A statistical hypothesis test acts on data to arrive at a decision to either reject or not reject a stated null hypothesis. One important focus is to design a test achieving a given alpha. Then there is the matter of drawing a picture to describe how the test is expected to perform. Finally, one can choose the sample size to achieve needed performance goals.

BE ALERT FOR CORRECTIONS TO THIS KEY

STT 315 Recitation Assignment Due 3-23-06

Read Chapter 7. The key points are these:

a. The null hypothesis H_0 and alternative hypothesis H_1 describe contrasting sets of parameters. Examples include

a1.	$H_0: \mu \leq 16$ oz.	$H_1: \mu > 16$ oz.	one-sided test
a2.	$H_0: \mu \geq 28$ mpg.	$H_1: \mu < 28$ mpg.	one-sided test
a3.	H_0: p ≤ 0.6	H_1: p > 0.6	one-sided test
a4.	H_0: p ≥ 0.3	H_1: p < 0.3	one-sided test
a5.	$H_0: \mu = 16$ oz.	$H_1: \mu \neq 16$ oz.	two-sided test
a6.	H_0: p $= 0.6$	H_1: p $\neq 0.6$	two-sided test

b. Test statistics are quantities used to measure evidence against the null hypothesis H_0. Let $P(T > t_\alpha) = \alpha$ and $P(Z > z_\alpha) = \alpha$.

b1. reject H_0 if (test statistic) (xBAR - 16) / (s / $\sqrt{n}$) $\geqslant z_\alpha$ or t_α

b2. reject H_0 if (xBAR - 28) / (s / $\sqrt{n}$) $\leq z_\alpha$ or t_α

b3. reject H_0 if (pHAT - 0.6) / ($\sqrt{0.6\ 0.4}$ / $\sqrt{n}$) $\geqslant z_\alpha$

b4. reject H_0 if (pHAT - 0.3) / ($\sqrt{0.3\ 0.7}$ / $\sqrt{n}$) $\leq z_\alpha$

b5. reject H_0 if | (xBAR - 16) / (s / $\sqrt{n}$) | $\geqslant z_{\alpha/2}$ or $t_{\alpha/2}$

b6. reject H_0 if | (pHAT - 0.6) / ($\sqrt{0.6\ 0.4}$ / $\sqrt{n}$) | $\geqslant z_{\alpha/2}$

Note: In b2 and b4 above, there should be a minus sign on each of zALPHA, and tALPHA.

c. The test descriptions of (b) above can be re-phrased in terms of significance value p_{sig} (the probability of more evidence against H_0 than has been observed in the data).

In terms of p_{sig} : reject H_0 if $p_{\text{sig}} \le \alpha$.

c1. p = P(T > test statistic) or p = P(Z > test statistic)
c2. p = P(T < test statistic) or p = P(Z < test statistic)
c3. p = P(Z > test statistic)
c4. p = P(Z < test statistic)
c5. p = P(|T| > |test statistic|) or P(|Z| > |test statistic|)
c6. p = P(|Z| > |test statistic|)

d. Every test must balance the errors of type I (rejecting null hypothesis when it is true) and type II (failing to reject null hypothesis when it is false). This balance is revealed by plotting

