STAT302: Secs 102 & 103
Summer I 1997
Final Exam
Form A

Instructor: Julie Hagen Carroll

1. **Don't even open this until you are told to do so.**
2. Be sure to mark your section number (102 or 103) and your test form (A or B) on the scantron!
3. Please PRINT your name at the top and sign your name where indicated on your scantron AND at the bottom of this sheet IF you want your grade posted!!!!!!
4. 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.
5. You will have 60 minutes to finish this exam.
6. 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.
7. This exam is worth 100 points and will replace the lowest of my exams scores, including this one.
8. Good luck!

Final grades will be available by 5PM, TUESDAY, July 8th. The Statistics office WON'T have them, however, so I would like my test score and final grade posted by the last four digits of my Social Security Number. This means it will appear on the Statistics Lab bulletin board no later than Tuesday at 5PM, and I won't call the office asking about it!

Printed Name:

Signature:

SSN:

1. Consider a set of 20 (x, y) pairs. Which of the following is true?

   A. If we add 5 to each of the x's, neither the slope nor the intercept will change.
   B. Any scale or shift change on x or y will cause the intercept to change.
   C. If we multiply the y's by 10, s_y and s_e will increase by a factor of 10, so R^2 will decrease.
   D. r is not affected unless we add a point to the data.
   E. All of the above are true.

2. The reason we use *simple random samples* is because our estimators

   A. will be unbiased.
   B. will have less variability.
   C. will be normally distributed.
   D. will sampled without replacement.
   E. Exactly two of the above are correct.

3. If we tested $H_0 : \mu = 10$ vs. $H_A : \mu \neq 10$ using the 95% confidence interval above, $(8.8, 10.4)$, what would be the appropriate conclusion?

   A. Fail to reject H_0 at the 1 and 5% levels of significance and conclude that there's not enough evidence to conclude that the true mean alcohol content is not 10.
   B. Fail to reject H_0 at the 5% level of significance, but not at the 1% and conclude that there's not enough evidence to conclude that the true mean alcohol content is not 10.
   C. Reject H_0 at the 10% level of significance and conclude that the true mean alcohol content is not 10.
   D. A and C are true.
   E. B and C are true.

4. Which test procedure is preferred when testing if three proportions are equal assuming we can approximate the data with a normal curve?

   A. a t-test for population proportions
   B. a z-test for the difference of two population proportions
   C. a χ^2 test
   D. a non-parametric test procedure
   E. There is no appropriate test procedure.

Source	Partl SS	df	MS	F	Pr > F
A	500.00	1	500.00	42.85	0.0006
B	3600.00	2	1800.00	154.24	0.0000
A*B	600.00	2	300.00	25.71	0.0011
Residual	70.00	6	11.67		
Total	4770.00	11	433.64		

5. What is the appropriate *alternative* hypothesis for testing factor A in the output above?

   A. $H_A : \mu_1 = \mu_2$
   B. $H_A : \mu_1 > \mu_2$
   C. $H_A : \mu_1 \neq \mu_2$
   D. $H_A : \mu_1 < \mu_2$
   E. None of the above are the correct alternative hypothesis.

6. What conclusions may be made from the output above?

   A. Since there is significant interaction, we cannot make conclusions about either main effect.
   B. At the 1% level, the interaction is not significant, but both A and B are.
   C. A, B and the interaction are all significant at the 10% level.
   D. Only factor B is significant at the 1% level.
   E. Exactly two of the above are valid conclusions.

7. What would be the value of the MSE if we hadn't included the interaction term?

   A. It would still be 11.67 since that's the best estimate for σ_e^2.
   B. $670/6 = 111.67$
   C. $670/8 = 83.75$
   D. $600/2 = 300$
   E. There's no way to calculate it by hand.

8. The assumptions necessary for inference (hypothesis testing) concerning the true slope of the simple linear regression line, β_1, are

   A. the e's are linear, *i.e.*, they fall along a straight line.
   B. the e's have a common variance, σ_e^2.
   C. the e's are independent and normally distributed with mean = 0.
   D. All of the above are necessary assumptions.
   E. Exactly two of the above are correct assumptions.

9. Suppose that a random sample of 50 bottles of a particular brand of cough medicine is selected and the alcohol content of each bottle is determined. Let μ denote the true average alcohol content for the population of all bottles of the brand under study. If the sample of 50 results in a 95% confidence interval for μ of $(7.8, 9.4)$, which of the following statements is correct?

   A. There is a 95% chance that the true mean alcohol content, μ, of this brand falls between 7.8 and 9.4.
   B. The estimate, $\overline{x}$, is within 95% of the true mean, μ.
   C. If the process of selecting a sample of size 50 and then computing the corresponding 95% confidence interval is repeated many times, approximately 95% of the resulting intervals will include μ.
   D. If the process of selecting a sample of size 50 and then computing the corresponding 95% confidence interval is repeated many times, approximately 95% of the resulting intervals will fall within $(7.8, 9.4)$.
   E. Both C. and D. are correct.

