Reference: Chapter 8 of J. L. Devore's 8th Edition **By S. Maghsoodloo**

TESTING a STATISTICAL HYPOTHESIS

A statistical hypothesis is an assumption about the frequency function(s) (i.e., pmf or pdf) of one or more random variables. Stated in simpler terms, a statistical hypothesis is an assumption about the parameter(s) of one or more population(s) such as μ, σ^2, $\mu_x - \mu_y$, p, $p_1 - p_2$ and σ_1^2 / σ_2^2, etc.

Examples. (a) H_0: $\mu = 100$, H_1: $\mu \neq 100$; (b) H_0: $\mu = 5000$ psi, H_1: $\mu > 5000$ psi; (c) H_0: $\sigma^2 = 0.05$, H_1: $\sigma^2 < 0.05$; (d) H_0: $\mu_x - \mu_y = 0$, H_1: $\mu_x - \mu_y \neq 0$; (e) H_0: $\sigma_x^2 / \sigma_y^2 = 1$, H_1: $\sigma_x^2 / \sigma_y^2 \neq 1$; (f) H_0: $\mu = 55$ dB, H_1: $\mu < 55$.

In the above examples, H_0 is called the null hypothesis while H_1 is called the alternative hypothesis (note that Devore uses the notation H_a for the alternative hypothesis, but his notation is not as prevalent in Statistical Literature as is H_1). Further, the above examples indicate that, just like confidence intervals (CIs), there are two types of hypotheses: two-sided and one-sided. The alternative H_1 (or H_a) always determines whether a hypothesis is one or 2-sided. If the statement under H_1 involves $\neq$, then the hypothesis is 2-sided; otherwise, the statement under H_1 will either include < for a left-tailed test, or will include > for a right-tailed test. Therefore, Examples (a), (d) and (e) above constitute 2-sided hypotheses, while Example (b) formulates a right-tailed test, and (c) and (f) are left-tailed tests. Finally, bear in mind that the equality sign "=" generally will be used only in the statement under H_0 (almost never under H_1). In the remaining of this chapter, we will study hypothesis testing about a single normal population parameter μ and σ^2, or about a population proportion p.

Hypothesis testing about parameters of two populations will be solely discussed in Chapter 9 of Devore(8e). Just like in the case of CIs (confidence intervals), the sampling distribution (SMD) of the point estimator of the parameter under H_0 must be used to conduct a parametric test of hypothesis. That is to say, the null hypothesis H_0: $\mu = \mu_0$ must be tested using the SMD of the point estimator $\bar{x}$ when σ^2 is known. The null hypothesis H_0: $\sigma = \sigma_0$ must be tested using the SMD of $(n-1)S^2/\sigma^2$, which is χ^2_{n-1}. The null hypothesis H_0: $\mu = \mu_0$, when σ^2 is unknown, must be tested using the SMD of $(\bar{x} - \mu)\sqrt{n}/S$, which is that of Student's t with $\nu = n - 1$ degrees of freedom (*df*). The critical (or rejection) region of a null hypothesis is that part of range space (or Support-Set) of the test statistic that corresponds to the rejection of H_0. In testing a statistical hypothesis, any one of the following four circumstances may occur.

LOS = Level of Significance, False Positive in Medical Sciences = Type I error in STAT

	H_0 is true	H_0 is false
Reject H_0	Event ≈ " Type I Error, False-alarm, or False Positive"; cell (1,1) ; Occurrence Pr = α = LOS	True Positive (or Sensitivity, or the Power-of-the-test) $\rightarrow$ Correct Decision; cell (1,2) Occurrence Pr = $1 - \beta$
Do not reject H_0 (or Accept H_0)	True Negative (or Specificity of the test in Medicine) $\rightarrow$ Correct Decision Occurrence Pr = $1 - \alpha$	Event ≈ " Type II Error, or False Negative", or accepting a false hypothesis Occurrence Pr = β

Note that once a decision about H_0 is made based on sample data, then the above 4 circumstances reduce to only two possibilities. For example, if a data set provides sufficient evidence to reject H_0, then the experimenter has either committed a type I error or has made a correct decision [cells (1, 1) and (1, 2), respectively]. And vice a versa when sample data does not provide sufficient evidence to reject H_0. Further, the Pr of rejecting H_0 given that H_0 is false is called the power of the statistical test (or the sensitivity of the test), which, from cell (1, 2) of the above table, and is clearly equal to $1 - \beta$.

Finally, the question arises as to when an experimenter should conduct a 2-sided test, and when a one-sided test is warranted? This author recommends a 2-tailed test for all nominal type parameters, the left-tailed test $H_1 : \theta < \theta_0$ for an STB type parameter, and the right-tailed test $H_1: \theta > \theta_0$, if θ is an LTB type parameter. However, exceptions to the above recommendations (especially in cases of the STB and LTB type parameters) do exist. Further, it will generally be best to set up the alternative hypothesis H_1 first by always placing the statement with a question mark (or a manufacturer's claim) under H_1, and the null (or negation) of the "question mark statement" will then be placed under H_0.

