+ 5 EC.

Try to keep your work on these sheets. Show or explain ALL work for full credit. <u>BOX in your answers please</u>. DATA TABLES are on the last pages.

1. (1.3: 10 pts) Secondhand smoke: A recent study compared the heart rates of 19 infants born to nonsmoking mothers with those of 17 infants born to mothers who smoked an average of 15 cigarettes a day while pregnant and after giving birth. The heart rates of the infants at one year of age were 20% slower on the average for the smoking mothers.

(a) What is the outcome variable?

2   **Heart Rates of infants.**

(b) What is the treatment variable?

2   **Second hand smoke**

(c) Was this a cohort study or a case-control study?

2   **Cohort** {case-controlled means you randomly ask mothers to smoke}

(d) Was the study prospective, cross-sectional, or retrospective?

2   **retrospective** { look into the past }

(e) Could the results be due to confounding? Explain.

2   Smoking mothers may not care for their diets either.
       So yes — mother's diets may contribute

2. (1.3) (5 pts) Eat your vegetables: In an observational study, people who ate four or more servings of fresh fruits and vegetables each day were less likely to develop colon cancer than people who ate little fruit or vegetables. True or false:

(a) The results of the study show that eating more fruits and vegetables (reduces) your risk of contracting colon cancer.

2   **False** – However, the ppl who ate fruits & vegs. also had lower colon cancer.

(b) The results of the study may be due to confounding, since the lifestyles of people who eat large amounts of fruits and vegetables may differ in many ways from those of people who do not.

3   **True** – Healthy eaters may work out/walk/sleep well

3. (1.4: 10 pts)   Specify the main type of bias involved.

(a) A bank sent out questionnaires to a simple random sample of 500 customers asking whether they would like the bank to extend its hours. Eighty percent of those returning the questionnaire said they would like the bank to extend its hours. Of the 500 questionnaires, 20 were returned.

2   Non-Response Bias — people who refuse to participate

(b) To determine his constituents' feelings about election reform, a politician sends a survey to people who have subscribed to his newsletter. More than 1000 responses are received.

2   Voluntary Response Bias — people chose to participate (opinionated).

(c) An e-store that sells phone accessories reports that 98% of its customers are satisfied with the speed of delivery.

2   Self-Interest Bias — poll creators drive questions in their way

(d) A sign in a restaurant claims that 95% of their customers believe them to have the best food in the world.

2   Self-Interest Bias — poll creators drive the questions in their way

(e) A television newscaster invites viewers to tweet their opinions about whether the U.S. Congress is doing a good job in handling the economy. More than 100,000 people send in an opinion.

2   Voluntary Response Bias — people choose to participate

10

4. (9.3: 10 pts) An antifungal medication was applied to the skin of eight adult subjects. One hour later, the amount of active ingredient that had been absorbed into the skin was measured for each subject. The results, in micrograms, were:

$$2.13 \quad 1.88 \quad 2.07 \quad 1.19 \quad 2.51 \quad 5.61 \quad 2.81 \quad 3.05$$

(a) Construct a boxplot for these data (use Statdisk). Is it appropriate to perform a hypothesis test?

2

Draw Box Plot here:

→ Not really, skewed right, which is OK if we had $n \geq 30$
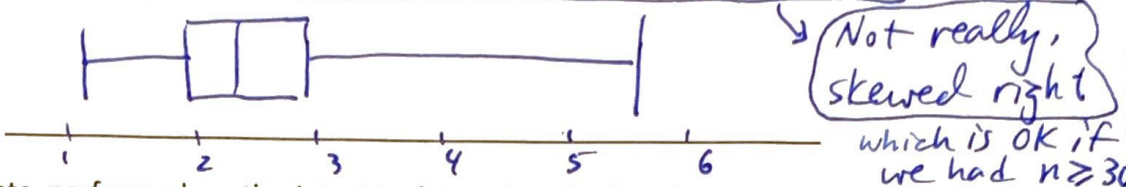
(b) Assuming it is appropriate, perform a hypothesis test to determine whether the mean amount absorbed is less than 3 micrograms. Use the $\alpha = 0.05$ level of significance. Populate the blanks

