have already been given questions on previous material. Further, the final exam is cumulative, so these questions are not fully representative of everything you are responsible for the final exam.

1. Which of the following statements about p -values are FALSE?

- (A) A p -value is a probability.
- (B) p -values are greater than 1 when the null hypothesis is true.
- (C) p -values are greater than 1 when the null hypothesis is false.
- (D) We reject the null hypothesis when the p -value is less than the significance level.
- (E) A p -value depends on statistics from a sample, and the null hypothesis.
- (F) A p -value is less than the significance level whenever the null hypothesis is false.

The next two questions refer to the following information. A researcher drew a random sample from a normally distributed population, and calculated a sample mean of 61.1. The population variance is known to be 13.9.

2. The researcher finds that a 95% confidence interval for the population mean is (60.36926, 61.83074). What was the sample size?
3. Assume the sample size was $n = 100$. Calculate an 85% confidence interval for the population mean.

The next two questions refer to the following information. A study investigated the monthly rent of off-campus University of Guelph students. Sixty-four students were sampled, and the sample mean and standard deviation were calculated to be \$360 and \$75, respectively. We wish to test the hypothesis that the mean rent of all University of Guelph students is \$400, against the alternative that it is different from \$400.

4. What is the p -value of the appropriate test?
5. What is an appropriate conclusion for the above test?

The next four questions refer to the following information. The reaction times of 4 students are measured before and after having several alcoholic beverages, with the following results.

Student	Before	After
1	.41	.78
2	.38	.63
3	.52	.94
4	.41	.47

Assuming the population of differences is normally distributed, we wish to test the null hypothesis that alcohol has no effect on reaction times, against the alternative that it slows reaction times.

6. What is the value of the test statistic?
 7. What is the p -value?
 8. What is an appropriate conclusion?
 9. What assumptions are necessary for the above test to be valid?
10. A random sample of 10 fourth year male students were asked how much time they spend on the internet per day. The sample mean and standard deviation were 2.1, and 1.1 hours, respectively. Test the null hypothesis that the true mean time is 1.0 hour, against the alternative it is different from 1.0. What is the p -value of the appropriate test?
11. The difference between a parameter and a statistic is that:
- (A) The value of a parameter will often change from sample to sample, whereas the value of a statistic usually remains fixed.
 - (B) Parameters have standard errors and statistics do not.
 - (C) We use parameters to estimate statistics.
 - (D) Statistics are numbers that are based on a sample, and are used to estimate parameters, which are characteristics of a population.
 - (E) All of the above

12. Which one of the following statements is FALSE?

- (A) If a 2 tailed hypothesis test rejects the null hypothesis at the 5% significance level, it must also reject at the 10% significance level.
- (B) The width of a confidence interval decreases as the sample size increases.
- (C) A 99% confidence interval for μ is narrower than the corresponding 95% interval.
- (D) The t distribution gets closer to the standard normal distribution as the degrees of freedom increase.
- (E) None of the above.

13. Which one of the following statements is TRUE?

- (A) For normally distributed data sets, the median is an unbiased estimator of the population variance.
- (B) If the probability of a success $p = 1$, the mean and variance of the binomial distribution are equal.
- (C) The variance of a data set is 0 whenever the mean equals the median.
- (D) The width of a confidence interval for μ increases as the margin of error increases.
- (E) The width of a confidence interval for μ decreases as the sample mean decreases.

14. Which one of the following statements is TRUE?

- (A) If X is a binomial r.v. with $n = 40$ and $p = .10$, then $P(X = 3) > P(X = 4)$.
- (B) The sample mean is a biased estimator of the population mean when n is small.
- (C) The normal distribution can be strongly skewed, but only for small sample sizes.
- (D) For skewed distributions, the mean is always smaller than the standard deviation.
- (E) None of the above.

15. Which of the following 3 statements are TRUE?

- I. If a 95% confidence interval for μ is (5, 12), we can be 95% confident μ equals either 5 or 12.
- II. We often use parameters to calculate confidence intervals for statistics.
- III. The standard deviation of the sampling distribution of the sample mean increases as the sample size decreases.

