Name_____

Mathematics 263 Section 6
Final Examination
Spring 2013

Please write your name *legibly* on every page!

Calculators and two note cards are allowed. Tables are supplied.

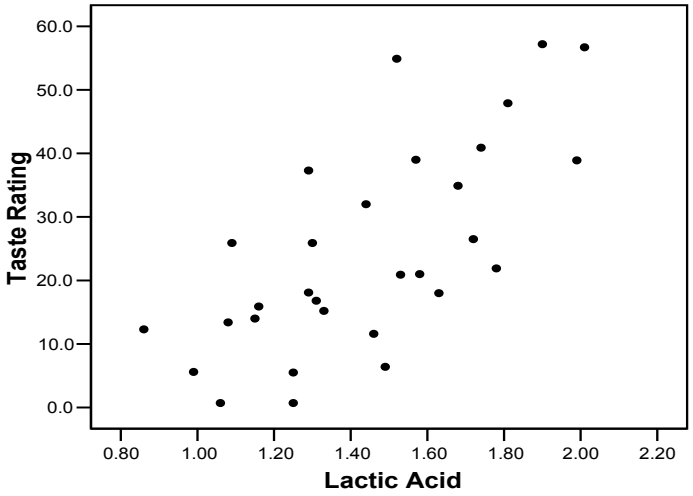
Good luck! Thanks for a great semester and enjoy your summer. Deb

Question	**Points**	**Score**
1-14	28	
15	3	
16	4	
17	11	
18	18	
19	13	
20	12	
21	11	
Total	**100**	

Name_____

In Problems 1-14, circle *one* answer. (2 points each)

1. As Swiss cheese matures, a variety of chemical processes take place. In a study, samples of cheese were analyzed for lactic acid concentration and were subjected to taste tests. A scatterplot of the data is below:



What is a plausible value for the correlation, r, between lactic acid concentration and taste rating?

(a) 0.999
(b) 0.7
(c) 0.07
(d) -0.07
(e) -0.7

2. In the last mayoral election in a city, 47% of the adults over the age of 65 voted Republican. A researcher wants to decide if the proportion of adults over 65 in the city who plan to vote Republican in the next mayoral election has changed. Let p represent the proportion of all adults over 65 in the city who plan to vote Republican in the next mayoral election. Which of the following null and alternate hypotheses should the researcher use?

(a) $H_0: p = 0.47$ versus $H_a: p < 0.47$
(b) $H_0: p = 0.47$ versus $H_a: p \neq 0.47$
(c) $H_0: p = 0.47$ versus $H_a: p > 0.47$
(d) $H_0: p = 0.5$ versus $H_a: p < 0.5$
(e) $H_0: p = 0.5$ versus $H_a: p \neq 0.5$
(f) $H_0: p = 0.5$ versus $H_a: p > 0.5$

3. A study collected data on two quantitative variables, x and y. A scatterplot of this data showed a straight-line relationship and the correlation, r, was -0.92. This tells us that

(a) All of the data values for the two variables lie on a straight line.
(b) There is little reason to believe that the two variables are linearly related.
(c) There is a strong linear relationship between the two variables with larger values of x tending to be associated with larger values of y.
(d) There is a strong linear relationship between x and y with smaller x values tending to be associated with larger values of y.
(e) There is a weak linear relationship between x and y with smaller x values tending to be associated with smaller values of y.

Name_____

4. The length of gestation until birth for pregnant women is a random variable that is normally distributed with a mean $\mu = 282$ days and standard deviation $\sigma = 11$ days. Births with gestation times of 258 days or less are considered to be premature. What is the probability that a randomly selected pregnant woman will give birth to a premature baby?

(a) 0.2182
(b) 0.0146
(c) 0.0018
(d) 0.0183
(e) Not within ± 0.001 of any of the above.

5. Scores on a standardized test are normally distributed, with mean 35 and standard deviation 14. Only 5% of the test takers scored higher than Alejandra. What was her score?

(a) 7
(b) 49
(c) 58
(d) 63
(e) 77

6. Suppose exactly 50% of a population supports a candidate, what is the probability that a random sample of 450 shows 52% support or more?

(a) 0.02
(b) 0.198
(c) 0.48
(d) 0.52
(e) 0.802

7. (Continuation of Problem # 6) Suppose the sample size for this poll was increased, but all the other quantities remained constant. How would the probability calculated in Problem #6 change?

(a) The new probability would be larger than before.
(b) The new probability would be smaller than before.
(c) The new probability would be the same as before.
(d) Need further information to tell.