Hypothesis Testing: Mean One Sample

Use Summary Statistics   **Use Data**

Alternative Hypothesis:
1) Population Mean [<] Claimed Mean

5

Significance: 0.05

Claimed Mean: 3.0

Population Standard Deviation: (if known)

Column Containing Sample Data: 1
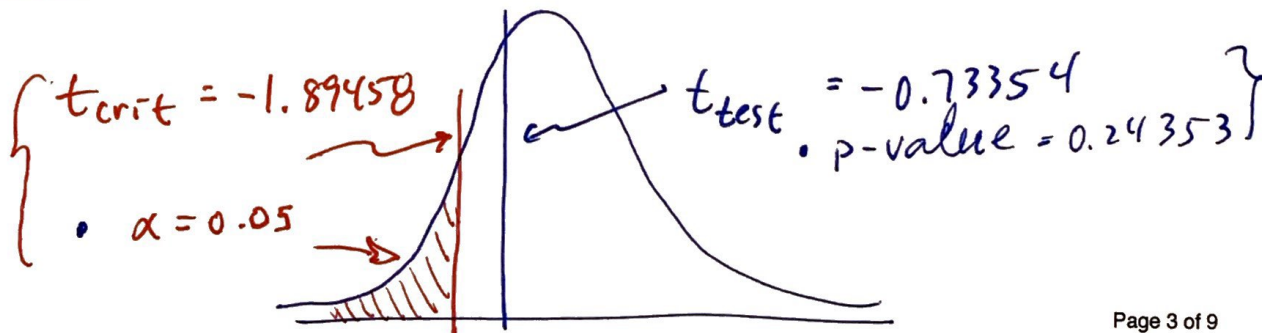
Evaluate

$t_{test} = -0.73354$

$t_{crit} = -1.89458$

p-value $= 0.24353$

{ there is a 24% chance that this could have occured given that the claim of L.T. 3mg is considered true }

(c) **What do you conclude?**

There is not enough evidence to suggest that the absorbed amount remains **below 3 mg**. What we see could be statistical variation.

3

$t_{crit} = -1.89458$

• $\alpha = 0.05$

$t_{test} = -0.73354$
• p-value $= 0.24353$

(0)

NoTe: The Conf. Intvl: $1.76841 < \mu < 3.54409$ Contains 3mg (so NOT unusual)

7. (11.3: 10 pts) At the intersection of Route 7 and North Shrewsbury in Clarendon, Vermont, 154 vehicles were observed to encounter a yellow light in the indecision zone, and 21 of them ran the red light. At the intersection of Route 62 and Paine Turnpike in Berlin, Vermont, 183 vehicles entered the intersection in the indecision zone, and 20 ran the red light. Can you conclude that the proportion of red light runners differs between the two intersections? $\alpha = 0.01$

① $21/154 = 0.1364$ ② $\frac{20}{183} = 0.1093$

STEP 0: **(a) Type of problem and table to use**
- HT for a proportion $\hat{p}$: 1- pop or (2 pop) (circle) then use a z-test statistic & z-table
- HT for means μ (σ unknown): 1- pop or 2 pop (circle) then use a t-test & t-table
- HT for matched pairs means μ (σ unknown): 1- pop or 2 pop (circle) then use a z-test
- goodness-of-fit test then use a χ^2-test statistic & χ^2-table
- contingency tests (independence or homogeneity) then use a χ^2-test & χ^2-table

(b) Assumptions   **Justification**

2

- SRS both   Independent Intersections
- $n < 10\%$ both   more than 1540 & 1830 cars respectively
- $X > 10$ both $21 > 10$ and $20 > 10$ ✓
- not $x > 10$ $154 - 21 > 10$ and $183 - 20 > 10$ ✓

STEP 1: State the Hypotheses and test tail type (if appropriate)

(a) H_o: __P_1__ = __P_2__ H_A: __P_1__ >, <, $\neq$ __P_2__ (circle)

(b) Tail: left | right (two-tail) (circle)
(c) Sketch the tail(s):

2   **STEP 2:** State the level of significance: $\alpha = 0.01$
Now look up the critical value in the appropriate table { revealed in STEP 0 (a) }
(z_c) or t_c or χ^2 (circle) = __2.576__ { last line of t-table - 2 tails }

STEP 3: Compute the test statistic. {for contingency tests Exp Val = (Row Total)(Col Total) / Grand Total }

