

STT 315

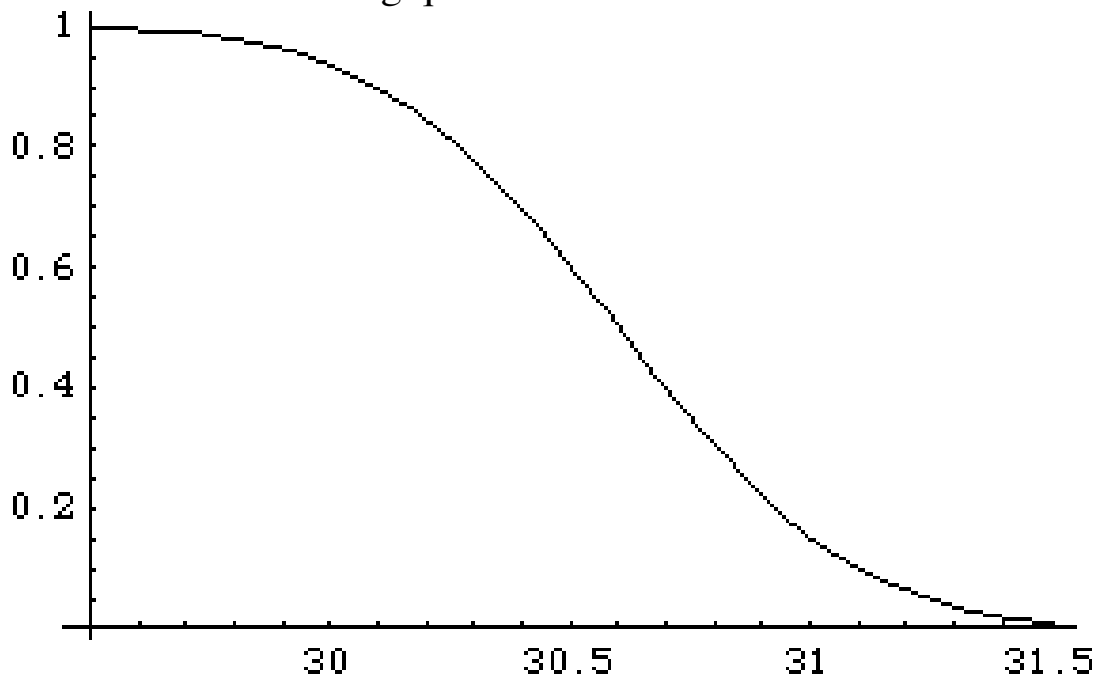
PrepKEY for Exam 3 (3-30-06) This will be covered MW in lec.
Keep track of your email for corrections to any errors present.

The exam will be run as usual. The exact number of questions is not known at this time. It will be announced by Wednesday lecture. The questions below do not constitute a representative exam but are certainly at the core of what we want you to be prepared to do. You can think of many variations on them.

1. H_0 : μ greater or equal to 31. H_1 : $\mu < 31$.

a. 1 or 2 sided? **ANS. ONE SIDED**

b. Sketch a curve $P(\text{reject } H_0 \mid \mu)$ representing the general shape seen for the z-test or, for normal population, the t-test. You will need to set up scales for the horizontal and vertical axes to deal with the following questions.



c. In your curve identify the approximate numerical value of the probability of type I error (i.e. error of first kind, i.e. α).
ANS. $\alpha \sim 0.15 = \text{height of this curve at } \mu = 31, \text{ the boundary between } H_0 \text{ and } H_1.$

d. In your curve identify the approximate numerical value of the probability of type II error (i.e. error of the second kind, i.e. beta) for $\mu = 30$.

ANS. ~ 0.06 , perhaps a little difficult to get accurately from the curve.

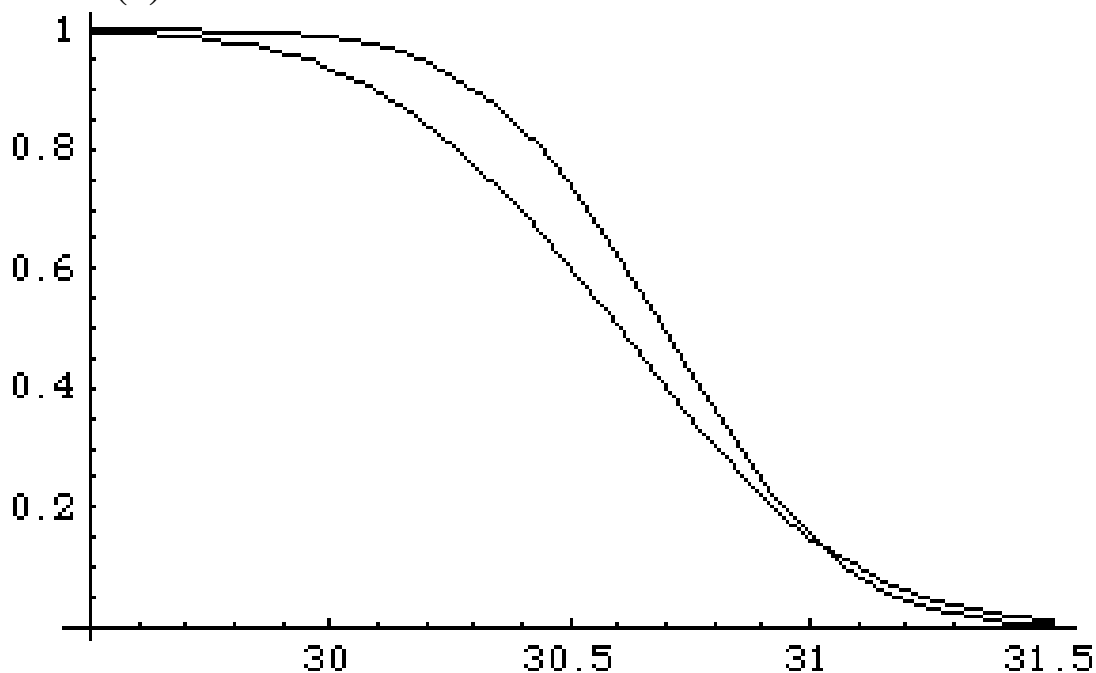
e. In (d) determine the approximate numerical value of the power of the test at $\mu = 30$.

ANS. height of curve at $\mu = 30 = 1 - \text{beta} \sim 0.94$.

f. Overlay on your sketch the ideal, but unattainable, plot of $P(\text{reject } H_0 \mid \mu)$.

ANS. = plot 1 over H_1 and 0 over H_0 .

g. Overlay on your sketch another plausible $P(\text{reject } H_0 \mid \mu)$ curve of the correct general shape, having the same alpha as your curve (b) but closer to the ideal for all other μ than 31.



h. There is no question (h).

i. For your curve (g) what is the power at $\mu = 30$? What is beta for $\mu = 30$?

ANS. These are difficult to obtain by eye. Power = 0.9895 (height at $\mu = 30$) and beta = $1 - \text{power} = 0.0105$.

j. Describe the two basic ways in which one may obtain a test with better performance as illustrated in (i)?

ANS. (Way 1) Increase the size of the sample. Each doubling of the sample steepens the curve's rise by a factor of two. (Way 2) Base the test on a better estimator of mu (i.e. one having a smaller sd). For example, a test based upon a regression estimator could help.

k. What is the test statistic based upon with-replacement samples $X_1 \dots X_n$?

ANS. $(\bar{x} - 31) / (s / \sqrt{n})$. Be SURE you are able to calculate the sample standard deviation s (see chapter 1).

l. For your curve (b), what is the criterion for rejecting H_0 ? Be explicit. Determine a threshold for rejection compatible with your alpha and the comparison with the test statistic that leads to rejection of H_0 .

