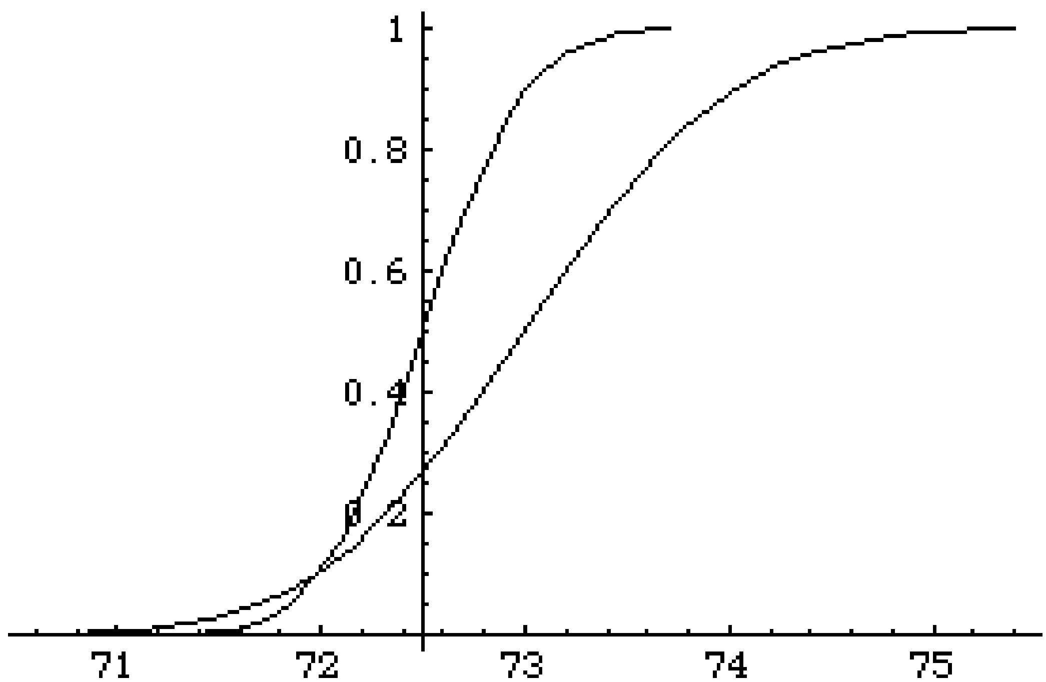
Monday 4-17-06 a representative of a local company will describe an internship opportunity for a person knowing some statistics. Course evaluation forms will be distributed. I will go over review questions (below). No review can cover everything, but this one touches a lot of bases.

1. Sales have traditionally averaged 72 per period. Consider a z-test of the hypothesis H0: {mu does not exceed 72} versus the alternative H1: {mu > 72}, with alpha = 0.1, based upon a with replacement sample of n = 50. Rejecting H0 requires some relatively strong evidence and so would indicate sales are likely rising.

a. Make a sketch illustrating the general appearance of P(reject H0 | mu) as a function of mu. Since we don't know the population sd sigma it is impossible to say how steeply the curve rises but be sure to get the general shape right and alpha right (identify alpha appropriately). **NEW: the steepest rise is roughly where the curve is at height 0.5 so, from now on, try to achieve that with your sketch to make it a bit more realistic.**



ANS. I set this up on a computer assuming a normal population and a population sigma = 5.67. It is the less steep curve of the two above, achieving its steepest rise at mu = 73. Alpha = 0.1 is seen by the heights of these two test curves at mu0 = 72. The ideal plot, obtainable only by census, is identically zero on H0: {mu less or equal 72} and identically one on H1: {mu > 72}. You will notice that these curves, which

are accurately drawn to reflect one sample size being 50 and the other 200 (four times greater) clearly show steepness that is greatest where the respective curves attain height 0.5. For the tests illustrated these are seen to be mu = 72.5 and mu = 73 respectively. If you look at the steepness of the respective slopes at these points you will see that one is twice as steep as the other, which is what happens when one sample is four times the other and we use the same alpha for both tests.

b. Overlay on your sketch (a) the IDEAL plot of P(reject H0 | mu), which cannot be achieved without a census of the population.