Thus far, we have laid down the foundation for testing a statistical hypothesis so that in the subsequent sections we will illustrate the specific procedures through many examples.

TEST of HYPOTHESIS ABOUT the MEAN of a De Moivre (or Normal) POPULATION WITH KNOWN (or SPECIFIED) σ^2

Example 38. A process manufactures ropes for climbing purposes whose breaking strength, X, is normally distributed with known variance σ^2 = 625 psi^2. A consumer group would like to determine if the manufacturer's process mean strength exceeds 1250 psi? (Note that this is the "question mark Statement" and almost always should be placed under H_1). This question leads to the following null and alternative hypotheses: H_0: $\mu \le 1250$, versus H_1: $\mu > 1250$. The null hypothesis in this case is best stated as H_0: $\mu = 1250$. A random sample of size n = 25 has given the value of $\bar{x}$ = 1260, and the question is "does this sample provide sufficient evidence to reject H_0 in favor of H_1 at a pre-assigned type I error probability (Pr) $\alpha = 0.05$? Recall that we have to make use of the sampling distribution of the point estimator $\bar{x}$ as depicted in Figure 13.

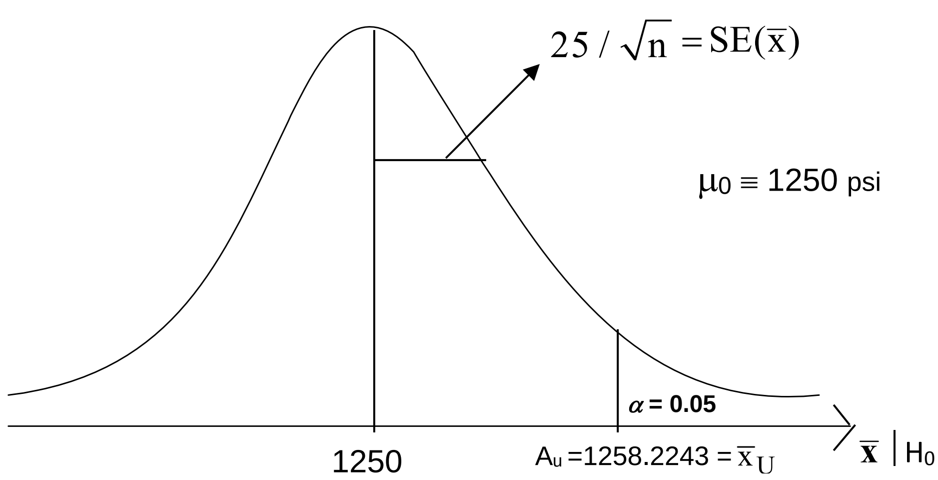

Figure 13. The Sampling Distribution (SMD) of $\bar{x}$ assuming that H_0 is true

Since larger values of $\bar{x}$ (well beyond 1250) lead to the rejection of H_0: μ = 1250 in favor of H_1: μ > 1250, we deduce that the rejection region for testing H_0 is on the right-tail of the $\bar{x}$ sampling-distribution. Therefore, the upper limit of acceptance interval, AI (or decision interval) is $\bar{x}_U$ = A_u = 1250 + 1.645×25/(25)$^{1/2}$ = 1258.2243, and the rejection (or critical) region consists of the interval $\overline{AI}$ = AI′ = [1258.2250, ∞) = 1258.2243 ≤ $\bar{x}$ < ∞. Since $\bar{x}$ = 1260 psi falls inside the rejection region $\overline{AI}$ = AI′ = 1258.22543≤ $\bar{x}$ < ∞, the sample provides sufficient evidence, at the 5% level, to conclude that μ > 1250 psi. The Pr of committing a type I error, α, is also called the level of significance (LOS) of the test, or the size of the critical (or rejection) region.

Exercise 64. For the above example, determine if the sample mean $\bar{x}$ = 1260 provides sufficient evidence to reject H_0 at the 1% level of significance. ANS: No.

The *P*-Value of a Statistical Test

The obvious question is: "Given the result of a sample, what is the smallest LOS at which H_0 can be barely rejected"? For the Example 38, we could reject H_0 at α = 0.05 but could not reject H_0 at α = 0.01 (from Exercise 64). Further, we can barely reject H_0 at the 2.5% level but not at the 2% LOS. After the test statistic is computed, the smallest LOS at which H_0 can be rejected, denoted by $\hat{\alpha}$, is also referred to as the probability level (or the *P-value*) of the test. That is, the *P-value* of a statistical test is the smallest LOS at which H_0 can be barely rejected after a random sample has been drawn and the corresponding sample statistic computed. Therefore, $\hat{\alpha}$ = P($\bar{x} \geq$ 1260) = P(Z ≥ 2) = 0.02275. Note that if H_0 is rejected, then $\hat{\alpha}$ is always less than the corresponding pre-assigned value of LOS α; further, the smaller *P*-value = $\hat{\alpha}$ is, the stronger the evidence is against the validity of H_0, i.e., the more strongly we can reject H_0.