ANS. The key point is that $\alpha \sim 0.15$ and the threshold is $-z_{\alpha} = -z_{0.15} \sim -1.03$. Reject H_0 if the test statistic is less than -1.03 . (For a one sided test with H_0 to the left of H_1 we would use $+1.03$. For a 2-sided test we would instead use $\alpha/2$.)

m. If the test statistic is evaluated for the data and is found to be 2.21 what action is taken by a z-test? Did you require any calculation to answer? Is \bar{x} greater than 31 or less than 31?

ANS. Since the test statistic is positive 2.21 it must be that the numerator $(\bar{x} - 31)$ is positive, so $\bar{x} > 31$. Our test is not going to reject H_0 if \bar{x} is actually in H_0 . After all, the test resists being moved off H_0 unless there is strong evidence to the contrary.

n. If the test statistic is evaluated for the data and is found to be -2.21 what action is taken by a z-test? Did you require any calculation to answer? Is \bar{x} greater than 31 or less than 31?

ANS. The test rejects H_0 if the test statistic is LESS than $-z_{0.15} \sim -1.03$. Since $-2.21 < -1.03$ we reject H_0 .

o. Repeat (m) for $n = 7$ and a t-test (assuming the population is normal or nearly so).

ANS. As in (m), if the test statistic is positive 2.21 that means the $\bar{x} > 31$, which means that \bar{s} is actually in H_0 . So the test (which resists being moved off H_0) will not reject H_0 .

p. Repeat (n) for $n = 7$ and a t-test (assuming the population is normal or nearly so).

ANS. Your t-table is not adequate to solve this problem. For $n = 7$ (DF 6) I've experimented with different values in tTail, with DF = 6, finding that $tTail[1.119, 6] \sim 0.15$. So the t-test is: reject H_0 if the test statistic is less than -1.119 . Since indeed the test statistic is $-2.21 < -1.119$ we reject H_0 .

Now, I will cast the very same test in p_{SIG} terms. Using our computer routines I find $tTail[2.21, 6] = 0.0313$ is the probability to the right of $t = 2.21$ for $n = 7$ (DF = 6). So $p_{SIG} = P(T < -2.21) = 0.0313$. Remember, p_{SIG} is defined to be the probability of EVEN MORE evidence against H_0 . We reject H_0 since $p_{SIG} = 0.0313 < 0.15 = \alpha$.

q. Once again refer to your test (b). Suppose that the test statistic is evaluated for a z-test and found to be -2.21 . What is the significance p_{SIG} ?

ANS. In this case $p_{SIG} = P(Z < -2.21)$ (does this make sense to you?) and from your z-table $P(Z < -2.21) = 0.0136$. Our computer routine gives $zTail[2.21] = 0.0135526$.

r. State the manner in which a test of any kind is always conducted when p_{SIG} and α are made available.

ANS. Regardless of the type of test, reject H_0 if $p_{SIG} < \alpha$. Remember, p_{SIG} is defined to be the probability of EVEN MORE evidence against H_0 than our data presented.

s. Referring to (q)(r), what action is taken by the z-test? Your answer must agree with (n).

ANS. Since $p_{SIG} = 0.0136 < \alpha = 0.15$, we reject H_0 . It is the same action taken in (n). This is just a different way of running the test. It has the advantage that the criterion for rejection "reject if $p_{SIG} < \alpha$ " is universal to all tests.

t. Suppose that a preliminary sample of $n_0 = 50$ is selected and the sample standard deviation of this sample of 50 scores is $s_0 =$

12.7. Determine a sample size n adequate to achieve both your α and your β from your curve (g). The required formula is in section 7-5 but you are substituting your preliminary estimate s_0 (obtained from a sample of $n_0 = 50$) for the true, unknown, population σ .

ANS.

$$n = ((|z_0| + |z_1|) s / (\mu_0 - \mu_1))^2$$

where $z_0 = -z_{\alpha} = -z_{0.15} \sim -1.03$ (had the test been two sided you'd have used $z_0 = z_{\alpha/2} = z_{0.075}$ and $z_1 = z_{\beta} = z_{0.06} \sim 1.56$ (even in the two sided case we do NOT use $\beta/2$). So

$$n = ((1.03 + 1.56) 12.7 / (31 - 30))^2 \sim 1082.$$

We'd need an additional $1082 - 50 = 1032$ samples.

u. Refer to (t). Suppose we continue sampling from $n_0 = 50$ up to the recommended sample size $n = 1082$. We are thus incorporating an additional 1032 samples as recommended by the formula. We then would need to evaluate the HYBRID test statistic

$$(\bar{x}_n - 31) / (s_0 / \sqrt{n}).$$

This test statistic is termed hybrid for the obvious reason that the sample standard deviation s_0 from the initial 50 is used but the sample mean \bar{x}_n of all $n = 1082$ is used, as is \sqrt{n} .

ANS. JUST UNDERSTAND THAT THE HYBRID TEST STATISTIC NEEDS TO BE USED IN THIS CASE, NOT THE TEST STATISTIC WITH THE s FROM ALL n . Also, it is the size of the INITIAL sample which determines the DF and whether the z-test is applicable.

v. The test achieving your α and β per (g) is then to reject H_0 if the HYBRID test statistic is less than the threshold you would have used for any z-test with your α of (g). Determine this $-z_{\alpha}$ and state the criterion for rejecting H_0 in terms of the value of the hybrid test statistic.

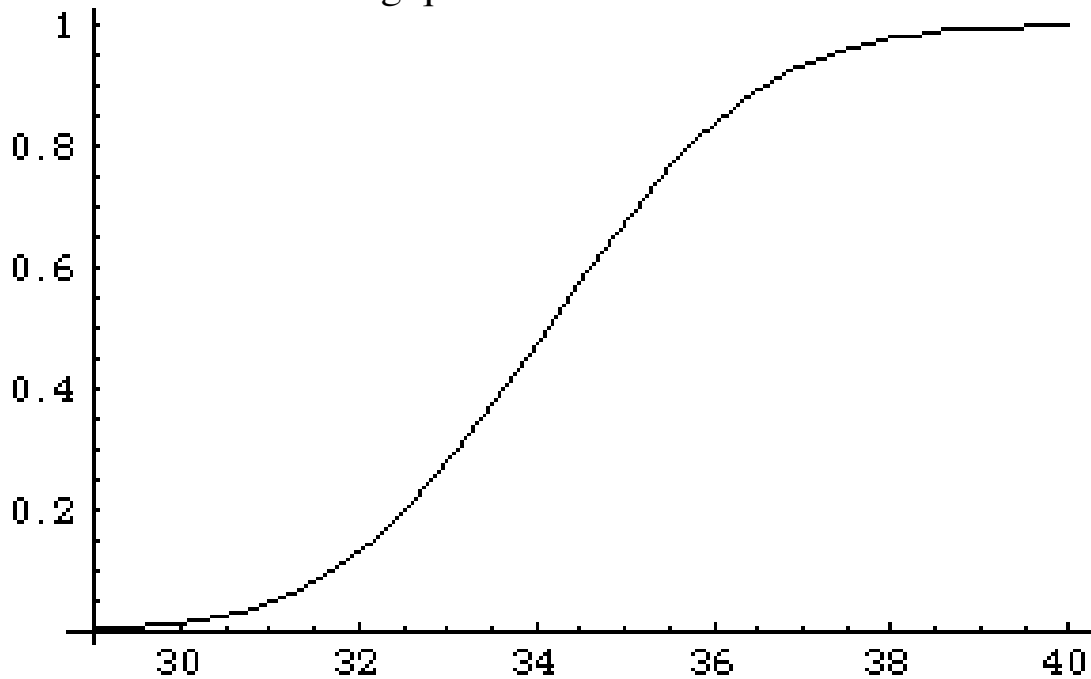
ANS. $-z_{\alpha} = -z_{0.15} = -1.03$. Reject H_0 if the hybrid test statistic is less than -1.03 .

2. Same as (1) but $H_0: \mu \leq 31$, $H_1: \mu > 31$, and we are interested in power at $\mu = 32$.

a. 1 or 2 sided?