- (A) Only I
- (B) Only II
- (C) Only III
- (D) I and III
- (E) II and III

16. To test the null hypothesis $H_0: \mu = 5000$ versus the alternative hypothesis $H_a: \mu < 5000$, you take a random sample of size 200 and obtain a sample mean of 4100 and a sample standard deviation of 9000. Which of the following is the best approximation to the p -value (to two decimal places) for this test?

- (A) 0.03
- (B) 0.08
- (C) 0.16
- (D) 0.84
- (E) 0.92

17. Which of the following statements is/are TRUE?

- (A) The sampling distribution of the sample mean is exactly normally distributed if the population that is sampled from is exactly normally distributed.
- (B) The margin of error for a $100 \times C\%$ confidence interval for a parameter increases as the sample size increases.
- (C) The Central Limit Theorem states that the sample mean is an unbiased estimator of the population mean.
- (D) If the assumptions of a procedure used to calculate a 90% confidence interval are not met, the interval may not contain the parameter of interest 90% of the time in repeated sampling.

18. In a particular large population of people, a researcher wished to test her hypothesis that the mean reaction time for a drug to start working was 10 minutes. Based on a random sample, a 95% confidence interval for the true mean reaction time of (9.2, 13.2) was obtained. Consider the following statements:

- I. The confidence interval provides insufficient evidence to reject the researcher's hypothesis.
- II. The confidence interval provides sufficient evidence to prove the researcher is correct.
- III. The confidence interval suggests that the true mean reaction time may be greater than 10 minutes.

Which of these statements are true?

- (A) I only
- (B) III only
- (C) I and II
- (D) I and III
- (E) II and III.

19. Which one of the following statements best describes the general philosophy behind the hypothesis tests that we have conducted?

- (A) The test is conducted under the initial assumption that the null hypothesis is true, and it is then determined if there is sufficient evidence to reject the null hypothesis and accept the alternative hypothesis.
- (B) The test is conducted under the initial assumption that the alternative hypothesis is true, and it is then determined if there is sufficient evidence to reject the alternative hypothesis and accept the null hypothesis.
- (C) The test is conducted under the initial assumption that the null hypothesis is true, and it is then determined if there is sufficient evidence to support the initial assumption and reject the alternative hypothesis.
- (D) The test is conducted under the initial assumption that the alternative hypothesis is true, and it is then determined if there is sufficient evidence to support the initial assumption and reject the null hypothesis.
- (E) The calculation of the test statistic does not require any initial assumption, it simply supplies evidence to choose between the null and alternative hypotheses.

20. In a hypothesis test, a p-value of 0.08 was obtained for the observed value of the test statistic. From this we can conclude:

- (A) There is little evidence against the null hypothesis, since the p-value is so low.
- (B) The null hypothesis can only be rejected if the level of significance is exactly 8%.
- (C) The null hypothesis should be rejected at the 10% level of significance but not at the 5% level of significance.
- (D) The null hypothesis should be rejected at the 5% level of significance but not at the 10% level of significance.
- (E) We cannot decide, because if a t test was used, we don't know whether it was a one-sided or two-sided test.

21. The population of Canada, although only 1/10 as large as the population of the United States, is nevertheless similar in many ways – including, for example, the proportion who smoke. A random sample of 1000 Americans gave an estimate for this proportion, with an associated standard deviation for the sample proportion. To get an estimate for Canada that has the same sample standard deviation, a random sample of Canadians would have to be:

- (A) 100 times smaller
- (B) 10 times smaller
- (C) of equal size
- (D) 10 times larger
- (E) 100 times larger

22. A researcher calculated that a sample size of 600 would assure that a population mean μ could be estimated within a bound of 4 with 95% confidence, even if $\mu = 5$. If this researcher wanted to lower the bound to 1, how many times larger would the sample size need to be?

- (A) 2
- (B) 4
- (C) 8
- (D) 16
- (E) 64

23. A 95% confidence interval for the mean of a certain population was calculated to be (56.2, 69.4). Consider the following statements:

- I. In repeated sampling, we know that 95% of possible sample means will fall in this interval.
- II. This confidence interval provides evidence to reject the hypothesis that the population mean is 55.
- III. This confidence interval either contains the population mean or it does not.

Which of these statements are TRUE?

