Good Morning! First of all, bad jokes are available on the web at

http://home.earthlink.net/~markharbison/harbisonweb/humor/

along with other resources for algebra and statistics students.
Or, my web page is the first one to come up under a "Google" search for "Mark Harbison".

Now, let's discuss a 4-step procedure to do a hypothesis test:

- step 1: translate the problem from words into symbols (H_o & H_1)
- step 2: find the "axis #" (test statistic z or t or χ^2 or F)
- step 3: find the "tail area" (pvalue normalcdf… or tcdf… or χ^2cdf… or Fcdf…)
- step 4: translate the result back into non-technical terms of the original problem.

I claim that statistics students' exam scores will increase significantly if questions on hypothesis tests are worded "nicely" to reflect the definition:

α = Prob(we conclude that the Null Hypothesis is false, when H_o is true in reality)

This definition requires that we assume that H_o is true. For example with H_o: $\mu = 5$ lbs ,

we will use the value "5" for μ. Thus, H_o will ALWAYS use "=", never any other symbol.

The alternate hypothesis must be the 'opposite' of "=". (Let's call it H_1 instead of H_a).
There are 3 ways to do this:

- if a keyword in the problem is like "greater" or "more", then use "$>$" (right-tailed)
- if a keyword is like "less" or "fewer" or "poorer", then use "$<$" (left-tailed)
- if there is no such 1-tail type of keyword given
  or if the keyword is "different" or "equal" or "same", then use "$\neq$" (2-tailed).

These 3 signs are the ONLY possible symbols allowed in H_1 since "=" was already used in H_o .

Finally for step 1, use the specific # used in the claim for the population—not the same # as given in the sentence with the word "sample" or "survey". Save all sample #s for step 2.

For step 2, most students are OK choosing the right "axis #" formula. However,
the average student does need extra practice getting parentheses in the right place for calculations.
For example, the key sequence 8 – 3 / 7 / √30 will not be the same as (8 – 3) / (7 / √ 30) .

On step 3, pvalues must be found either with a calculator or with alternate tables.
Traditional tables (e.g. t-table) include only famous areas like .01 or .05 . But an alternate table includes many other areas like .0483 and .0496 . Plus it is consistent with the "axis # in, area out" style of the z-table. Also, instead of basing a conclusion on a wide range of values,
the exact pvalue can give power to the reader to decide whether or not the result is significant.

Since most textbooks do not yet include alternate tables, then technology needs to be available to students. For example, on the TI-83 Stat TESTS menu, one button will find steps 2 & 3 all at once.

Mark Harbison. 2002. MisterHarbison@aol.com (916) 558-2687. Sacramento City College.

Once an exact pvalue has been found, then we are ready for step 4:

- if $\quad$ pvalue ≥ .10 then word the conclusion: “There is no evidence … (of a ‘difference’…)”
- if .05 ≤ pvalue < .10 then word the conclusion: “There is moderate evidence … (of a ‘diff’…)”
- if .01 ≤ pvalue < .05 then word the conclusion: “There is strong evidence … (of a ‘diff’…)”
- if $\quad$ pvalue < .01 then word the conclusion: “There is very strong evidence … (of a ‘diff’…)”.

The word ‘difference’ is good if H_1 used “≠” .
Otherwise, use the SAME 1-tail type word as already used (e.g. “more”, “fewer”, “better”, “worse”). Fill in the blanks with the same specific situation (e.g. $, gallons, inches) as were given in the original problem. Use the same population # as used in step 1 but do not use any sample #s.

(In most of mathematics, it is easier to find one counter-example to prove that something is false than to prove that something is true for many cases. And so it goes with hypothesis testing, also. We can gather evidence to show that a population parameter does NOT equal a given #, but we can not prove that the parameter does equal the given # since we only have sample data, not data for the entire population. Thus, never use the words “does equal” in step 4.)

A major objective of most course outlines is that students learn to “communicate their results”. This means writing non-technical conclusions that the “man on the street” can understand. It is NOT necessary for any part of a conclusion to use the word “claim” or “accept” or “reject” or “hypothesis”. For example (if $\mu > 5$ in step 1), it’s more <u>helpful</u> to say “there is strong evidence to conclude that the mean amount students spend per day for fast food is significantly more than $5” than to conclude simply “reject the null hypothesis”.

The following are some sample problems quoted from various current textbooks.

(from Moore’s “Basic Practice of Statistics, p. 330)
The NCHS reports that the mean systolic blood pressure for males 35 – 44 years of age is 128 and the standard deviation in this population is 15. The medical director of a large company looks at the medical records of 72 executives in this age group and finds that the mean systolic blood pressure in this sample is $\bar{x} = 126.07$. Is this evidence that the company’s executives have a different mean blood pressure from the general population?

<u>step 1</u> H_o: $\mu = 128$ & H_1: $\mu \neq 128$
<u>step 2</u> z = -1.09 by Stat TESTS #1 ZTest…
<u>step 3</u> pvalue = 2(0.1379) = 0.2758 > 0.10 or directly with ZTest (do not mult. by 2)
<u>step 4</u> There is no evidence that the mean systolic blood pressure of their executives is significantly different than 128 mm/Hg.

