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 $\alpha = 0.05$.

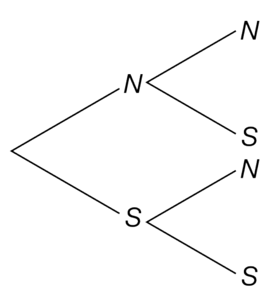
(a) How many t-tests will need to be conducted for the manager's investigation? List the pairs of groups for each test.

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.

Part A, B, C, and D

The primary goals of this question are to assess a student's ability to (1) count the number of t-tests required in a situation with multiple tests, (2) find the probability of a complement of an event, (3) calculate a family error rate for Type I error using a tree diagram, and (4) find a level of significance that will achieve a certain family error rate.

Each essentially correct (E) part counts as 1 point.

Each partially correct (P) part counts as $1/2$ point.

Scoring

Parts (a), (b), (c), and (d) are scored as essentially correct (E), partially correct (P), or incorrect (I).

If a response is between two scores (for example, $2\frac{1}{2}$ points), use a holistic approach to decide whether to score up or down, depending on the overall strength of the response and communication.

Reasons to score up:

· All notation is correct and clearly marked

· All explanations are clear

· No wrong information is included that was not part of the scoring (for example, saying sample size must be greater than 30 when that has nothing to do with the problem)

· No minor calculation errors are made, if they are not part of the scoring

· Interpretation parts are especially strong

Reasons to score down:

· Notation is not wrong, but is spotty and not clearly marked

· Explanations are not wrong, but are hard to follow

· Wrong or extraneous information is included but not part of scoring

· Minor calculation errors that are not part of the scoring are made

· Interpretation parts are scored an E but are considered a weak E

Unit 7 Progress Check: FRQ

0	1	2	3	4

Parts (a) through (d) sum to 4 points

OR

Parts (a) through (d) sum to $3^1/_2$ points AND a holistic approach is used to decide to score up

- Part (a) essentially correct
- Part (a) partially correct
- Part (a) incorrect
- Part (b) essentially correct
- Part (b) partially correct
- Part (b) incorrect
- Part (c) essentially correct
- Part (c) partially correct
- Part (c) incorrect
- Part (d) essentially correct
- Part (d) partially correct
- Part (d) incorrect

Solution

Part (a)

Three tests will be needed: A and B, A and C, B and C.

Scoring

Part (a) is scored as follows:

Essentially correct (E) if the response gives the three correct pairings. Stating 3 explicitly is not required.

Partially correct (P) if the response only answers 3 without the pairings

Incorrect (I) if the response does not meet the criteria for E or P

Note: If the response gives 6 pairings (taking order into account, for example, AB and BA), the response is scored P.

Unit 7 Progress Check: FRQ

Solution

Part (b)

(i) $P(\mathrm{N}) = 1 - P(\mathrm{S}) = 1 - 0.05 = 0.95$

(ii) The probability that at least one statistically significant result will occur equals 1 minus the probability that neither of the two tests give significant results (N and N).

In other words,
$$\begin{aligned} &P(\text{at least one S}) \\ &= P[(\text{S and N}) \text{ or } (\text{N and S}) \text{ or } (\text{S and S})] \\ &= 1 - P(\text{N and N}) \end{aligned}$$

From the tree, $P(\text{N and N}) = (0.95)^2 = 0.9025.$

Therefore, the family error rate for two *t*-tests is $1 - 0.9025 = 0.0975.$

OR

From the tree, we can see that at least one significant result is found with outcomes SN, NS, and SS. The sum of those probabilities is
$$\begin{aligned} &P(\text{S and N}) + P(\text{N and S}) + P(\text{S and S}) \\ &= (0.05)(0.95) + (0.95)(0.05) + (0.05)(0.05) \\ &= 0.0475 + 0.0475 + 0.0025 \\ &= 0.0975. \end{aligned}$$

Scoring

Part (b) is scored as follows:

Essentially correct (E) if the response satisfies the following three components:

· A response of 0.95 is given in part (b-i). (Work does not need to be shown.)

