

STAT303: Secs 508-510

Fall 2002

Exam #3

**Form A**

Instructor: Julie Hagen Carroll

1. **Don't even open this until you are told to do so.**
2. There are 20 multiple-choice questions on this exam, each worth 5 points. There is partial credit. Please mark your answers **clearly** on the scantron. Multiple marks will be counted wrong.
3. You will have 60 minutes to finish this exam.
4. If you are caught cheating or helping someone to cheat on this exam, you both will receive a grade of **zero** on the exam. You must work alone.
5. This exam is worth 100 points, and will constitute 20% of your final grade.
6. Good luck!

1. An SRS of 100 postal employees found that the average time these employees had worked for the postal service was  $\bar{x} = 7$  years with standard deviation  $s = 2$  years. Assume the distribution of the time the employees have worked for the postal service is approximately normal with mean  $\mu_x$ . Are these data evidence that  $\mu_x$  has changed from the value of 7.5 years of 20 years ago? To determine this we test the hypotheses  $H_0 : \mu_x = 7.5$  vs.  $H_a : \mu_x \neq 7.5$  using the one-sample  $t$ -test. The  $p$ -value is less than 0.01. If we had a sample of only 25 employees, instead of 100, with the same mean and standard deviation, the  $p$ -value would be
  - A. larger.
  - B. smaller.
  - C. unchanged because the sample data is the same.
  - D. unchanged because the hypothesized mean is the same.
  - E. indeterminable without actually recalculating the test statistic.
2. Referring to the previous question, suppose we were not sure if the distribution of the time the employees have worked for the postal service was normal. In which of the following circumstances would we NOT be safe using a  $t$  procedure in this problem?
  - A. The mean and median of the data are nearly but not exactly equal.
  - B. A histogram of the data shows moderate skewness.
  - C. A stemplot of the data has a large outlier.
  - D. The sample standard deviation is large.
  - E. We have to know the distribution is normal since the sample size is less than 30.
3. Ok, suppose that the true mean starting salary for Aggies is \$40,000 with a standard deviation of \$1500 and that of **t-sips** is \$30,000 and \$2000, respectively. What is the distribution of the difference in sample means based on samples of size 100?
  - A.  $N(10,000, 5^2)$
  - B.  $N(10,000, 500^2)$
  - C.  $N(10,000, 250^2)$
  - D.  $N(10,000, 2500^2)$
  - E.  $N(70,000, 500^2)$
4. An SRS of 100 of a certain popular model car in 1993 found that 20 had a certain minor defect in the brakes. An SRS of 400 of this model car in 1994 found that 50 had the minor defect in the brakes. Let  $\pi_1$  and  $\pi_2$  be the proportion of all cars of this model in 1993 and 1994, respectively, that actually contain the defect. A 90% confidence interval for  $\pi_1 - \pi_2$  is  $0.075 \pm 0.071$ . Suppose the sample of 1993 cars consisted of only seven cars, two of which had the minor brake defect. Also suppose the sample of 1994 cars consisted of only six cars, three of which had the minor brake defect. A 90% confidence interval for  $\pi_1 - \pi_2$  is now
  - A. the same as for the original sample of 100 and 400 cars.
  - B. much wider than that for the original sample of 100 and 400 cars.
  - C. the same as a 99% for the original sample of 100 and 400 cars.
  - D. unsafe to compute using the normal distribution to approximate the sampling distribution of  $p_1 - p_2$ .
  - E. This cannot be answered without recalculating the interval.
5. Are avid readers more likely to wear glasses than those who read less frequently? Three hundred men in the Korean army were selected at random and characterized as to whether they wore glasses and whether the amount of reading they did was above average, average, or below average. What type of test should we run to answer the question?
  - A. a  $\chi^2$  test for multiple proportions
  - B. an ANOVA  $F$  test for multiple means
  - C. a 2-sample  $t$  test for the mean of those wearing glasses vs. the mean of those who don't
  - D. a 2-sample  $z$  test for the proportion of those wearing glasses vs. the proportion of those who don't
  - E. a non-parametric test since we don't know if the data is normal or not

