Statistical Methods - Math 222 **Final Exam Review Problems**

These problems are mostly focused on ANOVA, so they are not a comprehensive review of all of the topics on final. Be sure to also review old tests, handouts, and practice problems.

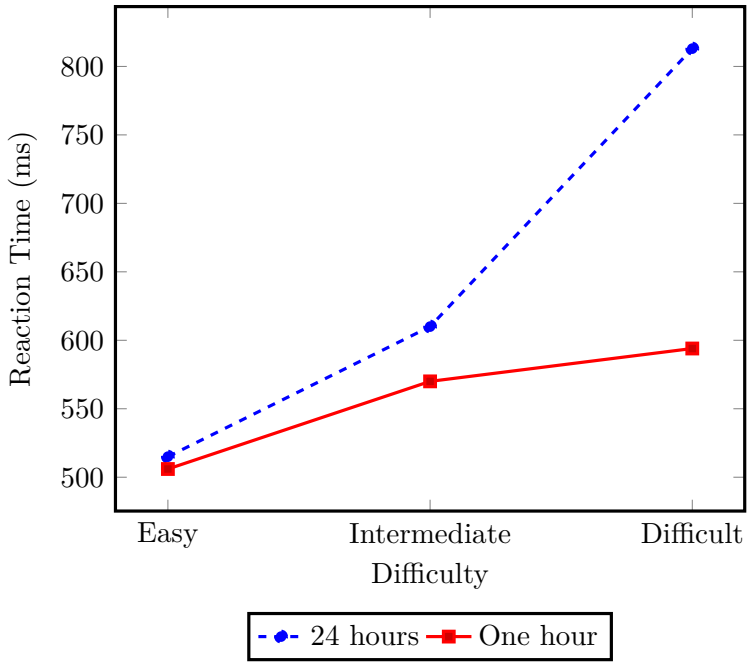
1. Sally gets a cup of coffee and a muffin every day for breakfast from one of the many coffee shops in her neighborhood. She picks a coffee shop each morning at random and independently of previous days. The average price of a cup of coffee is \$1.40 with a standard deviation of \$0.30, the average price of a muffin is \$2.50 with a standard deviation of \$0.15, and the two prices are independent of each other.
   (a) What is the mean, variance, and standard deviation of the amount she spends on breakfast daily?
   (b) What is the mean, variance, and standard deviation of the amount she spends on breakfast weekly (7 days)?
   (c) Which random variable will have a more normal distribution and why: The cost of Sally's breakfast each day or the total cost of Sally's breakfasts for a week?

2. The Behavioral Risk Factor Surveillance System(BRFSS) is an annual telephone survey designed to identify risk factors in the adult population and report emerging health trends. The following table displays the distribution of health status of respondents to this survey (excellent, very good, good, fair, poor) and whether or not they have health insurance.

	Health Status					
	Excellent	Very good	Good	Fair	Poor	Total
Has insurance	0.2099	0.3123	0.2410	0.0817	0.0289	0.8738
No insurance	0.0230	0.0364	0.0427	0.0192	0.0050	0.1262
Total	0.2329	0.3486	0.2838	0.1009	0.0338	1.0000

   (a) What is the probability that a randomly chosen individual has excellent health given that they have health coverage?
   (b) What is the probability that a randomly chosen individual has excellent health given that they don't have health coverage?
   (c) Do having excellent health and having health coverage appear to be independent?
   (d) What additional information would you need in order to do a Chi-square test of association with this data?

3. A study looked at how lack of sleep affects reaction times. Volunteers were randomly assigned to either complete a task one hour after waking up or after 24 hours without sleep. Reaction times were measured (in milliseconds) in a discrimination task. Three levels of task difficulty were used. The results are shown in the interaction plot below.



Use this plot to answer the following questions.

(a) Describe clearly the main effects of each factor in this experiment.

(b) Describe any interaction between the factors.

(c) What should we do to determine if the interaction is statistically significant?

4. Determine whether each statement below is True or False.

(a) In one way ANOVA the response variable is categorical and the explanatory variable is quantitative.

(b) Linear regression assumes that the residuals are normally distributed.

(c) One of the assumptions made in the application of the one-way ANOVA F test is homogeneity of variance (i.e., the variances for all populations are assumed to be the same).

(d) If the data in each group is strongly right skewed, it is okay to do an ANOVA F-test as long as the sample sizes are large.

(e) When testing differences between population means using the One-Way Analysis of Variance (ANOVA) statistical method, the region of rejection is always in the left tail of the F distribution.

(f) In two factor factorial design, factors A and B are said to have interaction if the effect of factor A is dependent on the level of factor B.

(g) If the null hypothesis is rejected when conducting a one-way ANOVA F-test, then there are statistically significant differences between all pairs of means.

5. Suppose you are performing one-way ANOVA to test for a difference in means for 4 groups. Each group contains 10 individuals that are randomly selected from a large population. Before conducting the test, you conduct a quick power computation for a specific alternative hypothesis where $\mu_1 = 10$, $\mu_2 = 11$, $\mu_3 = 12$ and $\mu_4 = 13$. You need to estimate σ for the computation, and so you choose $\sigma = 3$, which seems reasonable. Would the power be larger, smaller, or about the same if the true σ was actually larger than 3?

6. Do people from different cultures experience emotions differently? One study designed to examine this question collected data from 410 college students from five different cultures. 9 The participants were asked to record, on a 1 (never) to 7 (always) scale, how much of the time they typically felt eight specific emotions. These were averaged to produce the global emotion score for each participant. Here is a summary of this measure:

Culture	n	$\bar{x}$	SD
European American	46	4.39	1.06
Asian American	33	4.35	1.18
Japanese	91	4.72	1.13
Indian	160	4.34	1.26
Hispanic American	80	5.04	1.16

(a) Complete the ANOVA table below for these results by filling in the five missing entries:

	Df	SS	MS	F
Culture		31.268		
Residuals			1.4044	n/a
Total	409	600.04	1.4671	n/a

(b) What is are the null hypothesis and alternative hypothesis for this ANOVA test?

(c) It turns out that the p-value for the F-statistic above is 2.27×10^{-4}. What does that mean in this situation?

(d) Is it reasonable to used a pooled standard deviation for these data? Why or why not?

(e) Why don't we need to worry very much about whether the assumption of normality is met for this data?

(f) Recall that the confidence interval for the difference between the means of two groups is $\bar{x}_A - \bar{x}_B \pm t^{**} s_p \sqrt{\frac{1}{n_A} + \frac{1}{n_B}}$, where t^{**} is the adjusted critical value with the Bonferroni correction. According to the Bonferroni method, what adjusted confidence level should we use to be 95% certain that we capture the true difference in population mean for each pair of groups simultaneously? *You don't need to compute the confidence interval.*