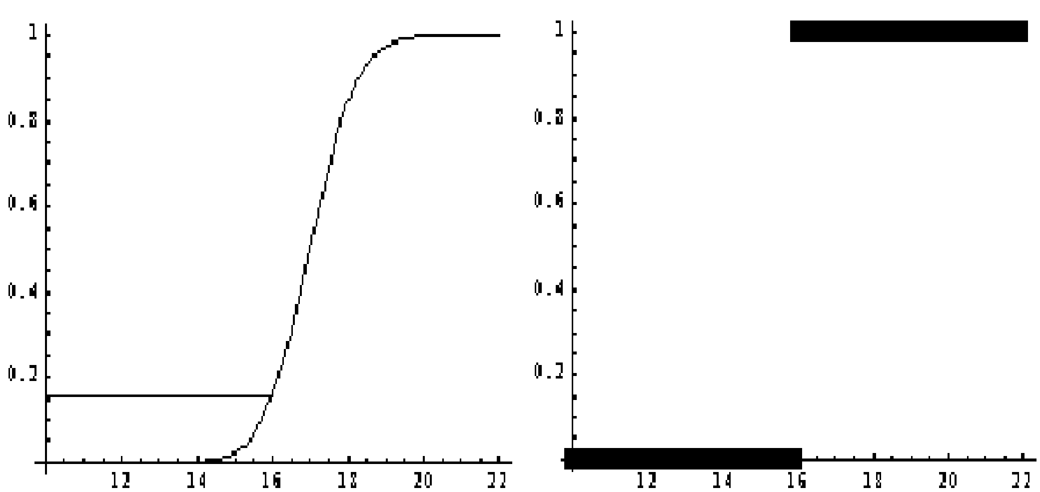
P(reject H_0 | μ) as a function of μ

or

P(reject H_0 | p) as a function of p

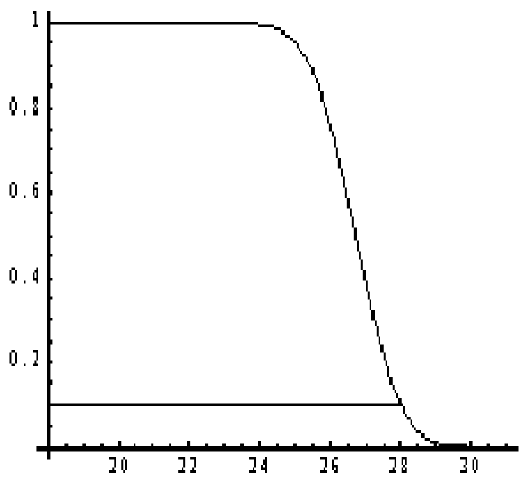
The following are typical shapes of these plots with α identified by a horizontal line above the boundary between the two hypotheses. Ideal plots are shown at right.

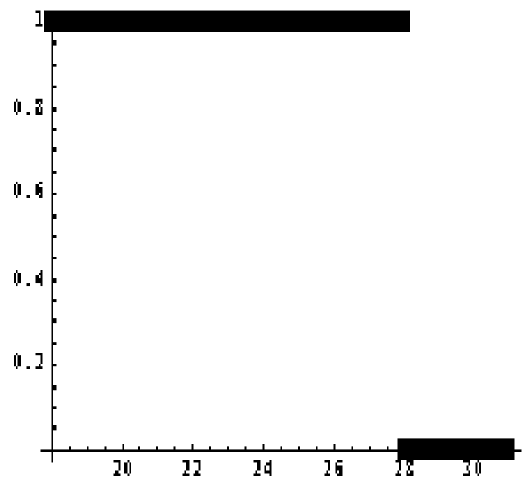
d1. Identify the null hypothesis, significance level alpha, power at mu = 18, ideal power at mu = 18 in the figures below. Label the axes mu and P(rej null | mu) respectively.



ANS. H0: mu less or equal 16, H1: mu > 16. Significance level alpha = 0.16. Power at mu = 18 is around 0.9 and beta = 0.1. Ideal power is 1 since 18 is not in H0.

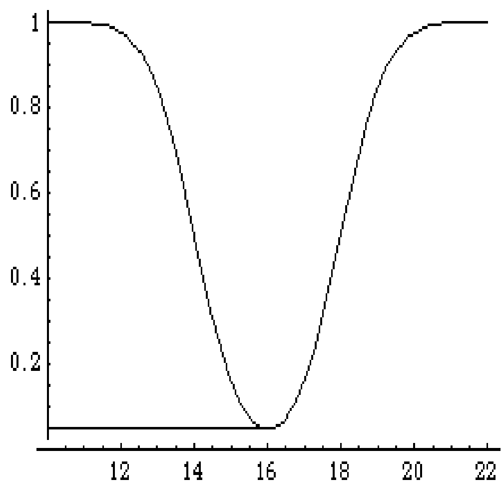
d2. Identify the null hypothesis, significance level alpha, power at mu = 26, ideal power at mu = 26 in the figures below. Label the axes mu and P(rej null | mu) respectively.