ANS. The ideal plot is zero everywhere on H0 and one everywhere on H1. You should draw it by hand in the above picture and also indicate H0, mu0, and H1 on the horizontal axis.

c. Overlay on your sketch (a) a sketch of P(reject H0 | mu) for a test based upon a sample of n = 200 (four times 50). Would it rise twice as fast as the original sketch? **NEW: the steepest rise is roughly where the curve is at height 0.5 so, from now on, try to achieve that with your sketch to make it a bit more realistic.**

ANS. It is the steeper curve, for sample size 200, which attains height 0.5 at mu = 72.5, nearer to mu0=72 than is the steepest point 73 of the original curve for sample size 50.

d. From your sketch (a) determine the approximate power at mu = 73.

ANS. It is the height at mu = 73, which happens to be 0.5.

e. From your sketch (a) determine the approximate probability of an error of type II (i.e. beta) at mu = 73.

ANS. It is one minus the power which happens to be beta = 1 – 0.5 = 0.5.

f. From the z-table, determine the threshold against which the test statistic will be compared when deciding whether or not to reject H0.

ANS. For a one sided test with H0 on the left of H1 it is z(alpha) = z(0.1) = 1.282 and we reject H0 if the test statistic exceeds 1.282.

Had H1 been to the left we would instead have used threshold –1.282 and reject H0 if the test statistic falls below –1.282.

Had the test been two sided we would instead use z(alpha/2) = z(0.05) = 1.645 and reject H0 is the ABSOLUE VALUE of the test statistic exceeds 1.645.

g. Suppose the sample mean is xBAR = 71. What action will be taken by any (reasonable) test? Why?

ANS. Since 71 belongs to H0 no reasonable test will reject H0. This is because we typically control alpha to low levels (i.e. the chance of rejecting H0 when it is true is kept small).

h. Suppose the sample mean of 50 is xBAR = 72.7 with sample standard deviation s = 8.6. What is the value of the test statistic and what action is taken by the z-test?

ANS. The z-test statistic is (72.7-72) / (8.6 / root(50)) = 0.575552 which does not exceed the threshold 1.282 so we fail to reject H0. This is slight evidence against H0.

i. Suppose the sample mean of 50 is xBAR = 75.2 with s = 2.5. What is the value of the test statistic and what action is taken by the z-test?

ANS. The z-test statistic is (75.2-72) / (2.5 / root(50)) = 9.05 which resoundingly exceeds the threshold 1.282 so we reject H0. This is extremely strong evidence against H0.

Note: A test based on a really large sample n will be almost certain to reject H0 for mu which are even slightly greater than 72, but that test will still have only alpha = 0.1 chance of rejecting H0 when mu is exactly 72 (unless we reset alpha to some lower value when making up the test).

j. If the population scores are normally distributed then the t-test is applicable. If the sample has instead n = 4 what is the t-threshold against which the test statistic will be compared when deciding whether to reject H0?

ANS. The threshold is t(alpha) = t(0.1) = 1.638 for degrees of freedom 4 – 1 = 3.

k. Refer to (j). Suppose the sample = {72.4, 72.9, 72.5, 73.6}. Calculate the test statistic and determine the action taken by the t-test.

ANS. (xBAR – 72) / (s / root(4)) = (72.85 – 72) / (0.544671 / root(4)) = 3.12 which exceeds the threshold 1.638 so we reject H0. If you look into the t-table for DF 3 you will see that pSIG is close to 0.025 which is rather strong evidence against H0.

l. Refer again to the z-test for n = 50 and data (h). Determine pSIG.

ANS. Data (h) give test statistic z = 0.58. Remember this test rejects H0 for large values of the test statistic so pSIG is P(Z > 0.58) = 0.5 – 0.219 = 0.281. So there is around a 28% chance we would see data more disagreeable with H0 than this just by luck of the draw (bad sample) when mu = 72. Again, there is not much evidence at all against H0.

Note. In terms of pSIG for ANY test, we reject H0 only if pSIG < alpha. In the present case we fail to reject H0 because 0.281 is not less than 0.1. This is just another way to execute the test and gives exactly the same result as (h).

You might try the pSIG approach for (i) as well. Of course z = 9.05 is off the z-table. But we have P(Z > 9.05) ~ $e^{-9.05^2}$ / (9.05 root(2 Pi) = 7.23 10^{-20} which is truly rare. Since pSIG is less than 0.1 we (resoundingly) reject H0 at the alpha = 0.01 level. This points out a reporting issue for tests with some fixed alpha, in that the data may have a lot more to say than simply "the hypothesis was rejected for alpha = 0.01." That is why many people like to have a look at pSIG.

m. Refer again to the z-test for n = 50 and data (h). Determine a sample size n (which would most often be greater than your 50 but will sometimes not require additional samples) required to achieve a z-test retaining alpha = 0.1 at mu = 72 but also achieving beta = 0.2 at mu = 73.

