Review for Final: Math 263 Section 6

1. The length of the western rattlesnake is normally distributed with mean 42 inches and standard deviation 2 inches. Consider the sampling distribution of mean lengths of samples of 15 western rattlesnakes.
   (a) What is the mean of the sampling distribution?
   (b) What is the standard deviation of the sampling distribution?
   (c) What is the probability that the mean length of a random sample of 15 western rattlesnakes is more than 42 inches?
   (d) What is the effect of increasing the sample size on the mean and standard deviation of the sampling distribution? (check one in each column)

Mean	Standard deviation
Increases__________	Increases__________
Decreases_________	Decreases_________
Unchanged________	Unchanged________
Need further information_______	Need further information_______

2. Women in labor are given injections to deaden the pain. A hospital[1] wants to determine whether the injections are more effective if they are given when the woman is lying down or when she is sitting up. They randomly assigned women to two groups and measured the number of segments of the spine that were anesthetized in each group (larger numbers mean the anesthetic is more effective). The following table gives results from the two samples.

	Sample size	Average	Standard deviation
Lying	48	8.8	4.4
Sitting	35	7.1	4.5

   Using the following steps and a 5% significance level, decide if the injections work better for women in one position or the other.

   (a) What is the null hypothesis?

   (b) What is the alternative hypothesis?

   (c) Calculate the statistic.

   (d) What is the distribution of the statistic?

   (e) What is the *P*-value of the statistic? What is its interpretation?

   (f) What is your conclusion?

3. A psychologist is studying the effect of drug and electroshock therapy on a subject's ability to solve simple tasks. The number of tasks completed in a 10-minute period is recorded for subjects who have been randomly assigned to compare four treatment groups: drug with electroshock, drug without electroshock, no drug with electroshock, and no drug/no electroshock. Sixty-four subjects were randomly selected students from a large local university.
   (a) State the null and alternative hypotheses for the researchers.
   (b) What can the psychologist conclude from the ANOVA output below? Explain how your conclusion follows from this output (assuming the technical conditions are met).

[1] National Women's Hospital in Auckland, NZ. Reported by C. Wild and G. Seber in *Chance Encounters*.

```
Analysis of Variance
Source     DF        SS        MS        F        P
Factor      3     58.13     19.38    11.01    0.000
Error      60    105.63      1.76
Total      63    163.75
```

4. An insurance company writes policies for a large number of newly-licensed drivers each year. Suppose 30% of these are low-risk drivers, 50% are moderate risk, and 20% are high risk. The company has no way to know which group any individual driver falls in when it writes the policies. None of the low-risk drivers will have an at-fault accident in the next year, but 10% of the moderate-risk and 20% of the high-risk drivers will have such an accident.

   (a) Fill in the following table for the number of accidents among 100 newly-licensed drivers in the next year.

	At-fault accident	**No at-fault accident**	**Total**
Low risk			
Moderate risk			
High risk			
Total			100

   (b) If a driver has an at-fault accident in the next year, what is the probability that he or she is high-risk?

   (c) Are having an at-fault accident and risk status independent? Why?

5. In March 2004, a California father, Dr. Michael Newdow, argued in front of the Supreme Court that the phrase "under God" should be removed from the Pledge of Allegiance because it violates the constitutional principle of separation of church and state. A June 2003 poll of 1000 adults reported that 26% believed that the phrase should be removed.[2]

   (a) Find the 90% confidence interval for the percentage of adults who agree with Dr. Newdow. (Give your answer in percentages to one decimal place.)

   (b) What sample size is needed for a margin of error of 3% at the 95% confidence level? (Give your answer as a whole number.)

[2] *USA Today*, March 25, 2004, page 3A.

6. A simple random sample of 15 apparently healthy children between the ages of 6 months and 15 years yield data on age and liver volume per unit of body weight (ml/kg).

SUMMARY OUTPUT

Regression Statistics	
Multiple R	0.788504
R Square	0.621738
Adjusted R Square	0.592641
Standard Error	5.866846
Observations	15

ANOVA

	df	*SS*	*MS*	*F*	*Significance F*
Regression	1	735.4748	735.4748	21.36773	0.000478
Residual	13	447.4585	34.41989		
Total	14	1182.933			

	Coefficients	*Standard Error*	*t Stat*	*P-value*	*Lower 95%*	*Upper 95%*
Intercept	48.53964	3.094431	15.68613	7.94E-10	41.85453	55.22476
Age	-1.5762	0.340982	-4.62252	0.000478	-2.31284	-0.83955

(a) Is age useful in predicting liver volume? Explain why or why not.
(b) What are the estimates of the slope and intercept for this model?
(c) How do you interpret them?
(d) How would you assess the fit of the model? Explain.

7. The oxygen demand in a lake or wetland is the concentration of oxygen needed for survival by the wildlife living in the area. An EPA (Environmental Protection Agency) report[3] on a wetland in Marin County, California, noted that the mean oxygen demand was 9.9 mg/liter and that 95% of the readings were between 6.5 mg/liter and 13.3 mg/liter. Assume that oxygen demand is normally distributed.
(a) Find the standard deviation of the oxygen demand. Give units with your answer.
(b) An oxygen demand of less than 8 mg/liter suggests that some wildlife may be dying. What is the probability that such a reading is observed?
(c) An oxygen demand of over 12 mg/liter suggests there may be too many organisms that endanger some types of plant life. What is the probability of such a reading?

8. Are the following statements true or false? Explain the reasons for your choice. You are given that

- The distribution of personal incomes in the US is skewed to the right.
- The weights of healthy newborn babies are normally distributed with mean 3.43 kg and standard deviation 0.48 kg.

(a) The distribution of mean incomes of samples of 1000 people is skewed to the right.