(from Triola’s “Elementary Statistics”, p. 385)
…the volumes (in ounces) of the Coke in a sample of 36 different cans… n = 36, $\bar{x}$ = 12.19 oz, s = 0.11 oz. Upon seeing these statistics, a line manager claims that the mean amount of Coke is greater than 12 oz, causing lower profits. Using a 0.05 significance level, test the manager’s claim that the mean is greater than 12 oz.

<u>step 1</u> H_o: $\mu = 12$ & H_1: $\mu > 12$
<u>step 2</u> z = 10.36 by Stat TESTS #1 ZTest…
<u>step 3</u> pvalue = 0.0001 by printed table or 2nd DISTR #2 normalcdf(10.36, 1000) 1.9×10^{-25}
<u>step 4</u> There is very very strong evidence that the mean amount of Coke is greater than 12 oz since pvalue < 0.01 (ignore $\alpha = 0.05$).

Mark Harbison. 2002. MisterHarbison@aol.com (916) 558-2687. Sacramento City College.

(from Mendenhall/Beaver/Beaver's "Brief Introduction to Probability and Statistics", p. 389)
…One manufacturer claims that a gallon of its paint will cover 400 square feet of surface area.
To test this claim, a random sample of ten 1-gallon cans of paint were used… [to get]
310 311 412 368 447 376 303 410 365 350 [square feet].
Do the data represent sufficient evidence to indicate that the average coverage differs from 400 ft^2?

<u>step 1</u> H_o: $\mu = 400$ & H_1: $\mu \neq 400$
<u>step 2</u> t = -2.27 by Stat TESTS #2 TTest… ('Data' option recommended)
<u>step 3</u> pvalue = 2(0.0235) = 0.0470 by an "alt. T table" or directly with TTest (do not mult. by 2)
or 2nd DISTR #5 2(tcdf(pos. 2.27, 1000, 9)) 2(0.0247) = 0.0494
<u>step 4</u> There is strong evidence that the average coverage does differ from 400 square feet
since $0.01 < 0.0494 < 0.05$.

(from Bluman's "Elementary Statistics", p. 440)
A dietician wishes to see if a person's cholesterol level will decrease if the diet is supplemented by a certain mineral. Six subjects were pretested and then took the mineral supplement for a six-week period [and then retested]… Can it be concluded that the cholesterol level has been lowered at $\alpha = 0.10$?
Assume that the variable is normally distributed.

<u>step 0</u> use the raw data (not shown here) to find the list of "before–after" Differences {20, 65, -2, 2, -1, 16}
<u>step 1</u> H_o: $\mu_d = 0$ & H_1: $\mu_d > 0$ [(a big 'before' #) minus (a small 'after' #) is positive]
<u>step 2</u> t = 1.61 by Stat TESTS #2 TTest… ('Data' option recommended)
<u>step 3</u> pvalue = 0.0852 by an "alt. T table" or 2nd DISTR #5 tcdf(1.61, 1000, 5) 0.0840
<u>step 4</u> There is moderate evidence that the mean cholesterol level did decrease
after taking the mineral supplement since $0.05 < 0.0852 < 0.10$.

(from Freund's "Modern Elementary Statistics", p. 318)
A random sample of 54 shirts worn by soldiers in a tropical climate has an average useful life of 63.9 washings with a standard deviation of 4.5 . Under non-tropical conditions, such shirts are known to have an average useful life of 81.6 washings. At the 0.01 level of significance, can we conclude that their use in a tropical climate reduces the average useful life of such shirts?

<u>step 1</u> H_o: $\mu = 81.6$ & H_1: $\mu < 81.6$
<u>step 2</u> t = -28.90 by Stat TESTS #2 TTest… ('Stat' option only)
<u>step 3</u> pvalue < 0.0045 by an "alt. T table (use df = 29)" (Ttest gives only zero pvalue)
or 2nd DISTR #5 tcdf(pos. 28.9, 1000, 53) 2.0×10^{-34}
<u>step 4</u> There is extremely tremendously humongously strong evidence that in a tropical climate,
the mean useful life of these shirts is significantly reduced from 81.6 washings.

(from Aliaga's "Interactive Statistics", p. 385)
In an experiment designed to test whether a subject has ESP (extrasensory perception), the subject is asked to guess the suit of each card drawn (spades, hearts, diamonds and clubs, in equal numbers, with replacement). If the subject gets 35 correct guesses out of 96 tries, then can we conclude this is higher than expected?

<u>step 1</u> H_o: $p = 1/4 = 0.25$ & H_1: $p > 0.25$ [note: $\hat{p} = 35/96$ 0.364583]
<u>step 2</u> z = 2.59 by Stat TESTS #5 1-PropZTest…
<u>step 3</u> pvalue 0.0048 by a printed table or 2nd DISTR #2 normalcdf(2.59, 1000)

Mark Harbison. 2002. MisterHarbison@aol.com (916) 558-2687. Sacramento City College.

step 4 There is very strong evidence that the subject has a higher than expected proportion of correct guesses. (But also be suspicious of lurking variables before concluding this a “proof” of ESP).

Mark Harbison. 2002. MisterHarbison@aol.com (916) 558-2687. Sacramento City College.