Statistical tests make a decision: either <u>reject</u> or <u>fail to reject</u> H0.

Important skills:
- **Design a one or two-sided test having a given alpha.**
- **Plot and interpret P(reject H0 | mu (or p)) vs mu (or p).**
- **Design one or two-sided tests having given alpha and beta.**
- **Execute tests based on sample data.**

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.	$H_0: \mu \le 16$ oz.	$H_1: \mu > 16$ oz.	one-sided test
a2.	$H_0: \mu \ge 28$ mpg.	$H_1: \mu < 28$ mpg.	one-sided test
a3.	$H_0: p \le 0.6$	$H_1: p > 0.6$	one-sided test
a4.	$H_0: p \ge 0.3$	$H_1: p < 0.3$	one-sided test
a5.	$H_0: \mu = 16$ oz.	$H_1: \mu \ne 16$ oz.	two-sided test
a6.	$H_0: p = 0.6$	$H_1: p \ne 0.6$	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 $P(Z > z_\alpha) = \alpha$. Be reminded that z_α is a special case of t_α for df $= \infty$. Except for very large df, we require a normal population if Student's t is used.

b1. reject H_0 if (test statistic) $(xBAR - 16)/(s/\sqrt{n}) \ge t_\alpha$

b2. reject H_0 if $(xBAR - 28)/(s/\sqrt{n}) \le -t_\alpha$

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

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

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

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

NOTE CHANGE: We will ONLY use root(p0 q0) NOT root(pHAT qHAT) in the denominator of tests about p since this must be done anyway when designing tests with given beta.

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} always|: reject H_0 if $p_{sig} \leq \alpha$.

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

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

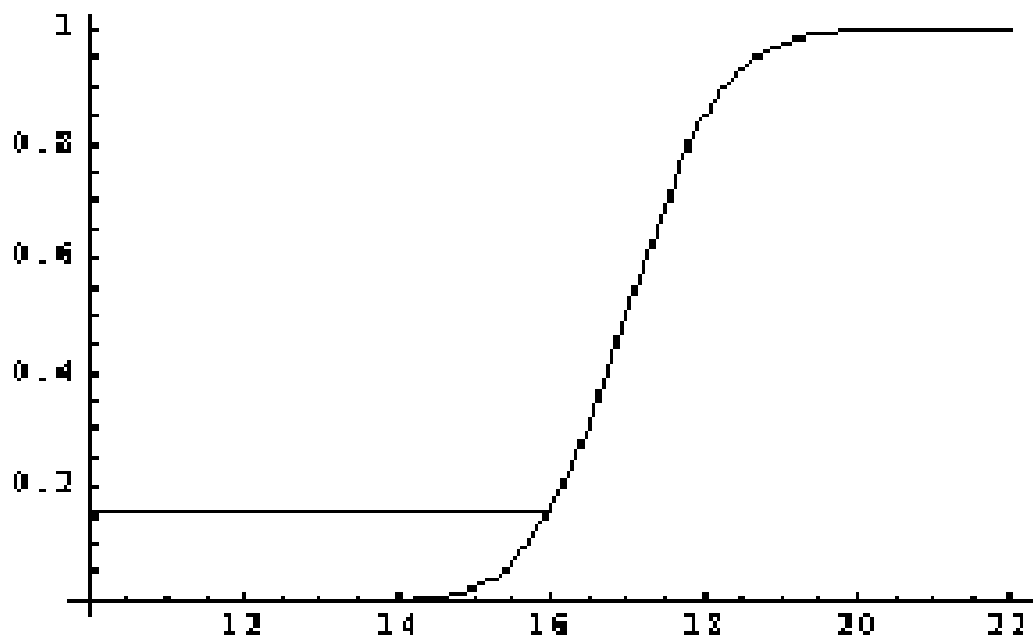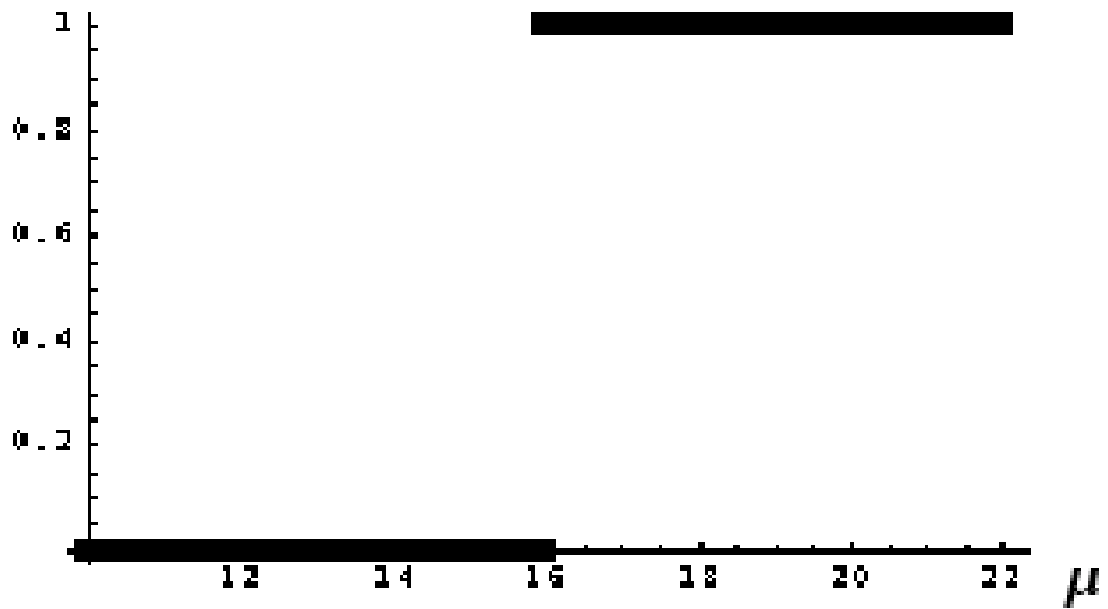$P(\text{reject } H_0 \mid \mu)$ as a function of μ

