Question 5

Intent of Question

The primary goal of this question was to assess students' ability to identify, set up, perform, and interpret the results of an appropriate hypothesis test to address a particular question. More specific goals were to assess students' ability to (1) state appropriate hypotheses; (2) identify the appropriate statistical test procedure and check appropriate conditions for inference; (3) calculate the appropriate test statistic and p-value; and (4) draw an appropriate conclusion, with justification, in the context of the study.

Solution

Step 1: States a correct pair of hypotheses.

Let μ_{diff} represent the population mean difference in purchase price (woman – man) for identically equipped cars of the same model, sold to both men and women by the same dealer, in the county.

The hypotheses to be tested are $H_0 : \mu_{\text{diff}} = 0$ versus $H_a : \mu_{\text{diff}} > 0$.

Step 2: Identifies a correct test procedure (by name or by formula) and checks appropriate conditions.

The appropriate procedure is a paired t-test.

The conditions for the paired t-test are:
1. The sample is randomly selected from the population.
2. The population of price differences (woman – man) is normally distributed, or the sample size is large.

The first condition is met because the car models and the individuals were randomly selected. The sample size ($n = 8$) is not large, so we need to investigate whether it is reasonable to assume that the population of price differences is normally distributed. The dotplot of sample price differences reveals a fairly symmetric distribution, so we will consider the second condition to be met.

Step 3: Correct mechanics, including the value of the test statistic and p-value (or rejection region).

The test statistic is $t = \dfrac{585 - 0}{\dfrac{530.71}{\sqrt{8}}} \approx 3.12$.

The p-value, based on a t-distribution with $8 - 1 = 7$ degrees of freedom, is 0.008.

Step 4: States a correct conclusion in the context of the study, using the result of the statistical test.

Because the p-value is very small (for instance, smaller than $\alpha = 0.05$), we reject the null hypothesis. The data provide convincing evidence that, on average, women pay more than men in the county for the same car model.

Scoring

Each of steps 1, 2, 3, and 4 were scored as essentially correct (E), partially correct (P), or incorrect (I).

Step 1 is scored as follows:

Essentially correct (E) if the response identifies the correct parameter *AND* states correct hypotheses.

Partially correct (P) if the response identifies the correct parameter *OR* states correct hypotheses, but not both.

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

Note: Defining the parameter symbol in context or simply using common parameter notation is sufficient.

Step 2 is scored as follows:

Essentially correct (E) if the response identifies the correct test procedure (by name or by formula) *AND* checks <u>both</u> conditions correctly.

Partially correct (P) if the response correctly completes two of the three components (identification of procedure, check of randomness condition, check of normality condition).

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

Note: The random sampling condition can be verified by referring to the random selection of car models or to the random selection of male and female car buyers.

Step 3 is scored as follows:

Essentially correct (E) if the response correctly calculates both the test statistic and the *p*-value.

Partially correct (P) if the response correctly calculates the test statistic but not the *p*-value;
    OR
if the response calculates the test statistic incorrectly but then calculates the correct *p*-value for the computed test statistic.

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

Note: If the response identifies a *z*-test for a mean as the correct procedure in step 2, then the response can earn a P in step 3 if both the test statistic and the *p*-value are calculated correctly.

Step 4 is scored as follows:

Essentially correct (E) if the response provides a correct conclusion in context, also providing justification based on linkage between the p-value and the conclusion.

Partially correct (P) if the response provides a correct conclusion with linkage to the p-value, but not in context;
    OR
if the response provides a correct conclusion in context, but without justification based on linkage to the p-value.

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

Notes:
- If the conclusion is consistent with an incorrect p-value from step 3 and also in context with justification based on linkage to the p-value, step 4 is scored as E.
- A response that performs a two-sample t-test with correct calculations should fail to reject H_0. A conclusion that is equivalent to "accept H_0" (such as "we conclude that women pay the same amount as men, on average"), either as a stated decision or as a conclusion in context, cannot be scored as E. Such a response will be scored as P provided that the conclusion is in context with linkage. Such a response will be scored as I if it lacks either context or linkage.

