Problem Solving

Question 1:

A startup's HR department wants to test if their new AI-based hiring tool unfairly rejects qualified female candidates. They form the hypothesis:

- H_0: The tool does not unfairly reject qualified female candidates.
- H_1: The tool does unfairly reject qualified female candidates.

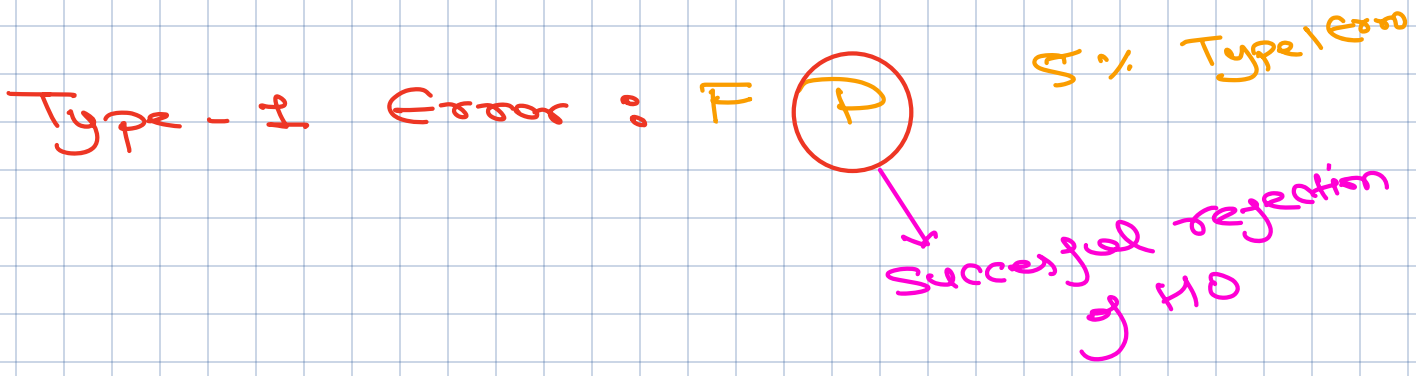
They set a significance level of 0.05 and ran a statistical test. The p-value comes out to 0.07.

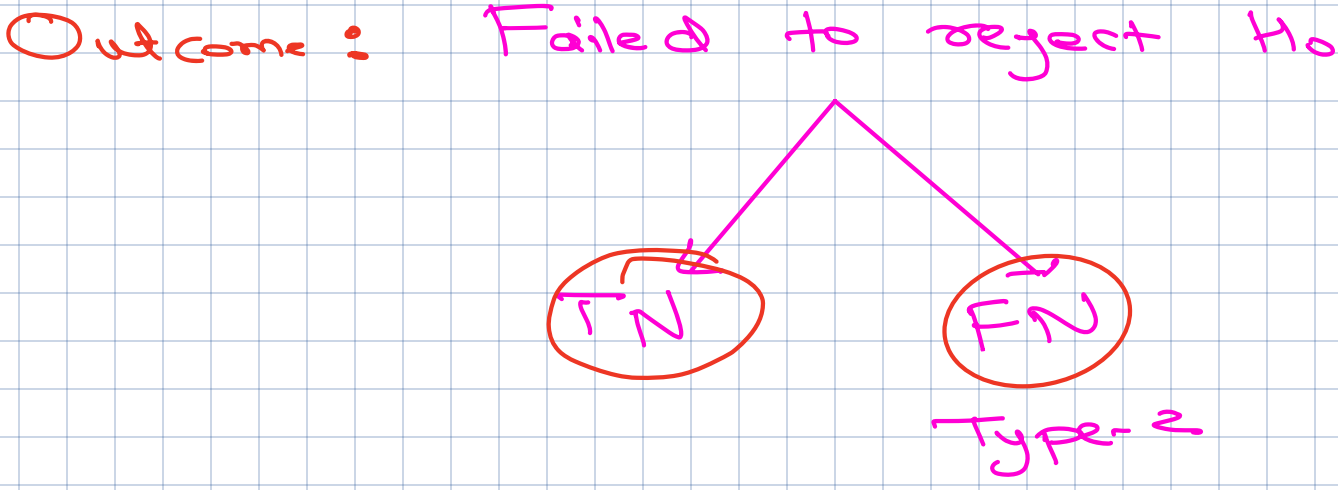
0.07 < 0.05

What is the best interpretation of their decision and possible error?