or

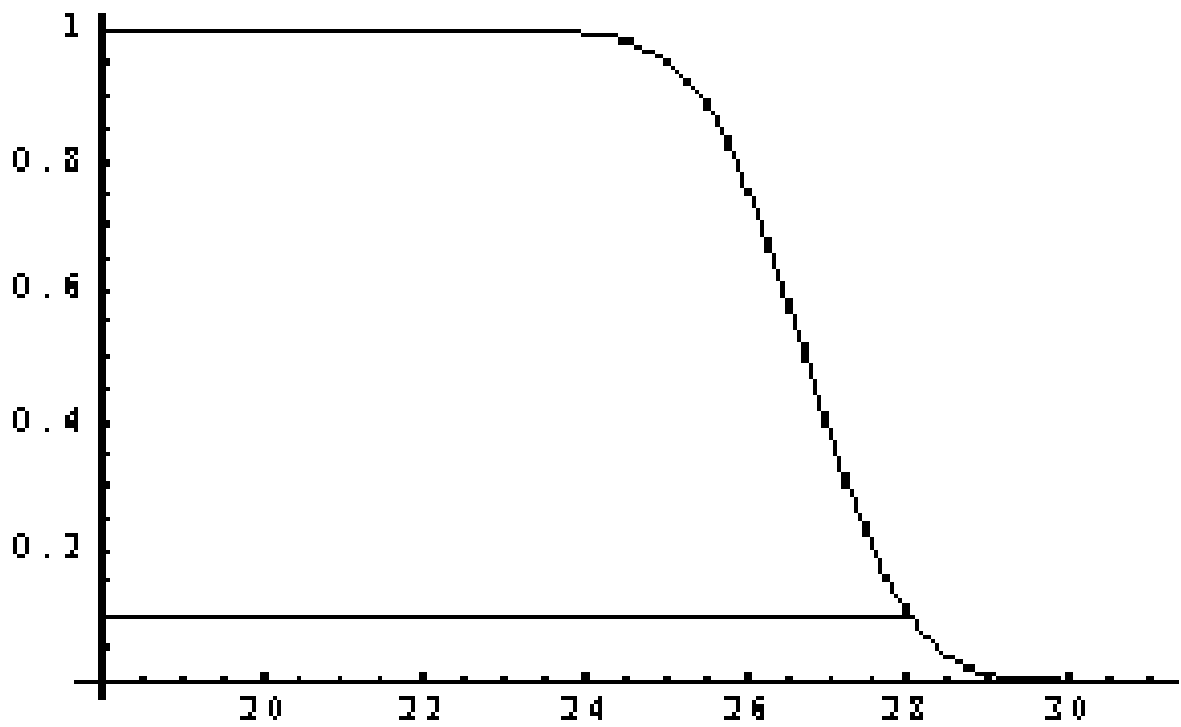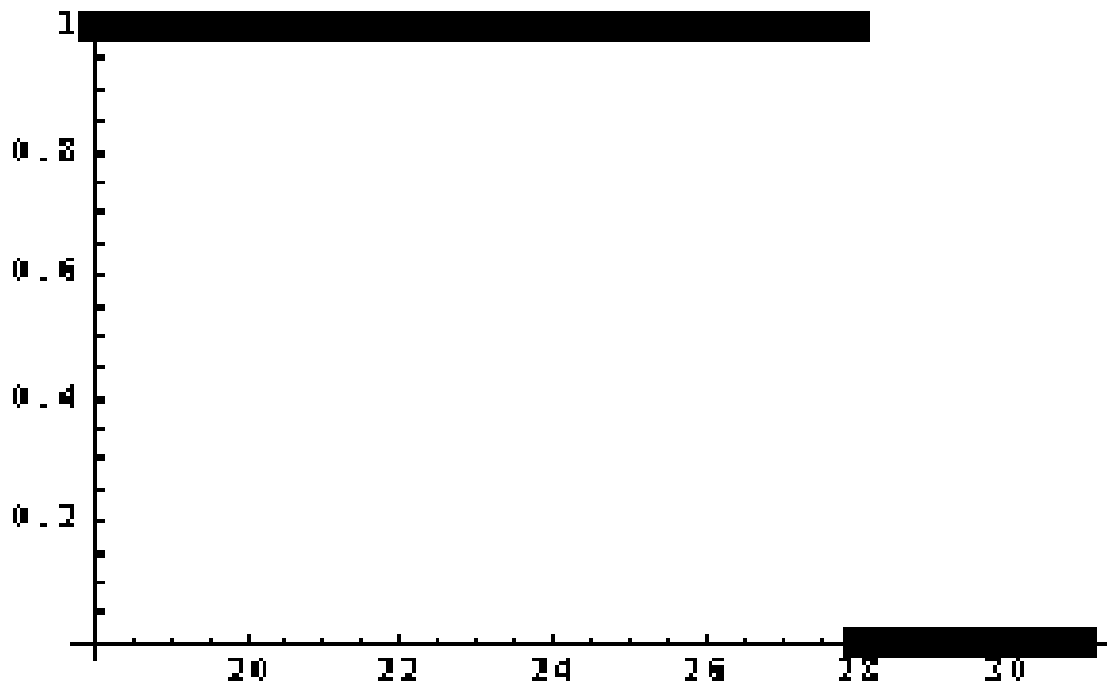$P(\text{reject } H_0 \mid p)$ as a function of p