ANS. The generic formula is n = ((| z_0| + | z_1|) s_0 / (mu_0 – mu_1) $)^2$ in which | z_0| = z(alpha) = 1.282 and | z_1| = z(beta) = 0.84 (enter 0.5 -0.2 = 0.3 to the body of the z-table and read off z). So n = ((1.282+ 0.84) 8.6 / (72 – 73) $)^2$ = 334.

Aside. For data (i) we'd instead have obtained n = ((1.282+ 0.84) 2.5 / (72 – 73) $)^2$ = 29. We already have n = 50 so our current test would, in case (i), not require increasing the sample, having already achieved desired alpha and beta specifications.

Note. For a two-sided test we instead use z_0 = z(alpha / 2) but retain z_1 = z(beta).

n. Sketch P(reject H0 | mu) for the test of (m). **NEW: the steepest rise is roughly where the curve is at height 0.5 so, from now on, try to achieve that with your sketch to make it a bit more realistic.**

ANS. Your curve must pass through (72, 0.1) and (73, 0.8) (power = 1 - beta remember) and must have its steepest rise where it attains height 0.5. It is increasing to the right as are the other curves above and has the characteristic appearance.

o. Refer to (m). Suppose that n = 1024 (it is not) and that you do continue sampling, taking an additional 1024 - 50. Suppose also that the overall sample mean of the 1024 is xBAR = 72.8. Determine the test statistic for the z-test (m) and the action taken by the test.

ANS. Remember, we use $s_0 = 8.6$ which is the sample sd from the initial sample of 50 reported in (h). The test statistic is

(xBAR – 72) / (s_0 / root(1024)) = (72.8 – 72) / (8.6 / root(1024)) = 2.98.

Since this exceeds z(0.1) = 1.282 we reject H0. Note that pSIG = P(Z > 2.98) = 0.5 – 0.4986 = 0.0014. Since pSIG = 0.0014 < alpha = 0.1 we reject H0 by this criterion as well (the two approaches to testing always agree).

p. **Refer to the t-test with data (k).** Determine a sample size n (which would most often be greater than your 4 but will sometimes not require additional samples) required to achieve a t-test retaining alpha = 0.1 at mu = 72 but also achieving beta = 0.05 at mu = 73.

ANS. The generic formula is n = $((|t_0| + |t_1|) s_0 / (mu_0 - mu_1))^2$ in which $|t_0|$ = t(alpha) = 1.638 and $|z_1|$ = t(beta) = 2.353, n = $((1.638+ 2.353)\ 0.544671 / (72 - 73))^2$ = 5. We already have n = 4 so we need only one additional sample to achieve the desired alpha and beta specifications.

Note. For a two-sided test we would instead have used t_0 = t(alpha / 2) but retained t_1 = t(beta).

q. Refer to (p). Suppose n = 1024 (it is not) and that you do continue sampling, taking an additional 1024 - 4. Suppose also that the overall sample mean of the 1024 is xBAR = 72.8. Determine the test statistic for the t-test (p) and the action taken by the test.

ANS. Remember, we use $s_0 = 0.544671$ which is the sample sd from the initial sample of 4 reported in (k). The test statistic is

(xBAR – 72) / (s_0 / root(1024)) = (72.8 – 72) / (0.544671 / root(1024)) = 47.