ANS. ONE SIDED

b. Sketch a curve $P(\text{reject } H_0 \mid \mu)$ representing the general shape seen for the z-test or, for normal population, the t-test. You will need to set up scales for the horizontal and vertical axes to deal with the following questions.



c. In your curve identify the approximate numerical value of the probability of type I error (i.e. error of first kind, i.e. alpha).

ANS. $\alpha \sim 0.05 =$ height of this curve at $\mu = 31$, the boundary between H_0 and H_1 .

d. In your curve identify the approximate numerical value of the probability of type II error (i.e. error of the second kind, i.e. beta) for $\mu = 32$.

ANS. ~ 0.87 . This is way too large.

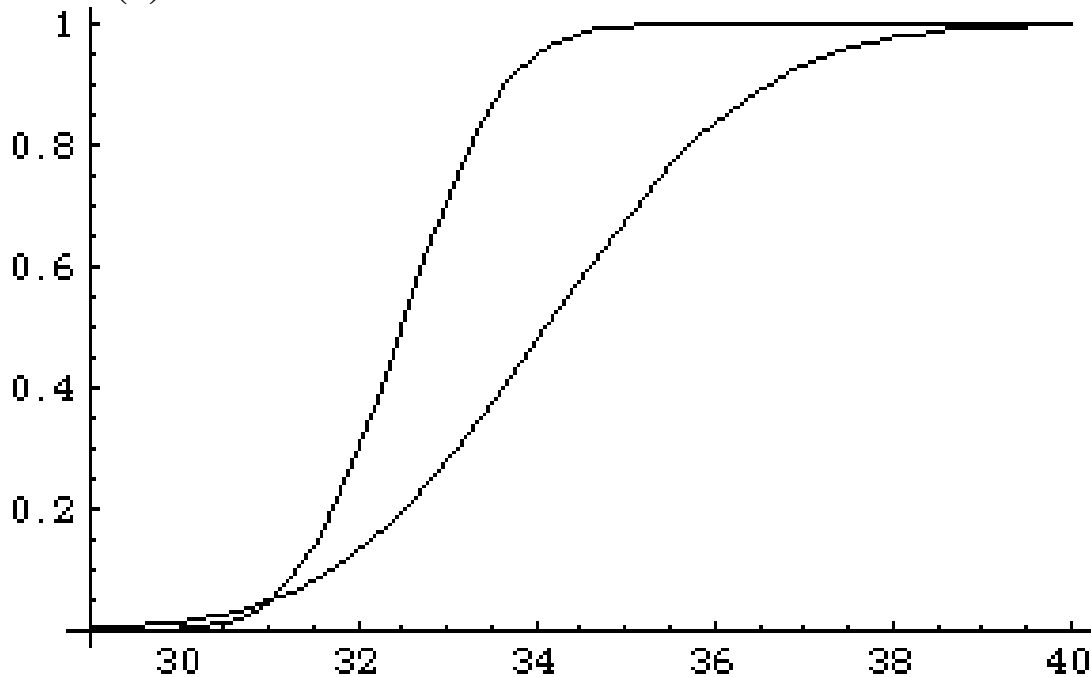
e. In (d) determine the approximate numerical value of the power of the test at $\mu = 32$.

ANS. height of curve at $\mu = 30 = 1 - \beta \sim 0.13$. We desire far larger power.

f. Overlay on your sketch the ideal, but unattainable, plot of $P(\text{reject } H_0 \mid \mu)$.

ANS. plot = 1 over H_1 and 0 over H_0 .

g. Overlay on your sketch another plausible $P(\text{reject } H_0 \mid \mu)$ curve of the correct general shape, having the same α as your curve (b) but closer to the ideal for all other μ than 31.



h. There is no question (h).

i. For your curve (g) what is the power at $\mu = 32$? What is beta for $\mu = 32$?

ANS. These are difficult to see by eye. Power = 0.29 (height at $\mu = 32$) and beta = $1 - \text{power} = 0.71$. Still not so good. We want far larger power and consequently far lower beta.

j. Describe the two basic ways in which one may obtain a test with better performance as illustrated in (i)?

ANS. (Way 1) Increase the size of the sample. Each doubling of the sample steepens the curve's rise by a factor of two. (Way 2) Base the test on a better estimator of μ (one having a smaller sd). For example, a test based upon a regression estimator.

k. What is the test statistic based upon with-replacement samples $X_1 \dots X_n$?

ANS. $(\bar{x} - 31) / (s / \sqrt{n})$.

l. For your curve (b), what is the criterion for rejecting H_0 ? Be explicit, determining the threshold for rejection compatible with your alpha and the comparison with the test statistic that leads to rejection of H_0 .

ANS. The key point is that $\alpha \sim 0.05$ and the threshold is $z_{\alpha} = z_{0.05} \sim 1.64$. Reject H_0 if the test statistic exceeds 1.64.

m. If the test statistic is evaluated for the data and is found to be 2.21 what action is taken by a z-test? Did you require any calculation to answer? Is \bar{x} greater than 31 or less than 31?

ANS. Since the test statistic is positive 2.21 it must be that the numerator $(\bar{x} - 31)$ is positive, so $\bar{x} > 31$. We reject H_0 because the test statistic exceeds threshold $z_{\alpha} = z_{0.05} \sim 1.64$.

n. If the test statistic is evaluated for the data and is found to be -2.21 what action is taken by a z-test? Did you require any calculation to answer? Is \bar{x} greater than 31 or less than 31?

ANS. Since the test statistic is negative we have $\bar{x} < 31$. As such the data leans towards H_0 not towards H_1 . So our test, which tends to sit on H_0 unless there is strong evidence to the contrary, will not reject H_0 .

o. Repeat (m) for $n = 7$ and a t-test (assuming the population is normal or nearly so).

ANS. Your t-table gives $t_{.05} = 1.943$. Likewise $t_{\text{Tail}}[1.943, 6] \sim 1.943$. The test rejects H_0 if the test statistic exceeds 1.943. Since $2.21 > 1.943$ we reject H_0 .

p. Repeat (n) for $n = 7$ and a t-test (assuming the population is normal or nearly so).

ANS. A test statistic of -2.21 tells us that $\bar{x} < 31$. This places \bar{x} in H_0 . We do not reject H_0 if \bar{x} is in H_0 .

q. Once again refer to your test (b). Suppose that the test statistic is evaluated for a z-test and found to be 2.21. What is the significance p_{SIG} ? Remember, p_{SIG} is defined to be the

probability of EVEN MORE evidence against H_0 . So in this case $p\text{SIG} = P(Z > 2.21)$ (does this make sense to you?).

ANS. $p\text{SIG} = P(Z > 2.21) = 0.0135$.

r. State the manner in which a test of any kind is always conducted when $p\text{SIG}$ and α are made available.

ANS. Reject H_0 if $p\text{SIG} < \alpha$.

s. Referring to (q)(r), what action is taken by the z-test? It must agree with (m).

ANS. Since $p\text{SIG} = 0.0135 < \alpha = 0.05$, we reject H_0 . It is the same action taken in (m). This is just a different way of running the test. It has the advantage that the criterion for rejection “reject if $p\text{SIG} < \alpha$ ” is universal to all tests.

t. Suppose that a preliminary sample of $n_0 = 50$ is selected and the sample standard deviation of this sample of 50 scores is $s_0 = 12.7$. Determine a sample size n adequate to achieve both your α and your β from your curve (g). The required formula is in section 7-5 but you are substituting your preliminary estimate s_0 (obtained from a sample of $n_0 = 50$) for the true, unknown, population σ .

