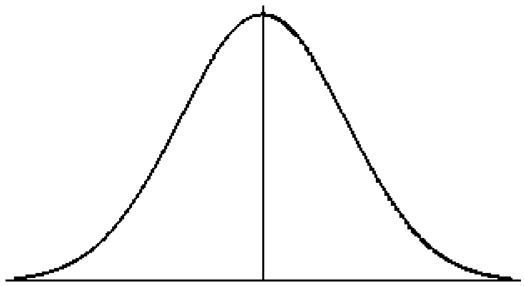
10.2 Tests of Significance

If we think the mean ACT here at FHN is 26, what are the chances of seeing results so low?



$\sigma = 2$

$n = 30$

sample mean = 25.2

Hypothesis: $\mu = 26$ assume

alternative $\mu < 26$

$P(\bar{x} \leq 25.2 \mid \mu = 26)$

| z | .00 | .01 | .02 | .03 | .04 | .05 | .06 | .07 | .08 | .09 |
|---|---|---|---|---|---|---|---|---|---|---|
| −3.4 | .0003 | .0003 | .0003 | .0003 | .0003 | .0003 | .0003 | .0003 | .0003 | .0002 |
| −3.3 | .0005 | .0005 | .0005 | .0004 | .0004 | .0004 | .0004 | .0004 | .0004 | .0003 |
| −3.2 | .0007 | .0007 | .0006 | .0006 | .0006 | .0006 | .0006 | .0005 | .0005 | .0005 |
| −3.1 | .0010 | .0009 | .0009 | .0009 | .0008 | .0008 | .0008 | .0008 | .0007 | .0007 |
| −3.0 | .0013 | .0013 | .0013 | .0012 | .0012 | .0011 | .0011 | .0011 | .0010 | .0010 |
| −2.9 | .0019 | .0018 | .0018 | .0017 | .0016 | .0016 | .0015 | .0015 | .0014 | .0014 |
| −2.8 | .0026 | .0025 | .0024 | .0023 | .0023 | .0022 | .0021 | .0021 | .0020 | .0019 |
| −2.7 | .0035 | .0034 | .0033 | .0032 | .0031 | .0030 | .0029 | .0028 | .0027 | .0026 |
| −2.6 | .0047 | .0045 | .0044 | .0043 | .0041 | .0040 | .0039 | .0038 | .0037 | .0036 |
| −2.5 | .0062 | .0060 | .0059 | .0057 | .0055 | .0054 | .0052 | .0051 | .0049 | .0048 |
| −2.4 | .0082 | .0080 | .0078 | .0075 | .0073 | .0071 | .0069 | .0068 | .0066 | .0064 |
| −2.3 | .0107 | .0104 | .0102 | .0099 | .0096 | .0094 | .0091 | .0089 | .0087 | .0084 |
| −2.2 | .0139 | .0136 | .0132 | .0129 | .0125 | .0122 | .0119 | .0116 | .0113 | .0110 |
| −2.1 | .0179 | .0174 | .0170 | .0166 | .0162 | .0158 | .0154 | .0150 | .0146 | .0143 |
| −2.0 | .0228 | .0222 | .0217 | .0212 | .0207 | .0202 | .0197 | .0192 | .0188 | .0183 |
| −1.9 | .0287 | .0281 | .0274 | .0268 | .0262 | .0256 | .0250 | .0244 | .0239 | .0233 |
| −1.8 | .0359 | .0351 | .0344 | .0336 | .0329 | .0322 | .0314 | .0307 | .0301 | .0294 |
| −1.7 | .0446 | .0436 | .0427 | .0418 | .0409 | .0401 | .0392 | .0384 | .0375 | .0367 |
| −1.6 | .0548 | .0537 | .0526 | .0516 | .0505 | .0495 | .0485 | .0475 | .0465 | .0455 |
| −1.5 | .0668 | .0655 | .0643 | .0630 | .0618 | .0606 | .0594 | .0582 | .0571 | .0559 |
| −1.4 | .0808 | .0793 | .0778 | .0764 | .0749 | .0735 | .0721 | .0708 | .0694 | .0681 |
| −1.3 | .0968 | .0951 | .0934 | .0918 | .0901 | .0885 | .0869 | .0853 | .0838 | .0823 |
| −1.2 | .1151 | .1131 | .1112 | .1093 | .1075 | .1056 | .1038 | .1020 | .1003 | .0985 |
| −1.1 | .1357 | .1335 | .1314 | .1292 | .1271 | .1251 | .1230 | .1210 | .1190 | .1170 |
| −1.0 | .1587 | .1562 | .1539 | .1515 | .1492 | .1469 | .1446 | .1423 | .1401 | .1379 |
| −0.9 | .1841 | .1814 | .1788 | .1762 | .1736 | .1711 | .1685 | .1660 | .1635 | .1611 |
| −0.8 | .2119 | .2090 | .2061 | .2033 | .2005 | .1977 | .1949 | .1922 | .1894 | .1867 |
| −0.7 | .2420 | .2389 | .2358 | .2327 | .2296 | .2266 | .2236 | .2206 | .2177 | .2148 |
| −0.6 | .2743 | .2709 | .2676 | .2643 | .2611 | .2578 | .2546 | .2514 | .2483 | .2451 |
| −0.5 | .3085 | .3050 | .3015 | .2981 | .2946 | .2912 | .2877 | .2843 | .2810 | .2776 |
| −0.4 | .3446 | .3409 | .3372 | .3336 | .3300 | .3264 | .3228 | .3192 | .3156 | .3121 |
| −0.3 | .3821 | .3783 | .3745 | .3707 | .3669 | .3632 | .3594 | .3557 | .3520 | .3483 |
| −0.2 | .4207 | .4168 | .4129 | .4090 | .4052 | .4013 | .3974 | .3936 | .3897 | .3859 |
| −0.1 | .4602 | .4562 | .4522 | .4483 | .4443 | .4404 | .4364 | .4325 | .4286 | .4247 |
| −0.0 | .5000 | .4960 | .4920 | .4880 | .4840 | .4801 | .4761 | .4721 | .4681 | .4641 |