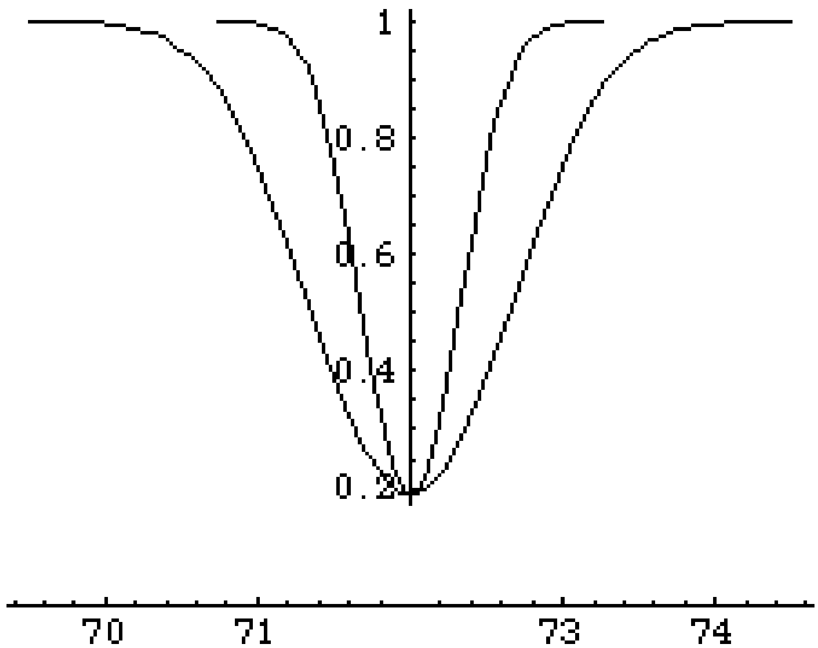
Since this far exceeds t(0.1) = 1.638 we resoundingly reject H0. Note that pSIG = P(t > 2.98) for DF = 3. This is of course tless than alpha = 0.1 so we resoundingly reject H0 by the pSIG method also.

Note. We saw from (p) that the actual sample size required was only n = 5 not 1024.

2. Workers are scored d = production this year – production last year. New production equipment introduced this year was justified as likely to improve production by 72 on average. Consider a z-test of the hypothesis H0: {mu = 72} versus the alternative H1: {mu is not equal to 72}, with alpha = 0.2, based upon a with replacement sample of n = 50. Rejecting H0 requires relatively strong evidence and so would indicate that the impact of the new equipment has likely differed from 72 in one direction or the other.

NOTE. We may lapse into referring to d scores as x-scores since difference scores d fall under the same analysis as typical x scores.

a. Make a sketch illustrating the general appearance of P(reject H0 | mu) as a function of mu. Since we don't know the population sd sigma it is impossible to say how steeply the curve rises but be sure to get the general shape right and alpha right (identify alpha appropriately). **NEW: the steepest rise is roughly where the curve is at height 0.5 so, from now on, try to achieve that with your sketch to make it a bit more realistic.**



ANS. The steeper curve is for n = 200; the shallower curve is for n = 50.

b. Overlay on your sketch (a) the IDEAL plot of P(reject H0 | mu), which cannot be achieved without a census of the population.

ANS. The ideal plot is zero for mu = 72 and is one for all other mu.

c. Overlay on your sketch (a) a sketch of P(reject H0 | mu) for a test based upon a sample of n = 200 (four times 50). Would it rise twice as fast as the original sketch? **NEW: the steepest rise is roughly where the curve is at height 0.5 so, from now on, try to achieve that with your sketch to make it a bit more realistic.**

ANS. Look at the places there the curves each rise to 0.5 height. The steeper one should be twice as steep (at its place) than the shallower curve is (at its place).

d. From your sketch (a) determine the approximate power at mu = 73.

ANS. It looks to be ~ 0.75, the approximate height of the curve (a) at mu = 73.

e. From your sketch (a) determine the approximate probability of an error of type II (i.e. beta) at mu = 73.

ANS. beta = 1 – power is around 0.25 at mu = 73.

f. From the z-table, determine the threshold against which the test statistic will be compared when deciding whether or not to reject H0.

ANS. For the two sided z-test it is z(alpha / 2) = z(0.2 / 2) = z(0.1) = 1.282.

g. Suppose the sample mean is xBAR = 72.00001. What action will be taken by any (reasonable) test? Why?

ANS. This xBAR is so close to H0 that we would not reject H0 unless the sample size n is so large as to be effectively a census.

h. Suppose the sample mean of 50 is xBAR = 69.1 with sample standard deviation s = 1.6. What is the value of the test statistic and what action is taken by the z-test?

ANS. (xBAR – 72) / (s / root(50)) = ((69.1 – 72) / (1.6 / root(50)) = -12.82.

We reject H0 because the ABSOLUTE VALUE of the test statistic exceeds threshold 1.282. In this case pSIG = 2 P(Z > 12.82) is extremely small. So we reject H0 (as well) on the grounds that pSIG < alpha = 0.2. This is extremely strong evidence against H0.

i. Suppose the sample mean of 50 is xBAR = 75.2 with s = 4.5. What is the value of the test statistic and what action is taken by the z-test?

