MS Module 10: Paired data and Two Population Proportions & Variances (overview 2nd edition)

(The attached PDF file has better formatting.)

(Readings from the second 2nd edition of the Devore text.)

Reading: §10.3: Analysis of paired data

Some actuarial scenarios have paired data, such as mortality rates for a husband and wife or motor insurance accident frequencies for a husband and wife if both are insured under the policy.

The arithmetic is simpler for paired data than for unpaired data. The textbook discusses the qualitative issues regarding use of paired vs unpaired procedures. Paired data are often positively correlated, as may be true for mortality rates and average claim frequencies. Once the differences are computed, the formulas for the paired t test (directly above Example 10.9) are like those for single sample data.

Examples using small data sets often show normal probability plots; small deviations from a straight line do not invalidate a normal distribution. Figure 10.8 looks slightly curved but not enough to warrant more complex statistical tests.

The textbook says that "whenever there is positive dependence within pairs, the denominator for the paired t statistic should be smaller than for t of the independent-samples test, resulting in a larger test statistic and a smaller P-value. … Similarly, when data is paired, the paired t CI will usually be narrower than the (incorrect) two-sample t CI. This is because there is typically much less variability in the differences than in the x and y values."

Review end of chapter exercises 40 (note the prediction interval in 40b), 41, 42 a and b, 43, and 44

Reading: §10.4: Inferences About Two Population Proportions

The variance of a proportion depends on the mean: variance = n × p × (1 − p), so the variance of the mean = p × (1 − p) / n. For large samples, the Z statistic has

!   estimated proportions in each sample for the numerator and
!   the estimated proportion in the combined sample for the denominator.

Note the formula for the variance in the equations above Example 10.11. This example is clear, with the seven step procedure carefully laid out.

Skip Example 10.13; it requires the lengthy equation 10.7, which is not tested on the final exam.

Read the section "A Large-Sample Confidence Interval for p1 − p2." Example 10.13 is clear and is tested.

Review end of chapter exercises 48, 49, 50, 51, and 52.

Reading: §10.5: Inferences About Two Population Variances

Inferences regarding differences of the variances of two samples use F tests. The confidence intervals use a critical F value on one side and the reciprocal of a critical F value on other side. The F-statistic is the ratio of the sample variances, which depends on which sample is the numerator and which is the denominator. The reciprocal of the F-statistic reverses the samples in the numerator and the denominator.

Expression 10.8 differs several ways from previous expressions.

!   It tests a ratio, not an absolute value or a difference.

! S_1 and S_2 are observed data; σ^2_1 and σ^2_2 are unknown variables.
! To test whether $\sigma^2_1 = \sigma^2_2$, we test whether σ^2_1 / σ^2_2 is close to one.
! The *F* distribution has three parameters, two of which are for degrees of freedom.

Figures 10.10 and 10.11 relate the three parameter *F* distribution to the single parameter *p* value.

Example 10.15 shows why variance tests are little used; few people care about variances in weights of bagels vs muffins. Actuaries are exceptions; they often deal with the variance in different classes. The variance in life expectancy is greater for men than for women, since men have higher death rates for accidents and homicides. Similarly, some psychologists believe that men and women have similar average IQ's but men have greater variance: more geniuses and more idiots.

Review end of chapter exercises 60 and 61 to make sure you know how to use the *F*-statistic. You can check the values with the Excel built-in functions or other statistical software. Review end of chapter exercises 64, 65, and 66.

§10.6 "Inferences Using the Bootstrap and Permutation Methods" is not on the syllabus for this course.