ANS.

$$n = ((|z_0| + |z_1|) s / (\mu_0 - \mu_1))^2$$

where $z_0 = z_{\alpha} = z_{0.05} \sim 1.65$ (had the test been two sided you'd have used $z_0 = z_{\alpha/2} = z_{0.025}$ and $z_1 = z_{\beta} = z_{0.71} = -0.55$ (even in the two sided case we would not use $\beta/2$). So

$$n = ((1.65 + 0.55) 12.7 / (31 - 30))^2 \sim 781.$$

We'd need an additional $781 - 50 = 731$ samples.

u. Refer to (t). Suppose we continue sampling from $n_0 = 50$ up to the recommended sample size $n = 781$. You are thus incorporating an additional 731 samples, per the recommendation of the formula. You would then evaluate the HYBRID test statistic

$$(x\text{BAR}_n - 31)/(s_0 / \text{root}(n)).$$

It is termed hybrid for the obvious reason that the sample standard deviation s_0 from the initial 50 is used but the sample mean $x\text{BAR}_n$ of all n is used as is $\text{root}(n)$. **NO ANSWER IS**

REQUIRED, JUST UNDERSTAND THAT THE HYBRID TEST STATISTIC NEEDS TO BE USED IN THIS CASE.

v. The test achieving your alpha and beta per (g) is then to reject H_0 if the HYBRID test statistic EXCEEDS the threshold you would have used for any z-test with your alpha of (g). Determine this z_{α} and state the criterion for rejecting H_0 in terms of the value of the hybrid test statistic.

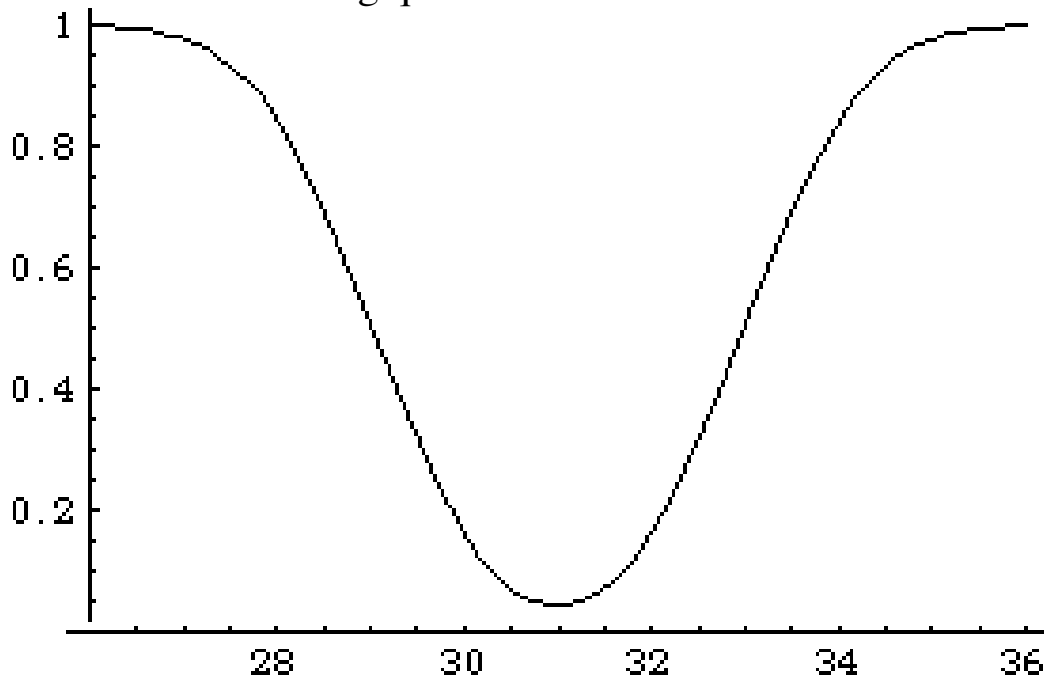
ANS. $z_{\alpha} = z_{0.05} = 1.65$. Reject H_0 if the hybrid test statistic exceeds this threshold.

3. Same as (1) but $H_0: \mu = 31$, $H_1: \mu$ is not 31, and we are interested in power at $\mu = 32$.

a. 1 or 2 sided?

ANS. TWO SIDED

b. Sketch a curve $P(\text{reject } H_0 \mid \mu)$ representing the general shape seen for the z-test or, for normal population, the t-test. You will need to set up scales for the horizontal and vertical axes to deal with the following questions.



c. In your curve identify the approximate numerical value of the probability of type I error (i.e. error of first kind, i.e. alpha).

ANS. $\alpha \sim 0.045 =$ height of this curve at $\mu = 31$, the boundary between H_0 and H_1 . It is not easy to get it this precisely by viewing the plot.

d. In your curve identify the approximate numerical value of the probability of type II error (i.e. error of the second kind, i.e. beta) for $\mu = 32$.

ANS. ~ 0.84 . This is much too large.

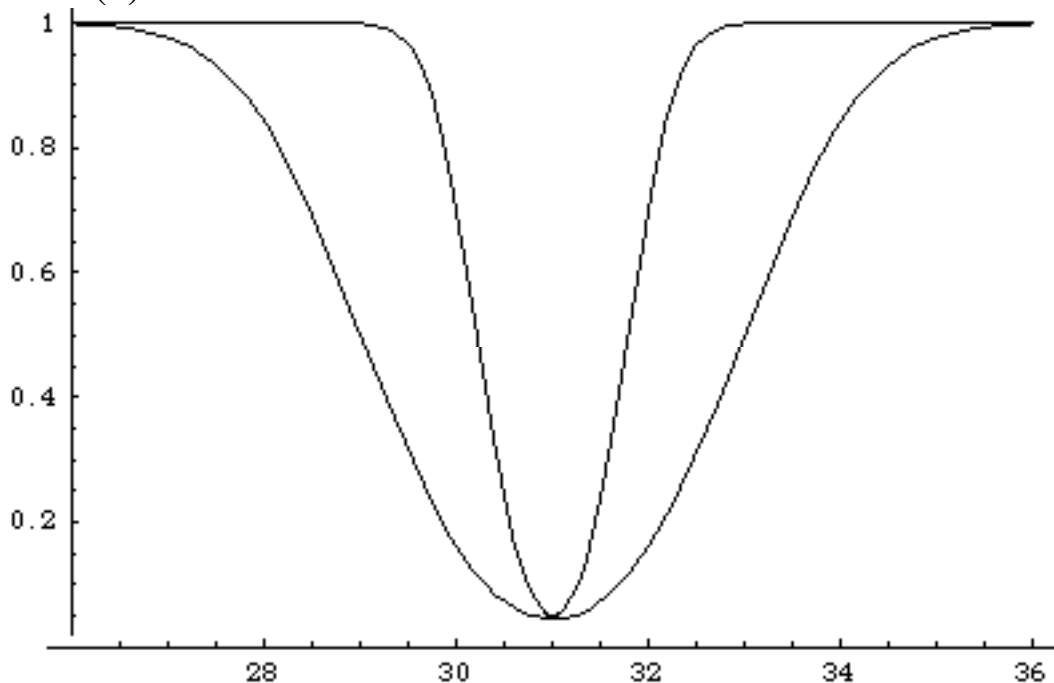
e. In (d) determine the approximate numerical value of the power of the test at $\mu = 32$.

ANS. Height of curve at $\mu = 30 = 1 - \beta \sim 0.16$. We desire far larger power.

f. Overlay on your sketch the ideal, but unattainable, plot of $P(\text{reject } H_0 \mid \mu)$.

ANS. = 0 at $\mu = 31$ and 1 for all other μ .

g. Overlay on your sketch another plausible $P(\text{reject } H_0 \mid \mu)$ curve of the correct general shape, having the same α as your curve (b) but closer to the ideal for all other μ than 31.



h. There is no question (h).

i. For your curve (g) what is the power at $\mu = 32$? What is beta for $\mu = 32$?

ANS. These are difficult to see by eye. Power = 0.7054 (height at $\mu = 32$) and $\beta = 1 - \text{power} = 0.2946$. We like larger power and consequently lower β .

j. Describe the two basic ways in which one may obtain a test with better performance as illustrated in (i)?