ANS. Test statistic (xBAR – 72) / (s / root(50)) = ((75.2 – 72) / (4.5/ root(50)) = 5.03. Reject H0 since the ABSOLUTE VALUE of the test statistic exceeds z(alpha / 2) = 1.282 for this two-sided z-test.

j. If the population scores are normally distributed then the t-test is applicable. If the sample size is instead n = 4 what is the t-threshold against which the test statistic will be compared when deciding whether to reject H0?

ANS. t(alpha / 2) = t(0.1) = 1.638 for DF = 3.

k. Refer to (j). Suppose the sample = {72.4, 72.9, 72.5, 73.6}. Calculate the test statistic and determine the action taken by the t-test.

ANS. (xBAR – 72) / (s / root(4)) = (72.85 – 72) / (0.544671 / root(4)) = 3.12 which exceeds the threshold 1.638 so we reject H0. If you look into the t-table for DF 3 you will see that pSIG is close to 0.025 < 0.2 = alpha, rather strong evidence against H0.

l-1. **Note, should refer to (i)**. Refer again to the z-test with n = 50 and data (**i**). Determine pSIG.

ANS. pSIG = 2 P(Z > 5.03) ~2 (0.0000003) using the closest table entry z = 5.00.

l-2. Determine a z-based 95% CI for mu(d) **from (i)**

ANS. dBAR +/- 1.96 s_d / root(50) = 75.2 +/- 1.96 4.5 / root(50) = {73.9527, 76.4473}.

m. **Note. Should have said retaining alpha = 0.2 at mu = 72.** Refer again to the z-test for n = 50 and data (h). Determine a sample size n (which would most often be greater than your 50 but will sometimes not require additional samples) required to achieve a z-test retaining alpha = **0.2** at mu = 72 but also achieving beta = 0.2 at mu = 73.

ANS. The generic formula is n = (($|z_0|$ + $|z_1|$) s_0 / (mu_0 – mu_1) $)^2$ in which, for the two-sided z-test, $|z_0|$ = z(alpha / 2) = z(0.1) = 1.282 and $|z_1|$ = z(beta) = z(0.2) = 0.84 (enter 0.5 -0.2 = 0.3 to the body of the z-table and read off z). So n = ((1.282+ 0.84) 1.6/ (72 – 73) $)^2$ = 12. We already have n = 50 so our current test does not require increasing the sample. It already achieves the desired alpha and beta specifications.

n. Refer to (n). Sketch P(reject H0 | mu) for the test of (m). **NEW: the steepest rise is roughly where the curve is at height 0.5 so, from now on, try to achieve that with your sketch to make it a bit more realistic.**

ANS. Sketch the two-sided curve form with alpha = 0.2 and passing through (71, 0.8) and (73, 0.8). It will fall between the other two curves.

o. Suppose that your n from test (m) is n = 1024 (it is not) and that you do continue sampling, taking an additional 1024 - 50. Suppose also that the overall sample mean of the 1024 is xBAR = 70.8. Determine the test statistic for the z-test (m) and the action taken by the test. **I refer to difference scores as x here.**

ANS. Test statistic (xBAR – 72) / (S_0 / root(1024)) = (70.8 – 72) / (1.6 / root(1024)) = -24. Reject H0 since the ABSOLUTE VALUE of the test statistic exceed (by far) threshold 1.282. pSIG = 2 P(Z > 24) is extremely tiny so this is overwhelming evidence against H0 and suggests in the most striking terms that mu_d is negative!

p. **Note. Retaining alpha = 0.2, not 0.1.** Refer to the t-test with data (k). Determine a sample size n (which would most often be greater than your 4 but will sometimes not require additional samples) required to achieve a t-test retaining alpha = **0.2** at mu = 72 but also achieving beta = 0.05 at mu = 73.

ANS. The generic formula is n = (($|t_0|$ + $|t_1|$) s_0 / (mu_0 – mu_1) $)^2$ in which, for the two sided test, $|t_0|$ = t(alpha / 2) = t(0.2 / 2) = t(0.1) = 1.638 and $|z_1|$ = t(beta) = t(0.05) = 2.353, n = ((1.638+ 2.353) 0.544671 / (72 – 73) $)^2$ = 5. We already have n = 4 so we need only one additional sample to achieve these specifications.

q. Suppose that your n from test (p) is n = 1024 (it is not) and that you do continue sampling, taking an additional 1024 - 4. Suppose also that the overall sample mean of the 1024 is xBAR = 73.8. Determine the test statistic for the t-test (p) and the action taken by the test.