6. Which of the following is an example of a Type II error in an ANOVA  $F$  test for the equality of means?
- You conclude that there is a significant effect when there really isn't any.
  - You fail to prove there is a significant effect even though one exists.
  - You claim the variances are all equal, but they're not.
  - You claim you can't run an ANOVA test because the variances are not all equal.
  - You use a 1%  $\alpha$  level when you should have used a 10%.
7. What is the advantage of the *paired t test* over the other 2 sample  $t$  tests?
- It uses less data.
  - It has less variability.
  - It has more degrees of freedom.
  - All of the above are true.
  - Exactly two of the above are true.
8. An SRS of 100 flights of a large airline (call this airline 1) showed that 64 were on time. An SRS of 100 flights of another large airline (call this airline 2) showed that 80 were on time. Let  $\pi_1$  and  $\pi_2$  be the proportion of all flights that are on time for these two airlines. 90, 95 and 99% confidence intervals for the difference  $\pi_1 - \pi_2$  are:  $(-0.259, -0.055)$ ,  $(-0.279, -0.035)$  and  $(-0.317, 0.003)$  respectively. What is the range of the  $P$ -value for testing  $H_0 : \pi_1 = \pi_2$  vs.  $H_a : \pi_1 \neq \pi_2$ ?
- $P\text{-value} > 0.10$
  - $0.10 > P\text{-value} > 0.05$
  - $0.05 > P\text{-value} > 0.01$
  - $P\text{-value} < 0.01$
  - You need a test statistic value to determine the  $P$ -value
9. Suppose we ran 10 hypothesis tests (using 20 different samples of data) on the airline data like in the problem above and found only 2 rejections for the test  $H_0 : \pi_1 = \pi_2$  vs.  $H_a : \pi_1 \neq \pi_2$ . Which of the following is most likely the truth?
- Since we had 2 rejections we are confident that the two airlines don't have the same proportion of on time arrivals.
  - Since we had only 2 rejections, these were Type I errors and there really isn't a difference in the proportions of on time arrivals for the two airlines.

- There is obviously some mistake since we should always get the same conclusion.
  - Ten runs is not enough to tell us if there is a difference in the two airlines' proportion of on time arrivals.
  - If we ran it ten more times, we'd have a total of 4 rejections.
10. Which of the following procedures is not robust to nonnormality?
- the one-sample  $t$  test
  - the  $t$  test for matched pairs
  - the two-sample  $t$  test
  - the  $F$  test for comparing two population standard deviations
  - All of the above are robust procedures concerning the normal assumption.

	female	male	Total
firearm	2559	16381	18940
poison	2110	3569	5679
hanging	803	3824	4627
other	623	1641	2264
Total	6095	25415	31510

11. The two-way table above categorizes suicides committed in 1983 by the sex of the victim and the method used. Which of the following statements is consistent with the table?
- There is absolutely no evidence of a relation between the sex of the victim and the method of suicide used.
  - Men display a greater tendency to use firearms to commit suicide than do women.
  - The correlation between method of suicide and sex of the victim is clearly positive.
  - The distributions for men and women appears to be the same.
  - Exactly two of the above are true.
12. To run a  $\chi^2$  test on the data above, you must first calculate the expected counts. These counts are based on what assumption?
- The data is normally distributed.
  - The true variances for men and women are equal.
  - The sample size is at least 30 for each cell.
  - The true variances for men and women are known.
  - The probability of being in a particular row is independent of which column you are in.

13. Some researchers have conjectured that stem-pitting disease in peach tree seedlings might be controlled with weed and soil treatment. If we wanted to create an experiment to compare peach tree seedling growth for soil and weeds treated with one of two herbicides and then compare the height of the seedlings at the end of the study period, how should we do it? Pick the best answer even if others would work.
- Take a random sample of at least 30 seedlings, give them one herbicide for the first half of the study and then change to the other herbicide for the second half of the study.
  - Take 2 random samples of at least 30 seedlings, give one of them one herbicide and the other the second herbicide.
  - Divide the plot of land into at least 30 subplots and plant 2 seedlings per subplot randomly assigning a herbicide to one of the two seedlings and then the second herbicide to the other.
  - Divide the plot of land into 2 sections and randomly assign one herbicide to one side and then the second herbicide to the other side.
  - Find two separate plots of land and randomly assign one herbicide to one and then the second herbicide to the other.
14. Referring to the previous question, suppose we wished to determine if there tended to be a difference in height for the seedlings treated with the different herbicides. We decide to test the hypotheses  $H_0 : \mu_2 - \mu_1 = 0$  vs.  $H_a : \mu_2 - \mu_1 \neq 0$ . If we have a 90% confidence interval for the difference of  $14.6 \pm 7.80$ ,
- we would not reject the null hypothesis of no difference at the 0.10 level.
  - we would reject the null hypothesis of no difference at the 0.10 level.
  - we would reject the null hypothesis of no difference at the 0.05 and 0.10 level.
  - we would reject the alternative and say there no difference at the 0.10 level.
  - we would not be able to make a decision since we don't have a  $p$ -value.