Test of significance based on sampling distributions (ch. 9) and standard scores (like z) and tell us how much evidence we have against some claim.

**US Court System**

1. Assume innocent.
2. Determine legal procedings.
3. Assess evidence against innocence.
4. Make a decision about guilt.

**Significance Testing**

1. Assume some claim is true.
2. Determine procedure to follow.
3. Find the probability of getting such a statistic.
4. Make a decision about the claim.

Hypothesis:

- claim about a population &/or parameter.
- what we assume is true or want to prove

mu = mean hours of sleep

p = % of chips that are white

μ = mean mpg of all cars

**Never** write a hypothesis about a statistic!

Writing a hypothesis about a statistic is immediately

**WRONG!**

Using parameter symbols but statistic vocabulary is still wrong!

x-bar = mean hours of sleep

p = % of chips that were white

μ = mean mpg of the 60 cars

Null Hypothesis $H_o$

always has an = sign

no change
no effect
no difference
the null is dull

The true population proportion or mean of what you are trying to study equals the guess of the parameter.

See page 565 for more about $H_o$.

Alternative Hypothesis $H_a$

always has one of these symbols: $>$ $<$ $\neq$

there is a change
there is an effect
there is difference

The true population proportion or mean of what you are trying to study is greater than, less than, or not equal to the guess of the parameter.

reject Ho

- implies there's sufficient evidence that the Ho is false
- guilty
- implies there's sufficient evidence to convict

fail to reject Ho

- implies there's insufficient evidence that the Ho is false
- not guilty
- implies there's insufficient evidence to convict