- (A) I and II
- (B) I and III
- (C) II and III
- (D) All three of the statements are true
- (E) Only one of the statements is true.

24. For a particular hypothesis test, the null hypothesis is rejected at the 5% level of significance. At the 1% level of significance, which one of the following is true?

- (A) The null hypothesis will definitely be rejected.
- (B) The null hypothesis will definitely be accepted.
- (C) The null hypothesis may or may not be rejected.
- (D) Whether or not the null hypothesis is rejected depends on whether the test is one-sided or two-sided.
- (E) Whether or not the null hypothesis is rejected depends on whether the population standard deviation is known or unknown.

25. Consider the following distributions: binomial, normal, t , uniform. For how many of these distributions is it impossible to obtain a negative value for the random variable?

- (A) 0
- (B) 1
- (C) 2
- (D) 3
- (E) 4

26. The t -table provided is not ideal for calculating a p-value for a t test statistic value. Nonetheless, we can use the t -table to obtain a range of possible values for the p-value. If a two-sided t test was conducted, the t test statistic had 16 degrees of freedom, and a test statistic value of $t_0 = 2.183$ was obtained, the t -table would allow us to say:

- (A) $0.005 < \text{p-value} < 0.01$.
- (B) $0.01 < \text{p-value} < 0.025$.
- (C) $0.02 < \text{p-value} < 0.05$.
- (D) $0.05 < \text{p-value} < 0.10$.
- (E) $0.01 < \text{p-value} < 0.02$.

27. A day's production of automobile engine oil was randomly sampled for quality control to estimate the volume of oil in a bottle. The sample mean was 0.947 L, and the corresponding 95% confidence interval was (0.946, 0.948). Suppose the production process doesn't change, so that the mean volume of oil in a bottle is constant from day to day. What can we say about similar random samples of the same size bottle to be taken over the next 10 years (2400 working days)?

- (A) About 95% of the sample mean will be equal to the population mean.
- (B) About 95% of the sample mean will fall in the interval (0.946, 0.948)
- (C) About 95% of the confidence intervals (similarly constructed) will cover the value 0.947.
- (D) About 95% of the confidence intervals (similarly constructed) will cover the population mean.
- (E) About 95% of the radiators will last at least 10 years without springing a leak.

28. Two groups of rats [1. normal and 2. adrenalectomized] were tested for blood viscosity. The summary statistics are:

	Group	
	1	2
sample size	11	9
sample mean	3.921	4.111
sample s.d.	0.527	0.601

If it was hypothesized, prior to obtaining the data, that adrenalectomized rats would have higher mean blood viscosity, then the appropriate alternative hypothesis for comparing the population mean of normal rats (μ_1) with the population mean of adrenalectomized rats (μ_2) would be:

- (A) $\mu_1 - \mu_2 > 0$
- (B) $\mu_1 - \mu_2 < 0$
- (C) $\mu_1 - \mu_2 \leq 0$
- (D) $\mu_1 - \mu_2 = 0$
- (E) $\mu_1 - \mu_2 \neq 0$

The next two questions refer to the following information.

Hand et al.'s *Handbook of Small Datasets* describes the following data set: "Eight people who each had one eye affected by glaucoma and one not had the corneal thicknesses (in microns) of both eyes measured."

Person	1	2	3	4	5	6	7	8
Affected eye	488	478	480	426	440	410	458	460
Not affected eye	484	478	492	444	436	398	464	476

29. To obtain a 95% confidence interval for the mean difference in corneal thickness between eyes, we would use the following value from the t -table.

- (A) 1.761
- (B) 1.895
- (C) 2.131
- (D) 2.145
- (E) 2.365

30. Suppose it is hypothesized that eyes with glaucoma would have thicker corneas, on average, compared to non-glaucoma eyes. If we calculated thickness differences from a random sample of glaucoma patients as "affected eye minus not affected eye", we would test the following set of null and alternative hypotheses regarding the mean difference, μ_d , in corneal thickness:

- (A) $H_0: \mu_d = 0$ versus $H_a: \mu_d < 0$.
- (B) $H_0: \mu_d = 0$ versus $H_a: \mu_d > 0$.
- (C) $H_0: \mu_d < 0$ versus $H_a: \mu_d = 0$.
- (D) $H_0: \mu_d > 0$ versus $H_a: \mu_d \neq 0$.
- (E) None of the above.

