TWELFTH LECTURE SUMMARY

NONPARAMETRIC TESTS

DON'T REQUIRE THE POPULATION(S) TO BE NORMAL, NOR SAMPLE SIZE(S) TO BE LARGE ($n > 30$)

ARE, IN GENERAL, LESS POWERFUL THAN THEIR PARAMETRIC COUNTERPARTS (I.E. THEY ARE LESS LIKELY TO REJECT H_0 WHEN THE POPULATION DIFFERENCES ARE NOT VERY PRONOUNCED)

WE DISCUSSED THREE SPECIFIC CASES:

SIGN TEST

USED IN THE PAIRED-SAMPLES (BEFORE AND AFTER) SITUATION, WHEN DIFFERENCES CANNOT BE CONSIDERED NORMAL (AND *n* # 30)

WE KEEP ONLY THE SIGN OF THE DIFFERENCES, DISCARDING N.D. (ZERO DIFFERENCE) OBSERVATIONS (CORRESPONDINGLY REDUCING THE VALUE OF *n*)

H_0 STATES THAT THE POPULATION PROPORTION OF THE + SIGNS EQUALS ½, H_1 CAN BE ANY OF THE USUAL 3 CHOICES

TEST STATISTIC: $$\frac{x-0.5}{\sqrt{\frac{0.25}{n}}}$$

WHERE x IS THE PROPORTION OF + SIGNS (THE OLD $\hat{p}$) AND n IS THE REDUCED SAMPLE SIZE

WE ALREADY KNOW THAT, WHEN $n > 10$ AND THE NULL HYPOTHESIS IS TRUE, THIS TEST STATISTIC HAS THE STANDARD NORMAL DISTRIBUTION (z); WE FIND THE CRITICAL VALUES ACCORDINGLY

RANK-SUM (MANN-WHITNEY) TEST

USED IN THE CASE OF TWO INDEPENDENT SAMPLES, TESTING $H_0 : :_1 = :_2$ AGAINST ONE OF THE USUAL ALTERNATES

ASSUMPTION: SAME-SHAPE POPULATIONS (DIFFERENT FROM TEXTBOOK)

WE REQUIRE EACH SAMPLE TO HAVE AT LEAST EIGHT OBSERVATIONS

COMPUTATION OF TEST STATISTIC:

RANK THE VALUES OF POOLED SAMPLES (FROM *1* TO n_1+n_2), KNOW HOW TO DEAL WITH TIES

SUM THE RANKS OF THE SAMPLE-ONE OBSERVATIONS (CALL THIS SUM R)

CONVERT TO $$\frac{R-\mu_R}{\sigma_R}$$

WHERE $$\mu_R = n_1 \frac{n_1 + n_2 + 1}{2}$$

(THE EXPECTED VALUE OF R WHEN H_0 HOLDS), AND $$\sigma_R = \sqrt{\frac{n_1 n_2 (n_1 + n_2 + 1)}{12}}$$

(THE CORRESPONDING STANDARD DEVIATION).

SPEARMAN (RANK) CORRELATION COEFFICIENT

USED IN SITUATIONS OF CHAPTER 10 (DOES THE RESPONSE VARIABLE y INCREASE OR DECREASE WITH x) WHEN THE RELATIONSHIP IS NOT NECESSARILY LINEAR

TO COMPUTE r_s WE FIRST NEED TO RANK, INDIVIDUALLY, x AND y OBSERVATIONS

(TO SIMPLIFY THE ISSUE, YOUR TEXTBOOK SKIPS THIS STEP, ASSUMING THAT BOTH SETS OF OBSERVATIONS ARE ALREADY PRESENTED IN THEIR RANKED FORM)

THEN SUBTRACT THE y RANKS FROM THE x RANKS (d STANDS FOR THE RESULTING DIFFERENCES) AND COMPUTE

$$r_s \equiv 1 - \frac{6\sum d^2}{n(n^2-1)}$$

WHERE n IS THE NUMBER OF THESE DIFFERENCES (THIS TIME, ZERO DIFFERENCES COUNT)

USING THIS r_s, WE CAN TEST (TABLE 9)

H_0: THERE IS NO CORRELATION BETWEEN x AND y

H_1: THERE IS A POSITIVE CORRELATION (OR THE OTHER TWO ALTERNATIVES)