Before leaving this section, we must state the fact that all CIs in the universe are tests of hypotheses in disguise for all possible values of the parameter under H_0. To illustrate this fact, recall the results of Example 36 (pp. 94 – 95 of my STAT 3610 notes), where we obtained the 95% lower one-sided CI for the mean breaking strength to be 1251.7750 ≤ μ < ∞. The rejection of our null hypothesis H_0: μ = 1250 versus (VS) H_1: μ > 1250 at α = 0.05 is consistent with the 95% CI: 1251.7750 ≤ μ < ∞ because the hypothesized process mean value $\mu_0 \equiv$ 1250 is outside the 95%

CI: [1251.7750, ∞). Then, it must be clear by now that the sample mean $\overline{x}$ = 1260 does not provide sufficient evidence at $\alpha = 0.05$ to reject H_0: $\mu = 1252$ VS H_1: $\mu > 1252$ because $\mu_0 \equiv 1252$ is inside the 95% CI $1251.7750 \leq \mu < \infty$.

Note that if one haphazardly obtains the upper one-sided 95% CI for μ, given by $(-\infty, \overline{x} + Z_{0.05} \times SE(\overline{x})] = (-\infty, 1268.225)$, s/he will observe that the hypothesized value of $\mu_0 \equiv 1250$ lies inside the 95% CI: $(-\infty, 1268.225)$, and hence in total contradiction to the decision of rejecting H_0: $\mu = 1250$ at the 5% level.

Exercise 65. Determine if the 99% CI of Exercise 55(a) on page 95 of my STAT3610 notes is consistent with the test result of Exercise 64?

COMPUTING TYPE II ERROR PROBABILITY β as a function a Parameter θ

Since β is the Pr of accepting H_0 if H_0 is false, then when testing H_0: $\mu = \mu_0$, $\mu_0 \equiv 1250$, the Pr statement for β of the above Example 38 is given by

β = Pr[$\overline{x}$ will lie within the acceptance interval: (0, 1258.225) | H_0 is false, i.e., $\mu > 1250$], where by the symbol | we mean "given that or conditioned on ".

To illustrate the computation of β, again consider the Example 38, where H_0: $\mu = 1250$ and H_1: $\mu >$ 1250 psi. Suppose the true process mean value is $\mu = 1255$, i.e., H_0 is now false. Then, what is the Pr of accepting H_0: $\mu = 1250$ with a random sample of size 25 before the sample is drawn given that the population mean $\mu = 1255$ psi $\neq \mu_0$? The SMD of $\overline{x}$ given that H_1 is true (or H_0 is false, i.e., $\mu \neq 1250$) is depicted in Figure 14. From Figure 14 we deduce that the value of

$$\beta(\text{at } \mu = 1255) = \Pr(-\infty < \overline{x} \leq 1258.2250 \mid \mu = 1255) = \Pr\left(\frac{\overline{x} - 1255}{5} \leq 0.6450\right)$$

$$= \Pr(Z \leq 0.6450) = \Phi(0.6450) = 0.74054.$$

Note that for the lower limit of the AI in the above computation of type II error Pr, you may use 0 in lieu of $-\infty$ because Breaking Strength can never be negative.

Figure 14 and the above computation of β(at $\mu = 1255$) show that as the specified value of μ changes, so does the value of β, i.e., Type II Error Pr is a function of the parameter under H_0 (in this case β is a function of μ). For example, if $\mu = 1258.2243$, then the type II error Pr is $\beta = 0.50$; if

μ > 1258.2243, then β < 0.50; if μ = 1270, β = 0.009261; if μ = 1280, β = 0.0000066534, and if μ = 1300, then β is practically 0. Note that the larger is the true value of μ than 1250, i.e.,

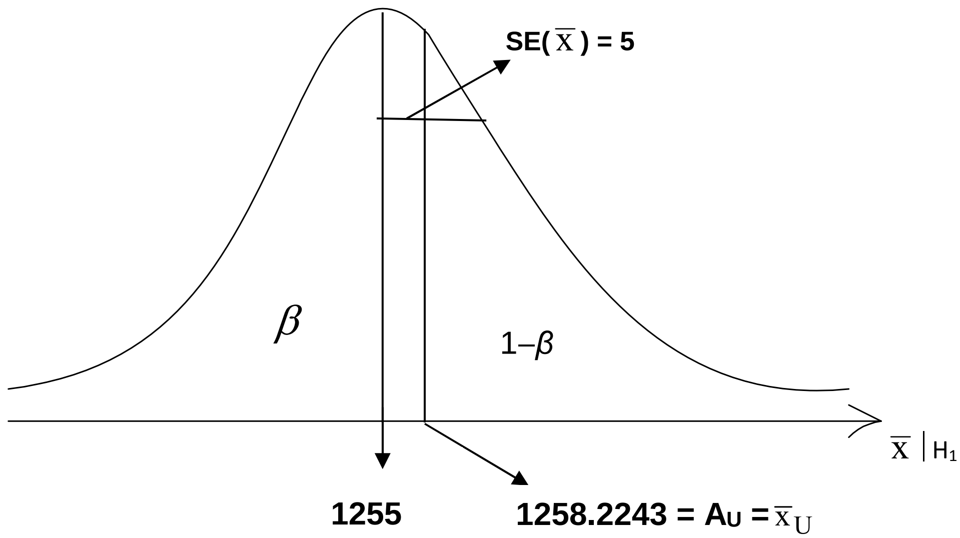

Figure 14. The Sampling Distribution (SMD) of $\bar{x}$ Given that μ = 1255 $\neq \mu_0$