The following are typical shapes of these plots with α identified by a horizontal line above the boundary value ($\mu 0$ or $p0$) where H0 touches H1. Ideal plots are also shown.
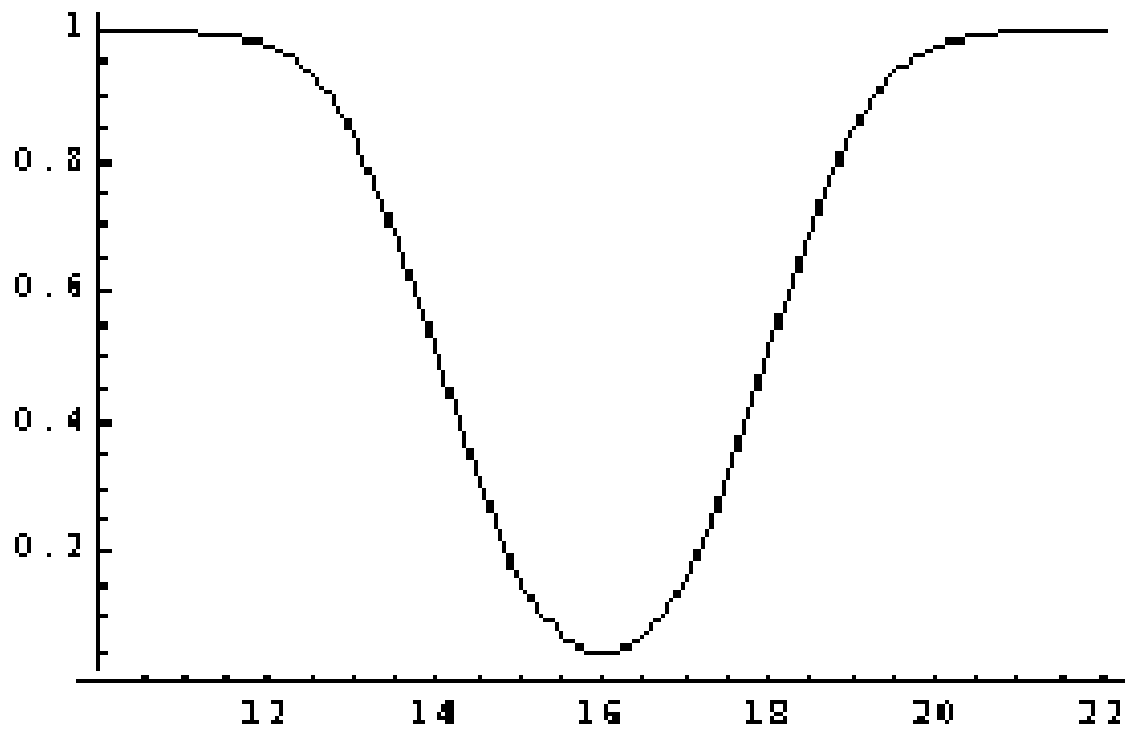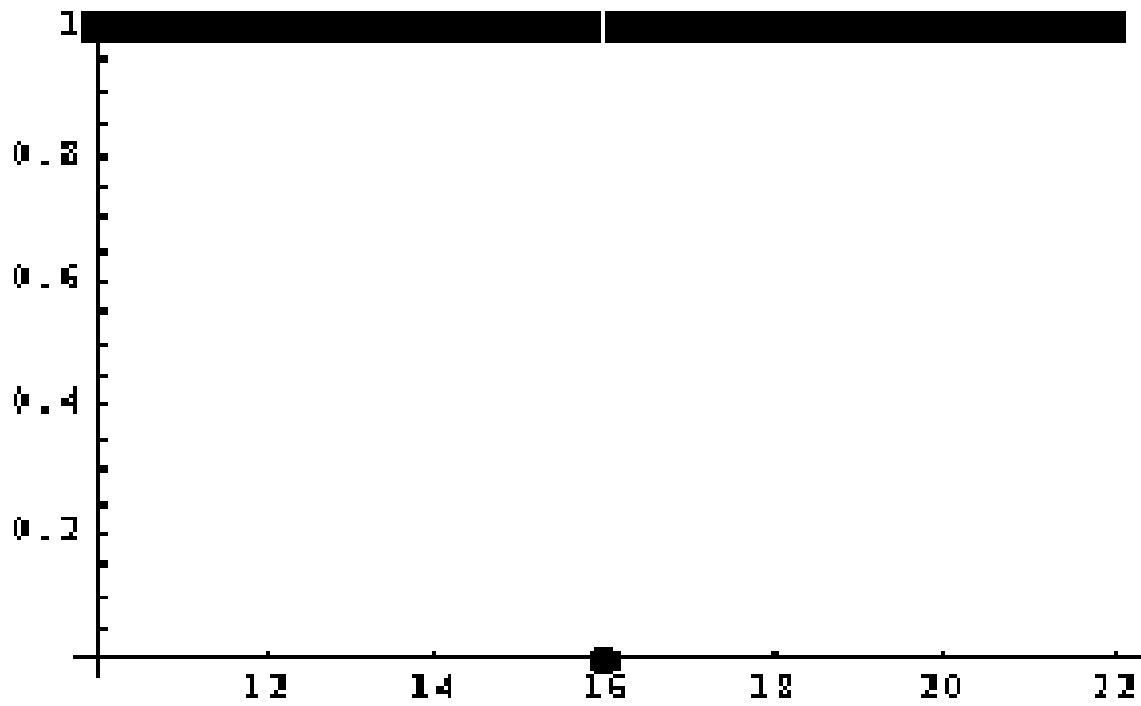
da1. (Ideal first).