Each essentially correct (E) step counts as 1 point. Each partially correct (P) step counts as ½ point.

    4   **Complete Response**

    3   **Substantial Response**

    2   **Developing Response**

    1   **Minimal Response**

If a response is between two scores (for example, 2½ points), use a holistic approach to decide whether to score up or down, depending on the overall strength of the response and communication.
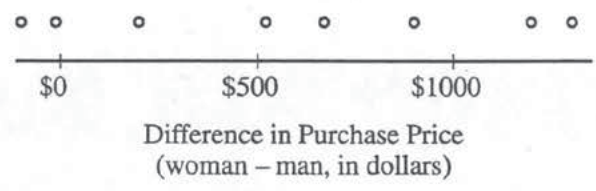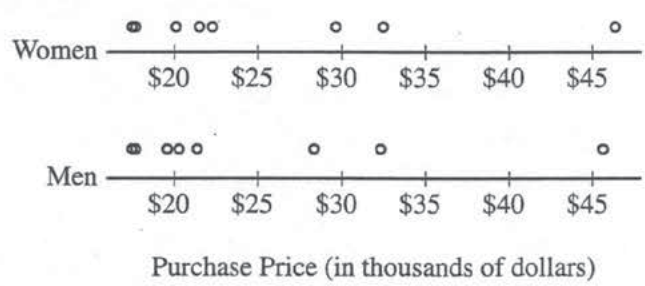
5. A researcher conducted a study to investigate whether local car dealers tend to charge women <u>more than</u> men for the same car model. Using information from the county tax collector's records, the researcher <u>randomly selected</u> one man and one woman from among everyone who had purchased the same model of an identically equipped car from the same dealer. The process was repeated for a total of 8 randomly selected car models.

The purchase prices and the differences (woman − man) are shown in the table below. Summary statistics are also shown.

Car model	1	2	3	4	5	6	7	8
Women	$20,100	$17,400	$22,300	$32,500	$17,710	$21,500	$29,600	$46,300
Men	$19,580	$17,500	$21,400	$32,300	$17,720	$20,300	$28,300	$45,630
Difference	$520	−$100	$900	$200	−$10	$1,200	$1,300	$670

	Mean	Standard Deviation
Women	$25,926.25	$9,846.61
Men	$25,341.25	$9,728.60
Difference	$585.00	$530.71

Dotplots of the data and the differences are shown below.

Purchase Price (in thousands of dollars)

Difference in Purchase Price
(woman − man, in dollars)

Do the data provide convincing evidence that, on average, women pay more than men in the county for the same car model?

If you need more room for your work in question 5, use the space below.

State: We want to test the following hypotheses at the assumed $\alpha = 0.05$ significance level.

$H_0: \mu_D = 0$

$H_A: \mu_D > 0$

where μ_D is the population mean difference (woman-man) in purchase prices.

Plan: If the necessary conditions (below) are met, we will use a matched-pairs t-test for mean difference.

Random: Stated that "the researcher randomly selected one man and one woman" and the 8 car models were "randomly selected."

Normal: Since the sample size is small (n<30), the CLT does not apply. Examining the given dotplot for the difference in purchase price, there does not appear to be outliers or skewness. There is no reason to doubt the normality of the distribution of differences.

Independent- 10% condition (sampling without replacement) $\rightarrow$ There are way more than 80 car models sold by local car dealers. $N \geq 10n$ $N \geq 10(8)$ $N \geq 80$ ✓
Also, we can assume that the purchased price of one car model does not affect another.