Options:

A. The null hypothesis is rejected; there's a 7% chance they made a Type I error. ✗
B. The null hypothesis is not rejected; they might be committing a Type II error. ✓
C. They failed to reject the alternative hypothesis; the AI is not unfair. ✗
D. Since the p-value is greater than 0.05, we accept the null hypothesis and the AI is fair. ✗





Question 2:

Scenario:
A popular pizza delivery chain claims their average delivery time at night (10 PM to 2 AM) is 30 minutes or less.
However, customers have recently started complaining on social media that the delivery is taking longer during night hours.
A data analyst decides to perform hypothesis testing using delivery time data from the past 2 weeks to investigate this claim.

Which of the following correctly represents the Null and Alternative Hypotheses for this test?

A.

- $H_0 : \mu = 30 \text{ minutes}$
- $H_1 : \mu \neq 30 \text{ minutes}$

B. ✓

- $H_0 : \mu \leq 30 \text{ minutes}$
- $H_1 : \mu > 30 \text{ minutes}$

C.

- $H_0 : \mu \geq 30 \text{ minutes}$
- $H_1 : \mu < 30 \text{ minutes}$

D.

- $H_0 : \mu \leq 30 \text{ minutes}$
- $H_1 : \mu < 30 \text{ minutes}$

Question 3:

Scenario:

A sports company has developed a new energy drink that it claims improves athlete performance during endurance training.
A group of researchers is conducting a controlled experiment to test whether the new drink actually increases the average endurance time (in minutes) compared to the current benchmark of 40 minutes.
They collect performance data from a sample of athletes and want to use hypothesis testing to validate the claim using the hypothesis framework.

1. Collect sample data
2. Decide on the significance level (α)
3. Calculate the test statistic and p-value
4. Define Null and Alternative Hypotheses (1)
5. Draw a conclusion (last step) (6)
6. Compare the p-value with α

Which of the following is the correct order of steps to set up the hypothesis testing framework in this case?

A. 2 -> 1 -> 3 -> 6 -> 5 -> 4
B. 4 -> 2 -> 1 -> 3 -> 6 -> 5
C. 4 -> 2 -> 5 -> 1 -> 3 -> 6
D. 1 -> 3 -> 6 -> 5 -> 4 -> 2

Question 4:

A health tech startup has developed an AI-based diagnostic tool to detect early-stage lung cancer using CT scan images.
The tool has been trained on historical data and will assist doctors in identifying potential cancer cases.
The company sets a risk threshold for classification. If the prediction score is above the threshold, the patient is flagged as having lung cancer. If it's below the threshold, the patient is classified as not having lung cancer.

Consider the following cases:

Case A:

- Reality: The patient has no cancer.
- Decision: The model correctly predicts no cancer.

Case B:

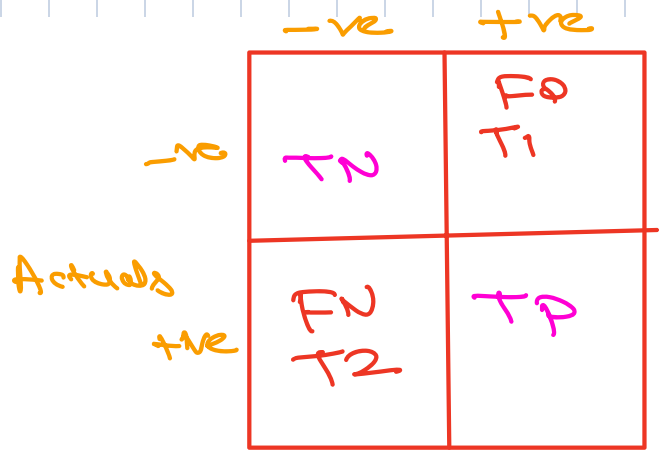
- Reality: The patient has cancer.
- Decision: The model fails to detect it and predicts no cancer.

Case C:

- Reality: The patient has no cancer.
- Decision: The model incorrectly flags the patient as having cancer.