P(reject H0 | μ)

da2.

da5.

The following examples will be covered in the lectures for the week of 10-30-06. A key will be posted Tuesday.

Exam 3 is Thursday, 11-2-06.

1. Identify the null hypothesis, significance level alpha, power and beta at mu = 18, ideal power at mu = 18 in the figures below. Label axes mu and P(rej H0 | mu) respectively.
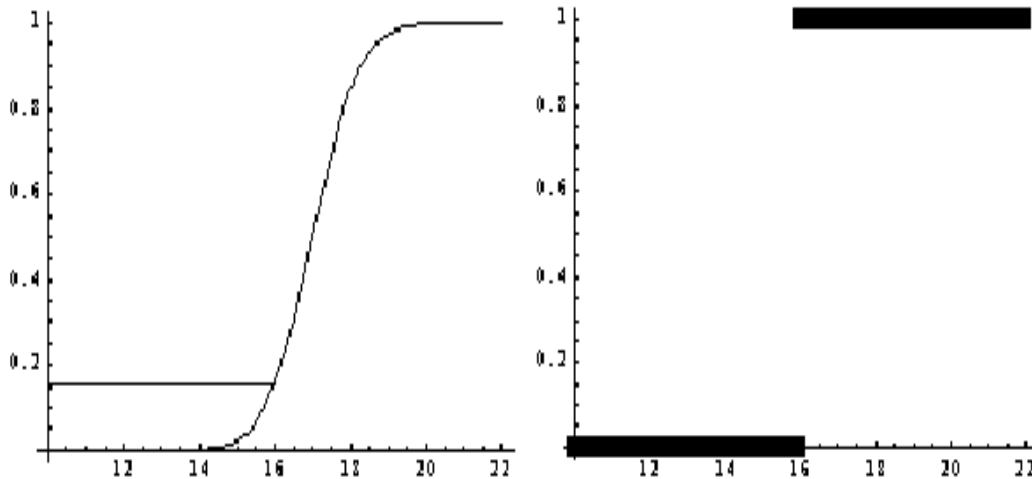
2. Identify the null hypothesis, significance level alpha, power and beta at mu = 26, ideal power at mu = 26 in the figures below. Label axes mu and P(rej H0 | mu) respectively.
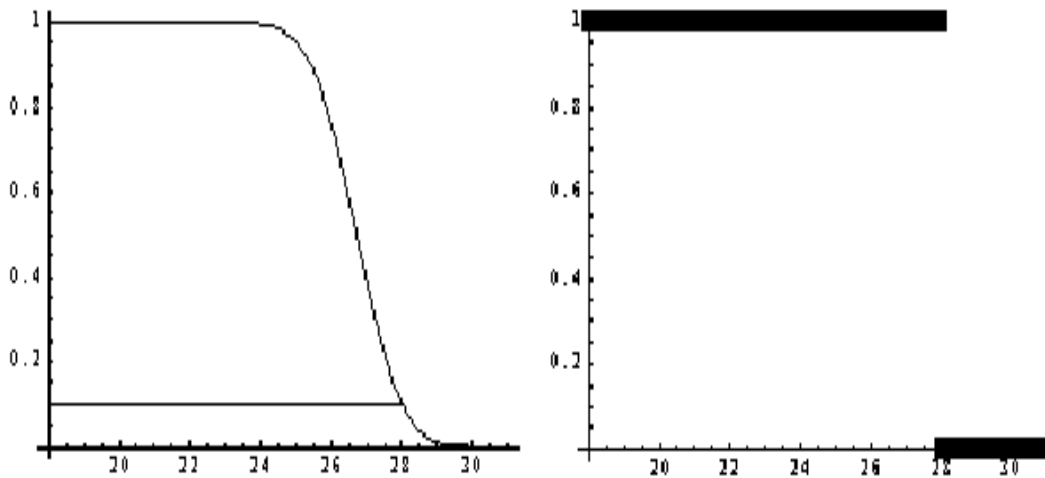
3. Identify the null hypothesis, significance level alpha, power at mu = 13, power at mu = 18, ideal power at mu = 13, ideal power at mu =18 in the figures below. Label axes.