DO· $\bar{x}_D = 585$ Test Statistic $t = \dfrac{\bar{x}_D - \mu_0}{\frac{S_D}{\sqrt{n}}}$ $n = 8$

$S_D = \$30.71$

$df = n-1 = 8-1 = 7$

$t = \dfrac{585 - 0}{\frac{530.71}{\sqrt{8}}} = \dfrac{585}{187.63} \approx 3.1178$

p-value$= P(t \geq 3.1178) = 0.0084$

conclude: Since our P-value of 0.0084 is smaller than our significance level, $\alpha = 0.05$. we reject the null hypothesis H_0. There is sufficient evidence to conclude that on average, women pay more than men in the county for the same car model.
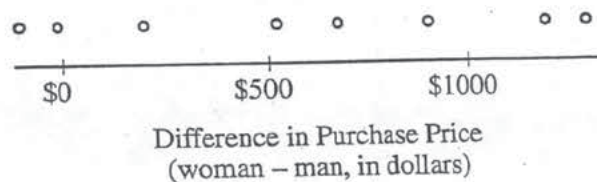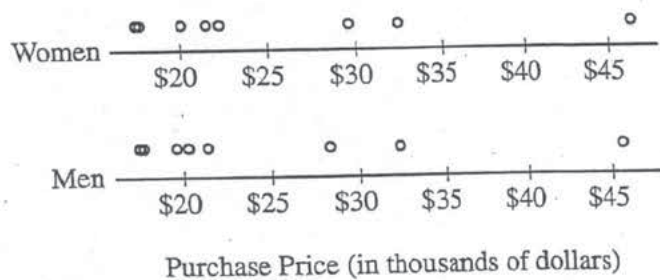
5. A researcher conducted a study to investigate whether local car dealers tend to charge women more than men for the same car model. Using information from the county tax collector's records, the researcher randomly selected one man and one woman from among everyone who had purchased the same model of an identically equipped car from the same dealer. The process was repeated for a total of 8 randomly selected car models.

The purchase prices and the differences (woman – man) are shown in the table below. Summary statistics are also shown.

Car model	1	2	3	4	5	6	7	8
Women	$20,100	$17,400	$22,300	$32,500	$17,710	$21,500	$29,600	$46,300
Men	$19,580	$17,500	$21,400	$32,300	$17,720	$20,300	$28,300	$45,630
Difference	$520	−$100	$900	$200	−$10	$1,200	$1,300	$670

	Mean	Standard Deviation
Women	$25,926.25	$9,846.61
Men	$25,341.25	$9,728.60
Difference	$585.00	$530.71

Dotplots of the data and the differences are shown below.

Purchase Price (in thousands of dollars)

Difference in Purchase Price
(woman – man, in dollars)

Do the data provide convincing evidence that, on average, women pay more than men in the county for the same car model?

If you need more room for your work in question 5, use the space below.

M_d: mean difference between purchase price of cars between women and men

H_0: $M_d = 0$

H_A: $M_d > 0$

The data is paired by car model. The men and women were randomly selected. 8 car models are less than 10% of all car models. The purchase price of one car is independent from the purchase price of another. The nearly normal condition is satisfied because the histograms are unimodal and somewhat symmetric.

Md   women   men
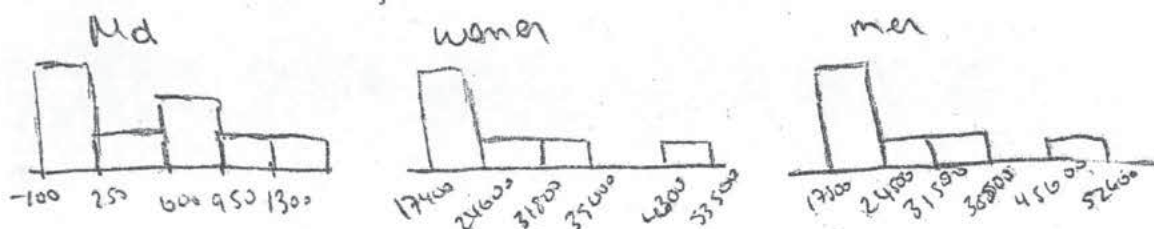
We will perform a matched pairs t test