the falser H_0 becomes, then the smaller is the Pr of committing a Type II error. For example, if μ = 1300, then the Pr($-\infty < \bar{x} \leq$ 1258.2243 $|\, \mu$ = 1300) = Pr(Z $\leq$ – 8.355) = $0.0^{11}60$, which is less than 1 in a billion.

The graph of β as a function of the parameter under H_0 is called an Operating Characteristic (OC) curve in the field of Statistics & Quality Control (QC). The OC curve for a 2-sided H_0: $\mu = \mu_0$ is symmetrical about μ_0 with maximum ordinate of $1 - \alpha$ which occurs only at $\mu = \mu_0$. For one-sided tests, $\beta = 1 - \alpha$ at $\mu = \mu_0$, but for right-tailed tests, $\beta > 1 - \alpha$ if $\mu < \mu_0$, and for a left-tailed test $\beta > 1 - \alpha$ if $\mu > \mu_0$.

Exercise 66. (a) For the null hypothesis H_0: μ = 1250 of the above Example 38, draw the OC curve by computing and tabulating the values of β at μ = 1245, 1250, 1253, 1258.225, 1260, 1263, and 1265. (b) Repeat part (a) using the LOS α = 0.01.

It must be clear by now that if H_0 is 2-sided, then the critical region ($\overline{\text{AI}}$ = AI′) of the test is divided equally at the left and right tails of the distribution of the test statistic, and for a left-tailed test the entire value of α is placed on the distribution's left tail and vice a versa for a right-tailed test. Further, always bear in mind that β gives the Pr over the AI, while α = Pr($\overline{\text{AI}}$ = AI′), and almost for all statistical tests 1– $\beta \geq \alpha$. A test for which 1– $\beta \geq \alpha$ is said to be unbiased, and a statistical test for which the value of 1 – $\beta \to 1$ as n $\to \infty$ is said to be consistent

Exercise 67. (a) Work Exercises 9-11 (pp. 309-310), 18-20, p.404 and Ex. 25 on p. 321 of Devore's 8th edition. (b) For the Exercise 10(d), p. 309 derive the minimum value of n if α = 0.05 and β = 0.01 at μ = 1350. Compute the critical (or Pr) levels, or *P*-values, of your tests in all cases.

TEST of HYPOTHESIS ABOUT THE MEAN of a NORMAL (or De Moivre) UNIVERSE WITH UNKNOWN VARIANCE σ^2

If a normal population variance σ^2 is unknown, then we may obtain a point unbiased estimator of σ^2 using the sample statistic S^2. Clearly, $(\overline{x}-\mu)/\sigma_{\overline{X}}$ = $(\overline{x}-\mu)\sqrt{n}\,/\sigma$ has a standardized normal distribution (whether σ is known or not), but herein σ is unknown and has to be estimated by S. However, $(\overline{x}-\mu)\sqrt{n}\,/S$ is not normally distributed, but its sampling distribution follows that of (William S. Gosset's) Student's-t with (n –1) *df*. We illustrate the procedure for testing H_0: $\mu = \mu_0$ at a prescribed LOS α through the following example.

Example 39. A manufacturer claims that the process loudness of its compressors is on the average less than 50 decibels. A random sample of n = 14 compressors has noise levels 53 dB, 48, 49, 57, 45, 46, 54, 47, 49, 50, 48, 44, 46, 51 dB. Our objective is to test H_0: μ = 50, at α = 0.05, VS the alternative H_1: μ < 50 dB, putting the burden of proof on the manufacturer. Note that a claim by a manufacturer should almost always be placed under H_1.

Since this is a left-tailed test, the rejection region of the test statistic $(\overline{x}-\mu)\sqrt{n}\,/S$ is $\overline{\text{AI}}$ = AI′= $(-\infty, t_{0.95,\,13}) = (-\infty, -1.7709)$. This test is left-tailed because smaller $\overline{x}$ values (relative to 50 dB), which in turn lead to smaller values of $(\overline{x}-50)\sqrt{n}\,/S$, will reject H_0 in favor of H_1: μ < 50. For our sample of n = 14 observations we have $\overline{x}$ = 49.07143, S_x = 3.66825 and sample *se*($\overline{x}$) = $S_x / \sqrt{14}$ = 0.9804 so that the value of our test statistic t_0 = ($\overline{x}$ – 50)/*se*($\overline{x}$) = (49.07143 –

50)/0.9804 = – 0.9472. Since $t_{0.95}(13) = -1.7709 < t_0$, we do not have sufficient evidence to reject H_0 at the 5% LOS. AS expected, the critical level of the test $\hat{\alpha}$ = $\Pr(T_{13} \le -0.9472) = 0.1804$ far exceeds the pre-assigned LOS $\alpha = 0.05$, because the sample did not provide sufficient evidence to reject H_0 at the 5% level.