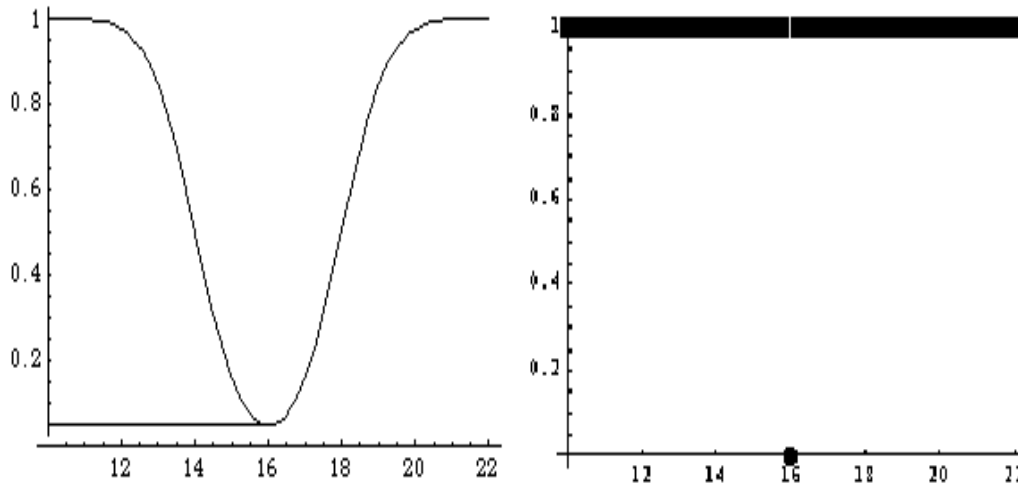4. For the null hypothesis H0 that mu is greater or equal IQ = 100 versus alternative mu < 100 and significance level alpha = 0.02, freehand sketch the **general** shape of the curve P(rej null | mu) versus mu, over the range 70 to 130. Identify alpha clearly and label axes properly. Impose upon your sketch the ideal curve P(rej H0 | 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 two curves (the ones achievable without a census) is for more data.

5. Do (4) but for null hypothesis mu = 100 versus alternative mu not equal to 100.

6. Refer to (4). 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? Hint: The score112.6 belongs to H0.

7. Refer to (4). 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 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 z0 value and you will have to reason as to whether should be positive or negative.

b. Calculate the numerical value of the test statistic. Compare it with the threshold z0 of part (a) and state the action taken (reject null or fail to reject null).

c. Determine the **statistical significance** pSIG 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 equal to –1.87 (it is not) you would report statistical significance equal to the area under the z-curve to the left of –1.87 (since more negative xBAR = more negative TS = more unfavorable to this H0).

d. Use your statistical significance pSIG to conduct the test. That is, reject H0 if the significance pSIG < alpha = 0.02. State the action taken. This should agree with (b), which conducted the test in the other, but equivalent, way.

e. Suppose, instead, that the sample was only n = 61 (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. Perform the t-test, stating the DF and consulting the t-table to determine the appropriate rejection threshold t0 for this test.

8. Let p denote the fraction of voters favoring a 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.

9. Let p denote the fraction of customers buying a "red label" product. Suppose that 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 H0 | 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 in your sketch. 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.