$M_d = 585$ $S_d = 530.7$ $n = 8$

$t = 3.11$
$p = 0.0084$

$t = \dfrac{M_d - 0}{\frac{S_d}{\sqrt{n}}} = \dfrac{585 - 0}{\frac{530.7}{\sqrt{8}}} = 3.11$
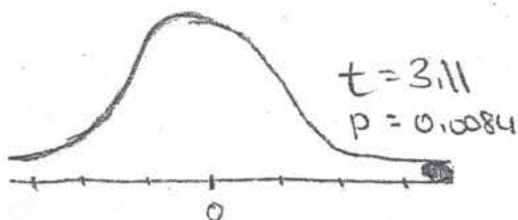
Since the p-value of 0.0084 is lower than the alpha level of 0.05, we reject the null hypothesis. There is sufficient evidence to conclude that women pay more than men in the county for the same model of car.
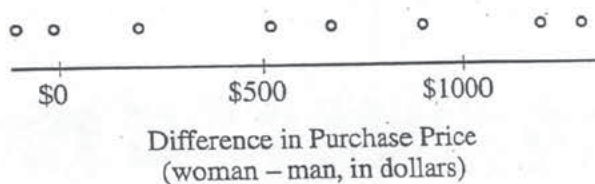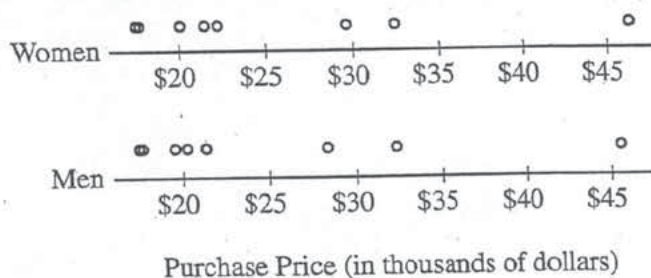
5. A researcher conducted a study to investigate whether local car dealers tend to charge women more than men for the same car model. Using information from the county tax collector's records, the researcher randomly selected one man and one woman from among everyone who had purchased the same model of an identically equipped car from the same dealer. The process was repeated for a total of 8 randomly selected car models.

The purchase prices and the differences (woman – man) are shown in the table below. Summary statistics are also shown.

Car model	1	2	3	4	5	6	7	8
Women	$20,100	$17,400	$22,300	$32,500	$17,710	$21,500	$29,600	$46,300
Men	$19,580	$17,500	$21,400	$32,300	$17,720	$20,300	$28,300	$45,630
Difference	$520	–$100	$900	$200	–$10	$1,200	$1,300	$670

	Mean	Standard Deviation
Women	$25,926.25	$9,846.61
Men	$25,341.25	$9,728.60
Difference	$585.00	$530.71

Dotplots of the data and the differences are shown below.

Purchase Price (in thousands of dollars)

Difference in Purchase Price
(woman – man, in dollars)

Do the data provide convincing evidence that, on average, women pay more than men in the county for the same car model?

-12-

If you need more room for your work in question 5, use the space below.

μ_m = true mean of price paid by men

μ_F = true mean of price paid by women

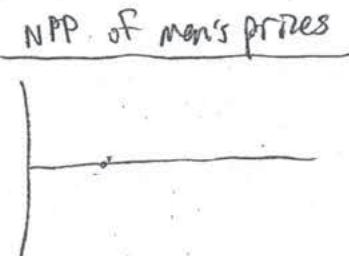H_0: $\mu_m = \mu_F$; Women pay the same as men in the county on average for the same car model.