Now suppose the true process mean noise level of all the manufacturer's compressors is μ = 48.5 dB. What is the Pr of rejecting H_0: $\mu = 50$, before the random selection of the sample of size n = 14, if the true mean μ were equal to 48.5 dB? The OC curves in Table A.17, page A-28, graph the values of β VS the abscissa d = $|(\mu-\mu_0)|/\sigma \cong |(\mu-\mu_0)|/S$. Since d = $|48.5-50|$/3.66825 ≅ 0.41, then $\beta \cong 0.57$ so that the Pr of rejecting H_0: $\mu = 50$ is approximately $1 - \beta = 0.43$, i.e., the power of the test at $\mu = 48.5$ is roughly 43%. The exact value of $1 - \beta$ from Minitab is 0.4228 (Open Minitab → STAT → Power and Sample Size → 1-sample t, in the Dialogue Box insert 14 for sample sizes, differences = –1.5, $\sigma \cong 3.67$ → Options → Alternative hypothesis → Less than, and then ok.

Exercise 68. (a) Work Exercises 24, 26 and 29 on pages 321-322 of Devore's 8th edition. (b) For Exercise 24, estimate the power of the test if the true population average $\mu = 2900$ from the OC curves in Table A.17 and then use Minitab to compute the exact value of $1-\beta$.

TEST of HYPOTHESIS ABOUT THE VARIANCE of a NORMAL POPULATION (J. L. Devore covers this test in Exercise 82, p. 344)

Since a process variance σ^2 is an STB type parameter, a CI of the type $0 < \sigma^2 \le \sigma_u^2$ is generally most meaningful, and hence a left-tailed test on σ^2 is the most appropriate. However, many authors quite often conduct a 2-sided test H_0: $\sigma^2 = \sigma_0^2$ VS H_1: $\sigma^2 \ne \sigma_0^2$, or even a right-tailed test on σ^2 where H_1: $\sigma^2 > \sigma_0^2$. For all three alternatives, the statistic $(n-1)S^2/\sigma_0^2$ is used to test H_0: $\sigma^2 = \sigma_0^2$. Recall that the SMD of $\chi_0^2 = (n-1)S^2/\sigma_0^2$ follows a χ^2 (chi-square) with (n –1) *df*, and hence for the alternative H_1: $\sigma^2 < \sigma_0^2$, the critical region is (0, $\chi^2_{1-\alpha,n-1}$); for the 2-tailed test, the acceptance interval for the test statistic (n –1)S^2/σ_0^2 is AI = ($\chi^2_{1-\alpha/2,n-1}$, $\chi^2_{\alpha/2,n-1}$), while for the right-tailed test the rejection region for χ^2_0 is $\overline{AI}$ = AI′ = ($\chi^2_{\alpha,n-1}$, ∞).

Example 40. The % of titanium in an alloy, X, in aerospace castings is N(μ, σ^2), and n = 51 randomly selected parts yielded the standard deviation for the sample S = 0.31, where S^2 = CSS/(n−1). Our objective is to test H_0: σ = 0.25 VS H_1: σ ≠ 0.25 at the LOS α = 0.05. The AI for the test statistic (n –1)S^2/σ_0^2 = (n –1)S^2/ 0.25^2 is AI = ($\chi^2_{0.975,50}$, $\chi^2_{0.025,50}$) = (32.3574, 71.4202), as depicted in Figure 15 below. The value of our test statistic is χ^2_0 = $50(0.31)^2/0.25^2$ = 76.8800, which is outside the AI = (32.3574, 71.4202), and hence we have sufficient evidence to reject H_0 in favor of H_1: σ ≠ 0.25 at the 5% level. Since the test is 2-tailed, the critical level of the test is $\hat{\alpha}$ = *P-Value* = 2×Pr(χ^2_{50} ≥ 76.8800) = 2(0.0086273) = 0.017255, which as expected, is less