---

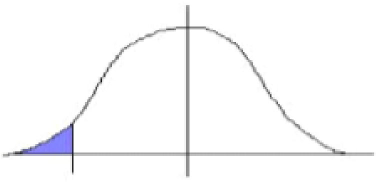
accept Ho

- implies proof that Ho is true
- innocent
- implies proof that one is innocent

Hypotheses for tests about the mean

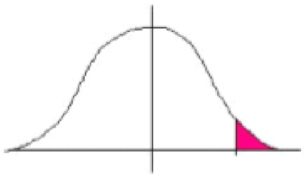
$H_o$: $\mu=\mu_o$
$H_a$: $\mu<\mu_o$

*Trying to prove the parameter is "less than" yields a left tailed test*



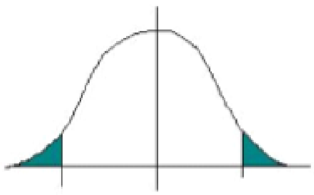
$H_o$: $\mu=\mu_o$
$H_a$: $\mu>\mu_o$

*Trying to prove the parameter is "more than" yields a right tailed test*



$H_o$: $\mu=\mu_o$
$H_a$: $\mu\neq\mu_o$

*Trying to prove the parameter is "not equal to" or "different than" yields a two tailed test*



P value:

- Probability the test statistic takes a value at least as extreme as that observed.
- A conditional probability
- P(seeing this statistic (or one more extreme) | $H_o$ is true)
- Probability of seeing a statistic at least this extreme if $H_o$ is true.
- When small, is strong evidence against $H_o$.
- page 567

significance level, $\alpha$:

- Greek letter "alpha"
- decision level of evidence.
- area of distribution's "critical" or "rejection" region(s).
- largest p value at which we reject $H_o$.
- frequently set at 0.05 or 0.01
- 1-confidence level
- when P-value<$\alpha$, we have statistically significant result.
- page 569

A Significance level is

• the p value that is the criterion for saying that $H_o$ is either true or not true.

• also called the $\alpha$ (alpha) level.

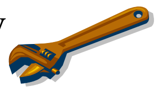
• generally set at .05 or .01.

• the area of the rejection region (also called the critical region) of a distribution.

From now on, **always** follow one of these methods:

**The inference toolbox** (page 571)

Step 1: ID the population and the parameter we want to draw conclusions about.



Step 2: Choose the appropriate inference procedure and verify the conditions for the procedure.

Step 3: If the conditions are met, find the test statistic and the P-value.

Step 4: Interpret the results in the context of the problem.

**PHANTOMS** (my usual method)

Population and parameter

Hypotheses in context

Assumptions/conditions verified

Name of the procedure

Test statistic found

Obtain a P value

Make a decision about $H_o$

State a conclusion in context

A comparison of PHANTOMS and the Inference Toolbox

| "PHANTOMS" | | **Inference toolbox:** |
|---|---|---|
| P | parameter | **1** |
| H | hypotheses | |
| A | assumptions | **2** |
| N | name of test | **3** |
| T | test statistic | |
| O | obtain p-value | |
| M | make decision | |
| S | state conclusion in context | **4** |

ex. Citing the problem of childhood obesity, a researcher believes the mean weight of children of a particular age has changed. The long-held belief is that children this age have a mean weight of 58 pounds. Assume the distribution of weights of children at a specific age are roughly normal with a standard deviation of 3 pounds. The researcher selects a SRS of 25 children and finds their mean weight to be 62 pounds. Is there evidence that the mean weight has changed?

Details! Details!

Show you verified assumptions/conditions.
Don't just list them and check them off.

Specify procedure, by name or formula.

Identify values or show them substituted into formula.