10. Refers to (9). Suppose a random with-replacement sample of $n = 100$ customers yields the sample fraction pHAT $= 0.47$ who favor the red label product.

a. <u>Calculate the test statistic</u> for the z-test of H0: $p = 0.38$ vs H1: p is not 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 be used for the CI. This is because doing so leads to a test having slightly more desirable performance against alternatives close to $p = 0.38$.**

b. <u>Use the z-table to determine a threshold z0 for rejection of the null.</u> 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.

c. <u>Conduct the test</u> stating the action taken (reject H0 or fail to reject H0).

d. For the given data, <u>calculate the statistical significance pSIG.</u> **Note: this is a two-sided test so, for a z-test statistic value of +1.87 (it is not) we would report significance level pSIG equal to the combined z-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."**

e. Use your statistical significance pSIG from (d) to (again) <u>conduct the test</u>, this time by comparing pSIG with alpha $= 0.05$. State the result of this comparison and the action taken (reject H0 or fail to reject H0). It will be the same action taken in (c) but the test is being performed this other way.

11. Refers to problem (10). Suppose we desire a test of H0: $p = 0.38$ vs H1: p is not 0.38, having alpha $= 0.05$ but also having power P(rej null | $p = 0.42$) $= 0.96$. That is, the probability of type 2 error is equal to beta $= 0.04$ (at $p = 0.42$). If this is 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.

a. Free hand sketch P(rej null | p) as p varies between 0 and 1 for such a test.

b. The formula at the top of page 319 may be used to determine a sample size that will support this desired test. 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 0.96 at p = 0.42 also. Evaluate the formula to determine this n.

c. For the "n" of (b) suppose we find that pHAT = (# buying red label)/n = 0.44. Calculate the test statistic for the alpha = 0.05 test of H0: p = 0.38 vs H1: "p is not 0.38" for this n.

d. Using the answer (c) conduct the test. Be sure to show the test statistic, its comparison with z0, and the decision taken by the test (either to reject H0 or not).

12. A company ships packages all over the world and its shipping costs can take unexpected swings due to economic changes but more particularly changes in its customer base, particularly international orders. Rather than attempt to model or monitor all of these factors it is decided to monitor shipping costs of individual packages directly on a continuing basis. Let's say that, as of today, we think we average 4.77 shipping cost per package. The following day a sample of packages will be selected and a z-test of H0: "mu = is less or equal to 4.77" vs H1: mu > 4.77 will be conducted at level alpha = 0.025.

a. Suppose we decide to use nINIT = 50 packages (with replacement). <u>Evaluate the test statistic, z0, and determine the decision taken by the z-test</u> if this initial sample of 50 finds xBAR = 5.31 with sample sd sINIT = 2.44 (it seems shipping costs vary a lot by package).

b. Concerned that a test sample of 50 may not be large enough to allow reliable detection of a change to (say) mu = 5, which would seriously affect total cost for the many thousands of packages shipped daily, it is decided to use a possibly larger n to achieve power 0.9 at mu = 5. That is, 90% of the time we would expect to reject H0 if the mean mu has shifted to 5. Use the formula of pg. 314, with sigma replaced by our estimate sINIT of (a), to <u>determine nFINAL</u>.

c. Suppose we then continue our sample-of-the-day to nFINAL finding that xBARfinal = 5.22. Give the form of, and evaluate, the test statistic for the new test (the hybrid test statistic continues to use sINIT but uses root(nFINAL) and xBARfinal).

d. Comparing (c) with z0 from (a) <u>determine the action taken by the hybrid test</u> (c).

Exam 3 will have around 12 questions of which 3 will be dedicated to determining confidence intervals for mu or p, or difference of mu's or p's, or CI with given precision (requiring determination of needed sample size). These bonus questions will be very close in content to some questions from exam 2.

Formulas given on exam 2 will also appear on exam 3.
Formulas of pages 314, 319 will appear on exam 3.
Formulas for test statistics will NOT appear on exam 3.

Tables of t and z will appear on exam 3.

Some questions may have a few parts, particularly with questions about a plot of P(reject H0 | mu (or p)) vs mu (or p).