8. An experiment was conducted to see if a new drug changed the mean time to recovery in comparison to the standard drug. The mean time to recovery for the standard drug is 26 days. Following a randomized experiment involving 65 patients, a 95% confidence interval was constructed for the mean recovery time (in days) for patients on the new drug; it was $(24.6, 27.8)$. Based on this interval, what can you conclude concerning the new drug relative to the standard drug?

(a) At the 0.05 significance level, there is evidence that the new drug is better than the standard drug.
(b) The experimenter should reject the claim that the new drug is the same as the standard drug with respect to mean recovery time.
(c) There is insufficient evidence to reject the claim that there is no difference between the new drug and the standard drug with respect to mean recovery time.
(d) The confidence interval contains the hypothesized value for μ and hence it is significant at the 0.05 level.
(e) There is reason to believe that the standard drug is better than the new drug and hence the new drug should not be prescribed.

Name_____

Questions 9 – 15 use the following information:
In a study of test scores of entering college freshmen, a random sample of colleges across the nation is selected and the average SAT Math score for each of their freshman classes is recorded. The colleges are categorized as Public, Private, or Church. We want to know whether the freshmen entering the three types of colleges do equally well on the SAT Math test. ANOVA computer output is shown below:

Source	Sum of squares	DF	Mean square	F	p-value
Between Groups	63906.2	2		5.696	0.005
Within Groups (Error)	353440.2	63	5610.2		
Total	417346.4	65			

9. How many colleges were included in the study?

(a) 3
(b) 63
(c) 64
(d) 65
(e) 66

10. What is the value of the Mean Squares for Between Groups?

(a) 5,610.2
(b) 31,953.1
(c) 63,906.2
(d) 127,812.4
(e) 353,440.2

11. Under the null hypothesis of equal population means in the different types of colleges, what is the distribution of the test statistic? [Note: $F(a,b)$ is the F-distribution with $df = a, b$, and $N(a,b)$ is the normal distribution with $\mu = a$ and $\sigma = b$, and $T(k)$ is the T-distribution with $df = k$.]

(a) $F(2,63)$
(b) $F(2,65)$
(c) $N(0,3)$
(d) $N(2,63)$
(e) $T(63)$

12. The value of the F-statistic in the ANOVA table is 5.696 and the p-value is 0.005. If we draw the F-distribution and mark the value of 5.696 on the x-axis, how do we indicate the p-value on the graph?

(a) The area under the curve to the left of 5.696.
(b) The area under the curve to the right of 5.696.
(c) The area under the curve between -5.696 and 5.696.
(d) The area under the curve to left of -5.696 together with the area to the right of 5.696.
(e) Twice the area under the curve to the right of 5.696.

13. At a significance level of 0.05, what is the conclusion about the average SAT Math scores?

(a) The average SAT Math scores for freshmen attending colleges with the three different affiliations appear to be the same.
(b) Each of the three average SAT Math scores for freshmen attending colleges with the three different affiliations appear to be different.
(c) It appears that freshmen attending at least one of the three different types of college have a different average SAT Math score.
(d) Freshmen at one type of affiliated college have a significantly better average SAT Math score than the other two.
(e) Two of the types of institutions have the same average SAT scores; the third is different.

Name_____

Problems 14 -15 refer to the ANOVA table on previous page

14. Which of the following statements about the SAT Math ANOVA table is (are) true?

(a) Sums of squares represent variation present in the data and they are calculated by summing squared deviations.
(b) There are three distinct sources of variation represented in the table: between groups, within groups (error), and total
(c) The sum of squares for total (SST) is composed of two parts, one due to groups (SSG) and one due to error (SSE).
(d) Mean squares are calculated by dividing the corresponding sum of squares by its degrees of freedom.
(e) All of the above are true.

15. (**3 points**) One of the assumptions in ANOVA is that the population standard deviations are equal. Mark each of the following statement as True (T) or False (F).

_____We could use side-by-side boxplots to assess if this assumption of equal population standard deviations seems reasonable.

_____As long as the ratio of the largest to the smallest sample standard deviation is greater than 2, then the assumption seems to be satisfied.

_____In the SAT study, an estimate for the common standard deviation s in the three populations equals 74.90.

16. (**4 points**) Which of the following are correct interpretations of the p-value? Mark each one True (**T**) or False (**F**).

_____The probability that the null hypothesis is false

_____The probability that the alternate hypothesis is true

_____The probability that, if the null hypothesis is true, we obtain by chance sample data is as extreme as we got or more so.

_____The probability that, if the alternate hypothesis is true, we obtain by chance sample data is as extreme as we got or more so.

