Answers for 2005 Super Questions

Question 1

Solution

Part (a):

The distribution is skewed to the left (or toward the lower v

Indicating a higher number of students did well on the exam that did poorly.

Part (b):

Since the distribution is skewed towards the lower values, the mean will be pulled in that direction. Thus, the instructor should report the median to motivate her stud

… the median is resistant to the distribution being skewed to lower values…

Part (c):

Step 1: Correct Mechanics:

$$\text{midrange} = \frac{64 + 95}{2} = 79.5$$

Step 2: Identify the midrange as a measure of center.

Remember you must **show all calculations**- this year especially!

Step 3: Correct rationale:

The maximum provides information about the upper tail, more specifically the upper extreme value. The minimum provides information about the lower tail, more specifically the lower extreme value. By averaging these two values and creating the midrange, we are creating a statistic that provides the halfway point between the two extremes. This statistic is a measure of center.

d) The probability that a student randomly selected from this group will have a score lees than 80 is **11/32 or 34.375%.**

e) Yes, it is reasonable to make generalizations about the population of all 150 students based on this data. The 32 students were randomly selected and can be considered a representative sample of the all 150 students.

f) The three conditions needed to create a confidence interval are:

Randomness --- The problem state that the 32 students were randomly selected. Condition satisfied.

Large Counts --- The CLT states that sample sizes of at 30 will create sampling distributions that are approximately normal. 32 > 30 Condition satisfied.

Independence --- Population size must be greater than 10 times the sample size. Our population is the 150 students who took the test. 150 is **NOT GREATER** than 10(32), therefore this condition is **not satisfied**.

Question 3

Solution

Part (a):

Each of the 100 selected people will be assigned a unique random number using a random number generator. A list of names and numbers will be created and sorted from smallest to largest by the assigned numbers (and carrying along the name). The first 50 people on the list will be asked to apply the new compound to their right arm and the other 50 people will be asked to apply the current compound to their right arm. The compounds will be put in identical, unmarked tubes so neither the participants nor the researchers know which compound is being applied. The analysts will be the only people who know which participants received which compound. Each person will be randomly assigned to a bin by assigning random numbers to bins using a random number generator. The first person on the list will be assigned to the bin with the smallest number, the second person on the list will be assigned to the bin with the second smallest number, and so on. After each person inserts his or her right arm into the assigned bin for one minute, the number of mosquito bites will be counted. The mean number of mosquito bites for the two compounds will be compared using a two-sample *t*-test and/or a confidence interval for the difference in means for two independent samples.

Part (b):

Each participant will be randomly assigned to a bin as described in part (a). The researchers will distribute two identical tubes, one labeled 1 and the other labeled 2, to each participant. One of those tubes will contain the new compound and the other will contain the current compound. Neither the researchers nor the participants will know which compound is in which tube. Only the analyst will have this information. Each participant will apply one compound to one arm and the other compound to the other arm. The assignment of the compounds to the arms is completed using randomization. A random number will be generated for each participant, and the participants with the 50 smallest random numbers will apply tube 1 to their right arm, and the remaining 50 participants will apply tube 2 to their left arm. Each participant will insert both arms into the assigned bin at the same time for one minute, and the number of mosquito bites will be counted on each arm. The analyst will compute the difference in the number of bites (new – current) for each of the 100 participants and use a one-sample *t*-test and/or construct a confidence interval for the mean of the differences to test the null hypothesis that the mean difference is zero.

Part (c):

The matched-pairs design in part (b) is better because one potential source of variation, person-to-person variability in susceptibility to mosquito bites, is controlled.

d) An appropriate graphical display to compare the treatments could be parallel box plots, back-to-back stem plots, parallel dot plots, two histograms using the same scale for the bins. **(You only have to give 1 example.)**

e) **(Remember matched pairs is a mean difference problem.)**

H_0: $\mu_{diff} = 0$ H_a: $\mu_{diff} < 0$ where μ_{diff} = true mean difference in mosquito bites on arms with new compound – arms with old compound

f) Using a significance level of $\alpha = 0.5$, since our p-value of 0.023 is less than $\alpha = 0.5$, we have evidence to reject the null hypothesis. There is evidence that it is plausible that the new repellent has a lower mean number of bites than the old repellent and therefore may be more effective. Our data was statistically significant.

Question 4

Solution

Part (a):

Step 1: Identify appropriate confidence interval by name or by formula.

One sample confidence interval for a mean (of the differences)

$$OR \quad \bar{x}_d \pm t^*_{n-1}\frac{s_d}{\sqrt{n}}$$

Step 2: Check appropriate conditions.

Assume the population of differences in growth is normally distributed. The information provided in the stem of the problem suggests that this condition is met. Because the 24 seeds were randomly chosen and randomly assigned to the containers, the differences are independent.

Step 3: Correct mechanics.

The 95% confidence interval for the mean difference in growth is

$$-2.015 \pm 2.201\frac{1.163}{\sqrt{12}} = -2.015 \pm (2.201)(0.336) = -2.015 \pm 0.7389$$

or (-2.7539, -1.2761).

In step 3, other confidence levels may be used, e.g.,

- 90% C.I. is $-2.015 \pm 1.796\frac{1.163}{\sqrt{12}} = -2.015 \pm 0.6030$ or (-2.618, -1.412)
- 99 % C.I. is $-2.015 \pm 3.106\frac{1.163}{\sqrt{12}} = -2.015 \pm 1.0428$ or (-3.0578, -.9722)

Step 4: Interpret the confidence interval in context.

We are 95% confident that the mean difference in the growth of the untreated and treated seeds is between –2.7539 and –1.2761.

Part (b):

Step 1: Identify a correct pair of hypotheses.

$H_0 : \mu_d = 0$ versus $H_a : \mu_d \neq 0$, where μ_d is the mean difference in the untreated and treated seeds.

Step 2: State the correct conclusion in context.

Since the 95% confidence interval does not include zero, the null hypothesis can be rejected at the $\alpha = 0.05$ significance level. In other words, we have statistically significant evidence at the $\alpha = 0.05$ level that there is a mean difference in the growth of untreated and treated seeds.

If a two-sample procedure is used, the highest possible score is (P). The 95% confidence interval for the difference in the two means is $(\bar{x}_1 - \bar{x}_2) \pm t^*_{\min(n_1-1, n_2-1)}\sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}} =$

$$(15.989 - 18.004) \pm 2.201\sqrt{\frac{(1.098)^2}{12} + \frac{(1.175)^2}{12}} = -2.015 \pm 1.0218 \text{ or } (-3.037, -0.993).$$

The incorrect two-sample confidence intervals from the calculator are:

- 95% C.I. is (–2.978, –1.052)
- 99% C.I. is (–3.324, –0.706)

c) *The center is at -2.015 cm, the sd (which we are estimating with a se) is 0.336 cm, and the shape may be assumed to be approximately normal. Although with a sample of only size 12, even though we are told the assumption of normality is reasonable, we are cautious about this conclusion.*

d) This is an experiment because the researcher randomly selected the seeds that would receive the treatment (additive.

e) *i)* The researcher used a matched pairs design for the experiment. One treated seed and one control seed were each placed in a container. This attempts to control any response that may be the result of the container itself.

ii) The experimental units are the seeds.

iii) The factor is the additive.

iv) The reposne variable is the growth of the plant.