Case D:

- Reality: The patient has cancer.
- Decision: The model correctly predicts cancer.



What labels would you assign to these 4 cases respectively?

A. True Negative, Type II Error, Type I Error, True Positive
B. True Negative, Type I Error, Type II Error, True Positive
C. True Positive, Type II Error, Type I Error, True Negative
D. Type II Error, True Positive, True Negative, Type I Error

Q) Find upper bound of CI (98%)

Question 5:

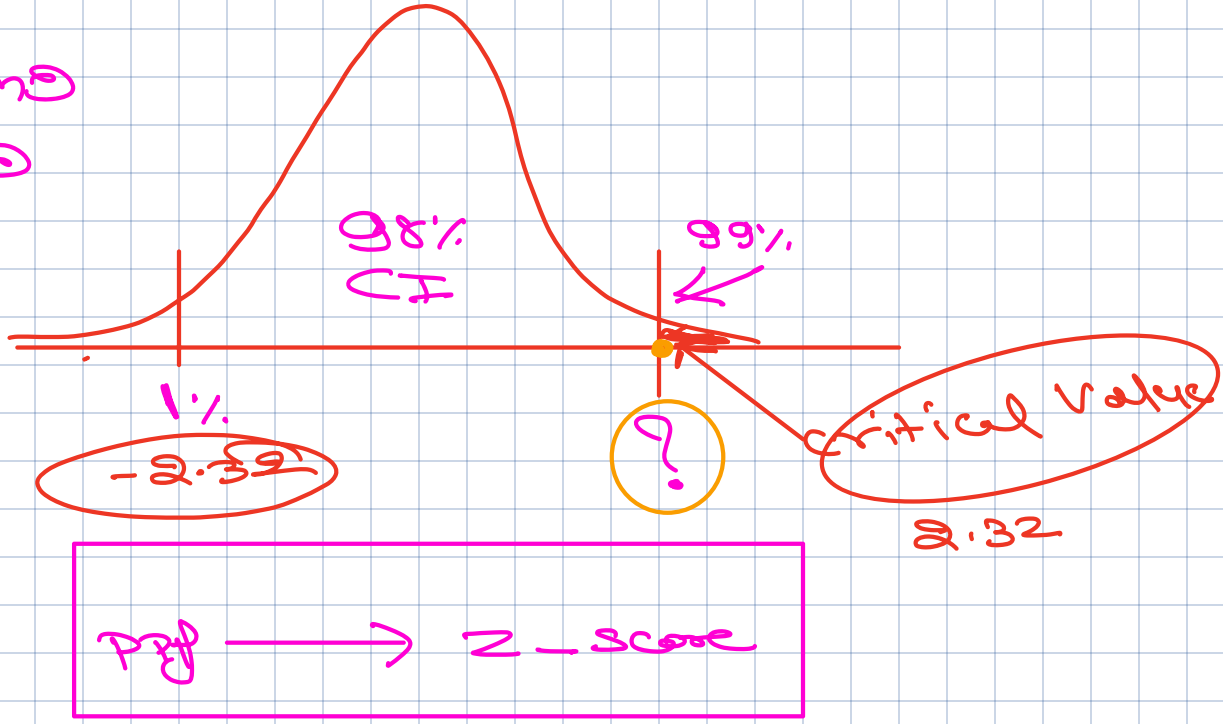
A manufacturing plant is evaluating the average diameter of metal rods produced by a new machine. The rods are expected to have a mean diameter of 5.0 cm, with a standard deviation of 0.2 cm.
To ensure quality, the plant manager wants to find the critical diameter range within which 98% of the rods should fall (i.e., the 98% confidence interval for rod diameter).

Using this information, calculate the critical value for the rod diameter based on a 98% confidence level, assuming the diameter distribution is normal.

Options:

A. 2.5318 cm
B. 0.6532 cm
C. 4.6153 cm
D. 5.4653 cm ✓

Pop - mean
Pop - std



ppf → z-score

Stats.norm.ppf

Actual Value → mean + z * std

z * std − z * std

Question 6:

An analyst claims the average number of website visitors per day is between 1500 and 1700 with 95% confidence.

Which of the following best interprets this confidence interval?