(a) SE Formula $\sqrt{\frac{p_0 q_0}{n}}$ $\left(\sqrt{\frac{\hat{p}\hat{q}}{n_1} + \frac{\hat{p}\hat{q}}{n_2}}, \ \hat{p} = \frac{y_1 + y_2}{n_1 + n_2}\right)$ $\frac{s}{\sqrt{n}}$ $\sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}}$ $\frac{s_d}{\sqrt{n}}$ (circle one):

$\hat{p} = \frac{21 + 20}{154 + 183} = \frac{41}{337}$

$SE = \sqrt{(0.1217)(0.8783)\left(\frac{1}{154} + \frac{1}{183}\right)} = \sqrt{(0.10686)(0.01196)} = \sqrt{0.001278}$

$\hat{p} = 0.1217$

$\hat{q} = 1 - 0.1217 = 0.8783$

SE Value = __0.03575__

(b) test statistic = $\dfrac{\text{sample data} - \text{pop claim}}{SE}$, For tables use $\sum \dfrac{(Obs - Exp)^2}{Exp}$

3   (z_{test}) or t_{test} or χ^2_{test} (circle) = $\dfrac{(0.1364 - 0.1093) - 0}{0.03575}$

test statistic = __0.758__

STEP 4: Compare the test statistic to the critical value:
the test-statistic is (<) > (circle) than the critical value
STEP 5: We therefore Reject | (Fail-to-reject) (circle) the claim
STEP 6: State a conclusion:

3   statistically there is no difference between the two intersections red light running.

difference could be zero

⑩ BTW:
StatDisk: $z_c = \pm 2.5758$, $z_{test} = 0.75238$, 99% CI : $-0.06566 < P_1 - P_2 < 0.11983$

5. (9.3: 5 pts) The Fair Isaac Corporation (FICO) credit score is used by banks and other lenders to determine whether someone is a good credit risk. Scores range from 300 to 850, with a score of 720 or more indicating that a person is a very good credit risk. An economist wants to <u>determine whether the mean FICO score is lower than the cutoff of 720</u>. She finds that a random sample of 100 people had a mean FICO score of 703 with a standard deviation of 92. *Can the economist conclude that the mean FICO score is less than 720?* **Use Statdisk.com** with α = 0.05 as your level of significance. Fill out your entries below.

(a) Statdisk menu path: Analysis -> _Hyp. Testing_ -> _One Sample Mean_

- Alternative Hypothesis: ⟨<⟩ > or =/= (circle)

- Significance: _0·05_

- Claimed Mean: _720_

- Sample Size (n): _100_
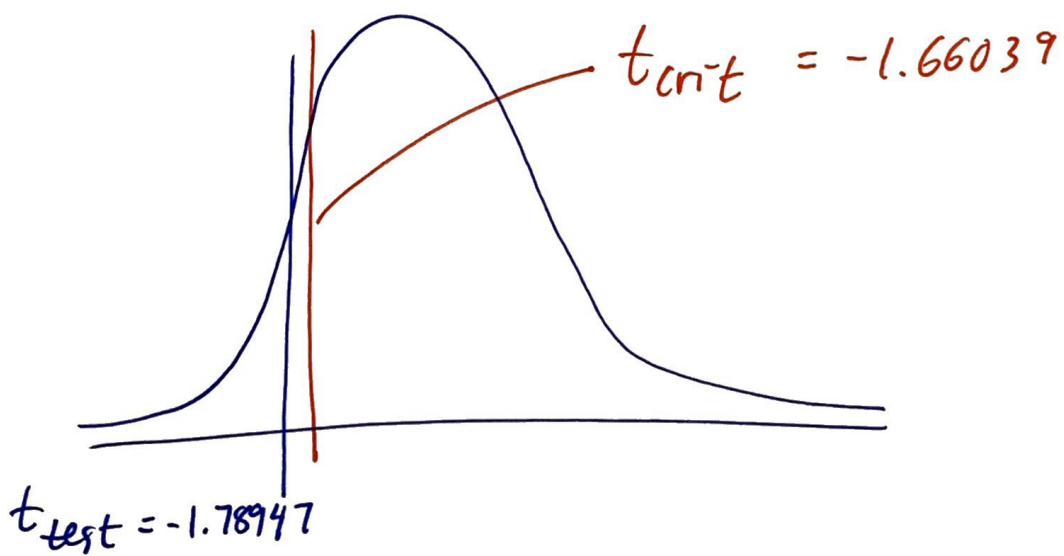
- Sample Mean : _703_
- Sample Std. Dev: _92_

(b) Results: p-value: _____0.03830_____

original
these blanks belonged to
a proportion problem ...
(No one recognized that) ⸪

(c) Answer the question:

The evidence supports the claim that the mean FICO score is less than 720 in this sample of 100.