Name_____

17. (**11 points**) A study of men who had had vasectomies (a contraceptive surgery) found 113 cases of prostate cancer in 22,000 men who had had vasectomies compared to the baseline rate of 70 cases per 22,000 men who have not had vasectomies. [1]

(a) Use this data to decide if there is a significant difference in the prostate cancer rate in the men who had had vasectomies and those who had not:

(i) State the null hypothesis

(ii) State the alternate hypothesis

(iii) Find the test statistic.

(iv) What can you say about the p-value?

(v) What is your conclusion about vasectomies?

(b) Can the differences in cancer rates be ascribed to the vasectomies?

Yes_____ No_____ (*check one*)

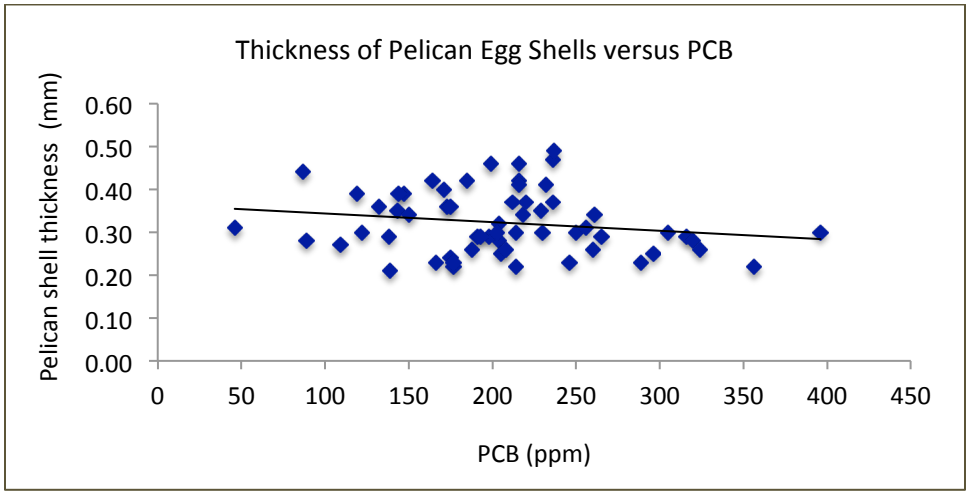
Reason:

[1] "A Retrospective Cohort Study of Vasectomies and Prostate Cancer in US Men" by E. Giovannucci, *JAMA*, 1993. Reported in Intro Statistics, 9th ed, N. Weiss (Pearson 2012).

Name_____

18. (**18 points**) Brown pelicans, the state bird of Louisiana, live in coastal waters and feed on fish. Since 1961, no pelicans have bred in Louisiana. Studies in other states suggested that the birds were failing to breed because their eggs had such thin shells that they broke before hatching. The cause was thought to be eating polluted fish, so data was gathered to investigate the relationship between chemicals in the eggs and the thickness of their shells.[2] The regression output and scatter plot of shell thickness, in mm, against concentration of polychlorinated biphenyl (PCB), in parts per million (ppm), is shown below:

	Coefficients	*Standard Error*	*t Stat*	*p-value*	*Lower 95%*	*Upper 95%*
Intercept	0.3639	0.0304	11.9627	0.0000	0.3030	0.4248
PCB (ppm)	−0.0002	0.0001	−1.4350	0.1567	−0.0005	0.0001



(a) What is the equation of the regression line shown? If you use x and y, please define them.

(b) What is the predicted shell thickness for an egg containing 300 ppm PCB? Show your calculation and give units with your answer.

(c) What are the units of the intercept, 0.3639?

(d) Interpret the intercept in terms of pelicans' eggs.

Continued on next page—->

[2] "Effects of Environmental Pollution Upon Animals Other than Man", by R. W. Risebrough. Reported in *Intro Statistics*, 9th ed, ny N. Weiss, (Pearson, 2012).

Name_____

Problem 18 continued

(e) What are the units of the coefficient -0.0002?

(f) Interpret the coefficient, -0.0002, in terms of pelicans' eggs. Include an interpretation of the sign.

(g) Draw a conclusion about the slope of the population regression line using a hypothesis test:
(i) What is the null hypothesis?

(ii) What is the alternative hypothesis?

(iii) What is the test statistic?

(iv) What is the p-value?

(v) What is your conclusion about the slope?

(h) Can you conclude that PCB has a significant effect on the thickness of pelican eggs shells?

Yes ____ No_____ (*check one*)

Reason:

Name_____

19. (**13 points)** The Rosement Copper Company has applied to begin open-pit copper mining in the Coronado National Forrest. The mine would bring economic benefits to Tucson, but also have an environmental impact. An April 2012 poll of 625 randomly selected residents of Pima, Cochise, and Santa Cruz counties reported that 287 of the 625 people were opposed to the mine.[3]

(a) What is the 95% margin of error of this poll?