· Value of the family error rate in part (b-ii) is consistent with the answer in part (b-i)

· Work shown on how the family error rate was calculated

Partially correct (P) if the response satisfies only two of the three components

Incorrect (I) if the response does not meet the criteria for E or P

Notes:

· A response that gives the probability of obtaining a statistically significant result for exactly one of the *t*-tests (0.095) is scored as partially correct.

· A response that gives the probability of obtaining a statistically significant result for none of the *t*-tests (0.9025) is scored as partially correct.

Solution

Unit 7 Progress Check: FRQ

Part (c)

As indicated in part (a), the investigation will have 3 tests: AB, BC, AC. The probability of no significant results is $P(\text{N and N and N}) = (0.95)^3 = 0.857375.$

Therefore, the family error rate for 3 t-tests is $1 - 0.857375 = 0.142625,$ or approximately 14%.

Scoring

Part (c) is scored as follows:

Essentially correct (E) if the response satisfies the following two components:

· Value for the family error rate that is consistent with the responses from parts (a) and (b-i)

· Work shown on how the value was calculated, for example, a well-labeled tree diagram or the use of a complement.

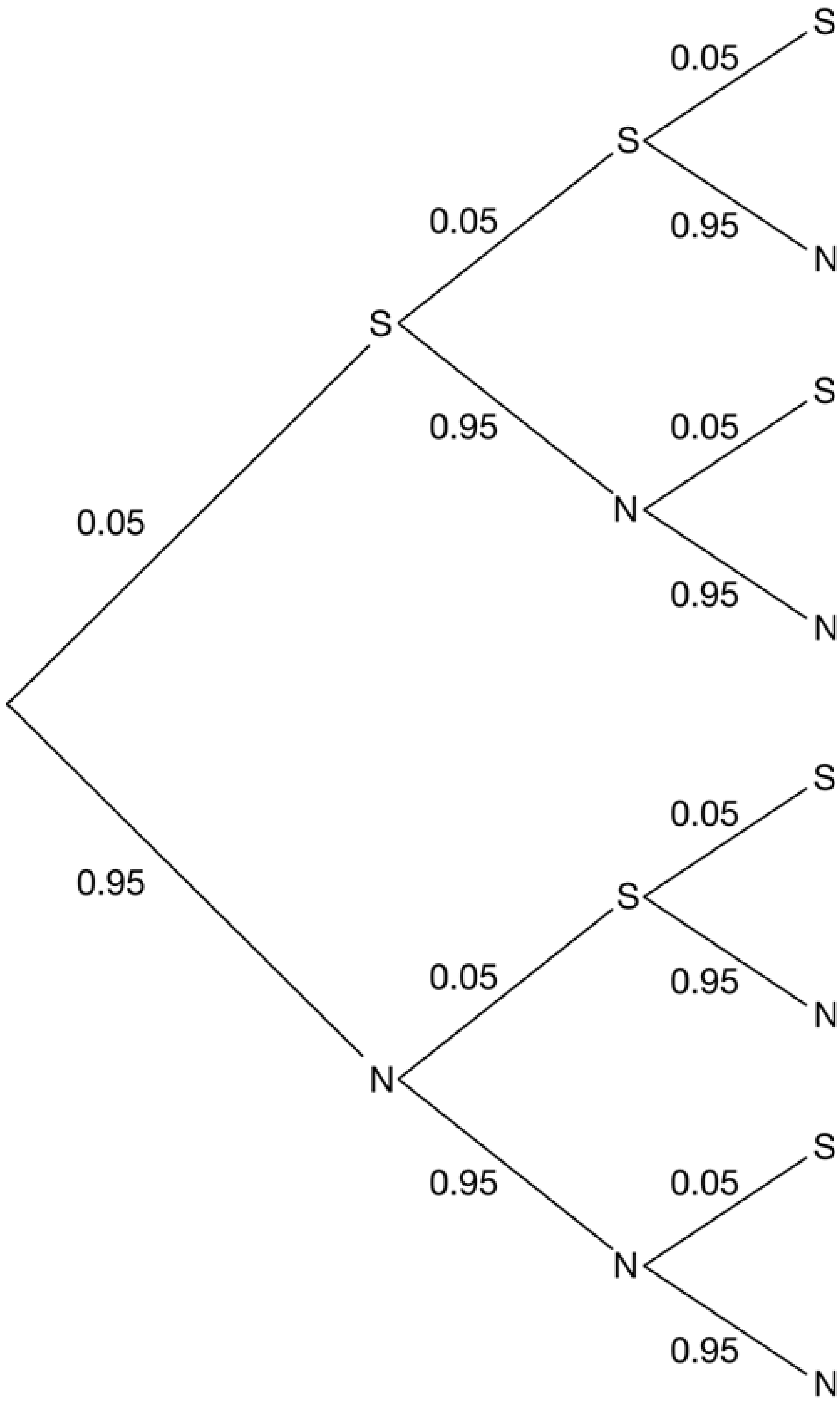
Partially correct (P) if the response satisfies only one of the two components