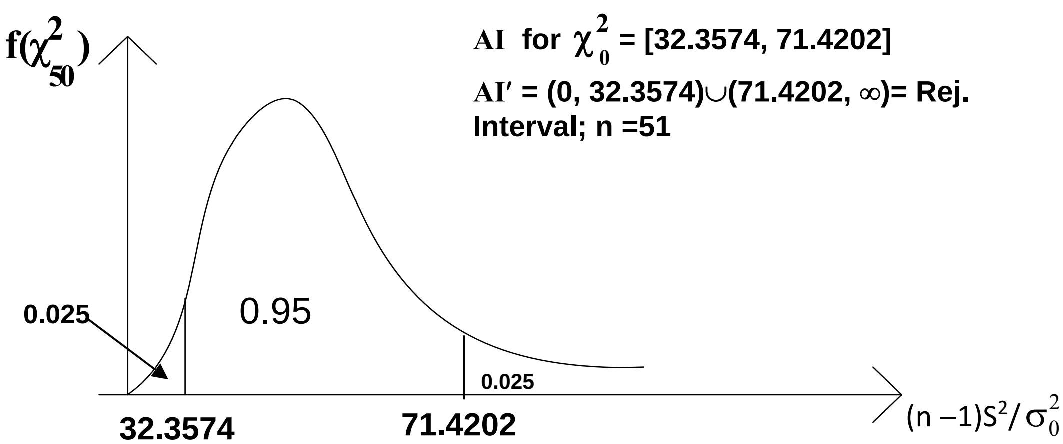

Figure 15. The Sampling Distribution (SMD) of (n –1)S^2/σ_0^2 under H_0

than the prescribed LOS α = 0.05. The corresponding 95% CI is obtained from σ_L^2 = CSS/71.4202 = 0.06728, and σ_U^2 = CSS/32.3574 = 4.8050/32.3574 = 0.14850. Note that the 95% CI: 0.2594 ≤ σ ≤ 0.3854 excludes the hypothesized value of σ = 0.25, congruent with rejection of H_0 at the 5% level.

Although, S is not a linear sum of independent normal variates (or rvs), yet its SMD very slowly approaches the N(σ, $\sigma^2/2n$)-pdf as n $\to \infty$. This property is depicted in Figure 16. It has

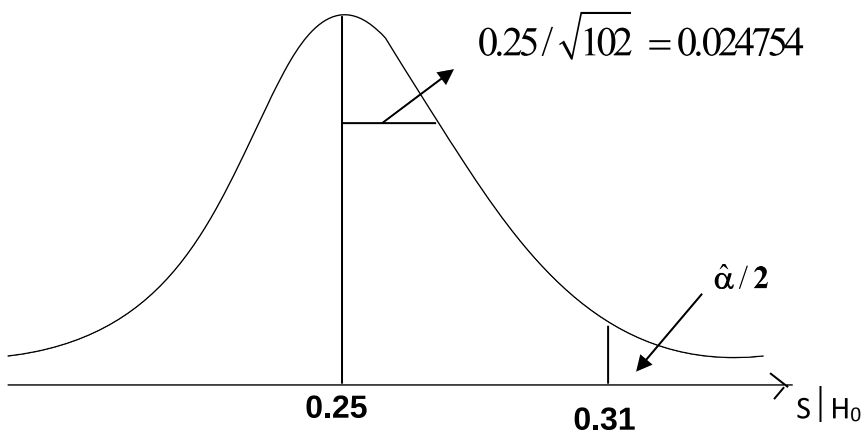

Figure 16. The Approximate SMD of S

been verified by this author that a better approximation for the exact V(S) = $(1-c_4^2)\sigma^2$ from a normal universe is V(S) $\cong \sigma^2/[2(n-0.745)]$, where the QC constant $c_4 = \sqrt{\frac{2}{n-1}} \times \frac{\Gamma(n/2)}{\Gamma(\frac{n-1}{2})}$ = $\sqrt{\frac{n-1}{2}} \times \frac{\Gamma(n/2)}{\Gamma(\frac{n+1}{2})}$, where by definition $\Gamma(n) = \int_0^\infty x^{n-1} e^{-x} dx$. For practical applications, when X ~ N(μ, σ^2), the normal approximation is fairly accurate for n > 60. However, if the underlying distribution is not Gaussian (or De Moivre), then sample sizes larger than 100 may be needed to attain adequate normal approximation to the SMD of S. From Figure 16, $\hat{\alpha} \cong 2\times\Pr(S \geq 0.31) = 2\Pr(Z \geq 2.4239) = 2\times\Phi(-2.4239) = 0.015356$, as compared to the exact value 0.017255. Please observe that even a random sample of size n = 51 is not sufficiently large for an adequate normal approximation of S-SMD.

We now use the above example to illustrate the computation of β when H_0: $\sigma = 0.25$ is False, where n = 51. To this end, suppose that the true value of $\sigma = 0.30 \neq \sigma_0 \equiv 0.25$. Then,

β (at σ = 0.30) = Pr(χ_0^2 is inside the AI | H_0 is false)

$$= \Pr(32.36 \le (n-1)S^2/\sigma_0^2 \le 71.42 \mid \sigma = 0.30)$$

$$= \Pr(32.36 \frac{\sigma_0^2}{\sigma^2} \le (n-1)S^2/\sigma^2 \le 71.42 \frac{\sigma_0^2}{\sigma^2} \mid \sigma = 0.30)$$

$$= \Pr(32.36 \times 0.6944\overline{4} \le \chi_{50}^2 \le 49.59722) \qquad \rightarrow$$

$$= \Pr(22.47222 \le \chi_{50}^2 \le 49.59722)$$

= The cdf of χ_{50}^2 at 49.59722 – the cdf of χ_{50}^2 at 22.47222

= 0.510526 – 0.000271 = 0.510261

The above exact Pr of type II error could also be approximated from the OC curves provided by some authors in the field of Pr & Statistics for testing H_0: $\sigma = \sigma_0$, when the underlying parent is that of De Moivre's. Since both the type II error Pr and power of testing H_0: $\sigma = \sigma_0$ can exactly be computed as shown above by me, such OC curves are not needed. Minitab 16 does provide power values for a χ^2 test on σ^2, and at λ = 1.20, Minitab16 gives $1-\beta$ = 0.489739.