(b) A particular random sample of 100 newborns is likely to have mean near 3.43 kg and standard deviation near 0.48 kg.

(c) The sampling distribution of the means of 100 newborns has mean 3.43 kg and standard deviation 0.48 kg.

[3]EPA Report 832-R-93-005, used in *Understandable Statistics* 10th ed, by Brase and Brase.

Solutions to Review

1.

(a) Mean is 42 inches.

(b) Standard deviation is $\frac{2}{\sqrt{15}} = 0.52$ inches.

(c) Since the mean of the sampling distribution is 42 inches, the probability is 50%.

(d) Mean is unchanged; standard deviation decreases.

2.

(a) The null hypothesis is $H_0: \mu_1 = \mu_2$.

(b) The alternative hypothesis is $H_a: \mu_1 \neq \mu_2$.

(c) The statistic we use is

$$t = \frac{(8.8 - 7.1) - (\mu_1 - \mu_2)}{\sqrt{\frac{4.4^2}{48} + \frac{4.5^2}{35}}} = 1.716$$

(d) This statistic has the *t*-distribution with 34 degrees of freedom.

(e) Using the tables with *df* = 30, the *P*-value is given by

$$P(|T| > 1.716) = 2P(T > 1.716) \approx 2(0.05) = 0.10.$$

From the calculator, with *df* = 34, the *P*-value is given by

$$P(|T| > 1.716) = 2\text{tcdf}(1.716{,}100{,}34) = 0.095.$$

Thus there is a 9.5% chance that the results of the two samples would differ by as much as this if the two methods were equally effective.

(f) We do not reject the null hypothesis. Since the *P*-value is more than 5%, we do not have evidence that the two methods (lying and sitting) differ significantly in their effectiveness.

3.

(a) The null hypothesis is that the mean number of tasks completed in the 10-minute period si the same for all four groups.

The alternate hypothesis is that all four means are not equal; that is, at least one is different.

(b) Since the *P*-value is small (0 to three decimal places), we reject the null hypothesis. This means that the at least one of the four groups completes a significantly different number of tasks.

We can conclude this because the *P*-value tells us that if the null hypothesis is true (that is, all means are the same), there is essentially no chance that we see the difference in means that was observed. Given this, it is more likely that the means are *not* all the same.

4.

(a)

	At-fault accident	No at-fault accident	Total
Low risk	0	30	30
Moderate risk	5	45	50
High risk	4	16	20
Total	9	91	100

(b) We want the proportion of people that had at-fault accidents that were high risk:

$$P(\text{High}|\text{At Fault}) = \frac{4}{9} = 44.4\%$$

(c) Not independent, since for example

$$P(\text{High}|\text{At Fault}) = \frac{4}{9} = 44.4\%$$

$$P(\text{High}) = 20\%$$

5. (a) Since $\hat{p} = 26\% = 0.26$, we have $SE_{\hat{p}} = \sqrt{\frac{0.26(1-0.26)}{1000}} = 0.0138$, so

Thus the confidence interval is
$(0.26 - 1.64 \cdot 0.0138,\ 0.26 + 1.64 \cdot 0.0138\ ) = (0.2374, 0.2826) = (23.74\%, 28.26\%)$.

(c) We have

$$0.03 = 1.96\sqrt{\frac{0.26(1-0.26)}{n}}$$

so

$$n = \left(\frac{1.96}{0.03}\right)^2 \cdot 0.26(1-0.26) = 821.2 = 822.$$

Thus, a sample of 822 or larger is needed. (Remember to round up.)

6.

(a) Yes; P-value of the coefficient is small (0.000478), so there is a significant relationship.

(b) Liver volume $= 48.54 - 1.57 \cdot$ Age.

(c) Slope is the average decreases in the proportion of liver volume for each additional year of age. The intercept is the proportion of liver volume in a child of age 0 (that is, a newborn).

(d) Since $R^2 = 0.62$, we know that 62% of the variation in liver volume is predicted by age. This is a substantial proportion.

7. The variable is quantitative (oxygen demand) and normally distributed. We are looking at the results of one sample.
   (a) For a normal distribution, 95% of the data is within 1.96 standard deviations from the mean. (Or approximately 2 standard deviations on either side of the mean.) Thus, the distance from 9.9 to 13.3, which is $13.3 - 9.9 = 3.4$ mg/liter, is 1.96 standard deviations, so
   $$1.96\sigma = 3.4$$
   $$\sigma = \frac{3.4}{1.96} = 1.73 \text{ mg/liter.}$$
   (b) To find the probability, find the z-value and use the table. We have
   $$z = \frac{8 - 9.9}{1.73} = -1.10.$$
   From the table $P(Z < -1.1) = 0.1357 = 13.57\%$.
   (c) To find the probability, find the z-value and use the table. We have
   $$z = \frac{12 - 9.9}{1.73} = 1.21.$$
   From the table $P(Z < 1.21) = 1 - P(Z > 1.21) = 1 - 0.8869 = 0.1131 = 11.31\%$.
   (d) From the table or calculator, the z-value at the 80th percentile is $z = 0.842,$ so
   $$0.842 = \frac{x - 9.9}{1.73}$$
   $$x = 9.9 + 0.842 \cdot 1.73 = 11.34 \text{ mg/liter.}$$

8. Using the information given:
   (a) False. The sampling distribution is close to normal with sample sizes of 30 or more. As the sample size increases, the approximation gets better, so the sampling distribution for samples of size 1000 is essentially normal, and hence not skewed.
   (b) True. A random sample will represent the population and have approximately the same mean and standard deviation.
   (c) False. By the Central Limit Theorem, the sampling distribution of the means has mean 3.43 kg and standard deviation $\frac{0.48}{\sqrt{100}} = 0.048$ kg.