Incorrect (I) if the response does not meet the criteria for E or P

Note:

A correct tree diagram, although not required, will look like the following:

Unit 7 Progress Check: FRQ



Solution

Part (d)

Several approaches are acceptable. Here are two:

Unit 7 Progress Check: FRQ

Approach #1

Anything less than $\alpha = 0.05$ will reduce the family error rate. Because there are 3 tests, we might try dividing alpha by 3 for a significance level of $\alpha = \frac{0.05}{3} \approx 0.017.$ In this case, $P(\text{S}) = 0.017,\ P(\text{N}) = 0.983,$ P(N and N and N) = $(0.983)^3 \approx$ 0.94986 and the family error rate would be $1 - 0.94986 = 0.05014.$.

OR

Approach #2

$$1 - [P(\text{N})]^3 = 0.05$$
$$1 - [1 - P(\text{S})]^3 = 0.05$$
$$[1 - P(\text{S})]^3 = 0.95$$
$$1 - P(\text{S}) = \sqrt[3]{0.95}$$
$$1 - P(\text{S}) \approx 0.98305$$
$$P(\text{S}) \approx 0.01695$$

A suggested significance level is approximately 0.01695.This level will lead to a family error rate of $1 - (1 - 0.01695)^3 \approx 0.04999$.

Scoring

Part (d) is scored as follows:

Essentially correct (E) if the response satisfies the following two components:

· Recognition that the significance level for a single test should be less than 0.05

· Work that shows the suggested significance level in the response reduces the family error to roughly 5%

Partially correct (P) if the response satisfies only one of the two components

Incorrect (I) if the response does not meet the criteria for E or P

Notes:

· A response that shows work that is consistent with earlier incorrect results, such as identifying 6 pairs of groups in part (a), can still be scored as essentially correct.

· A strong response in part (d) can be used in holistic grading to decide to score up.

Unit 7 Progress Check: FRQ

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 ($\mathrm{km/h}$), 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 $9.771\ \mathrm{km/h}$, and the standard deviation is $0.944\ \mathrm{km/h}$. Construct <u>and</u> 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.

4-part Inference scoring

The primary goals of this question are to assess a student's ability to (1) construct and interpret a confidence interval for a population mean; and (2) determine the appropriateness of generalizing the results of an inferential procedure.

<u>Scoring</u>

Part (a) has three scoring steps, where each is scored as essentially correct (E), partially correct (P), or incorrect (I). Part (b) is scored as essentially correct (E), partially correct (P), or incorrect (I). So, part (a) is worth 3 Es and part (b) is worth 1 E.

Each essentially correct (E) part counts as 1 point.

Each partially correct (P) part counts as $1/2$ point.

If a response is between two scores (for example, $2 1/2$ points), use a holistic approach to decide whether to score up or down, depending on the overall strength of the response and communication.