ANS. (Way 1) Increase the size of the sample. Each doubling of the sample steepens the curve's rise by a factor of two. (Way 2) Base the test on a better estimator of μ (one having a smaller sd). For example, a test based upon the regression estimator.

k. What is the test statistic based upon with-replacement samples $X_1 \dots X_n$?

ANS. $(\bar{x} - 31) / (s / \sqrt{n})$.

l. For your curve (b), what is the criterion for rejecting H_0 ? Be explicit, determining the threshold for rejection compatible with your α and the comparison with the test statistic that leads to rejection of H_0 .

ANS. The key point is that $\alpha \sim 0.045$ and SINCE TEST IS TWO SIDED the threshold is (using your z-table) is $z_{\alpha/2} = z_{0.0225} \sim 2.01$. Reject H_0 if the ABSOLUTE VALUE of the test statistic exceeds 2.01. (Using the program we'd find a more accurate $z_{\text{Tail}[2.0045]} = 0.022508$.)

m. If the test statistic is evaluated for the data and is found to be 2.21 what action is taken by a z-test?

ANS. Since the ABSOLUTE VALUE OF the test statistic 2.21 > 2.01 we reject H_0 .

n. If the test statistic is evaluated for the data and is found to be -2.21 what action is taken by a z-test?

ANS. Exactly the same as (m) since the absolute value of the test statistic is unchanged.

o. Repeat (m) for $n = 7$ and a t-test (assuming the population is normal or nearly so).

ANS. This is not available from your t-table. We find by experimentation with different values that $t_{\text{Tail}[2.5274, 6]} \sim .0225003$. So $t_{0.0225} \sim 2.5274$. Since the absolute value of the test

statistic = 2.21 < 2.5274 we fail to reject H0. This test is based upon a mere 7 sample size and simply lacks the strength of evidence to reject H0.

p. Repeat (n) for $n = 7$ and a t-test (assuming the population is normal or nearly so).

ANS. Same as in (p) since the absolute value of the test statistic is unchanged. Remember, the two sided test rejects H0 if the ABSOLUTE VALUE of the test statistic exceeds $z_{\alpha / 2}$ or $t_{\alpha / 2}$.

q. Once again refer to your test (b). Suppose that the test statistic is evaluated for a z-test and found to be 2.21. What is the significance pSIG? Remember, pSIG is defined to be the probability of EVEN MORE evidence against H0. So in this case $pSIG = P(|Z| > 2.21)$ (does this make sense to you?).

ANS. $pSIG = P(|Z| > 2.21) = 2(0.0135) \sim 0.027$.

r. State the manner in which a test of any kind is always conducted when pSIG and alpha are made available.

ANS. Reject H0 if $pSIG < \alpha$.

s. Referring to (q)(r), what action is taken by the z-test? It must agree with (m).

ANS. Since $pSIG = 0.027 < \alpha = 0.045$, we reject Ho. It is the same action taken in (m). This is just a different way of running the test. It has the advantage that the criterion for rejection “reject if $pSIG < \alpha$ ” is universal to all tests.

t. Suppose that a preliminary sample of $n_0 = 50$ is selected and the sample standard deviation of this sample of 50 scores is $s_0 = 12.7$. Determine a sample size n adequate to achieve both your alpha and your beta from your curve (g). The required formula is in section 7-5 but you are substituting your preliminary estimate s_0 (obtained from a sample of $n_0 = 50$) for the true, unknown, population sigma.

ANS.

$$n = ((|z_0| + |z_1|) s / (\mu_0 - \mu_1))^2$$

The test is two sided so $z_0 = z_{\alpha/2} = z_{0.0225} = 2.01$ and $z_1 = z_{\beta} = z_{0.2946} = 0.54$ (even in the two sided case we would not use $\beta/2$). So

$$n = ((2.01 + 0.54) 12.7 / (31 - 30))^2 \sim 1049.$$

So we'd need an additional $1049 - 50 = 999$ samples.

u. Refer to (t). Suppose we continue sampling from $n_0 = 50$ up to the recommended sample size $n = 1049$. You are thus incorporating an additional 999 samples, per the recommendation of the formula. You would then evaluate the HYBRID test statistic

$$(\bar{x}_{n} - 31) / (s_0 / \sqrt{n}).$$

It is termed hybrid for the obvious reason that the sample standard deviation s_0 from the initial 50 is used but the sample mean \bar{x}_n of all n is used as is \sqrt{n} .

JUST UNDERSTAND THAT THE HYBRID TEST STATISTIC NEEDS TO BE USED IN THIS CASE.

v. The test achieving your alpha and beta per (g) is then to reject H_0 if the HYBRID test statistic EXCEEDS the threshold you would have used for any z-test with your alpha of (g). Determine this z_{α} and state the criterion for rejecting H_0 in terms of the value of the hybrid test statistic.

ANS. $z_{\alpha} = z_{0.0225} = 2.01$. Reject H_0 if the ABSOLUTE VALUE OF the hybrid test statistic exceeds this threshold.

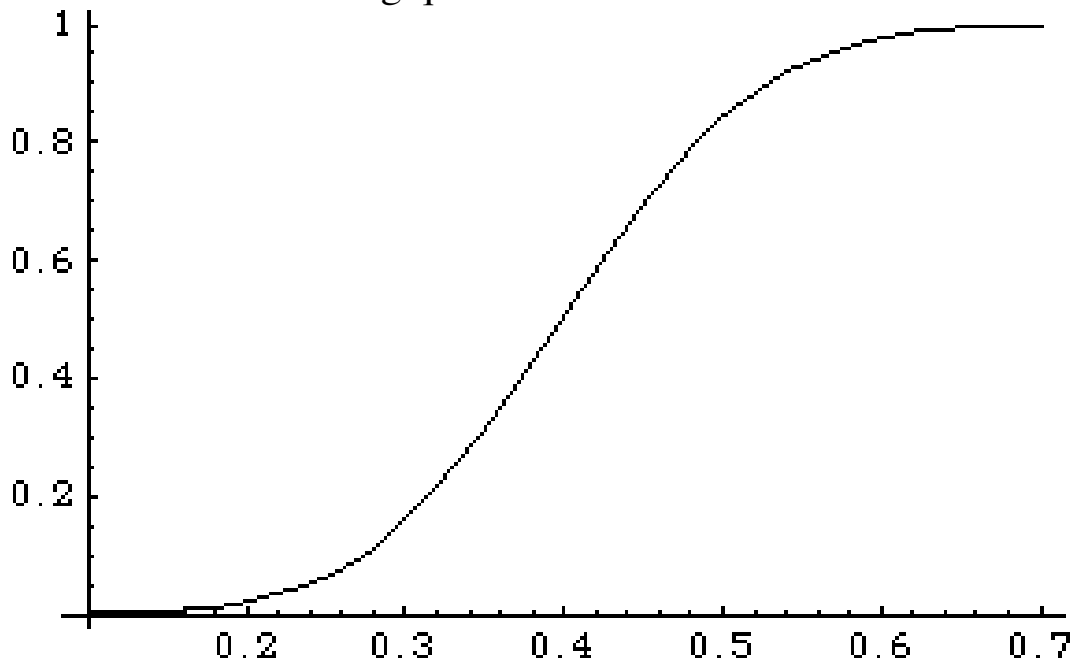
4. Same as (1) but $H_0: p$ is less or equal to 0.3, $H_1: p > 0.3$, and we are interested in power at $p = 0.4$. Parts (t)(u)(v) are simpler since the 0-1 data counterpart of the formula for n (section 7-5) does not require an unknown population sigma since $p_0 = 0.3$ determines an sd of $\sqrt{0.3 \cdot 0.7}$ whereas alternative $p_1 = 0.4$ determines an sd of $\sqrt{0.4 \cdot 0.6}$.

a. 1 or 2 sided?