ANS. (xBAR – 72) / (s / root(4)) = (73.8 – 72) / (0.544671 / root(1024)) = 105.75, which exceeds the threshold 1.638, so we reject H0. This is overwhelming evidence against H0.

3. a. A test has power 0.9 at mu = 6 (in H1). If mu = 6 we should reject H0 (since 6 is said to be in H1) but the test may instead make the error **of** failing to reject H0 owing to the random data (luck of the draw). What is the chance that the test will err and fail to reject H0?

ANS. The chance is beta = 1 – power = 0.1.

b. A test has boundary point mu0 = 45.5 and alpha = 0.01. If the true population mean is in fact 45.5 what is the chance that the test will fail to reject H0?

ANS. It is alpha = 0.01.

c. We are disappointed to learn that our sample shows a relatively large sample standard deviation s = 2.6. We'd have been happy if it had been around 1.3 since the precision of xBAR would have been acceptable for our purposes. If we can afford a larger sample size, around how many times larger will it have to be to achieve that precision?

ANS. For twice the precision the sample would have to be four times as large.

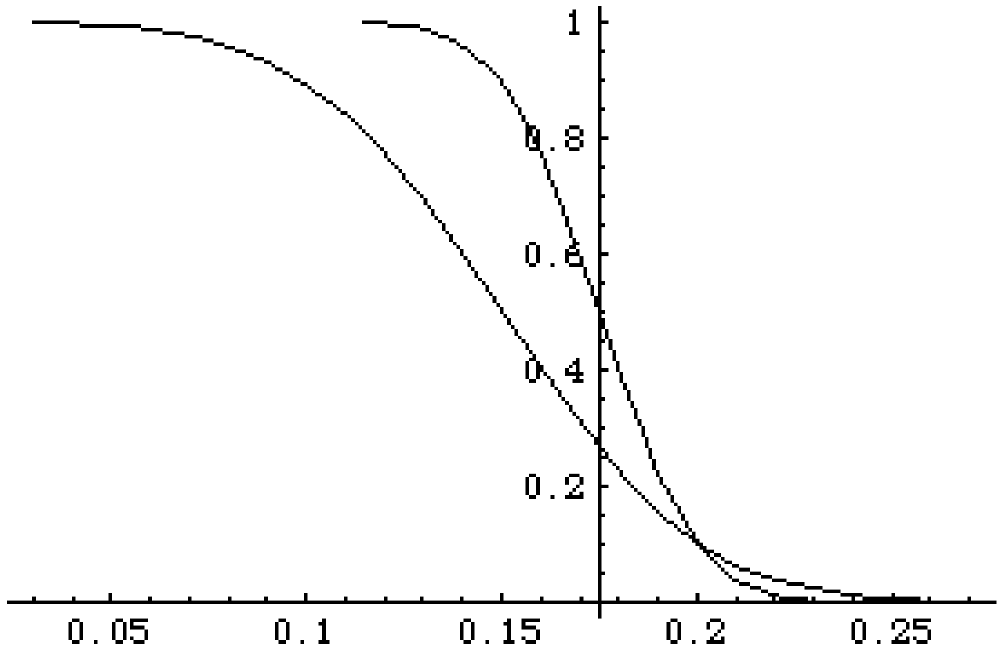
d. If we want greater precision but cannot afford more samples what remedies may be available?

ANS. Choose an estimator with smaller standard deviation. If applicable, a regression estimator, which reduces sd by the factor root(1 – rhoHAT^2) would act equivalently to a sample size increase by the factor 1 / (1 – rhoHAT^2). For example, if the sample correlation were rhoHAT = 0.8 then using the regression estimator in lieu of yBAR would effectively increase the sample size by the factor 1 / (1 – 0.64) = 2.8. So a regression estimator with n = 20 would in this case be equivalent to yBAR for n = 2.8 (20) = 56.

Other estimators can achieve similar improvement. We've not studied them, but estimators based upon stratification divide the sample into pieces that are random samples from different segments of the population. Stratified estimators reduce the variance of yBAR to the extent that the mean scores of the different population subgroups vary from one to the next. There are many other ways to reduce sampling variation.