Be sure your conclusion is in context.
Generic statements don't earn you points.

page 572

To test the hypothesis $H_o$: $\mu = \mu_o$ based on an SRS of size n from a population with unknown mean $\mu$ and known standard deviation $\sigma$, compute the one-sample z statistic

$$z = \frac{\bar{x} - \mu_0}{\frac{\sigma}{\sqrt{n}}}$$

The P-value for a test of $H_o$ against

$H_a$: $\mu > \mu_o$ is $P(Z>z)$

$H_a$: $\mu < \mu_o$ is $P(Z<z)$

$H_a$: $\mu \neq \mu_o$ is $2P(Z>z)$

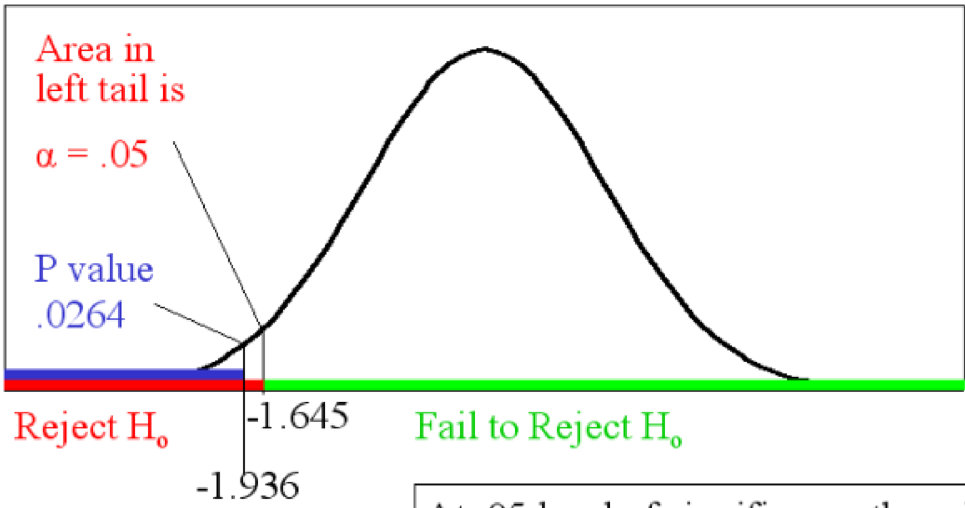
Where

$$z = \frac{\bar{x} - \mu_0}{\frac{\sigma}{\sqrt{n}}}$$

The P-values are exact if the population distribution is Normal. Otherwise, they are approximate for large n.

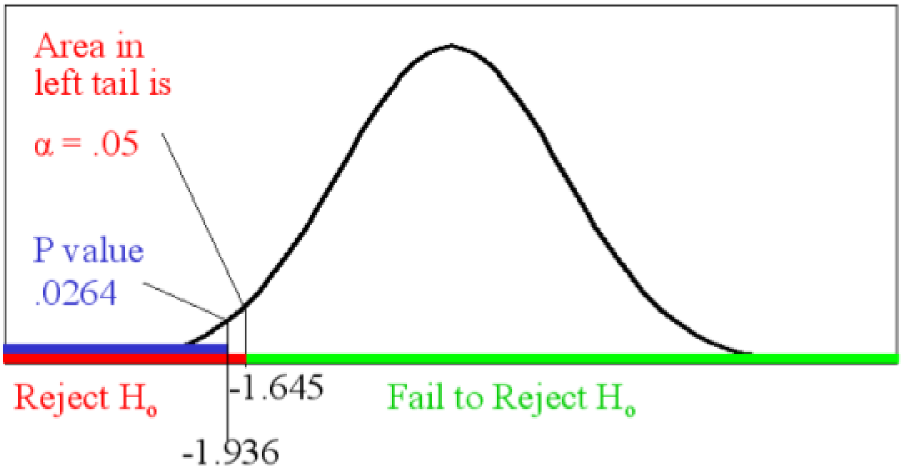
Ex. Company A uses bug repellent for its workers. The current repellent protects for 8 hours. A cheaper repellent is being considered. Assume $\alpha$=.05, that effective times are Normally distributed, and that $\sigma$=2 hours. Is there evidence that the cheaper repellent protects less than 8 hours? The cheaper repellent protected a SRS of 15 workers for an average of 7 hours.

# Hypothesis Testing




At .05 level of significance there is sufficient evidence to support the hypothesis that the average effective life of the cheaper repellent is less than 8 hours.