+5
ext.
credit

$t_{crit} = -1.66039$

$t_{test} = -1.78947$

6. (9.4: 5pts) Interpret computer output: A sample of college students was asked <u>whether they had a job outside of school</u>. The following MINITAB output presents the results of a hypothesis test regarding the proportion of college students who have a job outside of school.

```
Test of p = 0.6 vs p > 0.6
                             95%
                            Lower
    X     N    Sample p     Bound    Z-Value   P-Value
   539   871   0.618829    0.591760    1.13      0.129
```

(a) What are the null and alternate hypotheses? H_0: ___$p = 0.60$___   { current value is 0.60 or 60% work

H_A: ___$p > 0.60$___   { claim: more than 60% work outside of school }

3

(b) What is the value of the sample proportion? $\hat{p}$ = ___0.618829___

$\alpha =$
(c) Can H_0 be rejected at the 0.05 level? Y. or (N)

(d) State a Statistics Class (vs. a Newspaper article) conclusion.

2

We cannot conclude ── , at the $\alpha = 0.05$ level that the proportion of college students who have a job outside of school is greater than 0.6 since the data gathered* could have resulted from statistical variation – dictated by the fact that the p-value was 0.129, not less than α = 0.05

* 0.618829

8. (11.1: 10pts) An automobile manufacturer wants to compare the lifetimes of two brands of tire. She obtains samples of seven tires of each brand. On each of seven cars, she mounts one tire of each brand on each front wheel. The cars are driven until only 20% of the original tread remains. The distances, in thousands of miles, for each tire are presented in the following table.

matched pairs

Car	Brand A	Brand B
1	36.9	34.3
2	45.3	42.2
3	36.2	35.5
4	32.1	31.9
5	37.2	38.1
6	48.3	47.8
7	38.2	33.2

(a) Test the claim that brand A lasts longer than brand B. → ((mean differences)) > 0

i. State and justify the necessary conditions.

Conditions

3

1. matched
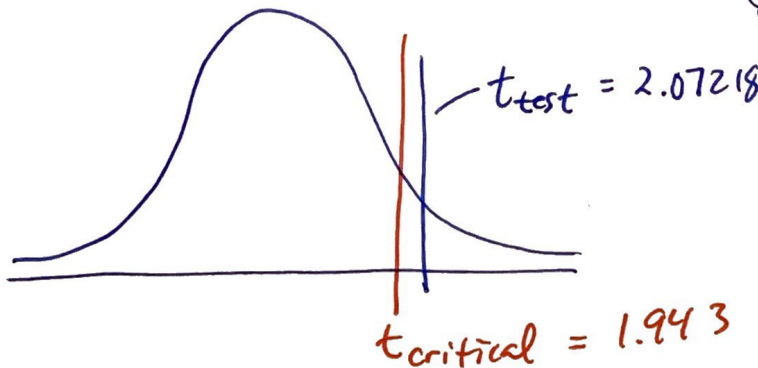2. SRS
3. $n < 10\%$
4. Bell shaped differences
   or
   $n > 30$

Justification

- stated in table
- assumed tires selected randomly
- samples are less than 10% of all tires produced
-

4

ii. Statdisk menu path: Analysis -> **Hyp. Testing** -> **Mean matched pairs**

iii. Statdisk Results (copy all digits): **-0.28934** < $\mu_1 - \mu_2$ < **3.4893** } Conf. Int'vl

barely contains 0 so we could say no difference based on the Conf. Int'vl

iv. Plot with the critical and test statistic:

$t_{test} = 2.07218$

$t_{critical} = 1.943$

(b) Conclusion.

3

Since our test statistic exceeds the critical value associated with $\alpha = 0.05$ we, but barely, we conclude that tire A does last slightly longer than tire B.

10