Options:

A. There's a 95% chance that any future sample mean will lie between 1500 and 1700. ✗

B. 95% of all daily visitors will be between 1500 and 1700. ✗

C. 95% of all confidence intervals constructed from similar samples will contain the true mean. ✓

D. There is a 95% chance that the sample mean is 1600. ✗

Doesn't guarantee to be true

We are Confident 95% that the mean Value will be between 1500 - 1700 ← Sampling

If I repeat the experiment 95% containing the Sample will pop mean

Multi Choice

FP & FN

Question 7:

A pharmaceutical company is testing whether a new drug reduces blood pressure more effectively than the current standard. The null hypothesis assumes no improvement. The alternative claims a 5 mmHg improvement. The significance level is set at 0.05, and the power of the test is 0.92.

Risk → 0.05 or 5%

What does this tell us?

Options:

A. There's an 8% chance of a Type 1 error

B. The probability of accepting a false null is 8% ✓

C. The test has a 92% chance of detecting a false null

D. The test has a 92% chance of making a Type I error ✗

Power = 1 − β ⇒ β = 8%

1 − β Error

FN incorrectly rejecting false Null 8%. FN

Question 8:

A pharmaceutical company wants to launch a new generic version of an existing cholesterol-lowering drug. To prove its bioequivalence, it must show that the mean reduction in LDL cholesterol in patients taking the new drug is not significantly different from the existing drug.
They run independent randomized trials for both drugs:

- Existing Drug: Mean reduction = 35.4 mg/dL, standard deviation = 4.2 mg/dL, n = 80
- New Drug: Mean reduction = 34.6 mg/dL, standard deviation = 4.0 mg/dL, n = 75

The company claims bioequivalence, i.e., there's no significant difference.
However, a regulatory body tests the hypothesis at $\alpha = 0.05$.

Which of the following interpretations is most accurate based on a test using p-value?

Options:

A. Fail to reject the null hypothesis. The new drug reduces LDL less than the old drug, but not significantly so.
B. Reject the null hypothesis. There is enough evidence that the new drug is more effective.
C. Fail to reject the null hypothesis. The $p - value$ is greater than alpha, suggesting no significant difference in effectiveness.
D. Reject the null hypothesis. The standard deviations are too close, making the Z-test invalid.

Question 9:

A fitness tracker company wants to test if there's any significant difference in average daily steps between Android users and iPhone users.

They collected random samples:

- Android users: mean = 7,800 steps, std = 1,200 steps, n = 20
- iPhone users: mean = 8,100 steps, std = 1,250 steps, n = 20
- A product manager argues the difference is negligible, while the analyst performs a test assuming equal variances at $\alpha = 0.05$.

Which interpretation is the most statistically accurate?

Options:

A. Reject the null hypothesis. Android and iPhone users walk significantly different steps.
B. Fail to reject the null hypothesis. The p-value shows no significant difference in daily steps.
C. Reject the null hypothesis. iPhone users walk more on average and the difference is practically large.
D. Fail to reject the null hypothesis. The equal variance assumption makes the test invalid.

Question 10:

A marketing analyst compares the effectiveness of two different email subject lines (A and B) used in a product launch campaign.
She records the open rates from two independent groups:

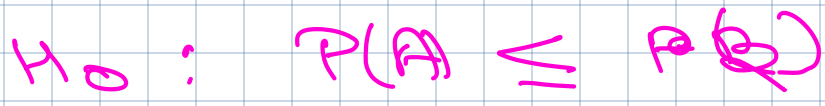
- Subject Line A: 310 out of 1000 recipients opened the email
- Subject Line B: 270 out of 900 recipients opened the email

She wants to test whether there's a significant difference in proportions at $\alpha = 0.05$.
However, her manager concludes, "Subject Line A is more effective when compared to Subject line B."

Which interpretation below is statistically valid based on the p-value from the test?

Options:

A. Reject the null hypothesis. Subject lines have significantly different open rates.
B. Fail to reject the null hypothesis. There is not enough evidence to suggest a significant difference in open rates.
C. Reject the null hypothesis. Subject Line A performs significantly worse.
D. Fail to reject the null hypothesis. The test is invalid due to large sample sizes.



Question 11:

A psychologist is testing whether a cognitive training program improves memory scores in participants. She measures the memory scores of 20 participants before and after completing the program.