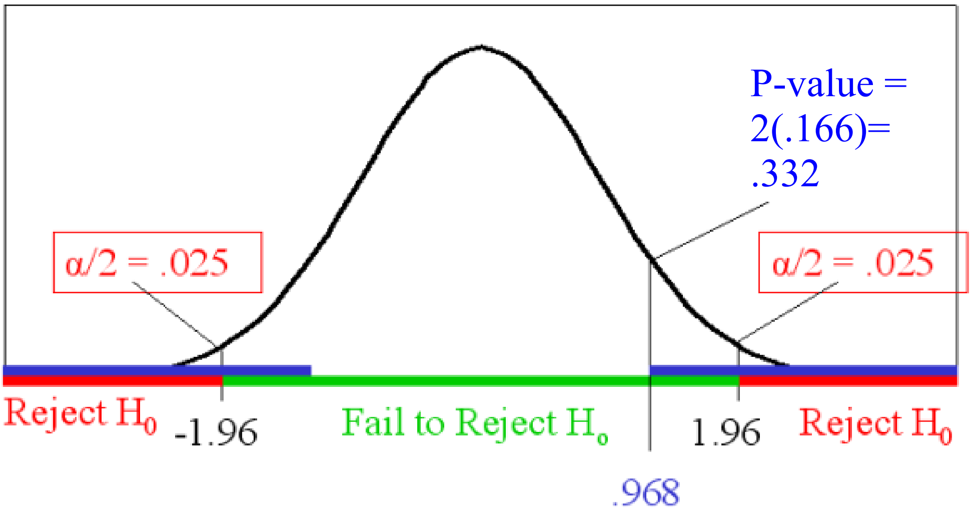
# Hypothesis Testing




.0264 is the p-value. A P-value is the lowest level (of significance) at which the observed value of the test statistic is significant. It is the lowest level at which you could have set α and still rejected the null hypothesis.

Ex. Company B also uses bug repellent for its workers. The current repellent protects for 8 hours. A plant-based repellent is being considered. Assume $\alpha$=.05, that effective times are Normally distributed, and that $\sigma$=2 hours. Is there evidence that the plant-based repellent is any different from the current one? The plant-based repellent protected a SRS of 15 workers for an average of 8.5 hours.

# Hypothesis Testing




The p-value is now 2(.166)= .332

This is the lowest level at which you could have set α and still rejected the null hypothesis.

At .05 level of significance there is no reason to believe that the average effective life of the plant-based repellent is different from 8 hours.

**<u>We can also use a CI for a significance test.</u>**

page 581

We reject $H_o$: $\mu = \mu_o$ at sig. level $\alpha$ in favor of $H_a$: $\mu \neq \mu_o$ if $\mu_o$ falls outside our $1-\alpha$ CI.

Ex. Survey of Study Habits and Attitudes (SSHA)

- a psychological test
- measures motivation, attitude toward school, & study habits
- scores range from 0 to 200
- for U.S. college students $\mu \approx 115$ and $\sigma \approx 30$

A teacher suspects that older students have better attitudes toward school & gives SSHA to 20 students over age 30. $\bar{x} = 135.2$. Assume $\sigma = 30$ for the older students. Construct a 90% confidence interval for the mean score $\mu$ for older students. Is there evidence that the mean score for older students differs from that of the general college population?