(b) Find the 95% confidence interval for the proportion opposed to the mine.

(c) Mark each of the following statements about the confidence interval as True (**T**) or False (**F**):

____ 95% of the sample means of all samples of size 625 lie in this interval; 5% do not.

____There's a 95% chance that the number of people opposing the mine in a sample of 625 lies in this interval

____95% of the intervals generated by this method contain the proportion of the population opposing the mine; 5% do not

____95% of the population proportions opposing the mine lie in this interval; 5% do not

____There's a 95% chance that the proportion of the population opposing the mine lies in this interval.

(d) The 625 respondents were then given a detailed description of the costs and benefits of the mine. After this, 344 out of the 625 were opposed. Find the new 95% confidence interval for the proportion opposed to the mine.

(e) Does this new confidence interval provide evidence that, when given detailed information, a majority of voters opposed the mine?

Yes _____ No_____ (*check one*)

Reason:

[3] Fairbank, Maslin, Maullin, Metz and Baselice "Summary of Rosemont Copper Mine Survey Findings" April 26, 2012.

Name_____

20. (**12 points**) The mean systolic blood pressure level in a population is 130 mm, with standard deviation 20 mm.

(a) If you take a single random sample of 200 people, what is the
(i) Approximate mean systolic blood pressure of the sample?

(ii) Approximate standard deviation of the sample?

(b) Consider the three sampling distributions for mean systolic blood pressure of samples of size 20, 50, and 200.
(i) What are the means of the three distributions? (*Answer all three*)

Mean for $n = 20$ is _____

Mean for $n = 50$ is _____

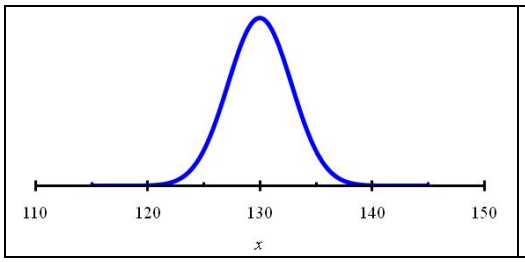
Mean for $n = 200$ is _____

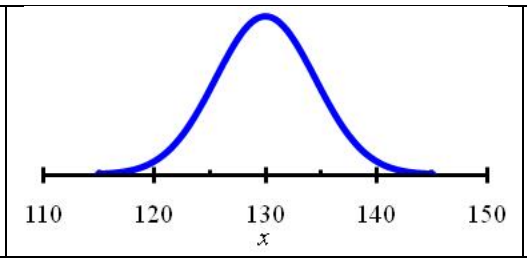
(ii) Which of the three distributions has the smallest standard deviation? (*Check one*)

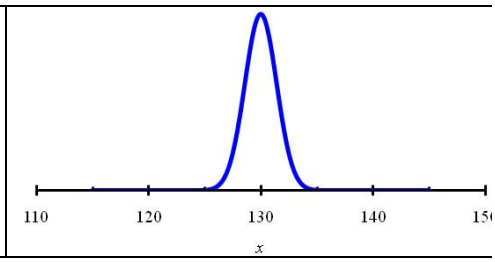
SD is smallest for $n = 20$_____ or $n = 50$_____ or $n = 200$_____

(iii) What is the standard deviation of the sampling distribution for $n = 200$? Include units in your answer.

(iv) Match the three systolic blood pressure sampling distributions with their graphs:







Sample size: _____ Sample size: _____ Sample size: _____

Name_____

21. (**11 points**) In the western US, some wheat farms depend on the winter snow and spring rains for their water. In bad years, when there is too little rain, the crop is poor and farmers suffer. The table shows the probability of having a given number of good years in a decade with enough rain for a good crop.[4]

Number of years with good crop	0	1	2	3	4	5	6	7	8	9	10
Probability	0.01	0.02	0.03	0.07	0.09	0.12	0.22	0.25	0.13	0.05	0.01

(a) What is the probability that there are eight or more good years in a decade?

(b) What is the probability that there is at least one bad year in a decade?

(c) For a decade, which is larger: (check one)

- The unconditional probability: $P(8 \text{ or more good years})$ _____
- The conditional probability $P(8 \text{ or more good years} | 4 \text{ or more good years})$_____

Give a reason for your choice that does not involve calculation:

(d) Find $P(8 \text{ or more good years} | 4 \text{ or more good years})$? Show your calculation.

[4] Based on *Agricultural Statistics*, USDA, as reported in *Understandable Statistics*, by Brase and Brase, 10th ed.