ANS. ONE SIDED

b. Sketch a curve $P(\text{reject } H_0 \mid \mu)$ representing the general shape seen for the z-test or, for normal population, the t-test. You

will need to set up scales for the horizontal and vertical axes to deal with the following questions.



c. In your curve identify the approximate numerical value of the probability of type I error (i.e. error of first kind, i.e. alpha).

ANS. $\alpha \sim 0.159 =$ height of this curve at $p = 0.3$, the boundary between H_0 and H_1 .

d. In your curve identify the approximate numerical value of the probability of type II error (i.e. error of the second kind, i.e. beta) for $p = 0.4$.

ANS. ~ 0.5 . This is way too large.

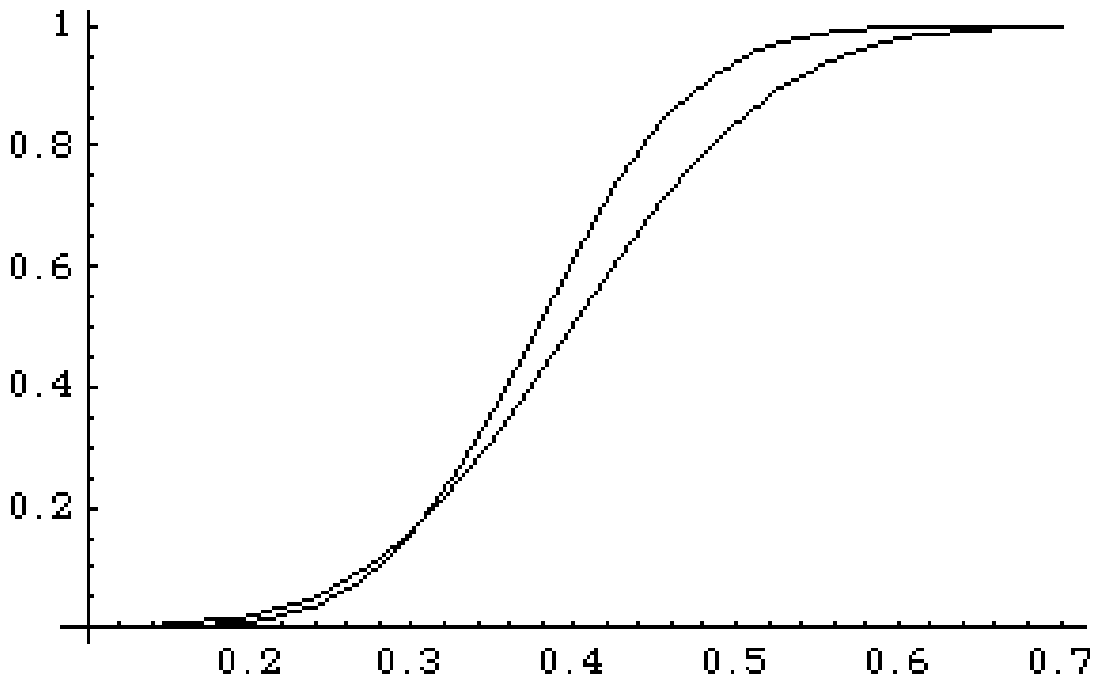
e. In (d) determine the approximate numerical value of the power of the test at $p = 0.4$.

ANS. height of curve at $p = 0.4 = 1 - \beta \sim 0.5$. We desire far larger power.

f. Overlay on your sketch the ideal, but unattainable, plot of $P(\text{reject } H_0 \mid \mu)$.

ANS. plot = 0 over H_0 and 1 over H_1 .

g. Overlay on your sketch another plausible $P(\text{reject } H_0 \mid \mu)$ curve of the correct general shape, having the same alpha as your curve (b) but closer to the ideal for all other p than 0.3.



h. There is no question (h).

i. For your curve (g) what is the power at $p = 0.4$? What is beta for $p = 0.4$?

ANS. These are difficult to see by eye. Power = 0.614 (height at $p = 0.4$) and beta = $1 - \text{power} = 0.386$. Still not so good. We want far larger power and consequently far lower beta.

j. Describe the two basic ways in which one may obtain a test with better performance as illustrated in (i)?

ANS. (Way 1) Increase the size of the sample. Each doubling of the sample steepens the curve's rise by a factor of two. (Way 2) Base the test on a better estimator of μ (one having a smaller sd). For example, a test based upon a regression estimator.

k. What is the test statistic based upon with-replacement samples $X_1 \dots X_n$?

ANS. $(\hat{p} - 0.3) / (\sqrt{0.3 \cdot 0.7} / \sqrt{n})$. Notice that we use $\sqrt{0.3 \cdot 0.7}$ and not the usual $\sqrt{\hat{p} \hat{q}}$ from confidence interval theory. The latter can make a test but for 0-1 data it is common to use $\sqrt{0.3 \cdot 0.7}$ (i.e. the boundary value 0.3 between H_0 and H_1). This gives a somewhat more accurate test for values of p near 0.3.

1. For your curve (b), what is the criterion for rejecting H_0 ? Be explicit, determining the threshold for rejection compatible with your alpha and the comparison with the test statistic that leads to rejection of H_0 .

ANS. The key point is that alpha \sim 0.159 and the threshold is $z_{\alpha} = z_{0.159} \sim 0.4368$ from $zTail[0.159] = 0.4368$. Reject H_0 if the test statistic exceeds 0.4368.

m. If the test statistic is evaluated for the data and is found to be 2.21 what action is taken by a z-test? Did you require any calculation to answer? Is $pHAT$ greater than 0.3 or less than 0.3?

ANS. Since the test statistic is positive 2.21 it must be that the numerator ($pHAT - 0.3$) is positive, so $pHAT > 0.3$. We reject H_0 because the test statistic exceeds threshold $z_{\alpha} = z_{0.159} \sim 0.4368$.

n. If the test statistic is evaluated for the data and is found to be -2.21 what action is taken by a z-test? Did you require any calculation to answer? Is $pHAT$ greater than 0.3 or less than 0.3?

ANS. Since the test statistic is negative we have $pHAT < 0.3$. As such, the data leans towards H_0 not towards H_1 . So our test, which tends to sit on H_0 unless there is strong evidence to the contrary, will not reject H_0 .

o. Repeat (m) for $n = 7$ and a t-test (assuming the population is normal or nearly so).

ANS. The 0-1 population is NOT normal and the t-test cannot be applied.

p. Repeat (n) for $n = 7$ and a t-test (assuming the population is normal or nearly so).

ANS. The 0-1 population is NOT normal and the t-test cannot be applied.

q. Once again refer to your test (b). Suppose that the test statistic is evaluated for a z-test and found to be 2.21. What is the significance $pSIG$? Remember, $pSIG$ is defined to be the probability of EVEN MORE evidence against H_0 . So in this case $pSIG = P(Z > 2.21)$ (does this make sense to you?).

ANS. $pSIG = P(Z > 2.21) = 0.0135$.

r. State the manner in which a test of any kind is always conducted when pSIG and alpha are made available.

ANS. Reject H0 if pSIG < alpha.

s. Referring to (q)(r), what action is taken by the z-test? It must agree with (m).

ANS. Since pSIG = 0.0135 < alpha = 0.159, we reject H0. It is the same action taken in (m). This is just a different way of running the test. It has the advantage that the criterion for rejection “reject if pSIG < alpha” is universal to all tests.

t. Determine a sample size n adequate to achieve both your alpha and your beta from your curve (g). The required formula is in section 7-5. There is no need for any preliminary sample (to establish a sigma) because for 0-1 data the sd are root(0.3 0.7) or root(0.4 0.6) for the two p-values 0.3 and 0.4.