<u>Exercise 69.</u> (a) For the above example, use MS Excel to compute the Pr of a type II error for testing H_0: σ = 0.25 (with n = 51) if the true value of σ were equal to 0.33. (b) For the Example 40, obtain a 95% CI for σ and compare your CI with the above test result and draw conclusions. ANS: $0.25938 \le \sigma \le 0.38540$. (c) For the Example 40, derive the necessary sample size n such that the Pr of type II error at $\lambda = \sigma/\sigma_0$ = 0.30/0.25 = 1.20 is reduced from 0.510526 to β = 0.20, where Pr = Probability.

TEST of HYPOTHESIS ABOUT a POPULATION PROPORTION (p)

To illustrate the procedure for testing H_0: $p = p_0$ VS one of the 3 alternatives H_1: $p < p_0$, H_1: $p \ne p_0$, or H_1: $p > p_0$, we consider Devore's Example 8.11 on pp. 335-6 of his 5th edition (not listed in his 8th ed.), which gives the results of a survey of 102 doctors, only 47 of whom knew the generic name for the drug Methadone. The objective was to determine if this sample of size n = 102 provided sufficient evidence to conclude, at the 5% LOS, that less than half of all Doctors knew the generic name for Methadone. Before proceeding, however, you will need to review my STAT 3610 notes pp. 99-101. As stated at the beginning of this chapter, the "? statement " should be placed under H_1. In this example, n = 102 doctors were randomly surveyed only 47 of whom knew the generic name for Methadone, i.e., there were 47 successes in n = 102 Bernoulli trials. Therefore, a point unbiased estimate of the population proportion (of all doctors), p, is $\hat{p}$ = 47/102 = 0.46078. The question is "does this sample provide sufficient evidence at the LOS $\alpha = 0.05$ to warrant the rejection of H_0: p = 0.50 in favor of H_1: p < 0.50"? The approximate large-sample SMD of $\hat{p}$ is depicted in Figure 17. Figure 17 clearly shows that $A_L = 0.50 - 1.645 \times 0.04951 = 0.4186$ so that the AI = (0.4186 , 1). Since the sample test statistic $\hat{p}$ = 0.461 is well inside this AI,

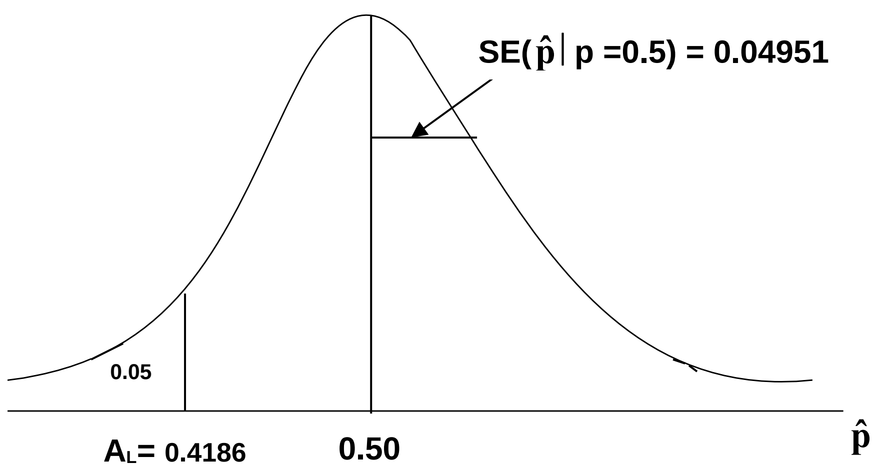

Figure 17. The Approximate SMD of $\hat{p}$ |H_0

the null hypothesis cannot be rejected at the 5% LOS. The critical (or Pr) level of this test is given by $\hat{\alpha}$ = *P*-value ≅ Pr($\hat{p} \le 0.46078$) = Pr(Z ≤ – 0.79221) = 0.21412, which as expected, far exceeds the pre-assigned LOS of 5%. The 95% CI for p is given by $0 \le p \le 0.5420$, which clearly agrees with the above test result because the hypothesized value of p, $p_0 \equiv 0.50$, is well inside this CI : $0 \le p \le 0.5420$. Note that *se*($\hat{p}$| p =0.5) = $(0.25/102)^{1/2}$ = 0.04951.