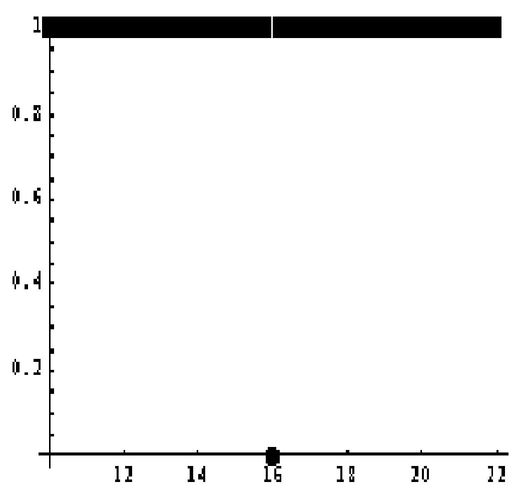




ANS. H0: mu greater or equal 28. H1: mu < 28. Alpha is around 0.1 while power at mu = 26 is around 0.76 with beta = 0.24. Ideal power is 1 since 26 is not in H0.

d5. Identify the null hypothesis, significance level alpha, power at mu = 13, power at mu = 18, ideal power at mu = 13, ideal power at mu =18 in the figures below. Label axes.





ANS. H0: mu = 16. H1: mu not 16. Alpha ~ 0.5, power at 13 ~ 0.85, beta ~ 0.15. Ideal power is 1 since 13 is not in H0.

Recitation assignment due 3-23-06.

1. For the null hypothesis that mu is greater or equal IQ = 100 versus alternative mu < 100 and significance level alpha = 0.02, freehand sketch the **general** shape of the curve P(rej null | mu) versus mu in the range 70 to 130. Identify alpha clearly and label axes properly. Impose on your sketch the ideal curve P(rej null | mu) over the same range of mu. Also impose upon your sketch another P(rej null | mu) curve for a better test, based upon **more** data, that shares the same alpha with your first one. Indicate which one of your curves is for more data.

ANS. Take d2 as your model.

2. Do (1) but for null hypothesis mu = 100 versus alternative mu not equal to 100.

ANS. Take d5 as your model.

3. Refer to (1). A random with-replacement sample of 400 persons from the population has sample mean IQ = 112.6. Without any calculation, what action is taken by the test? Why?

ANS. Since 112.6 belongs to H0 there is no possibility of rejecting H0.

4. Refer to (1). A random with-replacement sample of 400 persons from the population has sample mean IQ = 98.2 with sample standard deviation s = 14.6.

a. Give the math form of the test statistic and state under which circumstances the z-test will reject the null hypothesis. You will need to use the z-table to determine the appropriate z value and you will have to reason as to whether should be positive or negative.

ANS. (xBAR-mu0)/(s/root(n)) = (98.2-100)/(14.6/20) = -2.46575 reject H0 if this is less than -zALPHA = -z(.02) = -2.05. So we reject H0.

b. Calculate the numerical value of the test statistic. Compare it with the threshold of part (a) and state the action taken (reject null or fail to reject null).

ANS. See (4a).

c. Determine the **statistical significance** for this data. It is the probability that xBAR would have been more **unfavorable** to the null that it is for this data (calculated when mu is 100, the boundary between null and alternative). So if your test statistic is –1.87 (it is not) you would report statistical significance equal to the area under the z-curve to the left of –1.87 (more unfavorable than you observed).

ANS. pSIG = area left of –2.47 = 0.0068. So the same test (4ab) rejects H0 is pSIG < alpha. Since 0.0068 < 0.05 we (as above) reject H0.

d. From a lab, go to file bootstrap5.nb on the website you will find there a new function zTail which can calculate tail areas under the z-curve, e.g. zTail[1.96] will give 0.025. Use it to verify your answer (c).

ANS. zTail[2.47]~0.0068.

e. Use your statistical significance from (c) to conduct the test. That is, reject the null if the significance falls below alpha = 0.02. State the action taken. This should agree with (b), which conducted the test in another, but equivalent, way.

ANS. See (4c).

f. Suppose instead that the sample was only n = 50 (not n = 400) and the sample mean was xBAR = 98.2 with sample sd equal to s = 14.6. IF THE POPULATION IS KNOWN TO BE IN CONTROL (which is typical for IQ scores) we are entitled to conduct a t-test. Do so, stating the DF and consulting the t-table to determine a nearest entry for the rejection threshold for the test statistic. From a lab, go to the file bootstrap5.nb which has a function tTail[t, DF] for which the command tTail[1.87, 49] would give the statistical significance if the test statistic were –1.87 (it is not).

ANS. Test statistic = (98.2 – 100)/(14.6/20) = -2.46575. From the computer, pSIG = tTail[2.46575, 49] = 0.00861068. Reject H0 if pSIG < alpha = 0.05. Therefore reject H0.