4. **(NOTE change to n = 50, not 100).** Parts have a rejection rate p. Consider a z-test of the hypothesis H0: {p is not less than 0.2} versus the alternative H1: {p < 0.2}, with alpha = 0.1, based upon a with replacement sample of n = **50**. Rejecting H0 requires relatively strong evidence and would suggest that the rejection rate of parts is improving.

a. Make a sketch illustrating the general appearance of P(reject H0 | p) as a function of p in [0, 1]. Be sure to get the general shape right and alpha right (identify alpha appropriately). **NEW: the steepest rise is roughly where the curve is at height 0.5 so, from now on, try to achieve that with your sketch to make it a bit more realistic.**



ANS. I set this up to reflect a normal distribution for pHAT and a population sigma = root(0.2 0.8), but the curves are not truly accurate. Answer (a) is the less steep curve of the two above, achieving its steepest rise at p = 0.15. Alpha = 0.1 is seen by the heights of these two test curves at p0 = 0.2. The ideal plot, obtainable only by census, is identically zero on H0: {p greater or equal 0.2} and identically one on H1: {p < 0.2}. You will notice that these curves, drawn to reflect one sample size being 50 and the other 200 (four times greater), clearly show steepness that is greatest where the respective curves attain height 0.5. These are seen to be p = 0.15 and p = 0.175 respectively. If you look at the steepness of the respective slopes at these points you will see that one is twice as steep as the other, which is what happens when one sample is four times the other and we use the same alpha for both tests.

b. Overlay on your sketch (a) the IDEAL plot of P(reject H0 | p), which cannot be achieved without a census of the population of parts.

ANS. The ideal plot is zero everywhere on H0 and one everywhere on H1. You should draw it by hand in the above picture and also indicate H0, p0, and H1 on the horizontal axis.

c. Overlay on your sketch (a) a sketch of P(reject H0 | p) for a test based upon a sample of n = 200 (four times 50). Would it rise twice as fast as the original sketch? **NEW: the steepest rise is roughly where the curve is at height 0.5 so, from now on, try to achieve that with your sketch to make it a bit more realistic.**

ANS. It is the steeper curve, for sample size 200, which attains height 0.5 at p = 0.175, nearer to p0 = 0.2 than is the steepest point 0.15 of the original curve for sample size 50.

d. From your sketch (a) determine the approximate power at p = 0.1.

ANS. This is not exact as we are only interpreting the sketch as drawn. The power is the height at p = 0.1, which happens to be power ~ 0.9 in the sketch.

e. From your sketch (a) determine the approximate probability of an error of type II (i.e. beta) at p = 0.1.

ANS. Again, from the sketch, we have beta = 1 – power = 0.1 at p = 0.1. It is a coincidence.

f. From the z-table, determine the threshold against which the test statistic will be compared when deciding whether or not to reject H0.

ANS. Threshold is –z(alpha) = -z(0.1) = - 1.282. We reject H0 if the test statistic falls below –1.282.

g. Suppose a sample rate of part-rejections is pHAT = 0.25. What action will be taken by any (reasonable) test? Why? **Note, "rejections" refers to parts that are rejected. This use of "rejections" is not to be confused with rejecting H0.**

ANS. Since a rate 25% (of parts in the sample being rejected) is actually p = 0.25, which is in H0, there is no way a reasonable test would reject H0. The test usually rejects H0 only when the evidence is strong against H0. Often, rejecting H0 when it is true is regarded as a most serious error (error of the first kind, type one error).

h. Suppose a sample of 50 parts has 6 that must be rejected (this use of the word rejected has nothing to do with the test itself). So pHAT = 6/50. What is the value of the test statistic and what action is taken by the z-test? Be sure to use p0 in the denominator of the test statistic!

ANS. Test statistic is (pHAT – p0) / (root(p0 q0) / root(n))
= (0.12 – 0.2) / (root(0.2 0.8) / root(50)) = -1.4142
is less than the threshold –1.282, so we reject H0.