| z | .00 | .01 | .02 | .03 | .04 | .05 | .06 | .07 | .08 | .09 |
|---|---|---|---|---|---|---|---|---|---|---|
| 0.0 | .5000 | .5040 | .5080 | .5120 | .5160 | .5199 | .5239 | .5279 | .5319 | .5359 |
| 0.1 | .5398 | .5438 | .5478 | .5517 | .5557 | .5596 | .5636 | .5675 | .5714 | .5753 |
| 0.2 | .5793 | .5832 | .5871 | .5910 | .5948 | .5987 | .6026 | .6064 | .6103 | .6141 |
| 0.3 | .6179 | .6217 | .6255 | .6293 | .6331 | .6368 | .6406 | .6443 | .6480 | .6517 |
| 0.4 | .6554 | .6591 | .6628 | .6664 | .6700 | .6736 | .6772 | .6808 | .6844 | .6879 |
| 0.5 | .6915 | .6950 | .6985 | .7019 | .7054 | .7088 | .7123 | .7157 | .7190 | .7224 |
| 0.6 | .7257 | .7291 | .7324 | .7357 | .7389 | .7422 | .7454 | .7486 | .7517 | .7549 |
| 0.7 | .7580 | .7611 | .7642 | .7673 | .7704 | .7734 | .7764 | .7794 | .7823 | .7852 |
| 0.8 | .7881 | .7910 | .7939 | .7967 | .7995 | .8023 | .8051 | .8078 | .8106 | .8133 |
| 0.9 | .8159 | .8186 | .8212 | .8238 | .8264 | .8289 | .8315 | .8340 | .8365 | .8389 |
| 1.0 | .8413 | .8438 | .8461 | .8485 | .8508 | .8531 | .8554 | .8577 | .8599 | .8621 |
| 1.1 | .8643 | .8665 | .8686 | .8708 | .8729 | .8749 | .8770 | .8790 | .8810 | .8830 |
| 1.2 | .8849 | .8869 | .8888 | .8907 | .8925 | .8944 | .8962 | .8980 | .8997 | .9015 |
| 1.3 | .9032 | .9049 | .9066 | .9082 | .9099 | .9115 | .9131 | .9147 | .9162 | .9177 |
| 1.4 | .9192 | .9207 | .9222 | .9236 | .9251 | .9265 | .9279 | .9292 | .9306 | .9319 |
| 1.5 | .9332 | .9345 | .9357 | .9370 | .9382 | .9394 | .9406 | .9418 | .9429 | .9441 |
| 1.6 | .9452 | .9463 | .9474 | .9484 | .9495 | .9505 | .9515 | .9525 | .9535 | .9545 |
| 1.7 | .9554 | .9564 | .9573 | .9582 | .9591 | .9599 | .9608 | .9616 | .9625 | .9633 |
| 1.8 | .9641 | .9649 | .9656 | .9664 | .9671 | .9678 | .9686 | .9693 | .9699 | .9706 |
| 1.9 | .9713 | .9719 | .9726 | .9732 | .9738 | .9744 | .9750 | .9756 | .9761 | .9767 |
| 2.0 | .9772 | .9778 | .9783 | .9788 | .9793 | .9798 | .9803 | .9808 | .9812 | .9817 |
| 2.1 | .9821 | .9826 | .9830 | .9834 | .9838 | .9842 | .9846 | .9850 | .9854 | .9857 |
| 2.2 | .9861 | .9864 | .9868 | .9871 | .9875 | .9878 | .9881 | .9884 | .9887 | .9890 |
| 2.3 | .9893 | .9896 | .9898 | .9901 | .9904 | .9906 | .9909 | .9911 | .9913 | .9916 |
| 2.4 | .9918 | .9920 | .9922 | .9925 | .9927 | .9929 | .9931 | .9932 | .9934 | .9936 |
| 2.5 | .9938 | .9940 | .9941 | .9943 | .9945 | .9946 | .9948 | .9949 | .9951 | .9952 |
| 2.6 | .9953 | .9955 | .9956 | .9957 | .9959 | .9960 | .9961 | .9962 | .9963 | .9964 |
| 2.7 | .9965 | .9966 | .9967 | .9968 | .9969 | .9970 | .9971 | .9972 | .9973 | .9974 |
| 2.8 | .9974 | .9975 | .9976 | .9977 | .9977 | .9978 | .9979 | .9979 | .9980 | .9981 |
| 2.9 | .9981 | .9982 | .9982 | .9983 | .9984 | .9984 | .9985 | .9985 | .9986 | .9986 |
| 3.0 | .9987 | .9987 | .9987 | .9988 | .9988 | .9989 | .9989 | .9989 | .9990 | .9990 |
| 3.1 | .9990 | .9991 | .9991 | .9991 | .9992 | .9992 | .9992 | .9992 | .9993 | .9993 |
| 3.2 | .9993 | .9993 | .9994 | .9994 | .9994 | .9994 | .9994 | .9995 | .9995 | .9995 |
| 3.3 | .9995 | .9995 | .9995 | .9996 | .9996 | .9996 | .9996 | .9996 | .9996 | .9997 |
| 3.4 | .9997 | .9997 | .9997 | .9997 | .9997 | .9997 | .9997 | .9997 | .9997 | .9998 |