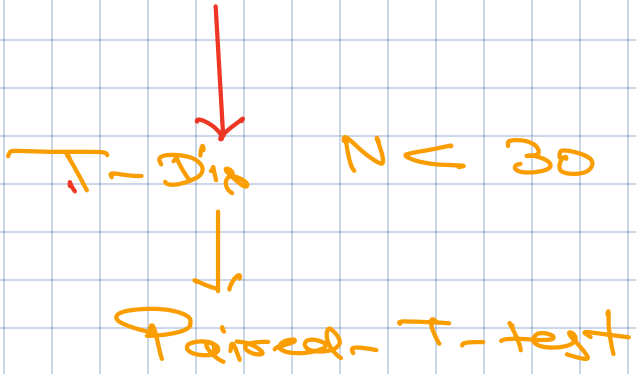
The results show:
before_scores = `[58, 62, 64, 54, 55, 59, 60, 63, 55, 56, 62, 61, 59, 64, 58, 57, 60, 63, 58, 59]`
after_scores = `[63, 67, 68, 57, 60, 64, 61, 66, 58, 59, 64, 63, 60, 67, 61, 60, 62, 65, 60, 61]`

The psychologist wants to performs a test at $\alpha = 0.05$ to test if the program significantly improves scores.

Which of the following interpretations is statistically valid based on the p-value from the test?

Options:

A. Reject the null hypothesis. There is strong evidence that the cognitive training program improves memory.
B. Fail to reject the null hypothesis. There is no statistically significant improvement in memory scores.
C. Reject the null hypothesis. The cognitive training program worsens memory.
D. Fail to reject the null hypothesis. The paired t-test is invalid due to the large differences in variances before and after.



DOF of Contingency Table

Question 12:

A researcher is investigating whether there is an association between education level and employment status among a sample of 400 people. The data collected is as follows:

Education Level	Employed	Unemployed	Total
High School	80	40	120
Bachelor's	100	50	150
Master's	50	30	80
PhD	30	20	50
Total	260	140	400

(num_rows-1) * (num_cols-1)

(4-1) x (2-1)
3x1 = 3

Question 13:

A local theater group wants to test whether the distribution of ticket sales for their five shows matches the expected uniform distribution. The expected number of sales for each show is 200 tickets. After selling tickets for a month, the actual sales for each show are as follows:

- Show 1: 190 tickets
- Show 2: 210 tickets
- Show 3: 195 tickets
- Show 4: 230 tickets
- Show 5: 175 tickets

Observed Data

200
200
200
200
200

They decide to perform a test at $\alpha = 0.10$ to see if the actual sales differ significantly from the expected uniform distribution (i.e., each show is expected to sell 200 tickets).

To check for distribution ?

↓

Goodness of Fit

Expected P%

Observed Data

Question 14:

A market researcher wants to analyze if product preference is independent of age group. The following table summarizes the preferences of 500 customers across three different age groups (18-30, 31-45, 46+ years):

Age Group	Product A	Product B	Product C	Total
18-30	100	50	30	180
31-45	90	60	20	170
46+	60	40	50	150
Total	250	150	100	500

3 Age Cat
3 product Cat

The researcher wants to test if product preference is independent of age group at the $\alpha = 0.05$ significance level.

Which of the following interpretations is statistically valid based on the p-value from the test?

Options:

A. Reject the null hypothesis. Product preference is significantly dependent on age group.
B. Fail to reject the null hypothesis. There is no significant evidence to suggest product preference is dependent on age group.
C. Reject the null hypothesis. Older age groups prefer Product A significantly more than younger age groups.
D. Fail to reject the null hypothesis. The chi-square test is invalid because the expected frequencies are too small.

Chi - Square → Test of Indepence

Chi - Square Ind test

- Test for two or more Categoric of Vars

Goodness of Fit test

- Given a distribution and Expected Value, check if observed data follows distribution

- T-test → $n < 30$ and Pop Std X
  - 1 Sample T-test
  - 2 Sample
    - Ind — Two different
    - Rel/Paired — Same Sample before → After
- Z proportion test
  - Data is given as prop
- Z-test → $n > 30$ or pop std

* T-test and Z-test will give similar results for $n > 30$