H_a: $\mu_m < \mu_F$; Women pay more than men in the county on average for the same car model.

i) SRS - stated ✓ ii) Independence - assume more than 80 cars bought total by both men and women in the county.

iii) Normality - NPP of Women's prices approximately linear → approximately normal ✓
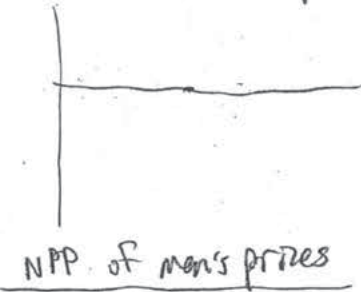
NPP of men's prices approximately linear → approximately normal ✓

2 - sample t-test: $df = 14$ $t = 0.1195$ $p = 0.453$

Since our p-value is higher than α (0.05) we do not have enough evidence to reject H_0. The data do not provide convincing evidence that women pay more than men in the county for the same car model. We conclude that women pay the same as men in the county on average for the same car model.

Question 5

Overview

The primary goal of this question was to assess students' ability to identify, set up, perform, and interpret the results of an appropriate hypothesis test to address a particular question. More specific goals were to assess students' ability to (1) state appropriate hypotheses; (2) identify the appropriate statistical test procedure and check appropriate conditions for inference; (3) calculate the appropriate test statistic and p-value; and (4) draw an appropriate conclusion, with justification, in the context of the study.

Sample: 5A
Score: 4

In Step 1 the student states a correct pair of hypotheses and defines the parameter appropriately, so this step was scored as essentially correct. In Step 2 the response correctly identifies a "matched pairs t-test" as the appropriate inference procedure. Because the random and normal conditions are verified appropriately, this step was scored as essentially correct. (Note that the comments about independence were viewed as extraneous.) In Step 3 the student correctly calculates the test statistic and p-value with appropriate supporting work. As a result, this step was scored as essentially correct. In Step 4 the student's conclusion begins with a correct decision about the null hypothesis, with linkage to the p-value. This is followed by a correct contextual conclusion about the mean difference in purchase price for women and men. Consequently, Step 4 was scored as essentially correct. Because all four steps were scored as essentially correct, the response earned a score of 4.

Sample: 5B
Score: 3

In Step 1 the student gives a reasonable description of the parameter and states a correct pair of hypotheses. As a result, this step was scored as essentially correct. In Step 2 the student identifies a matched pairs t-test as the appropriate procedure, and checks the random condition in an acceptable way. However, the student discusses all three dotplots when examining the normal condition. With two of the three required components satisfied, Step 2 was scored as partially correct. In Step 3 the student obtains the correct test statistic and p-value, but makes a notation error when referring to the sample mean as μ_d, so this step was scored as partially correct. In Step 4 the conclusion begins with a correct decision about H_0, with linkage to the p-value. This is followed by a contextual conclusion about women paying more than men that does not include the idea of on average. Due to this omission, Step 4 was scored as partially correct. Because one step was scored as essentially correct and three steps were scored as partially correct, this response was initially scored as a 2½. Looking holistically at the paper, the student demonstrates clear understanding of the process of performing a paired t-test. Therefore, this response earned a score of 3.

Sample: 5C
Score: 2

In Step 1 the response begins by stating a pair of hypotheses with two parameters, which gives no indication of the paired data setting in this question. Because the parameter definitions are reasonable, Step 1 was scored as partially correct. In Step 2, the student identifies an incorrect procedure (two-sample t-test) and checks the normality condition using two normal probability plots, which is not appropriate for a paired t-test. Although the random condition is verified, with only one of the three required components present, Step 2 was scored as incorrect. In Step 3 the student provides a correct test statistic and p-value for a two-sample t-test, so this step was scored as essentially correct. In Step 4, the first two sentences of the conclusion include all of the required elements for an essentially correct response. However, the last sentence is equivalent to accepting the null hypothesis, so Step 4 was scored as partially correct. Because one step was scored as essentially correct, two steps were scored as partially correct, and one step was scored as incorrect, the response earned a score of 2.