Were we to calculate pSIG is would be
P(Z < -1.41) ~ 0.07.
Since pSIG = 0.07 is less than alpha = 0.1 we reject H0 on that criterion also.

i. We desire a test with beta = 0.01 at p = 0.1 while retaining alpha = 0.1 at p = 0.2. Sketch the curve P(reject H0 | p) for such a test. **NEW: the steepest rise is roughly where the curve is at height 0.5 so, from now on, try to achieve that with your sketch to make it a bit more realistic.**

ANS. Sketch a curve like (a) with the same alpha = 0.1 at p = 0.2 but passing through (0.1, 0.01). Try to have steepest rise where the curve attains height 0.5.

j. **Note: I have re-phrased the question to reflect initial sample size 50 and to clarify that the 360 rejects are out of 2000. If you looked at this before it was proofed, the solution has been corrected.** Determine a sample size n adequate for the z-test (i). Suppose the required n is 2000 (it is not) and we sample the additional 2000 – **50** parts finding that among **all 2000** are 360 that must be "rejected." Determine the test statistic, the action taken by the z-test, and pSIG.

ANS. The basic formula is n = (($|z_0|$ root(p0 q0) + $|z_1|$ root(p1 q1)) / (p0 – p1))^2
= ((1.282 root(0.2 0.8) + 2.326 root(0.1 0.9)) / (0.2 – 0.1))^2 = 147.

But had we obtained 2000 instead of 147 we'd have continued to 2000 by selecting an additional 1950 samples. If we found 360 "rejects" among these 2000 we'd have an overall pHAT = 360 / 2000 = 0.18 and for our test statistic
(pHAT – 0.2) / (root(0.2 0.8) / root(2000)) = -2.24.
Since this is less than threshold –1.282 we would reject H0.

5. A population of workers is scored x1 = production using new equipment and x2 = production using old equipment. A sample of 50 workers is scored on the new equipment and an independent larger sample of 100 workers is scored on the old equipment, thought to yield more variable production scores. We find
x1BAR = 47, s1 = 33, n1 = 50 x2BAR = 43, s2 = 41, n2 = 100.

a. Determine a 95% CI for the difference mu1 – mu2.

ANS. (x1BAR – x2BAR) +/- 1.96 root($s_1^2 / n_1 + s_2^2 / n_2$)
= (47 – 43) +/- 1.96 root(33^2 / 50 + 41^2 / 100) = {-8.17569, 16.1757}.

Note. The above 95% CI overlaps {0}. So the usual alpha = 0.05 two-sided test of H0: {mu1 = mu2} would fail to reject H0. The usual two-sided test is equivalent to rejecting H0 only if the 95% CI fails to cover H0.

b. Determine the threshold for a z-test of H0: {mu1 = mu2} versus H1: {not equal} at level alpha = 0.05.

ANS. For the two-sided test the threshold is z(alpha / 2) = z(0.025) = 1.96.

c. From the data, determine the value of the test statistic and the decision reached by the test (b). If you choose not to pool declare that choice.

ANS. We have not studied the pooled test of mu1 = mu2 and will not pool. FYI the pooled approach assumes that the standard deviations of the two populations are the same (at least under the null hypotheses of no difference in the means).

The test statistic we will use does not assume $\text{sigma}_1 = \text{sigma}_2$. It is

(x1BAR – x2BAR) / root($s_1^2 / n_1 + s_2^2 / n_2$)

= (47 – 43) / root(33^2 / 50 + 41^2 / 100) = 0.64.

Since the ABSOLUTE VALUE of the test statistic does not exceed 1.96 we fail to reject H0.

d. Determine the threshold for a z-test of H0: {mu1 - mu2 is 2 or more} versus H1: {mu1-mu2 < 2} at level alpha = 0.05. The new equipment was touted as improving production by an average of 2 or more. Since rejecting H0 requires fairly strong evidence, we could be justified in claiming the equipment did not live up to its billing if the test rejects H0.

ANS. For the one sided test the threshold is –z(alpha) = - z(0.05) = -1.645. We reject H0 if the test statistic falls below this threshold.

e. Determine the test statistic for the test (d) and the action taken by the test.

ANS. The test statistic is

(x1BAR – x2BAR – 2) / root($s_1^2 / n_1 + s_2^2 / n_2$)

= (47 – 43) / root(33^2 / 50 + 41^2 / 100] = 0.32.

Since this is positve it actually supports the null hypothesis so we fail to reject H0.

f. Determine pSIG for the z-test (d).

ANS. It will be greater than 0.5 because the data actually supports H0. P(Z < 0.32) = 0.5 + P(0 < Z < 0.32) = 0.5 + 0.1255.

We would usually not bother to calculate pSIG because the data actually appears to support H0.