		GAMES		
		Regular	WSeries	Total
Hit	Count	2584	35	2619
	Expected	2593.2	25.8	2619.0
NoHit	Count	7280	63	7343
	Expected	7270.8	72.2	7343.0
Total	Count	9864	98	9962
	Expected	9864.0	98.0	9962.0

## Chi-Square Test

Value	df	Asymp. Sig. (2-sided)
4.536	1	.033

15. Reggie Jackson had a reputation for hitting better in the World Series than during the regular season. The table above shows his attempts and hits for his 21-year career. Do you believe that Reggie did perform better in the World Series?
- Yes, he had 35 hits in World Series when he should have only had 25 or 26.
  - Sort of. There is evidence of a relationship between the type of game and the number of hits at the 5% level, but not at the 1% level.
  - The **Expected Counts** don't all agree with the **Counts**, so he did do better in the World Series.
  - The **Expected Counts** don't all agree with the **Counts**, so he did NOT do better in the World Series.
  - It's only baseball, so who cares.

The following is for the next three questions:

Wild horse populations on federal lands have been protected since 1971. Since that time, the populations have grown large and need to be managed and kept to a supportable size. Management of the mustang population has been a controversial issue; one common method is periodic removal of the horses. Researchers were curious if a new method would work better. In 1985, 12 bands of horses were rounded up and male horses in each band treated. The number of foals in each band for three years was recorded. Year 1 was prior to treatment, year 2 was the year the treatment was applied, and year 3 one year after treatment. The mean number of foals per band along with the standard deviations are given below.

Year	Means	Std. Devs
1	15.25	7.10
2	15.64	14.14
3	15.25	14.03

The researchers did an ANOVA  $F$  test of the data and obtained the following results.

Source	Sums of Squares	df	Mean Square	F-ratio
Year	1.36	2	0.68	0.0045
Error	5017.65	33	152.05	
Total	5019.01			

16. For this example, which of the following conclusions is most reasonable?
- There is moderate evidence that the treatment is effective in reducing herd size for about one year, but then the effect appears to wear off.
  - An ANOVA  $F$  test is not appropriate for these data. Instead, the researchers should have done several tests to see if the proportion of successes differed for the three years. This analysis would have shown the treatment was effective.
  - The data provide strong evidence that the mean number of foals for the populations represented by the three years differ.
  - The data appear to provide little or no evidence that the treatment is effective in reducing herd size.
  - Without the  $P$ -value, it's impossible to tell what to conclude.
17. In this example, we notice
- the assumption that the data are independent for the three years is unreasonable because the same herds were observed each year.
  - the data show very strong evidence of a violation of the assumption that the three populations have the same standard deviation.
  - there is clear evidence of bias in the results and this is undoubtedly due to the lack of blinding on the part of the subjects.
  - ANOVA cannot be used on these data because the sample sizes are less than 30.
  - More than one of the answers above is correct.
18. Which of the following is/are true?
- If we had wanted to test whether the treatment reduced the number of foals per year, we should have included contrasts for testing  $H_a : \mu_1 > \mu_2$  and  $H_a : \mu_1 > \mu_3$ .
  - We would only use multiple comparisons if there was strong evidence that the treatment had a significant effect.
  - We would need to verify that the data was normal before applying the ANOVA procedure.
  - All of the above are correct.
  - Only two of the above are correct.
19. Did you take any of the Self Tests available on WebCT to help you study for this exam?
- No
  - Yes, some from each chapter.
  - Yes, some from a couple of chapters.
  - Yes, Chapter 7 only.
  - Yes, Ch 9 and 12 only since they're the newest material.
20. There will be an optional early final offered Thursday, Dec. 12th at 2PM in BLOC 165. Will you be taking it or the regularly scheduled one Wednesday, Dec. 18th at 8AM?
- the early one
  - the regular one
- 1A,2C,3C,4D,5A,6B,7B,8C,9B,10D,11B  
12E,13C,14B,15B,16D,17A,18D