ANS.

$$n = ((|z_0| \text{root}(0.3 \ 0.7) + |z_1| \text{root}(0.4 \ 0.6)) / (p_0 - p_1))^2$$

where $z_0 = z_{\alpha} = z_{0.159} \sim 0.4368$ (had the test been two sided you'd have used $z_0 = z_{\alpha/2} = z_{0.0795}$ and $z_1 = z_{\beta} = z_{0.5} \sim 0$ (even in the two sided case we would not use beta/2). So

$$n = ((0.4368 \text{root}(.3 \ .7) + 0 \text{root}(.4 \ .6)) / (0.3 - 0.4))^2 \\ \sim 4.$$

We knew from the poor beta that this test is not particularly good at discriminating between H0, H1 (i.e. P(reject H0 | p) doesn't rise sharply enough). Well, that kind of performance comes at a paltry $n = 4$? This is a little troubling since the calculations are based on normal approximations that are themselves untrustworthy for such a small n.

u. NO HYBRID TEST STATISTIC NEEDS TO BE USED IN THIS CASE.

v. Our sample size came in at $n = 4$ but we did not trust the calculation. Suppose it had come in at $n = 34$. The test achieving your alpha and beta per (g) is then to reject H0 if the usual test statistic $(\text{pHAT} - 0.3) / ((\text{root}(0.3 \ 0.7) / \text{root}(34))$

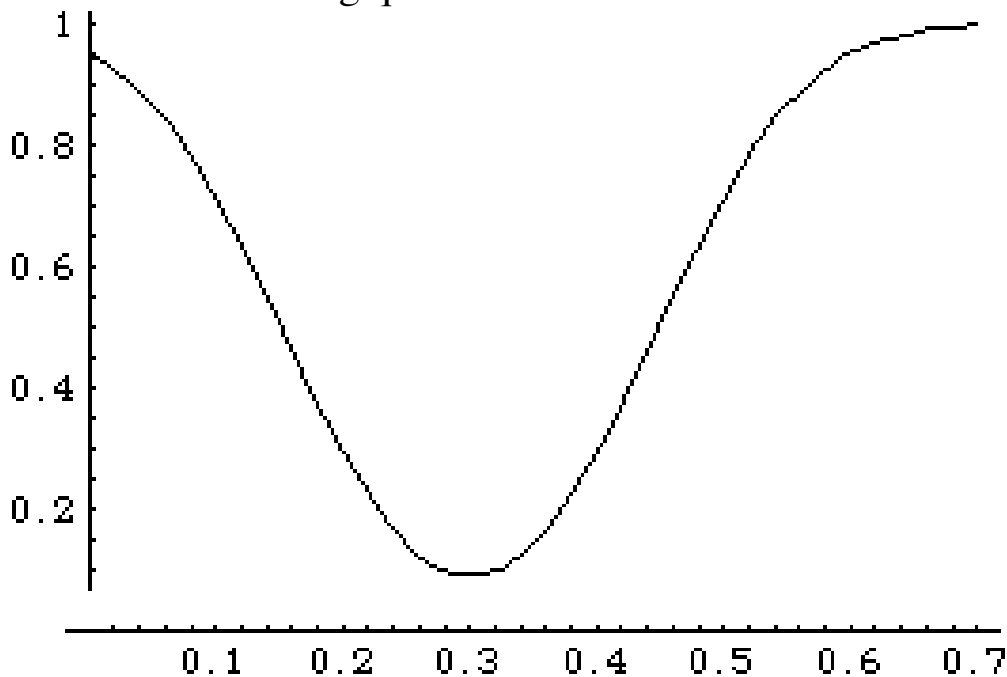
EXCEEDS the threshold 0.4368 you use for the z-test with alpha 0.159.

5. Same as (4) but $H_0: p = 0.3$, $H_1: p$ is not 0.3.

a. 1 or 2 sided?

ANS. TWO SIDED

b. Sketch a curve $P(\text{reject } H_0 \mid \mu)$ representing the general shape seen for the z-test or, for normal population, the t-test. You will need to set up scales for the horizontal and vertical axes to deal with the following questions.



Note: This curve was based on a normal approximation plot. You can see it is not correct for $p = 0$ since $P(\text{rej } H_0 \mid p = 0) = 1$.

c. In your curve identify the approximate numerical value of the probability of type I error (i.e. error of first kind, i.e. alpha).

ANS. $\alpha \sim 0.094 =$ height of this curve at $p = 0.3$, the boundary between H_0 and H_1 . It is hard to see from the curve.

d. In your curve identify the approximate numerical value of the probability of type II error (i.e. error of the second kind, i.e. beta) for $p = 0.4$.

ANS. ~ 0.71 . This is way too large.

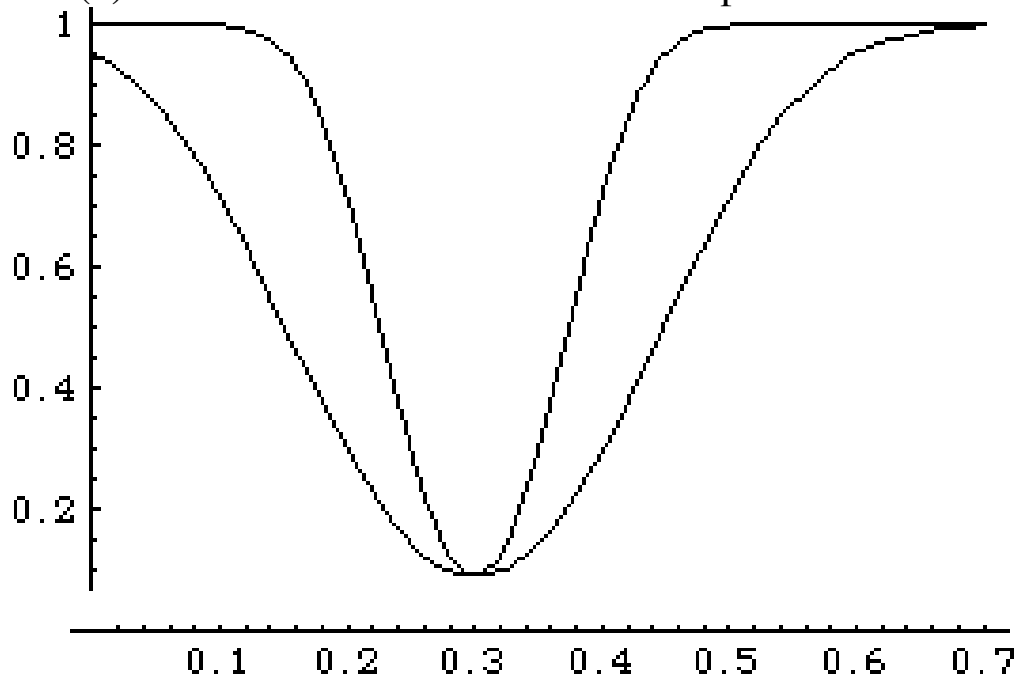
e. In (d) determine the approximate numerical value of the power of the test at $p = 0.4$.

ANS. height of curve at $p = 0.3 = 1 - \beta \sim 0.29$. We desire far larger power.

f. Overlay on your sketch the ideal, but unattainable, plot of $P(\text{reject } H_0 \mid \mu)$.

ANS. plot = 0 at $p = 0.3$ and 1 for all other p in $[0, 1]$.

g. Overlay on your sketch another plausible $P(\text{reject } H_0 \mid \mu)$ curve of the correct general shape, having the same alpha as your curve (b) but closer to the ideal for all other p than 0.3.



h. There is no question (h).

i. For your curve (g) what is the power at $p = 0.4$? What is beta for $p = 0.4$?

ANS. These are difficult to see by eye. Power = 0.712 (height at $p = 0.4$) and $\beta = 1 - \text{power} = 0.288$. Still not so good. We want larger power and consequently lower beta.

j. Describe the two basic ways in which one may obtain a test with better performance as illustrated in (i)?

ANS. (Way 1) Increase the size of the sample. Each doubling of the sample steepens the curve's rise by a factor of two. (Way 2) Base the test on a better estimator of μ (one having

a smaller sd). For example, a test based upon a regression estimator.

k. What is the test statistic based upon with-replacement samples $X_1 \dots X_n$?

