EIGHT LECTURE SUMMARY

TEST FOR $:_1 - :_2$

H_0: $:_1 - :_2 = 0$ (EQUIVALENTLY, $:_1 = :_2$)
.....

TEST STATISTIC:
(STANDARD NORMAL)

$$\frac{\bar{x}_1 - \bar{x}_2}{\sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}}}$$

ASSUMPTION: SAMPLES LARGE AND INDEPENDENT OF EACH OTHER

TEST FOR $:_1 - :_2$ (SMALL SAMPLES)

WE NEED TWO EXTRA ASSUMPTIONS:
i) BOTH POPULATIONS ARE NORMAL
ii) AND HAVE THE SAME F

THE TEST STATISTIC THAN CHANGES TO

$$\frac{\bar{x}_1 - \bar{x}_2}{\sqrt{\frac{1}{n_1} + \frac{1}{n_2}} \cdot \sqrt{\frac{(n_1 - 1)s_1^2 + (n_2 - 1)s_2^2}{n_1 + n_2 - 2}}}$$

AND HAS THE t DISTRIBUTION WITH
$n_1 + n_2 - 2$ DEGREES OF FREEDOM

TEST FOR $p_1 - p_2$

ASSUMPTIONS: LARGE INDEPENDENT SAMPLES

H_0: $p_1 = p_2$ THE USUAL NULL HYPOTHESIS

TEST STATISTIC:
$$\frac{\hat{p}_1 - \hat{p}_2}{\sqrt{\frac{1}{n_1} + \frac{1}{n_2}} \cdot \sqrt{\hat{p} \cdot \hat{q}}}$$

WHERE $\hat{p}$ IS THE POOLED SAMPLE PROPORTION, AND $\hat{q} \equiv 1 - \hat{p}$

WHEN H_0 IS TRUE, THE TEST STATISTIC HAS (APPROXIMATELY) THE STANDARD NORMAL DISTRIBUTION (Y CRITICAL VALUES).

PAIRED - DIFFERENCE TEST

FOR EACH SUBJECT, TWO OBSERVATIONS ARE TAKEN (USUALLY 'BEFORE' AND 'AFTER'), AND WE TEST WHETHER THE MEAN POPULATION DIFFERENCE μ_d IS EQUAL TO ZERO (NULL HYPOTHESIS), AGAINST ONE OF THE THREE ALTERNATIVES.

WE MUST FIRST COMPUTE THE DIFFERENCES d_i (AND FORGET THE ORIGINAL DATA).

TEST STATISTIC:

$$\frac{\bar{d}}{s_d/\sqrt{n}} = \frac{\bar{d}}{s_d}\sqrt{n}$$

HAS THE t DISTRIBUTION WITH n - 1 DEGREES OF FREEDOM (STANDARD NORMAL WHEN n IS LARGE).

ASSUMPTION: THE DIFFERENCES FOLLOW A NORMAL DISTRIBUTION (NEEDED FOR SMALL n ONLY).

EXAMPLE:

THE LABEL STATES THAT A BOTTLE OF SPRING WATER CONTAINS 1000 mL OF OUR FAVOURITE DRINK. HAVING A CLOSER LOOK AT A FEW BOTTLES ON A SUPERMARKET SHELF MAKES IT OBVIOUS THAT THE EXACT AMOUNT DIFFERS FROM BOTTLE TO BOTTLE. SO, WHAT THE LABEL MUST MEAN IS THAT THE AVERAGE AMOUNT OF WATER PER BOTTLE IS 1000 mL. WE DECIDED TO TEST THIS CLAIM AT THE 5% LEVEL OF SIGNIFICANCE, COLLECTED A RANDOM INDEPENDENT SAMPLE OF A HUGE SHIPMENT OF SPRING WATER (TECHNICALLY, ONLY THIS SHIPMENT IS THEN OUR POPULATION), AND OBTAINED THE FOLLOWING RESULTS:
995, 1002, 1003, 994, 999, 990, 1000, 987, 1005, 991, 996, 990

H_0: : = 1000 (THE SHIPMENT'S MEAN VALUE IS 1000 mL)

H_1: : < 1000 (THE MEAN VALUE IS LESS THAT 1000 Ml)

