Unit 7 Progress Check: FRQ

1. Show all your work. Indicate clearly the methods you use, because you will be scored on the correctness of your methods as well as on the accuracy and completeness of your results and explanations.

A bank categorizes its customers into one of three groups based on their banking habits. A random sample of 30 customers from each group was selected, and the number of times each customer visited the bank during the past year was recorded. The following table shows the summary statistics.

Group	n	$\bar{x}$	s
A	30	48	7
B	30	51	8
C	30	54	10

The bank manager will investigate whether there is a significant difference in mean numbers of bank visits for the groups. Multiple two-sample t-tests will be conducted, each at the significance level of $\alpha = 0.05$.

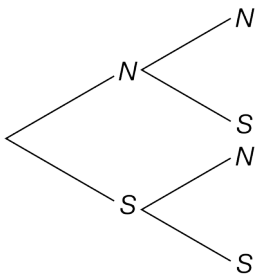
(a) How many t-tests will need to be conducted for the manager's investigation? List the pairs of groups for each test.

The significance level (α) of a single hypothesis test is the probability of making a Type I error. The manager wants to know the probability of making a Type I error for multiple t-tests, not just for a single t-test. This probability is called the family error rate for Type I error, which is also known as the family error rate.

(b) A t-test has two possible outcomes: reject or do not reject the null hypothesis. Suppose the null hypothesis is true. If the null hypothesis is rejected, the result is statistically significant, which would be a Type I error; if the null hypothesis is not rejected, the result is not statistically significant, which would not be a Type I error. Let S represent a statistically significant result, and let N represent a result that is not statistically significant.

(i) If $P(\mathrm{S}) = 0.05$, what is the value of $P(\mathrm{N})$?

The bank manager knows that the investigation will involve conducting multiple two-sample t-tests. The manager begins the investigation by considering two different t-tests as independent, successive trials. The possible outcomes of the trials, N or S, are shown in the following tree diagram.



Unit 7 Progress Check: FRQ

(ii) The family error rate is the probability of obtaining a significant result for at least one of the t-tests conducted, under the assumption that the null hypothesis is true. Use the tree diagram to determine the family error rate for two t-tests, each conducted at a level of $\alpha = 0.05$. Show your work.

(c) Determine the family error rate for the number of t-tests identified in part (a), each conducted at a level of $\alpha = 0.05$. Show your work.

(d) The manager wants the family error rate to be close to 0.05. Suggest a single significance level α that could be used for all of the individual t-tests that will bring the family error rate close to 0.05. Show work to support your suggested level.

2. Show all your work. Indicate clearly the methods you use, because you will be scored on the correctness of your methods as well as on the accuracy and completeness of your results and explanations.

The following stemplot shows the swimming speeds, in kilometers per hour (km/h), for a random sample of 31 emperor penguins.

Speed (km/h)	
7	8
8	3 4
8	6 7 9
9	0 0 1 3 4
9	5 5 6 7 8 8 9
10	0 1 1 2 3
10	5 8 8 8
11	0 2 3
11	5

Key: 7|8 = 7.8

(a) The mean of the sample is $9.771\ \text{km/h}$, and the standard deviation is $0.944\ \text{km/h}$. Construct and interpret a 95 percent confidence interval for the mean swimming speed of all emperor penguins in the population.

(b) Can the estimate of the mean swimming speed be generalized to all types of penguins? Explain your reasoning.