5. Let p denote the fraction of voters favoring the Republican candidate. For the null hypothesis that Republicans have at least 50% of the vote versus alternative p < 0.5, and significance level alpha = 0.04, freehand sketch the **general** shape of the curve P(rej null | p) versus p in the range 0 to 1 for a possible test of this null hypothesis. Identify alpha clearly and label axes properly. Impose on your sketch the ideal curve P(rej null | p) over the same range of p. Also impose upon your sketch another P(rej null | p) curve for a better test, based upon **more** data, that shares the same alpha with your first one. Indicate which one of your curves is for more data.

ANS. Take d2 as your model. The better curve is lower on H0 and higher on H1 while passing through the point (0.5, 0.04).

6. Let p denote the fraction of customers buying the red label product. Historically this has been at p = 0.38. For the null hypothesis p = 0.38, versus the alternative p not equal to 0.38, and significance level alpha = 0.05, freehand sketch the **general** shape of the curve P(rej null | p) versus p in the range 0 to 1. Identify alpha clearly and label axes properly. Impose on your sketch the ideal curve P(rej null | p) over the same range of p. Also impose upon your sketch another P(rej null | p) curve for a better test, based upon **more** data, that shares the same alpha with your first one. Indicate which one of your curves is for more data.

ANS. Take d5 as your model.

7. Refer to (6). Suppose a random with-replacement sample of n = 100 customers yields the sample fraction pHAT = 0.47 who favor the red label product.

a. Calculate the test statistic for the z-test of the hypothesis p = 0.38. **Note: when calculating the test statistic it is common practice to use root(.38 .62)/root(100), not the value root(.47 .53)/root(100) as would used for the CI. This is because doing so leads to a test having slightly more desirable performance against alternatives close to p = 0.38. Ordinarily, there will be little difference between doing it one way or the other since for fractions f between 0 and 1 the function root(f (1-f)) varies slowly as f is changed about, at least for values of f not too close to 0 or 1.**

ANS. (pHAT – p0)/(root(p0 q0)/root(n)) = (0.47 – 0.38)/(root(0.38 0.62)/root(100)) = +1.85419.

b. Use the z-table to determine a threshold for rejection of the null. Make sure to take into account that this is a two-sided test that will reject the null if the test statistic is either too large or too small.

ANS. z(ALPHA/2) = z(0.025) = +1.96. The test rejects H0 if the test statistic 1.85.. > 1.96. So we fail to reject H0.

c. Conduct the test stating the action taken (reject null or fail to reject null).

ANS. See (b) above.

d. For the given data, calculate the statistical significance. **Note: this is a two-sided test so, for a test statistic value of +1.87 (it is not) we would report significance level equal to the combined areas left of –1.87 and right of 1.87, or (simply put) twice the area right of 1.87 as "more extreme than observed."**

ANS. pSIG = 2 P(Z > 1.85) = 0.0643135.

e. Using your statistical significance (d) conduct the test by comparing it with alpha = 0.05. State the result of this comparison and the action taken (reject null or fail to reject null).

ANS. Reject H0 if pSIG < alpha. Since 0.0643135 < 0.05 we fail to reject H0.

8. Refer to problem (7). Suppose we desire a test having alpha = 0.05 and also power P(rej null | p = 0.42) = 0.96. If it can be achieved we then have a test whose chance of falsely rejecting p = 0.38 (if it is true) is only 0.05 but also one whose chance of rejecting p = 0.38 is 0.96 if truly p = 0.42. Free hand sketch P(rej null | p) as p varies between 0 and 1 for such a test.

ANS. Use p5 as your model.

9. The formula at the top of page 319 may be used to determine a sample size that will support the test described in (8). The formula tells us the required sample size n for these specifications. If we agree to use this recommended n then the test for alpha = 0.05 will automatically have the desired power also. Evaluate the formula using

$z_0 = z_{alpha/2} = z_{0.025} = 1.96$ (since the test is two-sided)

$z_1 = z_{beta} = z_{0.04} = 1.75$

$\mathbf{p_0 = 0.38 \quad p_1 = 0.47}$

sd known, or from a preliminary sample

The third to last line is not a typo since the chance of false rejection of null, for $p_1 = 0.47$, can safely ignore the small contribution of doing so by accidentally producing a pHAT far below 0.38.

ANS. (1.96 root(0.38 0.62) + 1.75 root(0.47 0.53))^2 / (0.38 – 0.47)^2 ~ 412.