ANS. $(\hat{p} - 0.3) / (\sqrt{0.3 \cdot 0.7} / \sqrt{n})$. Notice that we use $\sqrt{0.3 \cdot 0.7}$ and not the usual $\sqrt{\hat{p} \cdot \hat{q}}$ from confidence interval theory. The latter can make a test but for 0-1 data it is common to use $\sqrt{0.3 \cdot 0.7}$ (i.e. the boundary value 0.3 between H_0 and H_1). This gives a somewhat more accurate test for values of p near 0.3.

l. For your curve (b), what is the criterion for rejecting H_0 ? Be explicit, determining the threshold for rejection compatible with your alpha and the comparison with the test statistic that leads to rejection of H_0 .

ANS. From your z-table $z_{\alpha/2} = z_{0.047} \sim 1.68$. Reject H_0 if the ABSOLUTE VALUE of the test statistic (k) exceeds 1.68. Remember, this is a two sided test. (More accurately, $z_{\text{Tail}[1..6747]} = 0.0469966$.)

m. If the test statistic is evaluated for the data and is found to be 2.21 what action is taken by a z-test?

ANS. We reject H_0 because the ABSOLUTE VALUE of the test statistic exceeds threshold $z_{\alpha/2} = z_{0.047} \sim 1.68$.

n. If the test statistic is evaluated for the data and is found to be -2.21 what action is taken by a z-test?

ANS. Same as (m) since $|-2.21| = 2.21$.

o. Repeat (m) for $n = 7$ and a t-test (assuming the population is normal or nearly so).

ANS. The 0-1 population is NOT normal and the t-test cannot be applied.

p. Repeat (n) for $n = 7$ and a t-test (assuming the population is normal or nearly so).

ANS. The 0-1 population is NOT normal and the t-test cannot be applied.

q. Once again refer to your test (b). Suppose that the test statistic is evaluated for a z-test and found to be 2.21. What is the significance pSIG? Remember, pSIG is defined to be the probability of EVEN MORE evidence against H0. So in this case $pSIG = P(|Z| > 2.21)$ (does this make sense to you?).

ANS. $pSIG = P(|Z| > 2.21) = 2 (0.0135) = 0.027$.

r. State the manner in which a test of any kind is always conducted when pSIG and alpha are made available.

ANS. Reject H0 if pSIG < alpha.

s. Referring to (q)(r), what action is taken by the z-test? It must agree with (m).

ANS. Since $pSIG = 0.027 < \alpha = 0.094$, we reject H0. It is the same action taken in (m). This is just a different way of running the test. It has the advantage that the criterion for rejection “reject if pSIG < alpha” is universal to all tests.

t. Determine a sample size n adequate to achieve both your alpha and your beta from your curve (g). The required formula is in section 7-5. There is no need for any preliminary sample (to establish a sigma) because for 0-1 data the sd are $\sqrt{0.3 \cdot 0.7}$ or $\sqrt{0.4 \cdot 0.6}$ for the two p-values 0.3 and 0.4.

ANS.

$$n = ((|z_0| \sqrt{0.3 \cdot 0.7}) + |z_1| \sqrt{0.4 \cdot 0.6}) / (p_0 - p_1))^2$$

where $z_0 = z_{\alpha/2} = z_{0.047} \sim 1.68$ and $z_1 = z_{\beta} = z_{0.71} \sim -0.55$ (even in the two sided case we would not use beta/2). So

$$n = ((1.68 \sqrt{0.3 \cdot 0.7}) + 0.55 \sqrt{0.4 \cdot 0.6}) / (0.3 - 0.4))^2 \\ \sim 109.$$

u. NO HYBRID TEST STATISTIC NEEDS TO BE USED IN THIS CASE.

v. Our sample size came in at $n = 109$. Remember, this is a two sided test. The test achieving your alpha and beta per (g) is then to reject H0 if the ABSOLUTE VALUE of the usual test statistic $(\hat{p} - 0.3) / ((\sqrt{0.3 \cdot 0.7}) / \sqrt{109})$ exceeds $z_{\alpha/2} = z_{0.047} \sim 1.68$.

6. On the last exam we had a problem in the spirit of this one. A population has mean 55.4 and sd 17.9.

a. The sd of a r.v. X denoting a sample of one from this population.

ANS. It is the population sd = $\sqrt{0.3 \cdot 0.7} \sim 0.458$.

b. Is the population necessarily normal? Sketch two hypothetical population distributions having the given mean and sd.

ANS. The population is decidedly NOT normal. After all, the only possible population scores are 0 and 1. Sketch two distributions having mean 0.3 and sd 0.458. One could be normal (although it has no role here). Another is the distribution concentrating 0.3 on 1 and 0.7 on 0.

c. Give the mean (i.e. expectation) of the r.v. \bar{x} denoting the sample average of 60 selected with-replacement from this population.

ANS. $E \bar{x} = \mu = 0.3$. Remember, $\bar{x} = \hat{p}$ for 0-1 scores.

d. Give the standard deviation of r.v. \bar{x} from (c).

ANS. It is the sd of \hat{p} which is approximately the normal having mean 0.3 and $sd = \sqrt{0.3 \cdot 0.7} / \sqrt{60} \sim 0.059$.

e. Sketch the approximate distribution of \bar{x} offered by the central limit theorem. Label every important feature.

ANS. Sketch the normal with mean 0.3 and sd 0.059. Label the horizontal axis \hat{p} . Label the balance point 0.3 and the sd 0.059 (seen as the gap from the point of inflexion to the line issuing upwards from the mean).

7. In your sketch (6e) overlay the likely shape of a density prepared from a list of 1000 \bar{x} values, each one from an independent sample of 60 from the population.

ANS. It should resemble your sketch (6e). After all, (6e) is a description of what (all, or lots of) independently selected \bar{x} look like.

8. Suppose you have access to a particular random with-replacement sample SAM of 60 drawn from this population. In your sketch (6e) overlay the likely shape of a density prepared from your sample.

ANS. It should approximate the density of the population. that is, the sample resembles the population.

9. Refer to (8). In your sketch (6e) overlay the likely shape of a density prepared from a list of 1000 \bar{x} values, each one from an independent sample of 60 selected with-replacement from SAM. For the sake of argument we'll suppose SAM has sample mean 56.3 (fairly close to the population mean of 55.4).

ANS. sketch a nearly normal distribution centered on 55.4 and having approximately the sd of your sketch (7). That is, your sample, by plotting lots of \bar{x} , reveals (nearly) the shape of the density of \bar{x} except that it gets the center slightly wrong, putting it at \bar{x} instead of μ . Sample means of samples from your sample resemble sample means of samples from your population, except they get the mean somewhat wrong.

11. Identify the 95% bootstrap CI in your sketch (9).

ANS. In a sketch of the density portrait of a list of many \bar{x} , each of them a sample mean of 60 drawn independently with replacement from our original sample (of 60) SAM, the 95% CI for μ is simply the interval, symmetric and centered on the sample mean \bar{x} of SAM, that contains 95% of these \bar{x} .

Note: To be most illustrative, your pictures (8)(9) are centered on the sample mean of SAM whereas your pictures (6)(7) are centered on the population mean 55.4. Also, take care to clearly identify all of the above individual sketches in the overlay.

12. What is the symbol denoting the chance we reject a null hypothesis when it is true (more specifically, the chance calculated at the boundary between the null and alternative)? Would you say that rejecting H_0 requires a rather high degree of proof? Can you offer an argument against the use of the language “accept H_0 ” when H_0 is not rejected?

ANS. Alpha is the symbol. Typically yes, rejecting H_0 requires a high degree of proof because alpha is set low. Saying we accept H_0 when the test fails to reject H_0 may convey a sense that the truth of H_0 has been supported. But in many cases we may lack sufficient evidence to reject H_0 simply because our test has low power against alternatives. This could happen in t-tests with small n , or when the population sd is large.