Reasons to score up:

· All notation is correct and clearly marked

Unit 7 Progress Check: FRQ

· *All explanations are clear*

· *No wrong information is included that was not part of the scoring (for example, saying sample size must be greater than 30 when that has nothing to do with the problem)*

· *No minor calculation errors are made, if they are not part of the scoring*

· *Interpretation parts are especially strong*

Reasons to score down:

· *Notation is not wrong, but is spotty and not clearly marked*

· *Explanations are not wrong, but are hard to follow*

· *Wrong or extraneous information is included but not part of scoring*

· *Minor calculation errors that are not part of the scoring are made*

· *Interpretation parts are scored an E but are considered a weak E*

0	1	2	3	4

Scoring steps 1, 2, 3 and part (b) sum to 4 points

OR

Scoring steps 1, 2, 3 and part (b) sum to $3\frac{1}{2}$ points AND a holistic approach is used to decide to score up

- [ ] Scoring Step 1 Part (a) essentially correct
- [ ] Scoring Step 1 Part (a) partially correct
- [ ] Scoring Step 1 Part (a) incorrect
- [ ] Scoring Step 2 Part (a) essentially correct
- [ ] Scoring Step 2 Part (a) partially correct
- [ ] Scoring Step 2 Part (a) incorrect
- [ ] Scoring Step 3 Part (a) essentially correct
- [ ] Scoring Step 3 Part (a) partially correct
- [ ] Scoring Step 3 Part (a) incorrect
- [ ] Part (b) essentially correct
- [ ] Part (b) partially correct
- [ ] Part (b) incorrect

Solution

Unit 7 Progress Check: FRQ

Part (a)

Scoring part 1: Identify the correct procedure with conditions

The appropriate procedure is the one-sample t-interval for population mean: $\bar{x} \pm t^* \frac{s}{\sqrt{n}}$.

Conditions:

· Data were collected from a random sample as stated.

· The sample size $n = 31$ is large enough to assume that the sampling distribution of the sample mean is approximately normal.

Scoring

Scoring step 1 is scored as follows.

Essentially correct (E) if the response satisfies the following three components:

· The correct interval is identified, either by name or formula.

· The sample size condition is checked and indicates the sample size is greater than 30.

OR

It can be assumed that the population is roughly normal because the distribution of the sample, as shown in the stemplot, does not have strong skew or outliers.

· The random sampling condition is checked.

Partially correct (P) if the response satisfies only two of the three components

Incorrect (I) if the response does not meet the criteria for E or P

Note:

A response that checks the 10% condition, that it is reasonable to assume there are more than $10(31) = 310$ penguins, can be used in holistic grading to decide to score up.

Solution

Scoring part 2: Construct the interval

Since $\text{df} = n - 1 = 31 - 1 = 30$ for 95% confidence and 30 degrees of freedom, the correct t-critical value is $t^* = 2.042$.

The interval is $9.771 \pm 2.042 \frac{0.944}{\sqrt{31}} = 9.771 \pm 0.346,$ which produces the interval (9.425, 10.117).

Scoring

Scoring step 2 is scored as follows.

Unit 7 Progress Check: FRQ

Essentially correct (E) if the response calculates the correct interval with work

Partially correct (P) if the response calculates the correct interval with no work

OR

if the response gives an interval with a calculation error or with the wrong t-value

Incorrect (I) if the response does not meet the criteria for E or P

Solution

Part (a)

Scoring step 3: Interpret the interval.

We are 95% confident that the mean swimming speed of the population of emperor penguins is between 9.425 kilometers per hour and 10.117 kilometers per hour.

OR

We are 95% confident that the confidence interval (9.425, 10.117) captures the population mean swimming speed of emperor penguins.

Scoring

Scoring step 3 is scored as follows.

Essentially correct (E) if the response satisfies the following three components:

· A reasonable interpretation in context is given.

· The interpretation is clear that the interval estimates the population mean.

· The interpretation is given with 95% confidence.

Partially correct (P) if the response includes only two of the three conditions

Incorrect (I) if the response does not meet the criteria for E or P

Solution

Part (b)

It is not reasonable to generalize the estimate of the mean swimming speed to all types of penguins because the sample only consisted of emperor penguins.

Scoring

Part (b) is scored as follows.

Essentially correct (E) if the response satisfies the following two components:

· Identifies that the results cannot be generalized

Unit 7 Progress Check: FRQ

· A correct justification which indicates that only emperor penguins were sampled

Partially correct (P) if the response satisfies the first component but provides weak justification

Incorrect (I) if the response does not meet the criteria for E or P