If the normal approximation to the binomial is not adequate, then the binomial test on a population proportion must be conducted using the Bin(n, p) distribution directly. To illustrate, I will compute the exact critical level of the test for the Example 8.11 of J. L. Devore's 5th edition. The procedure consists of using MS Excel to compute $\hat{\alpha}$ from $\hat{\alpha}$ =

$\sum_{x=0}^{47} {}_{102}C_x\,(0.50)^x\,(0.50)^{102-x}$ = BINOMDIST(47, 102, 0.50, True) = 0.24417232 = Matlab's binocdf(47, 102, 0.50). Note that the approximate value of $\hat{\alpha}$ = *P-value* = 0.21412 from the normal approximation is fairly inadequate because the correction for continuity of 1/(2n) was not applied to obtain $\hat{\alpha}$ = 0.21412. If we apply the correction for continuity in the normal approximation to the binomial, we will obtain $\hat{\alpha} \cong$ Pr($\hat{p} \le$ 0.46078 + 1/204) = Pr($\hat{p} \le$ 0.4657) = Pr(Z $\le$ – 0.6931) = Φ (– 0.6931) = 0.24412, which is almost identical to the exact value of 0.24417. Therefore, if n < 50, use the exact procedure for the binomial test of proportion by computing the exact *P-Value* from MS Excel and compare $\hat{\alpha}$ against α = 0.05. If $\hat{\alpha}$ < 0.05, then reject H_0: p = p_0; otherwise, the data does not cast sufficient doubt about the validity of H_0.

Exercise 70. Work Exercises 37, 38, and 39 on p. 327 of the 8 th edition of Devore and Exercises 73 on page 343 of Devore.

Computing Type II Error Pr when H_0 : p = p_0 Is False

Suppose in the above Example, where X represented the number of doctors who knew the generic name for Methdone, the true value of p were equal to 0.40, i.e., H_0 was false. Then, from Figure 18 we deduce that β = Pr (0.4186 $\le \hat{p}$ < 1| p = 0.40) = Pr($\frac{0.4186-0.40}{0.0485} \le$ Z) = Pr(Z $\ge$ 0.3834) = Φ(– 0.3834) = 0.3507. Note that 0.0485 = $\sqrt{0.4(0.6)/102}$ = *se*($\hat{p} | p=0.40$), and that the power of the test at p = 0.40 is given by 1 – β = 0.6493.

You should always bear in mind that all CIs are simply tests of hypotheses in disguise! Therefore, to be consistent, the lower one-sided CI: $\theta_L \le \theta < \infty$ always corresponds to a right-tailed test on the parameter θ (i.e., the alternative is H_1= H_a: $\theta > \theta_0$) and vice a versa for an upper one-sided CI.

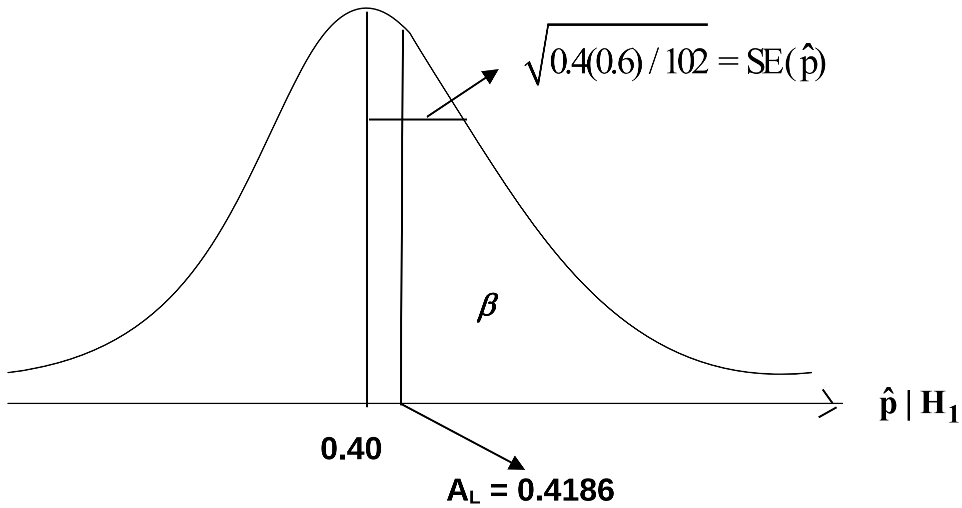

Figure 18. The Approximate SMD of $\hat{p}$ Given that the true value of of $E(\hat{p} \mid H_1)$ = p = 0.40, where A_L stands for lower-limit of decision interval

STAT3610/3611: Please note that throughout all my courses, I use SE for the population (or process) Standard-Error, while I use *se* as the Sample Standard-Error. The terminology SE (or *se*) must be used in lieu of STDEV when external variability is being estimated. For example, the variability of the estimator sample mean $\bar{x}$, of a process mean μ, must be referred to as the *se*(Xbar); referring to $S/\sqrt{n}$ as the STDEV of $\bar{x}$, would be totally inappropriate.

Finally, it can be statistically proven that sample sizes $2 \le n \le 20$ must be considered as small; those within $21 < n \le 60$ should to be considered as moderate. Sample sizes n > 60 should be considered as large.