31. A study was conducted in which 6 randomly sampled students were asked to estimate their exam mark prior to writing a test. The researcher's hypothesis was that students would, on average, tend to overestimate their actual mark. After the test, the actual mark was also recorded for each student.

Student	1	2	3	4	5	6
Estimate	72	85	60	68	50	95
Actual	76	81	88	74	53	93

With appropriate assumptions, the results would best be analyzed with:

- (A) a two-tailed t test for two independent samples.
- (B) a one-tailed t test for two independent samples.
- (C) a two-tailed t test for paired data.
- (D) a one-tailed t test for paired data.
- (E) a one-tailed Z test for paired data.

32. A physician compares the blood pressures of four randomly selected patients before and after treatment with a drug. The declines in blood pressures are as follows: 21, 10, 20, 15. Decline in blood pressure can be considered to be a normally distributed random variable. A 95% confidence interval for the mean decline in blood pressure is obtained by calculating the limits:

- (A) 16.5 ± 4.82
- (B) 16.5 ± 7.04
- (C) 16.5 ± 8.06
- (D) 16.5 ± 9.83
- (E) 16.5 ± 10.22

33. Based on a random sample of size 225 and a population standard deviation of $\sigma = 1.61$, a confidence interval for the population mean μ was calculated to have the limits 36.71 ± 0.25 . The confidence level, C , for this interval is:

- (A) 0.80
- (B) 0.90
- (C) 0.95
- (D) 0.98
- (E) 0.99

34. A random sample of size $n = 9$ is taken from a normal population with mean $\mu = 4$ and $\sigma = 15$. What is the probability that the sample mean will be negative?

- (A) almost 0
- (B) 0.2119
- (C) 0.3936
- (D) 0.6064
- (E) 0.7881

35. The width of a 95% confidence interval for a population mean (calculated with known population variance) was 0.042. If the same data set was used, the length of the 90% confidence interval would be:

- (A) 0.031
- (B) 0.035
- (C) 0.037
- (D) 0.039
- (E) impossible to say with the given information.

36. Benthic ecologists can spend a lot of time in frigid waters doing their sampling. A researcher in occupational health was concerned that the amount of time spent in frigid waters may contribute to various health problems later in life. As part of her research, she collected data on the amount of time spent (in hours) during a single field season in frigid waters (temperature below 5°C) by a random sample of 6 Canadian benthic ecologists. The data were: 18, 26, 41, 28, 57, 29

Assuming that the time spent in frigid waters for this season by Canadian benthic ecologists is a normal random variable (probably a flawed assumption, but we will make it anyhow), a 95% confidence interval for the population mean is:

- (A) (18.7, 47.7)
- (B) (19.4, 47.0)
- (C) (21.0, 45.4)
- (D) (21.5, 44.8)
- (E) (22.1, 44.2)

37. A 95% confidence interval for the population mean μ was calculated, based on a random sample of size $n = 8$ from a normal population. The population standard deviation σ was unknown and thus estimated by the sample standard deviation s . The confidence interval had the limits 13.20 ± 1.66 . The value of s used in the calculation of the confidence limits was:

- (A) 1.648
- (B) 1.789
- (C) 1.832
- (D) 1.910
- (E) 1.985

38. In order to estimate the mean diameter of trees (measured 4 feet above the ground) in a particular forest to within 2 inches of the true mean with probability 0.99, a random sample of at least 64 trees must be taken. What was the standard deviation of tree diameters that was used in this sample size calculation?

- (A) 4.8
- (B) 5.5
- (C) 6.2
- (D) 7.4
- (E) 8.2

39. Consider the calculation of a 95% confidence interval for estimating the population mean, based on a random sample from a normal population, when the population standard deviation must be estimated by the sample standard deviation s . Which of the following are valid reasons for why the confidence interval will become narrower (on average) as the sample size n increases?

- I. The t value used will be smaller.
- II. The variance of the sample mean will decrease.
- III. The variance of the population will decrease.

- (A) I and II only
- (B) I and III only
- (C) II and III only
- (D) I, II, and III are all valid reasons.
- (E) Only one of I, II, and III is a valid reason.