Estimation

*Please note: For this course, I'm content if you think "approximately 2", and simple use "1.96" in place of the precise t-statistic. Whenever you're using a computer-based statistical package, please **do** use the appropriate given t-statistic.*

1. A recent study on teenage alcohol use ("Alcohol Advertising in Magazines and Adolescent Readership" JAMA May 14, 2003) counted beer ads in magazines with young readership. In the 35 magazines studied, the mean number of beer advertisements was 34.3 and the standard deviation was 70.9. Construct a 95% confidence interval for the mean.

34.3 ± -T.INV(0.025,34)*70.9/SQRT(35), 34.3 ± 24.36, (The estimate potentially is off by quite a bit.)

2. According to the American Automobile Association (AAA), the average price of regular unleaded gasoline in the United States hit $3.971 on May 31, 2008 which was another in a string of consecutive all-time highs. This price is tallied from a national survey of "up to 100,000" service stations. Assuming a standard deviation of $0.328 and a sample size of 90,000, compute a 95% confidence interval for the true mean price of regular unleaded gasoline.

$3.971 ± -T.INV(0.025,89999)*$0.328/SQRT(90000), $3.971 ± $0.00214. (The sample size is large enough to provide a very accurate estimate.)

3. A survey for prospective job applicants[1] in economics found that academic job seekers were offered an average of 5.99 job interviews. The standard deviation was 6.62 and there were 211 job candidates in the study. Construct a 95% confidence interval for the mean number of interviews.

5.99 ± -T.INV(0.025,210)*6.62/SQRT(211), 5.99 ± 0.898.

4. The same survey from problem 4 reported the average starting salary offered to new economics Ph.D.s by 45 universities. The sample mean was $72,195 and the standard deviation was $11,596. a. Construct a 95% confidence interval for the mean starting salary

$72195 ± -T.INV(0.025,44)*$11596/SQRT(45), $72,195 ± $3484.

Sample Size

Based on national census data, the standard deviation of family income is believed to be $16,820. A researcher is interested in measuring the mean family income in a subset of the population that is typical of the national population. She plans to use a 95% confidence interval to estimate the true mean and she wants the margin of error to be $1,000, how many families should she include in her study?

She wants (roughly) $1.96 \cdot 16820 / \text{sqrt}(n) = 1000$, or $n \cong 1087$.

Hypothesis Testing

Please note that these problems can also be done using the "Hypothesis Testing Tool" in the course folder.

1. The marketing department at a big appliance maker wants to teach consumers about the energy efficiency of their products. Their top selling washing machine uses 29 kWh less energy than a typical washer. To gauge whether consumers are aware of the energy savings from the washer, they surveyed a random group of 114 people at an appliance retailer. The mean estimated energy savings was 20.5 kWh and the standard deviation was 37.8.

We're wondering if the mean consumer belief, μ, is truly less than the actual energy savings, so let's take as our null hypothesis that "$\mu \geq 29$", and see if the data contradicts this.

(20.5 – 29) / [37.8/sqrt(114)] = -2.401 (i.e., the sample mean is 2.401 standard deviations below the hypothesized mean). Using Excel, =T.DIST(-2.401,113,TRUE) gives us a (one-tailed) significance level of 0.90%. I'd interpret this as "extremely strong evidence against the null hypothesis, and consequently extremely strong evidence supporting the alternative, i.e., that consumers are indeed underestimating the energy savings available from our washing machine."

2. A large home builder wants to improve its cost estimation procedure. They take a sample of 57 recently completed projects and find that on average the actual costs were $1160 more than the estimated costs. The standard deviation of the sample is $4350.

a. Set up the appropriate null hypothesis for this test.
b. Does the data support a change in the procedure?

Define μ to be the true mean difference between actual and estimated costs. We'd want to change our cost estimation procedure if μ were different from 0, so we take "$\mu = 0$" as our null hypothesis.

(1160 – 0)/[4350/sqrt(57)] = 2.013 (i.e., the sample mean is 2.013 standard deviations above the hypothesized mean). Using Excel, =2* T.DIST(-2.01,56,TRUE) gives us a (two-tailed) significance level of 4.9%. I'd interpret this as "pretty strong evidence against the null hypothesis, and consequently pretty strong evidence supporting the alternative, i.e., that the mean actual-minus-estimated cost is non-zero and the procedure needs adjustment."

3. Television advertisers pay more to advertise to younger viewers whose tastes are generally more malleable than older viewers. The conventional wisdom believes viewers of "Big Wave Surfers" have an average age of 19.2 years. A sample of 219 random viewers has an average age of 21.1 years with a standard deviation of 17.7 years. Is the conventional wisdom correct?

A change in either direction would be of interest to advertisers, so let's take our null hypothesis to be "$\mu = 19.2$."

(21.1 – 19.2) / [17.7/sqrt(219)] = 1.589 (i.e., the sample mean is 1.589 standard deviations above the hypothesized mean). Using Excel,
=2*T.DIST(-1.589, 218,TRUE) gives us a (two-tailed) significance level of 11.4%. I'd interpret this as "a bit of evidence against the null hypothesis (i.e., against the conventional wisdom), but not even moderately strong evidence."