10. The p-value of a hypothesis test is

   A. the probability that H_0 is true, given the data.
   B. the probability that H_0 is not true, given the data.
   C. the probability of committing a Type I error, given the data.
   D. the probability of committing a Type II error, given the data.
   E. the power of the test.

11. Which of the following statements is NOT!!!! true?

   A. Neither Uniform nor Normal populations can have outliers.
   B. Since they can never be negative, we use distributions that are skewed right, (positively),*i.e.* the F and χ^2, to test hypotheses about variances.
   C. Discrete distributions have probabilities only at exact points; therefore, we can never use the Normal distribution to approximate them since it is continuous.
   D. All of the above are false statements.
   E. Exactly two of the above are false (excluding D.).

12. Increasing the sample size n,

   A. causes $\overline{x}$ to vary less about μ.
   B. causes the sample regression line to vary less about the population regression line.
   C. makes it easier to determine if there is a difference in means.
   D. All of the above are true.
   E. Exactly two of the above are true (excluding D.)

13. The reason the normal distribution is so important to Statistics is because

   A. all sample means, $\overline{X}$'s, can be modelled by the normal distribution.
   B. it is the only continuous distribution for which we can find probabilities.
   C. most sample means can be modelled by the normal distribution if n is large enough.
   D. All of the above are true.
   E. None of the above are true.

14. If $X \sim N(6, 3^2)$ and $Y \sim N(10, 4^2)$ and X and Y are independent, what are μ_{X-Y} and σ_{X-Y}?

   A. $\mu_{X-Y} = -4$ and $\sigma_{X-Y} = -1$
   B. $\mu_{X-Y} = 4$ and $\sigma_{X-Y} = 1$
   C. $\mu_{X-Y} = -4$ and $\sigma_{X-Y} = 1$
   D. $\mu_{X-Y} = -4$ and $\sigma_{X-Y} = 7$
   E. $\mu_{X-Y} = -4$ and $\sigma_{X-Y} = 5$

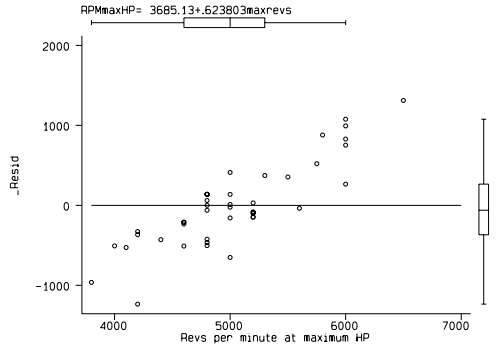
	Percentiles	Smallest
1%	51	51
5%	51	52
10%	51	54
25%	54	53
50%	56	
		Largest
75%	61	58
90%	75	61
95%	75	70
99%	75	75

15. How many outliers are there in the output above?

   A. None
   B. 1
   C. 2
   D. 3
   E. 4

16. Remembering that the *yellow* boxes represent the sample means, in the *Random Sampling* Lab, we saw that

    A. as the sample size n increases, the sample means increase.
    B. as the sample size n increases, the sample means' variance increases.
    C. as the sample size n increases, the sample means decrease.
    D. as the sample size n increases, the sample means look more normal.
    E. Exactly two of the above.



17. What does the RESIDUAL plot above tell us?

    A. The mean of the residuals is not 0.
    B. The data is linear.
    C. The variance of the residuals is not constant.
    D. There is an outlier in the data.
    E. There is another variable we could use to predict the y's.

18. For $X \sim N(10, 5^2)$, what is $P(\overline{X}_{25} < 5)$?

    A. 0
    B. 0.1587
    C. 0.4602
    D. 0.8413
    E. 1

19. Which type of graph *best* determines if data is approximately normal?

    A. if a histogram is bell-shaped
    B. if a residual plot is linear
    C. if a normal quantile plot is linear
    D. All of the above are equally good.
    E. Exactly two of the above (excluding D.)

20. If the grade distribution for this exam is approximately normal, $\overline{x} = 85$ and $s_x = 5$, approximately what grade must you make in order to be in the top 25% of the class? (Pick the closest)

    A. 85 + 0.75*5 = 89 rounded
    B. 85 - 0.675*5 = 82 rounded
    C. 85 + 0.675*5 = 88 rounded
    D. 85 - 0.675 = 84 rounded
    E. 85 + 0.75 = 86 rounded

21. What is the significance of the first Friday in July? (don't worry you'll get 5 points for this anyway)?

    A. It's the day I become independent of Statistics 302 (you can never be totally free from statistics, they're everywhere)
    B. It's America's Independence Day.
    C. It's actually an A&M holiday, which is suprising since we don't get Memorial Day nor Labor Day.
    D. It's the start of a 3-day weekend.
    E. All of the above.

Answers: 1. D 2. A 3. A 4. C 5. C 6. A 7. C 8. E 9. C 10. A 11. E 12. D 13. C 14. E 15. B 16. D 17. E 18. A 19. C 20. C