The mean score for U.S. college students on the SSHA is about 115, and the standard deviation is about 30. A teacher suspects that older students have better attitudes toward school. She gives the SSHA to 25 students who are at least 30 years old. Assume that scores in the population of older students are normally distributed with standard deviation $\sigma=30$. Suppose that the sample mean for the 25 older students is 118.6 on the SSHA. Conduct an appropriate hypothesis test.

The following pages are additional detailed information for this section, in case you want more examples or help.

More about hypotheses and claims.

The null hypothesis is just that-- "null" -- nothing interesting, nothing has changed, no difference, not effective, etc. The burden of proof resides in the alternative hypothesis.

Someone might claim that something has changed or that it has not. Either way the null is no change.

Ex. Suppose we have a medication that's proven effective, but has now been slightly reformulated.

Scenario 1: We've added another ingredient and now we claim the new version is even more effective.
$H_o$: The cure rate has not changed
$H_a$: The cure rate is higher

Scenario 2: We've deleted an ingredient that was causing upset stomach in some people and now we claim the new version is still just as effective.
$H_o$: The cure rate has not changed
$H_a$: The cure rate is lower

In both cases the null is the same, but in the first case the claim is in the alternative hypothesis and in the second it's in the null.

What's different is the strength of the conclusion we can reach. In the first case, rejecting the null offers strong evidence that the claim is true (not quite a definitive proof of the claim, but certainly leaning in that direction). In the second, failing to reject the null merely says we lack evidence things have changed. This is nowhere near proof that the claim is true. We'll continue to hold the conjecture of equivalent effectiveness, but we certainly haven't proven it (which is why we never "accept" a null hypothesis).

A couple ways to remember whether you should reject $H_o$:

**P-value method**

p α

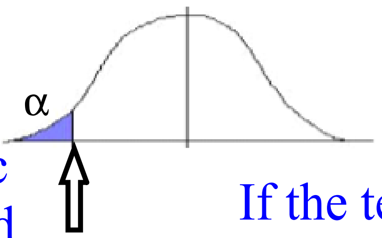
person already-set-up hurdle

> cleared the hurdle, so fail to reject Ho

≤ hit the hurdle, so reject Ho

---

**test statisic method**




If the test statistic falls in the shaded critical region (rejection region), then we reject Ho.

If the test statistic falls in the unshaded acceptance region, then we fail to reject Ho.

If the T(wisted) S(ister) goes shopping and the lights go out, she rejects her purchases; who would shop in the dark?

If the T(wisted) S(ister) goes shopping and the store is brightly lit, and not shaded, she keeps her purchases (fails to reject her purchases).

page 578 the critical value method

To test the hypothesis $H_0$: $\mu = \mu_o$ based on an SRS of size n from a population with unknown mean $\mu$ and known standard deviation $\sigma$, compute the one-sample z statistic.

Reject $H_o$ at significance level $\alpha$ against a one-sided alternative

Ha: $\mu > \mu_o$ if $z > z^*$

Ha: $\mu < \mu_o$ if $z < -z^*$

where z* is the upper critical value from Table C.

Reject $H_o$ at significance level $\alpha$ against a two-sided alternative

Ha: $\mu \neq \mu_o$ if $|z| > z^*$

where z* is the upper $\alpha/2$ critical value from Table C.