BEFORE WE CAN COMPUTE THE TEST STATISTIC, WE NEED: $\bar{x} = \frac{11952}{12} = 996$ AND

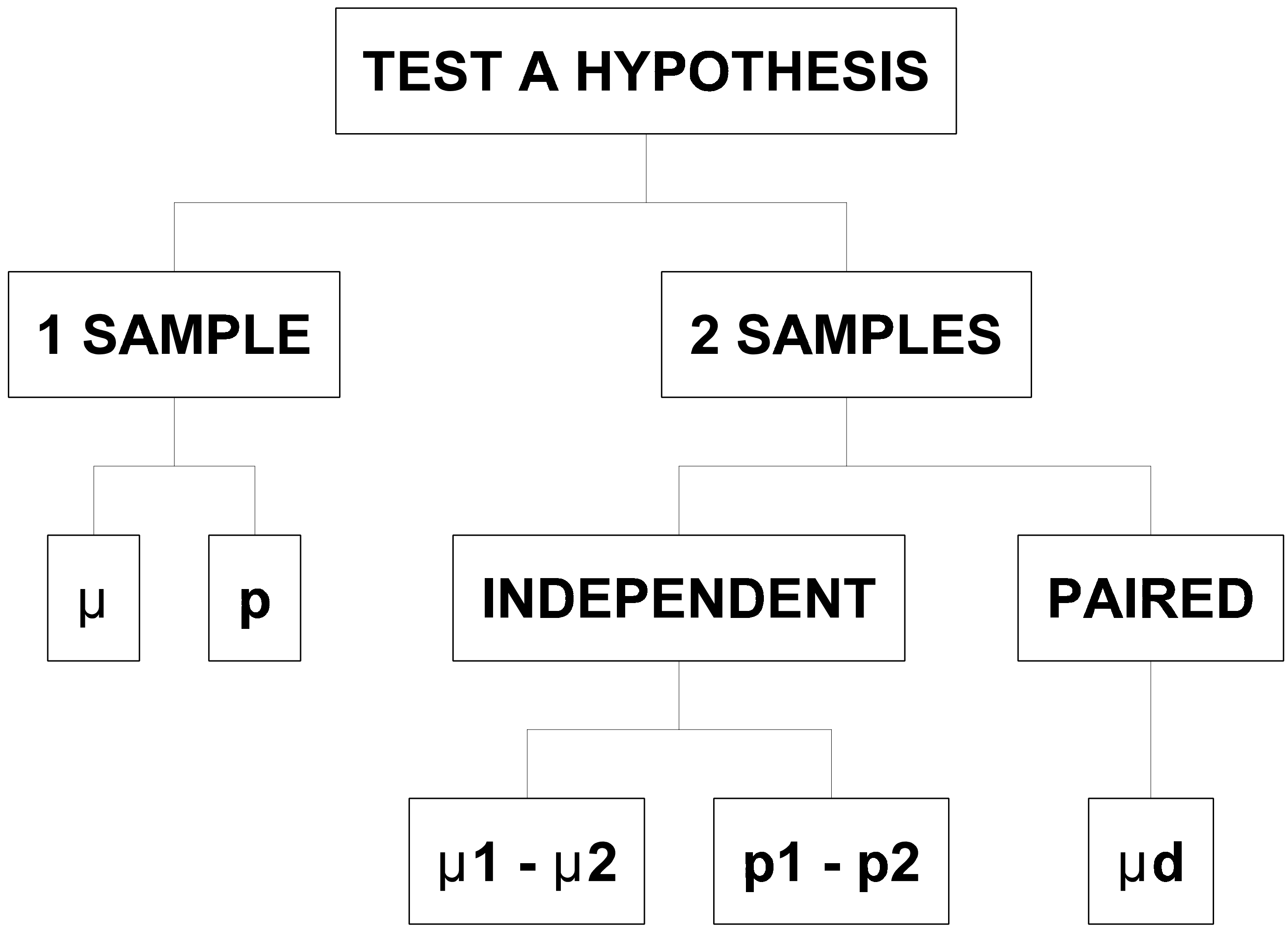
$$\frac{s}{\sqrt{n}} = \sqrt{\frac{11904566 - 11952^2/12}{11 \times 12}} = 1.68325$$

THE VALUE OF THE TEST STATISTIC IS THUS $\frac{996 - 1000}{1.68325}$ = -2.376, WELL BELOW THE CRITICAL VALUE OF -1.645

THE CORRESPONDING P VALUE EQUALS 0.018 (BASED MINITAB'S OUTPUT - WE CAN DO THIS ON OUR OWN ONLY WHEN n > 30). THIS, BEING SMALLER THAN 0.0500, LEADS TO THE SAME CONCLUSION (IT IS ALSO MORE INFORMATIVE, TELLING US THAT WE CAN REJECT H_0 EVEN AT 2% LEVEL OF SIGNIFICANCE - THE LOWER THE ", THAT MORE 'STRONGLY' WE REJECT).

CONCLUSION: WE HAVE A STATISTICALLY SIGNIFICANT EVIDENCE THAT, IN THIS PARTICULAR SHIPMENT, SPRING WATER COMPANY HAS SHORTCHANGED ITS CUSTOMERS.

NOTE THAT, IF THE SAMPLE MEAN TURNED OUT TO BE BIGGER THAN 1000, H_0 WOULD AUTOMATICALLY 'PASS' THE TEST.


TEST A HYPOTHESIS
1 SAMPLE
2 SAMPLES
μ
p
INDEPENDENT
PAIRED
μ1 - μ2
p1 - p2
μd